The Evolution of Sex Allocation and Hypodermic Insemination across the Flatworm Genus *Macrostomum*

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Jeremias Nepomuk Brand

aus Basel, 2020

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

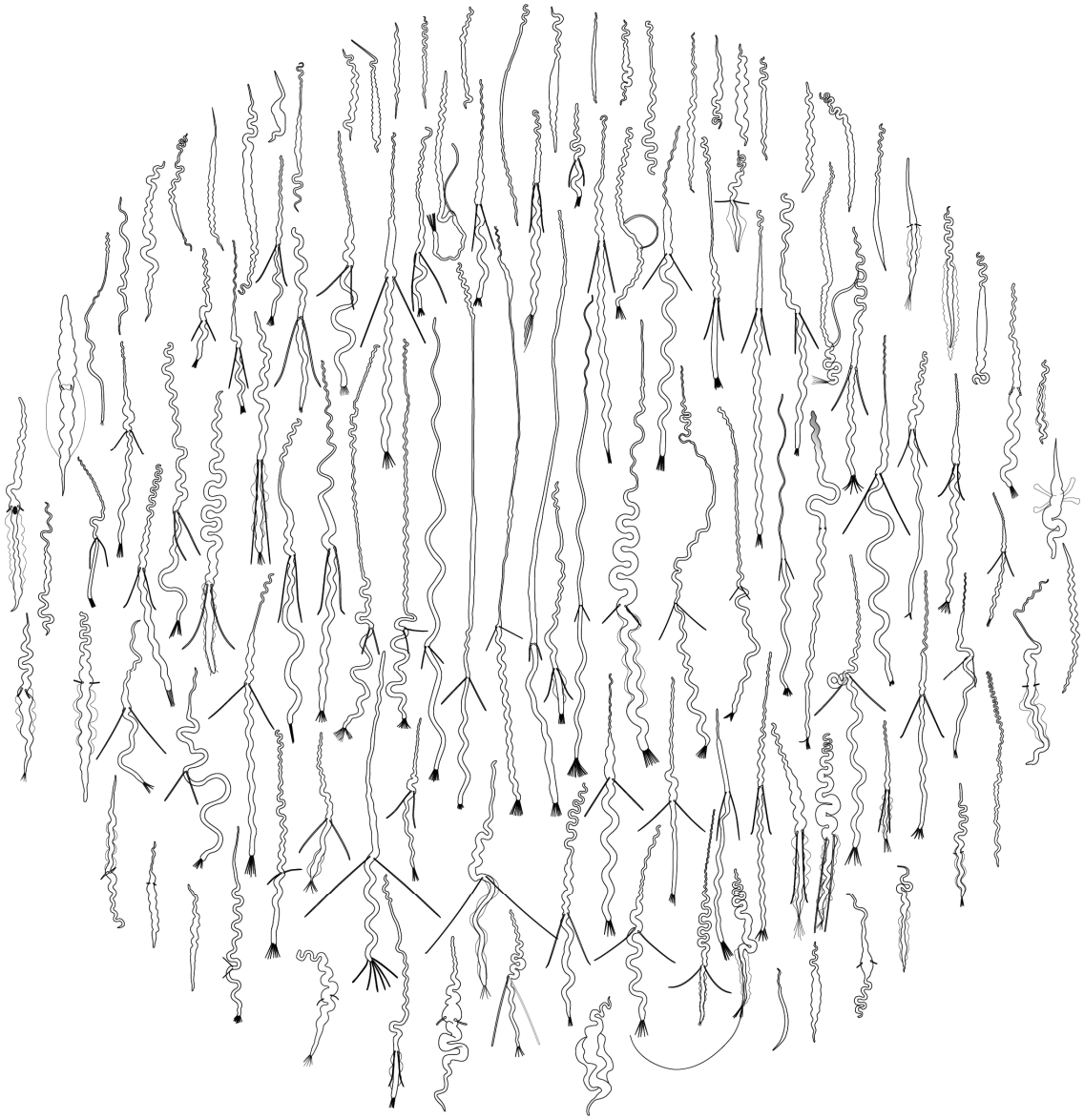
auf Antrag von:

Fakultätsverantwortlicher:	Prof. Dr. Dieter Ebert, Universität Basel
Dissertationsleiter:	PD Dr. Lukas Schärer, Universität Basel
Korreferent:	Prof. Dr. John Fitzpatrick, Universität Stockholm

Basel, den 23.06.2020

Prof. Dr. Martin Spiess
Dekan

The Evolution of Sex Allocation and Hypodermic Insemination across the Flatworm Genus *Macrostomum*



PhD Thesis - Jeremias N. Brand

Table of Contents

Abstract

Sexual selection in hermaphrodites is thought to result in conflict over mating roles, where individuals will be eager to act as the (sperm) donor but hesitant to act as the (sperm) recipient. Likely due to this conflict, some species engage in elaborate reciprocal mating behaviours such as egg trading or reciprocal copulation, often followed by intense postcopulatory conflict, to control the fate of the received ejaculates (e.g. sperm digestion or mechanical sperm removal). Besides reciprocal mating, another possible solution to the conflict about mating roles is hypodermic insemination, where the male copulatory organ is used to inject ejaculate directly into the tissue of the recipient. Such mating behaviour allows the donor to minimise the recipient's ability to control the fate of the received ejaculate, which potentially leads to a paternity benefit. Hypodermic insemination appears to be common in hermaphrodites, but it is unclear how frequently it evolves and through what routes it originates. In this thesis, I conducted a large-scale comparative analysis of the flatworm genus *Macrostomum,* to investigate the origins and the consequences of hypodermic insemination.

I have conducted field collections to expand both taxon-sampling and the geographic representation of the genus. By combining detailed morphological documentation and the first phylogenomic analysis of the genus, I documented 89 species that are new to science and present here three taxonomic descriptions that resulted from this work. I, therefore, showed that *Macrostomum* harbours large undiscovered biodiversity. Analysis of sperm and genital morphology, as well as the location of received sperm, shows that hypodermic insemination has evolved up to 13 times within the genus, thereby almost doubling the number of documented origins of this mating strategy across all hermaphrodites. These origins of hypodermic insemination are associated with consistent changes in the morphology of the male copulatory organ, the female sperm storage organ, and the sperm design. Such consistent correlations imply that these changes are adaptations to hypodermic insemination. I further show that hypodermic insemination likely evolved via initial internal wounding during copulation, leading to internal traumatic insemination and subsequently to the complete loss of copulation.

Since hypodermic insemination by-passes several processes that can decrease the level of sperm competition (e.g. cryptic female choice or sperm displacement), its evolution has been proposed to increase the proportion of reproductive resources allocated to sperm production. Contrary to this prediction, my work shows that in *Macrostomum*, hypodermic insemination is associated with reduced allocation towards sperm production. Most likely, this mating strategy is associated with the ability to self or other factors reducing the intensity of sperm competition. While no data on such a relationship exists in animals, this supports findings in plants, where selfing is associated with reduced investment into pollen production.

Finally, I here present evidence that reproduction-related genes evolve at an accelerated rate across the genus, as indicated both by sequence divergence and a decreased probability of identifying homologs with phylogenetic distance. This is the first documented case of such rapid evolution in hermaphrodites and supports a growing body of evidence that sexual selection can drive rapid gene evolution.

Chapter I
Thesis Introduction

Colorised scanning electron microscope images of the sperm of *Macrostomum spirale* (center) and *Macrostomum hystrix*. Images made by Martin Oeggerli (info@micronaut.ch).

Sexual selection and sexual conflict

"Sexual selection is the differences in reproduction that arise from variation among individuals in traits that affect success in competition over mates and fertilizations." (Andersson 1994). Sexual selection often takes the form of "Darwinian sex roles" (Parker 2014) with eager males competing for the opportunity to mate with choosy females. Note, however, that this is not universal since in some species the roles are reversed, and it is the males that are choosy (e.g. Maurer et al. 2011). The competition for mates can take the form of a direct, sometimes violent, contest between males or can play out through females choosing males based on, for example, their ornamentation, size or the resources they offer (Andersson 1994; Jennions and Kokko 2010; Hosken and House 2011). Besides these easily visible contests—given some level of promiscuity—sexual selection can continue after copulation within the female reproductive tract (Parker 1970; Charnov 1979; Eberhard 1996; Birkhead and Møller 1998; Birkhead and Pizzari 2002). Like precopulatory sexual selection, postcopulatory sexual selection can be classified into a competitive component, with sperm competition between sperm of unrelated males for access to the ova (Parker 1970, 1998; Simmons 2001; Birkhead and Pizzari 2002), and a choice component with some sperm being used preferentially for fertilisation by the female (i.e. cryptic female choice) (Charnov 1979; Eberhard 1996; Sirot and Wolfner 2015). While, differentiating between cryptic female choice and sperm competition is useful to generate testable predictions, it should be noted that disentangling them is challenging in practice since sperm competition (usually) takes place within the female genital tract and thus, just as cryptic female choice, involves the female to some degree (Eberhard 1996; Wedell and Hosken 2010).

Based on experiments with fruit flies, Bateman (1948) offered an explanation for the emergence of Darwinian sex roles. He observed that the reproductive success of male flies increased steeply with the number of mates, while the increase was less pronounced in female flies. The reproductive success of males thus depends mostly on the number of mates they can obtain, while female fecundity depends mostly on the amount of resources available for egg production (Bateman 1948). His observations were later summarised as Bateman's principles, which state that variance in both, reproductive success and mating success, should be higher in males and that the slope of a regression of reproductive and mating success—the Bateman gradient—should be larger in males (Charnov 1979; Arnold 1994). While some have cautioned against the overzealous use of Bateman gradients (Jennions et al. 2012), they have been examined extensively in gonochoristic (separate-sexed) animals, broadly confirming Bateman's observations (Janicke et al. 2016).

Except under very narrow conditions (i.e. lifetime monogamy, but see: Kokko and Jennions 2014), sexual reproduction will lead to a conflict of interests between the sexes (sexual conflict) (Arnqvist and Rowe 2005). Consequently, members of the

eager sex are expected to evolve traits to coerce members of the choosy sex to mate. Members of the choosy sex in turn will be selected to evolve resistance against coercion, potentially leading to antagonistic coevolution (Charnov 1979; Parker 1979; Holland and Rice 1998; Arnqvist and Rowe 2005). A well-studied example of sexual conflict leading to antagonistic coevolution occurs in water striders, where males and females engage in precopulatory struggles due to conflict over mating rate (Rowe et al. 1994). Males have exaggerated prolonged clasping genitalia and a flattened abdomen, allowing them to better attach to females during these struggles (Arnqvist 1989; Arnqvist and Rowe 2002b). Females, in contrast, exhibit morphological counteradaptations (e.g. longer abdominal spines and ventral orientation of the abdominal tip) that allow them to resist male attachment (Arnqvist and Rowe 1995, 2002b). Under antagonistic coevolution, we would expect a correlation between the size of the resistance and persistence traits, both within populations but also across species, on a macroevolutionary scale (Arnqvist and Rowe 2005). A comparative analysis of 15 species of water striders, has found such a macroevolutionary correlation, demonstrating that antagonistic coevolution can persist on these timescales (Arnqvist and Rowe 2002a).

Like sexual selection, sexual conflict does not necessarily end after mating, since males can evolve mechanisms to manipulate cryptic female choice and again females in turn can resist such manipulations (Parker 1979; Arnqvist and Rowe 2005; Wedell and Hosken 2010; Edward et al. 2015). For example, some transferred seminal fluids can have a profound effect on female behaviour, physiology and gene expression, and even influence life span and immune function (reviewed in Avila et al. 2010; Rodríguez-Martínez et al. 2011; Sirot et al. 2015). While most likely not all of these effects are due to sexual conflict, this certainly demonstrates that the intimate interactions between the sexes have implications even after mating (Parker 1979; Eberhard 1996; Sirot and Wolfner 2015). The interaction between the sexes, due to cryptic female choice and antagonistic coevolution, can lead to evolutionary chase dynamics similar to host-parasite coevolution (Holland and Rice 1998; Harrison et al. 2015; Wilkinson et al. 2015). Hence, we expect genes involved in these interactions to evolve rapidly, which has been shown in numerous taxa (reviewed in Swanson and Vacquier 2002; Wilburn and Swanson 2016).

Some authors have considered sexual conflict to be distinct from sexual selection and asked whether one or the other is superior in explaining some mating interactions (e.g. Shine et al. 2005). However, because sexual conflict is always present to some degree, it should be considered complementary to female choice and male-male competition (Hosken et al. 2009; Kokko and Jennions 2014).

Sexual selection in hermaphrodites

In a seminal paper, Charnov (1979) outlined that Bateman's insight can also be applied to hermaphrodites. In analogy to the argument in gonochorists, the male sex function of a hermaphrodite will be primarily limited by the number of matings it can obtain, while the female sex function will mostly be limited by the resources available for egg production (Charnov 1979). Unlike in gonochorists, Bateman's principle is, however, not only expected to lead to conflict over mating rate, but also to conflicts over mating role (Michiels 1998; Anthes 2010; Schärer et al. 2015). Most simultaneous hermaphrodites (hereafter simply called hermaphrodites) should mate more often for the opportunity to donate sperm rather than to receive it (Charnov 1979). While this logic is compelling, some have opposed it, arguing that, based on the Modern Portfolio theory, hermaphrodites should prefer to mate in the female role (Leonard 1999, 2005). According to this theory, mating as a female involves lower risk as fertilisation is assumed to be under the focal individual's control, while when mating as a male, fertilisation may be controlled by the mating partner (Leonard 1999, 2005). There is, however, little empirical support for a preference for the female role. Instead, most of the observed mating behaviour either indicates no strong preference or a preference for the male role (Michiels 1998; Anthes 2010; Schärer et al. 2015), which is also supported by estimates of Bateman gradients in hermaphrodites. Specifically, these studies have found the gradient of the male function to be significantly steeper compared to the gradient of the female function (plants: Johnson and Shaw 2016; Tonnabel et al. 2019; snails: Anthes et al. 2010; Pélissié et al. 2012), consistent with Charnov's prediction. However, except for one of these studies (Johnson and Shaw 2016), the gradients were estimated under tightly controlled conditions. Replication of one of the studies under *ad libitum* and restricted food conditions, only found significantly steeper Bateman gradients for the male function in the *ad libitum* treatment, while no significant Bateman gradients in any sex function were found under food restriction (Janicke et al. 2015). This indicates that environmental factors probably need to be considered to assess how reliably these estimates can be generalised to natural mating interactions (Janicke et al. 2015).

A general preference of hermaphrodites for one mating role will lead to sexual conflict because, while individuals will be eager to mate, they will not agree on who will donate and who will receive sperm (Charnov 1979; Michiels 1998; Michiels and Newman 1998; Anthes et al. 2006). Many hermaphroditic animals seem to resolve this conflict through conditional sperm receipt or reciprocity. For example, the black hamlet, *Hypoplectrus nigricans*, is thought to have a strong preference to mate in the male role (Leonard 1993) and to engage in egg trading behaviour (Fischer 1980). These reef fish are solitary and territorial during the day but will often leave their territory in the evening to migrate to the reef's edge for spawning. They always spawn in pairs, releasing their eggs in small parcels, usually alternating release and

fertilisation between the pair (Fischer 1980). While this behaviour appears to solve sexual conflict over mating role, by enforcing reciprocity ("conditional reciprocity", Leonard 2005), it could also simply occur due to random alignment of the egg-laying rhythms of the partners ("by-product reciprocity"). The actual conditionality of the exchange is difficult to show without experimental manipulation (Schärer et al. 2015). Nevertheless, observations of a marine polychaete, *Ophryotrocha diadema,* suggests that reciprocity in this egg trader is indeed conditional (Picchi et al. 2018).

Some copulating hermaphrodites seem to solve the conflict over mating role via reciprocal copulation, during which both partners donate and receive ejaculate in the same mating interaction (Michiels 1998; Schärer et al. 2015). Under these conditions, we expect sexual conflict to continue after mating, with (sperm) recipients exerting cryptic female choice to select favourable sperm or even remove all received sperm, thus effectively acting only as (sperm) donor. The receipt of unwanted ejaculate could be costly because it can increase, for example, the risk of polyspermy, lead to the transmission of sexually transmitted diseases or the receipt of manipulative seminal fluids (Schärer et al. 2015). Naturally, this can then lead to antagonistic coevolution where the donor will be selected to interfere with the recipient's choice. Due to the combination of individuals being eager to mate but preferring one mating role, it has been suggested that sexual selection in hermaphrodites is shifted towards the postcopulatory arena (Michiels 1998; Schärer and Pen 2013). Indeed, many hermaphrodites have intricate female genitalia with multiple chambers for sperm storage and organs for sperm digestions (Baur 1998; Michiels 1998; Nakadera and Koene 2013). For example, many pulmonate snails have specialised spermatophore receiving organs that can digest a received spermatophore (Nakadera and Koene 2013). Comparative analysis of spermatophore receiving organs and spermatophore size, across 51 snail species, has shown a positive correlation between them indicating coevolution (Koene and Schulenburg 2005). Similar evidence for coevolution has also been found in other hermaphroditic taxa, suggesting a general pattern among mating hermaphrodites (Beese et al. 2006; Anthes et al. 2008). While there are many intriguing examples of reciprocal mating in hermaphrodites, experimental manipulations remain rare and careful analysis will be needed to determine the precise underlying evolutionary dynamics (Michiels 1998; Schärer et al. 2015).

Traumatic mating

Another resolution to conflict over mating role is the unilateral enforcement of sperm donation. This can take the form of traumatic mating (Whitman 1891; Myers 1935; Apelt 1969; Angeloni 2003; Lange et al. 2013a), where animals will attempt to inseminate the partner but avoid being inseminated themselves (Charnov 1979; Michiels 1998; Schärer et al. 2011; Lange et al. 2013a). This possibly occurs in

polyclad flatworms which engage in a precopulatory behaviour called “penis fencing”. In these precopulatory bouts it appears that mating partners attempt to stab each other with their penises while simultaneously trying to avoid being stabbed themselves (Michiels and Newman 1998). Some animal species possess copulatory organs that damage the partner without the transfer of any substances (traumatic penetration, e.g. flour beetles: Crudgington and Siva-Jothy 2000 or some Drosophila: Kamimura 2007), possibly linked to an anchoring function. But traumatic mating can also involve the transfer of seminal fluids (traumatic secretion transfer). For example, during mating many pulmonate snails frequently stab their partner with a calcareous stylet, referred to as the “love dart”, which is coated in mucus gland products (Koene and Schulenburg 2005). The mucus can contain numerous substances, some of which have been shown to affect the physiology of the partner, influence their sex allocation, and increase paternity (Chase and Blanchard 2006; Nakadera et al. 2014; Lodi and Koene 2016; Stewart et al. 2016). Similarly, a species of Opisthobranch uses a secondary copulatory organ to inject secretions directly into the head of its partner, but it is currently unclear what effect these secretions have on the recipient (Lange et al. 2013b, 2014). Finally, traumatic mating can involve the transfer of ejaculate (traumatic insemination, transfer of sperm and seminal fluids). Traumatic insemination has been studied most thoroughly in bed bugs, which use their paramere to penetrate the external body wall of the female to inject sperm (Siva-Jothy 2006). Interestingly, females in some species of bed bugs have evolved a kind of secondary vagina that reduces the costs incurred due to traumatic insemination (Reinhardt et al. 2003; Siva-Jothy 2006).

Several hypotheses have been proposed to explain the evolution of these intriguing mating strategies (reviewed in Lange et al. 2013a; Reinhardt et al. 2015). Explanations based on sexual selection can be grouped into those that posit a direct benefit of the harm itself (Michiels 1998; Lessells 1999, 2005; Johnstone et al. 2000) and those that assume the harm constitutes a pleiotropic effect of some other benefit of traumatic mating (Parker 1979). Direct benefits could occur when the harm increases a recipient's mating latency or completely prevents them from remating, since this presumably results in more resources being allocated to the eggs fertilised by the donor (Johnstone et al. 2000). Harm could also stimulate short term fecundity of the recipient, if it is severe enough to trigger some kind of terminal investment (Michiels 1998; Lessells 1999, 2005). While direct benefits of harm have been frequently discussed, there is currently no good evidence in its support and instead it is more likely that harm is a by-product of other benefits (Morrow et al. 2003; Edvardsson and Tregenza 2005; Hotzy and Arnqvist 2009).

Traumatic mating could benefit donors because it can potentially allow the donor to overcome the partners pre- and postcopulatory choice by: (i) forcing copulation and (ii) by-passing the female genitalia of the recipient. Thus, traumatic mating can minimize the recipient’s ability to exert choice resulting in a paternity benefit

(Charnov 1979; Lange et al. 2013a; Tatarnic et al. 2014). As mentioned, in hermaphrodites, traumatic mating could also allow the enforcement of the preferred mating role. Indeed, in some species sperm transfer is unilateral and quick (Myers 1935; Apelt 1969; Borkott 1970) suggesting a "hit-and-run" mating strategy (Michiels 1998), but traumatic mating can also be reciprocal (Apelt 1969; Michiels and Newman 1998; Lange et al. 2014), suggesting enforcement of mating roles is not universally possible (Michiels 1998).

Traumatic mating could also be a result of natural selection since it may serve an anchoring function during mating in turbulent conditions (Lange et al. 2013a). The best candidates for an example of potential traumatic insemination due to natural selection can be found in Strepsiptera. The females of this insect order have a highly reduced neotenic morphology and are endoparasites of wasps (Kathirithamby et al. 2015). Traumatic mating has been suggested to allow faster mating, which could allow the male to avoid host aggression. However, it is unclear if traumatic mating in Strepsiptera is actually faster than normal copulation or if it has evolved because of the endoparasitic life-history of females (Kathirithamby et al. 2015).

Traumatic mating has evolved repeatedly across animals with at least 36 documented independent origins and it seems to occur more frequently in hermaphrodites (Lange et al. 2013a). Overrepresentation of this mating strategy could be a result of the above-mentioned sexual conflicts over the mating roles since traumatic mating could allow the enforcement of unilateral sperm transfer (Michiels 1998; Anthes 2010; Schärer et al. 2011, 2015). Additionally, hermaphrodites could be more likely to engage in harmful mating because negative consequences of mating in the female role can be compensated with benefits that the same individual receives from mating in the male role (Michiels and Koene 2006). Hermaphroditic taxa are thus good candidates for comparative work on traumatic mating since most other studied cases of traumatic mating represent only few independent origins of the behaviour (Lange et al. 2013a; Tatarnic et al. 2014).

Sex allocation in hermaphrodites

Since hermaphrodites unite both the male and the female functions (i. e. the production of sperm and eggs) within the same individual, they face the optimisation problem of allocating resources to both of them. These sex allocation decisions are expected to be profoundly impacted by the previously discussed dynamics of sexual selection and sexual conflict (Charnov 1982). Here I will briefly outline some of the most important factors thought to influence sex allocation in hermaphrodites.

Theoretical models predict hermaphroditism to be stable when at least one sex function shows a saturating fitness gain curve (Charnov et al. 1976; Charnov 1980; Fischer 1981). This means that investment into reproduction via that function at first gives high fitness returns, but then starts to yield lower returns per unit

invested, eventually favouring a reallocation of resources to the other sex function. It is usually assumed that the male fitness gain curve will be saturating, while the female curve, in accordance with Bateman's principle, will be linearly dependent on the resources allocated to egg production (Charnov 1979, 1982; Schärer 2009). The female gain curve is only expected to be saturating under some specialised conditions (e.g. when eggs are laid into a limited brood space, Heath 1979; Charnov 1982), while the male gain curve is expected to be saturating whenever there is competition between related sperm (termed local sperm competition, Schärer 2009). When related sperm compete, any additional resources invested in sperm production will mostly increase local sperm competition and lead to diminishing returns on the resources invested (Schärer 2009; Schärer and Pen 2013). To a large extent, the study of sex allocation in hermaphrodites is thus dedicated to the study of circumstances that lead to local sperm competition (Schärer 2009; Schärer and Pen 2013), but note that other factors such as body size can be relevant as well (reviewed in Schärer 2009).

High local sperm competition is expected in organisms with a high selfing rate, as sperm might often not compete with unrelated sperm and consequently selfing hermaphrodites are expected to have a strongly female-biased sex allocation (Charlesworth and Charlesworth 1981; Charnov 1982). This predicted relationship exists in plants where comparative work has shown an association between selfing rate and reduced investment into pollen (e.g. Charnov 1987; Barrett et al. 1996; Jürgens et al. 2002; Galloni et al. 2007; reviewed in Sicard and Lenhard 2011). There is less data on the effect of selfing in hermaphroditic animals, but the few studies that exist indicate the same relationship (Johnston et al. 1998; Winkler and Ramm 2018). Except under monogamy, there will be lower levels of local sperm competition in outcrossing hermaphrodites and its intensity will depend on the average number of donors that contribute to stored sperm in the recipients. This number can be influenced by the social group size, cryptic female choice, sperm displacement and random paternity skews (Figure 1; Charnov 1996; Schärer 2009; van Velzen et al. 2009; Schärer and Pen 2013). Especially, the influence of social group size has been evaluated in animals finding broad support for a relatively more male biased sex allocation in larger groups (e.g. Trouvé et al. 1999; Janicke et al. 2013). In plants, factors other than the selfing rate, such as pollination mode, the type of seed dispersal and plant size also influence sex allocation (reviewed in de Jong and Klinkhamer 2005). Furthermore, there is data on the shape of the fitness gain curves in some plants (Rademaker and de Jong 1998; Rademaker and Jong 1999; Campbell 2000; Elle and Meagher 2000), while these have only been evaluated in three animal species (Yund and McCartney 1994; McCartney 1997; Yund 1998).

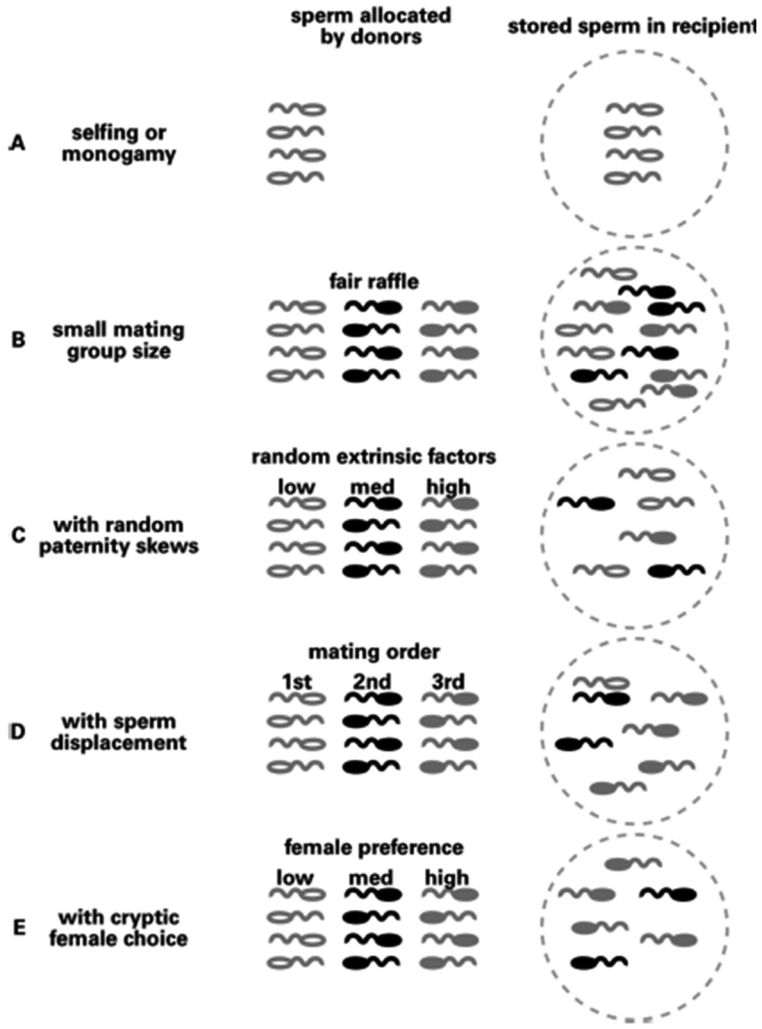


Figure 1. Conditions that are expected to lead to local sperm competition (LSC). (A) under monogamy or selfing only related sperm compete. (B) When all mating partners contribute equally to the stored sperm (fair-raffle) then the level of LSC is dependent on the number of partners contributing. A small number of partners will lead to LSC since a large proportion of stored sperm comes from the same individual. (C-E). Any process that leads to skews in how the received sperm will be represented in the female storage organ, will increase LSC since sperm from some partners will be underrepresented in storage. Figure from Schärer (2009).

The comparative method

The comparative method studies traits across species to gain insight into their form and function. Its application has a long-standing tradition going back as far as Aristotle (384–322 BC) and it is a mainstay of biological research. In the absence of any molecular data to support the theory of natural selection, Darwin made extensive use of the comparative method, by pointing to the abundance of homology across all of life (Darwin 1859). Darwin's discovery of common descent profoundly expanded the scope of comparative methods, giving them a historical dimension

because it led to the realisation that it should be possible to reconstruct the phylogenetic relationship of species. This eventually gave rise to phylogenetic comparative methods, which aim to statistically account for the common descent of species and infer the evolutionary history of traits as well as, the tempo and mode of their evolution (Harvey and Pagel 1991). These methods further aim to account for the pseudoreplication that results from the fact that closely related species are expected to be similar to each other simply due to common descent and therefore cannot be considered statistically independent samples (Felsenstein 1985; Harvey and Pagel 1991; Garamszegi 2014; Harmon 2019).

Felsenstein developed the first statistical comparative methods (1985), that were, due to the lack of accurate phylogenies, not widely used. Additionally, most available phylogenies at the time were based on cladistics (i.e. inferred using morphological traits), often making the applications of the method rather circular if one of the traits under study was used to infer the phylogeny. However, now with the broad availability of molecular data, the field of comparative biology is gaining increasing interest (Cooper et al. 2016) with the active development of new methods (Garamszegi 2014; Harmon 2019). Modern phylogenetic comparative methods are being applied throughout biology to study a wide array of topics such as speciation (Alfaro et al. 2009; Meredith et al. 2011; Louca et al. 2018), ecological adaptation (Losos 1998; Mahler et al. 2013; Cooney et al. 2017), development (Koyabu et al. 2011; Martín-Durán et al. 2018), metabolism (Uyeda et al. 2017) and sexual selection (Gage 1994; Pitnick et al. 1999; Stockley 2002; Fitzpatrick et al. 2009, 2012; Ramm and Stockley 2010).

The effectiveness of the comparative methods can be seen in studies on sperm-female coevolution in flies of the genus *Drosophila*. All *Drosophila* have two kinds of sperm storage organs: the paired spermathecae and the seminal receptacle (Sturtevant 1925). These structures are highly variable between species and were initially used for taxonomic purposes, as well as to infer phylogenetic relationships (Throckmorton 1966, 1975). In a comparative analysis including 113 species, Pitnick et al. (1999) showed that not all species use both organs for sperm storage. Instead, using ancestral state reconstruction, Pitnick et al. (1999) showed that the ancestor of the genus likely used both organs for storage but the use of the spermatheca has been lost convergently at least 13 times, while the loss of the use of the seminal receptacle has only occurred once. Further, Pitnick et al. (1999) showed that seminal receptacle length is correlated with sperm length, suggesting male-female coevolution (confirming previous findings that did not correct for phylogenetic relationship, Pitnick and Markow 1994). Finally, because seminal receptacle use has been lost convergently multiple times, Pitnick et al. (1999) were able to compare how strong the relationship between sperm length and seminal receptacle length is in species with, and without the use of the spermatheca. They found that the relationship differed between these two types of species, indicating that the loss of the

spermathecae modifies how sperm and the seminal receptacle coevolve. This example demonstrates that comparative methods can be particularly useful when we have multiple convergent events.

While phylogenetic comparative methods are powerful tools, one needs to keep some caveats in mind when applying them. First, they are necessarily correlational and therefore cannot demonstrate causation (Harvey and Pagel 1991; Garamszegi 2014). These methods parameterise natural history, which allows us to conduct exploratory studies of taxa and test predictions from theory. At the same time, we cannot exclude that an unobserved variable confounds our results. We therefore need to be cautious to not overinterpret results from such analyses and should, if possible, complement them with experiments. Second, although comparative methods aim to account for phylogenetic relationships, they, like all statistical approaches, can be prone to model misspecification or erroneous interpretation (Rabosky and Goldberg 2015; Cooper et al. 2016; Louca and Pennell 2020). For example, it has recently been pointed out that a commonly used tests for correlated evolution can reveal significant results when the traits in question have only evolved once, therefore, failing to correct for pseudoreplication (Maddison and FitzJohn 2015; Uyeda et al. 2018). Unfortunately, some of the traits that have drawn the interests of biologists have evolved only a few times. However, when nature is arranged so as to contain a multitude of convergent events, the prudent researcher should try to sample as many of them as possible.

The study organisms

For this thesis, I used the phylogenetic comparative method to investigate the evolution of sexual traits across the genus *Macrostomum* (Rhabditophora, Platyhelminthes). These free-living flatworms are hermaphrodites and possess several characteristics that make them ideal for investigations of the distribution and evolution of sex allocation and its interplay with traumatic mating.

All known *Macrostomum* species are small (0.5-3mm body length) and aquatic, occurring across the salinity spectrum. Marine species are part of the meiofauna, completing their life cycle in the interstitial spaces between sand grains, often on sheltered beaches. Freshwater species occur in ponds, lakes, and slow flowing rivers where they are mostly found attached to water plants or sometimes within the sediment. More than 200 species have been described, but the diversity is likely higher since the genus is understudied and taxonomic work has been geographically quite restricted. Furthermore, very little molecular information is currently available and sequence data on only exists for about two dozen species.

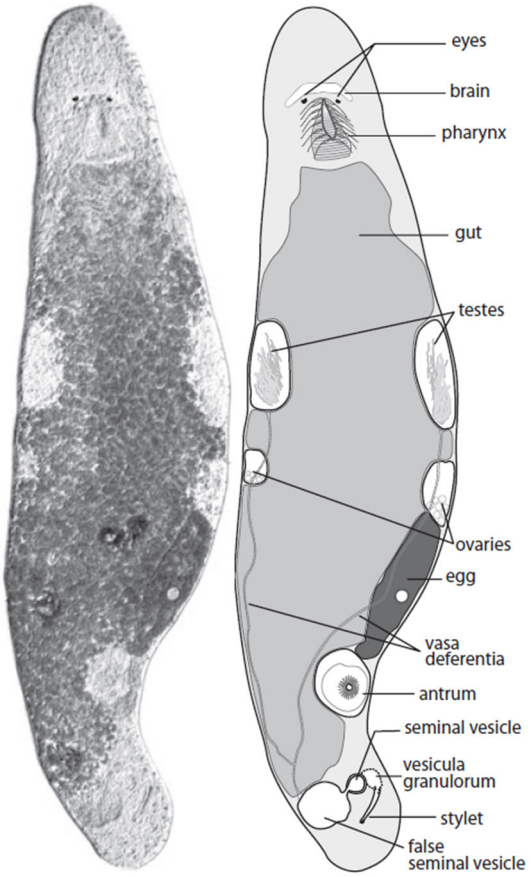


Figure 2 Micrograph and line drawing of a live adult specimen of *Macrostomum lignano* squeezed between two glass slides. The posterior part of the body is slightly twisted to better show the female antrum. The total length of this worm is approximately 1.8 mm. Figure from Vizoso et al. 2010.

Macrostomum flatworms are highly transparent allowing observation of their internal organs *in vivo* using a light microscope (Figure 2). They have paired testes that are connected via the vasa deferentia to the false seminal vesicle (absent in some species), which in turn is connected to a muscular true seminal vesicle. The true seminal vesicle is connected to the vesicula granulorum, which contains vesicles produced by neighbouring prostate gland cells. Fully formed sperm is stored in the seminal vesicle and during copulation passes through the vesicula granulorum into the male copulatory organ (stylet) before the ejaculate is then transferred to the partner. All species also have paired ovaries that contain the oocytes. Developing eggs exit the ovaries posteriorly and migrate towards the female sperm storage organ (female antrum). The mature eggs then enter the female antrum and are laid through a female genital opening. It is not known where fertilisation takes place, but presumably it occurs while the egg enters the female antrum since this is where sperm is deposited and stored, at least in the reciprocally copulating species (Ladurner et al. 2005; Vizoso et al. 2010).

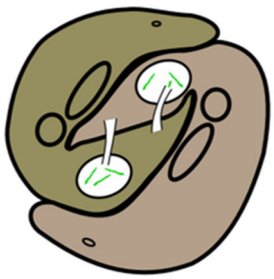


Figure 3 Stylised drawin of reciprocal copulation in *M. lignano*. Adapted from a drawing by D. Vizoso.

Macrostomum species can be assigned to two distinct mating syndromes, consisting of characteristic combinations of morphology and behaviour: Species with the **reciprocal mating syndrome** engage in a copulatory handshake where both partners insert their stylet into the female antrum of the partner (Figure 3). This type of mating can allow both partners to act as the sperm donor and the sperm recipient at the same time. If, as outlined above, individuals usually mate to donate rather than receive sperm, this mating

syndrome appears to be a mutualistic resolution to the sexual conflict over the mating role. Both partners allow sperm receipt in exchange for the opportunity to donate themselves (Michiels 1998; Vizoso et al. 2010; Schärer et al. 2011, 2015). However, after copulation many of these species perform the so-called "suck" behaviour, in which they place their mouth on top of their female genital opening, apparently in an attempt to remove the ejaculate they have just received (Schärer et al. 2004, 2011, 2020; Vizoso et al. 2010). This behaviour could constitute a female resistance trait involved in antagonistic coevolution between sperm donor and sperm recipient (Vizoso et al. 2010). Under this sexual conflict hypothesis, we expect persistence traits of the sperm donor to evolve and indeed there are good candidates for such traits in *Macrostomum*. Species within the reciprocal mating syndrome have complex sperm with an elongated anterior feeler and stiff lateral bristles (Figure 4A). These structures could be adaptations to counteract the suck behaviour and if this were the case, they would constitute male persistence traits (Vizoso et al. 2010; Schärer et al. 2011). Further, persistence traits do not necessarily need to be morphological in nature, since the ejaculate of one species (*M. lignano*) has recently been shown to contain products that manipulate the partners propensity to perform the suck behaviour (Patlar et al. 2020; Weber et al. 2020), as was proposed previously (Marie-Orleach et al. 2013).

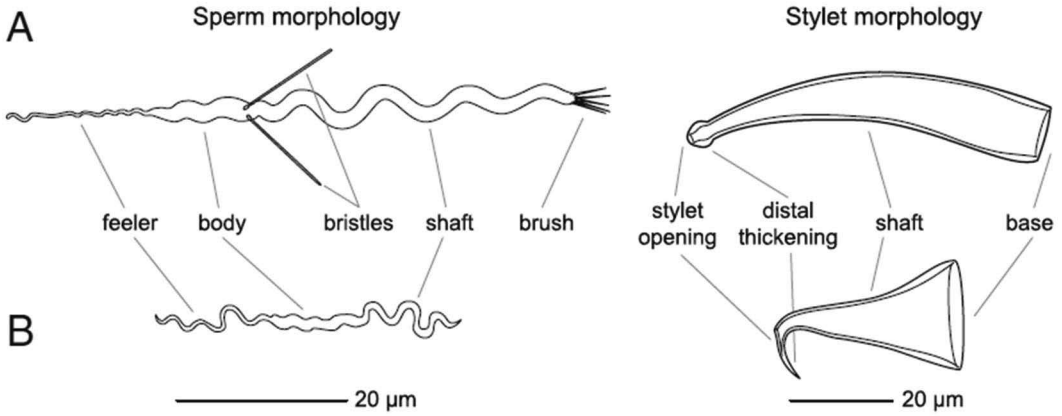


Figure 4 Morphology of the sperm and stylet of two *Macrostomum* species. (A) *M. lignano*, a species that represents the reciprocal mating syndrome. (B) *M. hystrix*, a species that represents the hypodermic mating syndrome (Schärer et. al. 2011).

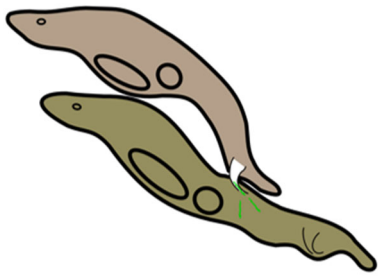


Figure 5 Stylised drawing of hypodermic insemination in *M. hystrix*. Adapted from a drawing by D. Vizoso.

Macrostomum species with the **hypodermic mating syndrome** mate via hypodermic insemination, a form of traumatic insemination during which a needle-like copulatory organ (Figure 4B) is used to inject ejaculate into the parenchyma of the partner (Figure 5). Due to the high transparency of *Macrostomum* it is possible to observe injected sperm within the tissue of these animals (Schärer et al. 2011; Ramm et al. 2012; Giannakara and Ramm 2017), which makes them ideal candidates for the study of traumatic insemination. Unlike species of the reciprocal mating syndrome, the sperm of species of the hypodermic mating syndrome is simple, lacking a brush and in most species also the bristles, while the bristles are greatly reduced in some other species (Figure 4B, Schärer et al. 2011). Further, the female antrum of species with the hypodermic mating syndrome is greatly simplified, while the female antrum of species with the reciprocal mating syndrome is thickened. This is most likely because in reciprocally mating species, the female antrum interacts with the stylet and sperm of the mating partner, while it is only used for egg laying in hypodermically mating species (Schärer et al. 2011).

A previous comparative analysis has revealed that the genus *Macrostomum* consists of two phylogenetically well-separated clades, of which one clade is thought to only contain species that exclusively mate through hypodermic insemination (Figure 6, Clade 1) and a second clade primarily contains reciprocally mating species (Figure 6, Clade 2). However, within the latter clade, hypodermic insemination has convergently evolved in *M. hystrix* (Schärer et al. 2011). Interestingly, while hypodermic insemination in *Macrostomum* likely evolves due to sexual conflict over mating roles, it is also associated with the ability to self-fertilise (Ramm et al. 2012; Giannakara and Ramm 2017). Selfing has been documented both in several species from Clade 1 as well as in *M. hystrix*, suggesting an evolutionary correlation of these traits (Ramm et al. 2012; Giannakara and Ramm 2017; Singh et al. 2020). Comparative investigations of hypodermic insemination in *Macrostomum* should thus ideally attempt to assess if further convergent origins of it are also linked to selfing.

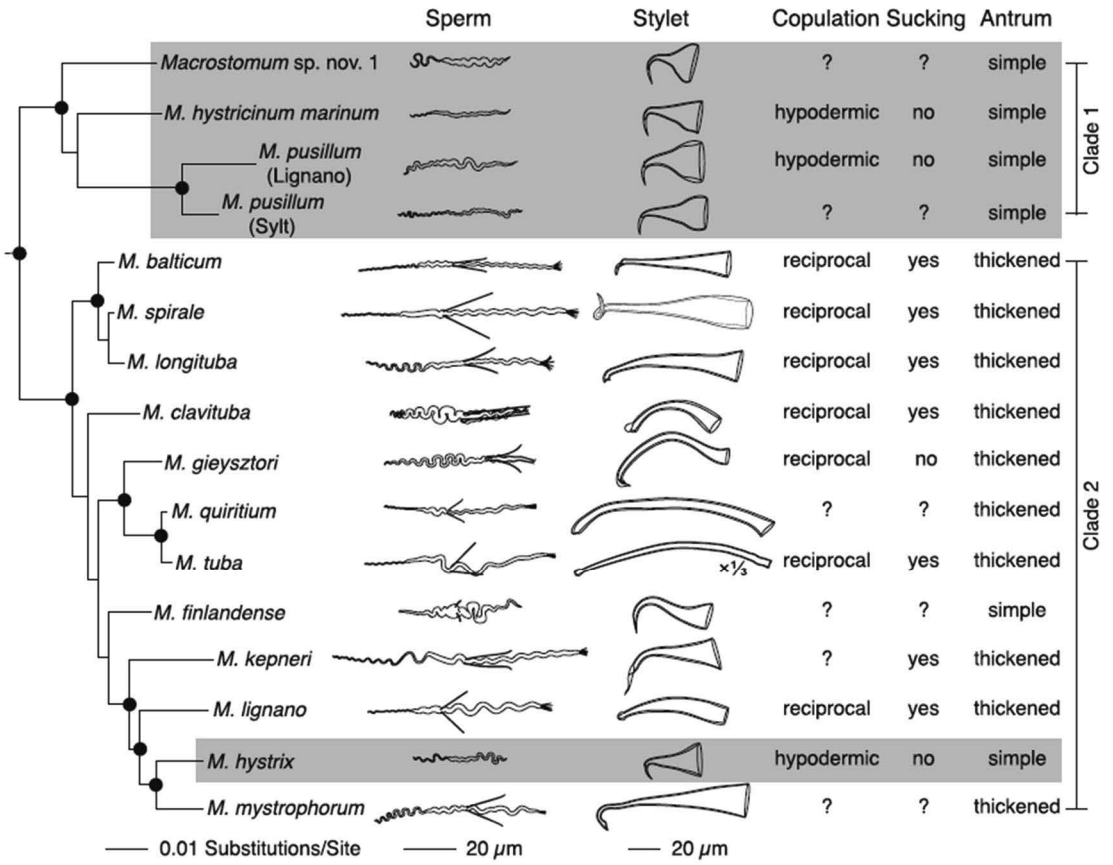


Figure 6 Phylogeny of 16 *Macrostomum* species and their morphological and behavioural traits. Nodes marked with circles had high support in the phylogenetic analysis of Schärer et al. (2011). Drawings illustrate the sperm and stylet diversity across the genus. Note, the similarity between the sperm and stylet morphology of *M. hystrix* and members of Clade 1, which is characteristic of the hypodermic mating syndrome. Given is also the type of copulatory behaviour, whether the suck behaviour has been observed, and the state of the female antrum. Figure from Schärer et al. (2011).

Thesis outline

In my thesis I made extensive use of the comparative method to study the evolution of sexual traits across the genus *Macrostomum*. My main work consisted in the detailed analysis of the rate of evolution of reproduction-related genes (Chapter II), investigations of the evolution of hypodermic insemination (Chapter III) and its correlations with sex allocation (Chapter IV). Finally, I present three species descriptions that result from our extensive field work (Chapter V).

Sexual selection is expected to drive rapid divergence of genes involved in reproduction. This has been shown in numerous taxa (Swanson and Vacquier 2002; Wilburn and Swanson 2016), but so far it has not been investigated in any flatworms. Because *M. lignano* is an established model organism that has been used to study a broad range of topics such as ageing (Mouton et al. 2009, 2018), stem cell biology (Grudniewska et al. 2016), bioadhesion (Lengerer et al. 2014, 2016; Wunderer et al. 2019), karyology (Zadesenets et al. 2016), and sexual selection (Janicke et al. 2013; Sekii et al. 2013; Marie-Orleach et al. 2016), there are numerous gene annotations available that are based on experimental and gene expression studies. Based on these annotations, it is possible to identify reproduction-related genes. In **chapter II** of my thesis, I transferred gene annotations from *M. lignano* to newly generated transcriptomes of three additional *Macrostomum* species. Specifically, I included representatives of both independent origins of hypodermic insemination, *M. pusillum* and *M. hystrix*, as well as a second species with reciprocal copulation (*M. spirale,* Figure 5). I then inferred robust orthologs between these transcriptomes, which in combination with RNA-Seq data from different life stages, allowed me to detect putative reproduction-related genes. I then determined whether these genes evolve at an accelerated rate within the genus.

Two independent origins of hypodermic insemination have been documented in *Macrostomum* (Schärer et al. 2011) and based on theory we expect such mating behaviour to evolve frequently in hermaphrodites (Charnov 1979; Michiels and Koene 2006; Schärer et al. 2015). Additionally, there is a species, *M. finnlandense*, within Clade 2 that has very short sperm bristles and a needle-like stylet. Schärer et al. (2011) did not have enough information to assign *M. finlandense* to a mating syndrome, but it seems plausible that it also mates via hypodermic insemination. If true this means that in the relatively small taxon sample analysed so far, hypodermic insemination has evolved three time independently. This suggests that many convergent events in the genus are not yet discovered. For **chapter III**, I conducted in total more than nine months of field work across four continents, to increase both the taxon sampling and the geographic representation of the genus. I collected extensive morphological and molecular information on 145 species to investigate how often hypodermic insemination has evolved in *Macrostomum* and what morphological changes accompany its evolution. To achieve this aim, I

conducted the first phylogenomic analysis of the genus, generating de novo transcriptomes for approx. 100 species, resulting in a robust phylogeny that allowed me to reconstruct the evolutionary history of hypodermic insemination and a plethora of other sexual traits.

Sex allocation is a trait of high importance in the study of sexual selection in hermaphrodites, but while it has been extensively studied in a comparative framework in plants, such data is sparse for hermaphroditic animals (Schärer 2009). The frequent origin of the hypodermic mating syndrome in *Macrostomum* is particularly interesting because hypodermic insemination could affect the mode and intensity of sperm competition and should influence the optimal sex allocation of species (Schärer and Janicke 2009; Schärer et al. 2011). In *Macrostomum*, sex allocation can be estimated by measuring gonad size of slightly squeezed animals and this technique has been applied in several species (Schärer and Ladurner 2003; Janicke et al. 2013; Giannakara and Ramm 2017; Winkler and Ramm 2018; Singh et al. 2019b). Sex allocation has been studied extensively in *M. lignano*, where individuals increase the relative allocation to testis in larger social groups, presumably because local sperm competition is less intense under these conditions (Janicke et al. 2013). The extent of plasticity is variable across the genus, with some species showing no plasticity, while others only respond to the availability of a mate but not the intensity of local sperm competition (Giannakara and Ramm 2017; Singh et al. 2019b). For **chapter IV**, I estimated the sex allocation of 120 field-collected *Macrostomum* species allowing me to apply statistical techniques to infer the tempo and mode of its evolution. Further, I combined these results with the morphological measures collected for chapter III to determine the morphological correlates of sex allocation and test if hypodermic insemination indeed influences it.

There is need for a new *Macrostomum* model species, because the current model organism, *M. lignano,* has been shown to harbour karyotype polymorphisms (Zadesenets et al. 2016). This represents a considerable roadblock for detailed investigations of the genomics of sexual traits using forward and reverse genetic approaches. Collecting species for chapter III and chapter IV allowed us to search for species that would be a suitable replacement for *M. lignano*. During our field work, we evaluated species based on whether they could be maintained in a laboratory setting and we investigated the karyology and genome size of the most promising candidates. In **chapter V,** we present the three taxonomic species descriptions that resulted from these efforts with one species, *Macrostomum cliftonensis*, representing the most promising candidate.

I illustrate the morphological diversity of the genus *Macrostomum*, by displaying the stylet and sperm of all the species included in this thesis in the margins of Chapter II-IV. The drawings are to scale between pages (scalebars below are 20µm); all drawings can also be found in Figure 2 in Chapter III.

Chapter II

RNA-Seq of three free-living flatworm species suggests rapid evolution of reproduction-related genes

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

Abstract

Background

The genus *Macrostomum* consists of small free-living flatworms and contains *Macrostomum lignano*, which has been used in investigations of ageing, stem cell biology, bioadhesion, karyology, and sexual selection in hermaphrodites. Two types of mating behaviour occur within this genus. Some species, including *M. lignano*, mate via reciprocal copulation, where, in a single mating, both partners insert their male copulatory organ into the female storage organ and simultaneously donate and receive sperm. Other species mate via hypodermic insemination, where worms use a needle-like copulatory organ to inject sperm into the tissue of the partner. These contrasting mating behaviours are associated with striking differences in sperm and copulatory organ morphology. Here we expand the genomic resources within the genus to representatives of both behaviour types and investigate whether genes vary in their rate of evolution depending on their putative function.

Results

We present *de novo* assembled transcriptomes of three *Macrostomum* species, namely *M. hystrix*, a close relative of *M. lignano* that mates via hypodermic insemination, *M. spirale*, a more distantly related species that mates via reciprocal copulation, and finally *M. pusillum*, which represents a clade that is only distantly related to the other three species and also mates via hypodermic insemination. We infer 23,764 sets of homologous genes and annotate them using experimental evidence from *M. lignano*. Across the genus, we identify 521 gene families with conserved patterns of differential expression between juvenile vs. adult worms and 185 gene families with a putative expression in the testes that are restricted to the two reciprocally mating species. Further, we show that homologs of putative reproduction-related genes have a higher protein divergence across the four species than genes lacking such annotations and that they are more difficult to identify across the four species, indicating that these genes evolve more rapidly, while genes involved in neoblast function are more conserved.

Conclusions

This study improves the genus *Macrostomum* as a model system, by providing resources for the targeted investigation of gene function in a broad range of species. And we, for the first time, show that reproduction-related genes evolve at an accelerated rate in flatworms.

Background

The genus *Macrostomum* (Platyhelminthes, Macrostomorpha) consists of small free-living flatworms and contains the model organism *Macrostomum lignano,* which has been used in numerous studies investigating a broad range of topics, ranging from sexual selection in hermaphrodites [1–3], ageing [4, 5] and stem cell biology [6], to bioadhesion [7–9] and karyology [10]. To enable this research many state-of-the-art tools have been established, such as an annotated genome and transcriptome [11, 12], efficient transgenesis [12], *in situ* hybridisation (ISH) [7, 13], and gene knock-down through RNA interference (RNAi) [3, 14]. The wealth and breadth of research on *M. lignano* make this species unique among the microturbellarians, for which research is generally restricted to taxonomic and morphological investigations.

Given the success of using *M. lignano* as a model system, it is now desirable to produce genomic resources for more species within the genus to test if insights gained in *M. lignano* can be generalised. This is especially relevant since two contrasting types of mating behaviour occur within this genus [15]. Some species, including *M. lignano* (Fig. 1), show the reciprocal mating syndrome. They mate via reciprocal copulation, where, in a single mating, both partners insert their male copulatory organ (the stylet) into the female sperm storage organ (the antrum), and simultaneously donate and receive sperm [15]. In addition, these reciprocally mating species possess stiff lateral bristles on their sperm, which are thought to be a male persistence trait to prevent the removal of received sperm [16]. Sperm removal likely occurs since, after copulation, worms of these species are frequently observed to place their pharynx over their female genital opening and then appear to be sucking, most likely removing seminal fluids and/or sperm from the antrum [17]. The sperm bristles could thus anchor the sperm in the epithelium of the antrum during this post-copulatory suck behaviour [16]. Other species within the genus, such as *M. hystrix*, show the hypodermic mating syndrome (Fig. 1). They mate via hypodermic insemination, where worms use a needle-like stylet to inject sperm into the tissue of the partner and the sperm then move through the tissue to the site of fertilisation [15, 18, 19]. Sperm of hypodermically mating species lack bristles entirely [15]. As a consequence of these contrasting mating behaviours there likely are differences in the function of reproduction-related genes between reciprocally and hypodermically mating species. Genomic resources of species with contrasting mating syndromes could, therefore, be used to identify these genes and investigate their function.

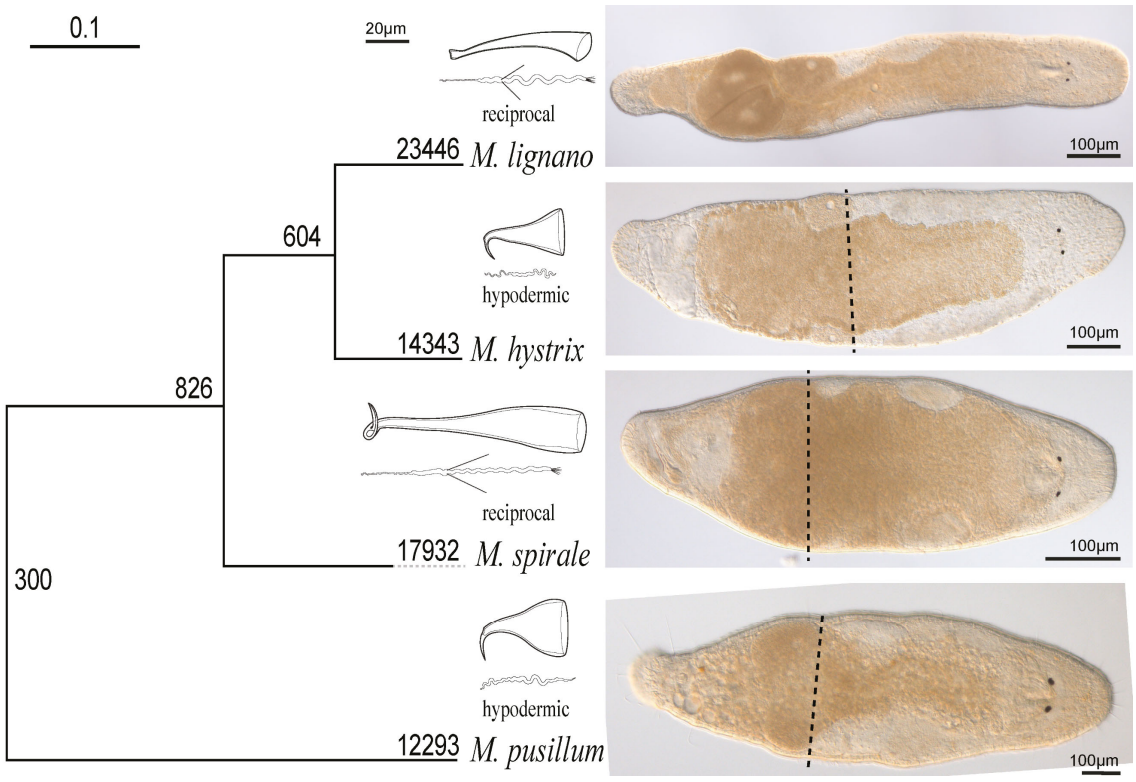


Fig. 1 Details of the phylogenetic relationships and the morphology of the species in this study. Phylogeny of the four species (left) next to line drawings of the male copulatory organs (stylets) and sperm, and light microscopic images of lightly squeezed live worms. The type of mating (reciprocal/hypodermic) is indicated above the species name. The phylogeny (see also Results) is rooted at the branch leading to *M. pusillum* since this represents the deepest split in the genus (see Janssen et al. 2015). The grouping of *M. lignano* with *M. hystrix* has maximal support (in both the ultrafast bootstrap as well as the Shimodaira–Hasegawa–like approximate likelihood ratio test), which suggests independent origins of the hypodermic mating syndrome in *M. hystrix* and *M. pusillum*. The scale bar represents substitutions per site, and the numbers next to the nodes give the number of gene duplications that occurred according to the OrthoFinder analysis (see also Methods; the amino acid alignment, the inferred phylogeny and the log file of the IQ-TREE analysis are provided in Additional file 1: "Amino acid alignment of one-to-one orthologs**"**; Additional file 2: "Maximum likelihood phylogeny" and Additional file 3: "IQ-TREE logfile"). The stippled lines on the light microscopic images show the intended cutting level for the regenerant treatment (see also Methods).

A range of empirical gene annotations derived from RNA-Seq experiments in *M. lignano* are available, with candidate gene sets that are differentially expressed (DE) between body regions [20], stages of tissue regeneration [21], social environments [22], animals of different ages [5], and between somatic cells and somatic stem cells (called neoblasts in flatworms) [6]. Identifying the homologs of genes with such empirical annotations in other *Macrostomum* species will allow us to investigate their function and rate of evolution in a broader phylogenetic context. For example, it can be assessed whether genes identified as being involved in neoblast function are conserved, and this may identify genes that are particularly important in flatworm regeneration.

Moreover, insights into the biology of these species can be gained by identifying rapidly evolving genes, since there is evidence that in a range of organismal groups reproduction-related genes evolve faster than genes serving other functions (reviewed in [23, 24]). Among the fastest-evolving genes are those encoding for proteins directly involved in molecular interaction with the mating partner, such as pheromone receptors (e.g. [25]), seminal fluid proteins (e.g. [26]), and proteins involved in gamete recognition and fusion (e.g. [27]). Groups of genes with biased expression in reproduction-related tissues, such as the testis and ovary, can also show elevated rates of evolution. Evidence for this comes both from sequence-based analysis of the rate of divergence and the increased difficulty of detecting homologs of reproduction-related genes[28, 29].

Here we present transcriptomes and differential expression (DE) datasets of three *Macrostomum* species (Fig. 1), namely i) *M. hystrix*, a close relative of *M. lignano* that mates via hypodermic insemination, ii) *M. spirale*, a somewhat more distantly related species that, like *M. lignano*, mates via reciprocal copulation, and finally iii) *M. pusillum*, which represents a clade that is deeply split from the other three species and which also mates via hypodermic insemination (see also [15] and [30] for the broader phylogenetic context). All three species are routinely kept in the laboratory and studies have been published using cultures of *M. hystrix* [10, 18, 19, 31], *M. pusillum* [32], and *M. spirale* [10]. Since the comparison to *M. pusillum* represents one of the largest genetic distances within the genus, it is an ideal choice to identify genes that are either conserved or evolve rapidly. The inclusion of two species with hypodermic insemination furthermore allows candidate selection for genes involved in determining differences in sperm morphology.

In all three species, we produced RNA-Seq libraries for adults (A), hatchlings (H), and regenerants (R), in order to capture the expression of as many genes as possible and to allow for DE analyses between these biological conditions (Fig. 2A, red labels). Since hatchlings lack sexual organs, genes with higher expression in adults compared to hatchlings can serve as candidate genes that are specific for those organs. Conversely, genes with higher expression in hatchlings are candidates for genes regulating early development. Finally comparing gene expression in adults vs. regenerants can identify regeneration-related candidate genes involved in the development of structures that are not actively forming in the adult steady state, such as the male genitalia (as demonstrated in [21]). Besides conducting the described DE analysis, we also determined groups of homologous genes (called orthogroups [OGs] throughout the text) between the three species presented here and *M. lignano* (Fig. 2). This allowed us to transfer the empirical annotations from three RNA-Seq experiments performed in *M. lignano* (Fig. 2B-D, red labels) to these inferred OGs and investigate whether OGs with particular annotations show signs of conservation or rapid evolution in patterns of protein sequence divergence and/or gene presence/absence.

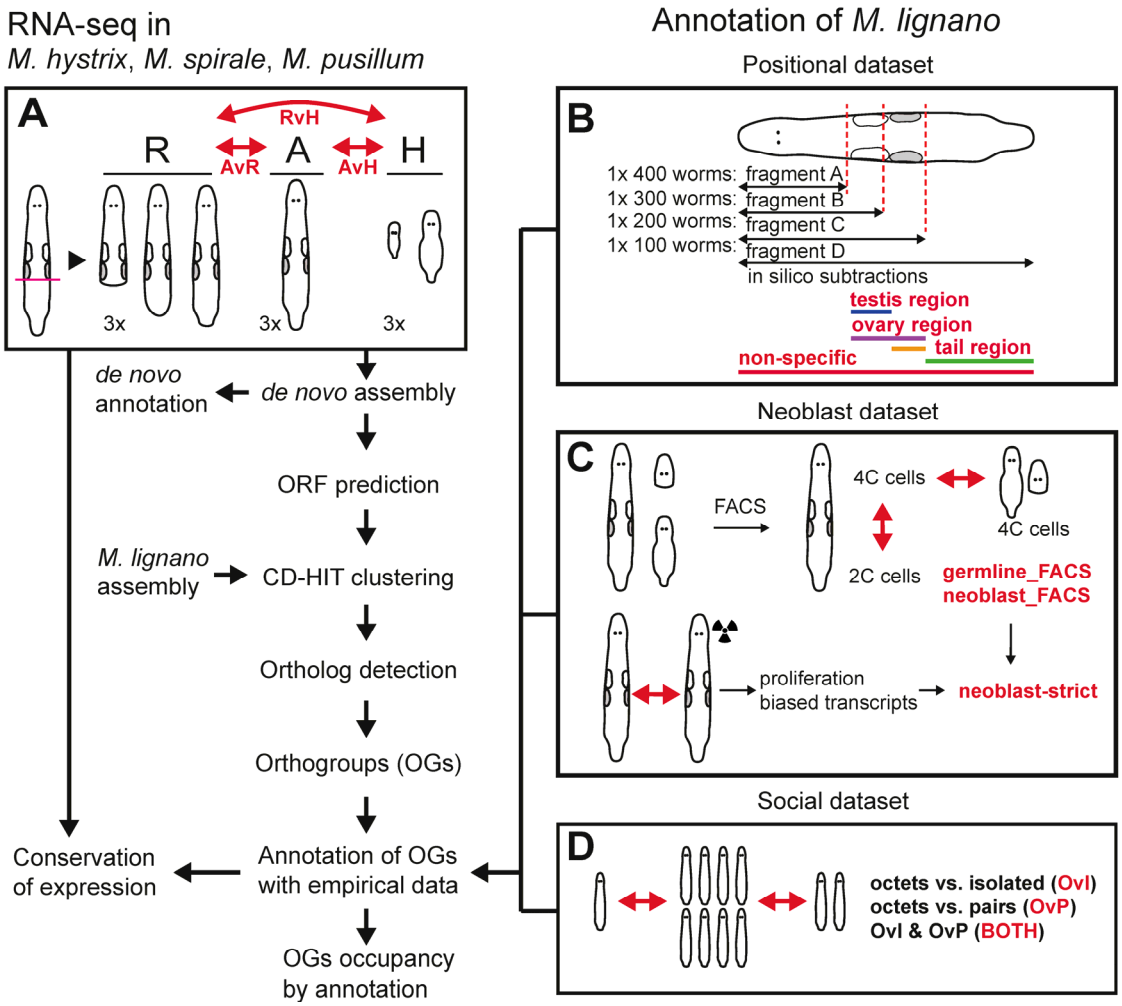
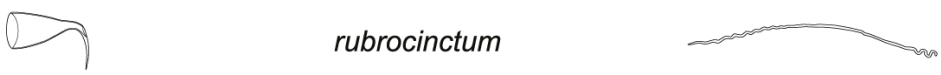


Fig. 2 Flowchart of the analysis steps in the manuscript. The red double arrows indicate DE analyses and red labels the resulting DE annotations. **A:** Details of the experiment conducted for this study (yielding three DE annotations: AvH, AvR, and RvH). **B:** Details on the positional dataset of Arbore et al. (2015). The stippled red lines on the schematic drawing of the worm indicate the levels at which worms were amputated to produce the four fragments indicated below. These fragments were then used to identify genes that were DE in the body regions shown in colour (yielding four DE annotations: non-specific, testis region, ovary region and tail region). **C:** Details on the dataset of Grudniewska et al. (2016). The top row shows the identification of candidates using FACS and the bottom row the approach using irradiation to remove proliferating cells, permitting the annotation of transcripts with germline- and neoblast-biased expression (yielding three DE annotations: germline_FACS, neoblast_FACS, and neoblast-strict). **D:** Details of the social dataset of Ramm et al. (2019). Comparisons between worms grown in different social group sizes permit identifying socially-sensitive transcripts (yielding three DE annotations: OvI, OvP, and BOTH).



Results

Transcriptome assembly and quality

We used >300 million paired-end reads per species—derived from adults (A), hatchlings (H), and regenerants (R)—to assemble the transcriptomes of *M. hystrix*, *M. spirale*, and *M. pusillum* (Tab. 1). All three transcriptomes were fairly complete in gene content when assessed using BUSCO, with more than 92.5% of all 978 core metazoan genes found either complete or as fragments in all species (Tab. 1). Moreover, the assemblies were a good representation of the reads used to infer them, with >87% and >46% of the reads mapping back to the raw and the (CD-HIT) reduced assembly, respectively (Tab. 2). TransRate scores were between 0.28 and 0.29 (Tab. 1), placing them above average when compared to 155 publicly available transcriptomes evaluated in [33] (which ranged from 0 to 0.52, with an average of 0.22). The *M. spirale* transcriptome contained almost twice as many transcripts as the other two, but although *M. spirale* had the highest absolute number of functional annotations (Tab. 1), it had the lowest percentage of transcripts with annotations. The *M. spirale* assembly could thus contain more redundant sequences, contain more poorly assembled contigs due to increased heterozygosity or contain more non-coding transcripts than the others (see Discussion).

Tab. 1. Transcriptome assembly statistics per species. The initial number of reads used, the number of reads after Trimmomatic processing, the number of initially assembled transcripts, the empirical mean insert size of the RNA-Seq libraries, the number of distinct 21-mers, the number of transcripts removed by CroCo, and the final number of transcripts, as well as the mean transcript length and number of bases in the final assemblies are shown. The BUSCO score is given as the percentage of complete (C) genes—divided into present as single copies (S) or duplicates (D)—and fragmented (F) genes of the 978 gene Metazoa set. The next three rows detail the TransRate score, the number of transcripts remaining after TransDecoder translation and CD-HIT clustering, and the number of transcripts considered in the DE analysis. Below this a summary of the results from the Trinotate annotations giving the number of transcripts (and the corresponding percentage of the whole transcriptome in brackets) with a given annotation: ORF, contains a predicted open reading frame; BLASTX, the predicted ORF and/or the entire transcript produced a hit in the Protein database; Pfam, a protein family domain was found; SignalP, a signal peptide was detected; TMHMM, a transmembrane helix is predicted.

Assembly statistics	*M. hystrix*	*M. spirale*	*M. pusillum*
Initial reads	320,462,680	347,532,862	315,510,916
Reads post trimming	297,398,416	320,497,034	295,230,930
Mean insert size	146	143	145
Distinct 21-mers	160,907,099	235,628,648	194,772,389
Assembled transcripts	169,758	296,658	177,453
Removed transcripts	217	156	274
Final transcripts	169,541	296,502	177,179
Mean transcript length	1,094	764	756
Number of bases	185,792,353	226,578,146	134,085,334
BUSCO score (Metazoa gene set)	C: 90.1 S: 49.3 D: 40.8 F: 3.4	C: 87.8 S: 37.3 D: 50.5 F: 4.7	C: 89.2 S: 55.8 D: 33.4 F:4.1
TransRate score	0.28	0.29	0.28
CD-HIT transcripts	53,132	74,135	53,416
DESeq2 transcripts	43,126	66,139	41,418
Annotation			
ORF	59,889 (35.3)	70,808 (23.9)	49,456 (27.9)
BLASTX	47,837 (28.2)	50,033 (16.9)	42,940 (24.2)
Pfam	42,330 (25.0)	43,840 (14.8)	34,726 (19.6)
SignalP	6,486 (3.8)	6,601 (2.2)	5,380 (3.0)
TMHMM	15,399 (9.1)	16,322 (5.5)	14,537 (8.2)

Tab. 2. Read mapping statistics. The average percentage of reads per species and condition, which could be mapped back to the raw or reduced transcriptome assemblies, respectively.

Species	Condition	Mapped to raw assembly (%)	Mapped to reduced assembly (%)
M. hystrix	Adult (A)	93.4	68.9
	Hatchling (H)	92.9	68.0
	Regenerant (R)	94.1	64.1
M. spirale	Adult (A)	88.1	48.3
	Hatchling (H)	87.0	51.1
	Regenerant (R)	88.7	46.0
M. pusillum	Adult (A)	90.8	74.1
	Hatchling (H)	89.0	73.0
	Regenerant (R)	91.7	74.1

Orthology detection

We used OrthoFinder to infer 23,764 OGs, with 11,331 of those OGs containing sequences from all four species, and 1,190 containing all species except for *M. lignano* (see Additional file 4: Tab. S1 for all inferred OGs). OGs were generally large with only 1,263 single-copy orthologs identified between all four species (these orthologs were used for the species tree inference depicted in Fig. 1, see also below). OrthoFinder provides a summary of the number of gene duplications that occurred on each node of the species tree (Fig. 1), and this analysis indicated that most of the gene duplications occurred on the terminal branches, with the highest number occurring in *M. lignano*.

DE Analysis

When comparing expression of adults vs. hatchlings (AvH), similar numbers of transcripts were DE in all three species, with about twice as many transcripts with higher expression in adults compared to hatchlings (Fig. 3A, see also Additional file 5: Tab. S2 for the DE results of the AvH comparison, and Additional file 6: Tab. S3 and Additional file 7: Tab. S4 for the AvR and RvH contrasts). *M. pusillum* showed slightly lower numbers of DE genes and a DE distribution that deviated from that of the other two species. Specifically, the distributions of DE genes in both *M. hystrix* and *M. spirale* shows a cloud of off-diagonal points, representing transcripts with high expression in adults, but low expression in hatchlings. In *M. pusillum,* this cloud of adult-biased transcripts also exists, but it is shifted up on the y-axis because many of these transcripts also show substantial expression in hatchlings.

We identified a total of 634 OGs that had at least one transcript from every species DE in the AvH contrast (Fig. 3B). 404 of these showed higher expression in adults, 117 showed higher expression in hatchlings, and 113 did not have a consistent signal. Again, we observed differences between *M. pusillum* and the other two species. All but two of the transcripts in those with higher expression in adults also had expression in hatchlings, while in *M. hystrix* and *M. spirale* many transcripts had no expression in hatchlings (see points with red colour at the bottom of the y-axis in Fig. 3B). We explore possible reasons for these observations in the Discussion.

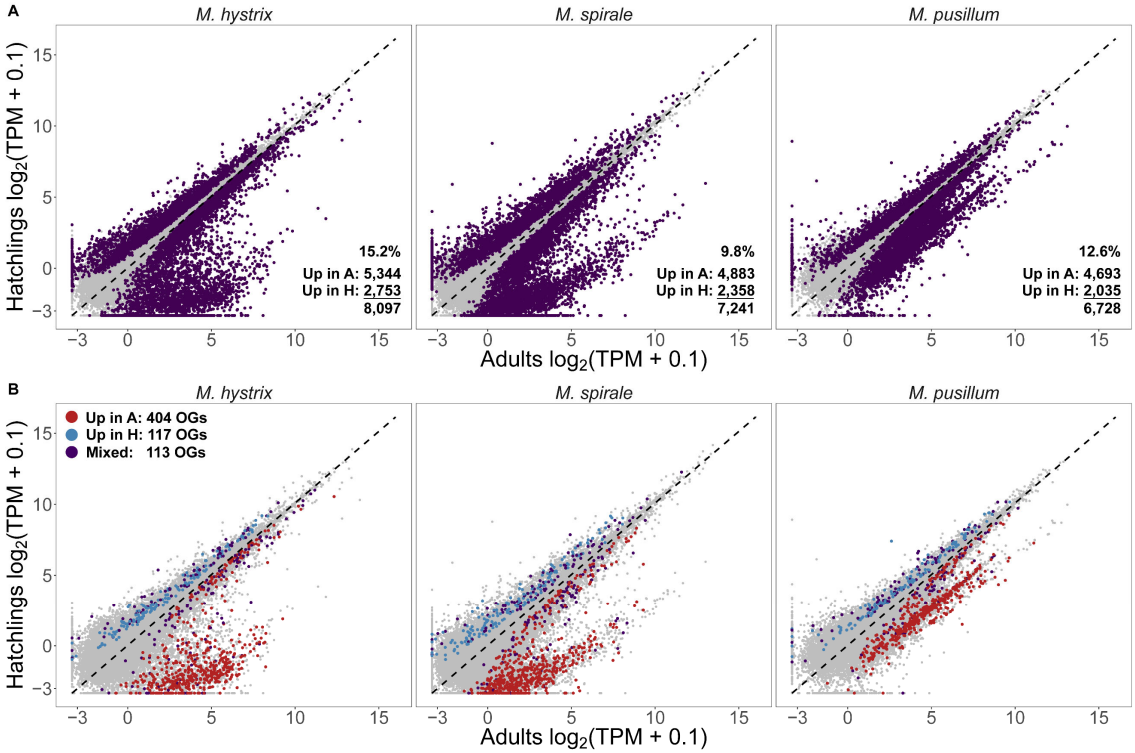


Fig. 3 Results of differential expression (DE) analysis between adults and hatchlings. A: Results of DE analysis comparing the expression in adults (shown on the x-axis) against expression in hatchlings (shown on the y-axis). Highlighted are transcripts that are significantly DE after adjusting for multiple testing (adjusted p-value < 0.05). The numbers at the bottom right of each panel refer to the total number of DE transcripts, and the percentage of DE transcripts out of all transcripts. **B:** The same plots but highlighting only transcripts from OGs that have representatives in all three species (but not necessarily a transcript from *M. lignano*) and in each species at least one transcript that is DE. Transcripts in red are significantly upregulated in adults, transcripts in blue are significantly upregulated in hatchlings, and transcripts in purple show an inconsistent signal within the OG.

Orthogroup annotation

18,938 OGs contained transcripts from *M. lignano* and could thus potentially carry over empirical annotations. Out of these, 6,119 OGs could be annotated with information from the positional (2,495 OGs), neoblast (1,924 OGs), or social (3,717 OGs) RNA-Seq datasets (see Additional file 8: Tab. S5 for the full annotations). In

the positional dataset 173 OGs contain Mlig_37v3 transcripts with conflicting positional information (e.g. tail region and testis region). We categorised these as "positional_mix" and did not consider them further in the downstream analysis since they contain multiple small groups with non-intuitive annotations. Similarly, in the neoblast dataset, we categorised 20 OGs as neoblast_mix because they contained transcripts with the germline annotation (germline_FACS) and transcripts with one of the two neoblast annotations (neobast_FACS and neoblast-strict). Finally, in the social dataset, we categorised 10 OGs as social_mix because they contained transcripts with the octets vs. isolated annotation (OvI) annotation and transcripts with the octets vs. pairs (OvP) annotation, but no transcript annotated from both contrasts (BOTH). We also excluded both the neoblast_mix and the social_mix annotations from the downstream analysis.

There was also overlap between the three RNA-Seq datasets, with several OGs being annotated from multiple sources. The most substantial overlap was between the germline_FACS and the testis region annotation, followed by the overlap between these two annotations and the octets vs. isolated (OvI) annotation (Fig. 4). This overlap was expected since testis region transcripts likely contain mostly transcripts expressed in the testis. Since the neoblast annotation was independent from our reanalysis of the positional dataset, the considerable overlap it shows with the positional and social data supports that these annotations are indeed reflecting biological reality. However, this overlap also made them highly redundant, and we thus excluded the germline annotation from the downstream analysis, retaining only the neoblast annotations. Within the social dataset, most OGs were either annotated as OvI or as BOTH, while only 42 OGs carried the OvP annotation. We also excluded the OvP annotation due to small sample size, leaving us with seven DE annotations in total for the downstream analysis (testis region, ovary region, and tail region; neoblast_FACS and neoblast-strict; and OvI and BOTH; but see Additional file 10: Tab. S6 for a complete annotation of the Mlig_37v3 transcriptome).

The distribution of secretory signals, as estimated by SignalP, was not uniform across the different positional annotations (chi-squared=18.0, df=4, p-value=0.001). The observed counts only differ substantially from the expected counts for the tail region OGs (54 observed vs. 32.9 expected, Tab. 3), indicating that OGs in the tail region are enriched in transcripts with a secretory signal.

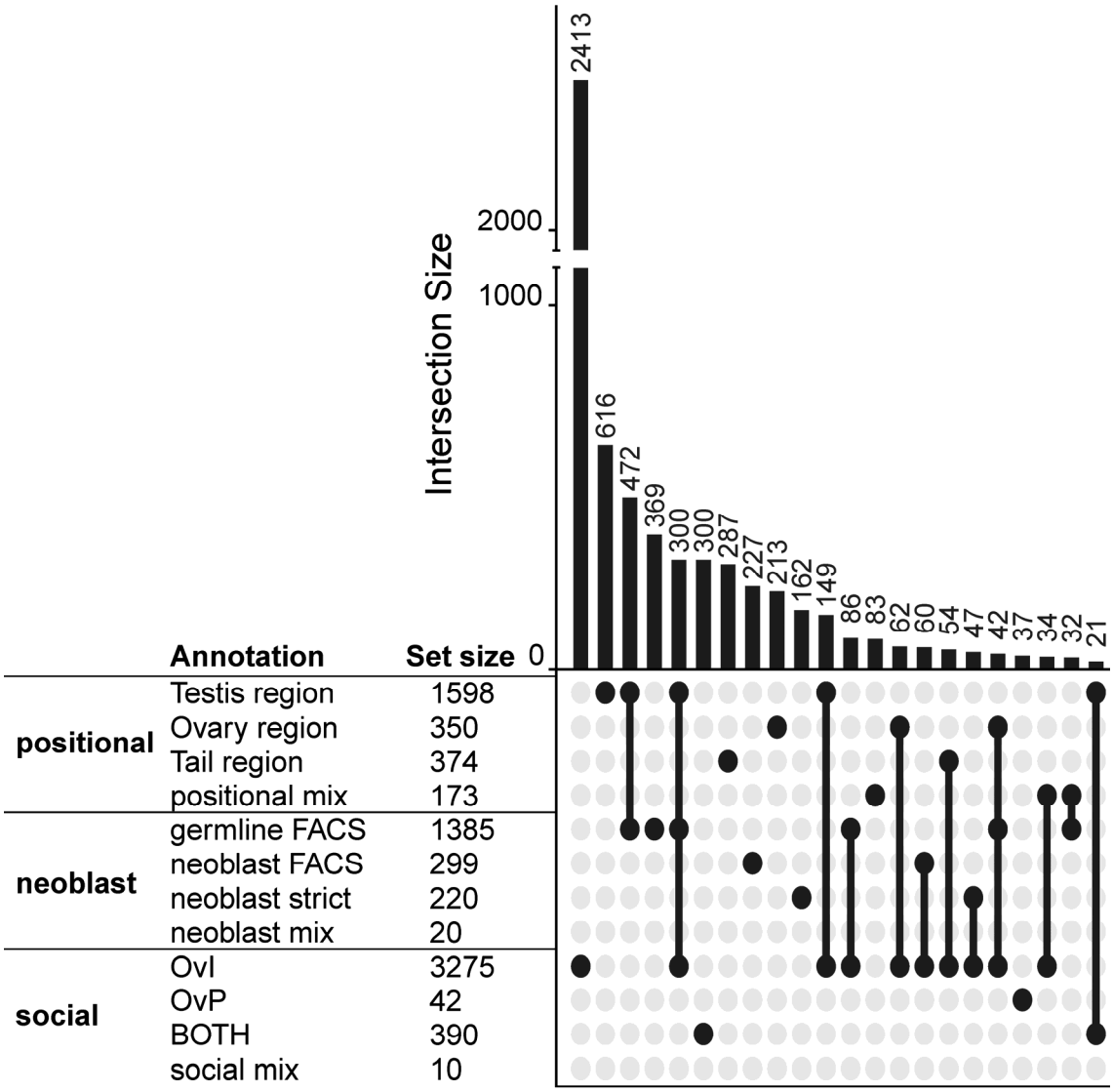


Fig. 4 Upset plot of the intersection of orthogroup (OG) annotations. Annotations are from the positional (testis region, ovary region, tail region, and positional_mix), neoblast (germline_FACS, neoblast_ FACS, and neoblast-strict), and social datasets (OvI, OvP and BOTH). The dots and lines on the bottom right show which intersection is represented by the bar plots above. The size of intersections is given above the bar plot. To the left of the intersection diagram, the absolute number of OGs per annotation is given. Note that only intersections with >20 OGs are displayed here, but that the set sizes reflect the sums of all OGs (for a complete plot see Additional file 9: Fig. S1).

Tab. 3. SignalP enrichment analysis. The number of complete OGs that contain transcripts with a SignalP hit, split by the positional annotation. The expected number of OGs with a SignalP is derived from the chi-square test.

Annotation	OGs with annotation	OGs with signalP	Expected signalP
testis region	728	130	128.5
ovary region	181	37	31.9
tail region	173	53	30.5
positional_mix	84	16	14.8
No annotation	10,165	1764	1794.2

Protein divergence and species composition of OGs differs by annotation

The majority (59.8 %) of OGs with a transcript from *M. lignano* contained all four species and 19.1% contained all species except *M. pusillum*, while only a few (1.2%) were shared just between *M. lignano* and *M. pusillum* (Additional file 11: Tab. S7). The protein divergence of OGs containing all four species differed depending on their annotation, with higher divergence in OGs with a positional annotation (one-sample Wilcoxon: all $p<0.001$, Fig. 5A) and lower divergence in OGs with the neoblast_FACS annotation (one-sample Wilcoxon: $p<0.001$), but not the neoblast-strict annotation (one-sample Wilcoxon: $p=0.2$, Fig. 5B) compared to OGs without an annotation from the respective sources. These patterns of divergence were also reflected in the species composition of OGs, with a smaller than expected percentage of OGs with a positional annotation containing all four species (Fig. 6), which is consistent with the more rapid evolution of these putative reproduction-related transcripts. Conversely, a substantially larger percentage of OGs with either of the neoblast annotations contained all four species (Fig. 6), suggesting that these genes are fairly conserved. Finally, while OGs annotated with the social dataset did not show a difference in protein divergence compared to OGs with no annotation (one-sample Wilcoxon: OvI: $p=0.34$, BOTH: $p=0.34$, Fig. 5 C) they contained a larger than expected percentage of OGs with all four species (Fig. 6). The difference between the expected and observed proportions was, however, quite small for the 'BOTH' annotation (Fig. 6), indicating a small effect size. Moreover, OGs annotated as testis or tail region contained a higher than expected percentage of OGs that were shared only between *M. lignano* and *M. spirale* (Fig. 6). Since both of these species mate through reciprocal copulation and have a characteristic sperm morphology with lateral bristles [15], these OGs are possible targets in the search for the genes underlying these traits. We explore these observations in more detail in the Discussion.

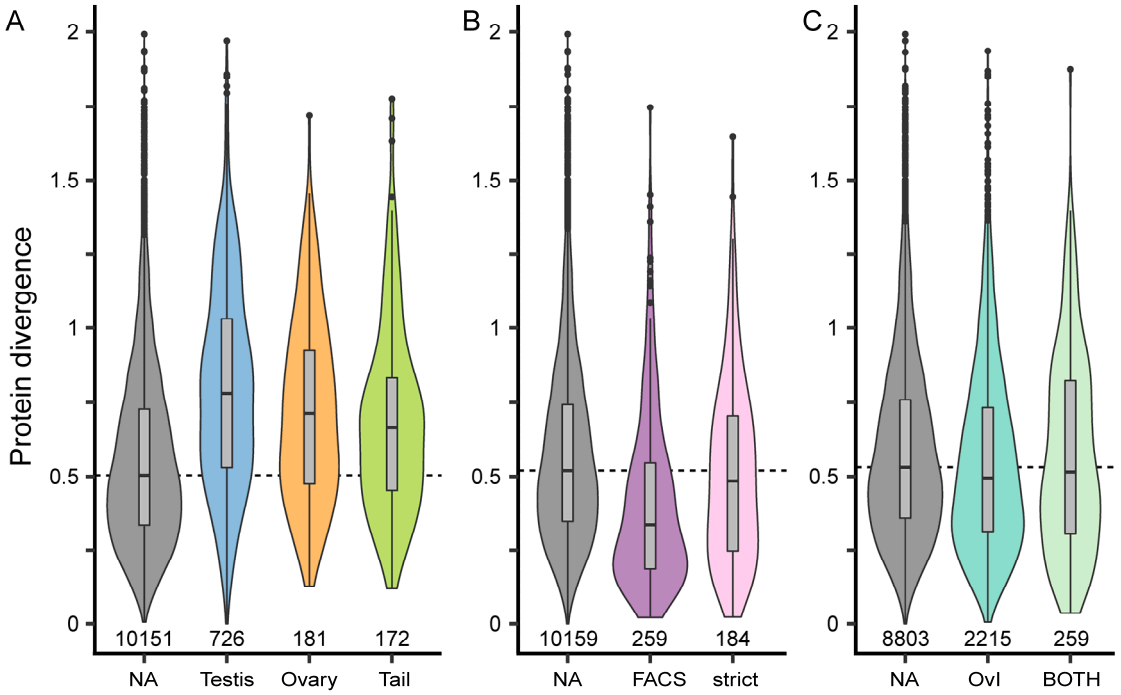


Fig. 5 Violin (mirrored density) plots and boxplots (median, box shows the interquartile range and whiskers extend up to 1.5 times the interquartile range) of the distribution of average protein divergence of OGs with various annotations. Numbers above the x-axis give the number of OGs in each group. A: OGs with a positional annotation (excluding the positional_mix annotation). B: OGs with a neoblast annotation: neoblast_FACS (FACS) and neoblast-strict (strict) (excluding the germline_FACS and neoblast_mix annotations). C: OGs with a social annotation (excluding the OvP and social_mix annotations). The stippled lines represent the median values of the respective OGs with no annotation (NA) against which the OGs with an annotation were tested (see Results).

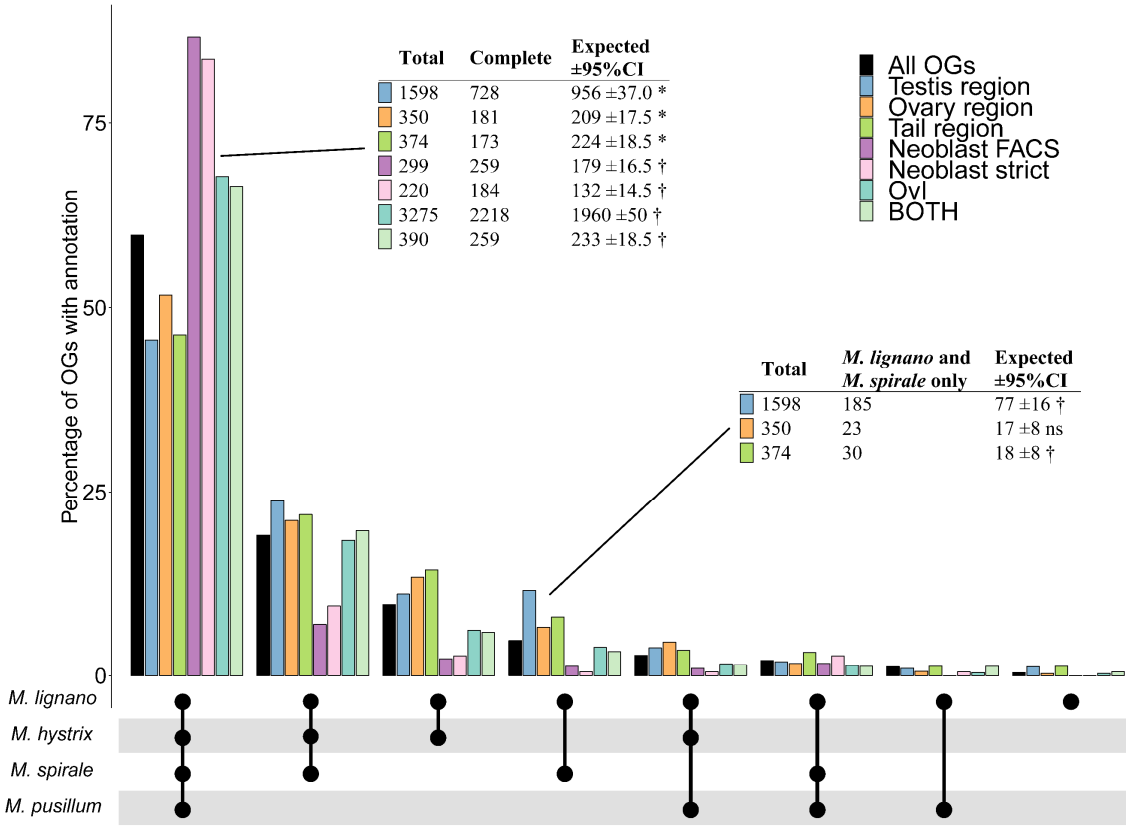


Fig. 6 Species composition of Orthogroups (OG) with different annotations. The bar charts indicate the percentage of OGs with the species composition drawn below. The colours identify OGs with different annotations (see legend top right). The exact numbers and percentages for each annotation and species composition can be found in Additional file 11: Tab. S7. The two inset tables give the results of permutation tests to investigate if fewer or more than the expected number of OGs with a particular annotation contain all four species (i.e. are complete, table on the left) or contain only *M. lignano* and *M. spirale* (table on the right) (see Methods for details). In both tables, the first column gives the total number of OGs with the annotation and the second column gives the number of OGs that are complete (left table) or restricted to *M. lignano* and *M. spirale* (right table). The third column gives the expected number of OGs derived from 100,000 samples from all OGs and the corresponding 95% confidence interval. Symbols indicate if the observed value deviates from the expected value (*=significantly smaller than expected, †=significantly larger than expected, ns=not significantly larger or smaller). All significant p-values are <0.01, all tests were two-tailed, and corrected for multiple comparisons.

OG validation using ISH

As a case study to show the relevance of the OGs across all four studied species, we analysed the expression of a gene that affects the sperm bristle phenotype in *M. lignano* (RNA815_7008 in the MLRNA110815 transcriptome) [20]. This transcript is exclusively expressed in the testes in *M. lignano* [20], and we thus expect its orthologs to also be expressed in the testes of the other species. We designed probes for the orthologs in *M. hystrix*, *M. spirale*, and *M. pusillum* and performed ISH experiments to test this prediction. In addition, we also repeated the ISH experiments in *M. lignano*. We detected a highly specific signal in the testes in

all four species (Fig. 7.; for sense control see Additional file 12: Fig. S2), which i) indicates that tissue specificity of this transcripts is conserved across the genus, and ii) demonstrates that our OGs can be used to identify orthologs and target them using molecular methods.

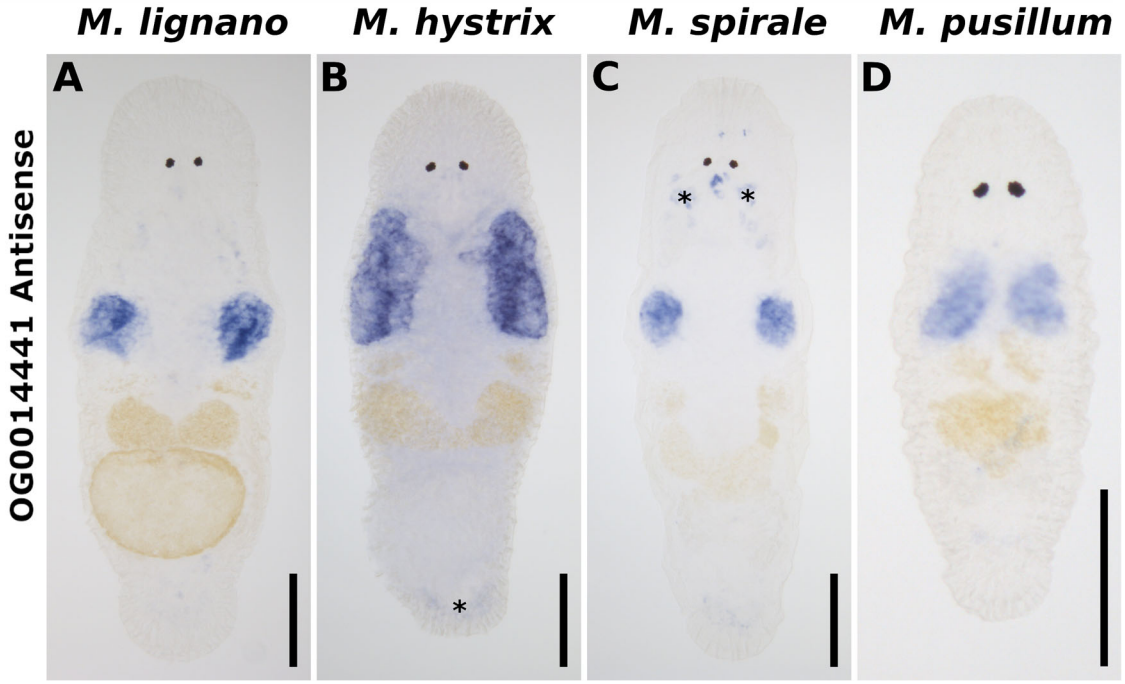


Fig. 7 Whole-mount *in situ* hybridisations for OG0014441. Expression of the transcript affecting the sperm bristle phenotype and its orthologs in the testes of in *M. lignano* (A), *M. hystrix* (B), *M. spirale* (C), and *M. pusillum* (D). Unspecific staining (asterisks) can be seen in pharyngeal glands and in the tail region (staining of pharyngeal glands is also present in control experiments with the sense probe; see also Additional file 12: Fig. S2). Transcripts RNA815_7008 (*M. lignano)*, Machtx_20180703@G07456_i1 (*M. hystrix*), Macspi_20180703@G161928_i1 (*M. spirale*), and Macpus_20180703@G35224_i1 (*M. pusillum*) of orthogroup OG0014441 were used for *in situ* probe generation. Scale bars: 100 µm.

Discussion

In the following section, we will first highlight some differences in the transcriptome assemblies and the DE results between the three species and their possible influence on our conclusions. Then we will focus on the differences in protein sequence divergence and species composition of OGs by annotation and discuss their implications. Note that we were only able to arrive at these results because we spent considerable effort on the reannotation of the *M. lignano* transcriptome. We discuss the majority of this work in Additional file 13: 'Reannotation of Mlig_37v3 transcriptome' (which also makes reference to Additional file 14: Tab. S8; Additional file 15: Tab. S9; and Additional file 16: Tab. S10) and we direct the reader to this document for a detailed explanation of all annotations.

Transcriptome assembly and quality

By performing an analysis of the transcriptome assemblies, we found that the transcriptome of *M. spirale* contained almost twice as many transcripts than the other two transcriptomes. In spite of this, a similar number of annotations was generated in *M. spirale* as in the *M. hystrix* transcriptome, indicating that many of the additional transcripts in the *M. spirale* transcriptome represent either redundant sequences, poor assembly, or non-coding RNA. Poor assembly could be caused by a shorter library insert size, which would make it more difficult to span repetitive sequences. However, insert size of the *M. spirale* libraries was not substantially smaller than that of the other libraries. We think it is more likely that the *M. spirale* libraries contain more heterozygosity, potentially leading to erroneous assembly of alleles as separate contigs. The *M. hystrix* specimens we used originate from the highly inbred SR1 line, and *M. pusillum* can self-fertilise, showing no signs of inbreeding depression [32]. In contrast, the *M. spirale* specimens stem from an outbred culture and might, therefore, retain higher rates of heterozygosity. This is also supported by the 21-mer diversity of the input reads, with the highest number of distinct 21-mers found in *M. spirale* (Tab. 1). The additional complexity of the *M. spirale* assembly resulted in a substantially lower number of reads mapping to the reduced assemblies (Tab. 2), potentially diminishing our ability to detect DE in genes with low expression. However, since we are primarily interested in genes that show strong expression differences between adults vs. hatchlings, this should not adversely affect our conclusions.

Orthology detection

The orthology detection indicated a high number of gene duplications, particularly with respect to the branches leading to each species. Duplications in *M. lignano* are expected, since it has recently been shown to be a hidden tetraploid, having undergone a whole-genome duplication [34, 35]. The 'normal' karyotype of *M. lignano* consists of three small and one large chromosome (2n=8), with some specimens having additional copies of the large chromosome (and more rarely also other aneuploidies) [34]. *M. lignano*'s large chromosome represents a fusion of much of the genetic material found on the three smaller chromosomes [35], suggesting that the majority of the genes are present in duplicate. However, there is no evidence for such a duplication in *M. hystrix* or *M. spirale* (with both species being 2n=6 with three small chromosomes, [10]), while the situation in *M. pusillum* is less clear (being 2n=12 with six small chromosomes, Zadesenets et al. unpublished). It is unclear why we detect such high levels of gene duplication. One possibility is the presence of isoforms or alleles that cannot be clustered using the CD-HIT algorithm. The second highest number of duplications is present in *M. spirale*, which likely also is most heterozygous. To investigate this further, it would be necessary to produce

genome assemblies of these species and to refine the gene models using genome-guided transcriptome assembly, as it was done for *M. lignano* [12, 36].

DE Analysis

The observed expression differences between adults and hatchlings are less pronounced in *M. pusillum* compared to the other two species, with fewer transcripts exclusively expressed in adults, and a general shift to higher expression in hatchlings for transcripts detected as DE in adults. *M. pusillum* has a shorter generation time than the other two species [32], and this could have presented challenges when we collected the hatchlings, leading us to collect worms that had already become subadults and started to express some reproduction-specific transcripts, albeit likely at a relatively low level. Since our hatchling samples were intended to contain various life-stages up to early juveniles, it would be interesting to generate additional RNA-Seq data of even younger *M. pusillum*. If we indeed collected some subadults in the *M. pusillum* hatchling pools, then we would expect the DE patterns between these pools and adults to be more similar compared to the patterns observed in *M. hystrix* and *M. spirale*.

Protein divergence and species composition of OGs differs by annotation

OGs annotated as specific for the testis, ovary, or tail region had higher average protein divergence than OGs with no annotation, suggesting a faster rate of evolution in these reproduction-related genes (Fig. 5). Additionally, OGs annotated with these reproduction-related annotations generally contained fewer species than a random subset of OGs (Fig. 6). While the patterns in species composition could also be produced by processes like gene duplication, gene gain/loss, or introgression, our findings are consistent with previous research showing that reproduction-related genes evolve more rapidly [23, 24] and that a major cause of homology detection failure across species is sequence evolution to the point where homology detection algorithms fail [37, 38]. More thorough testing of this hypothesis will require fitting explicit models of sequence evolution to the data. An important caveat is that the positional annotation does not identify the testis, ovary, and prostate directly, but merely the regions containing these organs. It could, therefore, be that we also include non-reproduction related genes in these analyses. This is undoubtedly the case for the tail region since the tail contains organs that are not present in other parts of the body, such as the adhesive organs as well as the shell and cement glands used for eggshell formation and adhesion of the eggs to the substrate [7, 9, 39]. But this can also be the case for the testis and ovary region since there appears to be a specialised gut epithelium in this area [20]. However, we think it is a valid assumption that reproduction-related genes represent a large proportion of transcripts with a positional annotation. The best evidence for this comes from ISH

experiments, which showed that the majority of the tested transcripts in the ovary and testis region are indeed expressed in the gonads [20]. This is further supported by our ISH experiment showing that the orthologs of a gene expressed exclusively in the testis in *M. lignano* are localised in the same way in *M. hystrix*, *M. spirale*, and *M. pusillum*. Furthermore, detailed screening of the tail region transcripts identified a large proportion as expressed in the prostate gland cells [40]. We also screened our gene annotations of OGs with a positional annotation for the presence of a signal peptide and found that the tail region is enriched in this way. This further supports that we are capturing reproduction-related genes since genes expressed in the prostate have also been shown to be enriched with secretion signals [40]. Finally, including non-reproduction related transcripts in the positional annotations should dilute the signal of increased sequence divergence and thus make our test more conservative, which further supports our finding.

Interestingly, we found an increased proportion of OGs with a testis or tail region annotation that only contained *M. lignano* and *M. spirale*. While the number of the latter is low and thus needs to be interpreted with some caution, the number of the former is quite large, so that this finding can be considered well supported. Since both of these species mate through reciprocal copulation and have a characteristic sperm morphology with lateral bristles [15], these OGs are possible targets in the search for the genes underlying these traits.

In contrast to the OGs with positional annotations, we showed that OGs with one of the neoblast-specific annotations (neoblast_FACS) had lower average protein divergence than the corresponding OGs with no annotation (Fig. 5) and that OGs annotated with both neoblast annotations were more likely to contain all four species (Fig. 6). This suggests that these transcripts are conserved across the genus, as one might expect for stem cell genes since they perform essential functions in homeostasis and regeneration. However, we were not able to place all of the clusters that were annotated with the neoblast dataset in *M. lignano* into an OG (Additional file 13: see Tab. A2 therein). This is despite the fact that many of these clusters appear to have a human homolog according to a BLAST search [6]. The failure to identify these transcripts in other *Macrostomum* species could be due to poor assembly of the homologous transcripts, but more likely is due to the challenges inherent in orthology detection. Orthology detection methods have to balance the trade-off between precision (the correct identification of orthologous relationships) with recall (the total number of genes grouped into OGs) [41, 42]. To avoid spurious grouping of transcripts, these methods, therefore, discard a substantial number of them, leading to a reduction in recall. We do not contend that transcripts in *M. lignano* that were not placed into an OG have no homologs in the other species. Rather, their lack of placement is likely a consequence of the decisions made within the OrthoFinder algorithm, which has been shown to have similar performance compared to other available orthology detection methods (see benchmarks on

orthology.benchmarkservice.org). Furthermore, we consider orthology detection methods like OrthoFinder that rely on explicit modelling of gene trees preferable to approaches relying on similarity only. Approaches using gene trees do more explicitly model evolutionary history and should be more accurate in the presence of gene duplications (as present in *M. lignano* [34]), gene loss, and incomplete assemblies [43, 44].

Finally, we could show that OGs annotated by the social dataset have higher species occupancy, indicating that they are more conserved compared to a random subset of OGs, while they did not show a difference in protein divergence. This is a somewhat counter-intuitive finding since these annotations are reproduction-related, showing the change in expression to the availability of a mate (OvI), the intensity of sperm competition (OvP), or to both (BOTH). Therefore, we would expect a large overlap between social annotations and the germline_FACS as well as the positional annotations. In the original publication, Ramm et al. [22] determined that a large proportion of transcripts with a positional annotation are DE in response to mating (see Fig. 4 in [22]), which is also what we find in our reanalysis of their data (Additional file 13: see Fig. A3 and Tab. A3 therein). Indeed, some OGs annotated as OvI show overlap with the testis region annotation and/or the germline_FACS annotation, but the majority have no overlap with other annotations (Fig. 4). Additionally, there is overlap between the OvI annotation and the neoblast dataset (Fig. 4). This overlap offers a possible explanation for the conservation of socially sensitive genes. In response to mating, multiple physiological changes occur, resulting in a general increase in metabolic activity, which could lead to a higher expression of transcripts involved in general maintenance of cellular processes. Or these transcripts could be regulating more general sensory or neurological traits used for the sensing of conspecifics. Transcripts involved in such fundamental processes are expected to be conserved, which would fit with our observation. OGs annotated with the social dataset are thus likely a heterogeneous population consisting both of reproduction-related genes and general metabolic genes.

Conclusions

The three high-quality transcriptomes and the accompanying DE data, in combination with an annotated set of OGs, will facilitate candidate selection for further investigations of gene function across the genus *Macrostomum*. Particularly interesting in this respect are the OGs with consistent DE across all species, as well as OGs that only contain *M. lignano* and *M. spirale*. These OGs are possible targets to identify reproduction-related genes and should be investigated using molecular techniques such as ISH and RNAi.

We show that reproduction-related genes evolve rapidly within the genus *Macrostomum*. To our knowledge, this is the first evidence for the rapid evolution of

such genes in flatworms and the first evidence for this phenomenon in hermaphroditic animals, since previous research has focused almost exclusively on separate sexed organisms. Future investigations should expand taxon sampling to validate this finding and expand analyses beyond documenting differences in the species composition of OGs and simple protein distances among species. Future studies could use sequence-based approaches such as the estimation of the rate of non-synonymous to synonymous substitution to identify particular genes that evolve rapidly [45].

Finally, our annotations are derived from *M. lignano* and are thus taxonomically biased. We suggest future work to replicate experiments conducted with *M. lignano* in other species, which would allow a more independent and balanced annotation of the OGs. This then would allow validation of the OG annotations as well as permit the annotation of genes that evolve too rapidly to be currently assigned to an annotated OG.

Methods

Animal cultures

The specimens of *Macrostomum hystrix* Ørsted 1843 sensu Luther 1905 used in this study originate from an inbred line derived from an outbred culture initially collected in May 2010 from the San Rossore Regional Park, near Pisa, Italy (N43.6848, E10.2838; note that the name *M. hystrix* is taxonomically problematic, as outlined in [15]). After the discovery that *M. hystrix* can self-fertilise [18], the inbred SR1 line was generated by enforcing selfing for eight generations, followed by several generations of sib-sib breeding (predicted inbreeding coefficient, F = 0.998, [31]). The specimens of *Macrostomum spirale* Ax 1956 derive from a long-term outbred laboratory culture initially collected in November 2004 from the Étang de Biguglia in Corsica, France (N42.6591, E9.4504). Finally, the specimens of *Macrostomum pusillum* Ax 1951 derive from a long-term outbred laboratory culture initially collected in April 2006 in Lignano Sabbiadoro, Italy (N45.6916, E13.1311; note that the name *M. pusillum* is also taxonomically problematic, as outlined in [15]).

All species were kept in replicated populations in glass Petri dishes and fed with the diatom *Nitzschia curvilineata* Hustedt 1922. In every generation, 20 juvenile animals were added to a dish and allowed to grow for several weeks (four to five weeks for *M. spirale* and *M. hystrix* and three to four weeks for *M. pusillum*), after which again 20 juveniles were transferred to a new set of Petri dishes to start the next generation. Worms were kept in artificial seawater (ASW, Wiegandt) at 32‰ salinity for *M. spirale* and *M. pusillum*, and at 8‰ salinity for *M. hystrix*. All animals used for the transcriptomes were not older than two months.

Experimental design

Since some genes may only be active during development or during regeneration, we wanted to obtain animals at various life-stages, so that a significant fraction of genes will be represented in the resulting transcriptomes. This also allowed us to identify genes that are DE between these life-stages and define candidate pools of genes relevant for specific functions (e.g. genes upregulated in adults vs. juveniles are good candidates for reproduction-related genes). We produced RNA samples for adults (A), hatchlings (H), and regenerants (R), using three biological replicates per condition and species, for a total of 27 RNA-Seq libraries (Fig. 2A). We defined adults as animals with clearly visible testes and collected 60 animals per replicate for *M. hystrix* and *M. spirale*, and 225 animals per replicate for *M. pusillum* (due to the smaller body size of this species, see Fig. 1). Hatchling samples consisted of a mixture of animals from various developmental stages, from freshly hatched flatworms up to early juvenile stages, but not having any visible gonads. We collected, on average about 330, 650, and 1100 hatchlings for each replicate of *M. hystrix*, *M. spirale*, and *M. pusillum*, respectively. Due to the large number of animals needed, hatchlings of *M. pusillum* and *M. spirale* were harvested at two time points, dissolved in Tri™ reagent (Sigma), and stored at -80°C until RNA isolation (see below). Animals used for the regenerant group were amputated at the level behind the ovaries (black dotted lines in Fig. 1) and then put in ASW with diatoms and allowed to regenerate for a variable amount of time before sampling to capture animals at various stages of regeneration. For *M. hystrix* and *M. spirale,* ten animals per replicate were amputated each day for six subsequent days, and on the seventh day, total RNA was isolated (6x10 = 60 animals per replicate). *M. pusillum* was treated in a similar way, but due to the smaller size and shorter regeneration time, five times 30 animals were amputated, and total RNA was isolated on the sixth day (5x30 = 150 animals per replicate).

RNA isolation, library preparation and sequencing

Before extraction, worms were starved for 24h to give them time to regurgitate consumed diatoms. Next, worms were gradually relaxed using a dilution series of 7.14% $MgCl_2$ in water and then directly dissolved in Tri™ reagent (Sigma) by pipetting up and down. Subsequently, the extraction was performed as to the manufacturer's recommendations, with the slight modification that we centrifuged the Tri™ reagent – Chloroform mixture for 20min instead of the recommended 15min. Quality checking, library preparation, and sequencing were performed by the Genomics Facility Basel at the Department of Biosystems Science and Engineering of the ETH Zürich in Basel. Libraries were prepared using the TruSeq® Stranded mRNA kit (Illumina) and sequenced as 101bp paired-end reads on a HiSeq2500 sequencer (using the HiSeq® SBS Kit v4, Illumina).

Transcriptome assembly

We used Rcorrector (commit 24940c9, [46]) with standard settings to correct for error due to inaccurate base calling, retaining reads that could not be corrected. Rcorrector is a k-mer-based correction technique that has been developed specifically for RNA-Seq data. It first constructs a De Bruijn graph of the reads and then assesses the coverage of the k-mers in the graph. K-mers with low coverage compared to other members of the path are likely due to sequencing error and are corrected [46]. After error correction, we trimmed sequencing adapters and low quality reads using Trimmomatic (version 0.36, command: :2:30:10:8:TRUE LEADING:5 TRAILING:5 SLIDINGWINDOW:4:15 AVGQUAL:30 MINLEN:36), removing low-quality regions and requiring an overall Phred score of 30. We then *de novo* assembled transcriptomes using Trinity (Version 2.6.6, [47]) , with a k-mer size of 25, digital read normalisation, and with settings for stranded libraries.

Removal of cross-contamination

All 27 libraries were multiplexed and ran on two lanes of the sequencer. To demultiplex reads, all libraries were assigned to their respective sample using dual combinatorial TruSeq kit indices (i.e. the 8bp i5 and i7 indices D501-D508 and D707-D710, respectively). Our samples were only unique in one index with all i5 and two of the four i7 indices used to tag multiple species. It has recently become evident that this kind of indexing can lead to demultiplexing errors, due to so-called index hopping [48, 49]. During this process highly expressed reads from one library can cross-contaminate another library on the same lane and then appear to occur in that library at low counts. Because we used such indices and had a high sequencing coverage, this is likely a concern in our data. To mitigate the issue, we cleaned our assemblies using CroCo with default settings [50]. CroCo uses information about the levels of expression across assemblies to detect cross-contamination [50]. We removed transcripts that CroCo detected as contaminations and retained transcripts with an ambiguous signal to be conservative.

Transcriptome quality assessment

To assess transcriptome quality, we ran TransRate (version 1.0.2, [33]), which maps the reads back to the assembly and calculates mapping metrics (e.g. if both read pairs map to the same transcript in the expected order), followed by BUSCO (version 2.0, [51]), which searches for the presence of a curated set of core conserved genes. Specifically, we ran the BUSCO analysis with the metazoan dataset consisting of 978 genes (version uploaded 2016-11-01). We determined the empirical insert size of our libraries by mapping the reads to the assemblies using bwa (version 0.7.17-r1188) and then extracting the mean insert size using Picard (version 2.20.2). We calculated the 21-mer distribution of the trimmed and corrected reads using jellyfish (option: -C, version 2.2.6) and recorded the number of distinct k-mers.

Transcriptomes used for orthology detection and DE analysis

Since the employed orthology detection method operates on amino acid sequences (see next section), we first inferred open reading frames (ORFs) and their corresponding amino acid sequences, for the generated transcriptomes of *M. hystrix*, *M. spirale*, and *M. pusillum*. We used TransDecoder (version 2.0.1, [52]) in combination with Pfam searches (version 32.0) to retain transcripts with predicted proteins and kept only one ORF per transcript using the "--single_best_only" option. We then reduced the resulting predicted proteins using the CD-HIT clustering algorithm (version 4.7, [53]), set to cluster amino acid sequences with at least 99.5% identity and extracted the coding sequences corresponding to the clustered amino acids for DE analysis. These simplified transcriptomes were then used in the orthology detection pipeline and the DE analysis.

For *M. lignano,* we used a previously published genome-guided transcriptome assembly (Mlig_RNA_3_7_DV1.v3 [36]; with the method described in [12]) as a starting point for our analysis. Since many transcripts in *M. lignano* are trans-spliced [11, 12], the initial assembly had been modified to refine the gene models, predict open reading frames (ORF) using TransDecoder, and remove non-coding and repetitive regions by mapping to the reference genome (Mlig_RNA_3_7_DV1.v3.coregenes method described in [12]). In addition, Grudniewska et al. [36] provide a file containing the amino acid sequences for only the best ORF per transcript (generated using TransDecoder, Mlig_RNA_3_7_DV1.v3.genes.bestORF.pep). This data corresponds most closely to the amino acid data we generated from our *de novo* assemblies. We therefore also clustered this version of the assembly with CD-HIT at 99.5% sequence identity and again extracted the corresponding coding sequences.

We thus used four amino acid sequence datasets represented by the best ORF per transcript for our orthology detection and the corresponding coding sequence datasets for the quantification of expression and DE analyses. We refer to these datasets as the reduced transcriptome assemblies in the case of the three *de novo* assemblies and as Mlig_37v3 for the *M. lignano* assembly.

Orthology detection

Our main aim with the orthology detection analysis was to identify homologous genes between the three species we sequenced for this study and the well-annotated transcriptome of *Macrostomum lignano*. We used OrthoFinder in amino acid mode (version 2.2.6, [41]), a method that infers whole sets of homologous transcripts (which we call orthogroups [OGs] throughout the text), based on a gene tree approach. We ran OrthoFinder with the "--msa" flag to use multiple sequence alignment instead of the default DendroBLAST. In this mode, OrthoFinder infers multiple sequence alignments for each cluster of putative homologs using MAFFT

pusillum a

and then infers a gene tree using FastTree. To assess the influence of the clustering on the detected orthologs, we ran OrthoFinder with both the raw and the reduced transcriptomes as input (raw refers to the transcriptomes before CD-HIT clustering). We found that most transcripts were shared between the two approaches, with only a smaller fraction exclusive to one method (shared: 218,367, raw only: 20,119, CD-HIT only: 8,768). For the following, we decided to use the CD-HIT clustered amino acid ortholog sets, since they are less complex and more amenable to downstream analysis.

DE analysis

We conducted DE analysis between all three of our biological conditions, i.e. adults (A), hatchlings (H), and regenerants (R). However, of the three possible resulting contrasts, we here primarily focus on the comparison between adults vs. hatchlings (AvH), since this comparison permits to identify candidate transcripts that are DE in the context of reproduction. The comparisons between adults vs. regenerants (AvR) and regenerants vs. hatchlings (RvH) are mainly dealt with in the additional information, with the former comparison permitting to identify candidate transcripts that are DE during regeneration, while the latter comparison does not *a priori* represent a very informative contrast.

We quantified the expression of transcripts in our reduced transcriptomes since this allows for easier comparison of expression between species. Specifically, we mapped the trimmed and corrected reads used for transcriptome assembly (see above) to the coding sequences of the reduced transcriptome assemblies, using Salmon in quasi-mapping mode (version 0.9, [54]) and then inferred DE using DESeq2 (version 1.24.0, [55]). Filtering to remove genes with low expression can improve the power of DE analyses [55–58], and we took the following two-step approach, which uses the independent filtering feature of DESeq2. First, all of the data was run through the DESeq2 pipeline and all pairwise contrasts tested (AvH, AvR and RvH). The overall mean count thresholds identified by DESeq2 for each contrast were collected, and the minimum of these thresholds was then used to filter genes for multiple test correction within each contrast. This ensured that the same criteria were used in each analysis. The number of transcripts remaining after this filtering procedure is given in Tab. 1. Thus, only these remaining transcripts were tested for DE. Note that estimates of DE using *de novo* assembled transcriptomes can lead to biased estimates and should be interpreted cautiously [59].

Detection of OGs with consistent DE signal between species

The inferred OGs can be combined with the DE analysis to identify gene families showing a consistent expression signal across all three species. We provide candidate gene sets from OGs with a consistent DE signal between adults and hatchling (Additional file 4: Tab. S1 and Additional file 8: Tab. S5 detailing the OGs and all

annotation information for the OGs, respectively). These genes then are not only conserved in sequence (as indicated by the fact that they are in the same OG) but also in aspects of their expression level, making them promising targets in the search for reproduction-related genes (if upregulated in adults) or genes essential for development (if upregulated in hatchlings). To be annotated, we required that an OG has at least one transcript per species that is DE in the contrast under investigation. We then categorised the OGs into those that showed a consistent signal between all DE genes versus those that showed conflicting signals.

Reannotation of the *Macrostomum lignano* transcriptome

Previous studies have used RNA-Seq in *M. lignano* to identify groups of genes involved in reproduction [20–22] or neoblast function [6], but not all of these studies used the same reference transcriptome. We have therefore combined information from three selected RNA-Seq studies (see Fig. 2B-D) and transferred their annotations to the most up-to-date transcriptome (see above). Specifically, we included a study [20] that generated expression data for candidate genes primarily expressed in specific body regions (i.e. the testis, ovary, and tail region; referred to as the **positional dataset**), a study [6] identifying genes expressed in both neoblasts and germline tissue (referred to as the **neoblast dataset**), and a study [22] that compared expression between worms held in different social group sizes (isolated, pairs and octets; referred to as the **social dataset**). We also identified transcripts that would be amplified using primers of existing ISH probes that had been designed based on previous transcriptome versions. In Additional file 13: "Reannotation of Mlig_37v3 transcriptome" we summarise the approaches used in these studies in some more detail and explain how we transferred these findings to the Mlig_37v3 transcriptome, to be subsequently transferred to our newly generated transcriptomes (see next section).

Transcriptome annotation

We performed *de novo* annotation of the three transcriptomes using Trinotate (version 3.1.1, [60]), which performs a BLASTX search against a protein database (in our case Swiss-Prot and UniRef90 from the 2018_11 release) to assign an annotation. The tool also assesses the presence of signal peptides with SignalP (version 4.1, [61]), transmembrane domains with TMHMM (version 2.0, [62], and domain content with HMMER (version 3.1b2), against the Pfam protein family database (version 32.0, [63]). We also transferred annotations derived from the three RNA-Seq experiments, and the different ISH experiments performed with *M. lignano* (see the previous section), using our inferred OGs. For this, we assumed that genes sharing an OG have a similar function and transferred the annotations from the *M. lignano* genes to the other genes in the group. We allowed an OG to carry multiple annotations from within and across datasets.

OG protein divergence by annotation

To investigate if certain OGs show more divergence at the sequence level, we estimated the protein divergence for each OG that contained all four species. We first filtered each OG alignment to only contain protein sequences that shared an aligned region (Additional file 17: Fig. S3A), leading to the removal of one OG that as a result contained only three species. For the remaining 11330 OGs, we then calculated all pairwise protein distances using the *protdist* function of PHYLIP (version 3.697, [64]), using the JTT substitution model, and retaining only the between-species values (Additional file 17: Fig. S3B). To avoid choosing one representative sequence per species and OG, we summarised the protein distances of each species pair as the mean of all pairwise comparisons between their sequences (Additional file 17: Fig. S3C). Finally, we averaged the protein distances for all species pairs to obtain one estimate of protein divergence per OG. For further analysis, we excluded 16 OGs with a protein divergence of more than 2 substitutions per site, since these likely are close to saturation or represent misalignments. We then compared the distributions of protein divergences of annotated OGs (positional: testis [n=726], ovary [n=259] and tail region [n=172]; neoblast: neoblast_FACS [n=259] and neoblast-strict [n=184]; social: OvI [n=2215] and BOTH [n=259]) against OGs without an annotation from each annotation source (positional [n=10151], neoblast [n=10159] and social [n=8803]). To partially mitigate the pseudoreplication that might result from treating the OGs as independent samples we calculated the median protein divergence for all OGs without annotation and then performed a one-sample Wilcoxon signed-rank test against this value, additionally correcting for multiple comparisons using the Benjamini-Hochberg procedure [65].

OG species composition by annotation

If reproduction-related genes evolve rapidly, then we expect few OGs with an *M. lignano*-derived annotation suggestive of sexual reproduction to be complete (i.e. contain all four species). In particular, we expect fewer OGs to contain *M. pusillum* since it is most distantly related to *M. lignano* (Fig. 1). In contrast, OGs with a neoblast-specific annotation are expected to be more conserved and we therefore expected them to be complete more often. To test this, we compared the proportion of complete OGs of each annotation type to a random sample of all OGs. Additionally, we compared the proportion of OGs with a positional annotation containing only *M. lignano* and *M. spirale* to a random sample of all OGs. These OGs are candidates for genes responsible for the morphological differences between reciprocally and hypodermically mating species, since the hypodermically mating *M. hystrix,* is phylogenetically closer to *M. lignano* than to *M. spirale* (Fig. 1). Particularly these OGs could contain genes controlling the presence or length of the sperm bristles and the sperm brush, structures that are absent in hypodermically mating species (see drawings in Fig. 1).

We used a resampling approach to test whether OGs differed in their species composition based on their annotation. We sampled from all OGs that contained *M. lignano*, whether or not they were annotated. For each annotation tested, we then drew a random sample, equal to the number of OGs with the annotation, and recorded species in them (e.g. OGs containing all four species). We repeated this procedure 100,000 times and compared the resulting distribution to the empirical value. We used the proportion of draws where the empirical sample was smaller or larger than the draw as the p-value testing whether the annotated OGs have a higher or smaller number of genes with a particular species composition than expected. Note, that transcripts cannot be considered completely independent since i) they can be part of the same co-expression network and ii) linkage disequilibrium between them can constrain their evolution. Both of these effects can effectively lead to pseudoreplication, and the p-values should thus be interpreted with this caveat in mind. To correct for multiple testing, we used the Benjamini-Hochberg procedure and applied it to all p-values generated for each test within OGs set with a particular species composition. Resampling was done in R (version 3.5.0, CRAN). Visualisation of intersections between annotations was done using UpSet plots [66], as implemented in the R package UpSetR [67].

Testing for enrichment in signal peptide

Since transcripts annotated as tail region specific in the positional dataset have previously been shown to contain a high number of seminal fluid proteins with evidence for secretion [40], we tested if OGs with a tail region annotation were enriched for a SignalP annotation. For this, we conducted a chi-square test, comparing the expected count of SignalP annotations (derived from the proportion of OGs with each annotation type) to the observed count. We then visually compared expected and observed counts visually to determine which annotation class was enriched.

Phylogenetics

We used 1,263 one-to-one orthologs identified by OrthoFinder (see above) to infer the phylogenetic relationship between our four species. We aligned the amino acid sequence of each ortholog using MAFFT (version 7.310 [68]) with the L-INS-i algorithm and concatenated the alignments with AMAS [69]. This resulted in an alignment of 615,314 amino acid sites with 13.6% missing data. We estimated a maximum likelihood phylogeny using IQ-TREE (version 1.5.5, [70]) with a separate partition for each gene and we inferred the best substitution model for each partition using ModelFinder [71] with the BIC criterion. We used ultrafast bootstrapping [72], combined with the Shimodaira–Hasegawa–like approximate likelihood ratio test [73] to assess support for bipartitions.

Whole-mount ISH

We performed ISH for members of OG0014441, which contained i) the best blastn hit for the RNA815_7008 transcript of *M. lignano* [20] (Maclig_37v3@Mlig016310.g1, see Additional file 18: Tab. S11), ii) Machtx_20180703@G07456_i1, iii) Macspi_20180703@G161928_i1, and iv) Macpus_20180703@G35224_i1. For RNA815_7008, we added the T7 and SP6 sequences to the 5' end of the published primers and we designed primers for the other species. PCR conditions were: 98°C 30s; 35 x [98°C 10s; 58°C 30s; 72°C 30s]; 72°C 120s, 4°C 15min. The resulting PCR products for *M. hystrix, M. spirale,* and *M. pusillum* were cloned into the pGEM-T vector (Promega, USA). Plasmids were extracted with the PureYield Plasmid Minipreps System (Promega, USA). PCR with the M13 primer set was performed on the plasmids to obtain the template DNA, which was cleaned up with the Wizard SV Gel and PCR Clean-Up System (Promega, USA) prior to probe synthesis. PCR fragments of *M. lignano* were purified with the Roche High Pure PCR Product Purification Kit (Sigma-Aldrich, USA). All four DNA templates were sequenced at Microsynth AG, Switzerland. The primer sequences and the sequenced data can be found in Additional file 19: Tab. S12. As a control for unspecific staining, we also performed all ISH using sense probes, which should not ligate to the target mRNA.

The ISH probes were synthesized with the Roche DIG RNA labelling kit (SP6/T7; Sigma-Aldrich, USA) using 6.5µl of the template DNA in a 10µl reaction. Probes were cleaned up with Micro Bio-Spin 6 Columns (Bio-Rad, USA), following the manufacturer's protocol. Probes were diluted to 5ng/µl in hybridization mix and stored at -80°C.Whole-mount ISH was performed according to the WISH protocol described in [14] with the following modifications: i) animal relaxation with 7.14% $MgCl_2$ x 6 H_2O was prolonged to 22 minutes for *M. pusillum* and 70 minutes for *M. spirale*, both on ice. *M. hystrix* was relaxed in 0.1% Chlorethone (1,1,1-Trichloro-2-methyl-2-propanol) in 8‰ artificial seawater for 20 minutes on ice; ii) a decreasing methanol series instead of ethanol series was used; iii) protease treatment was shortened to 15 minutes for *M. hystrix* and *M. spirale* and to 10 minutes for *M. pusillum*. iv) The heat-fixation in PBSw was prolonged to 30 minutes (for *M. hystrix, M. spirale,* and *M. pusillum*). v) The temperature of the stringent Hybmix/SSC-buffer washing steps was increased to 64 °C for *M. hystrix, M. spirale,* and *M. pusillum*. These changes reduced background staining, which should facilitate additional investigations in the future.

Availability of data and materials

The raw sequencing data generated for this study are available in the NCBI Sequence Read Archive repository with the following accession: PRJNA590781.

The assembled transcriptomes, annotation files, and other supporting information are available in a zenodo repository at DOI:10.5281/zenodo.3547572

Funding

This research was supported by Grant FWF P30347 of the Austria Science Fund to PL and Grants 31003A_162543 and 310030_184916 of the Swiss National Science Foundation (SNSF) to LS. PB was supported by the Fonds National de la Recherche Luxembourg (13569708) and the University of Innsbruck. The funding bodies had no role in the study design, the data collection and analysis, the decision to publish, or the preparation of the manuscript.

Author contributions

PL and LS conceived the study. RP and PL maintained all animal cultures. RP performed RNA isolation. CB supervised library preparation and sequencing. PB performed the ISH in consultation with PL and RP. RAWW performed primer mapping. RAWW and JNB performed the DE analysis. JNB assembled and annotated the transcriptomes, performed the orthogroup inference and annotation, performed all statistical analysis, and wrote the manuscript in consultation with LS. All authors have read and approved the manuscript.

Acknowledgements

We thank Steven A. Ramm for his help in conceiving and planning this study. We thank Dita B. Vizoso for the use of her drawing of the sperm of *M. lignano* and the drawings of sperm and stylets of *M. hystrix, M. spirale*, and *M. pusillum*. We thank Katja Eschbach of the Genomics Facility Basel for the library preparation. We thank Birte Mertens and Julia Wunderer for help with the ISH. We thank Julia Wunderer and Birgit Lengerer for providing previously unpublished primers used for ISH in *M. lignano*. We thank Lukas Zimmermann for IT support.

References

1. Janicke T, Marie-Orleach L, De Mulder K, Berezikov E, Ladurner P, Vizoso DB, et al. Sex allocation adjustment to mating group size in a simultaneous hermaphrodite. Evolution. 2013;67:3233–42.

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

3. Marie-Orleach L, Janicke T, Vizoso DB, David P, Schärer L. Quantifying episodes of sexual selection: insights from a transparent worm with fluorescent sperm. Evolution. 2016;70:314–28.

4. Mouton S, Willems M, Back P, Braeckman BP, Borgonie G. Demographic analysis reveals gradual senescence in the flatworm Macrostomum lignano. Front Zool. 2009;6:15.

5. Mouton S, Grudniewska M, Glazenburg L, Guryev V, Berezikov E. Resilience to aging in the regeneration-capable flatworm Macrostomum lignano. Aging Cell. 2018;17:e12739.

6. Grudniewska M, Mouton S, Simanov D, Beltman F, Grelling M, Mulder K de, et al. Transcriptional signatures of somatic neoblasts and germline cells in Macrostomum lignano. eLife. 2016;5:e20607.

7. Lengerer B, Pjeta R, Wunderer J, Rodrigues M, Arbore R, Schärer L, et al. Biological adhesion of the flatworm Macrostomum lignano relies on a duo-gland system and is mediated by a cell type-specific intermediate filament protein. Front Zool. 2014;11:12.

8. Lengerer B, Hennebert E, Flammang P, Salvenmoser W, Ladurner P. Adhesive organ regeneration in Macrostomum lignano. BMC Dev Biol. 2016;16:20.

9. Wunderer J, Lengerer B, Pjeta R, Bertemes P, Kremser L, Lindner H, et al. A mechanism for temporary bioadhesion. Proc Natl Acad Sci. 2019;116:4297–306.

10. Zadesenets KS, Vizoso DB, Schlatter A, Konopatskaia ID, Berezikov E, Schärer L, et al. Evidence for karyotype polymorphism in the free-living flatworm, Macrostomum lignano, a model organism for evolutionary and developmental biology. Plos One. 2016;11:e0164915.

11. Wasik K, Gurtowski J, Zhou X, Ramos OM, Delás MJ, Battistoni G, et al. Genome and transcriptome of the regeneration-competent flatworm, Macrostomum lignano. Proc Natl Acad Sci. 2015;112:12462–7.

12. Wudarski J, Simanov D, Ustyantsev K, de Mulder K, Grelling M, Grudniewska M, et al. Efficient transgenesis and annotated genome sequence of the regenerative flatworm model Macrostomum lignano. Nat Commun. 2017;8:2120.

13. Pfister D, De Mulder K, Philipp I, Kuales G, Hrouda M, Eichberger P, et al. The exceptional stem cell system of Macrostomum lignano: Screening for gene expression and studying cell proliferation by hydroxyurea treatment and irradiation. Front Zool. 2007;4:9.

14. Pfister D, De Mulder K, Hartenstein V, Kuales G, Borgonie G, Marx F, et al. Flatworm stem cells and the germ line: Developmental and evolutionary implications of macvasa expression in Macrostomum lignano. Dev Biol. 2008;319:146–59.

15. Schärer L, Littlewood DTJ, Waeschenbach A, Yoshida W, Vizoso DB. Mating behavior and the evolution of sperm design. Proc Natl Acad Sci. 2011;108:1490–5.

16. Vizoso DB, Gunde Rieger, Lukas Schärer. Goings-on inside a worm: functional hypotheses derived from sexual conflict thinking. Biol J Linn Soc. 2010;99:370–383.

17. Schärer L, Joss G, Sandner P. Mating behaviour of the marine turbellarian Macrostomum sp.: these worms suck. Mar Biol. 2004;145:373–80.

18. Ramm SA, Vizoso DB, Schärer L. Occurrence, costs and heritability of delayed selfing in a free-living flatworm. J Evol Biol. 2012;25:2559–68.

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

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

21. Lengerer B, Wunderer J, Pjeta R, Carta G, Kao D, Aboobaker A, et al. Organ specific gene expression in the regenerating tail of Macrostomum lignano. Dev Biol. 2017;433:448–60.

22. Ramm SA, Lengerer B, Arbore R, Pjeta R, Wunderer J, Giannakara A, et al. Sex allocation plasticity on a transcriptome scale: Socially sensitive gene expression in a simultaneous hermaphrodite. Mol Ecol. 2019;00:1–21.

23. Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins. Nat Rev Genet. 2002;3:137–44.

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

25. Grus WE, Zhang J. Rapid turnover and species-specificity of vomeronasal pheromone receptor genes in mice and rats. Gene. 2004;340:303–12.

26. Dean MD, Clark NL, Findlay GD, Karn RC, Yi X, Swanson WJ, et al. Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (Mus domesticus). Mol Biol Evol. 2009;26:1733–43.

27. Dorus S, Wasbrough ER, Busby J, Wilkin EC, Karr TL. Sperm proteomics reveals intensified selection on mouse sperm membrane and acrosome genes. Mol Biol Evol. 2010;27:1235–46.

28. Jagadeeshan S, Singh RS. Rapidly evolving genes of Drosophila: Differing levels of selective pressure in testis, ovary, and head tissues between sibling species. Mol Biol Evol. 2005;22:1793–801.

29. Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ram KR, Sirot LK, et al. Evolution in the fast lane: Rapidly evolving sex-related genes in Drosophila. Genetics. 2007;177:1321–35.

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

31. Winkler L, Ramm SA. Experimental evidence for reduced male allocation under selfing in a simultaneously hermaphroditic animal. Biol Lett. 2018;14:20180570.

32. Giannakara A, Ramm SA. Self-fertilization, sex allocation and spermatogenesis kinetics in the hypodermically inseminating flatworm Macrostomum pusillum. J Exp Biol. 2017;220:1568–77.

33. Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Res. 2016;26:1134–44.

34. Zadesenets KS, Schärer L, Rubtsov NB. New insights into the karyotype evolution of the free-living flatworm Macrostomum lignano (Platyhelminthes, Turbellaria). Sci Rep. 2017;7:6066.

35. Zadesenets K, Ershov N, Berezikov E, Rubtsov N. Chromosome evolution in the free-living flatworms: first evidence of intrachromosomal rearrangements in karyotype evolution of Macrostomum lignano (Platyhelminthes, Macrostomida). Genes. 2017;8:298.

36. Grudniewska M, Mouton S, Grelling M, Wolters AHG, Kuipers J, Giepmans BNG, et al. A novel flatworm-specific gene implicated in reproduction in Macrostomum lignano. Sci Rep. 2018;8:3192.

37. Jain A, Perisa D, Fliedner F, von Haeseler A, Ebersberger I. The Evolutionary Traceability of a Protein. Genome Biol Evol. 2019;11:531–45.

38. Vakirlis N, Carvunis A-R, McLysaght A. Synteny-based analyses indicate that sequence divergence is not the main source of orphan genes. eLife. 2020;9:e53500.

39. Ladurner P, Schärer L, Salvenmoser W, Rieger RM. 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. 2005;43:114–26.

40. Weber M, Wunderer J, Lengerer B, Pjeta R, Rodrigues M, Schärer L, et al. A targeted in situ hybridization screen identifies putative seminal fluid proteins in a simultaneously hermaphroditic flatworm. BMC Evol Biol. 2018;18:81.

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

42. Altenhoff AM, Boeckmann B, Capella-Gutierrez S, Dalquen DA, DeLuca T, Forslund K, et al. Standardized benchmarking in the quest for orthologs. Nat Methods. 2016;13:425–30.

43. Gabaldón T. Large-scale assignment of orthology: back to phylogenetics? Genome Biol. 2008;9:235.

44. Yang Y, Smith SA. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. Mol Biol Evol. 2014;31:3081–92.

45. Wilkinson GS, Breden F, Mank JE, Ritchie MG, Higginson AD, Radwan J, et al. The locus of sexual selection: moving sexual selection studies into the post-genomics era. J Evol Biol. 2015;28:739–55.

46. Song L, Florea L. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaScience. 2015;4:48.

47. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.

48. Costello M, Fleharty M, Abreu J, Farjoun Y, Ferriera S, Holmes L, et al. Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. BMC Genomics. 2018;19:332.

49. MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, Giorda K, et al. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. BMC Genomics. 2018;19:30.

50. Simion P, Belkhir K, François C, Veyssier J, Rink JC, Manuel M, et al. A software tool 'CroCo' detects pervasive cross-species contamination in next generation sequencing data. BMC Biol. 2018;16:28.

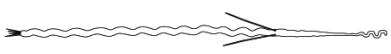
51. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO Applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2018;35:543–8.

52. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512.

53. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.

54. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nat Methods. 2017;14:417–9.

55. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.



56. Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. Proc Natl Acad Sci. 2010;107:9546–51.

57. Sha Y, Phan JH, Wang MD. Effect of low-expression gene filtering on detection of differentially expressed genes in RNA-seq data. IEEE Eng Med Biol Soc Annu Conf. 2015;2015:6461–4.

58. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17:13.

59. Vijay N, Poelstra JW, Künstner A, Wolf JBW. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. Mol Ecol. 2013;22:620–34.

60. Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, et al. A tissue-mapped Axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Rep. 2017;18:762–76.

61. Petersen TN, Brunak S, Heijne G von, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6.

62. Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol. 1998;6:175–82.

63. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279–85.

64. Felsenstein J. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics. 1989;5:164–6.

65. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995;57:289–300.

66. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. UpSet: Visualization of intersecting sets. IEEE Trans Vis Comput Graph. 2014;20:1983–92.

67. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017;33:2938–40.

68. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

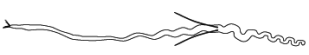
69. Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 2016;4. doi:10.7717/peerj.1660.

70. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.

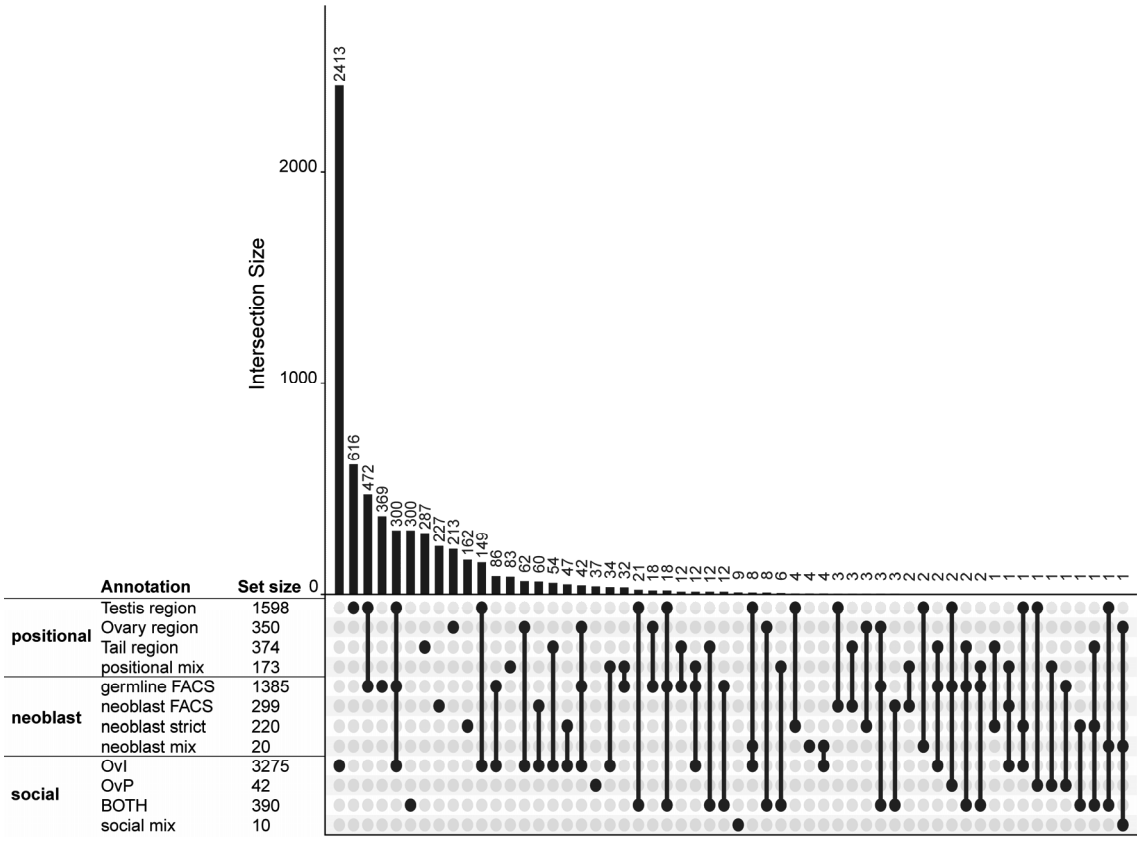
71. Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler A von, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.

72. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30:1188–95.

73. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.



Supporting Information



Additional file 9: **Fig. S1.** Upset plot of the intersection of orthogroup (OG) annotations from the positional, neoblast, and social datasets. The dots and lines on the bottom right show which intersection is represented by the bar plots above it. The size of intersections is given above the bar plot. To the left of the intersection diagram, the absolute number of OGs per annotation is given.

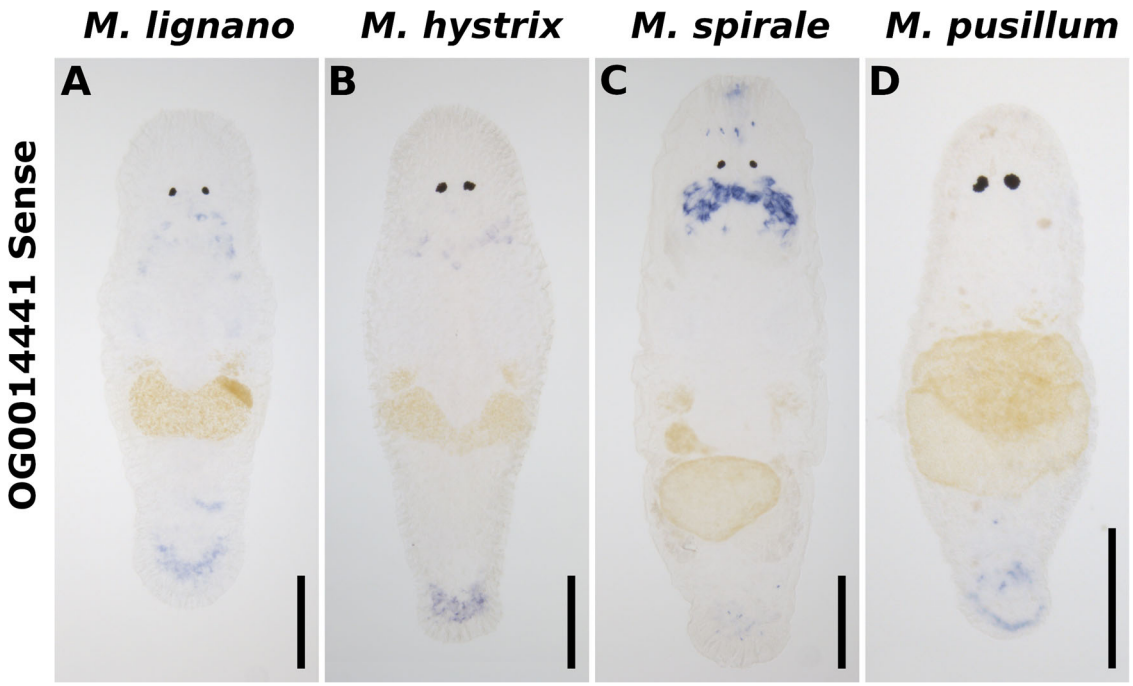
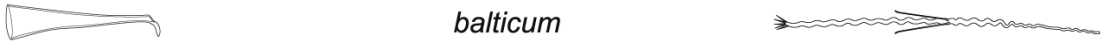


Fig. S2. Sense probe control ISH. Unspecific staining in pharyngeal glands and in the tail regions. Transcripts RNA815_7008 (*M. lignano)*, Machtx_20180703@G07456_i1 (*M. hystrix*), Macspi_20180703@G161928_i1 (*M. spirale*), and Macpus_20180703@G35224_i1 (*M. pusillum*) of the orthogroup OG0014441 were used for sense in situ probe generation. Scale bars: 100 µm.



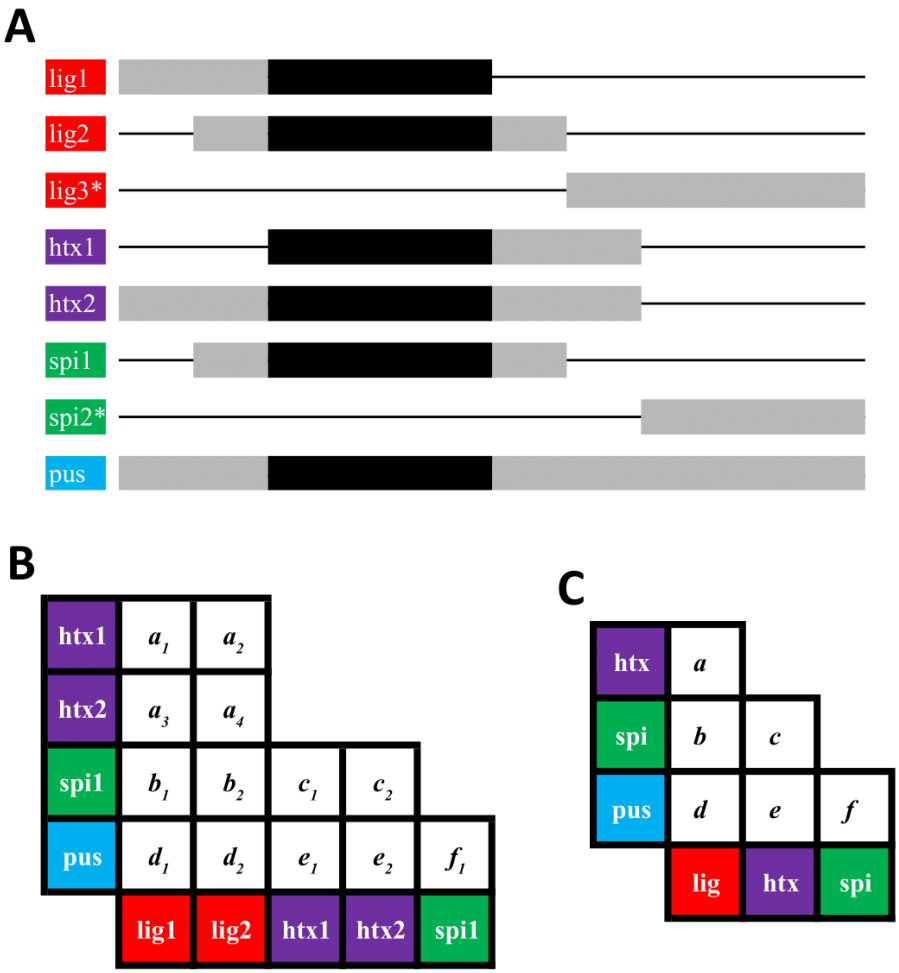


Fig. S3. Representation of the method used to estimate protein divergence for each OG. Species are abbreviated as: lig = *M. lignano*, htx = *M. hystrix*, spi=*M. spirale* and pus=*M. pusillum*. A: Hypothetical protein alignment of an OG containing all four species. Protein divergence was calculated between all sequences that share an aligned region (indicated in black), thus excluding sequences that do not overlap (indicated with an asterisk, i.e. lig3*, spi2*). B: Matrix of all pairwise comparisons between the overlapping sequences in the OG, with letters denoting divergences between particular species pairs (e.g. $a_{1\text{-}4}$ represent the protein divergence between the sequences of *M. lignano* and *M. hystrix*). C: Average protein divergences between species pairs in the OG (e.g. a is the average of all a-values in panel B). The divergence for the whole OG is then calculated as the average protein distance over all six species pairs.

Additional file 13: Reannotation of Mlig_37v3 transcriptome

Introduction

In this document, we explain how we transferred the annotations that were made in *M. lignano* based on RNA-Seq and ISH using previous transcriptome versions of that species to the most up-to-date transcriptome version used in this study. Specifically, the MLRNA110815 transcriptome (available at http://www.macgenome.org/download/MLRNA110815; Simanov 2014), which has been used by Arbore et al. (Arbore et al. 2015), Ramm et al. (Ramm et al. 2019), and many follow-up ISH screens, has not previously been linked to the Mlig_RNA_3_7_DV1.v3 transcriptome assembly, so we have performed DE using these data using our reduced version of this assembly (Mlig_37v3) as a reference. In addition, we have transferred gene expression pattern annotations from ISH probes developed from older transcriptomes to transcripts of the Mlig_37v3 transcriptome. Finally, the transcripts of the transcriptome used in Grudniewska et al. (Grudniewska et al. 2016), assembly MLRNA150904) have been linked to the transcripts in the Mlig_RNA_3_7_DV1.v3 transcriptome using a genome-guided approach (Grudniewska et al. 2018; with the method described in Wudarski et al. 2017). We were thus able to transfer the annotations of Grudniewska et al. (Grudniewska et al. 2016) to the Mlig_37v3 transcriptome. We used the Mlig_37v3 transcriptome because it was also used for the orthology detection and estimating expression directly on these transcripts, thus makes comparisons across species easier. The detailed annotation of Mlig_37v3 can be found in Additional file 10 Tab. S6.

Positional dataset

The positional dataset of Arbore et al. (Arbore et al. 2015) consists of four RNA-Seq libraries derived from animals cut into different fragments along the anterior-posterior body axis. The first sample was amputated anterior to the testes, the second sample anterior to the ovaries, the third sample was amputated posterior to the ovaries, and the fourth sample contained complete animals (fragments A-D, see main text Fig. 2). By comparing the expression in adjacent fragments (e.g. A vs. B), it is then possible to identify genes with higher expression in the body region contained only in the larger sample, thus allowing the identification of expression that is specific for the testis, ovary, and tail region.

DE analysis was performed with the same contrasts used in the original study (Arbore et al. 2015). Reads were trimmed using Trimmomatic, corrected using

Rcorrector, expression quantified using Salmon (version 0.9, Patro et al. 2017) in quasi-mapping mode with k-mer length 31, and then DE inferred using DESeq2 (Love et al. 2014). Due to the low quality of some 36bp single-end reads we employed a gentle trimming setting, a conservative minimum read length of 25, and an average quality score of 20 (2:30:10:8:TRUE LEADING:5 TRAILING:5 SLIDINGWINDOW:4:5 AVGQUAL:20 MINLEN:25). We calculated the expression level per body region as the percentage of reads mapping to a transcript to account for differences in sequencing depth among fragments, calculated identically to Arbore et al. 2015 as $\log_2$(((reads mapping to transcript/total reads mapping within the segment) x 100) + 0.00001). We then assigned transcripts to the positional classes as defined in Arbore et al. (Arbore et al. 2015).

In our reanalysis, we focused on genes with higher expression in the testis, ovary, and tail region but also annotated genes as non-DE or as "other" if the pattern did not match the previous classes. In Arbore et al. (Arbore et al. 2015) two types of ovary region genes were defined, those that were DE only in the comparison of the B vs. C fragment and those that were additionally DE in the A vs. B fragment. We think that this latter category may have resulted from some ovary tissue having ended up in the second fragment, in spite of the fact that Arbore et al. (Arbore et al. 2015) deliberately tried to cut those fragments somewhat anterior of the testis/ovary boundary, so as to avoid any ovary tissue being included. Moreover, this kind of pattern seems more likely for genes that are highly expressed in the ovary. Since both of these classes should yield ovary region candidates, we combined these two ovary classes into one class for our downstream analysis.

The reanalysis of the positional data resulted in a marked difference compared to the original results (Fig. A1A, Tab. A1). Arbore et al. (Arbore et al. 2015) classified 93.8% of all transcripts as not DE, while our analysis only assigned 80.7% to that class. Most of this difference resulted from the fact that our analysis assigned 9.7% of all transcripts to classes referred to by Arbore et al. (Arbore et al. 2015) as "other", while the value for their study was only 0.63%. We also assigned more transcripts to the testis, ovary, and tail regions, with the difference being most pronounced for one of the ovary classes and the tail region class (Tab. A1).

One source of disagreement appears to stem from the fact that transcripts with low expression are unreliably placed in the different classes. Indeed, when we excluded transcripts that had fewer than 50 reads mapped to them in the entire worm (called the D fragment in Arbore et al. 2015), the differences between the two analysis were greatly reduced, with only 343 (0.59%) transcripts assigned to the "other" classes and the absolute numbers for the different positional classes becoming similar between (Arbore et al. 2015) and our analyses (Fig. A2, Tab. A1).

Estimating expression of lowly expressed transcripts is notoriously difficult (Zhang et al. 2017), and this problem is compounded by the lack of replication in Arbore et

al. (Arbore et al. 2015), which makes the use of statistical tools to call differential expression impossible. Differences between the analyses are likely also due to the fact that Salmon (used in this study) and RSEM (used in Arbore et al. 2015), employ a different approach to calculate the effective length of transcripts when the transcript length is similar or smaller than the empirical fragment length (i.e. the insert size of the sequencing library). RSEM reduces the effective length of these transcripts to 1 and thus does not quantify their expression (Li and Dewey 2011), while Salmon still quantifies such short transcripts (Patro et al. 2017). To investigate this, we filtered our data to exclude all transcripts shorter than 300bp (with an empirical fragment length at 250bp on average), but this did not qualitatively change the observed differences between the analyses.

Finally, an important difference between the two analyses is that Arbore et al. used the options "--best --strata" when mapping the reads with bowtie. This option will only return the best alignment for each read and thus does not allow one to account for multi-mapping. Arbore et al. (Arbore et al. 2015) correctly state that RSEM can account for multi-mapping in the analysis, but this is actually prevented by their bowtie settings. To verify this, we mapped all the reads in the D fragment to the Mlig_37v3 transcriptome and found that indeed all reads mapped uniquely (data not shown).

In conclusion we can say that even though there are considerable differences between the analyses, they are mainly affecting low-expression transcripts, which should be treated with caution anyway, especially because we are dealing with an unreplicated experiment. We retained all transcripts in our further analysis since it is easier for investigators to filter our results rather than repeating them. However, we caution against the use of candidates without prior inspection of their read counts. To facilitate this, we provide a column in the results table of the reanalysis grouping transcripts by count number (Additional File 14 Tab. S8).

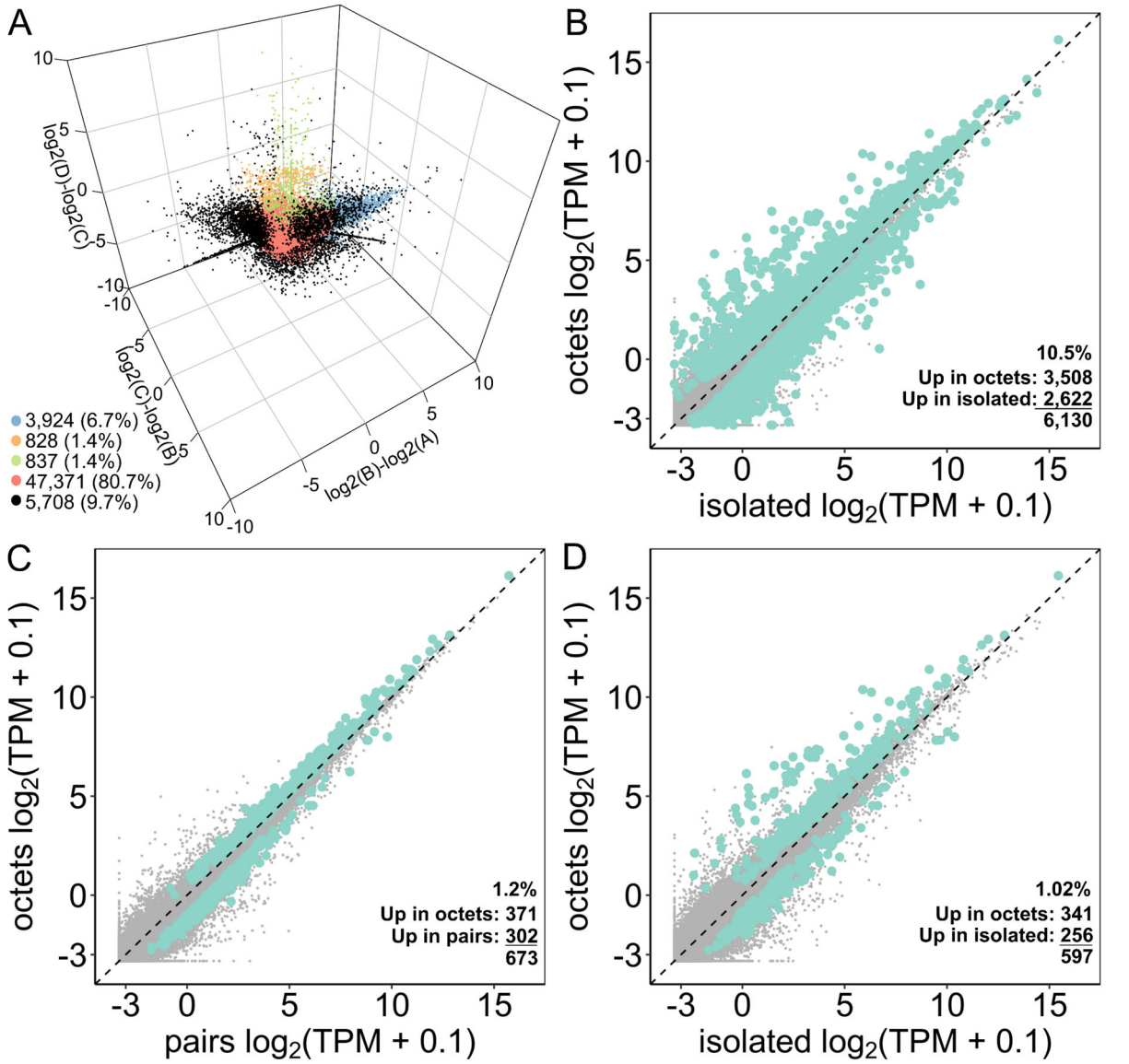
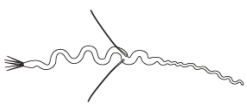


Fig. A1 Reanalysis of the positional and social RNA-Seq datasets. A: Differences in expression between different fragments for each gene. Each axis represents the *in silico* subtractions that were performed. Points are coloured by the positional classes to which the genes are assigned; blue – testis region, orange – ovary region, green – tail region, red – non-specific expression, black – other expression patterns. The inset numbers give the number and percentage of all genes belonging to each positional class. **B:** mean expression for transcripts in octets vs. isolated (OvI). **C:** mean expression for transcripts in octets vs. pairs (OvP). Coloured points in **B** and **C** show the genes with significantly different expression values. **D:** same points as in **B**, but coloured points show the transcripts that are differentially expressed both in OvI and OvP (BOTH). Inset values in the bottom right of panels **B**, **C**, and **D** give the number of genes that are more highly expressed in isolated, pairs or octets, as well as the total number of genes that were called as DE in each category. The values in the top left of **B**, **C**, and **D** give the percentages of DE genes out of the total (58,668).

sp. 2



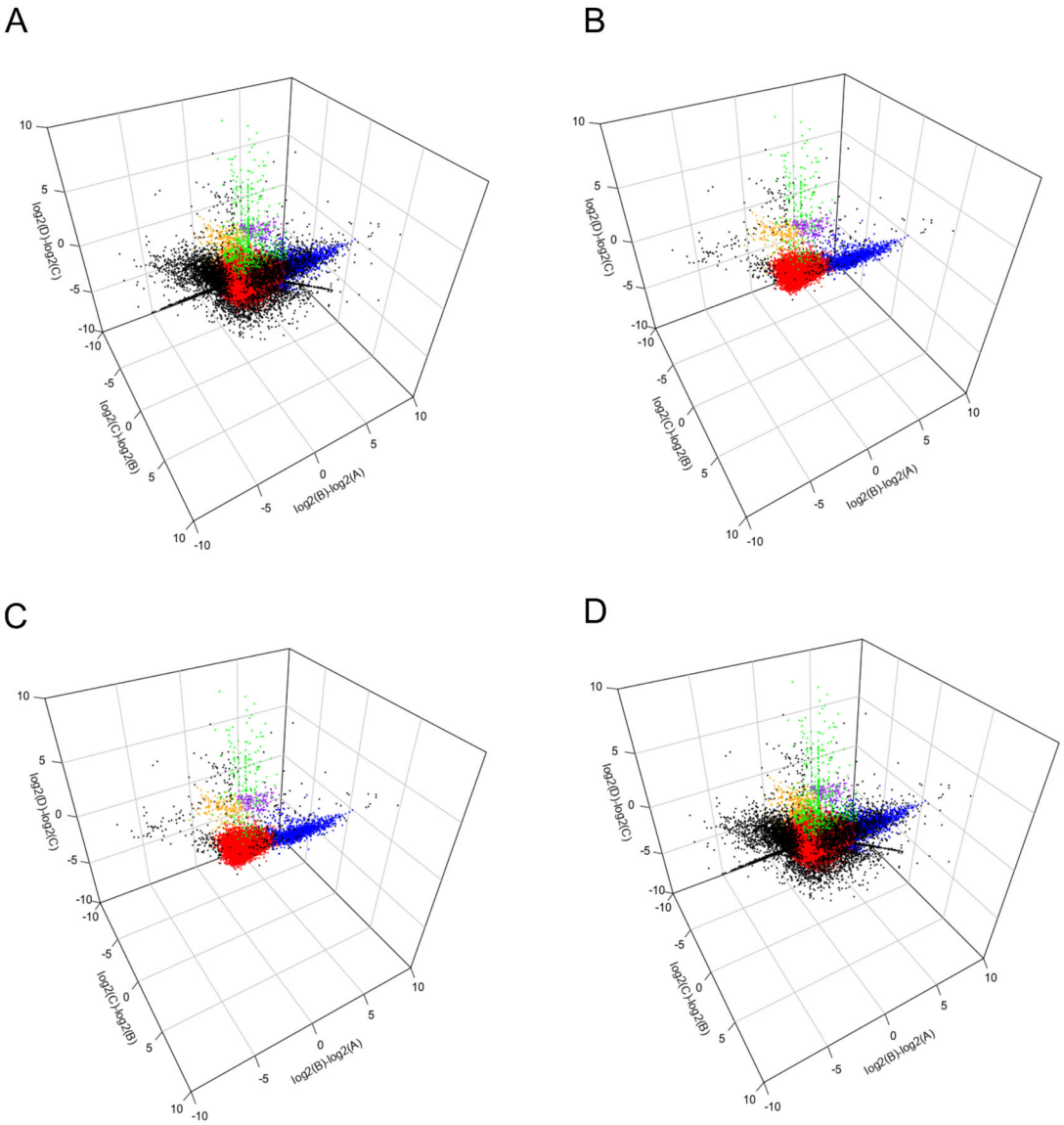


Fig. A2 Reanalysis of positional RNA-Seq dataset with different filtering options. Each plot shows the differences in expression between different fragments for each transcript. Each axis represents the *in silico* subtractions that were performed. **A:** All transcripts are shown. **B:** transcripts are subset to exclude those with low expression (<50 counts in the whole worm). **C:** transcripts are subset to exclude those with low expression and lengths <300bp. **D:** Transcripts are subset to exclude those with lengths < 300bp. See Table A1 for counts of genes in **A** and **B**.

Tab. A1 Counts and proportions of genes in each positional annotation class as defined in the original Arbore et al., 2015 paper. Counts in whole animal: Subset of data given by thresholds of the read counts in the "D fragment" (I.e. the whole worm) , meanBA: Mean difference in expression values between B and A fragments, meanCB: Mean difference in expression values between C and B fragments, meanDC: Mean difference in expression values between D and C fragments, Class: The code for the positional annotation category (as defined in Arbore et al., 2015), class_name: The name of the positional annotation category (as defined in Arbore et al., 2015), count: The count of the number of transcripts in this positional category, % of total: The percent of the total number of Mlig_37v3 transcripts (N = 58,668) that occur in the expression category, % of subset: The percent of the total within level of the "D_count_bin" column that occur in the expression category

Counts in whole animal	class	class_name	meanBA	meanCB	meanDC	count	% of total	% of subset	Arbore counts	Arbore % of total
[0 – max]	0,0,0	non_diff	-0.24	-0.12	0.02	47,371	80.74	80.74	70,064	93.78
	+,0,0	testes	4.10	-0.28	-0.35	3,924	6.69	6.69	3,360	4.50
	+,+,0	ovary2	3.11	3.82	0.43	187	0.32	0.32	127	0.17
	0,+,0	ovary1	-0.10	3.22	-0.04	641	1.09	1.09	323	0.43
	0,0,+	tail	-0.09	-0.41	3.33	837	1.43	1.43	366	0.49
	Other	other	-0.91	-0.25	-0.02	5,708	9.73	9.73	468	0.63
[0 – 49]	0,0,0	non_diff	-0.23	-0.19	-0.03	24,431	41.64	75.31		
	+,0,0	testes	3.29	-0.40	-0.34	1,559	2.66	4.81		
	+,+,0	ovary2	2.64	2.35	-0.87	13	0.02	0.04		
	0,+,0	ovary1	-0.33	2.82	-0.29	426	0.73	1.31		
	0,0,+	tail	-0.11	-0.52	2.70	645	1.10	1.99		
	Other	other	-0.92	-0.28	-0.19	5,365	9.14	16.54		
[50 – max]	0,0,0	non_diff	-0.24	-0.05	0.07	22,940	39.10	87.46		
	+,0,0	testes	4.64	-0.20	-0.36	2,365	4.03	9.02		
	+,+,0	ovary2	3.15	3.93	0.53	174	0.30	0.66		
	0,+,0	ovary1	0.37	4.01	0.46	215	0.37	0.82		
	0,0,+	tail	0.00	-0.05	5.43	192	0.33	0.73		
	Other	other	-0.77	0.24	2.61	343	0.58	1.31		

Neoblast dataset

Grudniewska et al. (Grudniewska et al. 2016) used two strategies to identify transcripts upregulated in proliferating cells. First, they gamma-irradiated adult worms, thus killing all proliferating cells (De Mulder et al. 2010) and compared expression before and after irradiation to find genes with lower expression after irradiation. Second, they sorted cells stained for DNA from adults, hatchlings, and amputated heads using fluorescence-activated cell sorting (FACS), with gates designed to capture cells with a 2C and 4C DNA content. Cells with a 4C DNA content are late S-, G2-, and M-phase cells and thus proliferating. They compared expression between cells collected with the 2C and 4C gates to find genes upregulated in 4C cells. By then further comparing expression between 4C cells of adults against 4C cells of hatchlings and heads, they divided these candidates into genes with biased expression in germline stem cells or neoblasts (somatic stem cells). Grudniewska et al. (Grudniewska et al. 2016) thus, used three annotations: germline-specific genes identified using FACS sorting (germline_FACS) and two groups of neoblast specific genes (neoblast_FACS and neoblast-strict). The first neoblast group was also identified using FACS sorting only, but the second group was cross-referenced with irradiation results, and they concluded that these are the most stringent neoblast candidates.

Here we detail how we transferred the neoblast dataset annotations to the Mlig_37v3 transcriptome. Grudniewska et al. (Grudniewska et al. 2016) performed DE analysis on transcripts clustered with Corset, which likely represent isoforms. For each transcript per Corset cluster, we first searched for the corresponding transcript(s) in the Mlig_RNA_3_7_DV1.v3. coregenes.bestORF.pep file (this corresponds to the transcriptome used in this study prior to CD-HIT clustering; see section “Transcriptomes used for orthology detection and DE analysis” in the main text) using the annotation provided by (Grudniewska et al. 2018) and merged the annotation into the Mlig_37v3 transcript representing these hits. This was done to avoid losing annotations because an annotated Mlig_RNA_3_7_DV1.v3.coregenes.bestORF.pep transcript was clustered with an unannotated one by CD-HIT. When several transcripts were linked to one Mlig_37v3 transcript, we concatenated all annotations, potentially resulting in multiple annotations for the same transcript.

The majority of annotated transcripts could be assigned to an OG (Tab. A2). A small proportion of transcript clusters did not result in an annotation, either because we were not able to identify the transcripts in our Mlig_37v3 transcriptome or because multiple clusters were linked to the same Mlig_37v3 transcript (Tab. A2). But the majority of transcript clusters that did not result in an annotation did so because they were not assigned to an OG by Orthofinder (Tab. A2).

Since multiple transcripts could hit the same Mlig_37v3 transcript and then again, multiple Mlig_37v3 transcripts could be assigned to the same OGs, this resulted in some overlap between annotations. When an OG contained transcripts with a neoblast_FACS and transcripts with a neoblast-strict annotation, we annotated the OG as neoblast-strict. 20 OGs contained transcripts annotated as neoblast (neoblast_FACS and neoblast-strict) and also transcripts with the germline_FACS annotation. We annotated these OGs as neoblast_mix and excluded them from downstream analysis.

Tab. A2 Summary of the annotation transfer from the neoblast dataset. For the three types of annotations we transferred, we list the initial number of transcript clusters annotated by Grudniewska et al. (2016), followed by the total number of transcripts the clusters contained. We further list how many of these transcripts were represented in the Mlig_RNA_3_7_DV1.v3.coregenes assembly of Grudniewska et al. (2018) and how many transcripts in the Mlig_37v3 assembly were annotated based on this. The second to last row indicates how many transcripts in Mlig_37v3 were assigned to an OG with the percentage of transcripts in brackets. Finally, many transcripts were placed in the same OG and thus we give the total number of OGs annotated from the neoblast dataset. Note that if transcripts were annotated from multiple sources, we have counted them in each category with the exception of the final row, where these OGs were designated neoblast_mix as indicated by the last column.

No. of transcript clusters, transcripts or OGs	**germline_FACS**	**neoblast_FACS**	**neoblast-strict**	**neoblast_mix**
Initially annotated clusters	2,739	567	357	
Transcripts in the clusters	3,760	759	494	
Hit in Mlig_RNA_3_7_DV1.v3.coregenes	3,747	757	493	
Annotated transcripts in Mlig_37v3	3,429	716	405	
Assigned to an OG	2,650 (77.3%)	452 (63%)	371 (91.6%)	
Final OGs	1,385	299	220	20

Social dataset

The social dataset of Ramm et al. (Ramm et al. 2019) consists of four treatments: isolated animals (I), animals that were isolated but then joined with a partner for 24h prior to RNA extraction (J), animals kept in pairs (P), and animals kept in octets (O). Each treatment consists of four biological replicates containing the RNA of 57 worms on average. In our reanalysis, we excluded the J treatment and compared expression between octets and isolated as well as octets and pairs.

The social dataset consists of high-quality, paired-end 100bp reads, which were trimmed using Trimmomatic, corrected using Rcorrector, expression quantified using Salmon (version 0.9, Patro et al. 2017) in quasi-mapping mode with k-mer length 31 and then DE inferred using DESeq2 (Love et al. 2014). Using these analyses, we could annotate transcripts as DE in octets vs. isolated (OvI), in octets

vs. pairs (OvP), or in both comparisons (BOTH). We performed minimal pre-filtering to remove genes with no expression in any sample (i.e. a count of 0 in all samples), keeping ~97% of all transcripts for both the OvI and OvP comparisons. To correct for multiple testing, we set our false discovery rate (FDR) at a conservative 0.05. We used the independent filtering feature of DESeq2, which had not yet been implemented in the DESeq version used by Ramm et al. (Ramm et al. 2019). We also conducted the analysis with an FDR of 0.1, and without the independent filtering since this corresponds more closely to the analysis in Ramm et al. (Ramm et al. 2019).

Overall, the proportion of DE transcripts for the OvI comparison are similar to the previously reported figures (i.e. 10.5%, compared to the 9.9% reported by Ramm et al. (Ramm et al. 2019); Figure 1B; Additional file 15: Tab. S9). Breaking the data down into patterns for each tissue region class, the absolute numbers of DE transcripts are also qualitatively similar to the previous values (Fig. A3, Tab. A3). We sought to determine whether the remaining differences could be explained by our analysis decisions. Specifically, we retrieved results for the OvI comparison using an FDR threshold of 0.1 and without applying the independent filtering in DESeq2. Although the values obtained with these settings are in some cases more similar to those reported by Ramm et al., they do not substantially differ, and we conclude that these decisions did not have a meaningful impact on the results (Tab. A3).

Fewer transcripts were found to be significantly DE in the OvP comparisons (Fig. A1C, Additional file 16: Tab. S10), and the overlap of the OvI and OvP datasets (i.e. BOTH) was quite substantial (Fig. 1AD). These results imply a relatively small number of transcripts that are exclusively DE in the OvP comparison (i.e. transcripts that respond in expression to an increase in social group size, rather than to the presence/absence of a mating partner). Since only a few transcripts were DE in OvP only, we focussed on transcripts in the OvI and the BOTH groups in our downstream analysis.

Similar to the neoblast dataset, we had some overlap between OG annotations because multiple Mlig_37v3 transcripts could be assigned to the same OG. When at least one transcript in an OG had a BOTH annotations, we annotated the OG as BOTH since this indicates that at least one transcript is sensitive to mating opportunity and the intensity of sperm competition. Ten OGs contained transcripts that were annotated OvI and additional transcripts annotated as OvP. It would be possible to annotate these OGs as BOTH but instead, we annotated them as social_mix to highlight that no single transcript within this OG had the BOTH annotation and excluded these OGs from downstream analysis.

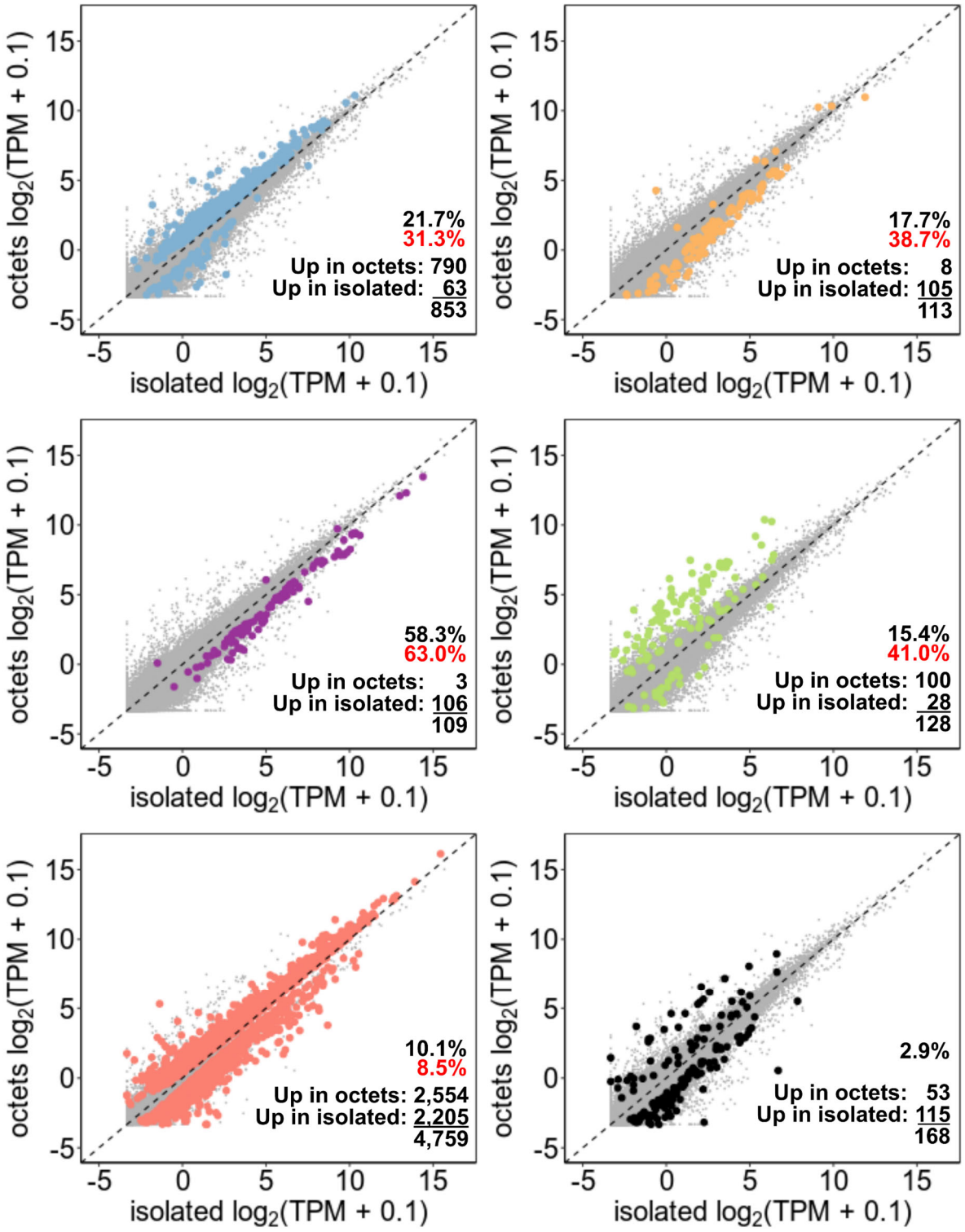
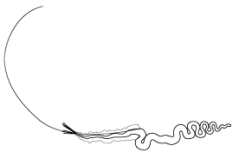


Fig. A3 Comparison of mean expression for transcripts in octets vs. isolated. Each panel shows genes called as DE in this comparison and also occur in the positional classes from the positional dataset. The colours are; blue – testis, orange – ovaries class 1, purple – ovaries class 2, green – tail, red – non-specific expression, black – other expression patterns. The inset values in the bottom right of each panel give the number of genes that are more highly expressed in octets and isolated worms, respectively, as well as the total number of genes that were called as DE from each positional class. The values in black in the top left of each panel give the percentage of genes from each positional class that were called as DE, in red are the percentages from the original study of [3] for comparison.

sp. 30

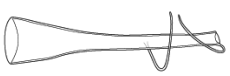


Tab. A3 Counts of DE genes from an analysis of worms kept in octets vs worms kept isolated (OvI). In order to compare results more closely to those of Ramm et al. (Ramm et al. 2019), the analysis was also performed using an FDR threshold of 0.1 (as in Ramm et al., (Ramm et al. 2019), the default of DESeq2) and without the independent filtering procedure of DESeq2. Given are the total number of tested transcripts in each group, the number and percentage that were either higher in octets or isolated, as well as the percentage of the total that were called as DE. We also give the breakdown of genes assigned to different tissue regions that are called as DE in the OvI comparison. Where possible, the corresponding overall proportions reported in Ramm et al. (Ramm et al. 2019) are also reported in the final column.

Category	Count	Higher in octets (%)	Higher in isolated (%)	% DE	% DE (from Ramm et al., 2019)
Total	56,915	3,938 (6.9)	3,200 (5.6)	12.5	9.9
testes region	3,924	875 (22.3)	85 (2.2)	24.5	31.3
ovary 1 region	641	12 (1.9)	117 (18.3)	20.1	38.7
ovary 2 region	187	3 (1.6)	112 (59.9)	61.5	63.0
tail region	837	111 (13.3)	41 (5.9)	18.2	41.0
non-specific	47,371	2,861 (6.0)	2,673 (5.6)	11.6	8.5
Other	5,708	76 (1.3)	172 (3.0)	4.3	-

ISH data

Because most ISH studies have been conducted with the previous transcriptome assembly, we took a mapping approach to annotate the new *M. lignano* transcriptome using existing primer information. Primers used for designing ISH probes were collected from the literature (Arbore et al. 2015; Lengerer et al. 2017; Weber et al. 2018; Ramm et al. 2019) and from personal correspondence with authors. Specifically, we collected information on i) the original assembly and transcript that was used to design the primers, ii) the estimated length of the PCR product, iii) the forward and reverse primer sequences, and iv) the ISH pattern observed. We focused on reproduction-related genes that showed expression in the gonads, prostate glands, and the antrum. In total 247 sets of primers representing 204 original transcripts were collected for which complete information was available (Additional file 18: Tab. S11). Many transcripts (92) had more than one set of primers. We mapped primer pairs to the reduced *M. lignano* transcriptome assembly (Mlig_37v3) with bowtie2 (version: 2.3.3.1; Langmead and Salzberg 2012). Maximum and minimum insert sizes were set as the original estimated product length +/- 100bp. No mismatches were allowed in the seed alignment. Mismatches may occur in the final alignment; three alignments had two mismatches, two alignments had indels with respect to the Mlig_37v3, and no alignments had more than two mismatches. The resulting .sam file was parsed to identify primer pairs that mapped



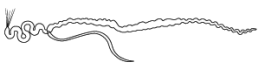
concordantly to only a single transcript. Because gene models are presumably improved in the newer transcriptome assembly and new isoforms may be identified, we treated each pair of primers as independent, even if they were designed from the same original transcript. Note that several primer sets have multiple annotations and the Mlig_37v3 transcripts inherited all annotations.

For those primer annotations that could not be transferred in the above way, we used blastn (version: 2.6.0; Camacho et al. 2009) to identify the best hit in Mlig_37v3 of the original source transcript. We performed a one-way blastn with Mlig_37v3 as the database and MLRNA110815 as the query. We then ranked all hits first by e-value and used the bit score to break ties. For each MLRNA110815 transcript with ISH primer pairs that had not already been mapped to a Mlig_37v3 transcript, we identified the best blast hit and assigned the observed ISH pattern as an annotation to the Mlig_37v3 transcript.

Between 42.9 and 57.1% of primer pairs could be mapped to the Mlig_37v3 transcriptome (Tab. A4) and additional annotations could be transferred to Mlig_37v3 by blastn. The number of Mlig_37v3 transcripts that received an annotation was lower than the number of source transcripts. This is because i) different primer pairs from the same source transcript often mapped to the same Mlig_37v3 transcript, and ii) primer pairs from different source transcripts also often mapped to the same Mlig_37v3 transcript. For example, six separate primer pairs (RNA815_14562.a, RNA815_14562.b, RNA815_18395.a, RNA815_18395.b, RNA815_39625.b, and RNA815_64228.b) map concordantly to one Mlig_37v3 transcript (Mlig005822.g1, see also Additional file 18: Tab. S11). This Mlig_37v3 transcript was also the best blastn hit of another transcript (RNA815_42719), resulting in seven annotations of this single Mlig_37v3 transcript. The number of Mlig_37v3 transcripts that were finally assigned to each annotation category is given in Tab. A4. Note that some Mlig_37v3 transcripts will have multiple annotations and are represented multiple times, in total 147 Mlig_37v3 transcripts are annotated (Additional file 18: Tab. S11). The annotated transcripts occur in 77 orthogroups (Additional file 8: Tab. S5). The majority of orthogroups have representatives from all species except for those with antrum and prostate annotations (Tab. A4). While the sample size is quite small, this pattern is expected, since especially prostate transcripts are assumed to be under positive selection due to antagonistic coevolution.

Tab. A4 Summary of the annotation transfer from ISH patterns in *M. lignano*. Counts are given for each annotation category of how many primer pairs and transcripts there are, as well as how many could be transferred by primer mapping. Also given are the number of Mlig_37v3 transcripts and OGs that receive and annotation as well as how many of these OGs contain all species. Where relevant, the transcriptome assembly is noted in brackets. Note that some transcripts and orthogroups (OGs) are counted twice because they have multiple annotations.

Annotation	**Testes**	**Ovaries**	**Gonads**	**Antrum**	**Prostate**
No. primer pairs (MLRNA110815)	21	28	46	25	148
No. original transcripts (MLRNA110815)	20	27	43	22	110
No. (%) primer pairs that map	12 (57.1)	12 (42.9)	25 (54.3)	12 (54.5)	63 (43.6)
No. transcripts that receive an annotation from primer mapping (Mlig_37v3)	11	16	23	10	40
No. transcripts that receive an annotation from blastn (Mlig_37v3)	5	9	14	10	20
No. OGs that receive an annotation	14	19	24	11	16
No. (%) OGs that contain all species	9 (64.3)	14 (73.7)	14 (58.3)	5 (45.5)	7 (43.8)



References Supporting Information

1. Simanov D. Genomic resources for the flatworm model organism Macrostomum lignano. PhD dissertation. Utrecht University; 2014.

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

3. Ramm SA, Lengerer B, Arbore R, Pjeta R, Wunderer J, Giannakara A, et al. Sex allocation plasticity on a transcriptome scale: Socially sensitive gene expression in a simultaneous hermaphrodite. Mol Ecol. 2019;00:1–21.

4. Grudniewska M, Mouton S, Simanov D, Beltman F, Grelling M, Mulder K de, et al. Transcriptional signatures of somatic neoblasts and germline cells in Macrostomum lignano. eLife. 2016;5:e20607.

5. Grudniewska M, Mouton S, Grelling M, Wolters AHG, Kuipers J, Giepmans BNG, et al. A novel flatworm-specific gene implicated in reproduction in Macrostomum lignano. Sci Rep. 2018;8:3192.

6. Wudarski J, Simanov D, Ustyantsev K, de Mulder K, Grelling M, Grudniewska M, et al. Efficient transgenesis and annotated genome sequence of the regenerative flatworm model Macrostomum lignano. Nat Commun. 2017;8:2120.

7. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nat Methods. 2017;14:417–9.

8. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

9. Zhang C, Zhang B, Lin L-L, Zhao S. Evaluation and comparison of computational tools for RNA-seq isoform quantification. BMC Genomics. 2017;18:583.

10. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.

11. De Mulder K, Kuales G, Pfister D, Egger B, Seppi T, Eichberger P, et al. Potential of Macrostomum lignano to recover from γ-ray irradiation. Cell Tissue Res. 2010;339:527–42.

12. Lengerer B, Wunderer J, Pjeta R, Carta G, Kao D, Aboobaker A, et al. Organ specific gene expression in the regenerating tail of Macrostomum lignano. Dev Biol. 2017;433:448–60.

13. Weber M, Wunderer J, Lengerer B, Pjeta R, Rodrigues M, Schärer L, et al. A targeted in situ hybridization screen identifies putative seminal fluid proteins in a simultaneously hermaphroditic flatworm. BMC Evol Biol. 2018;18:81.

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

15. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

Chapter III

Frequent shifts to hypodermic insemination and correlated evolution of sperm and genital morphology in a genus of flatworms

Manuscript in preparation as: Frequent shifts to hypodermic insemination and correlated evolution of sperm and genital morphology in a genus of flatworms. Jeremias N. Brand, Gudrun Viktorin, Christian Beisel, Luke J. Harmon and Lukas Schärer

Abstract

During hypodermic insemination, the (sperm) donor injects an ejaculate through the epidermis of the (sperm) recipient using a traumatic intromittent organ and sperm subsequently move through different tissues to the site of fertilization. This striking mating behaviour likely evolves to resolve sexual conflict between donor and recipient over the fate of the transferred ejaculate, since it allows the donor to circumvent the recipient's genitalia, thereby reducing the potential for cryptic female choice. Hypodermic insemination may evolve more readily in hermaphrodites because costs an individual incurs as a recipient could be directly compensated by also acting as a donor. While some striking examples of this behaviour have been studied, our understanding of the frequency of its evolution, and the routes this takes, remain poorly understood. We present comparative work on convergent evolution of hypodermic insemination in a genus of free-living hermaphroditic flatworms, by collecting morphological information on sexual traits of 145 *Macrostomum* species and analysing them in the context of a robust molecular phylogeny. The results show at least nine, but up to 13 convergent transitions to hypodermic insemination in *Macrostomum* and we describe a suite of morphological changes that are associated with these transitions. Particularly striking are changes in different aspects of the sperm design that likely represent functional adaptations to the need for sperm to move within the tissues of the recipient, as opposed to adaptations of the sperm in copulating species that are linked to their function inside of the female genitalia. Moreover, sperm of species with hypodermic insemination are much shorter, either to aid in their movement through tissue or because the circumvention of the recipient's sperm receiving organ alters the mode of sperm competition. Further, we present evidence that transitions to hypodermic insemination may initially occur through internal wounding of the female sperm receiving organ and we discuss several morphological traits likely important in the evolution of the genus.

Introduction

There frequently is a difference in mating propensity between the sexes, because male fecundity is often limited by the number of matings a male can achieve, while female fecundity is often limited by the number of resources a female can invest into offspring (Bateman 1948; Arnold 1994). The resulting sexual conflict over the mating rate can therefore lead to antagonistic coevolution between female resistance traits and male persistence traits (Rice 1996; Arnqvist and Rowe 2002, 2005). Moreover, such antagonistic coevolution is not restricted to precopulatory traits, but often also involves interactions of the female genital tract with the male intromittent organs and the received ejaculate (Birkhead and Pizzari 2002; Wedell and Hosken 2010; Hare and Simmons 2019). Such coevolution can drive the emergence of male traits that inflict considerable harm to females (Arnqvist and Rowe 2002; Morrow and Arnqvist 2003; Morrow et al. 2003).

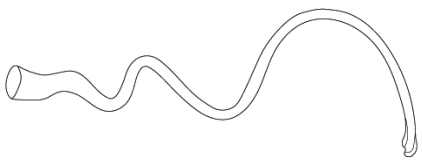
A striking example that implicates such traits is traumatic mating, which occurs in some internally fertilizing species and involves the infliction of a wound to the integument of the recipient (traumatic penetration) through which the donor then transfers seminal fluids (traumatic secretion transfer) or sperm (traumatic insemination, which can also include the transfer of seminal fluids) (Lange et al. 2013). Since traumatic mating occurs frequently in both gonochoristic (separate-sexed) and hermaphroditic species (Lange et al. 2013), we in the following use the more general terms (sperm) donor and (sperm) recipient to refer to the two sexual roles, with no loss of generality (Schärer et al. 2015). Although traumatic mating can result in costs to the recipients (Morrow and Arnqvist 2003; Reinhardt et al. 2003, 2015; Benoit et al. 2012; Lange et al. 2013; Tatarnic 2018) it has evolved repeatedly across a range of animal groups (Lange et al. 2013). While natural selection might play a role in its evolution in some taxa—especially the endoparasitic *Strepsiptera* (Tatarnic et al. 2014; Kathirithamby et al. 2015)—it has been suggested that it evolves because it allows donors to bypass pre- and post-copulatory mechanisms of the choice by the recipient. Specifically, traumatic insemination can allow donors to force copulation (Morrow and Arnqvist 2003), bypass the recipient's genitalia and deposit sperm closer to the site of fertilization (Kathirithamby et al. 2015; Peinert et al. 2016) or even directly within the relevant tissue (Stutt and Siva-Jothy 2001; Morrow and Arnqvist 2003), and thus minimize the control that the recipient can exert over the received ejaculate (Charnov 1979; Lange et al. 2013). In this view, hypodermic insemination represents a form of sexual conflict resolution, since it allows the donor to avoid sexual conflicts over the fate of the ejaculate it transferred to the recipient (though this, of course, will often lead to new sexual conflicts).

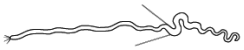
Alternatively, traumatic insemination could also evolve due to sperm competition, since in many internally fertilising species sperm of unrelated donors compete within

the female genital tract for fertilisation of the recipient's eggs (Birkhead and Møller 1998; Parker 1998). Traumatic insemination might allow donors to bias sperm competition in their favour and prevent competing donors from removing their donated sperm. There is some evidence for this in a family of spiders, where sperm precedence is biased towards the first male in a species with traumatic insemination, while it is biased towards the second male in its non-traumatically mating relatives (Řezáč Milan 2009; Lange et al. 2013; Tatarnic et al. 2014). Interestingly, since traumatic insemination likely also affects important factors shaping sperm competition (i.e. cryptic female choice, sperm removal and sperm density), it could also influence the mode of sperm competition. Specifically, by reducing the scope for cryptic female choice and by making sperm removal by competitors more difficult, it could eventually lead to more fair-raffle like sperm competition, leading to a reduction in sperm size (Schärer and Janicke 2009; Schärer et al. 2011).

Traumatic insemination is expected to evolve more readily in hermaphrodites because costs incurred by a recipient can potentially be compensated by direct benefits from the same individual acting as a donor. A hermaphrodite may thus also engage in matings that are involve harmful effects, as long as the net benefits for engaging in mating are positive (Michiels 1998; Michiels and Koene 2006). Additionally, sexual conflict in hermaphrodites could be more intense due to conflict over the mating role in addition to the mating rate (Charnov 1979; Michiels 1998; Anthes 2010; Schärer et al. 2015). Indeed, traumatic mating is more common in simultaneous hermaphrodites with a sevenfold higher incidence of transitions to traumatic mating in hermaphroditic compared to gonochoristic taxa (14 out of 36 cases: Lange et al. 2013 even though hermaphrodites make up approximately ~6% of animals Jarne and Auld 2006). Hermaphroditic animals are thus ideal targets for comparative studies investigating the evolution of traumatic mating, since—while it has been studied in some charismatic systems (Michiels and Newman 1998; Morrow and Arnqvist 2003; Koene and Schulenburg 2005; Koene et al. 2005; Kamimura 2007; Tatarnic and Cassis 2013; Peinert et al. 2016)—we know little about how frequently it evolves or what routes this takes when it originates (Lange et al. 2013; Reinhardt et al. 2015; Tatarnic 2018).

Here we present comparative work on the evolution of traumatic insemination across the genus *Macrostomum*, a species-rich taxon of hermaphroditic free-living flatworms. In *Macrostomum*, traumatic insemination is generally called hypodermic insemination (HI), since in several species the donor uses a needle-like stylet (Figure 1) to inject sperm through the epidermis of the mating partner and sperm then move through the body of the recipient to the site of fertilization (Schärer et al. 2011; Ramm et al. 2012, 2015). Injected sperm can readily be observed inside the parenchymal tissues of these highly transparent animals using a light microscope (Schärer et al. 2011; Ramm et al. 2012, 2015; Winkler and Ramm 2018), making it feasible to screen a large number of species for the convergent evolution of HI. While





we here present evidence that not all traumatically mating *Macrostomum* species inject sperm though the external epidermis, we will nevertheless use the term HI throughout for consistency with the *Macrostomum* literature.

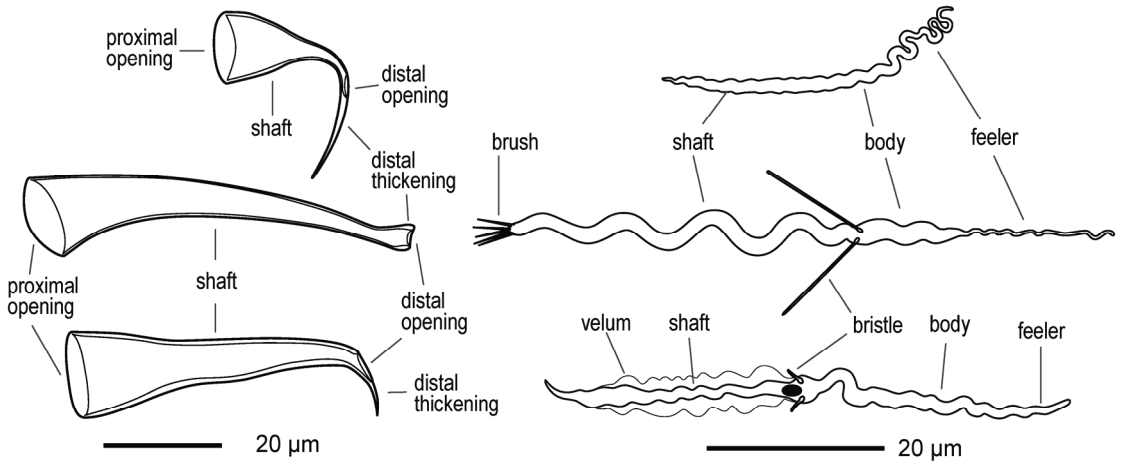


Figure 1 Representative drawings of stylets (left) and sperm (right) of three Macrostomum species. Top: M. sp. 92, a hypodermically mating species from the hypodermic clade, with a typical needle-like stylet and a simple sperm morphology. Middle: The well-studied model organism M. lignano with the typical morphology for reciprocally-mating species, showing a stylet with blunt distal thickenings and a complex sperm with an anterior feeler, bristles, and a terminal brush. Bottom: M. sp. 9 representing one of the convergent origins of hypodermic insemination in the reciprocal clade, showing a stylet with a highly asymmetric and sharp distal thickening and sperm with reduced sperm bristles, no brush, but a thin velum along the shaft. Note, that it is difficult to delimit all of the sperm morphology traits originally defined in M. lignano in some of the other species.

The genus consists of two phylogenetically well-separated clades, with one clade thought to only contain species that exclusively mate through HI ("hypodermic clade", referred to as Clade 1 in Schärer et al. 2011) and a second clade primarily containing reciprocally mating species ("reciprocal clade", referred to as Clade 2 in Schärer et al. 2011). However, within the reciprocal clade, HI has previously been shown to have convergently evolved in *M. hystrix* (Schärer et al. 2011). During reciprocal copulation the worms engage in a copulatory handshake where both partners insert their male intromittent organ (stylet) via the female genital opening into the female sperm storage organ (female antrum) of the partner, so that both can donate and receive sperm in the same mating (Schärer et al. 2004). After copulation, individuals in many of these species place their mouth over their own female genital opening and suck (Singh et. al. in preparation; Schärer et al. 2011, 2020), presumably in an attempt to remove components of the received ejaculate from the female antrum (termed the suck behaviour; Schärer et al. 2004; Vizoso et al. 2010). Removal of ejaculate could target manipulative seminal fluids, as it has been shown that the ejaculate of *M. lignano* contains substances that influence the mating partners propensity to perform the suck behaviour (Patlar et al. 2020; Weber et al. 2020). And it could serve to reduce the number of stored sperm (e.g. to lower the risk of polyspermy), constitute a form of cryptic female choice (e.g. to favour donors of

higher quality), and/or represent a resistance trait in sexual conflict over mating roles (i.e. to reverse unwanted sperm receipt). If the suck behaviour evolved due to sexual conflict it would represent a resistance trait of the recipient, so that we could expect the evolution of persistence traits of the donor, potentially leading to antagonistic coevolution (Arnqvist and Rowe 2005). Indeed, the sperm of copulating species generally have a thin anterior feeler and two stiff lateral bristles that have been suggested to represent such persistence traits, since they may permit sperm to anchor themselves in the female antrum to prevent being removed during the suck behaviour (Figure 1) (Vizoso et al. 2010; Schärer et al. 2011). In contrast, sperm of the species performing HI (i.e. the hypodermic clade and *M. hystrix*) lack lateral bristles and have a simplified morphology, presumably because they no longer need to persist against the suck behaviour (Vizoso et al. 2010; Schärer et al. 2011), and may instead be adapted to efficiently moving through the tissues of the partner (Figure 1). And while species with reciprocal copulation have a female antrum with a thickened epithelium, the species with HI were shown to have a simple female antrum, presumably because it no longer interacts with the donor's stylet and sperm, and instead is only used for egg laying (Schärer et al. 2011). Based on these findings, reciprocal copulation and HI have previously been described as representing two different mating syndromes, i.e. the reciprocal and hypodermic mating syndrome, respectively, since they each constitute a specific combination of morphological (sperm, stylet and female antrum) and behavioural traits (Schärer et al. 2011).

If HI evolves as a resolution of sexual conflict over the mating roles, then we would expect it to occur frequently, but it is unclear whether HI has convergently evolved more than once within the reciprocal clade, since little molecular data is available for *Macrostomum* and description of essential structures and mating behaviours is often neglected in the literature. We have assembled molecular and morphological information on 145 *Macrostomum* species with the aim to identify additional independent origins of HI and to quantitatively assess the convergent changes in sperm design and genital morphology that accompany shifts to HI. Further, we highlight several *Macrostomum* species that represent possible transitional states in the evolution to HI. And to investigate the different demands that HI might place on sperm, we check if shifts to HI result in changes in sperm size. Finally, we investigate antagonistic coevolution of reproductive traits within species showing the reciprocal mating syndrome, by checking for macroevolutionary correlations between different traits and by discussing the emergence of additional morphological traits that could constitute previously unknown resistance or persistence traits in this genus.

Results

Species collected and phylogenetics

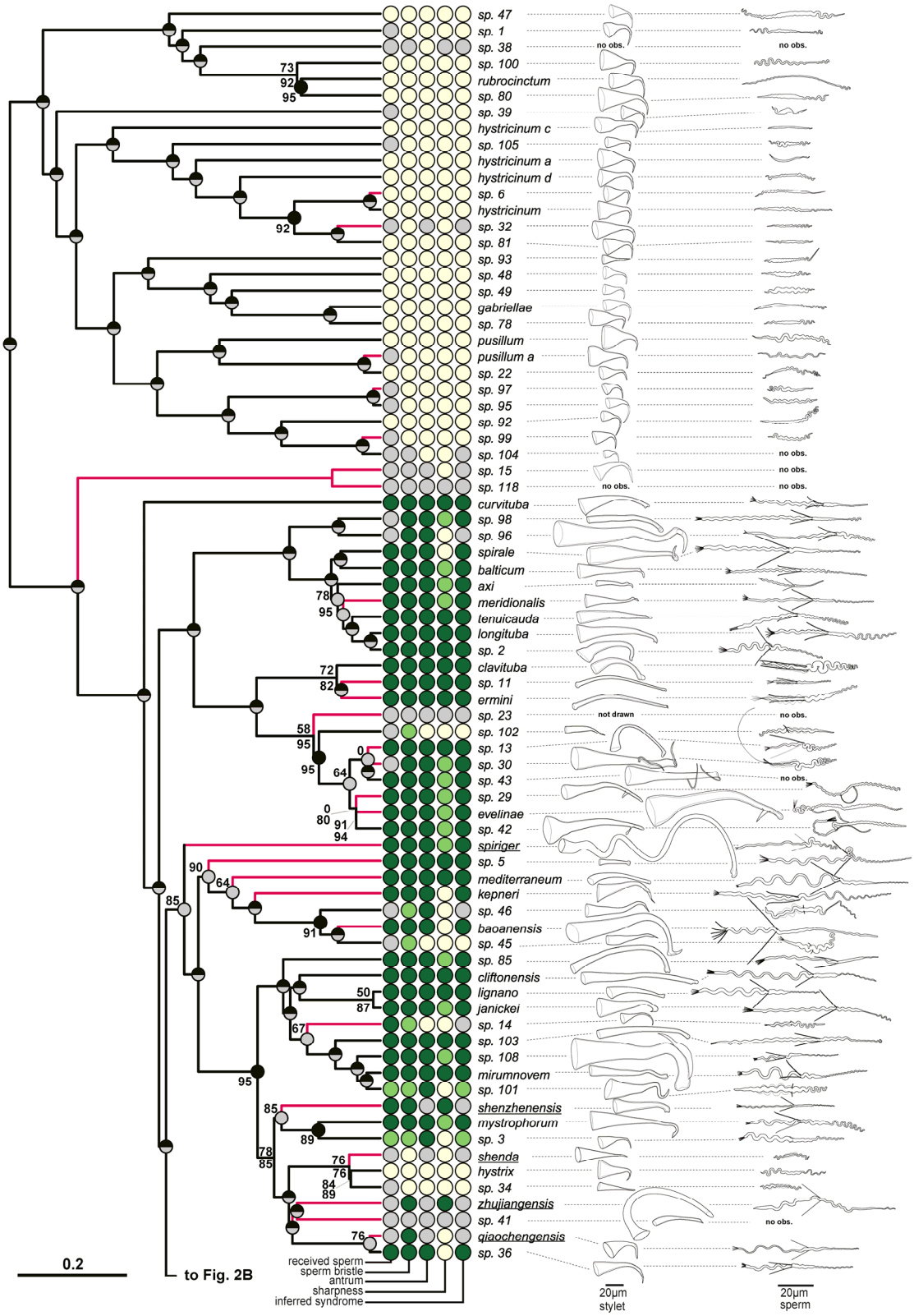
We included 145 *Macrostomum* species in our phylogeny, which was inferred from a combination of whole transcriptome data (98 species) and a partial *28S rRNA* gene sequence (permitting to add 47 additional species) (for details in outgroup species see methods). Except for 14 species we have collected and systematically documented all species ourselves. Based on an integrative approach—using detailed documentation of *in vivo* morphology and *28S rRNA* sequences—we were able to identify 51 species as previously described, leaving a striking 94 species that are new to science (see methods for details on species delimitation, Table 1 for a summary of all collected species, and Table S1 for sample sizes for morphological measures).

Table 1 Summary of the taxonomic status of the included Macrostomum species. Species inferred based on single specimens are listed separately. The high number of new species is largely due to our increased geographic sampling.

Species in this study	N = 1	N > 1	Total
All	18	127	145
Described	5	46	51
Undescribed & immature	3	2	5
Undescribed, mature & no transcriptome	5	17	22
Undescribed, mature & transcriptome	5	62	67

The molecular diversity within the hypodermic clade was high, with many species showing high molecular divergence in spite of the fact that they are difficult or even impossible to distinguish morphologically, since the stylets are extremely similar among nearly all members of this clade (Figure 2A, top). Stylets of nearly all of these species consist of a short proximal funnel that tapers to a curved bottle shape and then ends in a drawn-out asymmetrical needle-like thickening. It is possible to distinguish some species from others by their general habitus, and these differences are reflected by the four deeply split clades that are represented, respectively, by *M. rubrocinctum*, *M. hystricinum*, *M. gabriellae*, and *M. pusillum* (although species in the latter two clades are also quite similar). In cases where morphological similarity was very strong, we have indicated this by appending a letter to the species name (e.g. *M. hystricinum a*, *d*, *c* and *M. pusillum a*). Investigations of species in the hypodermic clade without molecular data require considerable caution and it needs to be clearly defined what types of specimens were collected. And as previously stated, it may often not even be clear which species the name-bearing type specimens belong to (Schärer et al. 2011).

sp. 46


sp. 47
sp. 1
sp. 38
sp. 100
rubrocinctum
sp. 80
sp. 39
hystricinum c
sp. 105
hystricinum a
hystricinum d
sp. 6
hystricinum
sp. 32
sp. 81
sp. 93
sp. 48
sp. 49
gabriellae
sp. 78
pusillum
pusillum a
sp. 22
sp. 97
sp. 95
sp. 92
sp. 99
sp. 104
sp. 15
sp. 118
curvituba
sp. 98
sp. 96
spirale
balticum
axi
meridionalis
tenuicauda
longituba
sp. 2
clavituba
sp. 11
ermini
sp. 23
sp. 102
sp. 13
sp. 30
sp. 43
sp. 29
evelinae
sp. 42
spiriger
sp. 5
mediterraneum
kepneri
sp. 46
baoanensis
sp. 45
sp. 85
cliftonensis
lignano
janickei
sp. 14
sp. 103
sp. 108
mirumnovem
sp. 101
shenzhenensis
mystrophorum
sp. 3
shenda
hystrix
sp. 34
zhujiangensis
sp. 41
qiaochengensis
sp. 36
no obs.
not drawn
0.2
to Fig. 2B
received sperm
sperm bristle
antrum
sharpness
inferred syndrome
20µm
stylet
20µm
sperm

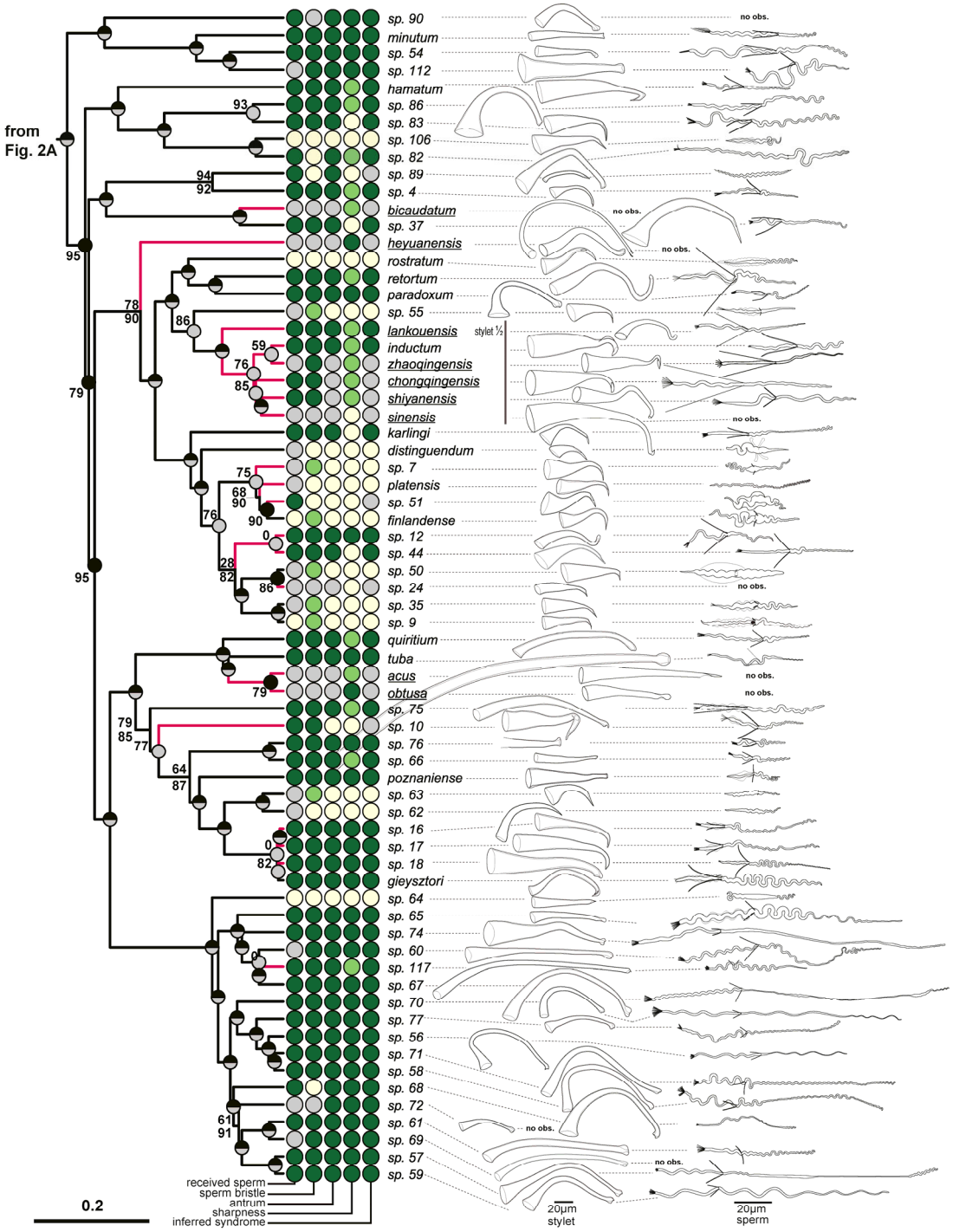

from Fig. 2A
sp. 90
minutum
sp. 54
sp. 112
hamatum
sp. 86
sp. 83
sp. 106
sp. 82
sp. 89
sp. 4
bicaudatum
sp. 37
heyuanensis
rostratum
retortum
paradoxum
sp. 55
lankouensis
inductum
zhaoqingensis
chongqingensis
shiyanensis
sinensis
karlingi
distinguendum
sp. 7
platensis
sp. 51
finlandense
sp. 12
sp. 44
sp. 50
sp. 24
sp. 35
sp. 9
quiritium
tuba
acus
obtusa
sp. 75
sp. 10
sp. 76
sp. 66
poznaniense
sp. 63
sp. 62
sp. 16
sp. 17
sp. 18
gieysztori
sp. 64
sp. 65
sp. 74
sp. 60
sp. 117
sp. 67
sp. 70
sp. 77
sp. 56
sp. 71
sp. 58
sp. 68
sp. 72
sp. 61
sp. 69
sp. 57
sp. 59
stylet ½
no obs.
0.2
received sperm
sperm bristle
antrum
sharpness
inferred syndrome
20µm
stylet
20µm
sperm

Figure 2 Phylogeny of the genus Macrostomum with drawings of stylets and sperm. The ultrametric phylogeny (C-IQ-TREE) includes all species (with 77 species depicted in Figure 2A and 68 species in Figure 2B). Branch supports are ultrafast bootstraps (top, black if 100) and approximate likelihood ratio tests (bottom, grey if 100). Species without available transcriptomes that were added based on a 28S rRNA fragment are indicated with red branches. Columns indicate the states of 5 reproductive characters from light to dark (i.e. yellow, light green and dark green for trinary states; or yellow and dark green for binary states; grey indicates missing data): received sperm location (hypodermic, both, in antrum), sperm bristle state (absent, reduced, present), antrum state (simple, thickened), sharpness of stylet (sharp, neutral, blunt), inferred mating syndrome (hypodermic, intermediate, reciprocal). Stylet and sperm are drawn based on our live observation except for species with underlined names, which were redrawn based on the species description: *M. acus*, *M. obtusa* and *M. sinensis* (Wang 2005); *M. heyuanensis*, *M. dongyuanensis* and *M. bicaudatum* (Sun et al. 2015); *M. chongqingensis* and *M. zhaoqingensis* (Lin et al. 2017a); *M. shiyanensis* and *M. lankouensis* (Lin et al. 2017b); *M. shenzhenensis* and *M. qiaochengensis* (Wang et al. 2017); and *M. spiriger* and *M. shenda* (Xin et al. 2019). Note that the stylet of *M.* sp. 15 is not drawn to scale and the stylet of *M.* sp. 23 is not drawn since it was incomplete (specimen MTP LS 913). Unobserved structures are marked as no observation (no. obs.).

Within the hypodermic clade, we discovered only one species that had a stylet that differed from the stereotypical form, namely *M.* sp. 93, which differed by having a small proximal funnel extending into a straight and obliquely-cut tube (Figure 2). This stylet shape is similar to the stylets of *M. shenda*, *M.* sp. 34 and *M.* sp. 64 (as well as *M. orthostylum*, for which we currently have no phylogenetic placement). Since we observed hypodermic received sperm in both *M.* sp. 93 and *M.* sp. 64 (Figure 2A & B), this shape appears to be adapted for HI and while we did not observe hypodermic received sperm in *M. orthostylum* and *M.* sp. 34 (nor was such sperm reported for *M. shenda* by Xin et al. 2019), these species nevertheless likely also mate through HI. Moreover, stylets with similar shapes have been documented to be involved in HI in related flatworms (Janssen et al. 2015).

We inferred phylogenies based on two sets of trimmed protein alignments, one aimed at including a large number of genes but having a lower occupancy (*L* alignment with 8218 genes, Table 2) and one with a higher occupancy but fewer genes (*H* alignment with 385 genes, Table 2). To both alignments, we applied analyses using maximum likelihood (L-IQ-TREE, H-IQ-TREE, using IQ-TREE) and summary methods (L-ASTRAL and H-ASTRAL, using ASTRAL III). We performed summary methods because they account well for incomplete lineage sorting and the gene tree – species tree conflict it can potentially cause. In addition, we also constructed a combined maximum likelihood phylogeny where we added information from a *28S rRNA* sequence to the *H* alignment (referred to as C-IQ-TREE, see Methods), in order to include species for which we lacked transcriptomes. And finally, we performed two types of Bayesian analysis on the alignment, one that like the maximum-likelihood analysis operates on partitioned amino acids (H-ExaBayes) and a second performed on the unpartitioned DNA coding sequence alignment to better account for rate heterogeneity across sites (H-PhyloBayes).

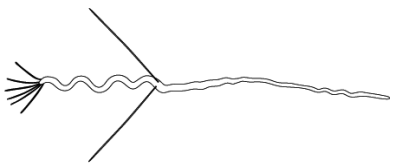
Table 2 Characteristics of the protein alignments used. We used one alignment aimed at a high number of genes (*L*) and one aimed at high occupancy (*H*). Alignments were trimmed to only include regions with a high probability of being homologous using ZORRO. Read statistics are given for the alignment before and after trimming. The top section gives information on characteristics of the entire alignment while the bottom gives the median occupancy per gene.

Whole alignment metrics	***L* untrimmed**	***L***	***H* untrimmed**	***H***
No. genes	8218	8218	385	385
No. amino acids AA	5,057,157	1,687,014	200,729	94,625
No. variable sites	3,263,955	1,157,689	135,887	74,175
No. parsimony informative sites	2,287,246	934,803	103,425	63,066
Missing data (%)	78.1	59.3	55.7	22.9
Per gene metrics (median %)				
Species occupancy	44	44	88	88
Transcriptome occupancy	40	40	84	84
Occupancy	-	35	-	78

Across all phylogenies, all species with more than one transcriptome (i.e. 15 x 2 transcriptomes, 3 x 3 transcriptomes, and 1 x 9 transcriptomes) were monophyletic, with the notable exception of *M. lignano*. We had *M. lignano* transcriptomes from the DV1 inbred line from the type locality in Lignano, Italy and from an outbred population from Thessaloniki, Greece (Zadesenets et al. 2016; Schärer et al. 2020). *M. lignano* was monophyletic in the L-IQ-TREE, H-PhyloBayes, and both ASTRAL phylogenies, but the Greek population was sister to *M. janickei* in the H-IQ-TREE and H-ExaBayes phylogenies (but node support for this split was low for H-IQ-TREE, Figure S1). *M. janickei* and *M. lignano* can hybridise in a laboratory setting (Singh et al. 2019) and it is thus not unexpected that methods not accounting for ILS or hybridisation might have errors. ASTRAL accounts for ILS and does show *M. lignano* as monophyletic. Closer inspection of the morphology of the Greek population also revealed that they had a considerably larger stylet and slightly longer sperm. We still consider *M. lignano* a proper species, but a closer investigation of the Greek populations would be interesting.

Next, we discuss agreements and discrepancies between the inferred phylogenies. Grouping of the major clades was largely consistent across all phylogenetic approaches (Figure 3 and Figure S1). With six large and very distinct species groups (hypodermic, "spirale", "lignano", "finlandense", "tuba", and "tanganyika" clade), two smaller species groups ("minutum" and "hamatum" clade), and two consistent species pairs (*M.* sp. 45 + 46 and *M.* sp. 4 + 89). However, the backbone, and the position of some species with long branches (*M.* sp. 37, *M.* sp. 39, *M.* sp. 90 and *M. curvituba)*, differed depending on the alignment and method used. But despite these discrepancies, the Robinson-Foulds distances between the trees were low (Table S2) indicating high agreement between the methods. In all phylogenies, the hypodermic clade was deeply split from the reciprocal clade, with large phylogenetic

baoanensis



distances also appearing within the hypodermic clade. Also consistent is the inferred relationship between the tanganyika, tuba and finlandense clade (Figure 3 and Figure S1) and the hamatum and minutum clade together with *M.* sp. 4 + 89 always were the next closest relatives to the former three clades. The exact relatedness patterns were fairly uncertain, however, with *M.* sp. 4 + 89 being most closely related, followed by the hamatum clade and then the minutum clade in the L-IQ-TREE, H-IQ-TREE and H-ExaBayes phylogenies, but in both ASTRAL and the H-PhyloBayes phylogenies, the hamatum clade was more closely related to the minutum clade (with the latter nested within the former in case of H-PhyloBayes) and both were sister to the grouping of tanganyika, tuba and finlandense. The exact branching at the base of the reciprocal clade is also not resolved. The spirale clade splits off first in the H-IQ-TREE and H-PhyloBayes phylogenies, while the lignano clade splits off first in the other phylogenies. Consistent with the conflict between methods, the quartet support from the ASTRAL analysis indicated strong gene tree – species tree conflict at most nodes in the backbone of the reciprocal clade (Figure S2). These internal nodes are separated by short branches suggestive of rapid speciation events, such as can occur during adaptive radiations (e.g. Irisarri et al. 2018), where substantial incomplete lineage sorting is expected. This pattern is also consistent with ancient hybridisation, which is a distinct possibility, as there is evidence for hybridisation in the genus under laboratory conditions (see above).

The topology of the C-IQ-TREE was identical to the H-IQ-TREE topology when we removed all the additional species (Figure 2). The addition of these species thus did not negatively influence the overall shape of the tree. Support for nodes in the C-IQ-TREE phylogeny was lower, as expected since the placement of these additional species solely based on *28S rRNA* sequences is difficult. Nevertheless, we think that the added inferential power obtained from a ~50% increase in species representation is highly worthwhile.

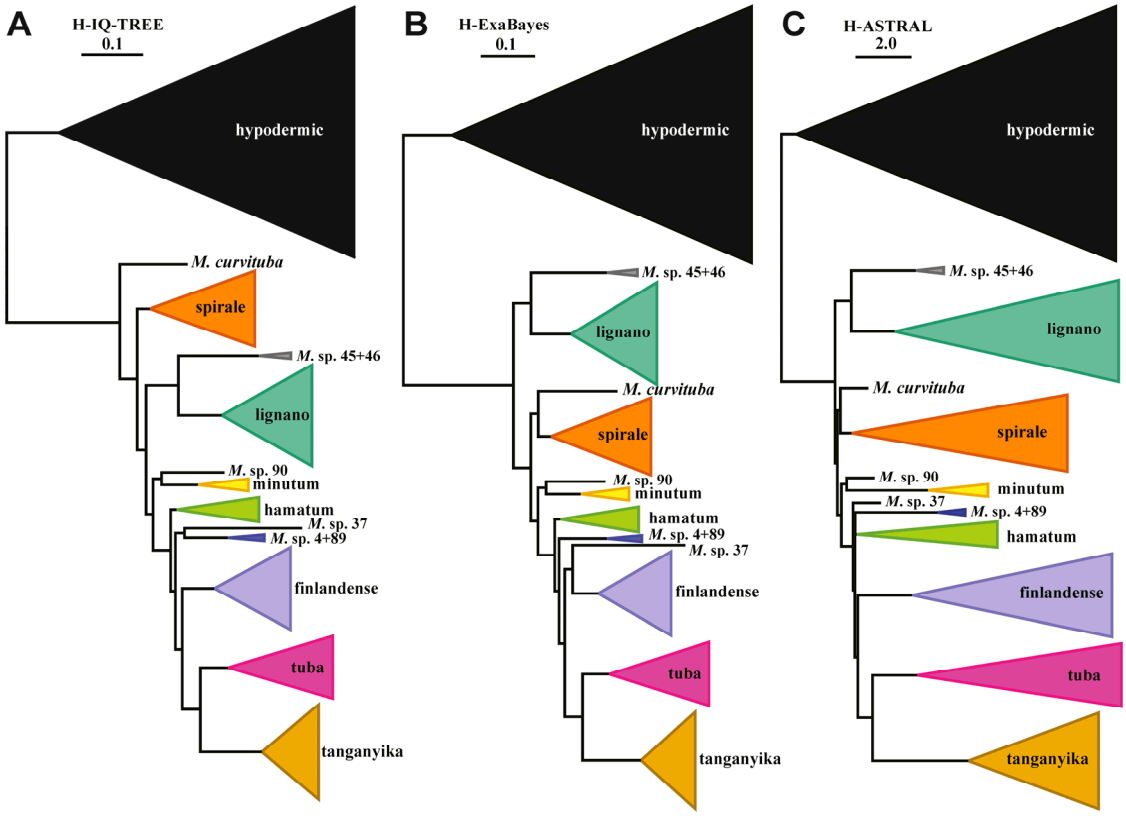


Figure 3 Simplified phylogenies of the genus Macrostomum with clades collapsed with the with proportional to the number of species contained. When not preceded by the genus abbreviation names represent clades. Not shown is that the placement of *M.* sp. 39 within the hypodermic clade was variable. A: Maximum-likelihood phylogeny generated with IQ-TREE (H-IQ-TREE) B: Bayesian phylogeny inferred with ExaBayes (H-ExaBayes). C: Summary method phylogeny inferred using ASTRAL III (H-ASTRAL). Branch length is in substitutions per site in A & B and in coalescent units in C. Support values for these groupings are generally high, for details see Figure S1.

Convergent evolution of hypodermic insemination

We here first present results from ancestral state reconstructions that allow us to infer the number of convergent transitions to HI. Then we discuss the morphological consequences that the evolution of HI entails, both through tests of correlated evolution and through a multivariate approach using principal component analysis. Finally, we highlight several species with interesting morphologies that could represent intermediate steps in the macroevolutionary transition to HI.

Ancestral state reconstruction

We reconstructed the ancestral states of received sperm location, sperm bristle state, antrum thickness and inferred mating syndrome. All reconstructions indicate a high number of transitions, which is expected if characters evolve convergently. In all reconstructions with trinary states an ordered transition model without gains was preferred, but other models sometimes received substantial support. For all analyses with binary states, a model without gains was preferred, but models permitting

gains often also received some support (Table 3). The ancestral state of received sperm location with the trinary state inferred a large number of changes (Table 3 and Figure S3) into and out of the intermediate state, with eight changes to hypodermic insemination and nine changes into the intermediate state. These frequent changes into the intermediate state are driven by the ordered model's requirements to transition through this state. When scoring received sperm location as binary trait, we estimate nine independent transitions to hypodermic sperm (Table 3 and Figure S3).

Results for the sperm bristle state are similar, with 12.2 losses and 17.5 reductions in the trinary model and 18.8 reduction/losses in the binary model (Table 3 and Figure S3). The reconstruction indicates a dynamic evolution of the bristles in the "finlandense" clade, which contains five species with reduced bristles, two species without bristles and nested within them two species with long bristles (*M.* sp. 12 and *M.* sp. 44: Figure 2, Figure S3). The most recent common ancestor of these species with bristles is predicted to have had reduced bristles, which supports the interpretation that bristle reduction can be a transitionary state, allowing loss but also secondary elongation. The reconstruction for antrum state again reveals 14.7 losses (Table 3 and Figure S3). Finally, the reconstruction for the inferred mating syndrome predicts 13.7 changes to HI and 13.2 changes to the intermediate state. Again, the many changes to the intermediate state are driven by the ordered model and in particular the trait distribution in the finlandense clade, because the reciprocally mating species *M.* sp. 12 and *M.* sp. 44 are nested within species that we classified as having HI.

Based on the ancestral states we can give a lower bound of the number of convergent events, by only counting independent losses that are separated from other losses though nodes with more than 95% posterior probability of having that state. This gives us nine transitions to hypodermic received sperm, 15 losses/reductions in sperm bristles, 13 simplifications of the female antrum, and finally, at least 12 transitions to the hypodermic or intermediate mating syndrome. However, since a model without losses is preferred in all traits, it is more likely that the finlandense clade represents two convergent transitions to HI, with *M.* sp. 12 and *M.* sp. 44 retaining the ancestral state, which would then raise the minimum number of convergent transitions to 13 (Figure S3).

Table 3 Ancestral state reconstruction of sexual characters. Several MK-models were compared based on their AIC weight (ER: equal rate, SYM: symmetrical rate, ORD-Dollo: ordered model without gains once the trait is in state 0, Dollo: model without gains, ORD: ordered model, ARD: all rates different). For models with AIC weight >0.15, we estimated the number of transitions between states using stochastic character mapping with 1000 posterior samples. Given are the number of transitions between the states and their 95% credible interval in brackets. Results are based on the C-IQ-TREE phylogeny, for the results with the other two phylogenies see Table S3.

Trait	model	dfs	log lik	AICc	ΔAICc	AICc weight	$N_{Species}$	$N_{Changes}$	0 →1	1→0	1→2	2→1	0→2	2→0
Received sperm location	ER	1	-43.3	88.6	10.1	0.004								
0: hypodermic only	SYM	3	-40.9	88.1	9.6	0.005								
1: hypodermic &	ORD-Dollo	3	-36.1	78.5	0	0.579	102	21.2 (16, 32)	-	8 (7, 13)	4 (0, 11)	9.3 (7, 13)	-	-
in antrum	Dollo	4	-36.1	80.6	2.2	0.196	102	10.8 (9, 17)	-	0.8 (0, 4)	0.9 (0, 4)	2.6 (2, 5)	-	6.5 (4, 8)
2: in antrum only	ORD	4	-36.1	80.6	2.2	0.196	102	23.4 (16, 38)	0.7 (0, 4)	8.4 (6, 13)	4.8 (0, 15)	9.6 (7, 15)	-	-
	ARD	6	-36.1	85.1	6.6	0.021								
Received sperm location	ER	1	-36.8	75.6	3.1	0.135								
0: hypodermic	Dollo	1	-35.2	72.5	0	0.638	102	9.3 (9, 11)	-	9.3 (9, 11)				
1: in antrum only	ARD	2	-35.2	74.5	2.1	0.227	102	9.9 (9, 14)	0.9 (0, 4)	9 (8, 12)				
Sperm bristle state	ER	1	-82.9	167.8	27.2	0								
0: absent	SYM	3	-81.7	169.5	28.9	0								
1: reduced	ORD-Dollo	3	-67.2	140.6	0	0.578	131	36.3 (29, 48)	-	12.2 (11,	6.7 (1, 16)	17.5 (14, 23)	-	-
2: present	Dollo	4	-67.2	142.7	2.1	0.200	131	29.3 (22, 41)	-	7.1 (3, 13)	5.3 (1, 13)	12.4 (8, 18)	-	4.6 (0, 8)
	ORD	4	-67.2	142.7	2.1	0.199	131	37.1 (29, 50)	0.8 (0, 5)	13 (10, 20)	6 (1, 14)	17.3 (14, 22)	-	-
	ARD	6	-67.2	147.1	6.5	0.023								
Sperm bristle state	ER	1	-57.6	117.3	5.7	0.038								
0: absent or reduced	Dollo	1	-54.8	111.6	0	0.635	131	18.8 (18, 22)	-	18.8 (18,				
1: present	ARD	2	-54.4	112.9	1.3	0.327	131	20.5 (17, 29)	2.7 (0, 9)	17.8 (15,				
Antrum thickness	ER	1	-48.1	98.2	3.8	0.086								
0: simple	Dollo	1	-46.2	94.4	0	0.577	127	14.7 (14, 17)	-	14.7 (14,				
1: thickened	ARD	2	-45.7	95.5	1.1	0.337	127	15.2 (13, 21)	2.1 (0, 6)	13.1 (11,				
inferred mating syndrome	ER	1	-59.3	120.6	16.2	0								
0: hypodermic	SYM	3	-54.4	114.9	10.4	0.003								
1: intermediate	ORD-Dollo	3	-49.1	104.5	0	0.578	119	34.1 (27, 47)	-	13.7 (12,	7.1 (2, 16)	13.2 (10, 19)	-	-
2: reciprocal	Dollo	4	-49.1	106.6	2.1	0.199	119	16.6 (14, 25)	-	0.7 (0, 5)	1.1 (0, 6)	2.8 (2, 6)	-	11.9 (9,
	ORD	4	-49.1	106.6	2.1	0.198	119	35.4 (27, 50)	1 (0, 5)	14.3 (11,	7.2 (2, 16)	12.9 (9, 18)	-	-
	ARD	6	-49.1	111	6.5	0.022								

Test for correlated evolution

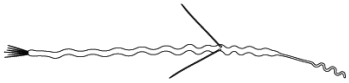
Our analyses found strong support for correlated evolution of both the sperm bristle state and the antrum state with the received sperm location (Table 4). This is in line with previous findings that a shift to HI is associated with a simplification of the female sperm storage organ and a change in sperm design (Schärer et al. 2011). Both sperm bristle and antrum state can therefore be used as markers for the mating syndrome when observations of received sperm could not be made. Besides drastically increasing taxon sampling and the number of sampled convergent events, we also expand on the previous analysis by providing evidence for the correlated evolution between the sperm bristle and antrum states (Table 4), which was implied in (Schärer et al. 2011), but not formally evaluated. This lends support to the interpretation that sperm design is shaped by the interaction of sperm with the female antrum and a transition to HI influences sperm design by modifying the selective environment of sperm.

The analyses were moderately sensitive to the chosen priors, with the largest Bayes factors for the uniform prior, slightly lower values for the exponential prior, and substantially lower values for the reversible-jump hyperprior in all tests. The reversible-jump hyperprior in some cases resulted in posterior distributions of the likelihood with multiple peaks, especially in runs using only the species with a transcriptome (see Supporting information Figure SI1-SI12). These models also showed two peaks in the distributions in the rates of transitions that were very rare (e.g. the transition away from hypodermic received sperm, which likely did not occur in these datasets) and were thus difficult to estimate and more sensitive to the prior. This phenomenon can also be seen in the posteriors of these rates using the uniform and the exponential priors, since the posteriors are very similar to the priors. Reversible-jump hyperpriors resulted in more well-behaved posteriors when running on the C-IQ-TREE phylogeny, likely because more data was available to influence the posteriors and because of the presence of species with thickened antrum state in the finlandense clade (Figure 2), which allows for all rates to be estimated more easily.

Principal component analysis

Next, we used phylogenetically corrected principal component analysis (pPCA) to investigate if these convergent transitions to HI also coincide with changes in other morphological traits, such as the length and shape of the stylet. The first two principle component vectors, PC1 and PC2, together captured nearly half of the variation in the analysed reproductive traits, followed by smaller vectors (with the first ten PCs capturing 90% of the variation, Table S4). Specifically, PC1 captured 32.6% of the variation and the loadings were defining two distinct morphological mating syndromes (Figure 4). Smaller values of PC1 were associated with a change in sperm phenotype, i.e. a reduction in sperm and bristle length, and an increased chance for the sperm to lack a brush. PC1 also captured a change in stylet phenotype,

janickei



with smaller values indicating shorter, more curved stylets that are also more asymmetric and sharper. Finally, low values of PC1 were associated with female antra that are less thickened, have a less pronounced anterior cellular valve, and a less complex internal structure. PC2 captured 13.8% of the variation and had a less clear interpretation, with high values indicating animals that are larger and have a stylet with larger proximal and distal openings.

As expected, due to their morphological similarity, the species from the hypodermic clade had very similar values in PC1 and were mainly diverged only in PC2. Interestingly species from the reciprocal clade that were categorised as showing the hypodermic mating syndrome grouped closely with the hypodermic clade (Figure 4 and Figure S4), indicating that these species have not only reduced the size of their sperm bristles, but that they have also evolved to occupy a similar morphospace with respect to stylet, sperm and antrum morphology. PC1 also separated species based on the received sperm location, with hypodermic received sperm only found in species with low values, supporting the interpretation that the suite of morphological traits defined by PC1 captures a morphology necessary for HI (Figure 4). Almost all species with reduced or absent sperm bristles grouped closely together, with the notable exception of *M.* sp. 68 and *M.* sp. 82 (Figure 4), which cluster together with other species we categorised as showing the reciprocal mating syndrome. We have observed sperm in the female antrum of both these species (two out of seven specimens in *M.* sp. 68 and 16 out of 21 specimens in *M.* sp. 82) and the female antrum of both species is similar, with a long muscular duct that performs a 90° turn towards the anterior before it enters a second antrum chamber that is strongly muscular. Both species also have a similar U-shaped stylet with a blunt tip (Figure 5), which makes it extremely unlikely that they mate through HI. This is reflected in their PC1 score indicating they are dissimilar from hypodermically mating species in the suite of traits captured by PC1 (Figure 4). We discuss the significance of this observation, also in light of their interesting sperm morphology (Figure 5), below.

Table 4 Pagel's test for correlated evolution of sexual characters. Test are conducted using BayesTraits on the three phylogenies and with three priors each. Given is the marginal likelihood of models where character transitions are dependent or independent of the other character's state. Significance of the dependent model was evaluated using Bayes factors.

Variables	Phylogeny	Prior	Marginal LH independent	Marginal LH dependent	Bayes Factor
Received sperm location - Bristle state	H-IQ-TREE	uniform	-79.5	-55.8	47.3
		exp	-71.3	-49.6	43.5
		rjhp	-67.7	-54.7	25.9
	H-ExaBayes	uniform	-79.1	-55.7	46.9
		exp	-71.0	-49.3	43.3
		rjhp	-67.3	-54.4	25.8
	C-IQ-TREE	uniform	-94.0	-68.8	50.4
		exp	-85.8	-61.3	49.0
		rjhp	-80.5	-64.7	31.6
Received sperm location - Antrum state	H-IQ-TREE	uniform	-71.1	-47.8	46.5
		exp	-63.0	-42.2	41.6
		rjhp	-57.2	-47.0	20.4
	H-ExaBayes	uniform	-70.9	-47.7	46.4
		exp	-62.7	-42.0	41.4
		rjhp	-57.0	-46.8	20.3
	C-IQ-TREE	uniform	-86.4	-66.1	40.5
		exp	-78.2	-58.3	39.8
		rjhp	-71.5	-64.4	14.1
Bristle state - Antrum state	H-IQ-TREE	uniform	-94.9	-70.4	49.0
		exp	-86.8	-63.1	47.4
		rjhp	-84.5	-65.9	37.1
	H-ExaBayes	uniform	-94.4	-70.3	48.3
		exp	-86.4	-63.0	46.9
		rjhp	-84.1	-65.6	36.8
	C-IQ-TREE	uniform	-117.6	-86.4	62.5
		exp	-109.6	-78.9	61.4
		rjhp	-106.8	-88.1	37.4

Possible pathways to hypodermic insemination

While most species could be assigned to a mating syndrome (119 of 145) there were some species that were not easily identified as performing either type of mating behaviour (Table 5). We categorised two species as intermediate because we observed sperm both in the female antrum and embedded inside the recipient's tissues. Both of these species (*M.* sp. 3 and *M.* sp. 101) have a sharp stylet and sperm with reduced bristles, fitting with the hypodermic mating syndrome, but they also have a thickened antrum wall, which is more indicative of the reciprocal mating syndrome, and these species therefore have intermediate values of PC1 (species indicated in light green in Figure 4). Multiple observations of hypodermic sperm in *M.* sp. 3 show them to be embedded deeply in the anterior wall of the female antrum, as well as more deeply in the tissue lateral to the body axis and extending up to the ovaries, and finally we also observed sperm in the tail plate (Figure 6). In *M.* sp. 101 we did not observe sperm as deeply in the recipient's tissues, but some were completely embedded in the cellular valve and just anterior of it, and thus close to the developing eggs (Figure 7). One explanation for these findings would be that during mating the stylet of both species pierces the antrum wall and sperm is traumatically injected into the body internally. Unfortunately, no copulations were observed in mating observations of *M.* sp. 101, but this species has been seen performing the suck behaviour, as expected if ejaculate is, at least partially, deposited in the female antrum (Singh et al. in preparation). We lack mating observations for *M.* sp. 3 and further investigations of the behaviour and antrum histology of both species would be highly desirable.

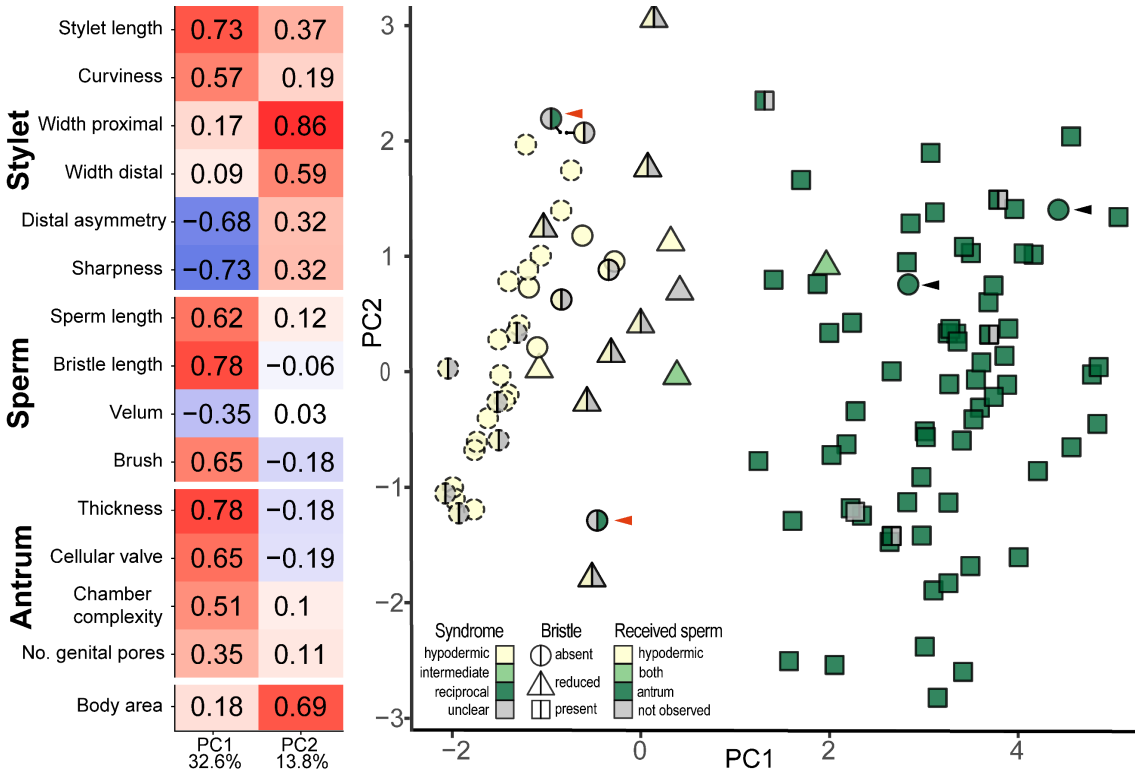
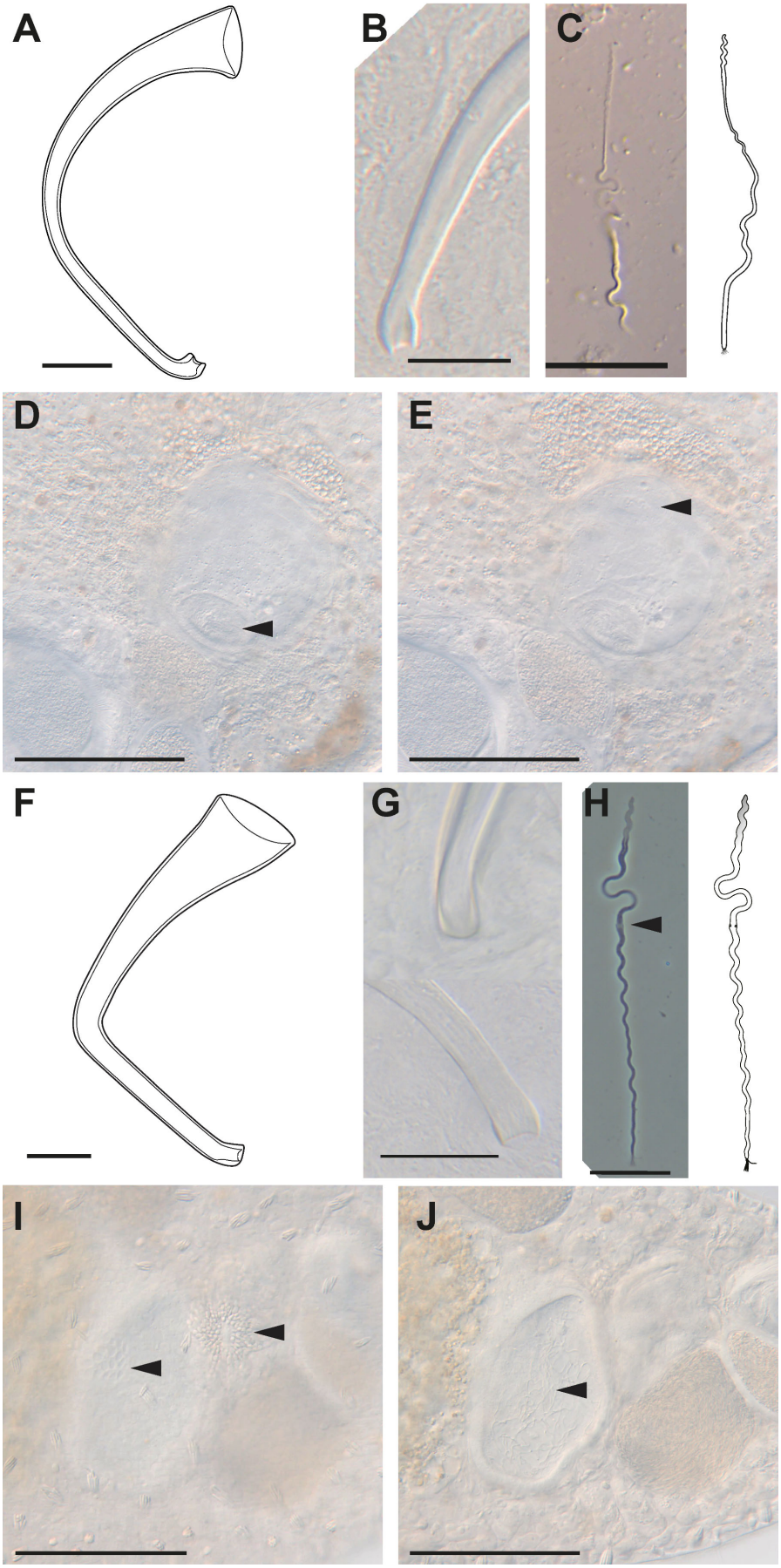


Figure 4 Results of pPCA of the analysed reproductive traits. Left: Loadings of PC1 and PC2, with the percentage of variance explained at the bottom. Right: 2D morphospace defined by PC1 and PC2. As indicated by the legend, shape represents the bristle state, while the colours represent the inferred mating syndrome (left side) and the received sperm location (right side). Species from the hypodermic clade are outlined with stippled lines. Red arrows indicate two species (*M.* sp. 51 and *M.* sp. 89) that cluster closely with species assigned to the hypodermic syndrome but we observed received sperm in the female antrum. Black arrows indicate two species (*M.* sp. 68 and *M.* sp. 82) assigned to the reciprocal syndrome, which have no discernible sperm bristles. The phylogenetic relationships of these species are represented as a phylomorphospace animation in Figure S4.

Figure 5 (Next page) Details on the reproductive morphology of *M.* sp. 68 (A-E) and *M.* sp. 82 (F-J); *M.* sp. 68 (A) Stylet drawing showing the blunt distal thickening; (B) Distal stylet tip in a smash preparation (specimen MTP LS 2611). (C) Sperm image (specimen MTP LS 2686) and drawing showing no apparent sperm bristles. (D-E) Details of the antrum (specimen MTP LS 2562) indicating the muscular connection between the female genital opening and the antrum (arrow in D) and the anterior second chamber (arrow in E). *M.* sp. 82 (F) Drawing of the stylet showing the slight blunt distal thickening. (G) Distal stylet tip in situ (top, specimen MTP LS 2845) and in a smash preparation (bottom, specimen MTP LS 2846). (H) Sperm image (specimen MTP LS 2877) and drawing indicating the modified anterior part of the sperm (shaded grey) and a less dense area approximately 1/3 along the sperm, which could be a vestigial bristle anchor location (arrow). (I-J) Details of the antrum (specimen MTP LS 2848) indicating the anterior opening (I, left arrow) next to the posterior opening (I, right arrow) both connecting into a large chamber (J, arrow). Scale bars represent 100 µm in the antrum images and 20 µm otherwise.


A
B
C
D
E
F
G
H
I
J

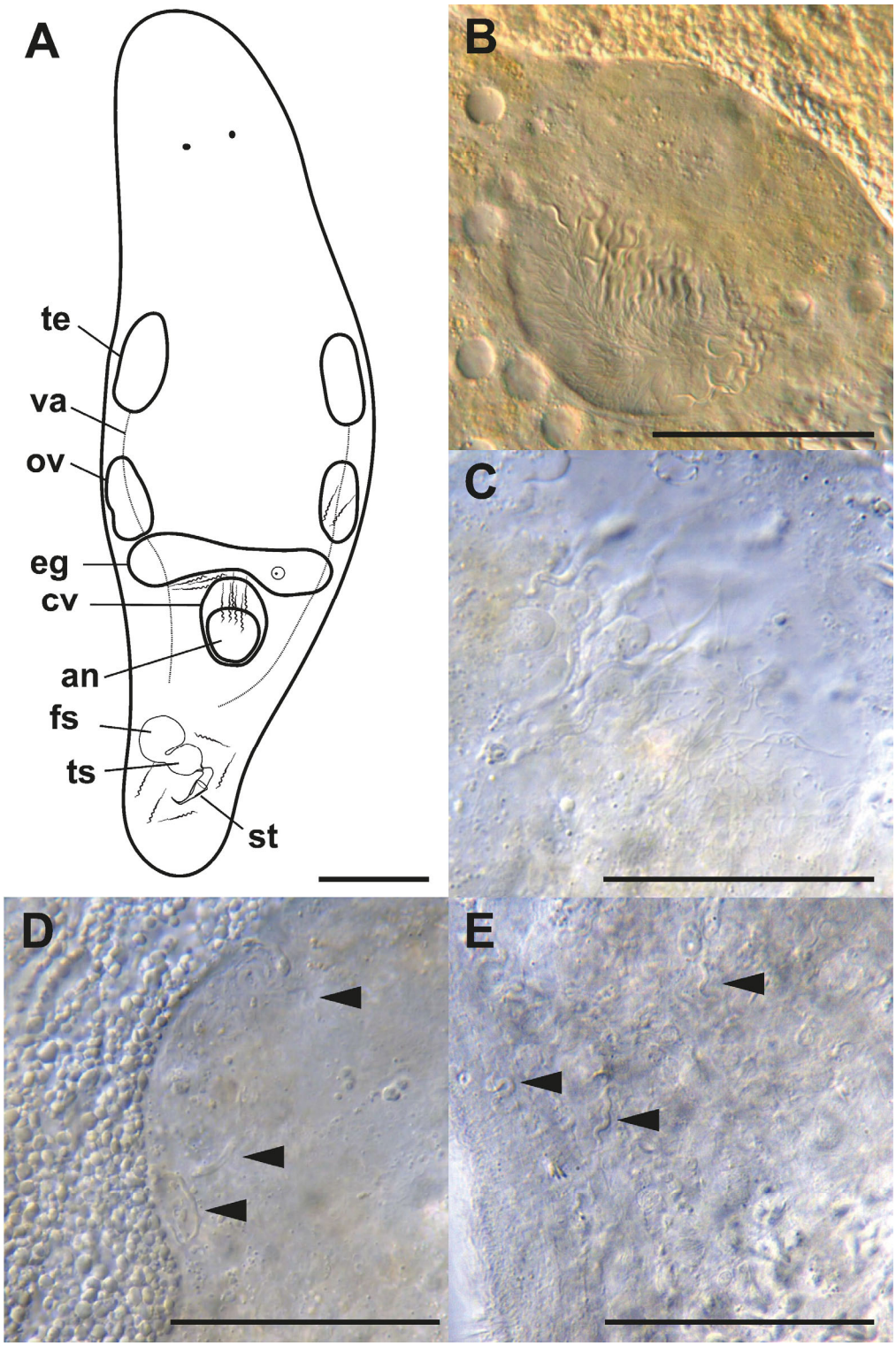


Figure 6 Detailed observations of received sperm location in *M.* sp. 3. (A) Drawing of the sexual organs of *M.* sp. 3, (te: testis, va: vas deferens, ov: ovary, eg: egg, cv: cellular valve, an: female antrum, fs: false seminal vesicle, ts: true seminal vesicle, st: stylet), with sperm drawn at locations where they were observed. (B) sperm in the female antrum (specimen MTP LS 3286) embedded in the cellular valve (anterior); (C) sperm in cellular valve with loose tissue (specimen MTP LS 3314); (D) sperm close to developing egg (specimen MTP LS 3314); (E) sperm on top of ovary (specimen MTP LS 3317). Scale bars represent 100µm in (A) and 50 µm otherwise.

sp. 108

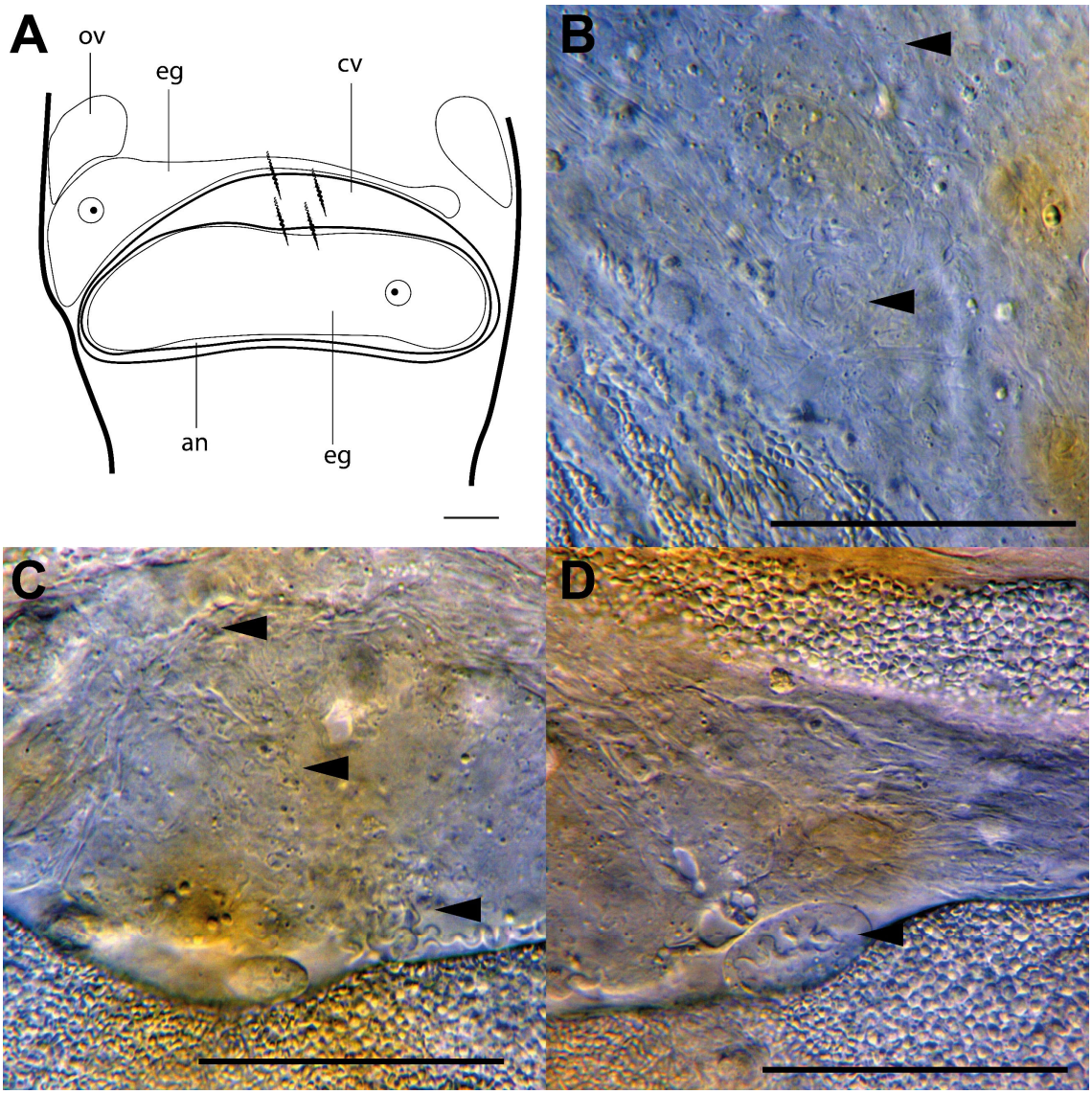
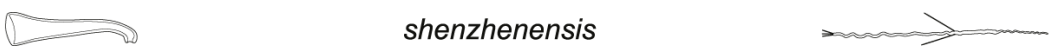


Figure 7 Detailed observations of received sperm location in *M*. sp. 101 (A) Drawing of antrum region of *M*. sp. 101, (ov: ovary, eg: egg, cv: cellular valve, an: female antrum), with sperm drawn at locations where they were observed. (B-D) show sperm in situ. (B) sperm (arrow) embedded in cellular valve (specimen MTP LS 3206); (C) sperm (arrow) in cellular valve and antrum (specimen MTP LS 3127); (D) Same specimen as C with dislodged sperm (arrow) surrounded by what appears to be cellular valve derived tissue. Scale bars represent 100μm in (A) and 50μm otherwise.



Three species (*M.* sp. 14, *M.* sp. 51 and *M.* sp. 89) were difficult to classify because, although their morphology indicates HI, we observed received sperm in the female antrum. *M.* sp. 51 and *M.* sp. 89 both grouped with the hypodermically mating species in PC1 (black arrows in Figure 4), while we did not include *M.* sp. 14 in this analysis due to missing data for sperm bristle length. We found sperm within the female antrum of only one specimen in *M.* sp. 51 (four specimens) and *M.* sp. 89 (12 specimens), and it is thus possible that sperm is hypodermically injected and may enter the female antrum when an egg passes the cellular valve before egg laying. But these species could also represent an intermediate state between the mating syndromes. We observed sperm in the female antrum of three out of five specimens in *M.* sp. 14 and therefore it is unlikely that sperm entering during the transition of the egg into the female antrum is the cause of its presence. Instead, sperm is likely deposited in the female antrum by the mating partner during copulation. However, based on its general morphology, we predict that closer investigations of this species will reveal hypodermic received sperm in a similar location as found in the intermediate species (*M.* sp. 3 and *M.* sp. 101).

Finally, we were not able to assign *M.* sp. 10 to a mating syndrome because, although we found received sperm in the female antrum and its sperm carry long bristles, it also has a sharp stylet and a simple female antrum. From our previous findings, we would expect this species to have a thickened female antrum due to its interaction with sperm and the stylet of the mating partner. This discrepancy could be attributed to misclassification of the antrum morphology, since *M.* sp. 10 has very pronounced shell glands, making it difficult to clearly see the anterior part of the female antrum, possibly obscuring a thickening or cellular valve (see specimens MTP LS 788 and MTP LS 801 for a possible thin cellular valve).

Hypodermic insemination and sperm design

We tested whether HI is associated with a change in sperm length using phylogenetic least squares (PGLS) regression. We used received sperm location, bristle state, antrum state, and the inferred mating syndrome as predictors and the $\log_{10}$ transformed sperm length as the response variable. In all cases, the states indicating reciprocal copulation were associated with longer sperm, with the largest effect size for antrum state, followed by the inferred mating syndrome (Figure 8). This is expected since, their bristle type falsely classifies *M.* sp. 68 and *M.* sp. 82 as hypodermically mating and the received sperm location includes fewer species, while antrum state has a higher sample size than the inferred mating syndrome. The predictive value of the PGLS models was high, indicating that a large proportion of the variation in sperm length is explained by the phylogeny and indicators of reciprocal copulation (Figure 8). Note that despite the clear effect, there is considerable overlap in sperm length between the groups, with some reciprocally mating species having short sperm (Figure 8).

mystrophorum

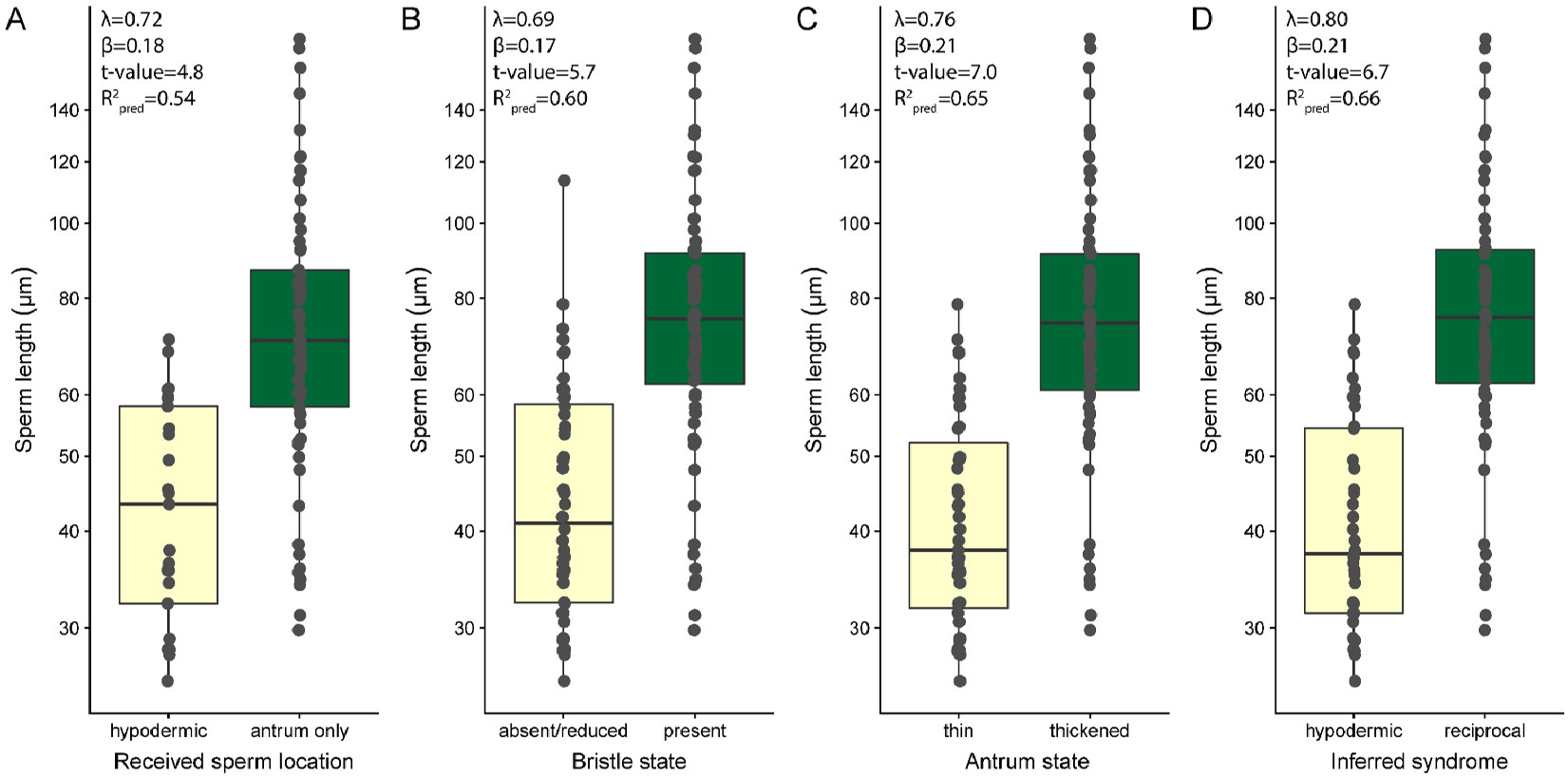


Figure 8 Sperm length of species dependent on (A) sperm bristle state; (B) received sperm location; (C) antrum state; (D) inferred mating syndrome. Values are slightly jittered in the x direction, and the y-axis is on a log-scale. Within each panel the main results of PGLS analysis are given. In all tests the slopes were significant at $p < 0.001$. Detailed results are in Table S5.

Evolution of reproductive traits within the reciprocal mating syndrome

We tested for signatures of coevolution between male and female genital traits in reciprocally mating species by performing PGLS analysis of antrum complexity and stylet length. We found no significant association between these traits, irrespective of whether we included body size as a covariate (Figure 9, Table S6), suggesting no matching of trait size. This is counter to documented examples of coevolution in hermaphrodites (e.g. Koene and Schulenburg 2005; Anthes et al. 2008) and it seems possible that we find no evidence for coevolution because the measurements of both traits are quite crude and do not adequately capture the complexity of the structures involved (as discussed in the Methods section).

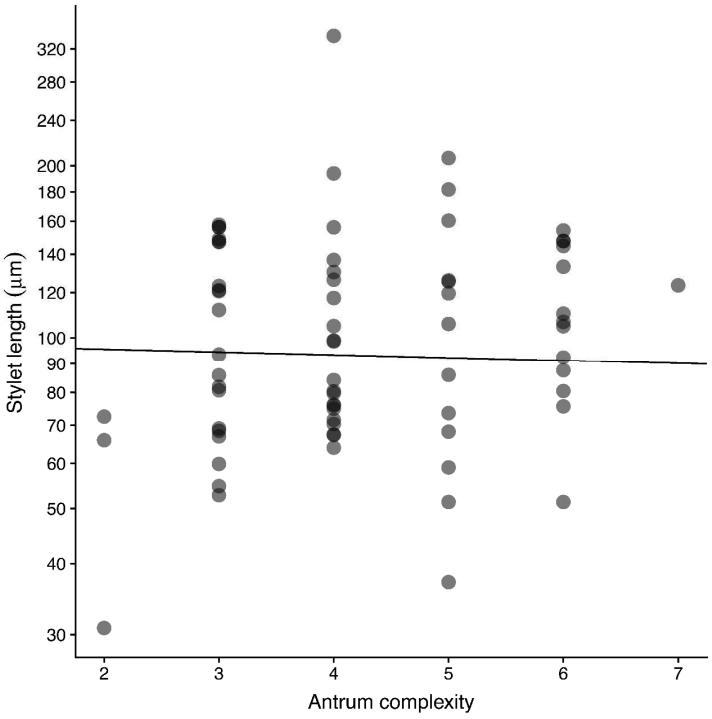


Figure 9 Stylet length and antrum complexity. The line gives the fit from PGLS analysis showing not relationship between the variables. Detailed results are in Table S6.

However, we want to highlight several morphological traits that could be important for sexual selection and sexual conflict within the genus. First, the phylogeny clearly shows that a second female genital pore has evolved at least four times within the genus, being present in *M. spiriger, M. gieysztori*, and *M. paradoxum* (previously called *Promacrostomum paradoxum* (An-der-Lan 1939), but we suggest that it should now be referred to as *Macrostomum paradoxum,* see discussion) as well as in four additional species presented here, namely three close relatives of *M. gieysztori* (*M.* sp. 16, *M.* sp. 17 and *M.* sp. 18), and *M.* sp. 82 ,which also has a peculiar sperm morphology. Second, several species in the tanganyika clade have extraordinarily long sperm. These traits likely are important for sexual selection, as we will outline in the discussion.

Discussion

Species collected and phylogenetics

In all phylogenies, *Promacrostomum paradoxum* (An-der-Lan 1939) was placed within the reciprocal clade as sister to *M. retortum* (Figure 2A) and we suggest that it should from now on be referred to as *Macrostomum paradoxum*. With this reclassification, the genus *Promacrostomum* will be rendered monospecific, only containing *P. palum* (Sluys 1986), which was assigned to this genus because it, like *M. paradoxum,* has two female genital openings. However, as outlined above this character has evolved independently at least four times and it thus clearly is not a synapomorphy of *Promacrostomum*. Moreover, it was recently suggested that another species with two female genital openings, *M. gieysztori*, also belongs into the genus *Macrostomum* (Schärer et al. 2011), in spite of it having been placed into another genus.

The addition of *28S rRNA* sequences revealed that *M. zhaoqingensis* (Lin et al. 2017b) is sister to *M. inductum* and the haplotype analysis shows that sequences from both species cluster closely (see Supporting information Figure SI29). Further investigations thus might reveal that *M. zhaoqingensis* is a re-description of *M. inductum*. The C-IQ-TREE phylogeny further places *M.* sp. 15 and *M.* sp. 118 as sisters that are deeply split from the other species and placed at the base of the reciprocal clade. While we have no observations of the stylet morphology of *M.* sp. 118, the morphology of *M.* sp. 15 is suggestive of HI and closer investigations of this clade would be interesting since this could allow inference about the ancestral state leading to the reciprocal clade.

Convergent evolution of hypodermic insemination

Ancestral state reconstruction

Our data clearly shows that HI has evolved independently at least 9 times when assessed based on the location of received sperm but may have evolved up to 13 times when our more inclusive inferred mating syndrome is considered. This greatly increases the number of such convergent events documented (only 14 in hermaphrodites according to (Lange et al. 2013), which includes the two previously documented in *Macrostomum*), thus nearly doubling the number of documented cases among hermaphrodites resulting from a detailed analysis of a single genus of free-living flatworms. Moreover, three additional origins of HI have also been documented in the Macrostomorpha (Janssen et al. 2015), the parent group of the genus *Macrostomum*, suggesting that the frequent origins of HI are not restricted to this genus, but likely very widespread in the Macrostomorpha, and maybe other groups of free-living flatworms. Considering the large number of new species we were able to collect, it seems likely that we have not sampled all convergent events

within the genus *Macrostomum*, since a large proportion of the diversity in the genus *Macrostomum* is probably yet to be discovered. As a case and point, the deeply split placement of *M.* sp. 15 and *M.* sp. 118 (see above) suggests the existence of an additional clade representing large amounts of additional molecular diversity, and many regions of the world remain extremely poorly studied in terms of free-living flatworms.

Frequent convergent evolution of this extreme type of mating bolsters the interpretation that this is an adaptive resolution to sexual conflict over mating rate, the mating role or both (Charnov 1979; Michiels 1998; Michiels and Newman 1998; Vizoso et al. 2010; Schärer et al. 2011, 2015). HI likely is an alternative strategy in an ongoing evolutionary arms race between donor and recipient, with donor persistence traits, such as complex sperm with bristles and manipulative seminal fluids, and recipient resistance traits, such as the suck behaviour and complex female genitalia in constant antagonistic coevolution, and both repeatable and novel outcomes occurring with great frequency.

Test for correlated evolution

The analysis of correlated evolution clearly shows that there is a very strong association between both the sperm bristle and antrum state with the received sperm location. While this finding might initially be conceived as simply a confirmation of previous findings (Schärer et al. 2011), albeit with more species and consequently a larger Bayes factors, it actually represents an important qualitative improvement of the evidence for these associations. This is due to how tests of correlated evolution generally, and also the Pagel test implemented in the BayesTraits analysis specifically, are set up. Tests of correlated evolution evaluate evidence for convergence by testing if a model where the evolution of the state of one trait is dependent on the state of the other trait is more likely than a model where it is not. Counterintuitively, however, while these tests supposedly correct for phylogenetic independence, they can support the dependent model of evolution even with only a single (unreplicated) origin of the trait states in question (as outlined in Maddison & FitzJohn 2015, Uyeda et al. 2018), provided that they map to the appropriate nodes in the phylogeny. However, descendants of a clade often share multiple traits, many of which may not be functionally linked (as in the example given by Maddison & FitzJohn 2015 of the joint presence of fur and middle ear bones in mammals). But one aim of such a correlation analysis of course is to explore whether there might be a possible causal relationship between the traits. Thus, while the previous findings with two independent origins of HI could be considered evidence for correlated evolution, that evidence with two origins is not as decisive as the large Bayes factors may have suggested. By sampling many more convergent events we here could remedy the limitations of the earlier results and can now have substantially more confidence in that our findings indeed may indicate a causal link.

Furthermore, sampling more convergent events also affords us the opportunity to test the consequences of HI more quantitatively (see next sections).

Principal component analysis

The frequent origins of hypodermic mating allow us to quantitatively assess the predictions made by Schärer et. al. (2011) concerning the nature of the two mating syndromes. They posited that combinations of adaptations are necessary for efficient HI, consisting of a needle-like stylet, sperm adapted for movement through tissue, and as a consequence, a reduction of antrum complexity, since it is now no longer interacting with the sperm and stylet of the partner, but is simply used for egg-laying. While their data was suggestive, the issue was that the second hypodermic origin consisted only of *M. hystrix*, which made quantitative analysis difficult. The principal component analysis we performed here shows that species with HI have similar values of PC1, with this state corresponding tightly to the mating syndromes described by Schärer et. al. (2011) (but note that we had to somewhat alter their definitions since we lacked behavioural observations for most species). We thus show that HI is indeed associated with a distinct syndrome of reproductive traits that has convergently evolved numerous times in this genus and thus these traits likely constitute predictable adaptations to this type of mating.

Possible pathways to hypodermic insemination

The species with an unclear or intermediate inferred mating syndrome suggest that a possible route to HI is via the initial evolution of a sharp stylet, which could arise to provide anchorage during mating copulation (similar to stylets that could serve such a function e.g. *M. spirale* or *M. hamatum*), as has been suggested for traumatically mating bedbugs. There the attachment during mating is a two-step process, indicating that traumatic insemination was evolutionarily preceded by traumatic penetration for attachment (Lange et al. 2013). Similar transitions have also been proposed for *Drosophila* in the melanogaster group, where extragenital structures that are used in some species for anchorage are modified in traumatically inseminating species to pierce the integument of the mating partner (Kamimura 2007, 2010).

A sharp stylet may also serve to stimulate the partner during copulation, aid in the destruction or removal of rival sperm already present in the partner's female antrum, or help to embed sperm in the antrum wall to prevent its removal, either by rival mating partners or by the recipient during the suck behaviour. Irrespective of the initial selective advantage that internal wounding may confer, it could then evolve further to complete internal traumatic insemination, and eventually complete avoidance of the female genitalia and thus hypodermic insemination via the epidermis. Accidental sperm transfer due to copulatory wounding has generally been suggested to be a possible route from copulation to traumatic insemination (Lange et al. 2013; Reinhardt et al. 2015), though the fact that we often observe stylets with

blunt distal thickenings suggests that such internal wounding may not be advantageous for the donor in many *Macrostomum* species (Schärer et al. 2011).

Hypodermic insemination and sperm design

We further confirm the findings of Schärer et al. (2011) that HI is associated with a loss of the sperm bristles. Furthermore, we, for the first time, document hypodermic received sperm in species with reduced bristles. This indicates that at least in those species HI precedes the complete loss of bristles. This is supported by the ancestral state reconstructions using three sperm bristle states, which showed the strongest support for an ordered evolutionary model, suggesting that a transition via an intermediate state is statistically supported. If bristles only confer a selective advantage in reciprocally mating species, then we could expect them to be lost through drift once HI evolves. Moreover, sperm bristles might be actively selected against in hypodermically mating species if they result in costs for the donor, such as a reduced rate of spermatogenesis of complex sperm or a reduced motility of hypodermic sperm carrying bristles (Schärer et al. 2011). Indeed, spermatogenesis of the complex sperm with bristles of *M. lignano* takes about 6 days, two days longer than the development of simple sperm in *M. pusillum* (Schärer and Vizoso 2007; Giannakara et al. 2016; Giannakara and Ramm 2017). But note that this could also be due sperm length differences between these species. The fact that there are hypodermically mating species with reduced bristles suggests that selection might only be efficient at reducing bristle size until they no longer result in strong costs. Short bristles might not hinder movement much and once the bristles are reduced to such an extent that they do not protrude from the sperm, they could be retained across longer evolutionary times. Indeed, the sperm of a species from the *M. pusillum* species-complex in the hypodermic clade, contains electron-dense bodies (Rohde and Faubel 1997) that are similar to the anchor structures of the sperm bristles identified in reciprocally mating species *M. tuba* and *M. lignano* (Rohde and Watson 1991; Willems et al. 2009). If these structures are indeed remnants of bristles, this would imply that the ancestor of all *Macrostomum* had bristles (in agreement with our ancestral state reconstruction) and that these remnants have been retained for a long time, implying they might not be costly or detrimental to sperm function. More widespread investigations of sperm ultrastructure of species in the hypodermic clade will be needed to confirm this hypothesis.

Our data further suggest that the correlation between reduced/absent bristles and the hypodermic mating syndrome is not perfect, because we document two species without sperm bristles, that very likely mate through reciprocal copulation (*M.* sp. 68 and *M.* sp. 82). In both species, we have observed sperm in the female antrum, which in both is very muscular and complex (Figure 5). We speculate that sperm is deposited very deeply inside the female antrum, so that sperm bristles may possibly no longer serve an anchoring function in these species. Additionally, the

zhujiangensis

sperm of *M.* sp. 82 has a peculiar feeler (Figure 5), which could possibly represent an adaptation to this complex antrum morphology. It is also not clear if these species perform the suck behaviour, as this was not observed in mating observations of *M.* sp. 82, while we currently have no data at all on *M.* sp. 68 (Singh et al. in preparation). Finally, we discovered several monophyletic species that had particularly elongated sperm, presumably for similar reasons. We will discuss these separately.

Our PGLS analysis clearly showed that, on average, the sperm of species with the hypodermic mating syndrome are shorter. There are several possible explanations for these findings. First, HI avoids the partners genitalia, and these potentially allow both cryptic female choice (e.g. via the suck behaviour), and sperm displacement or removal by competing donors. HI could therefore alter the mode of postcopulatory sexual selection and particularly sperm competition. This is because these mechanisms can introduce paternity skews (Charnov 1996; van Velzen et al. 2009; Schärer and Pen 2013), which likely results in lower sperm competition compared to a "fair-raffle" type sperm competition where sperm mix more freely (Parker 1982, 1993, 1998), which seems likely under HI. If sperm size trades-off with sperm number (Parker 1978, 1982, 1993) then more intense sperm competition under HI will favour the evolution of smaller sperm (Schärer and Janicke 2009; Schärer et al. 2011). Second, sperm in *Macrostomum* is quite large compared to the female antrum and intimately interacts with its epithelium, often being partially embedded in the cellular valve with the feeler (Ladurner et al. 2005; Schärer et al. 2011). And sperm is also in close contact with rival sperm when animals mate multiply (Janicke et al. 2013; Marie-Orleach et al. 2016). Under such conditions of high sperm density, when sperm displacement is likely (e.g. Miller and Pitnick 2002; Lüpold et al. 2012; Manier et al. 2013), sperm are predicted to be bigger compared to species in which the sperm storage organ is substantially larger than the sperm (Parker et al. 2010; Immler et al. 2011). While in species with HI the sperm still intimately interact with the tissue of the partner, it might encounter rival sperm less easily and run the risk of being diluted. Third, sperm size could simply decrease because small sperm are better adapted for movement through the tissue of the mating partner (Schärer et al. 2011). We know very little about how sperm move within the recipient's tissues, but it seems they move similarly to sperm within the female antrum via undulation of the sperm body using cortical microtubules. Presumably, smaller sperm will need to exert less energy to overcome the resistance of the tissue and thus might be able to move more efficiently. Note, that these explanations are not mutually exclusive, and their relative importance might depend on the physiology and ecology of each species.

Evolution of reproductive traits within the reciprocal mating syndrome

While we did not find quantitative evidence for coevolution of male and female reproductive traits, we describe several morphologies that could constitute resistance or persistence traits in sexual conflict over the fate of received ejaculate. A prime candidate for a resistance trait is the evolution of a second female genital opening. All species with such a second opening have one that is associated with the shell glands and one that is not (referred to as the gonopore and bursa pore, respectively An-der-Lan 1939) and it is therefore assumed that the gonopore is homologous to the single gonopore in all other *Macrostomum,* since that is also surrounded by shell glands (Sluys 1986). The bursa pore is always associated with a small chamber (bursa) that has strong circular musculature and is located anterior to the gonopore in *M. spiriger*, *M. paradoxum* and *M.* sp. 82, but posterior to it in *P. palum* and *M. gieysztori*, further supporting their independent evolution. Moreover, in *M. paradoxum* the bursa is described as being connected to the gut via a genito-intestinal-tract (An-der-Lan 1939), which could function for sperm digestion. However, we were not able to locate this duct in our *in vivo* observations of specimens and may have to resort to histology. Even without such a duct, the bursa could potentially allow cryptic choice, since it might be used to eject sperm through muscular contraction. Such contraction is also at play during the suck behaviour, where at least in *M. hamatum* sperm can be seen ejected even before the worm places its mouth on the gonopore (Singh et al. in preparation). Finally, the exaggeratedly long sperm in most species of the "tanganyika" clade warrant further study, since they could constitute a novel persistence trait serving a similar function as is hypothesised for the sperm bristles. It appears possible that these long structures can embed themselves in the female antrum of the recipient and thus anchor the sperm, preventing removal, which would make bristles less important to resist the suck behaviour. This interpretation is supported by the fact that several of these species have quite short bristles, in spite of showing the reciprocal mating syndrome (including the species *M.* sp. 68 discussed above lacking bristles entirely).

At this point, these highlighted morphological traits simply represent intriguing characters the likes of which are frequently found among various groups of flatworms. However, the resources we have provided here uniquely position *Macrostomum* as a genus in which detailed investigations of such characters can be conducted. Cultures of several species are routinely maintained in the laboratory, making them available for experimental investigation. Further, *M. lignano* is an established model organism where several genomic resources are available that will facilitate investigations of gene function (Wudarski et al. 2017). Future research will be capable of, not only expanding on the comparative aspects of this work, but also using the provided transcriptomic resources to tackle experimental investigations of the molecular basis of these traits.

Conclusion

We have dramatically increased our understanding of the genus *Macrostomum*, by collecting and collating detailed molecular and morphological information on 145 species. Most of these species are new to science, highlighting that much diversity is still to be discovered in this interesting genus. Through increased taxon sampling and a robust phylogenomic phylogeny we have almost doubled the number of documented cases of convergent evolution of HI among hermaphrodites. HI is thus likely a prevalent mating strategy in flatworms and in hermaphrodites in general. We were also able to determine several species that represent intermediate morphologies in the evolutionary transition to HI, suggesting it evolves via initial hypodermic insemination through the antrum epithelium. Due to the large number of sampled convergent events we were able to confirm previous findings that HI is associated with a simplification of the female antrum. Furthermore, we were able to show that HI is associated with a clearly defined morphological syndrome and coincides with a strong reduction in sperm length. This indicates that HI fundamentally changes the functional requirements on sperm design, possibly due to a change in the mode of sperm competition. Finally, we have highlighted and discovered several reproductive morphologies that likely represent important characters in the evolution of sexual traits within the genus. Thus, this work establishes the genus *Macrostomum* as an ideal system for further investigations of sexual selection, sexual conflict, and the evolution of HI on a macroevolutionary scale.

Material and Methods

Field collections and documentation

Almost all specimens were collected from the field in freshwater, brackish or marine habitats, either from sediment samples and/or from water plants (details on all sampled specimens and their measurements in Table S7). We documented specimens extensively with digital photomicrography, as previously described (Ladurner et al. 2005; Schärer et al. 2011; Janssen et al. 2015), using light microscopes (Leica DM2500, Olympus BH2, Leitz Diaplan, Zeiss Axioscope 5) with differential interference contrast (DIC) and equipped with digital cameras (Sony DFW-X700, Sony DFK41 and Ximea xiQ). We collected both images and videos at various magnifications (40x to 1000x), documenting the general habitus and details of internal organs, which is possible due to the small size and high transparency of these organisms. To document sperm we amputated a worm's tail slightly anterior of the seminal vesicle, ruptured the seminal vesicle (as described in Janicke and Schärer 2010), and documented the released sperm in a smash preparation using DIC (and sometimes also phase-contrast microscopy). When possible, we also

prepared whole-mount permanent preparations of these amputated tails to preserve the male intromittent organ (Schärer et al. 2011; Janssen et al. 2015). Finally, we preserved the entire animal, or its anterior portion when amputated, for molecular analysis, in either RNAlater solution (Sigma, stored at 4°C up to a few weeks and then at -80°C) or in absolute ethanol (stored cool for up to a few weeks and then at -20°C).

Morphological data

To characterise the morphology of reproductive traits, we collected both quantitative (Q) and categorical (C) data from the detailed images and videos of the collected specimens (or from the taxonomical descriptions of the few species we did not collect ourselves). Categorical data were determined on a per species basis, while quantitative data were taken per individual. We measured body size (Q) as the total body area and either measured or scored various aspects of the stylet (Q: length, curviness, width of the proximal opening, width of the distal opening, and asymmetry of the distal thickening; C: sharpness of the distal thickening), the sperm (Q: total length, bristle length; C: bristle state and presence of a brush or a velum) and female antrum (Q: number of genital pores; C: antrum thickness, presence and thickness of anterior cellular valve, antrum chamber complexity, and an overall score of antrum complexity). We refer the reader to the "Supporting information" for detailed explanations of these measures. Morphometric analyses were performed using the software ImageJ (version 1.51w) and the plugin ObjectJ (version 1.04r), which allows marking structures in the original images in a non-destructive manner. The pixel length of structures was converted into µm, by calibrating the different microscope setups using a stage micrometer. For comparative analysis we transformed body area ($\log_{10}$ of the square-root) and $\log_{10}$ transformed all linear measure (stylet length, width of the proximal opening, width of the distal opening, sperm length, bristle length). The sample sizes for all quantitative measurements are given in Table S1.

Sequence data generation

We extracted both DNA and RNA from the RNAlater samples using the Nucleospin XS Kit in combination with the Nucleospin RNA/DNA Buffer Set (Macherey-Nagel) and we extracted DNA from the ethanol samples using the DNeasy Blood and Tissue kit (Qiagen, Germany). Extracted DNA and RNA was stored at -80°C. We amplified a partial *28S rRNA* sequence from DNA samples using PCR primers ZX-1 and 1500R and, for some fragments, additional nested PCR using primers ZX-1 and 1200R, or 300F and 1500R, with polymerases and cycling conditions as previously described (Schärer et al. 2011, 2020). We sequenced the resulting fragments from both sides using the amplification primers (Microsynth, Switzerland). For some sequences, we obtained additional sequences using internal primers 1090F, ECD2 (both Schärer et

al., 2011). To generate the RNA-Seq library for *M. clavituba* we extracted RNA from 40 pooled animals using Tri™ reagent (Sigma) and then prepared the library using the TruSeq® Stranded mRNA kit (Illumina). We generated RNA-Seq libraries for all other selected RNA samples using SMART-Seq v4 (Clontech Laboratories, Inc.) in combination with the Nextera XT DNA library preparation kit (Illumina). When at least 5ng mRNA was available we preformed the SMART-Seq v4 protocol with 12 preamplification steps and otherwise used 1ng and performed 15 preamplification steps. Libraries were checked for quality using a Bioanalyzer (Agilent) and then sequenced as 101 paired-end reads on the HiSeq2500 platform (using the HiSeq® SBS Kit v4, Illumina) at the Genomics Facility Basel of the Department of Biosystems Science and Engineering of the ETH Zürich in Basel.

Species delimitation and assignment to an inferred mating syndrome

Most specimens collected for this study could not be assigned to previously described *Macrostomum* species (http://turbellaria.umaine.edu) and they are thus likely new to science. While we have recently described three *Macrostomum* species (Schärer et al. 2020) and are currently working on describing several more (Brand et al. in prep), the focus of our study is not on taxonomy and thus we do not aim at formally describing all these new species here, but instead treat them as operational taxonomic units. However, since we deposit extensive image, video, geographic and molecular data for all specimens collected and their operational species assignment (see Table S7), we think that it will be possible for taxonomists to formally describe these species in the future.

Our species delimitation approach integrates both morphological and molecular data. Although we have whole transcriptome level information for a representative of most species, it was not possible to conduct RNA-Seq on all >1600 collected specimens. Additionally, some specimens were stored in EtOH, from which RNA-Seq is not possible. When we considered the morphology of a species to be diagnostic, we sequenced several specimens, when possible from different sample locations, to confirm that they were indeed molecularly similar and then assigned additional specimens based on morphology only. When no unambiguous diagnostic characters could be defined (as was the case for many species of the hypodermic clade) we sequenced all specimens collected for molecular assignment.

To molecularly delimit species, we constructed haplotype networks from partial *28S rRNA* sequences using the TCS algorithm (Templeton et al. 1992) implemented in the TCS software (Clement et al. 2000). This fragment is widely used as a DNA barcode for flatworms (e.g. Chambrier et al. 2004, 2015; Janssen et al. 2015; Scarpa et al. 2015). We selected 668 sequences representing the available sequence diversity within the genus *Macrostomum*. Most sequences used were generated for this study

(Accessions: MT428556-MT429159), but we also included 64 publicly available sequences. Moreover, we also included 15 sequences from seven species of our chosen outgroup genus *Psammomacrostomum*. We aligned all *28S rRNA* sequences using mafft („—genafpair –maxiterate 1000“) and generated haplotype networks for the recovered clades. We chose to ignore indels to avoid the generation of cryptic species solely based on these, since scoring indels is non-trivial and they are thus frequently treated as missing data (Simmons et al. 2007). To ignore indels we removed all columns from the alignment that contained at least one gap before running TCS, leaving us with an alignment with 787 sites and 385 variable bases.

We delimit species if they have >3 mutational differences in the haplotype network. Such a difference in this *28S rRNA* fragment is highly indicative of distinct species among microturbellarians (Scarpa et al. 2015). However, this fragment does not evolve fast enough to clearly distinguish very recent divergences within *Macrostomum* and COI is frequently used as an additional marker (e.g. Schärer et al. 2011; Janssen et al. 2015). In spite of considerable efforts to also sequence COI, we were unable to develop universal primers (a common issue in flatworms, see e.g. Vanhove et al. 2013) and individual primer optimization (Schärer et al. 2020) for this large number of species was not feasible. Instead, we therefore also delimit species with ≤3 differences if they showed clear diagnostic differences in morphology, in particular in the morphology of the male copulatory stylet or the morphology of the sperm or both. These operational species definitions allow us to classify many species (with multiple specimens) for downstream analysis. Further investigations with more variable molecular markers will possibly recover a higher number of species. As such we err on the side of lumping specimens rather than splitting them, with species with shallow molecular and no diagnostic morphological divergence being assigned to the same species. We provide haplotype networks for all species accompanied by drawings of the diagnostic features (see Supporting information: Figure SI16 – SI32).

To integrate the available morphological information and assign each species to the most likely mating syndrome we adapted the original definitions of Schärer et. al. (2011), since they involved both morphological and behavioural traits and we here lacked behavioural data for most species. Instead, we relied on scoring the location of received sperm in the body in combination with observations on morphological characters to assign species to a morphologically-inferred mating syndrome (a study of 64 species has recently shown that received sperm location is a good predictor of mating behaviour, Singh et al. in preparation). While observing received sperm in the female antrum of an animal may not exclude occasional traumatic mating, it is a strong indication of the reciprocal mating syndrome, especially when it occurs in a species with a blunt stylet, since HI requires a sufficiently sharp stylet to penetrate the tissue of the partner. We therefore assigned all species with received sperm in the female antrum and a blunt stylet to the reciprocal mating syndrome (Table 5).

Nevertheless, there are some species that mate reciprocally but still have a sharp stylet (e.g. *M. spirale*), which could possibly wound the partner internally during mating (pers. obs.). Because of this we also assigned species with a sharp stylet to the reciprocal mating syndrome, provided that we observed received sperm in the female antrum, and that they had sperm with bristles and a thickened antrum (Table 5). Further, we assigned two species to the reciprocal mating syndrome, even though they did not have visible sperm bristles, because they had a very muscular antrum and a blunt stylet (as well as somewhat unusual sperm morphologies, see Results and Discussion). We assigned species to the hypodermic mating syndrome if we found hypodermic received sperm exclusively (as opposed to both hypodermic sperm and received sperm in the female antrum, which we classify as intermediate) as this is strong evidence for HI. Because hypodermic sperm can often be difficult to observe, especially in species with low investment into sperm production, we also assigned species with no received sperm observation to the hypodermic mating syndrome based on their morphology alone (Table 5). We do this since our analysis of correlated evolution showed a strong association between received sperm location and sperm bristle state as well as antrum type (see Results). Since the inferred mating syndrome is based on morphology alone in these cases, it should be considered a more inclusive classification, in contrast to the more conservative assignment based on received sperm location only.

Table 2 Assignment of the morphologically-inferred mating syndrome. Species were assigned based to a mating syndrome based on the female antrum state (simple or thickened), the shape of the distal thickening of the stylet (sharp or blunt), and the sperm bristle state (absent, reduced or present) as well as the location of received sperm in the body (An: in the female antrum only, Hy: hypodermic only, Both: in the female antrum and hypodermic, NA: no observation). 26 species with either not enough (22 species) or contradictory (four species) information were not assigned. Note, that all species assigned to the reciprocal mating syndrome had a thickened female antrum, but this was not a condition for their assignment.

	Morphology			**Received sperm location**				
Syndrome	**Antrum**	**Stylet**	**Sperm bristle**	**An**	**Hy**	**Both**	**NA**	**N**
Reciprocal	(Thickened)	Blunt	Any state	61			6	67
Reciprocal	(Thickened)	Sharp	Present	8				8
Intermediate	Thickened	Sharp	Reduced			2		2
Hypodermic	Simple	Sharp	Reduced/absent		24		18	42
Unclear	Other combinations			7			19	26
			Total	76	24	2	43	145

Phylogenetics

For our phylogenomic analysis we used 134 transcriptomes representing 105 species, including four distant outgroups, *Haplopharynx rostratum*, *Microstomum lineare*, *Myozonaria bistylifera*, and *Karlingia* sp. 1, three species from the sister genus,

sp. 112

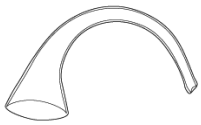
Psammomacrostomum, and 98 *Macrostomum* species (see also Janssen et al. 2015). This included four publicly available high-quality transcriptomes of *M. hystrix*, *M. spirale*, and *M. pusillum* (Brand et al. in preparation), and the main model species *M. lignano* (Wudarski et al. 2017; Grudniewska et al. 2018), as well as 130 newly generated and *de novo* assembled transcriptomes. Transcriptomes were assembled as previously described (Brand et al. in preparation) and most derived from single specimens, while nine were generated by combining RNA-Seq data sets from several animals or pooling animals into one sample that were collected at the same location and assigned to the species based on our taxonomic expertise (for details on the transcriptomes used see Table S8). We assessed transcriptome quality using TransRate (version 1.0.2, Smith-Unna et al. 2016), which maps the reads back to the assembly and calculates mapping metrics, and BUSCO (version 2.0, Waterhouse et al. 2017), which searches for the presence of a curated set of core conserved genes. Specifically, we ran the BUSCO analysis with the metazoan dataset consisting of 978 genes (version uploaded 2016-11-01). These BUSCO scores were also used to select one representative transcriptome when multiple transcriptomes were available for a species (see below). We determined the empirical insert size of our libraries by mapping the reads to the assemblies using SNAP (version 1.0, Zaharia et al. 2011) and then extracting the mean insert size using Picard (version 2.20.2).

To infer a set of orthologous genes (orthologs) for phylogenomic analysis we first used TransDecoder (version 2.0.1, Haas et al. 2013) to predict open reading frames for each transcriptome using Pfam searches (version 32.0) to retain transcripts with predicted proteins and kept only one ORF per transcript using the "--single_best_only" option and then clustered the resulting predicted proteins using the CD-HIT clustering algorithm (version 4.7, Fu et al. 2012), set to cluster amino acid sequences with at least 99.5% sequence identity. The resulting amino acid sequences where then processed with OrthoFinder (version 2.2.6, Emms and Kelly 2015) using the „-os" option to perform a length-adjusted reciprocal BLAST searches followed by MCL clustering. We then processed the resulting set of homologous genes (homogroups) using modified scripts from the phylogenomic dataset construction (PDC) workflow (Yang and Smith 2014). We aligned all sequences of homogroups that contained at least 10 species using MAFFT (version 7.310, „–genafpair –maxiterate 1000", Nakamura et al. 2018), inferred the best substitution model with ModelFinder (Kalyaanamoorthy et al. 2017), and the gene tree using IQ-TREE (version 1.5.5, Nguyen et al. 2015, command: "-mset DCMut, JTTDCMut, LG, mtZOA, PMB, VT, WAG –mfreq FU,F –mrate G"). Then we trimmed the gene trees using "trim_tips.py" to remove tip branches that were longer than two substitutions per site and "maks_tips_by_taxonID_transcripts.py" to remove monophyletic paralogs by choosing the sequence with the best representation in the alignment. We then split off subtrees with long internal branches using

"cut_long_internal_branches.py" and finally inferred our orthologs using the rooted outgroup method in "prune_paralogs_RT.py". This method uses known outgroup taxa to root the phylogenies, which then allows to infer the history of speciation and duplication and extract the most inclusive set of orthologs. We defined *Haplopharynx rostratum*, *Microstomum lineare*, *Karlingia sp. 1* and *Myozonaria bistylifera* as the outgroup and all *Macrostomum* and *Psammomacrostomum* as an ingroup (following the phylogeny of Janssen et al. 2015), since the algorithm does not include the defined outgroup in the output and we could thus use *Psammomacrostomum* to root our final ortholog trees. We realigned the orthogroups using MAFFT and trimmed the alignment with ZORRO (Wu et al. 2012), discarding any columns in the alignment with a score of less than five and filtering alignments that were shorter than 50 amino acids after trimming. Finally, we inferred ortholog gene trees with 100 non-parametric bootstraps with IQ-TREE, by inferring the best fitting model „-mset DCMut, JTTDCMut, LG, mtZOA, PMB, VT, WAG -mfreq FU,F -mrate E,I,G,I+G“. These best fitting substitution models were later also used for the partitioned maximum-likelihood analysis.

We generated two gene matrices. One matrix containing many genes but a relatively moderate occupancy (called *L* for low occupancy) and one with a lower gene number but a higher occupancy (called *H* for high occupancy, Table 2). We conducted most analysis on the *H* alignment for computational tractability and since missing data can have a negative influence on tree inference, particularly when using Bayesian methods (Roure et al. 2013, but see Tan et al. 2015), and we only applied the faster summary and maximum-likelihood methods to *L* alignment. For both alignments, we calculated a maximum likelihood species phylogeny using IQ-TREE with a partition for each gene and using the best substitution model for each gene, as previously determined for the gene trees (see above). We calculated 1000 ultrafast bootstraps and conducted an approximate-likelihood ratio test to assess branch support. We further used IQ-TREE to apply a maximum likelihood approximation of the CAT model using posterior mean site frequency (Wang et al. 2018) allowing for 20 rates since this should reveal if the partitioned alignment is biased due to model misspecifications.

We further inferred Bayesian phylogenies using ExaBayes (version 1.5, Aberer et al. 2014) for the *H* alignment (H-ExaBayes) with partitions for each gene, equal prior probability on all available amino acid substitution models, and with gamma models for all partitions. We ran four independent chains retaining every 500th iteration and discarding the first 365000 iterations as burn-in. We terminated the analysis after 1.46 million generations since the average deviation of split frequencies between all the chains was below 1% indicating convergence. We further assessed convergence using Gelman's R of the likelihoods with the R package coda, which showed that three chains had converged, while one appeared stuck on a local peak. Since all chains quickly converged on the same topology, the local peak probably



occurs due to differences in the substitution models applied with the HIVB model being present at a low rate in the three converged chains but being absent in the fourth. We combined the three converged chains, discarded the fourth and calculated a consensus tree using quadratic path distance optimisation (using the ls.consensus function in phytools).

To account for potential issues caused by model misspecification, we also performed an unpartitioned analysis using the CAT model implemented in PhyloBayes (H-PhyloBayes, version 1.5, Lartillot et al. 2013). Because a full GTR model of the amino acids was too parameter rich we ran the tool on the DNA coding sequence of the *H* alignment and fit the CAT+GTR model. Two chains were run in parallel on 400 CPUs for two weeks. Due to the high cost of the analysis, we terminated the run after 23311 iterations, at which point the chains showed an identical topology (after removal of 10000 iterations as a burn-in). At this point the realised difference in the likelihoods between the two H-PhyloBayes chains was high (0.21 with an ESS of 858) suggesting the chains had not converged. However, Gelman diagnostics for the likelihoods suggested convergence. Our chains could potentially be at a local peak where the hypodermic clade is paraphyletic, possibly due to poor mixing, which could in turn bias these phylogenetic results. Therefore, while we discuss the H-PhyloBayes phylogeny, we exclude it from the comparative analysis.

To construct phylogenies while accounting for potential incomplete lineage sorting, we ran the quartet-based method implemented in ASTRAL III (version: 5.6.3, Zhang et al. 2018) on both alignments. We assessed the level of gene tree - species tree conflict at each node of the phylogeny using the quartet support score, which gives the proportion of quartet trees induced by the gene trees that are supporting the species tree topology as opposed to the two possible alternatives (Zhang et al. 2018). Strong support for the species tree partition can be interpreted as little disagreement between the gene trees, while support for the alternate topologies indicates strong gene tree – species tree conflict.

In order to incorporate information on species for which we lacked transcriptomes, we first, for each transcriptome, chose a representative *28S rRNA* sequence. When available we used the sequence derived from the same sample from which the RNA was extracted, but in few cases we had to choose a sequence from another specimen (see Table S). We then added the sequences from species for which we did not have transcriptomes, choosing a representative sequence from our haplotype network analysis (see above). We aligned the sequences using MAFFT ("--maxiterate 1000 –globalpair") and then trimmed the start and end using trimAl ("–nogaps –terminalonly", Capella-Gutierrez et al. 2009). We inferred the best fitting substitution model for this alignment using ModelFinder with the BIC criterion and then inferred a maximum-likelihood phylogeny, further referred to as C-IQ-TREE (called *C* for combined), which combined this DNA alignment with the *H* amino acid

alignment, with each gene set as its own partition and analyzed it with the best fitting substitution model using IQ-TREE. We again assessed support for each node using both ultrafast bootstraps and an approximate likelihood ratio test. For all comparative analysis we transformed the phylogenies (C-IQ-TREE, H-IQ-TREE and H-ExaBayes) to be ultrametric and with a root depth of 1 using a penalized marginal likelihood approach (Sanderson 2002) implemented in the software TreePL (Smith and O'Meara 2012) since it automates the optimization of the smoothing parameter using cross validation.

To explore possible effects of phylogenetic uncertainty we performed all follow-up analyses on the H-IQ-TREE and the H-ExaBayes phylogenies. Furthermore, we performed the analyses on the C-IQ-TREE phylogeny, since this allowed us to include more species. We pruned the phylogenies to include only one representative tip per species by choosing the most complete transcriptome according to a BUSCO assessment. We present the results obtained with all three phylogenies but discuss only the ones obtained with the C-IQ-TREE phylogeny since all results are quantitatively similar and qualitatively identical.

Convergent evolution of hypodermic insemination

Ancestral state reconstruction

We estimated ancestral states of the three morphological traits that have been associated with the mating syndromes, namely received sperm location, sperm bristle state and antrum thickness, scored as binary traits equivalent to how they were used in the tests for correlated evolution (see below). Since we predicted that losses/reductions of the trait would transition through an intermediate state in the case of received sperm location and sperm bristle state we also reconstructed ancestral states with trinary traits. In this analysis we also included the inferred mating syndrome (see above) also scored as trinary. To estimate ancestral states and the likely history of transitions we used stochastic character mapping (Bollback 2006) implemented in the R package phytools (Revell 2012). First we determined the appropriate transition matrix to be used by fitting MK-models with either all state transitions with equal rates (ER), symmetric rates (SYM), all rates different (ARD), and a model without the possibility of gains once the trait is lost (Dollo). For traits with three states we additionally fit an ordered model where transitions have to go through the intermediate state (ORD) and an ordered Dollo model without the possibility of gains once the trait is completely lost, but allowing transitions back from the intermediate state (ORD-Dollo). We reconstructed ancestral states for all models with a corrected AIC weight >0.15 (Table 3). We used the fully Bayesian implementation of stochastic character mapping with a gamma prior on each transition ($\alpha = 1$, $\beta = 1$, this results in a low prior on the number of transitions) and simulated 10,000 character histories based on those sampled values, with 10,000 simulations burn-in and retaining every 10th simulation. To summarise the number

of transitions we calculated the average number of changes for each transition as well as the 95% credible interval.

Test for correlated evolution

We aimed to test for an association between HI and a simplified female antrum. The best indication for HI is the direct observation of hypodermic received sperm in the tissues of an animal, but since we do not have direct observation of received sperm in all species, we first conducted a correlation test between sperm bristle state and received sperm location, and then tested for correlated evolution between both of these variables and the antrum type. We scored all traits as binary (see "Morphological data" and "Supporting information") and applied Pagel's correlation test (Pagel 1994) as implemented in BayesTraits3 (available for download at: http://www.evolution.rdg.ac.uk/BayesTraitsV3.0.2/BayesTraitsV3.0.2.html). The test compares the marginal likelihood of a model where the transition probability of the first trait depends on the state of the second trait with a model where the two traits are independent. For all analysis we ran four independent MCMC chains for 510 million iterations, discarding the first 10 million iterations as burn-in and retaining every 1000th iteration. We calculated the marginal likelihood of the models using the stepping stone sampler of BayesTraits3 with 1000 power posteriors estimated with 10000 iterations each. We assessed convergence of the chains by calculating Gelman's R using the coda R package (Plummer et al. 2006) and upon confirming convergence merged the chains for further analysis. To evaluate sensitivity to the prior we ran all analysis with a uniform prior (U 0 100), an exponential prior (exp 10) and a reversible-jump hyperprior with a gamma distribution between 0 and 1 for both the rate prior and the hyperprior. We included the last prior for comparison, since it was used in a previous study (Schärer et al. 2011). Models were compared using Bayes factors using the marginal likelihoods calculated using the stepping stone sampler ($BF=2(logLH_{dependent} - logLH_{independent})$). To assess the influence of the phylogeny we conducted these tests on three different phylogenies. This setup resulted in 54 model runs consisting of 216 MCMC chains (three traits with three priors on three topologies with four chains for the independent and the dependent model: 3 x 3 x 3 x 2 x 4 = 216). The number of species included was largest for the correlation between sperm bristle state and antrum type (C-IQ-TREE: 124, H-IQ-TREE & H-ExaBayes: 95) and similar for the other two tests (received sperm location and antrum type: C-IQ-TREE: 100, H-IQ-TREE & H-ExaBayes: 77; received sperm location and sperm bristle state: C-IQ-TREE: 101, H-IQ-TREE & H-ExaBayes: 76).

Principal component analysis

We conducted multivariate analysis to investigate whether the convergent evolution of HI is associated with changes in morphological traits. For this we summarized all our data on sperm, stylet and antrum morphology (both continuous and categorical data, Table S9) using principal component analysis. We then visualised the principal

components and coloured each species according to the mating syndrome it was assigned to and according to received sperm location. Principal component analysis assumes independence of observation, an assumption violated by the phylogenetic relationship of species (Revell 2009). We therefore calculated a phylogenetically corrected principal component (pPCA) using the phyPCA function in phytools with the lambda model. Since we combined data with different scales, we used the correlation matrix for all calculations. When discussing loadings of principal components, we apply an aggressive threshold of 0.5 since although this results in erosion of power it keeps false-positive rate within expectations (Peres-Neto et al. 2003).

Hypodermic insemination and sperm design

To test the influence of HI on sperm length, we performed phylogenetically corrected ordinary least squared regression (PGLS) with the *gls* function in the R package *nlme* (version 3.1). We used *gls* because it allowed us to simultaneously incorporate phylogenetic signal in the residuals and account for variation the number of measured specimens by using the sample size of the response as weights. We determined the best fitting evolutionary model for the covariance in the residuals by comparing corrected AIC of PGLS fitted with Brownian motion, lambda or Ornstein-Uhlenbeck models. We assessed if the assumptions of the PGLS were met by checking the distribution of the phylogeny-corrected residuals for normality and profiled the likelihood of the parameter of the correlation structure (i.e. lambda or alpha). Since R-squared values are problematic for OLS models (Ives 2019) we calculated R_{pred} (Ives and Li 2018) to show model fits. As predictors we used the binary traits included in the test of correlated evolution (received sperm location, sperm bristle state and antrum state) since they all are strong indicators of HI. We also included the inferred mating syndrome as a predictor but coded it also as binary (hypodermic and reciprocal, excluding the intermediate syndrome) due to the low sample size for the intermediate group. Grouping species based on all four of these predictors is rather redundant but we present results from all of them because, while received sperm location is the most direct evidence for HI it results in a reduced sample size (see above) and the other groupings allow the inclusion of more species.

Evolution of reproductive traits within the reciprocal syndrome

If male and female genitalia are coevolving in species of the reciprocal syndrome, we would expect a tight correlation or matching in aspects of their morphology. Ideally this would be detected by identifying homologous male and female structures in all species and correlating their shape evolution (e.g. Simmons and Fitzpatrick 2019). However, this is difficult in *Macrostomum* because the stylets within the reciprocal clade are highly variable making the placement of consistent landmarks between species challenging. Capturing the diversity of the female storage organs is also

challenging because these soft structures are often difficult to see within live specimens due to their high transparency. While they can be seen in histological sections, they are often distorted and contracted due to the fixation, making it impossible to measure length or volume of structures like the female genital opening or the antrum accurately. With these caveats in mind, we performed PGLS (see above) analysis on the combined antrum complexity measure (see Supporting information) with total stylet length. Because stylet length scaled with body size, we also fit a PGLS including body area as a covariate in the model. We further surveyed all collected species for traits that could be relevant in the context of sexual selection and sexual conflict such as aspects of sperm, stylet, and antrum morphology.

Acknowledgment

We thank Ralf A. W. Wiberg for helpful discussions and comments on the manuscript. We thank the numerous people that have helped with field work. Especially, we are grateful for the help of: Benny Glasgow, Peter Ladurner, Gregor Schulte, Christopher Laumer, Walter Salzburger, Adrian Indermaur, Bernd Egger, Fabrizia Ronco, Heinz Büscher, Victoria Huwiler, Philipp Kaufmann, Michaela Zwyer, Stefanie von Fumetti, Joe Ryan, Mark Q. Martindale, Marta Chiodin, John Evans, Leigh Simmons, Mauro Tognon, Piero Tognon, Christiano Tognon, Ralf A. W. Wiberg, Pragya Singh, Nikolas Vellnow, Christian Felber, Ulf Jondelius, Sarah Atherton, Tim Janicke, Georgina Rivera-Ingraham, Ben Byrne, Yvonne Gilbert and Rod Watson. We thank Katja Eschbach for preparing and sequencing RNA-Seq samples. We thank Peter Fields and Lukas Zimmermann for IT advice and support. We thank Dita Vizoso for the use of sperm and stylet drawings for some of the previously described species and for providing inspiration for the new drawings. We thank Yu Zahng and his collaborators for kindly providing specimens of *M. baoanensis*. We thank Nick Goldman and Ziheng Yang for organising the excellent summer school on Computational Molecular Evolution, which has helped and inspired the first author to explore phylogenetics.

bicaudatum

References

Aberer, A. J., K. Kobert, and A. Stamatakis. 2014. Exabayes: Massively parallel bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31:2553–2556.

An-der-Lan, H. 1939. Zur rhabdocoelen Turbellarienfauna des Ochridasees (Balkan). Sitzungsberichte Mathem-Naturw Kl Abt 148 5:196–254.

Anthes, N. 2010. Mate choice and reproductive conflict in simultaneous hermaphrodites. Pp. 329–357 *in* P. Kappeler, ed. Animal Behaviour: Evolution and Mechanisms. Springer Berlin Heidelberg, Berlin, Heidelberg.

Anthes, N., H. Schulenburg, and N. K. Michiels. 2008. Evolutionary links between reproductive morphology, ecology and mating behavior in opisthobranch gastropods. Evolution 62:900–916.

Arnold, S. J. 1994. Bateman's principles and the measurement of sexual selection in plants and animals. Am. Nat. 144:S126–S149.

Arnqvist, G., and L. Rowe. 2002. Antagonistic coevolution between the sexes in a group of insects. Nature 415:787–789.

Arnqvist, G., and L. Rowe. 2005. Sexual conflict. Princeton University Press, Princeton, N.J.

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

Benoit, J. B., A. J. Jajack, and J. A. Yoder. 2012. Multiple traumatic insemination events reduce the ability of bed bug females to maintain water balance. J. Comp. Physiol. B 182:189–198.

Birkhead, T. R., and A. P. Møller (eds). 1998. Sperm competition and sexual selection. Acad. Press, San Diego.

Birkhead, T. R., and T. Pizzari. 2002. Evolution of sex: Postcopulatory sexual selection. Nat. Rev. Genet. 3:262–273.

Bollback, J. P. 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7:88.

Brand, J. N., R. A. W. Wiberg, R. Pjeta, P. Bertemes, C. Beisel, P. Ladurner, and L. Schärer. in preparation. RNA-Seq of three free-living flatworm species suggests rapid evolution of reproduction-related genes.

Capella-Gutierrez, S., J. M. Silla-Martinez, and T. Gabaldon. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

Chambrier, A., M. Zehnder, C. Vaucher, and J. Mariaux. 2004. The evolution of the Proteocephalidea (Platyhelminthes, Eucestoda) based on an enlarged molecular phylogeny, with comments on their uterine development. Syst. Parasitol. 57:159–171.

Chambrier, A. de, A. Waeschenbach, M. Fisseha, T. Scholz, and J. Mariaux. 2015. A large 28S rDNA-based phylogeny confirms the limitations of established morphological characters for classification of proteocephalidean tapeworms (Platyhelminthes, Cestoda). ZooKeys 500:25–59.

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

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

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9:1657–1659.

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

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152.

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

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

Grudniewska, M., S. Mouton, M. Grelling, A. H. G. Wolters, J. Kuipers, B. N. G. Giepmans, and E. Berezikov. 2018. A novel flatworm-specific gene implicated in reproduction in *Macrostomum lignano*. Sci. Rep. 8:3192.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, and A. Regev. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–1512.

Hare, R. M., and L. W. Simmons. 2019. Sexual selection and its evolutionary consequences in female animals. Biol. Rev. 94:929–956.

Immler, S., S. Pitnick, G. A. Parker, K. L. Durrant, S. Lüpold, S. Calhim, and T. R. Birkhead. 2011. Resolving variation in the reproductive tradeoff between sperm size and number. Proc. Natl. Acad. Sci. 108:5325–5330.

Irisarri, I., P. Singh, S. Koblmüller, J. Torres-Dowdall, F. Henning, P. Franchini, C. Fischer, A. R. Lemmon, E. M. Lemmon, G. G. Thallinger, C. Sturmbauer, and A. Meyer. 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. Nat. Commun. 9:3159. Nature Publishing Group.

Ives, A., and D. Li. 2018. rr2: An R package to calculate R^2s for regression models. J. Open Source Softw. 3:1028.

Ives, A. R. 2019. R^2s for Correlated Data: Phylogenetic Models, LMMs, and GLMMs. Syst. Biol. 68:234–251.

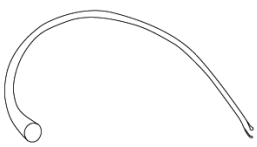
Janicke, T., L. Marie-Orleach, K. De Mulder, E. Berezikov, P. Ladurner, D. B. Vizoso, and L. Schärer. 2013. Sex allocation adjustment to mating group size in a simultaneous hermaphrodite. Evolution 67:3233–3242.

Janicke, T., and L. Schärer. 2010. Sperm competition affects sex allocation but not sperm morphology in a flatworm. Behav. Ecol. Sociobiol. 64:1367–1375.

Janssen, T., D. B. Vizoso, G. Schulte, D. T. J. Littlewood, A. Waeschenbach, and L. Schärer. 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.

Jarne, P., and J. R. Auld. 2006. Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. Evolution 60:1816–1824.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14:587–589.



heyuanensis

Kamimura, Y. 2010. Copulation anatomy of *Drosophila melanogaster* (Diptera: Drosophilidae): wound-making organs and their possible roles. Zoomorphology 129:163–174.

Kamimura, Y. 2007. Twin intromittent organs of *Drosophila* for traumatic insemination. Biol. Lett. 3:401–404.

Kathirithamby, J., M. Hrabar, J. A. Delgado, F. Collantes, S. Dötterl, D. Windsor, and G. Gries. 2015. We do not select, nor are we choosy: reproductive biology of Strepsiptera (Insecta). Biol. J. Linn. Soc. 116:221–238.

Koene, J. M., T. Pförtner, and N. K. Michiels. 2005. Piercing the partner's skin influences sperm uptake in the earthworm *Lumbricus terrestris*. Behav. Ecol. Sociobiol. 59:243.

Koene, J. M., and H. Schulenburg. 2005. Shooting darts: co-evolution and counter-adaptation in hermaphroditic snails. BMC Evol. Biol. 5:25.

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

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

Lartillot, N., N. Rodrigue, D. Stubbs, and J. Richer. 2013. PhyloBayes MPI: Phylogenetic Reconstruction with Infinite Mixtures of Profiles in a Parallel Environment. Syst. Biol. 62:611–615.

Lin, Y., W. Zhou, P. Xiao, Y. Zheng, J. Lu, J. Li, and A. Wang. 2017a. Two new species of freshwater *Macrostomum* (Rhabditophora: Macrostomorpha) found in China. Zootaxa 4329:267.

Lin, Y.-T., W.-T. Feng, F. Xin, L. Zhang, Y. Zhang, and A.-T. Wang. 2017b. Two new species and the molecular phylogeny of eight species of *Macrostomum* (Platyhelminthes: Macrostomorpha) from southern China. Zootaxa 4337:423.

Lüpold, S., M. K. Manier, K. S. Berben, K. J. Smith, B. D. Daley, S. H. Buckley, J. M. Belote, and S. Pitnick. 2012. How multivariate ejaculate traits determine competitive fertilization success in *Drosophila melanogaster*. Curr. Biol. 22:1667–1672.

Manier, M. K., S. Lüpold, J. M. Belote, W. T. Starmer, K. S. Berben, O. Ala-Honkola, W. F. Collins, and S. Pitnick. 2013. Postcopulatory sexual selection generates speciation phenotypes in *Drosophila*. Curr. Biol. 23:1853–1862.

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

Michiels, N. K. 1998. Mating conflicts and sperm competition in simultaneous hermaphrodites. Pp. 219–254 *in* Sperm Competition and Sexual Selection. Elsevier.

Michiels, N. K., and J. M. Koene. 2006. Sexual selection favors harmful mating in hermaphrodites more than in gonochorists. Integr. Comp. Biol. 46:473–480.

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

Miller, G. T., and S. Pitnick. 2002. Sperm-female coevolution in *Drosophila*. Science 298:1230–1233.

Morrow, E. H., and G. Arnqvist. 2003. Costly traumatic insemination and a female counter-adaptation in bed bugs. Proc. R. Soc. B Biol. Sci. 270:2377–2381.

Morrow, E. H., G. Arnqvist, and S. Pitnick. 2003. Adaptation versus pleiotropy: why do males harm their mates? Behav. Ecol. 14:802–806.

Nakamura, T., K. D. Yamada, K. Tomii, K. Katoh, and J. Hancock. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics 34:2490–2492.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268–274.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc R Soc Lond B 255:37–45.

Parker, G. A. 1978. Selection on non-random fusion of gametes during the evolution of anisogamy. J. Theor. Biol. 73:1–28.

Parker, G. A. 1998. Sperm competition and the evolution of ejaculates: towards a theory base. Pp. 3–54 *in* Sperm competition and sexual selection. Academic Press, London, UK.

Parker, G. A. 1993. Sperm competition games: sperm size and sperm number under adult control. Proc. R. Soc. Lond. B Biol. Sci. 253:245–254.

Parker, G. A. 1982. Why are there so many tiny sperm? Sperm competition and the maintenance of two sexes. J. Theor. Biol. 281–294.

Parker, G. A., S. Immler, S. Pitnick, and T. R. Birkhead. 2010. Sperm competition games: Sperm size (mass) and number under raffle and displacement, and the evolution of P2. J. Theor. Biol. 264:1003–1023.

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

Peinert, M., B. Wipfler, G. Jetschke, T. Kleinteich, S. N. Gorb, R. G. Beutel, and H. Pohl. 2016. Traumatic insemination and female counter-adaptation in Strepsiptera (Insecta). Sci. Rep. 6:25052.

Peres-Neto, P. R., D. A. Jackson, and K. M. Somers. 2003. Giving meaningful interpretation to ordination axes: assessing loading significance in principal component analysis. Ecology 84:2347–2363.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnosis and output analysis for MCMC. R News 6:7–11.

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

Ramm, S. A., D. B. Vizoso, and L. Schärer. 2012. Occurrence, costs and heritability of delayed selfing in a free-living flatworm. J. Evol. Biol. 25:2559–2568.

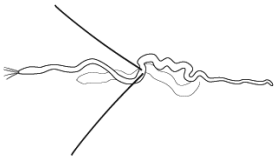
Reinhardt, K., N. Anthes, and R. Lange. 2015. Copulatory wounding and traumatic insemination. Cold Spring Harb. Perspect. Biol. 7:a017582.

Reinhardt, K., R. Naylor, and M. T. Siva-Jothy. 2003. Reducing a cost of traumatic insemination: female bedbugs evolve a unique organ. Proc. R. Soc. B Biol. Sci. 270:2371–2375.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things): phytools: R package. Methods Ecol. Evol. 3:217–223.

Revell, L. J. 2009. Size-correction and principal components for interspecific comparative studies. Evolution 63:3258–3268.

retortum



Řezáč Milan. 2009. The spider *Harpactea sadistica*: Co-evolution of traumatic insemination and complex female genital morphology in spiders. Proc. R. Soc. B Biol. Sci. 276:2697–2701.

Rice, W. R. 1996. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. Nature 381:232–234.

Rohde, K., and A. Faubel. 1997. Spermatogenesis of *Macrostomum pusillum* (Platyhelminthes, Macrostomida). Invertebr. Reprod. Dev. 32:209–215.

Rohde, K., and N. Watson. 1991. Ultrastructure of spermatogenesis and sperm of *Macrostomum tuba*. J Submicrosc Cytol Pathol 23:23–32.

Roure, B., D. Baurain, and H. Philippe. 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. Mol. Biol. Evol. 30:197–214.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

Scarpa, F., P. Cossu, D. Sanna, T. Lai, J. L. Norenburg, M. Curini-Galletti, and M. Casu. 2015. An 18S and 28S-based clock calibration for marine Proseriata (Platyhelminthes). J. Exp. Mar. Biol. Ecol. 463:22–31.

Schärer, L., J. N. Brand, P. Singh, K. S. Zadesenets, C.-P. Stelzer, and G. Viktorin. 2020. A phylogenetically informed search for an alternative *Macrostomum* model species, with notes on taxonomy, mating behavior, karyology, and genome size. J. Zool. Syst. Evol. Res. 58:41–65.

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

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

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

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

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

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

Simmons, L. W., and J. L. Fitzpatrick. 2019. Female genitalia can evolve more rapidly and divergently than male genitalia. Nat. Commun. 10:1312.

Simmons, M. P., K. Müller, and A. P. Norton. 2007. The relative performance of indel-coding methods in simulations. Mol. Phylogenet. Evol. 44:724–740.

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

Singh, P., D. Ballmer, M. Laubscher, and L. Schärer. 2019. Successful mating and hybridisation in two closely related flatworm species despite significant differences in reproductive morphology and behaviour. Animal Behavior and Cognition.

Sluys, R. 1986. First representative of the order *Macrostomida* in Australia (Platyhelminthes, Macrostomidae). Rec. South Aust. Mus. 19:399–404.

paradoxum

Smith, S. A., and B. C. O'Meara. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics 28:2689–2690.

Smith-Unna, R., C. Boursnell, R. Patro, J. M. Hibberd, and S. Kelly. 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Res. 26:1134–1144.

Stutt, A. D., and M. T. Siva-Jothy. 2001. Traumatic insemination and sexual conflict in the bed bug *Cimex lectularius*. Proc. Natl. Acad. Sci. 98:5683–5687.

Sun, T., L. Zhang, A.-T. Wang, and Y. Zhang. 2015. Three new species of freshwater *Macrostomum* (Platyhelminthes, Macrostomida) from southern China. Zootaxa 4012:120–134.

Tan, G., M. Muffato, C. Ledergerber, J. Herrero, N. Goldman, M. Gil, and C. Dessimoz. 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. Syst. Biol. 64:778–791.

Tatarnic, N. J. 2018. Traumatic Insemination and Copulatory Wounding. Pp. B978-0-12-809633-8.20730-9 *in* Reference Module in Life Sciences. Elsevier.

Tatarnic, N. J., and G. Cassis. 2013. Surviving in sympatry: Paragenital divergence and sexual mimicry between a pair of traumatically inseminating plant bugs. Am. Nat. 182:542–551.

Tatarnic, N. J., G. Cassis, and M. T. Siva-Jothy. 2014. Traumatic insemination in terrestrial arthropods. Annu. Rev. Entomol. 59:245–261.

Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132:619–633.

van Velzen, E., L. Scharer, and I. Pen. 2009. The effect of cryptic female choice on sex allocation in simultaneous hermaphrodites. Proc. R. Soc. B Biol. Sci. 276:3123–3131.

Vanhove, M. P. M., B. Tessens, C. Schoelinck, U. Jondelius, D. T. J. Littlewood, T. Artois, and T. Huyse. 2013. Problematic barcoding in flatworms: A case-study on monogeneans and rhabdocoels (Platyhelminthes). ZooKeys 355–379.

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

Wang, A.-T. 2005. Three new species of the genus *Macrostomum* from China (Platyhelminthes, Macrostomida, Macrostomidae). Acta Zootaxonomica Sin. 30:714–720.

Wang, H.-C., B. Q. Minh, E. Susko, and A. J. Roger. 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst. Biol. 67:216–235.

Wang, L., F. Xin, C.-Y. Fang, Y. Zhang, and A.-T. Wang. 2017. Two new brackish-water species of *Macrostomum* (Platyhelminthes, Macrostomida) from mangrove wetland in southern China. Zootaxa 4276:107.

Waterhouse, R. M., M. Seppey, F. A. Simão, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva, and E. M. Zdobnov. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol., doi: 10.1093/molbev/msx319.

Weber, M., B. Patlar, and S. A. Ramm. 2020. Effects of two seminal fluid transcripts on post-mating behaviour in the simultaneously hermaphroditic flatworm *Macrostomum lignano*. J. Evol. Biol. jeb.13606.

Wedell, N., and D. J. Hosken. 2010. The evolution of male and female internal reproductive organs in insects. P. *in* J. L. Leonard and A. Córdoba-Aguilar, eds. The evolution of primary sexual characters in animals. Oxford University Press, Oxford ; New York.

Willems, M., F. Leroux, M. Claeys, M. Boone, S. Mouton, T. Artois, and G. Borgonie. 2009. Ontogeny of the complex sperm in the macrostomid flatworm Macrostomum lignano (Macrostomorpha, Rhabditophora). J. Morphol. 270:162–174.

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

Wu, M., S. Chatterji, and J. A. Eisen. 2012. Accounting for alignment uncertainty in phylogenomics. PLOS ONE 7:e30288.

Wudarski, J., D. Simanov, K. Ustyantsev, K. de Mulder, M. Grelling, M. Grudniewska, F. Beltman, L. Glazenburg, T. Demircan, J. Wunderer, W. Qi, D. B. Vizoso, P. M. Weissert, D. Olivieri, S. Mouton, V. Guryev, A. Aboobaker, L. Schärer, P. Ladurner, and E. Berezikov. 2017. Efficient transgenesis and annotated genome sequence of the regenerative flatworm model *Macrostomum lignano*. Nat. Commun. 8:2120.

Xin, F., S.-Y. Zhang, Y.-S. Shi, L. Wang, Y. Zhang, and A.-T. Wang. 2019. *Macrostomum shenda* and *M. spiriger*, two new brackish-water species of *Macrostomum* (Platyhelminthes: Macrostomorpha) from China. Zootaxa 4603:105.

Yang, Y., and S. A. Smith. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. Mol. Biol. Evol. 31:3081–3092.

Zadesenets, K. S., D. B. Vizoso, A. Schlatter, I. D. Konopatskaia, E. Berezikov, L. Schärer, and N. B. Rubtsov. 2016. Evidence for karyotype polymorphism in the free-living flatworm, *Macrostomum lignano*, a model organism for evolutionary and developmental biology. Plos One 11:e0164915.

Zaharia, M., W. J. Bolosky, K. Curtis, A. Fox, D. Patterson, S. Shenker, I. Stoica, R. M. Karp, and T. Sittler. 2011. Faster and more accurate sequence alignment with SNAP. ArXiv Prepr. ArXiv11115572.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19.

lankouensis

Supporting Figures and Tables

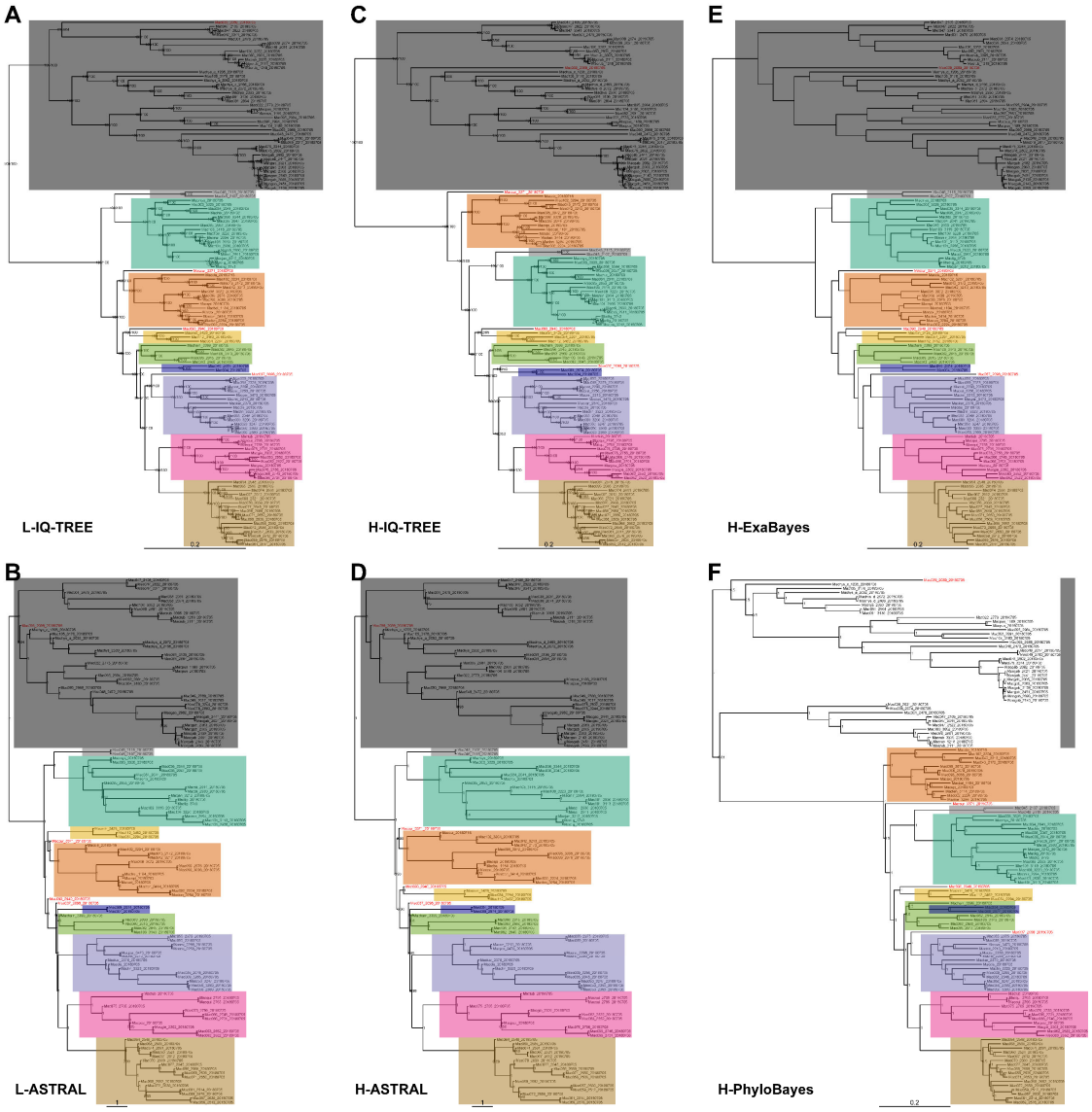


Figure S1 Phylogenetic trees inferred using various methods on the *L* and *H* alignments. Clades are coloured consistently between panels and transcriptomes that are placed differently between approaches are highlighted in red. Species assignments are given as a short abbreviation in the name of the transcriptomes but see Table S8 for details.

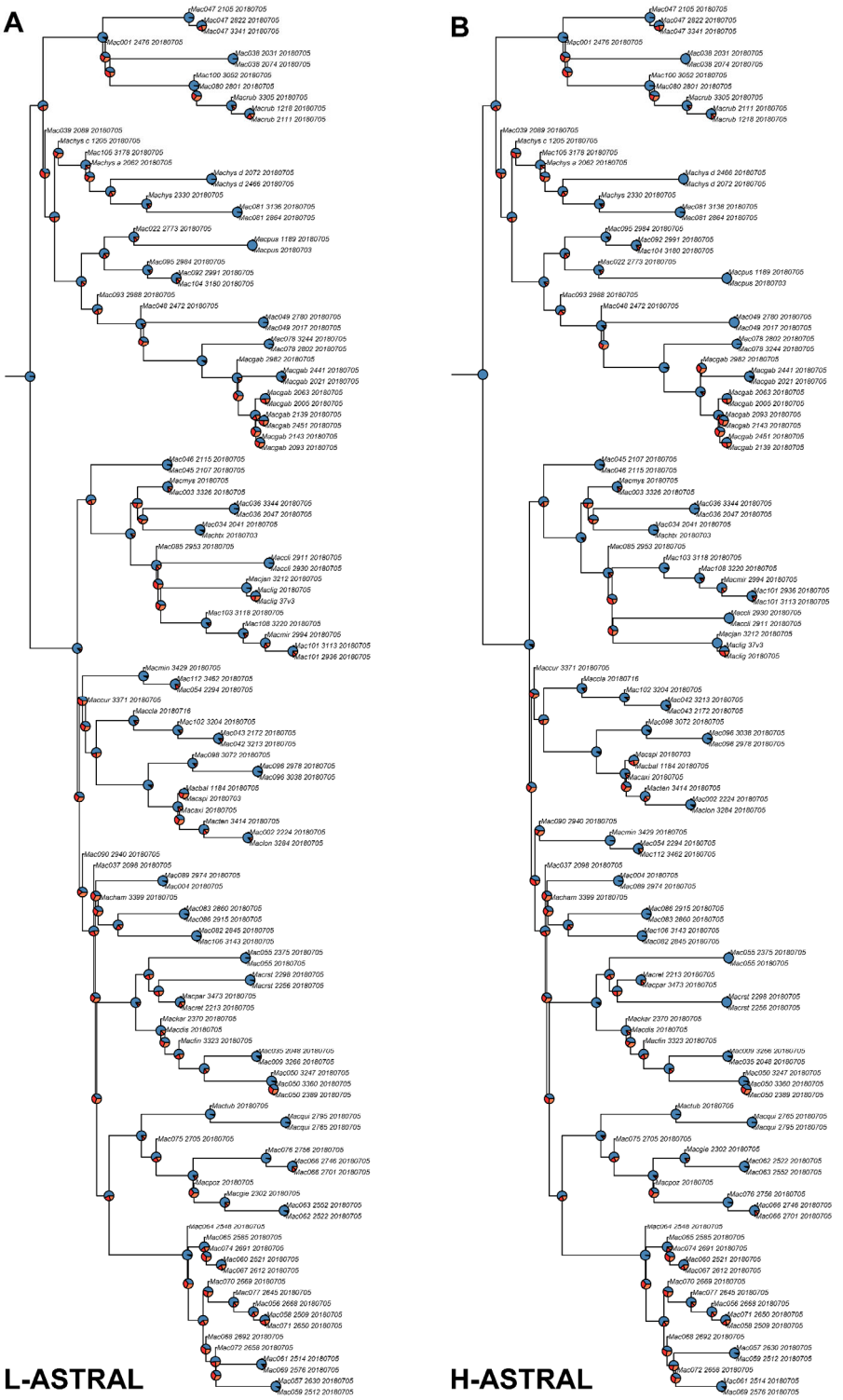
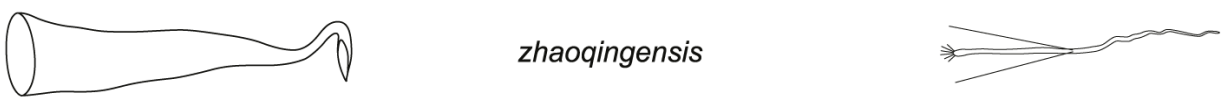
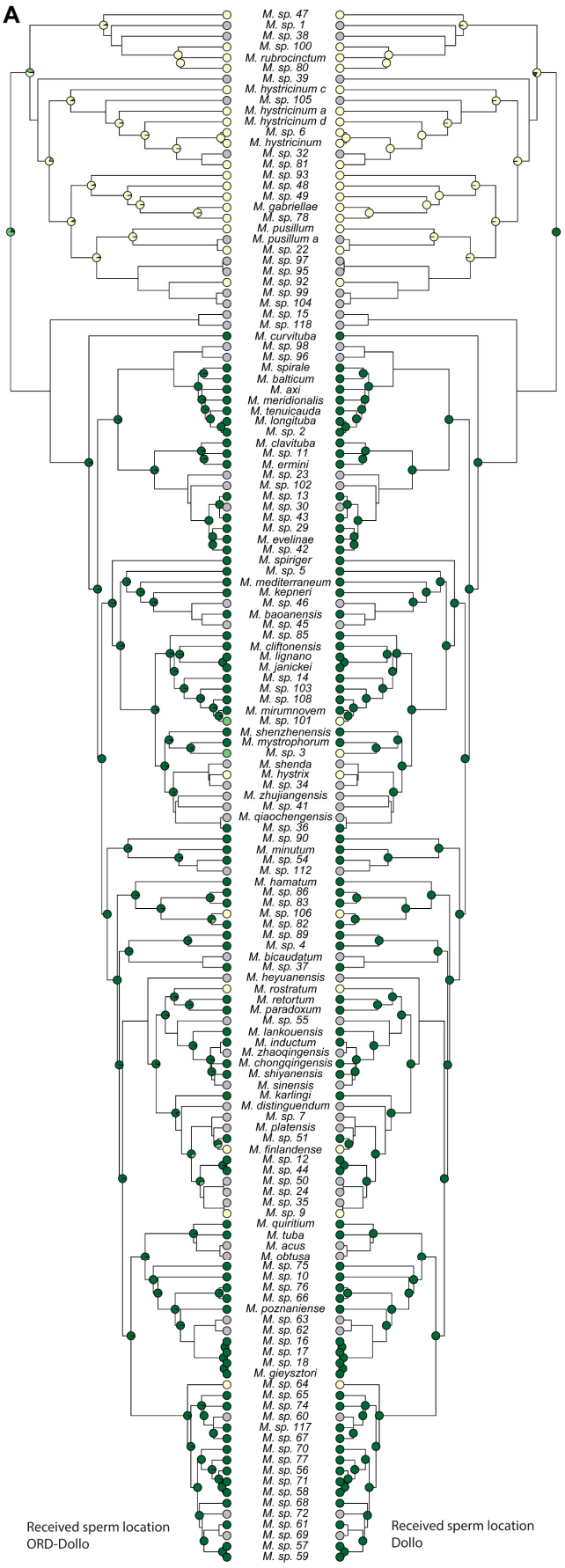
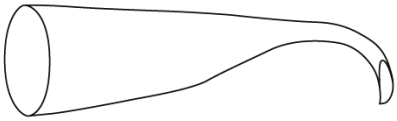


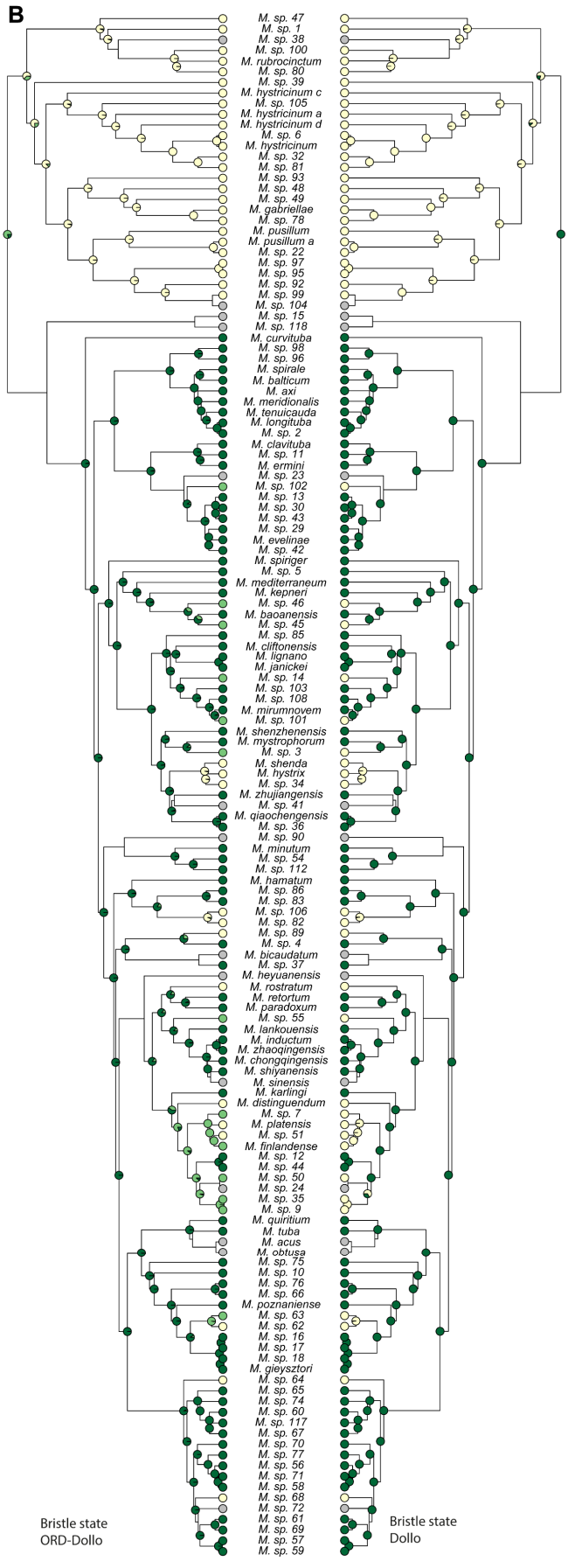
Figure S2 The two ASTRAL phylogenies with quartet support shown for each node. A: L-ASTRAL phylogeny, B: H-ASTRAL phylogeny. The pie charts at the nodes represent the quartet support for the three possible partitions at that node. Blue represents support for the species tree topology and orange and red represent support for one of the two alternative partitions. Mostly blue pie charts indicate little conflict. This analysis indicates that there is substantial gene tree – species tree conflict throughout the genus in particular in the backbone of the reciprocal clade.

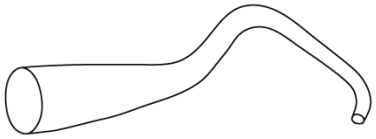




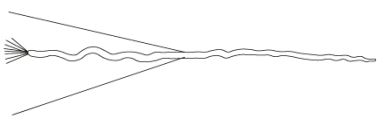


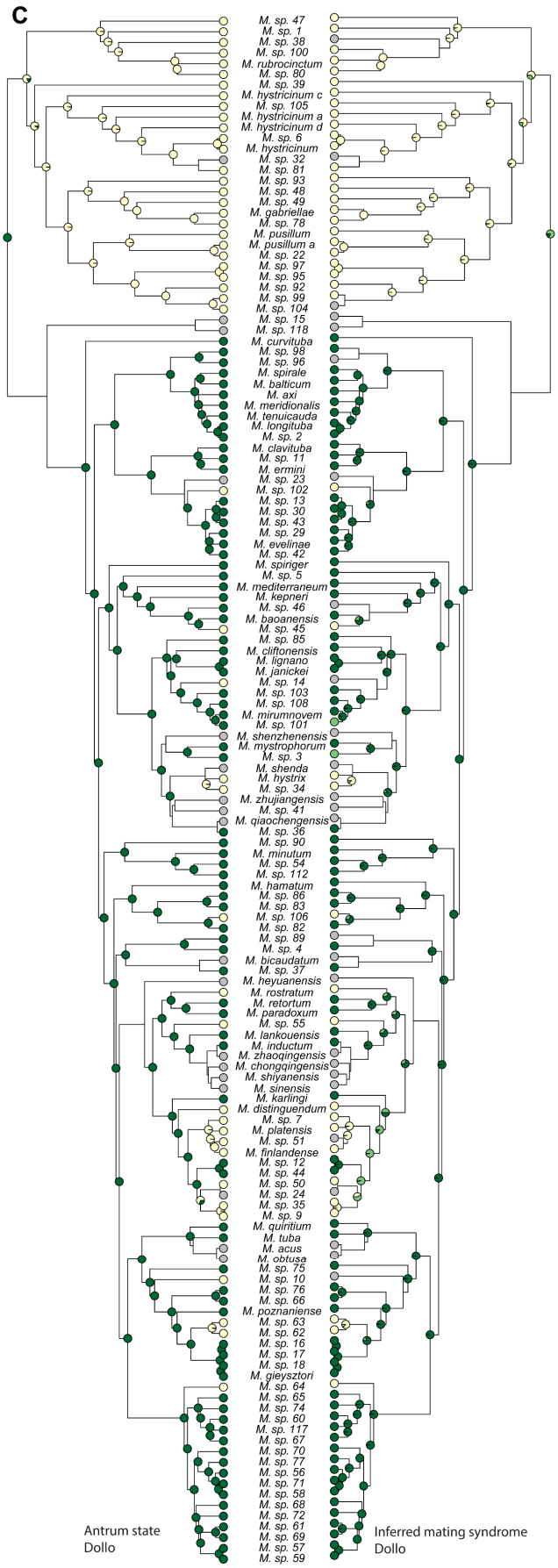
chongqingensis





shiyanensis




C
M. sp. 47
M. sp. 1
M. sp. 38
M. sp. 100
M. rubrocinctum
M. sp. 80
M. sp. 39
M. hystricinum c
M. sp. 105
M. hystricinum a
M. hystricinum d
M. sp. 6
M. hystricinum
M. sp. 32
M. sp. 81
M. sp. 93
M. sp. 48
M. sp. 49
M. gabriellae
M. sp. 78
M. pusillum
M. pusillum a
M. sp. 22
M. sp. 97
M. sp. 95
M. sp. 92
M. sp. 99
M. sp. 104
M. sp. 15
M. sp. 118
M. curvituba
M. sp. 98
M. sp. 96
M. spirale
M. balticum
M. axi
M. meridionalis
M. tenuicauda
M. longituba
M. sp. 2
M. clavituba
M. sp. 11
M. ermini
M. sp. 23
M. sp. 102
M. sp. 13
M. sp. 30
M. sp. 43
M. sp. 29
M. evelinae
M. sp. 42
M. spiriger
M. sp. 5
M. mediterraneum
M. kepneri
M. sp. 46
M. baoanensis
M. sp. 45
M. sp. 85
M. cliftonensis
M. lignano
M. janickei
M. sp. 14
M. sp. 103
M. sp. 108
M. mirumnovem
M. sp. 101
M. shenzhenensis
M. mystrophorum
M. sp. 3
M. shenda
M. hystrix
M. sp. 34
M. zhujiangensis
M. sp. 41
M. qiaochengensis
M. sp. 36
M. sp. 90
M. minutum
M. sp. 54
M. sp. 112
M. hamatum
M. sp. 86
M. sp. 83
M. sp. 106
M. sp. 82
M. sp. 89
M. sp. 4
M. bicaudatum
M. sp. 37
M. heyuanensis
M. rostratum
M. retortum
M. paradoxum
M. sp. 55
M. lankouensis
M. inductum
M. zhaoqingensis
M. chongqingensis
M. shiyanensis
M. sinensis
M. karlingi
M. distinguendum
M. sp. 7
M. platensis
M. sp. 51
M. finlandense
M. sp. 12
M. sp. 44
M. sp. 50
M. sp. 24
M. sp. 35
M. sp. 9
M. quiritium
M. tuba
M. acus
M. obtusa
M. sp. 75
M. sp. 10
M. sp. 76
M. sp. 66
M. poznaniense
M. sp. 63
M. sp. 62
M. sp. 16
M. sp. 17
M. sp. 18
M. gieysztori
M. sp. 64
M. sp. 65
M. sp. 74
M. sp. 60
M. sp. 117
M. sp. 67
M. sp. 70
M. sp. 77
M. sp. 56
M. sp. 71
M. sp. 58
M. sp. 68
M. sp. 72
M. sp. 61
M. sp. 69
M. sp. 57
M. sp. 59
Antrum state
Dollo
Inferred mating syndrome
Dollo

karlingi

Figure S3 Ancestral state reconstructions. Stochastic character mapping is summarised using pie charts representing the proportion of stochastic maps with the character state. Shown is the reconstruction using the best fitting model, which is the ordered model without losses in all cases. Light colours indicate loss of a trait (i.e. yellow, light green and dark green for trinary states; or yellow and dark green for binary states; grey indicating missing data). A: received sperm location, left: trinary (hypodermic, both, in antrum), right: binary (hypodermic, in antrum); B: bristle state left: trinary (absent, reduced, present), right: binary (absent/reduced, present); C: left: antrum state (simple, thickened), right: inferred mating syndrome (hypodermic, intermediate, reciprocal).

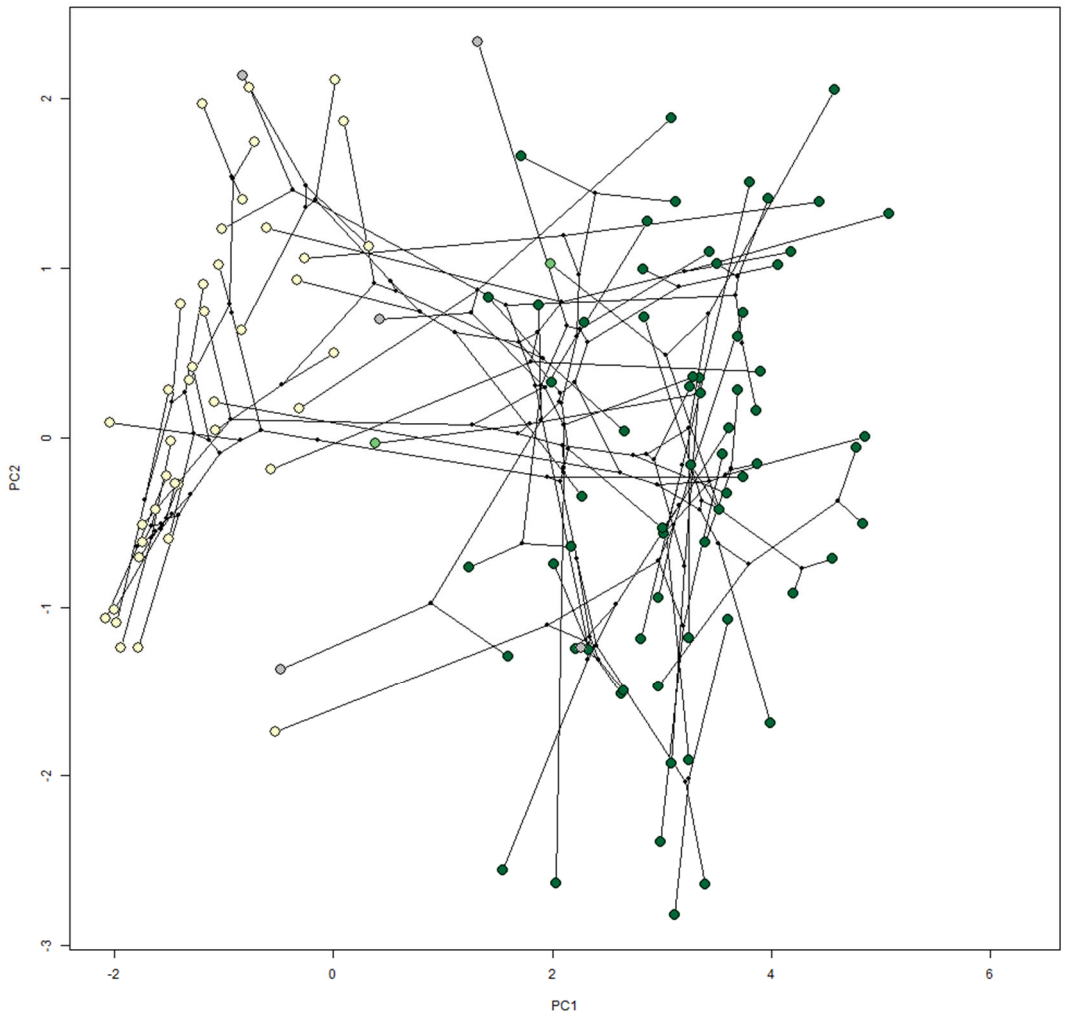
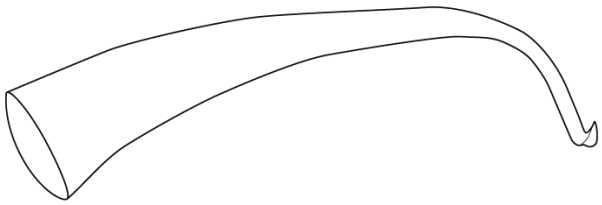


Figure S4 Phylomorphospace represented by PC1 and PC2 of the species in the C-IQ-TREE phylogeny. Calculated using the phylomorphospace function in phytools.



sinensis

Table S1 The sample size for all the included quantitative traits. For each species with measurements the sample size is given for body area (Area), aspects of the stylet (Len: length, Cur: curviness, Prox: width proximal opening, Dist: width distal opening), sperm, and sperm bristles. N_{Sp}: Number of specimens for which the trait was measured, N_{Total}: Number of individual objects measured across all specimens.

		Stylet				Sperm		Bristle	
Species	Area	Len	Cur	Prox	Dist	N_{Sp}	N_{Total}	N_{Sp}	N_{Total}
Macrostomum sp. 1	4	4	4	4	4	3	17	0	0
Macrostomum sp. 2	17	5	5	5	5	3	15	3	17
Macrostomum sp. 3	10	8	8	8	8	3	16	3	17
Macrostomum sp. 4	17	7	7	9	8	8	19	7	19
Macrostomum sp. 5	1	1	2	1	1	1	4	1	5
Macrostomum sp. 6	2	2	2	2	2	1	3	0	0
Macrostomum sp. 7	1	1	1	1	1	1	2	1	1
Macrostomum sp. 9	6	4	5	4	4	4	18	2	9
Macrostomum sp. 10	16	13	13	13	13	5	8	5	15
Macrostomum sp. 11	2	2	2	2	1	1	7	1	5
Macrostomum sp. 12	3	2	2	2	2	1	1	1	2
Macrostomum sp. 13	4	4	4	4	3	1	2	1	2
Macrostomum sp. 14	5	3	3	3	3	2	3	0	0
Macrostomum sp. 16	2	2	2	2	2	1	3	1	2
Macrostomum sp. 17	3	2	2	2	2	3	6	2	9
Macrostomum sp. 18	2	1	1	1	1	3	8	3	7
Macrostomum sp. 22	15	14	14	14	14	10	43	0	0
Macrostomum sp. 23	1	1	1	1	1	0	0	0	0
Macrostomum sp. 24	0	0	2	0	0	0	0	0	0
Macrostomum sp. 29	4	4	4	6	5	3	8	1	3
Macrostomum sp. 34	14	10	10	10	10	3	15	0	0
Macrostomum sp. 35	16	6	6	6	6	3	4	2	2
Macrostomum sp. 36	23	13	13	19	19	4	12	5	13
Macrostomum sp. 37	8	6	6	7	7	2	7	2	6
Macrostomum sp. 38	1	0	0	0	0	0	0	0	0
Macrostomum sp. 39	2	2	2	2	2	1	4	0	0
Macrostomum sp. 41	1	1	1	1	1	0	0	0	0
Macrostomum sp. 42	9	7	7	7	7	2	9	1	6
Macrostomum sp. 43	4	3	3	4	4	1	1	1	1
Macrostomum sp. 44	7	4	4	6	6	3	8	3	8
Macrostomum sp. 45	1	1	1	1	1	1	1	1	1
Macrostomum sp. 46	3	3	3	3	3	1	1	1	1
Macrostomum sp. 47	22	12	12	13	13	8	31	0	0
Macrostomum sp. 48	8	5	5	5	5	4	25	0	0
Macrostomum sp. 49	7	5	5	5	5	3	11	0	0
Macrostomum sp. 50	24	16	16	19	19	9	32	10	33
Macrostomum sp. 51	3	1	1	1	1	1	1	0	0
Macrostomum sp. 54	5	4	4	4	4	2	4	2	4
Macrostomum sp. 55	19	13	13	16	16	7	20	3	8
Macrostomum sp. 56	3	1	1	1	1	0	0	1	1

		Stylet				Sperm		Bristle	
Species	Area	Len	Cur	Prox	Dist	N_{Sp}	N_{Total}	N_{Sp}	N_{Total}
Macrostomum sp. 57	3	1	1	2	2	1	1	1	1
Macrostomum sp. 58	4	4	4	3	4	3	7	2	5
Macrostomum sp. 59	11	10	10	10	10	3	3	2	2
Macrostomum sp. 60	2	2	2	2	2	1	1	1	1
Macrostomum sp. 61	18	18	18	18	18	3	5	3	4
Macrostomum sp. 62	10	6	6	6	6	4	8	0	0
Macrostomum sp. 63	7	7	7	7	7	5	10	1	2
Macrostomum sp. 64	7	7	7	7	7	2	5	0	0
Macrostomum sp. 65	19	14	14	15	15	2	2	3	3
Macrostomum sp. 66	15	10	10	11	11	2	8	2	6
Macrostomum sp. 67	10	9	9	9	9	2	8	3	7
Macrostomum sp. 68	6	5	5	5	5	1	4	0	0
Macrostomum sp. 69	1	1	1	1	1	0	0	1	1
Macrostomum sp. 70	1	1	1	1	1	1	2	1	2
Macrostomum sp. 71	7	6	6	6	6	1	4	1	3
Macrostomum sp. 72	2	2	2	2	2	0	0	0	0
Macrostomum sp. 74	15	12	12	14	14	4	12	4	14
Macrostomum sp. 75	7	3	3	4	4	2	2	2	2
Macrostomum sp. 76	9	5	5	6	6	2	2	2	2
Macrostomum sp. 77	2	1	1	2	2	1	1	1	1
Macrostomum sp. 78	13	11	11	11	11	6	30	0	0
Macrostomum sp. 80	12	7	7	7	7	5	17	0	0
Macrostomum sp. 81	16	10	10	11	10	7	56	0	0
Macrostomum sp. 82	12	5	5	9	9	4	11	0	0
Macrostomum sp. 83	17	7	7	13	13	5	15	5	16
Macrostomum sp. 85	14	10	10	11	11	4	11	4	12
Macrostomum sp. 86	19	10	10	14	14	5	15	5	15
Macrostomum sp. 89	12	7	7	8	8	4	12	0	0
Macrostomum sp. 90	1	1	1	1	1	0	0	0	0
Macrostomum sp. 92	18	16	16	16	16	12	57	0	0
Macrostomum sp. 93	14	14	14	14	14	6	32	0	0
Macrostomum sp. 95	4	4	4	4	4	1	3	0	0
Macrostomum sp. 96	11	6	6	9	9	6	20	6	19
Macrostomum sp. 97	1	1	1	1	1	1	2	0	0
Macrostomum sp. 98	12	9	9	11	11	6	16	6	15
Macrostomum sp. 99	2	2	2	2	2	2	14	0	0
Macrostomum sp. 100	9	7	7	7	7	4	25	0	0
Macrostomum sp. 101	12	7	7	9	9	5	37	5	33
Macrostomum sp. 102	15	14	14	14	14	4	18	4	17
Macrostomum sp. 103	13	12	12	12	10	6	29	6	29
Macrostomum sp. 104	2	0	0	0	0	0	0	0	0
Macrostomum sp. 105	6	6	6	6	6	4	12	0	0
Macrostomum sp. 106	23	15	15	16	16	7	17	0	0
Macrostomum sp. 108	9	5	5	7	7	4	21	4	16
Macrostomum sp. 112	1	1	1	1	1	1	3	1	4

sp. 7

		Stylet				Sperm		Bristle	
Species	Area	Len	Cur	Prox	Dist	N_{Sp}	N_{Total}	N_{Sp}	N_{Total}
Macrostomum sp. 117	9	8	8	9	7	4	9	4	6
M. axi	23	13	13	16	16	3	8	3	7
M. balticum	12	7	7	8	8	5	12	5	25
M. baoanensis	5	5	5	5	5	1	3	1	3
M. clavituba	20	15	15	16	16	7	24	7	22
M. cliftonensis	21	11	11	16	16	6	20	6	20
M. curvituba	17	11	11	17	17	8	26	8	27
M. distinguendum	30	26	26	26	26	7	16	0	0
M. ermini	4	4	4	4	4	2	10	2	12
M. evelinae	1	1	1	1	1	1	3	0	0
M. finlandense	17	13	13	13	13	8	27	8	25
M. gabriellae	47	39	39	39	39	26	108	0	0
M. gieysztori	30	16	16	16	16	5	18	4	19
M. hamatum	15	14	14	15	14	6	27	6	33
M. hystrix	26	8	8	13	13	8	21	0	0
M. hystricinum	15	12	12	12	12	10	65	0	0
M. hystricinum a	1	1	1	1	1	0	0	0	0
M. hystricinum c	6	3	3	4	4	1	5	0	0
M. hystricinum d	12	8	8	10	10	5	17	0	0
M. inductum	7	1	2	1	1	2	6	2	16
M. janickei	9	6	6	8	8	3	11	2	26
M. karlingi	8	7	7	7	7	3	13	3	12
M. kepneri	7	7	7	7	7	6	27	5	24
M. lignano	27	21	23	21	21	12	43	12	42
M. longituba	21	16	16	17	17	7	22	6	23
M. mediterraneum	6	5	5	5	5	2	5	1	4
M. meridionalis	1	0	0	1	1	1	2	1	6
M. minutum	17	20	20	21	21	6	22	5	28
M. mirumnovem	19	11	11	14	13	10	28	10	27
M. mystrophorum	12	11	11	11	11	4	12	4	18
M. paradoxum	39	26	26	27	27	9	14	12	23
M. platensis	1	1	1	1	1	1	3	0	0
M. poznaniense	23	15	15	17	17	6	13	5	11
M. pusillum	11	8	8	8	8	6	22	0	0
M. quiritium	18	7	7	10	10	8	22	8	29
M. retortum	12	5	5	9	9	3	8	4	13
M. rostratum	35	20	20	23	23	14	50	0	0
M. rubrocinctum	17	13	13	13	13	8	26	0	0
M. spirale	18	8	8	10	10	4	17	3	15
M. tenuicauda	28	14	14	14	14	11	37	11	55
M. tuba	16	8	8	8	8	7	16	6	21
Total	1382	949	956	1053	1043	504	1765	294	1021

platensis

Table S2 Robinson-Foulds distance matrix between the various phylogenetic inference methods, the first row gives the alignment used for inference (*L*/*H*) and the second row the inference software used.

		L		***H***			
		IQ-TREE	ASTRAL	IQ-TREE	ASTRAL	ExaBayes	PhyloBayes
L	IQ-TREE	-	16	16	12	8	24
	ASTRAL	16	-	18	12	16	28
H	IQ-TREE	16	18	-	16	8	26
	ASTRAL	12	12	16	-	14	28
	ExaBayes	8	16	8	14	-	24
	PhyloBayes	24	28	26	28	24	-

Table S3 Ancestral state reconstruction using stochastic character mapping.

(Table too large for print, available from the Author)

Table S4 Scores and loadings from the phylogenetically corrected principal component analysis.

(Table too large for print, available from the Author)

Table S5 Results of PGLS analysis of states indicating reciprocal copulation versus hypodermic insemination (HI) on sperm length. All predictors were binary, with the reference level being the state indicating HI.

Predictor	Phylogeny	Variable	N	λ	Est.	SE	*t*	*P*	R^2_{pred}
Received sperm location	H-IQ-TREE	Intercept	70	0.81	1.67	0.04	41.13	<0.001	0.65
		ß			0.20	0.04	4.90	<0.001	
	H-ExaBayes	Intercept	70	0.82	1.67	0.04	41.10	<0.001	0.65
		ß			0.20	0.04	4.87	<0.001	
	C-IQ-TREE	Intercept	91	0.72	1.68	0.03	49.86	<0.001	0.54
		ß			0.18	0.04	4.77	<0.001	
Bristle state	H-IQ-TREE	Intercept	88	0.68	1.67	0.04	45.79	<0.001	0.64
		ß			0.17	0.03	5.10	<0.001	
	H-ExaBayes	Intercept	88	0.69	1.67	0.04	45.73	<0.001	0.64
		ß			0.17	0.03	5.06	<0.001	
	C-IQ-TREE	Intercept	113	0.69	1.67	0.03	55.83	<0.001	0.6
		ß			0.17	0.03	5.69	<0.001	
Antrum state	H-IQ-TREE	Intercept	88	0.83	1.68	0.04	46.39	<0.001	0.71
		ß			0.23	0.04	6.51	<0.001	
	H-ExaBayes	Intercept	88	1	1.70	0.04	45.63	<0.001	0.73
		ß			0.27	0.04	7.77	<0.001	
	C-IQ-TREE	Intercept	113	0.76	1.67	0.03	58.24	<0.001	0.65
		ß			0.21	0.03	6.99	<0.001	
Inferred mating syndrome	H-IQ-TREE	Intercept	83	1	1.70	0.04	46.00	<0.001	0.73
		ß			0.28	0.04	7.94	<0.001	
	H-ExaBayes	Intercept	83	0.9	1.69	0.04	45.01	<0.001	0.72
		ß			0.24	0.04	6.74	<0.001	
	C-IQ-TREE	Intercept	105	0.8	1.67	0.03	55.96	<0.001	0.66
		ß			0.21	0.03	6.66	<0.001	

Table S6 Results from PGLS correlating antrum complexity with stylet length.

Predictor	Phylogeny	Variable	N	λ	Est.	SE	*t*	*P*	R^2_{pred}
Antrum complexity	H-IQ-TREE	Intercept	55	0.23	1.987	0.11	18.73	<0.001	0.07
		ß			-0.005	0.02	-0.24	0.81	
	H-ExaBayes	Intercept	55	0.24	1.988	0.11	18.89	<0.001	0.06
		ß			-0.005	0.02	-0.23	0.82	
	C-IQ-TREE	Intercept	71	0.22	1.990	0.09	21.74	<0.001	0.06
		ß			-0.005	0.02	-0.25	0.81	
Antrum complexity + body area	H-IQ-TREE	Intercept	55	0.3	-0.165	0.46	-0.36	0.72	0.32
		Complexity			0.007	0.02	0.35	0.73	
		Body area			0.796	0.17	4.74	<0.001	
	H-ExaBayes	Intercept	55	0.33	-0.178	0.46	-0.38	0.70	0.32
		Complexity			0.007	0.02	0.34	0.74	
		Body area			0.802	0.17	4.77	<0.001	
	C-IQ-TREE	Intercept	71	0.43	-0.220	0.43	-0.52	0.61	0.33
		Complexity			0.007	0.02	0.40	0.69	
		Body area			0.820	0.15	5.33	<0.001	

Table S7 Details on all specimens included in this study.
(Table too large for print, available from the Author)

Table S8 Details of the transcriptomes used in this study.
(Table too large for print, available from the Author)

Table S9 Mean species values for all morphological variables.
(Table too large for print, available from the Author)

Supporting Information

Correlated evolution

We performed tests for correlated evolution on three trait combinations with three trees and three priors for each test. Here we first show the mean transition rates inferred (Figure SI1-SI3) and then give the posterior distributions for all transition rates and the likelihood for each combination of parameters (Figure SI4-SI12).

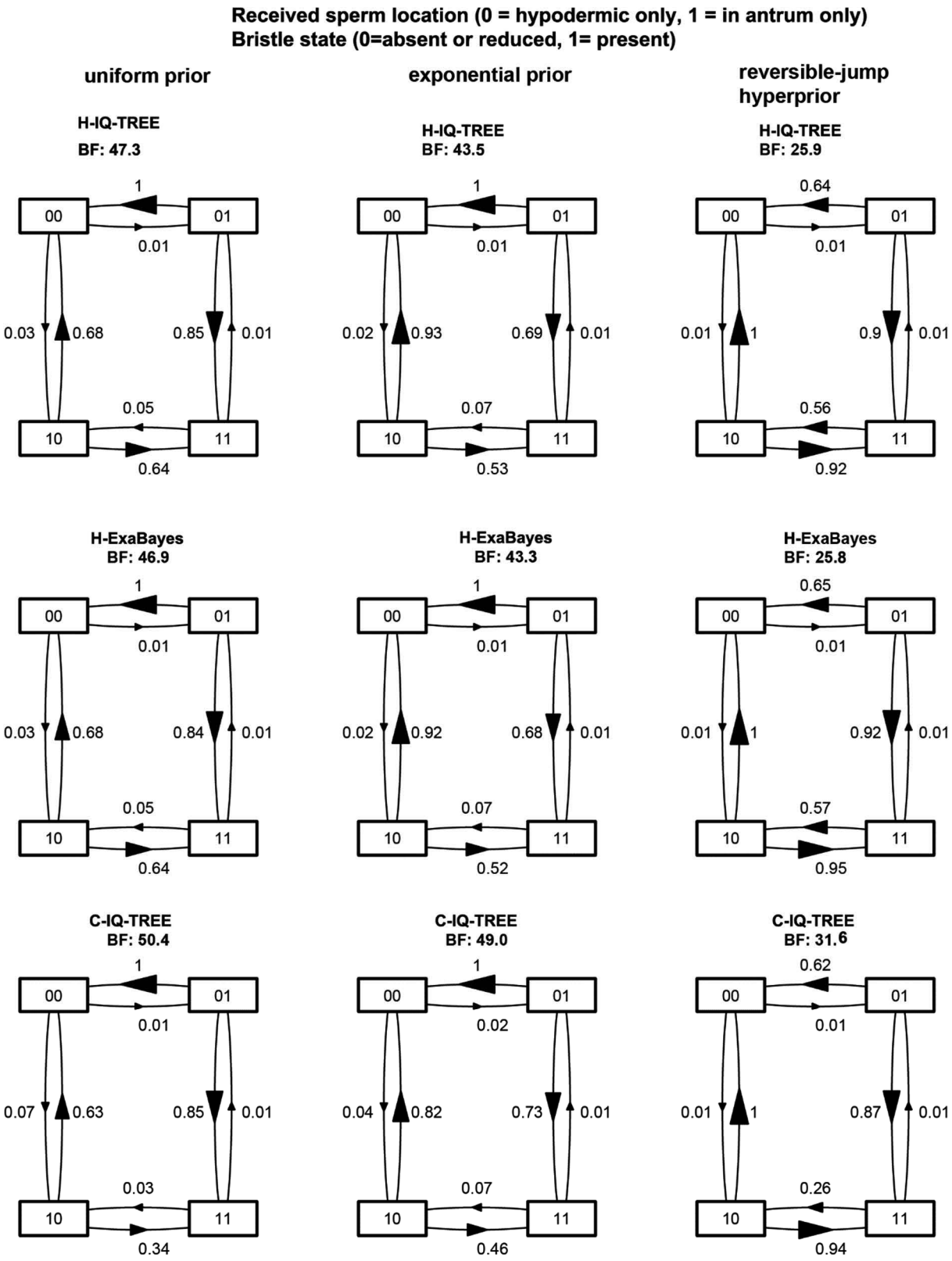


Figure SI1 Visualisation of the transition matrices for received **sperm location** and **bristle state**. Transition rates are scaled so that the largest is unity. Rows show the different phylogenies and columns show different priors.

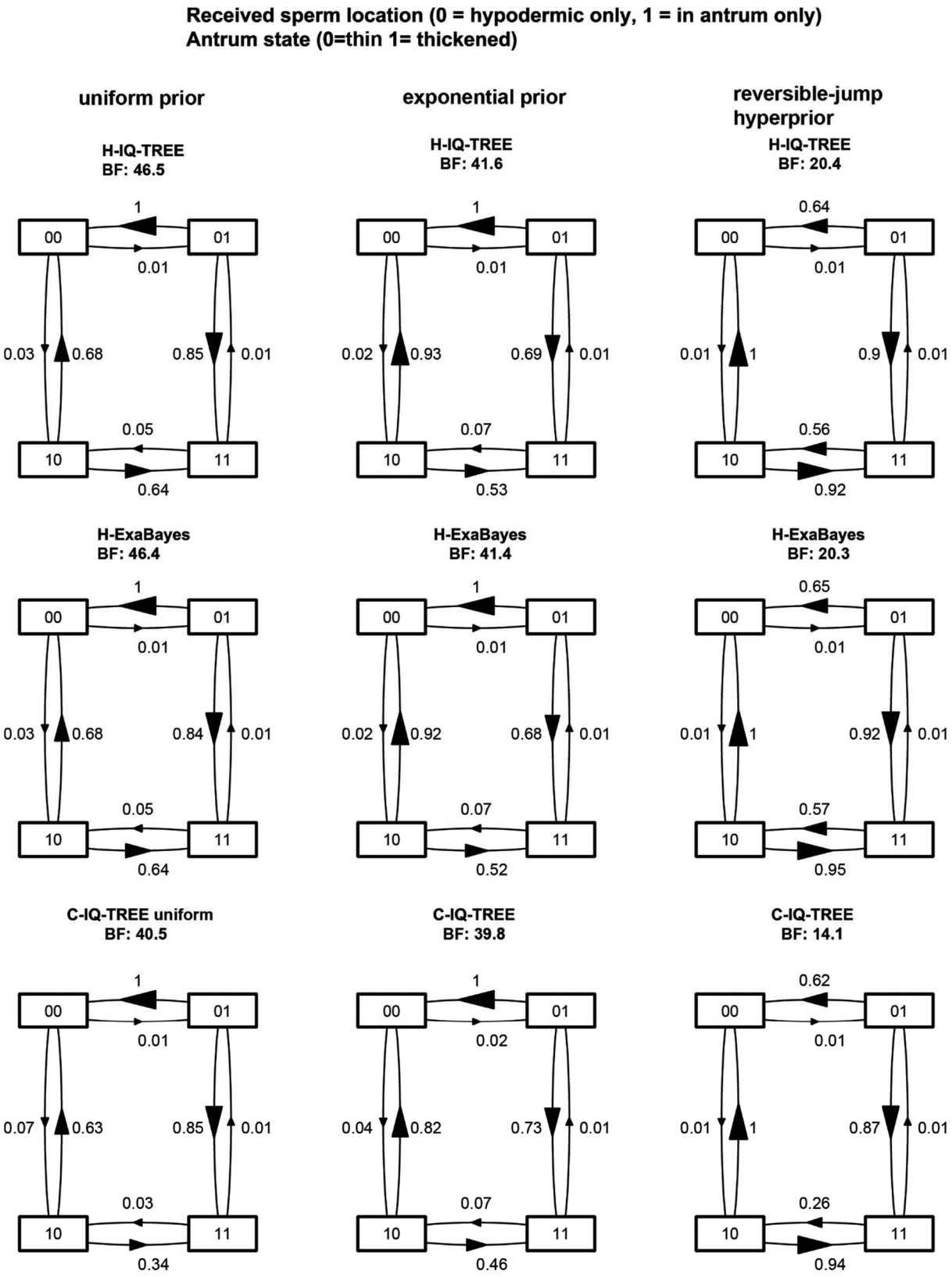


Figure SI2 Visualisation of the transition matrices for received **sperm location** and **antrum state**. Transition rates are scaled so that the largest is unity. Rows show the different phylogenies and columns show different priors.

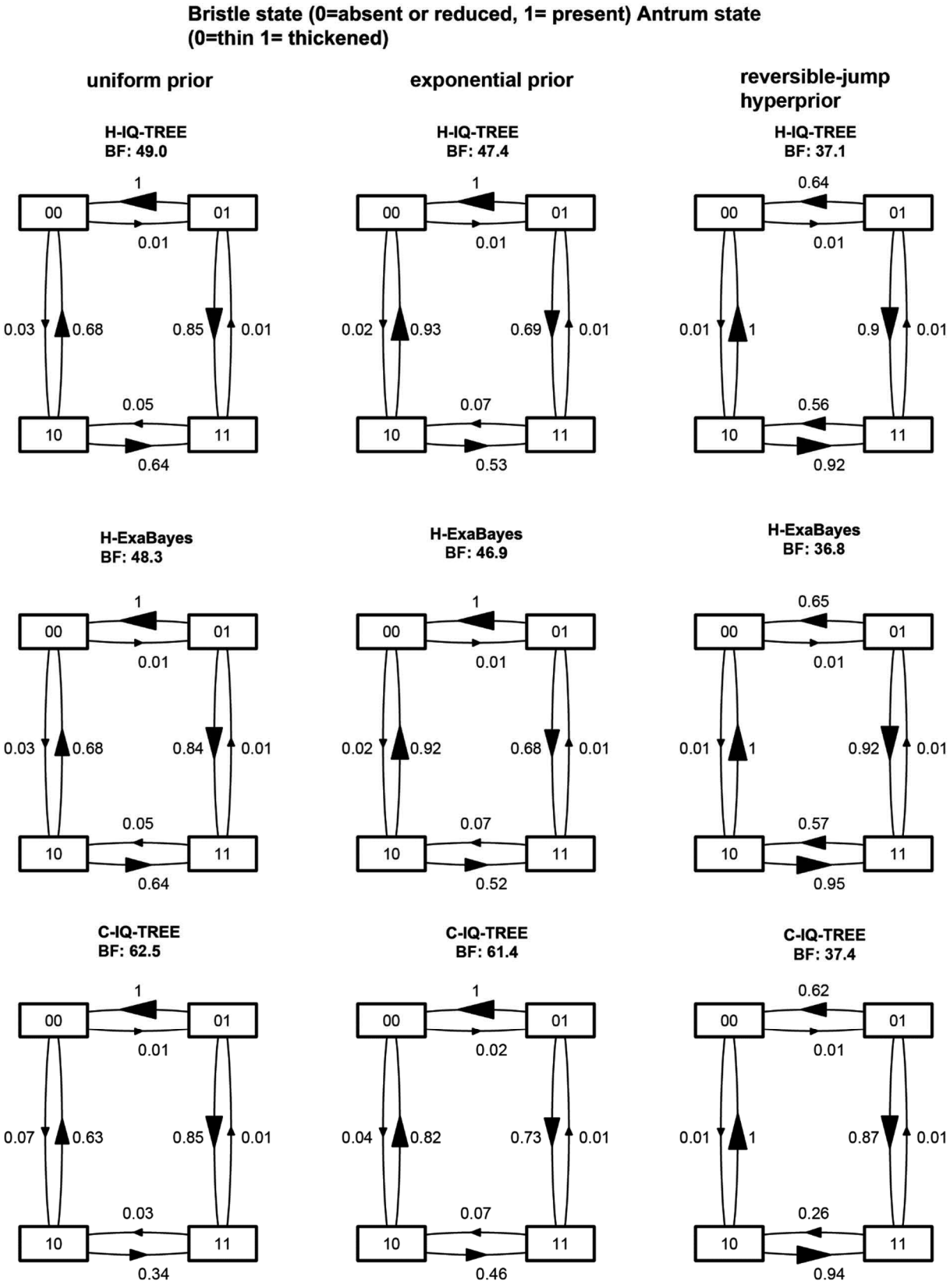


Figure SI3 Visualisation of the transition matrices for **bristle state** and **antrum state**. Transition rates are scaled so that the largest is unity. Rows show the different phylogenies and columns show different priors.

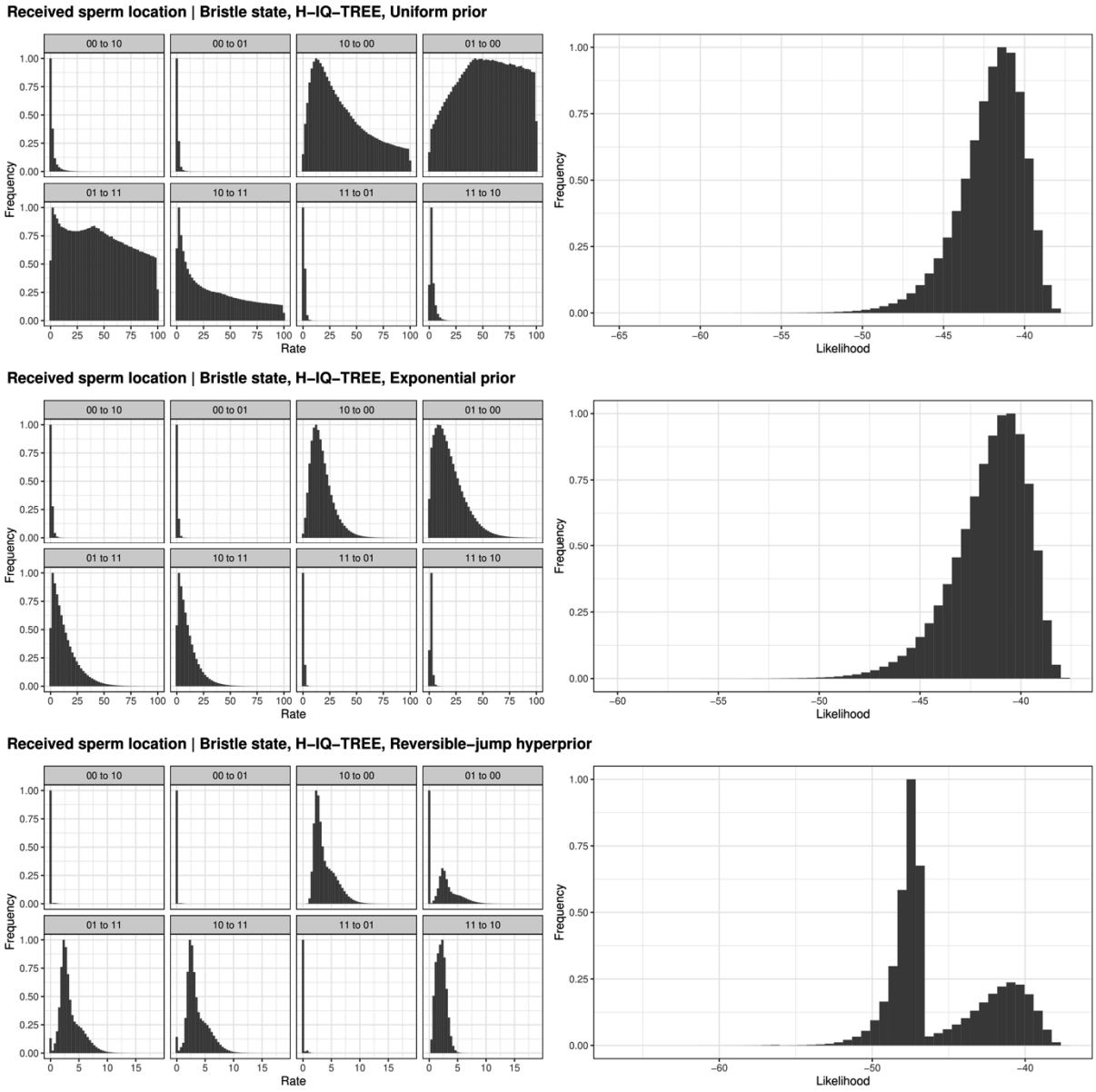
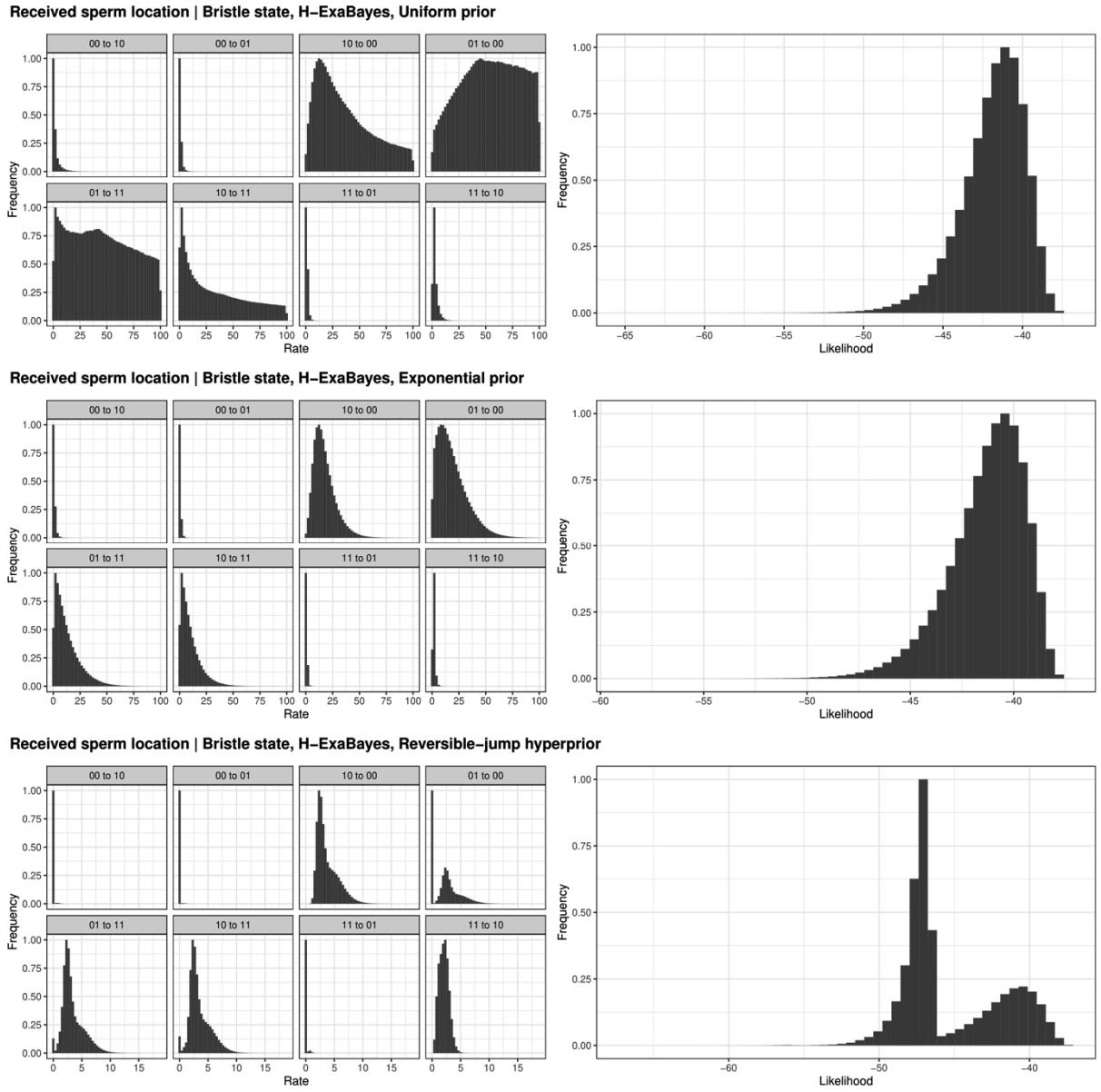


Figure SI4 Posterior distributions of the rate parameters (left) and likelihoods (right) for BayesTraits runs with the H-IQ-TREE phylogeny and received sperm location and bristle state. Histograms show values from $2x10^6$ samples from four converged chains and are scaled so that the largest bin has a height of unity. Rate panels are arranged so that rates that would be equal in the independent model are arrange vertically (e.g. 00 to 10 would be equal to 01 to 11).



SI5 Posterior distributions of the rate parameters (left) and likelihoods (right) for BayesTraits runs with the H-Exabayes phylogeny and received sperm location and bristle state. Layout identical to previous figure.

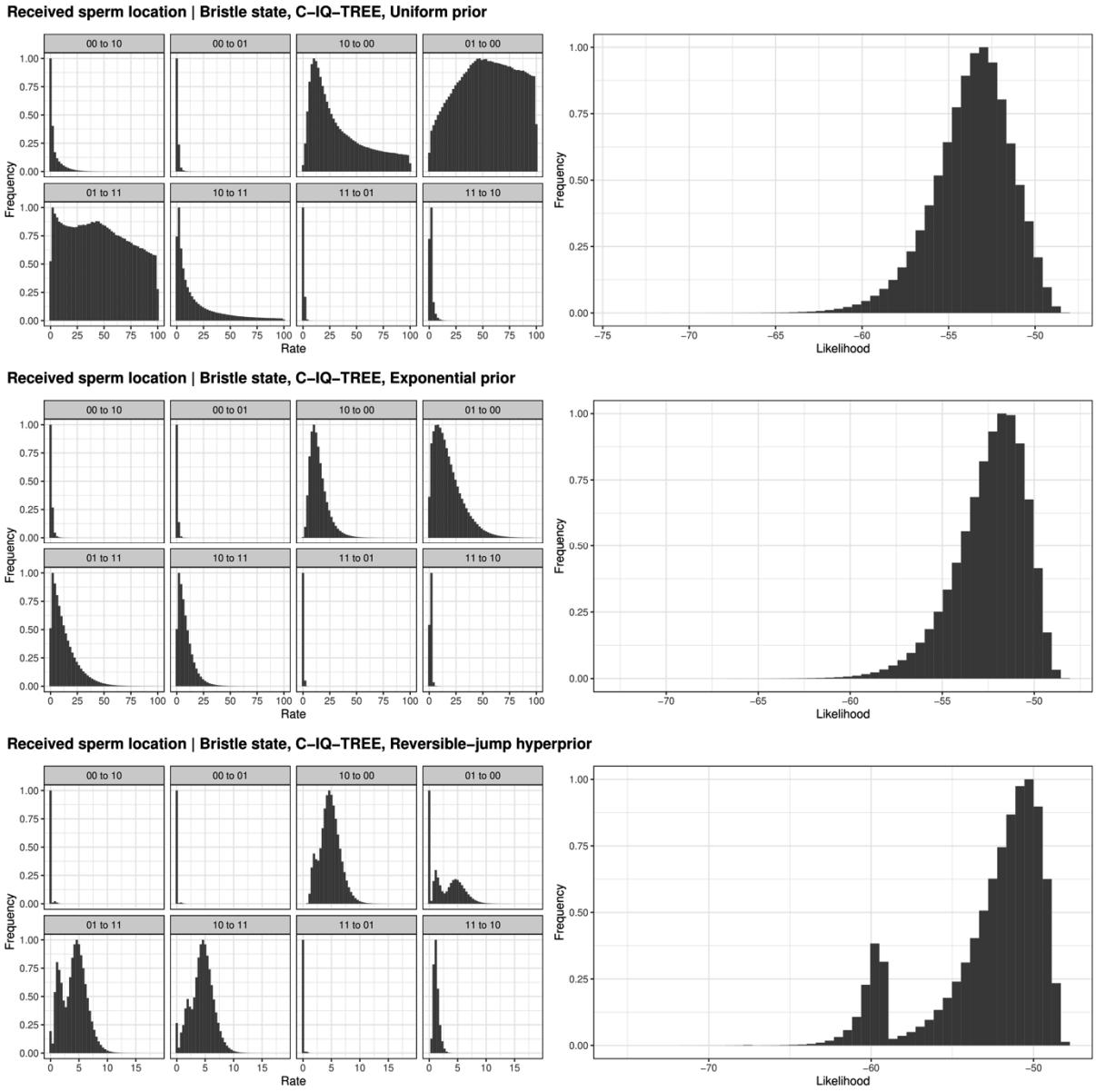


Figure SI6 Posterior distributions of the rate parameters (left) and likelihoods (right) for BayesTraits runs with the C-IQ-TREE phylogeny and received sperm location and bristle state. Layout identical to previous figure.

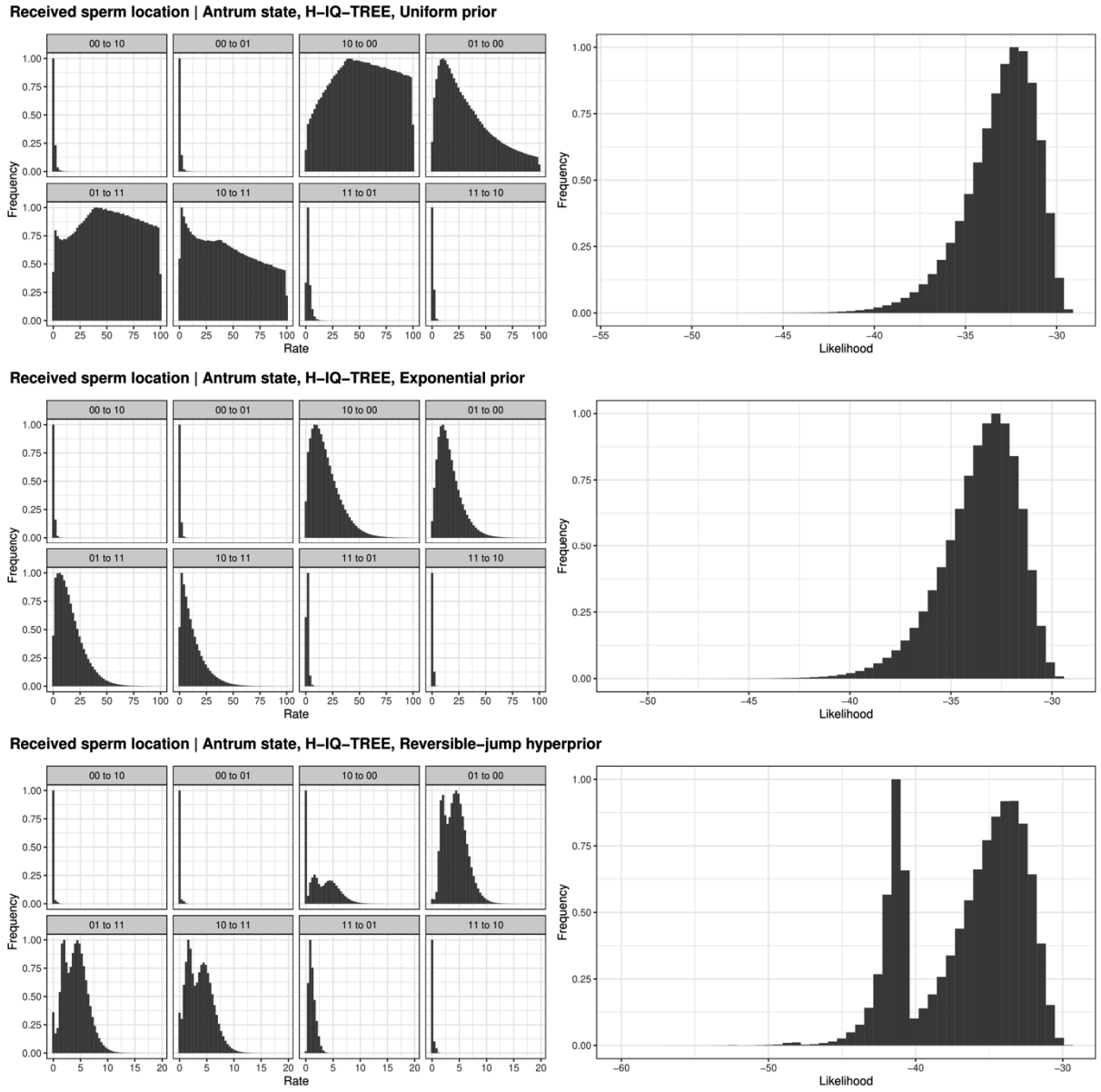
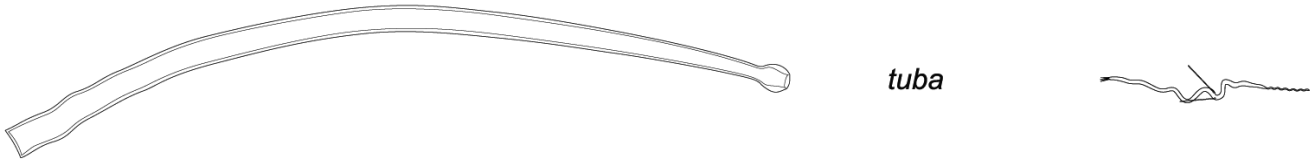


Figure SI7 Posterior distributions of the rate parameters (left) and likelihoods (right) for BayesTraits runs with the H-IQ-TREE phylogeny and received sperm location and antrum state. Layout identical to previous figure.



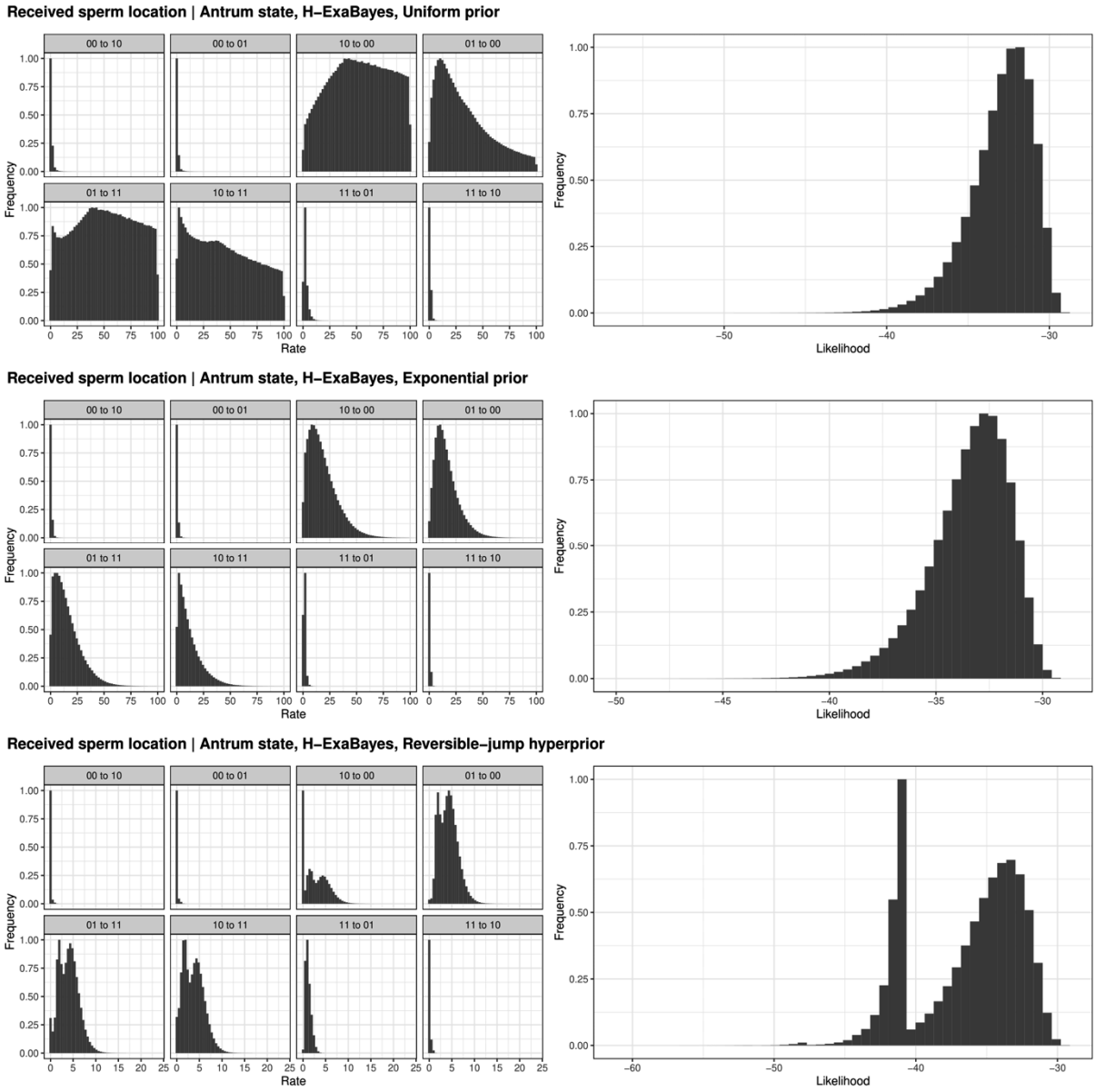


Figure SI8 Posterior distributions of the rate parameters (left) and likelihoods (right) for BayesTraits runs with the H-ExaBayes phylogeny and received sperm location and antrum state. Layout identical to previous figure.

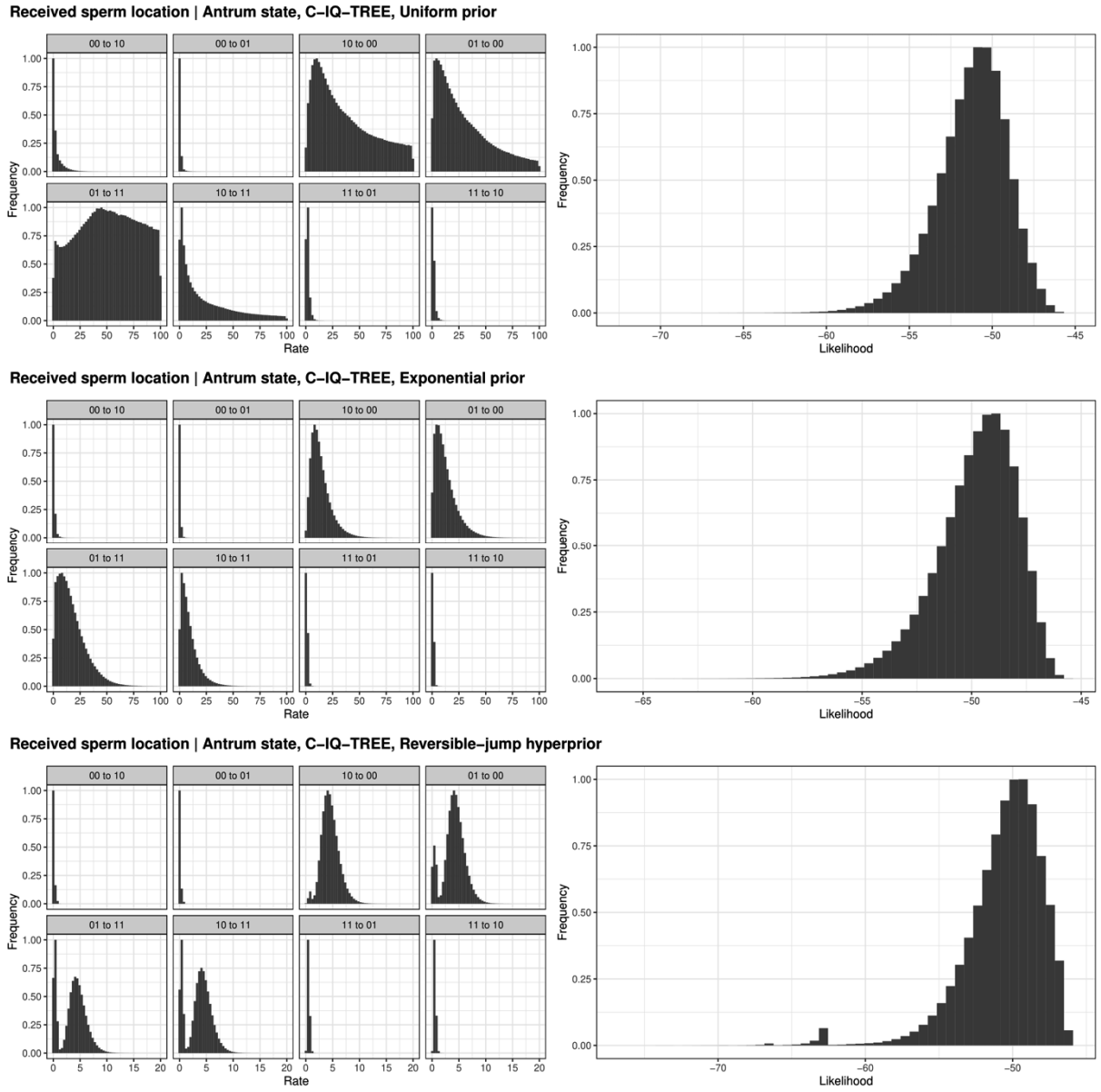


Figure SI9 Posterior distributions of the rate parameters (left) and likelihoods (right) for BayesTraits runs with the C-IQ-TREE phylogeny and received sperm location and antrum state. Layout identical to previous figure.

obtusa

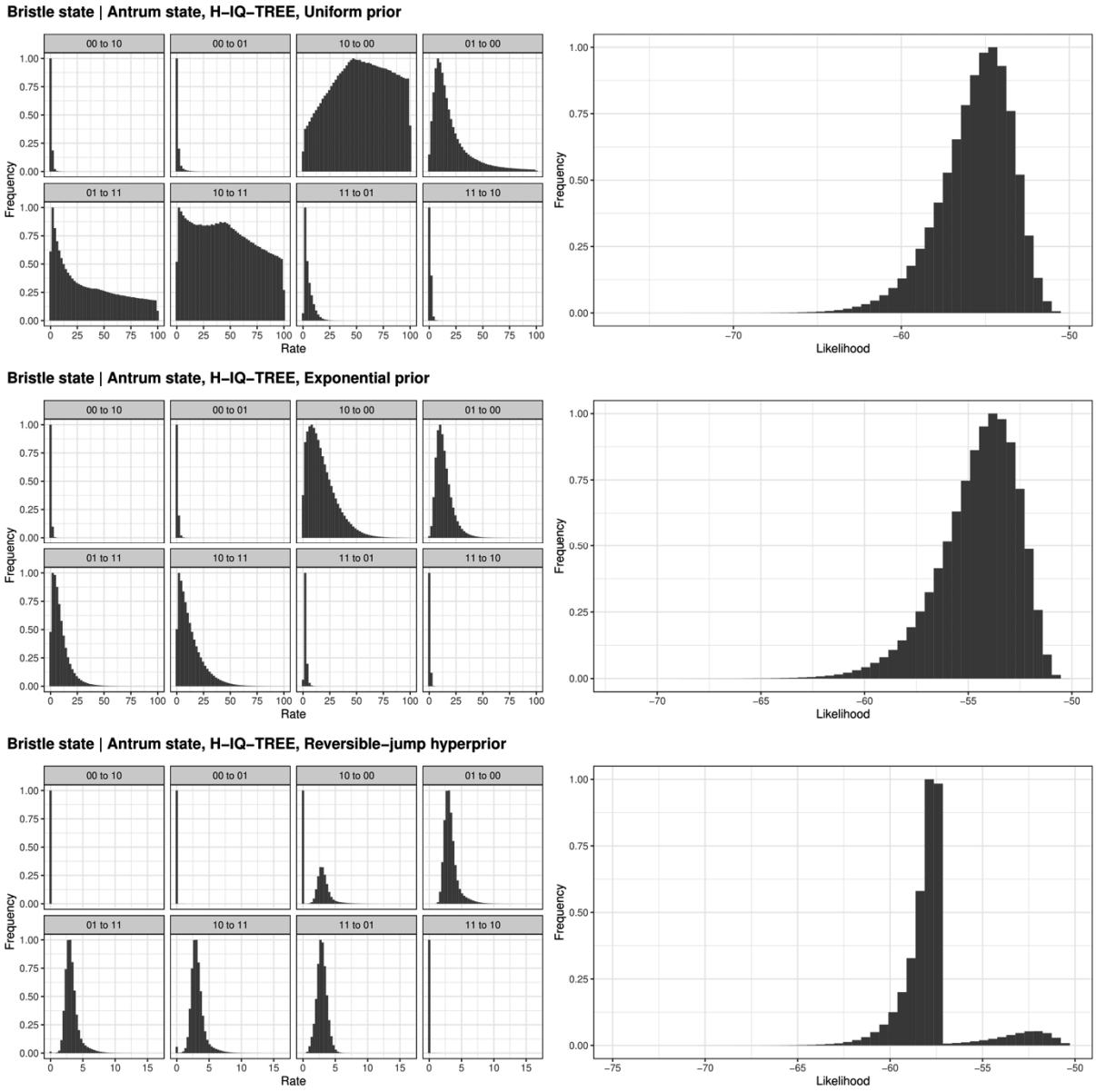


Figure SI10 Posterior distributions of the rate parameters (left) and likelihoods (right) for BayesTraits runs with the H-IQ-TREE phylogeny and bristle state and bristle state. Layout identical to previous figure.

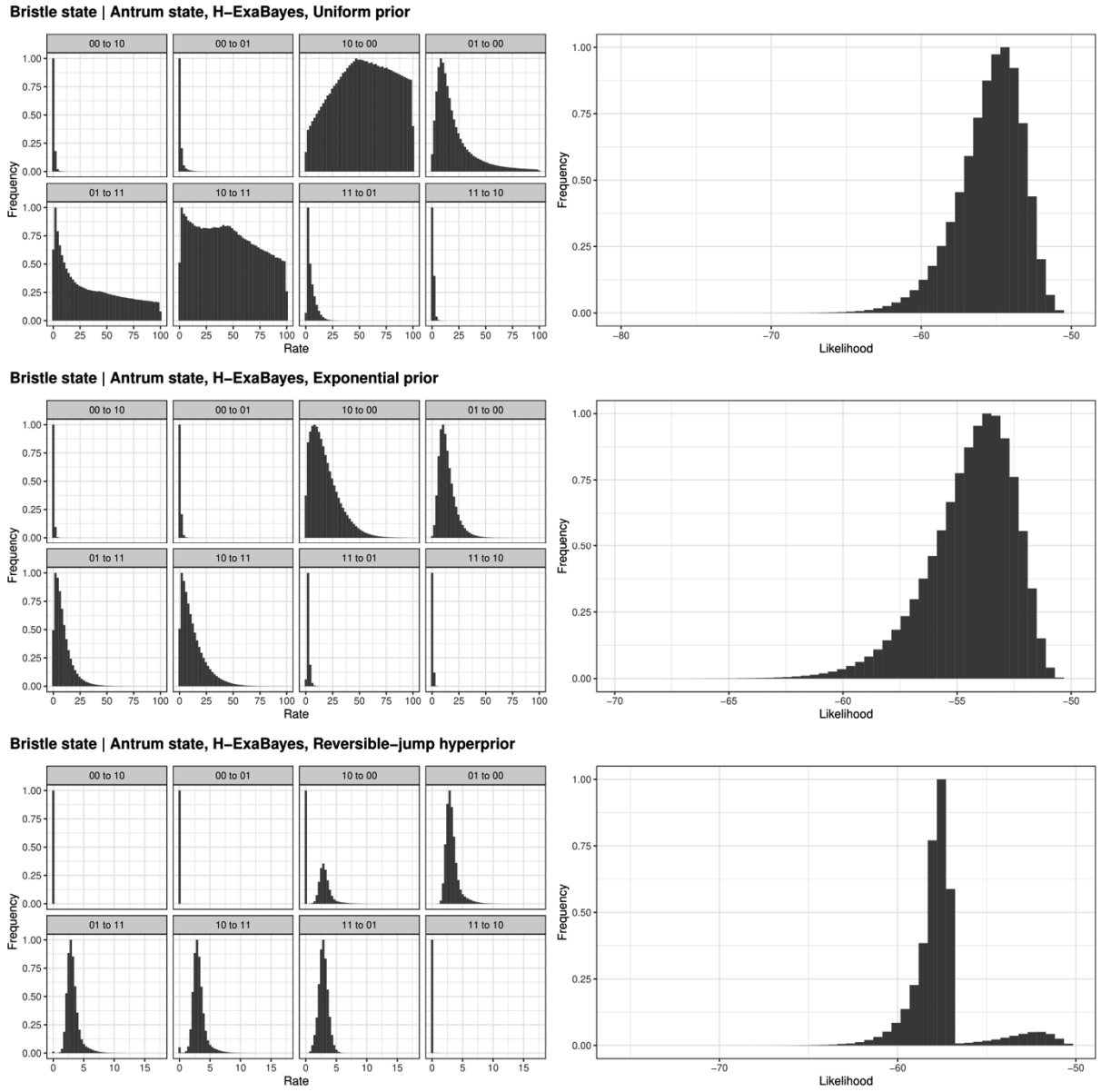


Figure SI11 Posterior distributions of the rate parameters (left) and likelihoods (right) for BayesTraits runs with the H-ExaBayes phylogeny and bristle state and bristle state. Layout identical to previous figure.

sp. 10

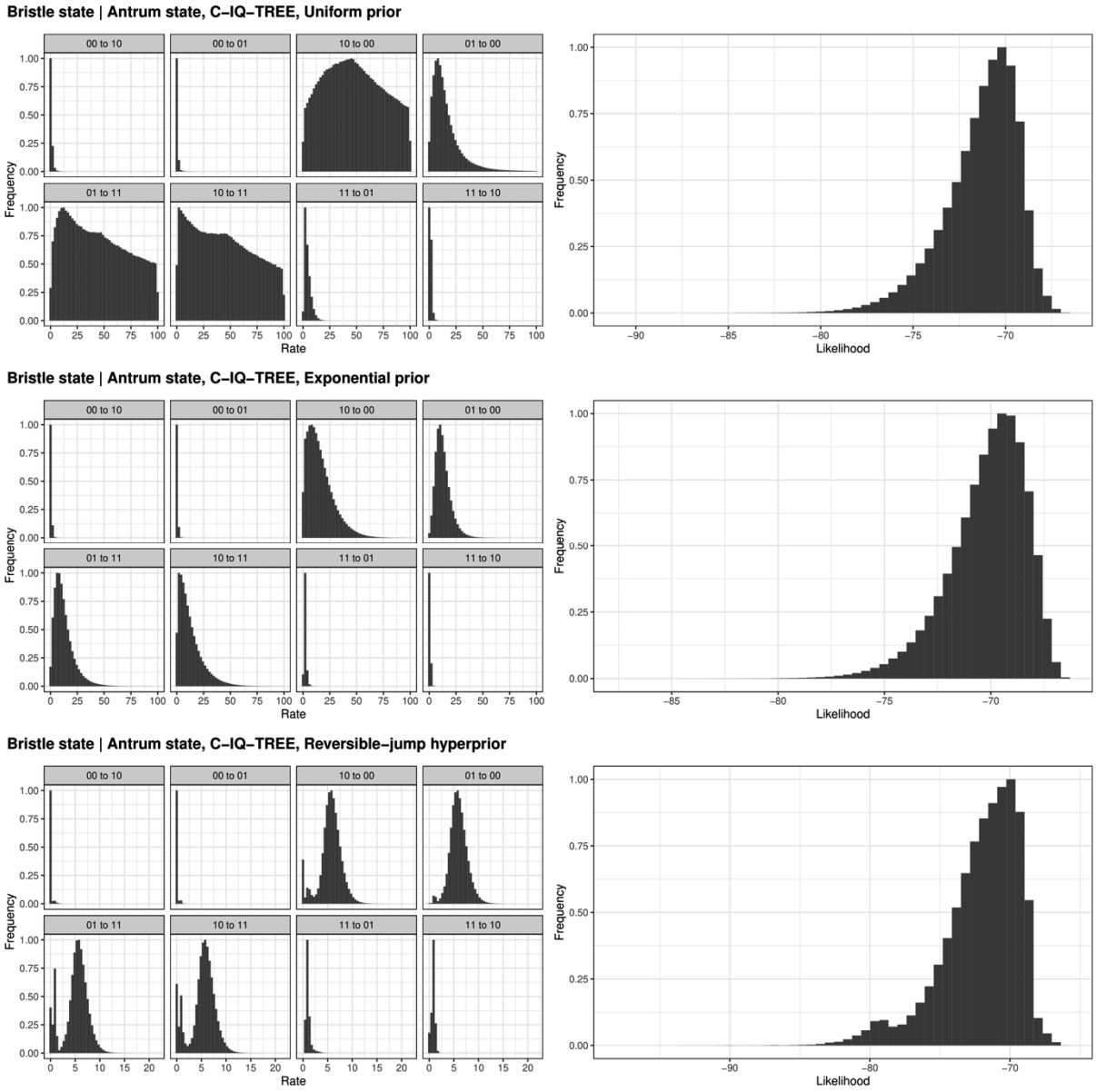


Figure SI12 Posterior distributions of the rate parameters (left) and likelihoods (right) for BayesTraits runs with the C-IQ-TREE phylogeny and bristle state and bristle state. Layout identical to previous figure.

Morphological data

Stylet morphology

We measured stylet length by placing two segmented lines along its sides (Figure SI13) and then taking the average of both. We also used this line to measure the curviness of the stylet by summing all the angles between the neighbouring line segments for each line and again taking the average. Due to this, longer stylets may tend to be curvier simply because they have more line segments and thus more angles that are added. However, we decided against standardizing by length since this would lead to autocorrelation in our principal component analysis and since such standardisation is easy to implement for future use. We also measured the width of the proximal and distal stylet openings (Figure SI13). Defining the exact position of the distal opening can be challenging, since it often is highly asymmetrical and the transition from stylet to the thickening can be hard to define (see Figure SI13 B-D for an illustration). We addressed this by defining the longer section of an obliquely ending stylet to be part of the distal thickening. To measure the distal opening, we thus placed a marker on the side of the stylet that terminates first and drew an orthogonal line to the last line segment to define the termination point of the second length line (Figure SI13). In cases where this orthogonal line did not intersect with the other side of the stylet, the point was instead placed at the last possible position. The distal thickenings were measured by connecting the sides of the distal opening to the furthest possible positions on the thickening, which is akin to increasing the line width of the length lines and placing the point on the structure that is obscured last. As a measure of distal asymmetry, we then took the absolute difference between the size of the two distal thickenings. Finally, we categorized the stylet according to how sharp its distal end is on a per species basis (Table SI1).

Measurements of stylets were performed on *in situ* images and videos of live worms (590 stylets) and from images of smash preparations (462 stylets). We assessed if the preparation method introduced a measurement bias by measuring 103 specimens across 59 species both *in situ* and in smash preparations. We found that *in situ* estimates were somewhat shorter (resulting in on average 3.1 µm shorter stylets (median 3.1% shorter), paired t-test: $t=5.8$, $df=102$, $p< 0.001$). Even though we introduced a slight bias by measuring under both setups, this allowed us to greatly increase sample size, since stylets can be challenging to measure *in situ* in certain orientations. In cases where both estimates were available, we used the mean between the two. In some of the specimens it was not possible to measure the length of the stylet, but we still measured the width of the proximal and distal openings (103 cases), conversely a few times we were able to measure the stylet length but not the width of the distal opening (11 cases), the width of the proximal opening (one case) or both (two cases). Finally, we could not spatially calibrate the images of 7 specimens, but still collected curviness information on these stylets.

Sperm morphology

We measured sperm length of 1765 sperm across 504 specimens by placing a segmented line along the midline, starting from the most distal part of the feeler and terminating just anterior of the brush (if present) (Figure SI14). Note that this measure of total sperm length is different from a previous study on *M. lignano* where the sperm feeler was excluded (Janicke and Schärer 2010), but equivalent to the definition used in (Schärer et al. 2020). This was necessary since the boundary between the feeler and the sperm body is less clear in many species compared to *M. lignano*. The feeler was excluded in (Janicke and Schärer 2010) since they determined that repeatability was lower when it was included and its inclusion in the current study potentially captures some interesting variation.

We measured sperm bristle length by placing a line along its axis. If possible, we measured both bristles of a sperm and then averaged them per sperm before taking the average across all sperm measured per specimen, and then finally averaging per species (1021 bristles measured across 294 specimens). We also categorised sperm on a per species basis according to their bristle type and whether they carried a velum or a brush (Table SI1). Among the species in which the sperm bristles are visible in the light microscope, these have a quite continuous distribution. However, it is likely that very small bristles do not serve a function, or at least are too small to anchor sperm in the antrum wall during the suck behaviour. In some species the sperm bristles are so small that they do not even protrude outside of the sperm (e.g. *M.* sp. 50). We consider these bristles as reduced and categorised them as such to highlight our functional hypothesis. Since the trait is continuous, a clear cut-off point is hard to define and we decided that *M. poznaniense* is the species with the smallest sperm bristles (3.2 um) that we do not consider reduced (Figure SI15).

Female antrum morphology

The morphology of the female antrum was also categorised by examining all available material from a species and scoring the species as a whole. We examined the thickness of the overall antrum wall, the presence of a visible anterior thickening of the female antrum, the so-called cellular valve, the complexity of the chambers, and we counted the number of female genital pores (Table SI2). For the analysis of correlated evolution and PGLS on sperm length we scored antrum thickness as binary and refer to it as antrum state. When the antrum wall is thickened throughout, it is more difficult to see if there is a cellular valve and potentially there is some conflict between the two scorings. However, this would occur only in specimens that already achieve a high score on antrum complexity and as such might just introduce a slight downward bias for high values. Finally, we summed the thickness, cellular valve, and chamber scores to get one summary antrum complexity score (Table SI2).

poznaniense

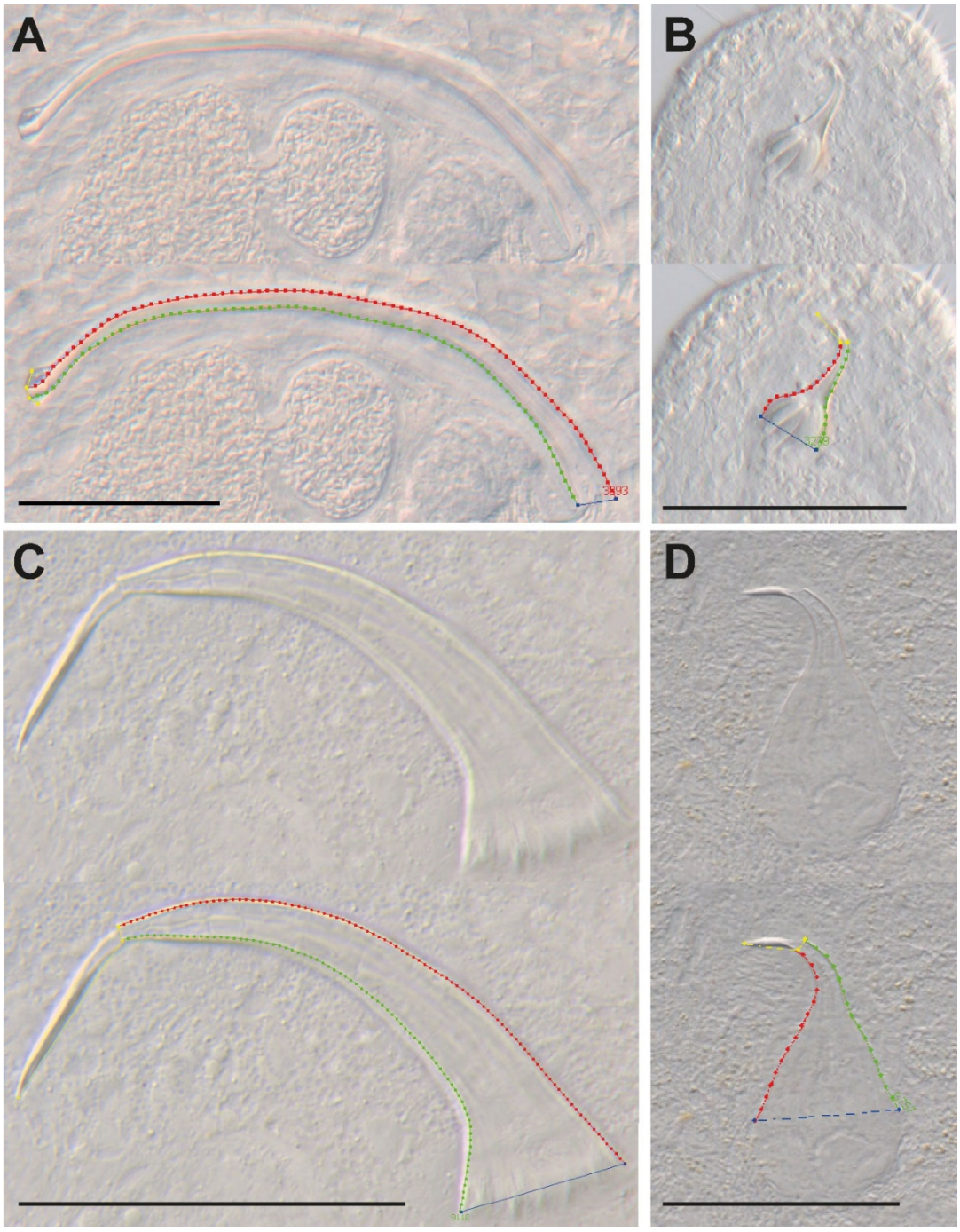


Figure SI13 Details on stylet morphometrics. In each panel, the top shows the raw image and the bottom the measured lines. The segmented red and green stylet length lines are used to determine stylet length and curviness, and the blue line measures the width of the proximal opening. Each yellow segmented line consists of four points (a, b, c, and d) with a to b and c to d measuring the size of the 'red' and 'green' distal thickenings, respectively, and with b to c measuring the width of the distal opening. To determine points b and c, the stylet length line that terminates first (green in panel D) is selected and a line orthogonal to its last line segment drawn through its endpoint (now considered point c). Where this orthogonal line intersects with the longer stylet length line (red in panel D) is now considered point b. This intersection then also defines the endpoint of the longer stylet length line (red in D). The size of the distal thickenings was then measured as the distance from the endpoint of the orthogonal line (e.g. relatively symmetrical distal thickening in A and asymmetric thickening

in B-D). A in situ stylet of *M.* sp. 67 (specimen MTP LS 2613). B in situ stylet of *M.* sp. 48 (specimen MTP LS 2128). C smashed stylet of *M.* sp. 37 (specimen MTP LS 2082). D smashed stylet of *M. hystricinum* (specimen MTP LS 2133). Scalebars in A, C and D are 50 µm and 25 µm in B.

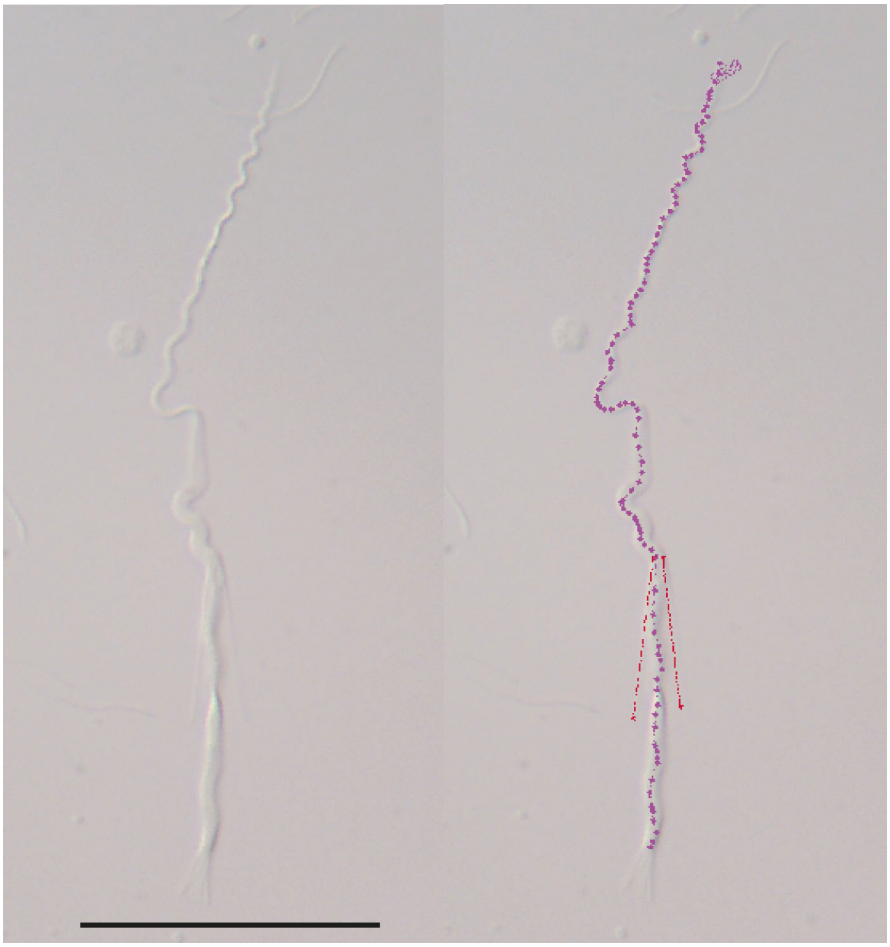


Figure SI14 Details on sperm morphometrics. Sperm image of *M. karlingi* (specimen MTP LS 3357) on the left, with an example of the segmented sperm length line (purple) and the lines used to measure sperm bristle length (red). Scalebar is 25 µm.

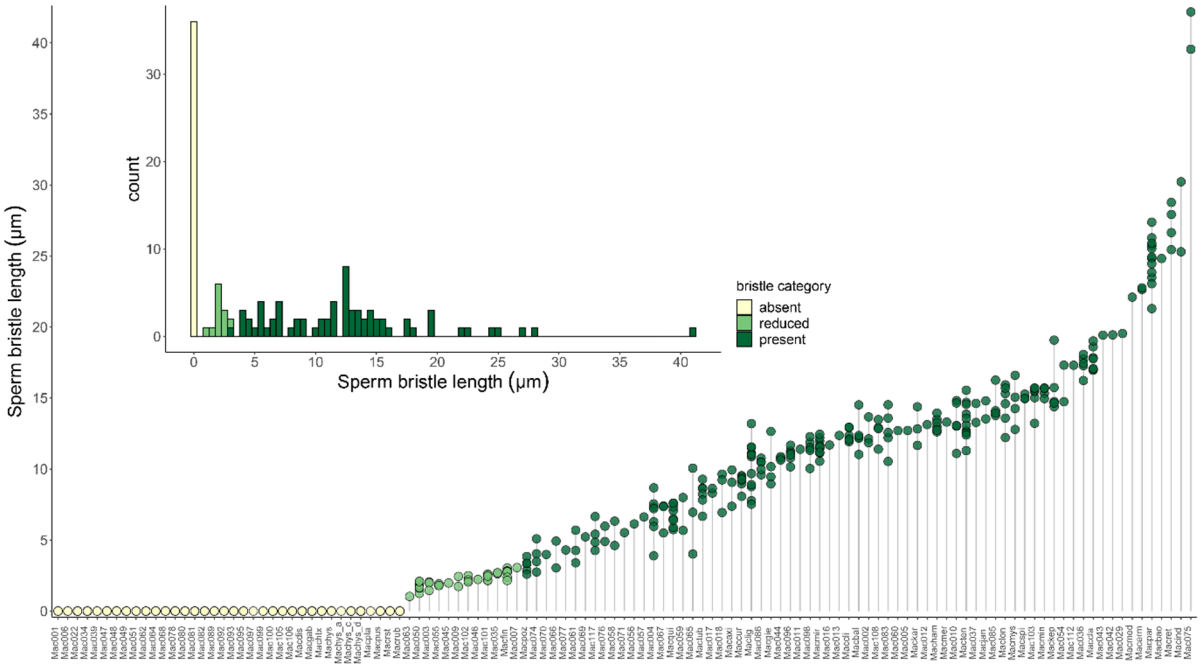


Figure SI15 Distribution of sperm bristle length across 117 *Macrostomum* species. The large plot show the individual data points of the species ordered by mean bristles length and the inset a histogram of mean bristles length. The colours indicate sperm bristle state.

sp. 16

Table SI1 Categorisation of male reproductive morphology and the received sperm location. Given are the state definitions and how they were translated into ordinal coding.

Stylet sharpness		**Bristle state**			**Sperm brush**		**Sperm velum**		**Received sperm location**	
State	Coding	State	Trinary coding	Binary coding	State	Coding	State	Coding	State	Binary coding
Rounded/blunt distal thickening	0	**Absent**	0	0	**Absent**	0	**Absent**	0	**Hypodermic only**	0
No clear blunt or sharp distal thickening	1	**Reduced** (<3.2 um)	1	0	**Present**	1	**Present**	1	**Hypodermic and antrum**	0
Sharp distal thickening (terminating in edge or point)	2	**Present** (>3.2 um)	2	1					**Antrum only**	1

Table SI2 Categorisation female reproductive morphology. Given are the state definitions and how they were translated into ordinal coding.

Antrum thickness		**Cellular valve**		**Antrum state binary**		**Antrum chamber**		**Antrum complexity**	**Genital pores**	
State	Coding	State	Coding	State	Coding	State	Coding	Coding	State	Coding
Antrum wall **very thin**, only visible when egg is present	0	**No clear** anterior thickening	0	Thickness <2 AND cellular valve = 0	0	**One simple chamber**	0	∑ (antrum thickness, cellular valve, antrum chamber)	One pore	1
Thin wall, but structure also visible without egg present	1	**Clear** anterior thickening	1	Thickness >1 OR cellular valve > 0	1	**Elaborated chamber and/or vagina**	1		Two pores	2
Clear thickening	2	Thickening **extends far anterior**	2			**Multiple chambers**	2			
Strong thickening	3									

Haplotype networks

We constructed haplotype networks for all 668 partial *28S rRNA* sequences included in this study. We inferred haplotype networks for subgroups (indicated by numbers in Figure SI16) and visualised them including the morphology of the stylet (and if diagnostic the sperm) for each cluster. As described in the methods we delimit species with as distance in the network >3 mutational steps. Since in some cases networks become quite large, we severed all connection larger than 12 steps and simply display the separated clusters. In all the displays (Figure SI5 – S20) differences between the sequences are represented by the edges in the graph and the scale bar represents 20 micrometres.

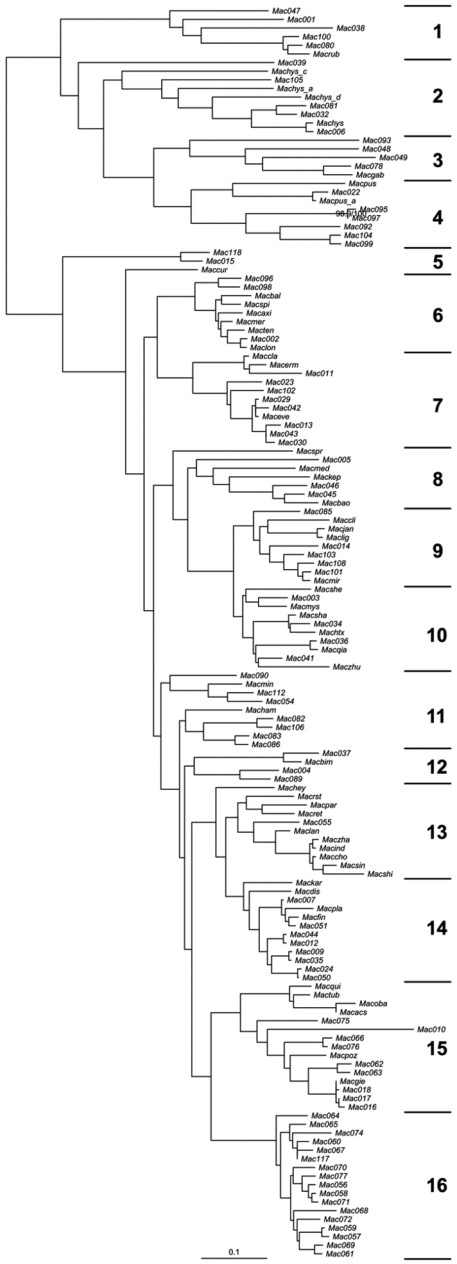


Figure SI16 C-IQ-TREE phylogeny indicating the subgroups haplotype networks were constructed for. Scale bar represents substitutions per site and large numbers denote the subgroups (Figure SI5-SI20).

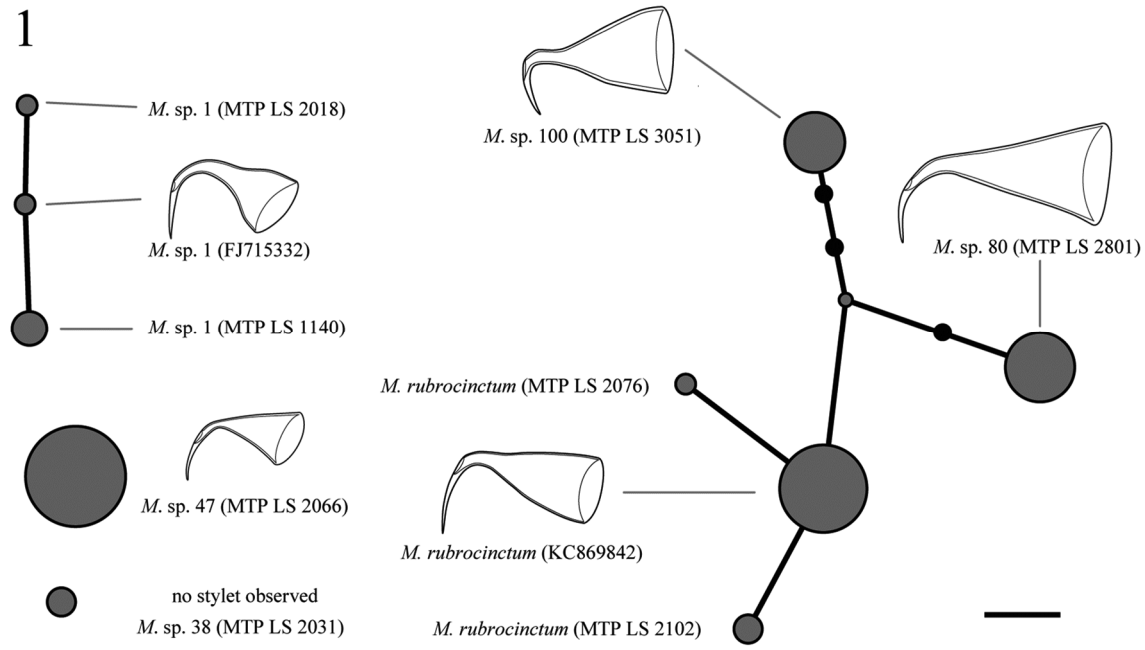


Figure SI17 TCS haplotype network of group 1. While molecularly distinct we did not observe a stylet for *M.* sp. 38.

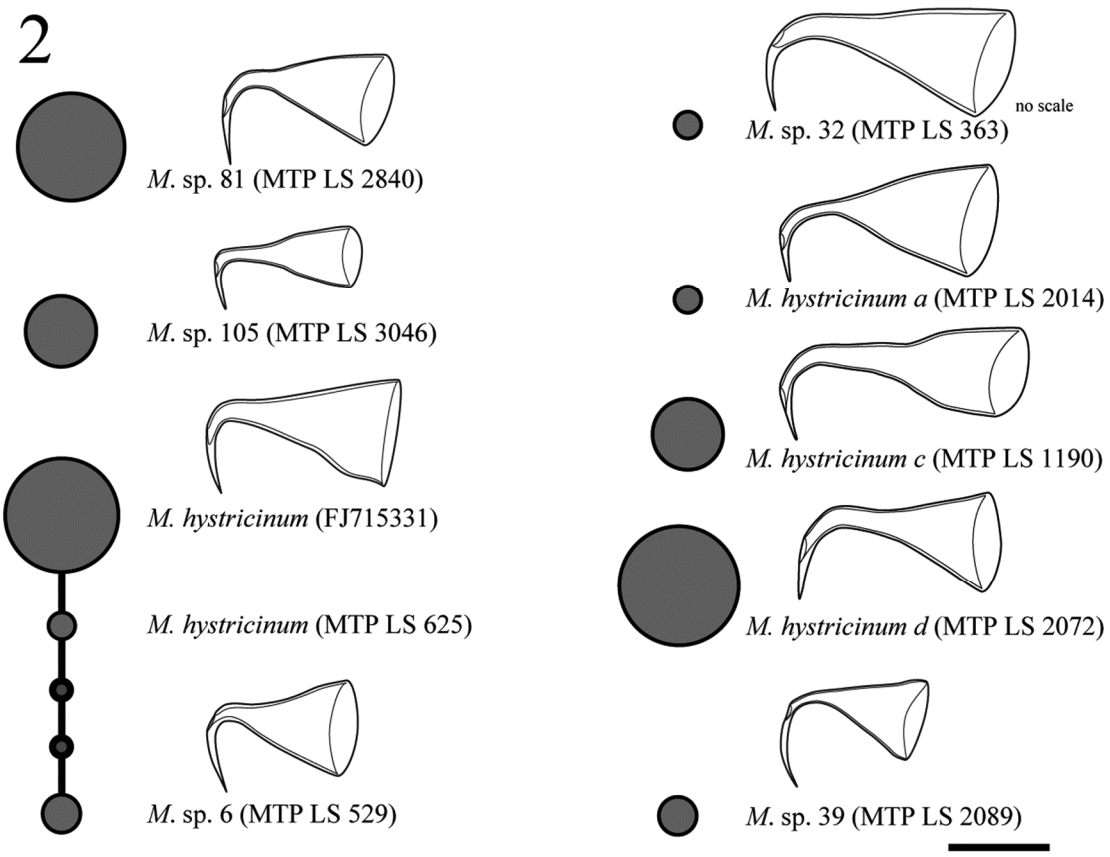


Figure SI18 TCS haplotype network of group 2.

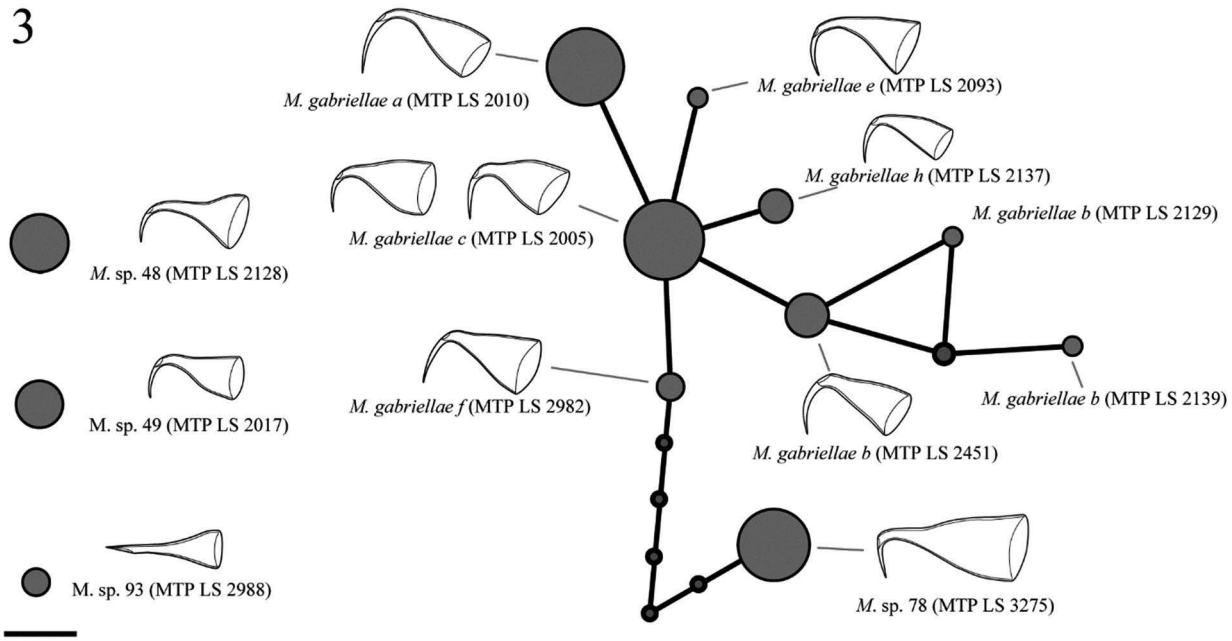


Figure SI19 TCS haplotype network of group 3. All species denoted with *M. gabriellae* have been grouped into one species. While some differences in the stylets are apparent in these drawings the variation within each cluster is substantial and overlapping with the other morphologies and hence, we have not been able to consistently assign clusters based on morphology.

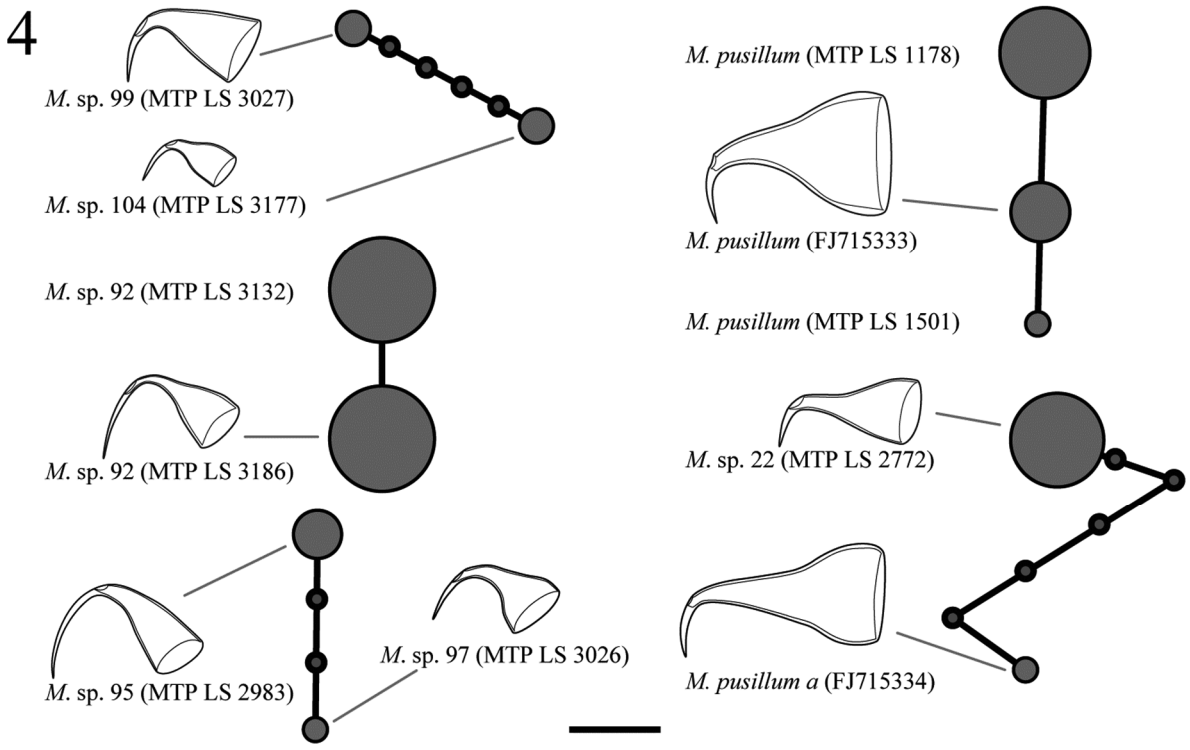


Figure SI20 TCS haplotype network of group 4. Note that *M. pusillum a* which is so named because it has a close resemblance with *M. pusillum* clusters more closely with *M.* sp. 22.

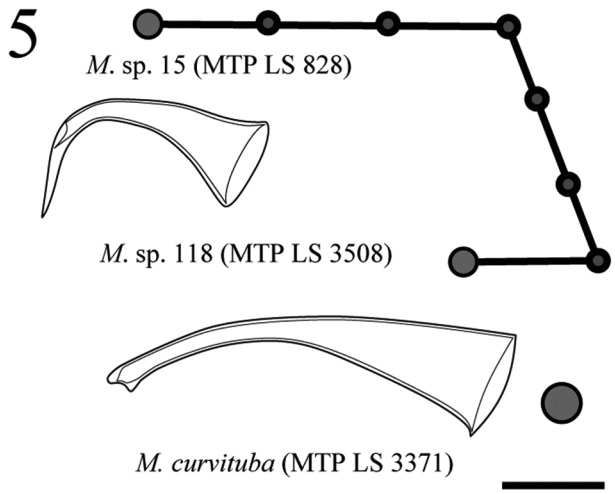


Figure SI21 TCS haplotype network of group 5. *M.* sp. 118 was sexually mature when sampled and we thus have no morphology of the stylet. We do not have microscope scale information for *M.* sp. 15 and thus no drawing.

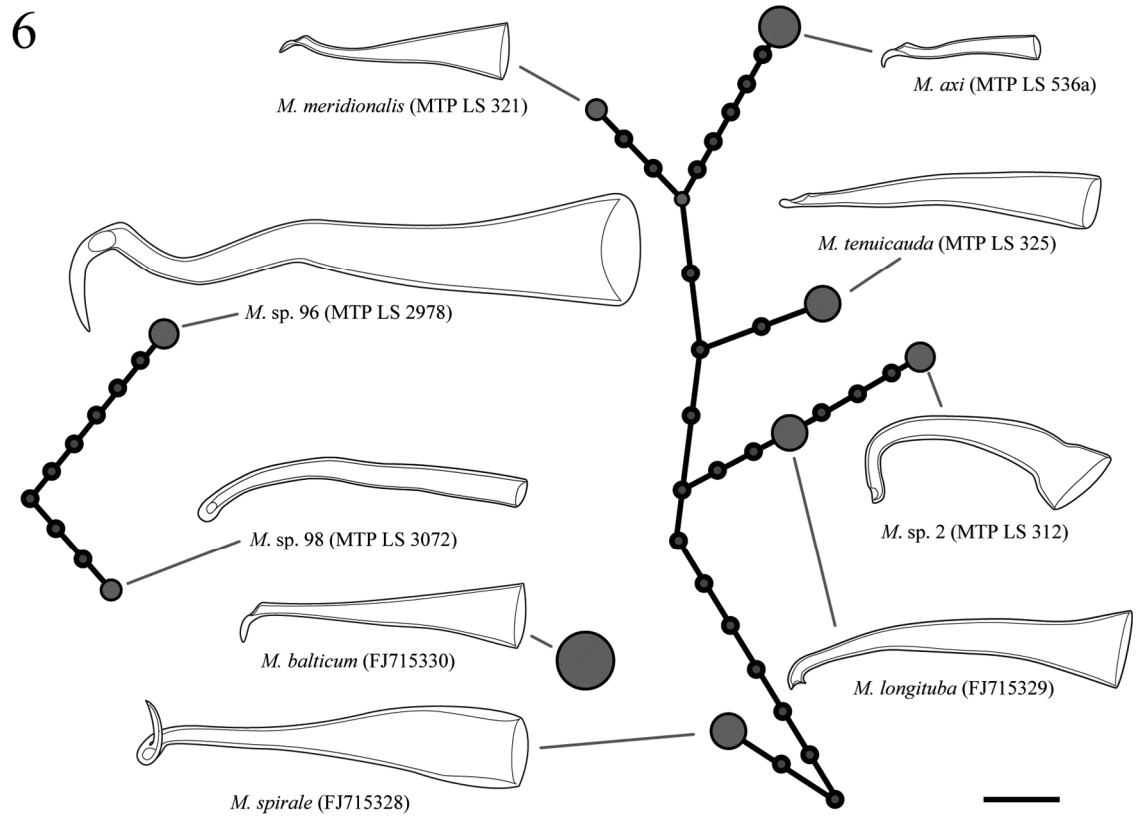


Figure SI22 TCS haplotype network of group 6.

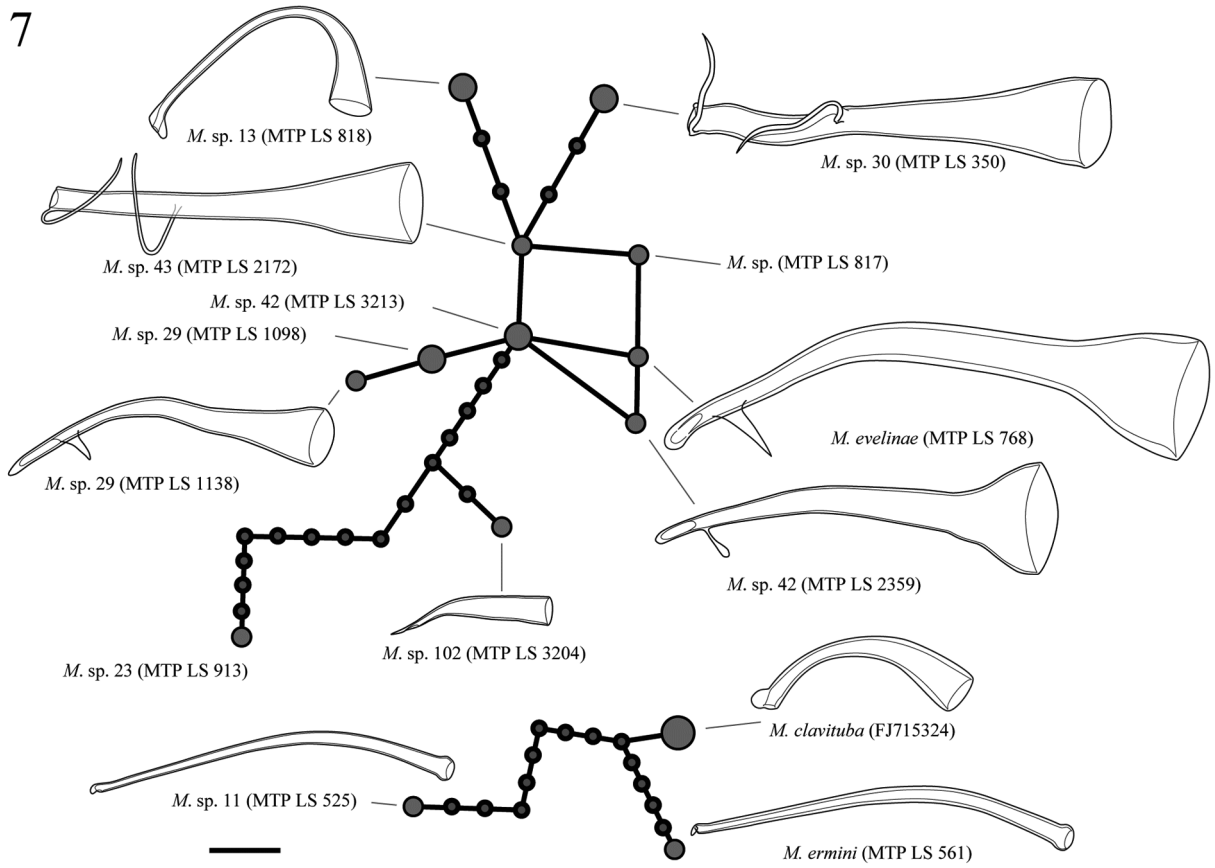


Figure SI23 TCS haplotype network of group 7. *M.* sp. (MTP LS 817) was not sexually mature and we thus have no morphological information on the stylet. Since it clusters closely with some of the other species, we have not assigned it a provisional species number. *M.* sp. 29, *M.* sp. 42, *M.* sp. 43 and *M. evelinae* are all connected with <3 differences but clearly distinct morphologically *M.* sp. 42 is distinct due to its funnel shaped proximal opening and a rounded lateral protrusion just proximal from the opening. While *M. evelinae* is superficially similar the stylet is larger, and the lateral protrusion is sharp. *M.* sp. 29 also has a sharp protrusion but it is located further proximal, the distal opening is larger, and the stylet has an additional curve. *M.* sp. 43 bares little resemblance to the other three with a straight stylet that has two thin lateral protrusions.

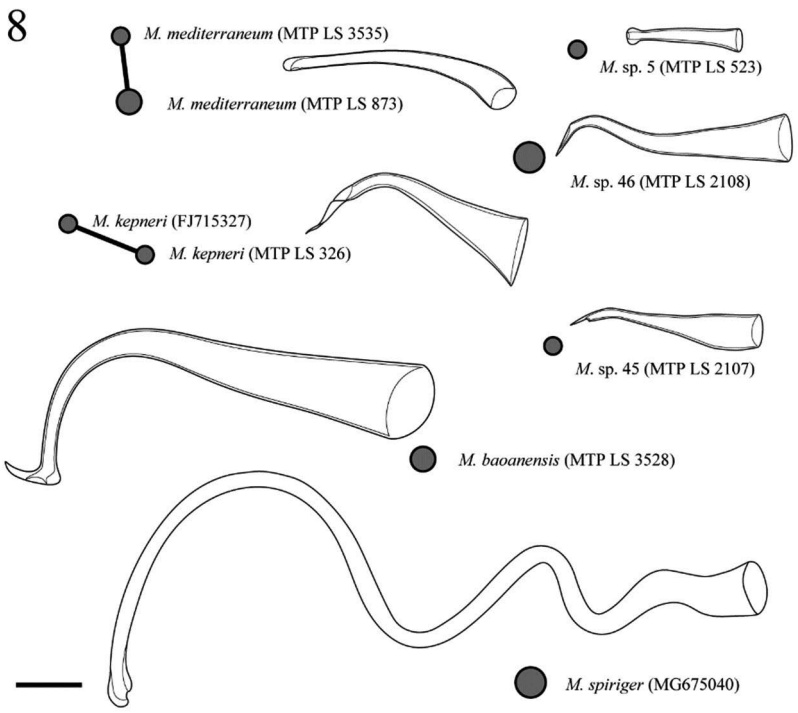


Figure SI24 TCS haplotype network of group 8.

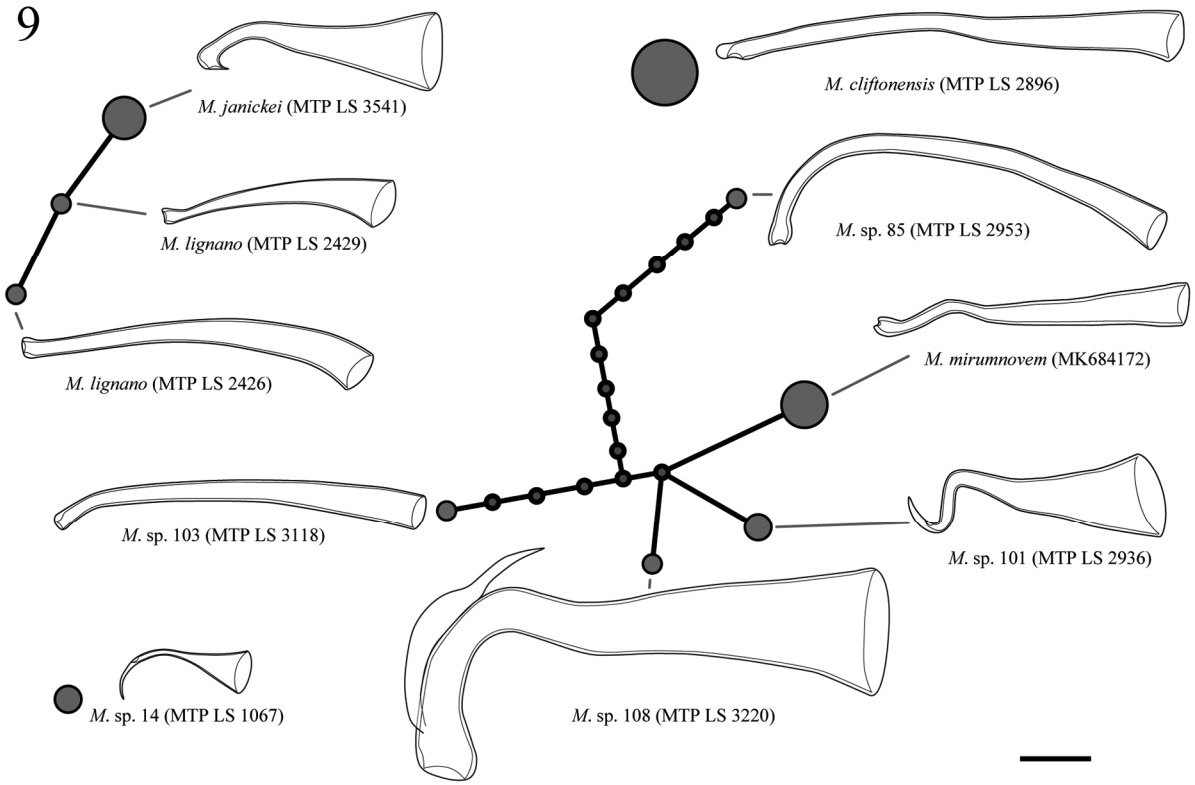


Figure SI25 TCS haplotype network of group 9. MTP LS 244 represents the M. lignano population from Italy while MTP LS 2426 represents a population from Greece. The Greek population has a larger stylet. The shape variation within the Italian population can include morphologies like the one drawn for the Greek population. Note that our 28S sequence of *M. mirumnovem* is identical to GenBank accession KC869843. *M. mirumnovem*, *M.* sp. 101, and

M. sp. 108 are clearly distinct with much larger size in *M.* sp. 108 and a sharp twisted distal region in *M.* sp. 101.

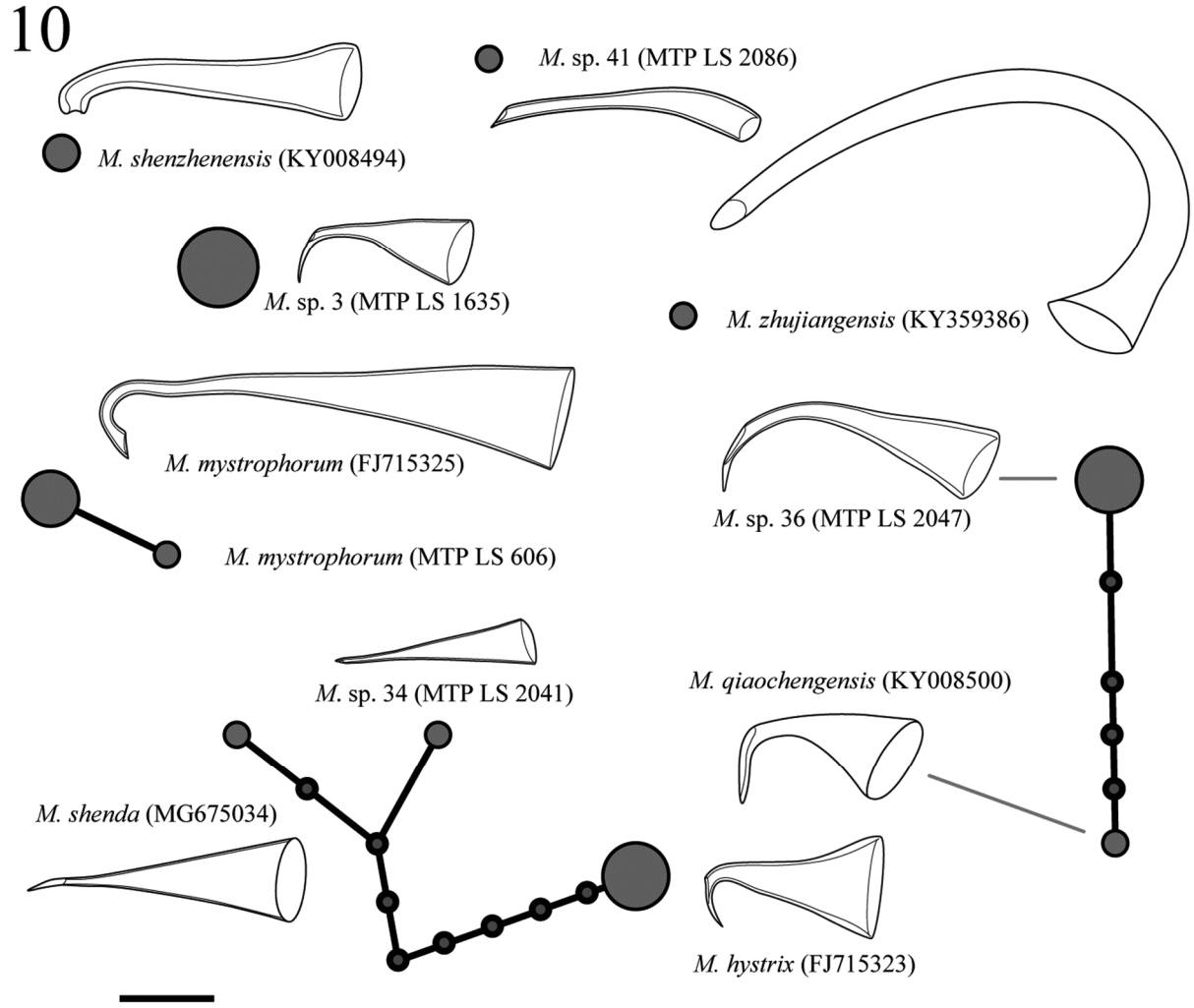


Figure SI26 TCS haplotype network of group 10. *M. sp. 34* and *M. shenda* are similar in stylet morphology and general habitus, however, *M. shenda* stylet is larger.

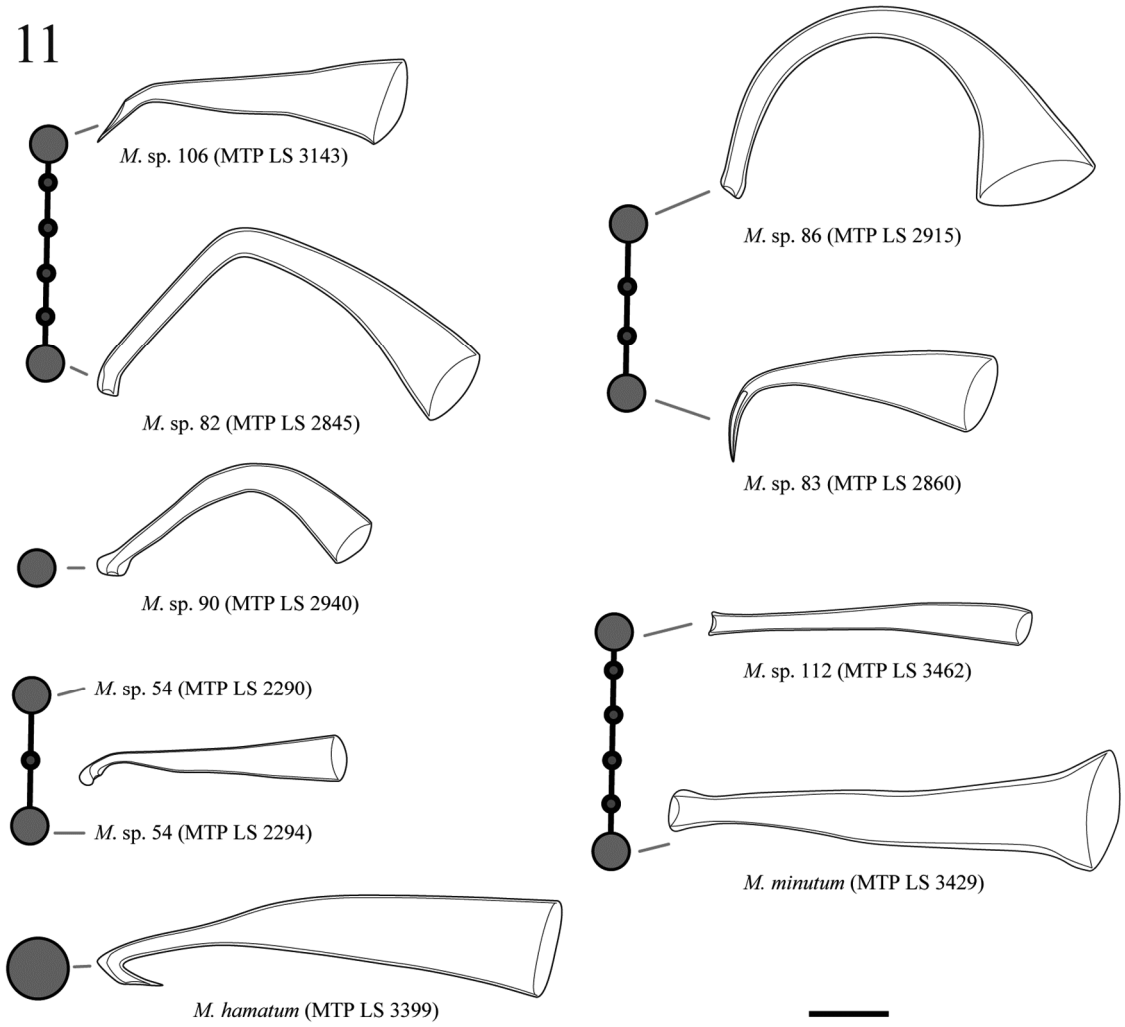


Figure SI27 TCS haplotype network of group 11. Stylet of *M. minutum* and *M.* sp. 112 differ int that the former is longer, wider and has a clear distal blunt thickening, while the latter does not have this thickening and does not have a funnel shaped proximal opening.

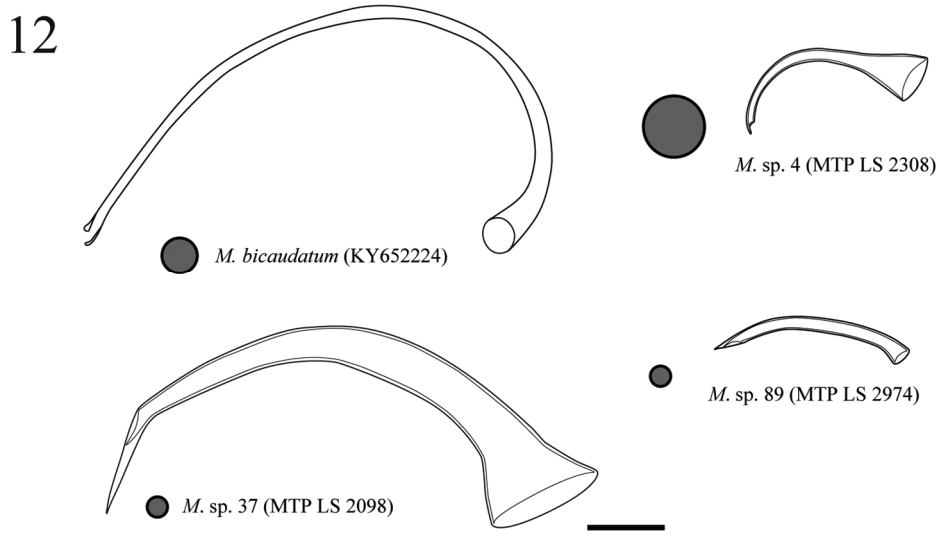


Figure SI13 TCS haplotype network of group 12.

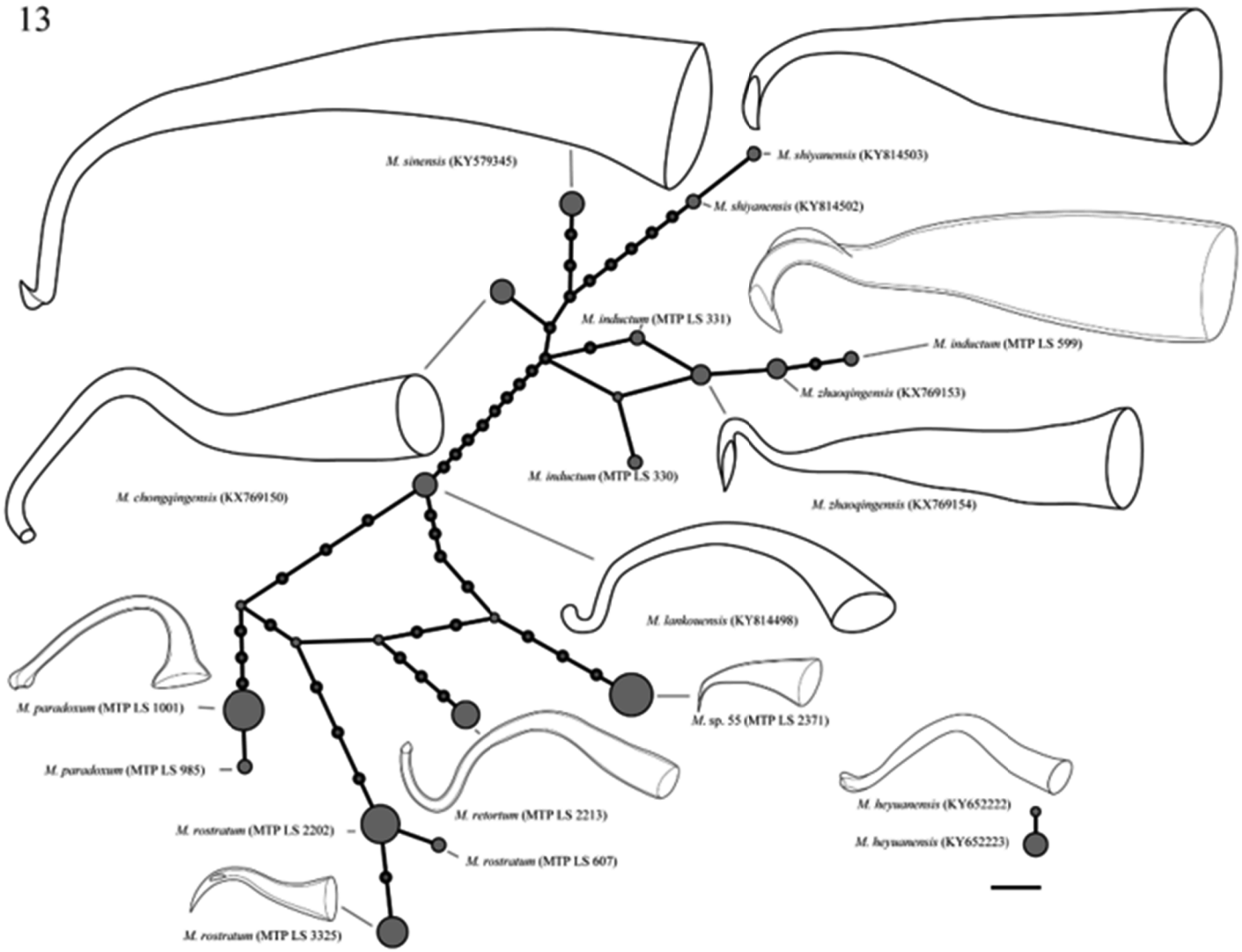


Figure SI29 TCS haplotype network of group 13. Our *M. inductum* samples cluster with *M. zhaoqingensis* and they are also very similar in stylet and sperm morphology, the latter could thus be a re-description of the former and should be further investigated taxonomically.

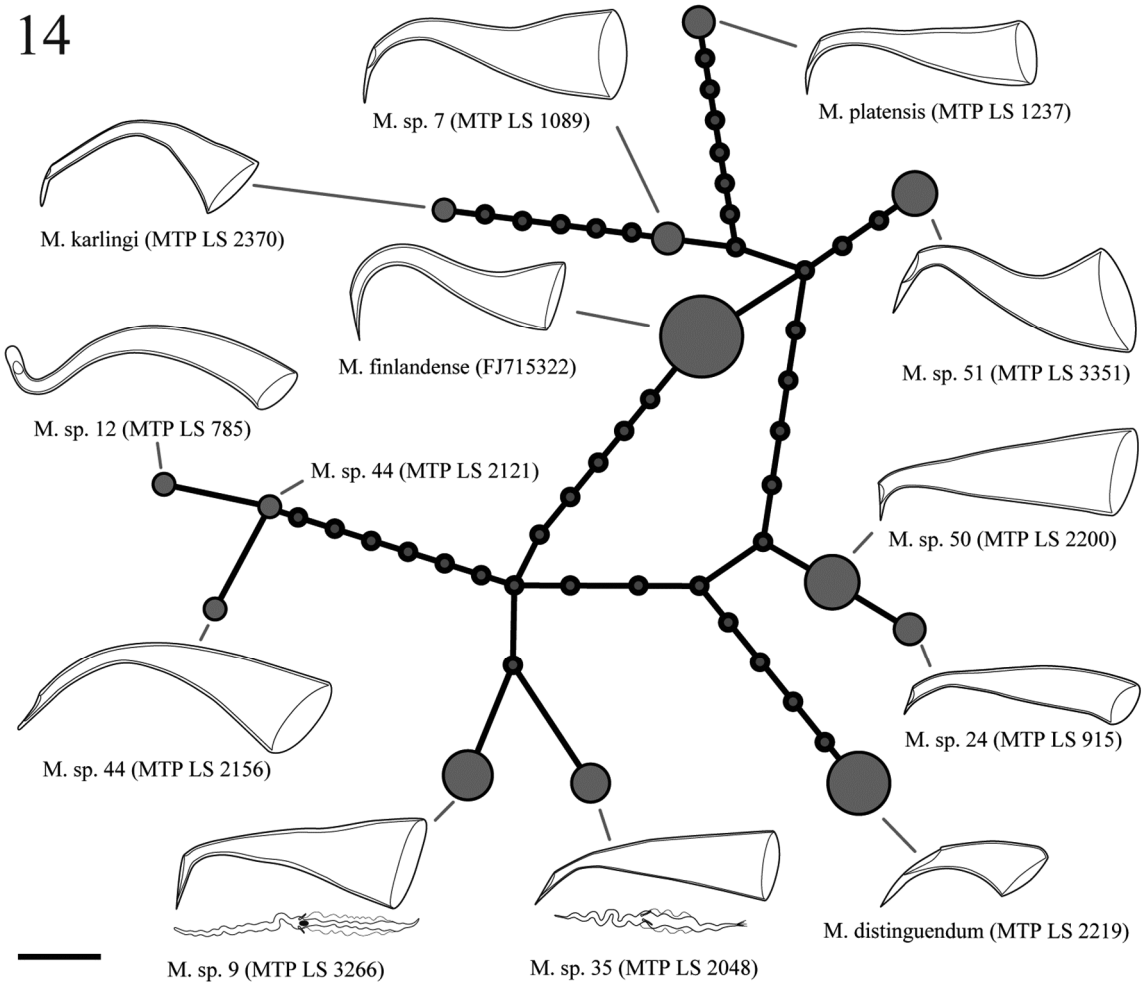


Figure SI29 TCS haplotype network of group 14. Stylets are similar in shape between *M.* sp. 9 and *M.* sp. 35 but with a longer drawn out distal tip and a broader base in *M.* sp. 9. The sperm between the two species is diagnostic with a clearly visible dense body close to the bristles in *M.* sp. 9. This structure, which could be the nucleus, is not visible in *M.* sp. 35. and the sperm of *M.* sp. 35 has a brush. The stylets of *M.* sp. 12 and *M.* sp. 44 differ in the width of the base, being much wider in the latter and ending in an oblique cut while there is an additional curve in the former ending in a blunted subterminal opening. The stylets of *M.* sp. 24 and *M.* sp. 50 differ in the width of the base, being much wider in the latter with a more pronounced 90° turn of the distal thickening.

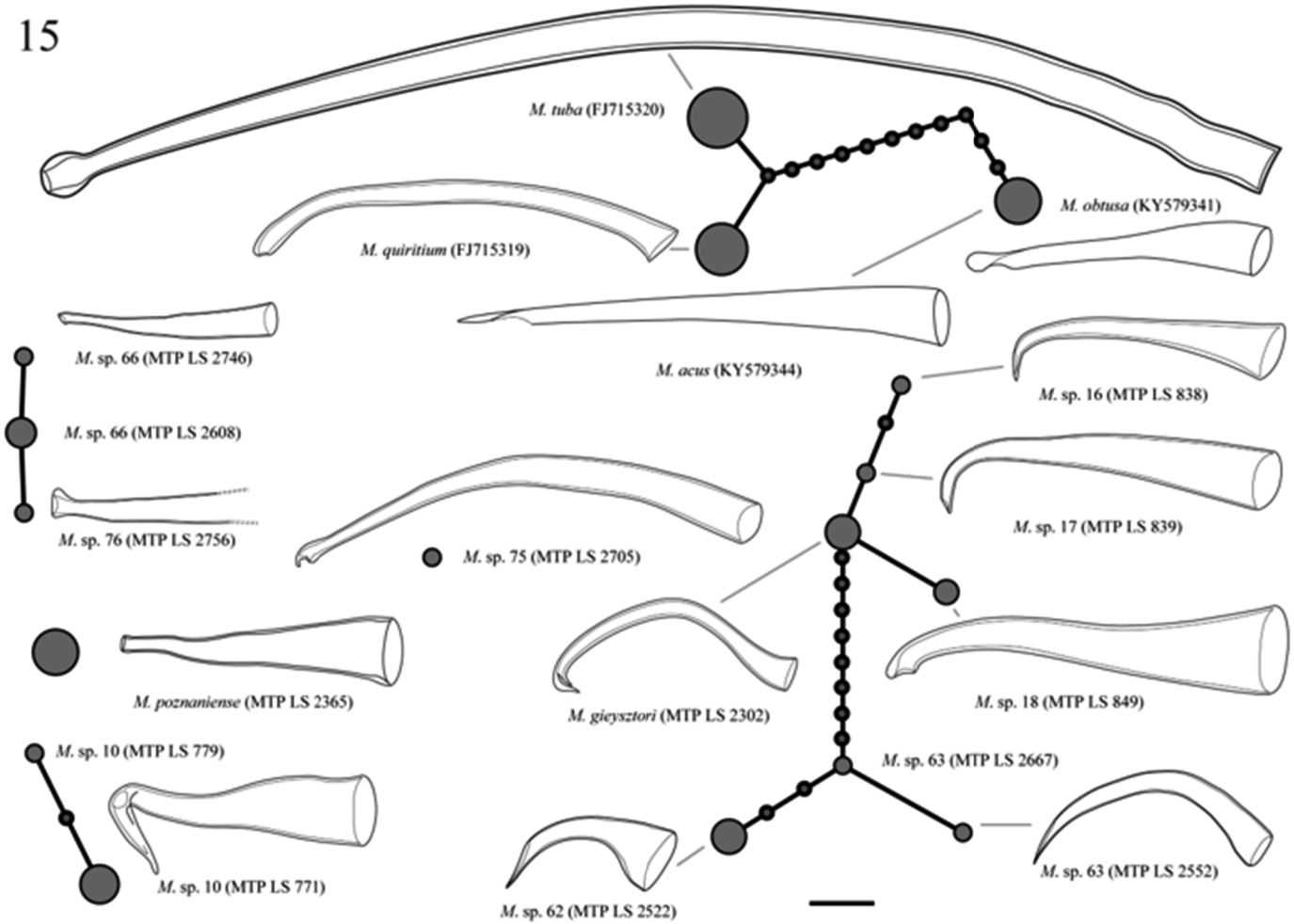


Figure SI31 TCS haplotype network of group 15. The stylet of *M.* sp. 16 is shorter than *M.* sp. 17 and the distal curve of the stylet is sharper in the former ending in a strongly oblique tip, while the curve is gentler in the latter and ends much less oblique. *M.* sp. 18 differs from the former two since it has a longer wider stylet that terminates subterminally and has a slight thickening on the convex part of the distal curve. All three differ from *M. gieysztori* which has a thinner stylet with a pronounced additional curve and a subterminal opening pointing proximally and a clear anterior thickening.

References Supporting Information

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

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

sp. 17

Chapter IV

The evolution of sex allocation and its correlates in a genus of simultaneous hermaphrodites

Manuscript in preparation as: The evolution of sex allocation and its correlates in a genus of simultaneous hermaphrodites. Jeremias N. Brand, Luke J. Harmon and Lukas Schärer

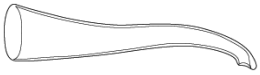
Abstract

Sex allocation theory is one of the success stories of evolutionary biology, accurately predicting the investment into reproduction via eggs or sperm depending on a species' ecology and reproductive biology. Sex allocation has focused extensively on separate-sexed organism but has also been scrutinised in simultaneously hermaphroditic animals. This research has revealed that many simultaneous hermaphrodites show plasticity in sex allocation in response to the intensity of sperm competition they experience in their social group. In contrast, few studies have investigated whether these patterns are also reflected at a macroevolutionary scale. We here present a comparative study of the evolution of sex allocation across 120 species in the free-living flatworm genus *Macrostomum*, revealing the tempo and mode of the evolution of sex allocation. And by combining these sex allocation estimates with a rich dataset on sexual traits we can identify important morphological and genetic correlates of sex allocation evolution. We find that sex allocation in *Macrostomum* shows a bimodal distribution, with the peaks corresponding to macroevolutionary optima of relatively male or relatively female-biased allocation that species evolve towards. We also test some of the fundamental assumptions of sex allocation theory, by showing a negative correlation between the investment into the two sexes and a positive correlation between different aspects of male allocation. Finally, we show that hypodermic insemination, a type of traumatic mating that has convergently evolved at least nine times within the *Macrostomum* genus, is associated with a more female-biased sex allocation, likely because it is accompanied by an increased propensity for selfing in hypodermic species.

Introduction

Sex allocation theory predicts how organisms allocate resources to reproduction via the male or female function and it has been argued to be one of the most successful branches of evolutionary biology, due to its ability to accurately predict sex allocation patterns of many organisms (West 2009). Sex allocation has been primarily studied in gonochorists (separate-sexed organisms), where it is relatively easily measured as the offspring sex ratio (Charnov 1982; West 2009). In gonochorists, sex allocation is mostly determined by the intensity of local mate competition (Hamilton 1967), in which sons compete for the access to females and thus, from the perspective of the mother, waste reproductive resources that could be used for the production of daughters. Local mate competition occurs in structured populations where only a few females lay eggs in a patch (e.g. fig wasps Frank 1985; West and Herre 1998), favouring mothers that produce a female-biased sex ratio, since this will reduce competition between, and provide more potential mating partners for, their sons (Hamilton 1967; Taylor 1981). In hermaphrodites, sex allocation concerns the allocation of resources of a single individual into its male or female function (e.g. the production of sperm or eggs) either separated temporally in sequential hermaphrodites or at the same time in simultaneous hermaphrodites (SH).

SH is expected to be stable within a population when at least one of the sex function shows a saturating fitness gain curve (Charnov et al. 1976; Charnov 1980; Fischer 1981), meaning that investments into reproduction via that function initially results in high fitness returns, but then starts to yield lower returns (Figure 1), eventually favouring a reallocation of resources to the other sex function. In most cases, it is assumed that it is the male function that shows saturating returns (Charnov 1979, Figure 1). The main reason thought to lead to a saturating male fitness gain curve is low levels of sperm competition, leading to competition between related sperm in what has been termed local sperm competition in analogy to local mate competition in gonochorists (Schärer 2009). High local sperm competition is, for example, expected in organisms with a high selfing rate, where sex allocation is consequently expected to be female-biased (Charlesworth and Charlesworth 1981; Charnov 1982). Low population density will also lead to local sperm competition, since the reduced encounter probability will decrease the average number of donors contributing their ejaculates to a recipient (Charnov et al. 1981; Schärer 2009). But even when densities are high, paternity skews can decrease sperm competition through sperm removal, sperm displacement or cryptic female choice (Charnov 1996; Pen and Weissing 1999; Schärer 2009; van Velzen et al. 2009; Schärer and Pen 2013).



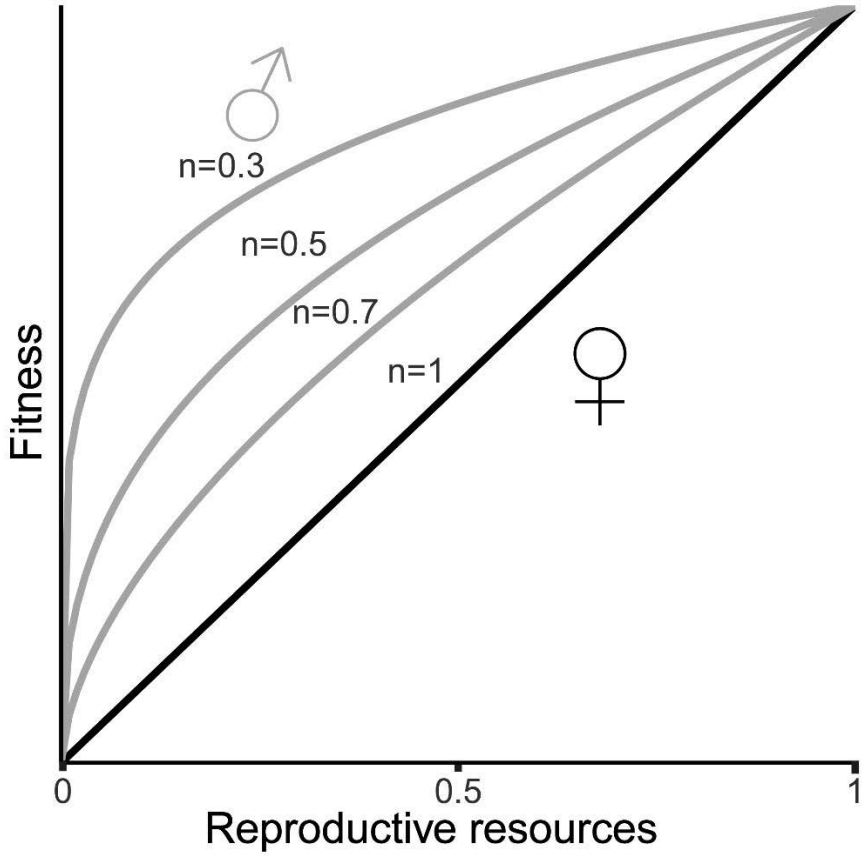


Figure 1 Theoretical fitness gain curves for investment into the male (grey) or female (black) function. The diminishing curves for the male investment (resources^{n}) represent reduced fitness returns per unit investmented in small mating groups due to local sperm competition. As mating groups size increases the male gain curve becomes more linear (and n-> approaches 1). The female curve is generally assumed to show linear returns.

Sex allocation in SH has been studied extensively in plants and it has been shown that many species plastically adjust their sex allocation in relation to body size or ecological conditions, in general in accordance with predictions from theory (reviewed in Goldman and Willson 1986; de Jong and Klinkhamer 2005). Further, for plants there is data on the exact relationship between investment into a function and the resulting fitness returns (Rademaker and de Jong 1998; Rademaker and Jong 1999; Campbell 2000; Elle and Meagher 2000). Similarly, there is a growing body of literature on sex allocation in SH animals that shows plastictiy in response to changes in factors causing local sperm competition (reviewed in Schärer 2009) and studies attempting to estimate the shape of the male fitness gain curve (Yund and McCartney 1994; McCartney 1997; Yund 1998). There is, however, a large discrepancy between the plant and animal literature in terms of availability of cross species comparisons of sex allocation. While sex allocation has been studied across numerous taxa in plants (reviewed in Campbell 2000; de Jong and Klinkhamer 2005; Sicard and Lenhard 2011), we are aware of only four small comparative studies on animals (Singh and Schärer in preparation; Petersen 1991; St. Mary 2000; Maxfield

et al. 2012). Further, only the two studies have accounted for phylogenetic relationships (this is also true for most studies in plants), which is problematic since species cannot be considered independent samples (Felsenstein 1985; Garamszegi 2014). Additionally, microevolutionary data exists from the flatworm *Dugesia polychroa,* showing reduced male allocation in populations with sperm-dependent parthenogenesis compared to outcrossing populations (Weinzierl et al. 1998, 1999).

Here we present a comparative study of sex allocation in a clade of free-living SH flatworms of the genus *Macrostomum*. The genus currently contains more than 200 described species (http://turbellaria.umaine.edu), with a large undescribed species diversity (Brand et al. in preparation) and large variation in sperm and genital morphology (Ferguson 1954; Luther 1960; Ax 2008; Schärer et al. 2011). These flatworms are ideal organisms to study sex allocation due to their transparency, which allows the estimation of testis and ovary size, and thus also sex allocation, *in vivo* (e.g. Ladurner et al. 2005; Janicke et al. 2013) and because a robust phylogenetic framework has recently been established using whole transcriptome data (Brand et al. in preparation). Most importantly, *Macrostomum* flatworms show interesting variation in their reproductive biology that potentially allow to evaluate the effect of both self-fertilisation and postcopulatory sexual selection on sex allocation. Next, we will explain two contrasting mating strategies that occur within the genus and how they might affect sex allocation. These are referred to as mating syndromes, since they constitute a repeatable combination of morphological characters (Schärer et al. 2011).

Macrostomum species with the reciprocal mating syndrome (Singh et al. in preparation; Schärer et al. 2004a, 2011), insert their male copulatory organ (stylet) into the recipient's female storage organ (antrum), in which sperm can then be stored for future use (Luther 1947). Interestingly, many species with reciprocal copulation perform a postcopulatory behaviour, where worms put their mouth on their own female genital opening in the so-called “suck” behaviour, probably to remove received ejaculate (Singh et al. in preparation; Schärer et al. 2004a; Vizoso et al. 2010; Patlar et al. 2020; Weber et al. 2020). This behaviour likely represents a female resistance trait that has evolved in the context of sexual conflict over the fate of the received ejaculate (Vizoso et al. 2010), which is supported by the presence of a possible male persistence trait in the form of stiff lateral bristles on the sperm, which could in turn serve an anchoring function during the suck behaviour and thereby counteract the attempt at sperm removal by the recipient (Vizoso et al. 2010) (Figure 2 top). Several prostate secretions have been identified in *M. lignano* that reduce the recipient's propensity to perform the suck behaviour (Patlar et al. 2020), which further supports this sexual conflict interpretation. *M. lignano* also plastically adjusts its sex allocation in response to the social group size, with animals raised in larger groups having a more male-biased sex allocation (Schärer et al. 2005; Janicke et al. 2013), as predicted by theory. The plasticity in sex allocation is likely due to changes in

sperm competition, since with increased social group, more worms contribute to stored sperm in the antrum (Janicke et al. 2013). It has further been determined that testis size is a good indicator of the sperm production rate (Schärer and Vizoso 2007) and that sex allocation is positively correlated with the mating rate (Janicke and Schärer 2009).

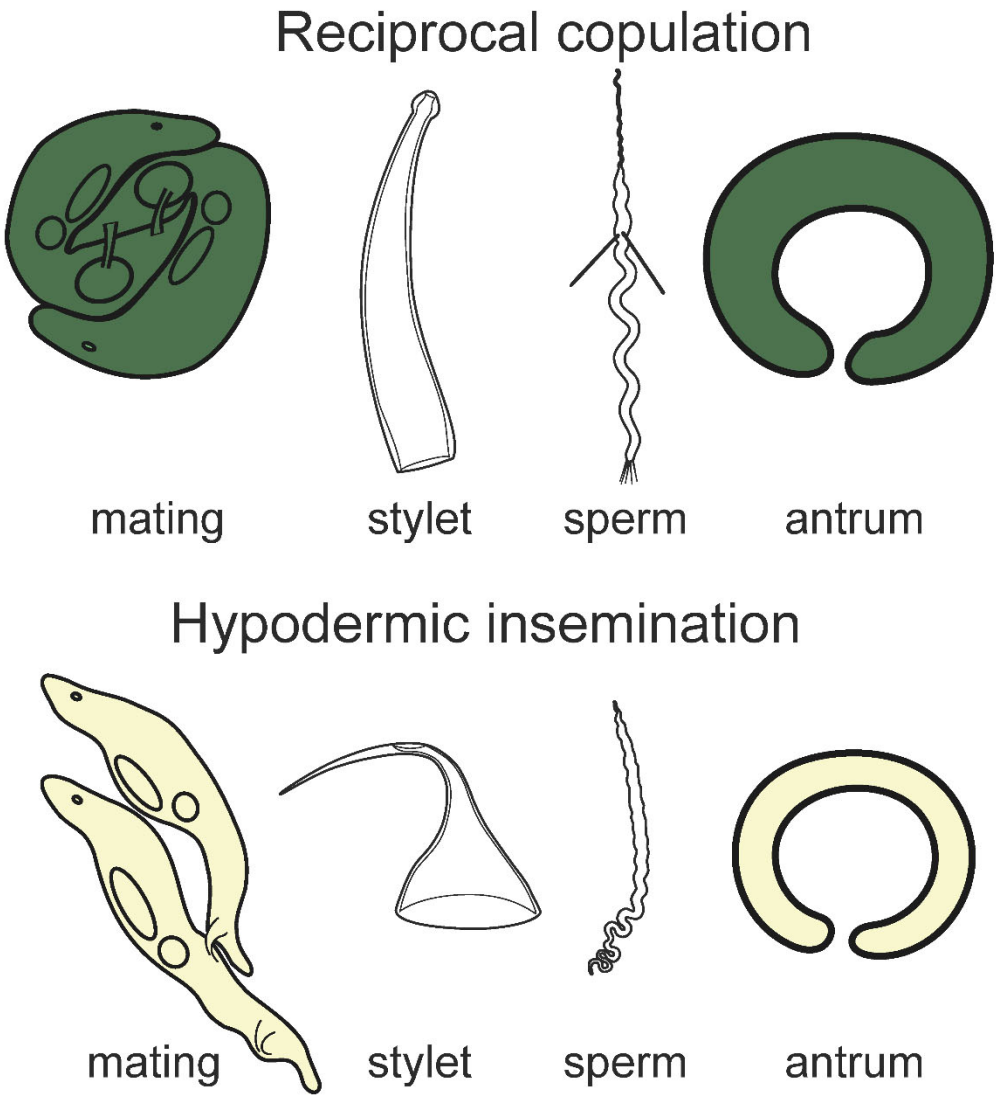


Figure 2 Stylized representation of the mating syndromes. Reciprocal copulation (top) with a non-invasive stylet, sperm with lateral bristles and a thickened antrum . Hypodermic insemination (bottom) with a needle-like stylet, simple sperm and a simple antrum. Note that these colours are used throughout the text to represent species that exhibit the reciprocal vs. hypodermic mating syndrome, (with species that were assigned to the intermediate syndrome and species that cannot be assigned to either depicted as light green and grey, respectively).

In contrast, *Macrostomum* species with the hypodermic mating syndrome, mate via hypodermic insemination (HI), during which donors traumatically inject sperm into the tissue of the recipient using a needle-like copulatory organ. Comparative

analyses have shown that HI has convergently evolved at least 9 times, but possibly up to 13 times (Brand et al. in preparation) and that its emergence results in correlated changes in the male and female genital morphology, and changes in the sperm design (Brand et al. in preparation; Schärer et al. 2011). Species with HI have a simplified antrum, shorter and simpler sperm that either lack or carry drastically reduced lateral bristles, and they have shorter stylets that are adapted for HI (Figure 2 bottom). The many independent origins of HI within the genus make *Macrostomum* ideal to study its effect on sex allocation. HI likely has consequences for the intensity and mode of sperm competition and is thus expected to influence the optimal sex allocation (Schärer and Janicke 2009; Schärer et al. 2011). In reciprocally mating *Macrostomum* species the donated sperm is stored inside the antrum, where it can be exposed to both the suck behaviour by the recipient and sperm displacement by competing donors, both of which are likely increase the realised level of local sperm competition (Charnov 1996; van Velzen et al. 2009; Schärer and Pen 2013). In contrast, in species with HI the donor circumvents the antrum, thus possibly minimizing the recipient's ability to gain control over the fate of the received ejaculate, and competing donors may also have less access to sperm already in storage. Sperm removal or sperm displacement is thus less likely and HI could therefore lead to less local sperm competition (Parker 1990) and result in a relative increase in allocation to the male function (Schärer and Janicke 2009; Schärer et al. 2011).

While HI might lead to a reduction in local sperm competition, it has also been associated with the ability to self-fertilise (Singh and Schärer in preparation; Ramm et al. 2012, 2015; Giannakara and Ramm 2017; Winkler and Ramm 2018). Selfing would lead to a drastic increase of local sperm competition and consequently reduced investment into the male function (Charlesworth and Charlesworth 1981; Charnov 1982). The exact route through which selfing occurs is unclear in most species, but it has been studied in the hypodermically mating *M. hystrix*, where is achieved by injecting sperm hypodermically into the head region (Ramm et al. 2015). However, it is unclear if all species with HI are capable of selfing and at least one reciprocally mating species, *M. mirumnovem*, can also self, indicating that HI is not essential for selfing (Singh et al. 2019).

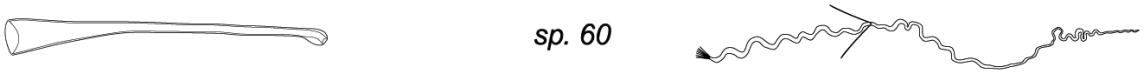
A previous comparative study has investigated sex allocation in seven *Macrostomum* species (Singh and Schärer in preparation). While no significant effects were found, the data tentatively suggested that both the mating syndrome and the ability to self might be correlated with sex allocation, with two out of the three hypodermically inseminating species and three of the four selfing species having the lowest investment into the male function (Singh and Schärer in preparation). However, another hypodermically mating and selfing species, *M. hystrix*, had the most male biased sex allocation of all the analysed species. It is therefore unclear what effect HI has on sex allocation.

We here substantially increase taxon sampling and include more independent origins of hypodermic mating to determine which of these species has the more typical sex allocation for HI. We further test key assumptions of sex allocation theory, namely that i) various aspects of allocation to a function should trade-off (e.g. investment into sperm or copulatory organ length could trade-off with gonad size or male-specific behaviour Michiels et al. 2009; Schärer and Pen 2013), and ii) that investment between the male and female functions trade-off. The latter assumption is central to most sex allocation models (Charnov et al. 1976; Charnov 1980; Fischer 1981; Schärer and Pen 2013), but it has rarely been tested and only a few studies have found evidence for it (reviewed in Schärer 2009). Finally, we infer the tempo and mode of evolution of sex allocation across the genus and determine, whether investment into the male function is correlated with traits indicative of intense sperm competition.

Methods

Study species and morphological measurements

Specimens were primarily collected during multiple field sampling trips across the globe. We have previously given a detailed account of the collected specimens, including image and video documentation (Brand et al. in preparation). Briefly, for documentation animals were mounted on a microscope slide and lightly squeezed dorsoventrally using a coverslip with plasticine feet. Due to the high transparency of these animals, it is then possible to observe the testes, the ovaries, the seminal vesicle, and the male and female genitalia *in vivo*. We have previously reported on numerous measurements that we collected from these specimens and how we transformed them for multivariate analysis (Brand et al. in preparation). Briefly, we measured body area, sperm length, sperm bristle length, stylet length, stylet width, stylet curvature and the asymmetry of the stylet tip by tracing the structures in ImageJ software (NIH, Figure 3). Images were calibrated using a stage micrometre. We further categorised sperm, stylet and the female genitalia on a per species basis, by scoring if sperm had a brush or velum, if the stylet tip was sharp or blunt, and we scored various aspects of the antrum (Figure 3). We summed the scores of the antrum measures to produce a compound complexity score, with increasing values representing more complex genitalia. Finally, we previously assigned species to a mating syndrome according to their genital morphology and the location of received sperm. 40 species were assigned to the hypodermic mating syndrome, 69 species to the reciprocal mating syndrome, two species were assigned to an intermediate mating syndrome and nine species could not be assigned because of insufficient or contradictory information (Brand et al. in preparation).



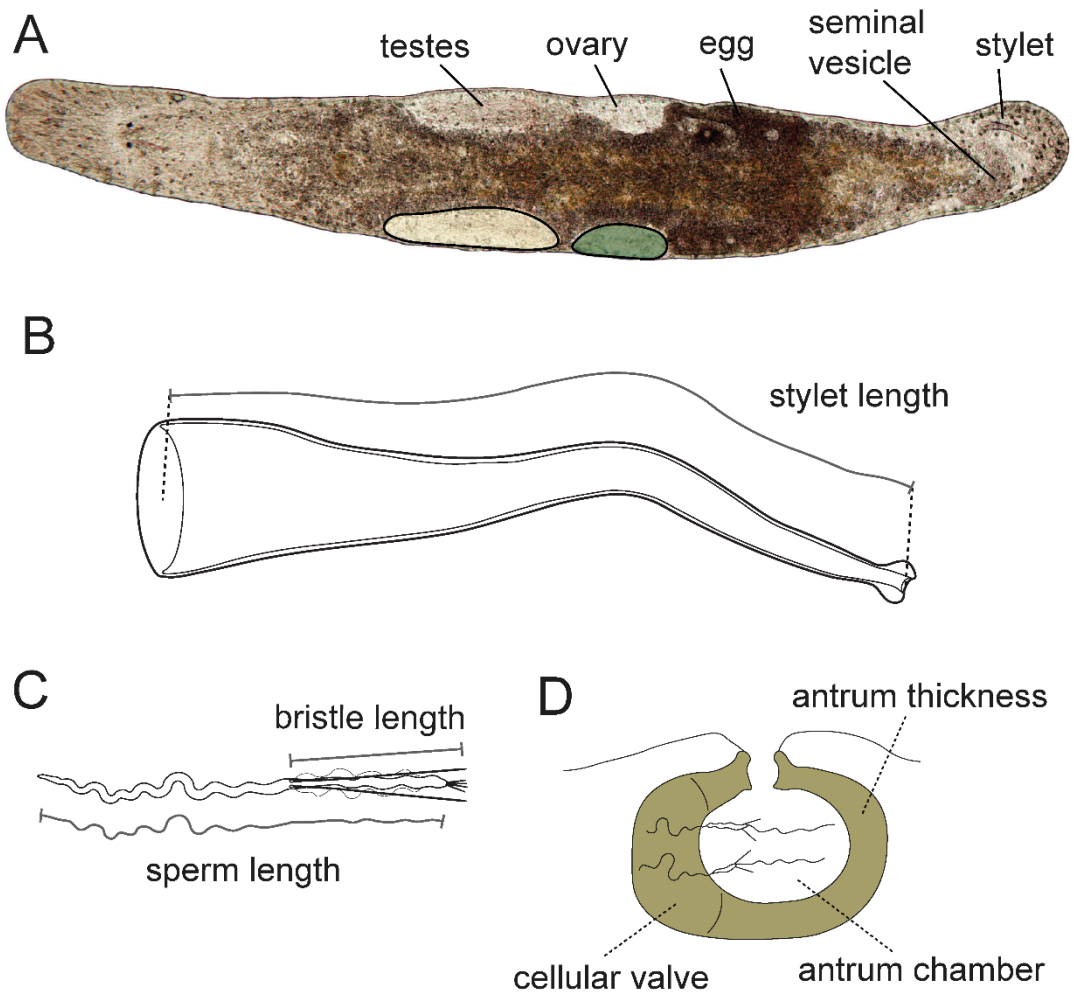
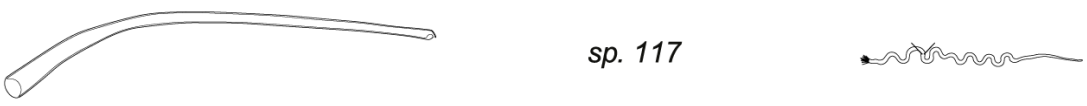


Figure 3 Details on the general morphology of Macrostomum and the specific measurements used. A: M. lignano with visible internal organs and an indicated outline of how the gonads were measured. Below are drawings of a stylet (B), a sperm (C) and an antrum (D) with representations of the measurements taken or the structures scored. Linear quantitative measurements were taken as indicated in B and C. Such measures were not possible for the antrum and we thus scored the indicated structures on a ordinal scale and summed these scores to arrive at an overal measure of antrum complexity (high values indicate a more complex antrum).

In addition to these data, we, for the present study, estimated gonadal sex allocation (GSA, testis area / (testis area + ovary area)) for 1092 specimens across 120 *Macrostomum* species, with most specimens being documented within a few days after extraction. To estimate testis and ovary size we traced the outline of the gonads manually using ImageJ (Figure 3). Whenever possible we traced both paired testes and ovaries, and used the sums as the datum. When the position of the animal made it impossible to clearly see the outline of one gonad, but we could confirm that the second gonad was actually present, then we doubled the value of the gonad we could measure. When it was unclear if the second gonad was present then we only used the estimate for one gonad as the datum (all described *Macrostomum* species have paired gonads). Gonads were generally traced at 400x magnification (1459 testes and



1366 ovaries), but traced at lower magnification if they were too large (100x: 17 testes, 20 ovaries, 200x: 378 testes, 387 ovaries), and traced at 1000x when they were very small (74 testes, 71 ovaries). In most cases, we measured gonads at the same magnification within an individual (1001 out of 1102 cases). We only estimated GSA for specimens that had active gonads, which we judged by testes having visible elongating spermatids towards the lumen, and ovaries having visible oocytes, yolk granules at the posterior end of the ovary and in most cases developing eggs. Moreover, we also required that specimens showed a complete male reproductive system, consisting minimally of a seminal vesicle, a vesicula granulorum, and the stylet. We calculated GSA on untransformed values, but to the variable body sizes of the species and specimens, we did not standardize the thickness of the squeeze preparation, as it is usually done when measuring GSA in *Macrostomum* (Schärer and Ladurner 2003; Janicke et al. 2013; Giannakara and Ramm 2017). Because of this it is thus problematic to directly compare estimates of gonad size between species, however, we have determined that GSA estimates are unaffected by the squeezing intensity in four species (Brand et al. unpublished). To compare gonad area directly between species, we t calculated the residuals of a linear regression of the gonad size on the body size across all species (all square-root transformed and then taken the $\log_{10}$, testis: $F_{1,119}=37.02$, $\beta = 0.64$, $p < 0.001$, adjusted $R^2 = 0.24$; ovary: $F_{1,118}=179.8$, $\beta = 1.04$, $p < 0.001$, adjusted $R^2 = 0.60$).

Phylogeny

For the comparative analyses we used a recently generated ultrametric molecular phylogeny of the genus *Macrostomum* that is based on a combination of whole-transcriptome RNA-Seq (98 species) and Sanger sequenced 28S data (47 species, Brand et al. in preparation). While the major phylogenetic groupings were consistent between different inference methods and the sequence alignment used, the previous study found some discrepancies in the backbone of the phylogeny (Brand et al. in preparation). We present results from a phylogeny made from an alignment containing 385 genes from the 98 species with a transcriptome plus a 28S fragment including an additional 47 species and calculated with a maximum likelihood approach (C-IQ-TREE). To assess whether tree uncertainty had an influence on our conclusions we also performed all analysis on two phylogenies that only include species with transcriptomic data available, calculated using maximum likelihood (H-IQ-TREE) or with a Bayesian approach (H-ExaBayes). The analyses on the alternate topologies were quantitatively similar and qualitatively identical (this was also the case in the analyses in Brand et al. in preparation) and we therefore only present results from analysis performed on the C-IQ-TREE topology.

Evolution and distribution of gonadal sex allocation

We calculated the empirical density of GSA using the "density" function in R (version 3.5.1) and determined the bandwidth using the method described in (Sheather and Jones 1991). We estimated phylogenetic signal (both lambda and Bloomberg's K) in the distribution of GSA using the function *phylosig* from the R package *phytools* (version 0.6, Revell 2012) and conducted likelihood ratio tests to determine if the signal differed from zero or Brownian motion. Next, we calculated ancestral states of GSA using the *contMap* function from *phytools*, which calculates maximum-likelihood estimates for all internal nodes using the method by Felsenstein (Felsenstein 1985). To explore whether peaks in the density of GSA corresponded to distinct peaks in a macroevolutionary landscape we fitted three evolutionary models to the data using the R package BBMV (version 2.1, Boucher 2019). This package allows the fitting of multi-peak evolutionary landscapes using the Focker-Plank-Komolgorov diffusion equation, comparing flat landscapes (equivalent to Brownian motion, BM), to single peak landscapes (equivalent to a single peak Ornstein-Uhlenbeck model, OU), to bounded Brownian motion models with evolutionary potential with up to two peaks (BBMV, Boucher et al. 2018). These landscapes are modelled using a polynomial equation with three terms ($ax^4 + bx^2 + cx$) and calculated over discretized sections of the values. We bounded GSA (0,1) and fit all models with 200 discretization windows (Npts = 200) and then compared the AIC-weights of the full model (estimating a, b and c), the OU model ($a = 0$) and the BM model ($a = 0$, $b = 0$, $c = 0$). Since the full model has two peaks when a and b have opposite signs, we calculated the 95% confidence interval for the polynomial terms using the "Uncertainty_FPK" function.

Phylogenetic least-squared regression

We performed phylogenetically corrected ordinary least squared regression (PGLS) with the *gls* function in the R package *nlme* (version 3.1). We incorporated phylogenetic signal in the residuals of the regression and simultaneously accounted for the varying sample sizes using the number of specimens for which we had GSA estimates as weights ("~1/N_{GSA}"). We determined the evolutionary model for the covariance in the residuals by fitting Brownian motion, lambda and Ornstein-Uhlenbeck models and then chose the model with the lowest corrected AIC. We assessed if the assumptions of the PGLS were met by checking the distribution of the phylogeny-corrected residuals for normality and profiled the likelihood of the parameter of the correlation structure (i.e. lambda or alpha). In some smaller subsets of the data, we had issues with the convergence of the models because the likelihood profile for lambda increased towards negative values. In these cases, we restricted lambda to zero, making the model equivalent to a weighted OLS. In a few cases, both the lambda and OU models fit equally well, in which case we used the lambda model.

Since R-squared values are problematic for OLS models (Ives 2019) we calculated R_{pred} (Ives and Li 2018) to show model fits.

To test if there was a trade-off between allocation to the male and the female function, we performed a weighted PGLS of residual testis area on the residual ovary area. We further identified evolutionary correlates of GSA as the dependent variable with a number of different analyses. To maximize the number of species available for each predictor we initially performed univariate analysis. We used four binary categorical predictors that are closely associated with the two mating syndromes (Brand et al. in preparation) namely the location of received sperm (at least some sperm hypodermic vs. sperm in antrum only), the state of the sperm bristles (no bristles or bristles reduced vs. bristles present), the morphologically inferred mating syndrome (hypodermic vs. reciprocal, excluding the two species classified as intermediate) and the thickness of the antrum (thin vs. thickened). We further used one variable summarising investment into female genitalia (antrum complexity), several traits that may represent non-gonadal aspects of male allocation (stylet length, sperm length, sperm bristle length) and also body area. Finally, we performed a univariate PGLS analysis assessing whether GSA was correlated with total investment into gonads (gonadosomatic index: (testis area+ ovary area)/body area) and an estimate of per-site heterozygosity of the species for which we have a transcriptome (see next section).

In order to explore more complex effects of specific trait combinations we also performed multivariate analysis for GSA with the five morphological traits as predictors. Because the mating syndrome was significantly correlated with GSA and since we were not able to assign all species to a syndrome, we performed all analysis across the entire dataset and also restricted to species classified as belonging to the hypodermic or reciprocal syndrome. We do not apply a correction for multiple testing since the data used varies slightly between tests, but we note that the p-values were usually low and would, in most cases, remain significant after adjustment.

Estimates of heterozygosity

As outlined in the introduction, sex allocation theory predicts that selfing should have a strong effect on GSA and we have some evidence that selfing may be more widespread in species with hypodermic insemination. It would therefore be interesting to have data on the selfing rates of the different species. Unfortunately, selfing rates have not been estimated for any *Macrostomum* species to date, and for the mostly field-collected specimens we analysed here obtaining such estimates was outside of the scope of our study. However, estimates of heterozygosity can be obtained from transcriptome data, which could give some indication of genetic diversity across species, which we would expect to be reduced in species that perform selfing regularly (although selfing is of course not the only process that can lead to

the loss of heterozygosity, but so may biparental inbreeding and small population sizes).

We estimated per site heterozygosity using a superTranscript approach (Davidson et al. 2017), as integrated with the Trinity suite of tools. A superTranscript represents a linear concatenation of all the assembled isoforms of a gene, which are then used as a reference for read mapping using a splice aware mapper. We generated superTranscripts for each of the transcriptomes assembled for Brand et al. (in preparation), excluding transcriptomes from species without GSA estimates or for which we pooled RNA samples from many specimens, leaving us with 130 transcriptomes across 90 species. We then called haplotypes using the GATK 3.8 RNA variant calling pipeline (Van der Auwera et al. 2013). Briefly, we mapped the trimmed reads to the superTranscripts using STAR in 2-pass mode, post-processed the alignment using Picard, translated mapping scores using GATK *SplitNCigarReads*, and finally called variants using the GATK *HaplotypeCaller*. We discarded SNP clusters (three or more SNPs within a 35-base sliding window) and SNPs with a Fisher Strand value larger than 30 and/or a quality depth score below 5 (This represents quite more aggressive filtering than the standard pipeline, Van der Auwera et al. 2013). We calculated the per-site heterozygosity as the number of SNPs divided by the total number of bases across all superTranscripts. Most assemblies derived from a single individual and we thus simply mapped the reads used for the assembly. Eight of the transcriptomes were derived from sequencing data from multiple individuals per species (seven species with three samples and one species with two samples), collected from the same location. In these cases, we called variants for each individual separately by mapping the reads to the species reference transcriptome and averaged estimates for each species. We assessed the repeatability of the estimates by calculating the Intraclass Correlation Coefficient (ICC) for species with more than one sample using the R package *ICC* (Wolak et al. 2012). Repeatability was quite low (N=22, k=2.78, ICC=0.32, 95%CI=0.04-0.6) but this was mainly due to variation between different sampling site since the ICC was large when we compared specimens collected from the same site (N=13, k=2.53, ICC=0.87, 95%CI=0.69-0.96, Figure S1). We tested whether hypodermic insemination is associated with a reduction in heterozygosity, by performing a PGLS with heterozygosity as the dependent variable and the mating syndrome as the predictor.

Results

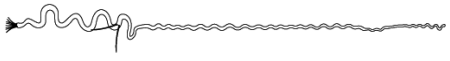
Evolution and distribution of gonadal sex allocation

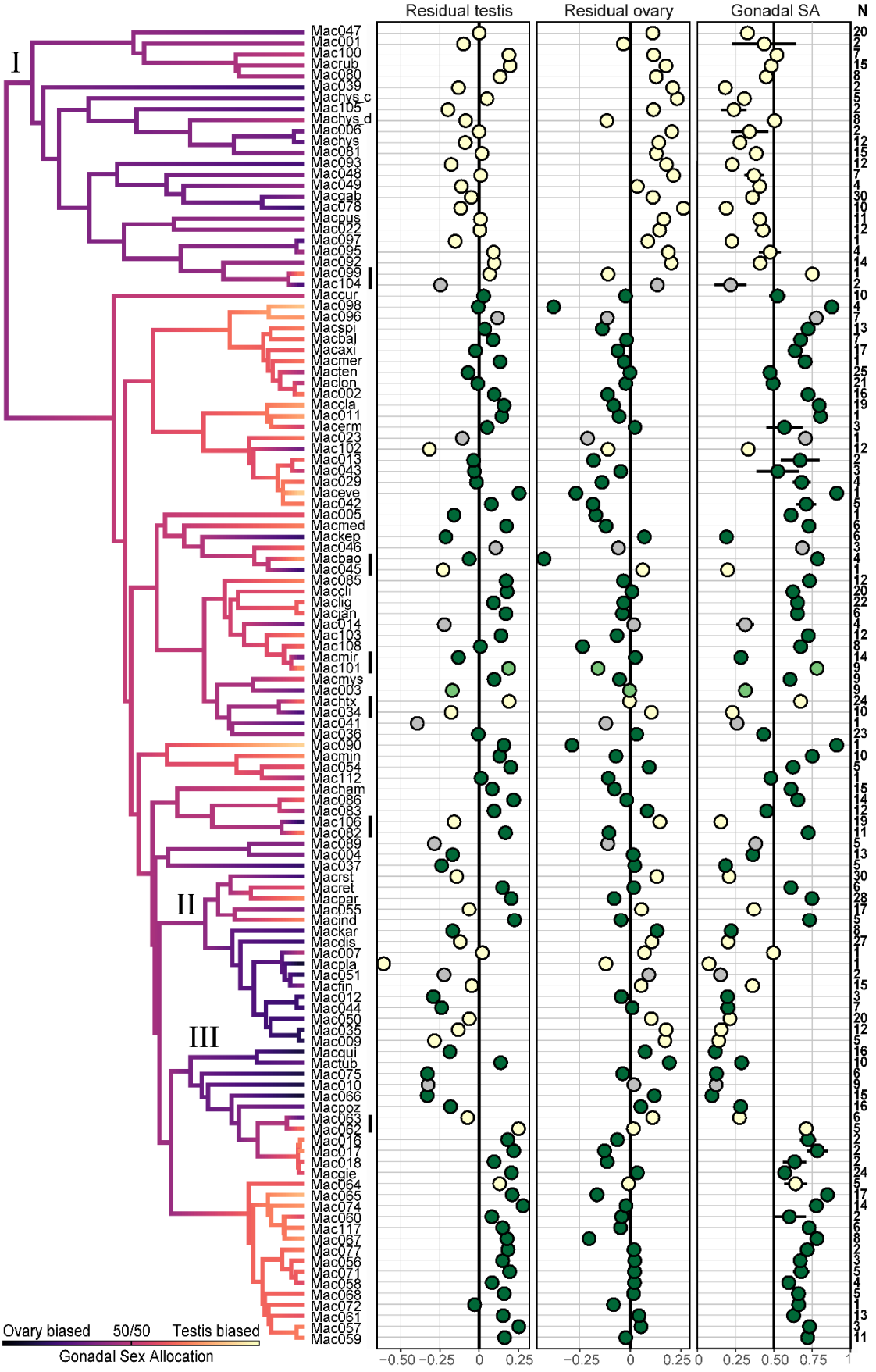
The obtained estimates of gonadal sex allocation (GSA) varied considerably across the 120 *Macrostomum* species in the phylogeny (Figure 4), and they showed significant phylogenetic signal, indicating that closely related species had a tendency

sp. 56

to have similar GSA, but showing less covariance than expected under Brownian motion (λ=0.64, $p_{>0}$<0.001, $p_{<1}$<0.001; K=0.33, $p_{>0}$<0.001, $p_{<1}$<0.001). This was also true for residual testis size (λ=0.47 $p_{>0}$<0.001, $p_{<1}$<0.001; K=0.23, $p_{>0}$<0.001, $p_{<1}$<0.001), residual ovary size (λ=0.55 $p_{>0}$<0.001, $p_{<1}$<0.001; K=0.47, $p_{>0}$<0.001, $p_{<1}$=0.01) and body size (λ=0.80 $p_{>0}$<0.001, $p_{<1}$<0.001; K=0.39, $p_{>0}$<0.001, $p_{<1}$=0.01), with only body size showing covariance close to the expectation under Brownian motion. The mean of GSA across all species was 0.50 (median 0.52, both not corrected for phylogenetic dependence) but species ranged from strongly ovary-biased (0.08) to strongly testis-biased (0.91) thus showing variation across the whole possible spectrum.

The ancestral state reconstruction showed that the large clade containing only hypodermically-mating species (I in Figure 4) likely had ancestors with a more ovary-biased GSA and within this clade, only one species (*M. sp. 99*) had a GSA that was substantially larger than 0.5. But this estimate is based on only one specimen and could, therefore, be unrepresentative of this species (Figure 4). However, although we had few specimens for 22% of the collected species (14 with one and 12 with two samples), we were able to collect larger number for most species (N<5: 13 species, N=5-10: 37 species, N>10: 44 species) and standard errors for the GSA estimates are generally low (mean SE: 0.04), which suggests single specimens are quite useful estimates of species means. The ancestral state reconstruction further showed the other clade, consisting chiefly of reciprocally-mating species, to have a more testis-biased GSA. There are also two clades that also tend to show a more ovary-biased GSA (II and III in Figure 4). Interestingly, however, both of these clades also contain species with more testis-biased GSA (e.g. *M. paradoxum* and *M. inductum*, and the small subclade with *M. gieysztori*, respectively). Furthermore, there are several closely related sister species that differ substantially in GSA, suggesting that it can evolve rapidly over short evolutionary timeframes (Figure 4). What is striking about these sister species, is that at least one of them is always either assigned to the hypodermic or intermediate syndrome suggesting that this mating type influences GSA (see our formal test of this below).




Residual testis
Residual ovary
Gonadal SA
N
I
II
III
Mac047
Mac001
Mac100
Macrub
Mac080
Mac039
Machys_c
Mac105
Machys_d
Mac006
Machys
Mac081
Mac093
Mac048
Mac049
Macgab
Mac078
Macpus
Mac022
Mac097
Mac095
Mac092
Mac099
Mac104
Maccur
Mac098
Mac096
Macspi
Macbal
Macaxi
Macmer
Macten
Maclon
Mac002
Maccla
Mac011
Macerm
Mac023
Mac102
Mac013
Mac043
Mac029
Maceve
Mac042
Mac005
Macmed
Mackep
Mac046
Macbao
Mac045
Mac085
Maccli
Maclig
Macjan
Mac014
Mac103
Mac108
Macmir
Mac101
Macmys
Mac003
Machtx
Mac034
Mac041
Mac036
Mac090
Macmin
Mac054
Mac112
Macham
Mac086
Mac083
Mac106
Mac082
Mac089
Mac004
Mac037
Macrst
Macret
Macpar
Mac055
Macind
Mackar
Macdis
Mac007
Macpla
Mac051
Macfin
Mac012
Mac044
Mac050
Mac035
Mac009
Macqui
Mactub
Mac075
Mac010
Mac066
Macpoz
Mac063
Mac062
Mac016
Mac017
Mac018
Macgie
Mac064
Mac065
Mac074
Mac060
Mac117
Mac067
Mac077
Mac056
Mac071
Mac058
Mac068
Mac072
Mac061
Mac057
Mac059
Ovary biased
50/50
Testis biased
Gonadal Sex Allocation
−0.50 −0.25 0 0.25
−0.25 0 0.25
0 0.25 0.50 0.75 1

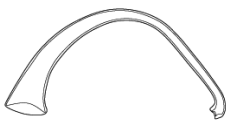
sp. 58

Figure 4 Evolution and distribution of gonadal sex allocation (GSA) across 120 species of Macrostomum. Colours along the branches show results of an ancestral state reconstruction for GSA. Some clades are marked with numerals (see Results) and sister species with strong differences in GSA are marked with a vertical bar to the right of the species abbreviation. The three panels show, from left to right, residual testis size, residual ovary size, and GSA, with dots representing species means and bars representing standard errors (SE) for the latter while there is not SE given for the former two since the residuals were calculated across all species. Dot colour indicates the mating syndrome species are assigned to (yellow: hypodermic, light green: intermediate, gree: reciprocal, grey: unclear). The last column (N) gives the number of specimens with a GSA estimate and no SE are drawn for species with only one sample (14 of 120, 12%). In some specimens we measured only one gonad type, thus N for residual testis and ovary size can be larger than the value for GSA.

Across all species, GSA had a multimodal distribution (Figure 5, solid black line), with one peak at 0.21 representing species with relatively ovary-biased values, a second peak of higher density at 0.7 representing species with relatively testis-biased values, and a local minimum at 0.53 representing that there are few species with relatively equal-sized ovaries and testes. Separating species by inferred mating syndrome suggests that the ovary-biased and testis-biased peak correspond to the hypodermic and reciprocal mating syndrome, respectively (Figure 5). Almost all species in the hypodermic syndrome have a GSA <0.5, while the majority of species in the reciprocal syndrome have values >0.5. Fitting of models of the mode of evolution using BBMV revealed that a macroevolutionary landscape with two peaks was a better fit to the data than a single peak OU model or a Brownian motion model (Table 1). This multimodal model had one peak at an ovary-biased GSA of 0.22 (evolutionary potential: 3.34, dashed black line) that closely corresponds to the first peak of the empirical GSA distribution (Figure 5). Moreover, it identified a second peak at a testis-biased GSA of 0.76 (evolutionary potential: 1.11, dashed black line) that was slightly higher than the large peak in the empirical GSA (Figure 5). The evolutionary potential of the ovary-biased peak was more than three times larger than the testis-biased peak, indicating that species quickly evolve towards low relative investment into testis, while they do not increase their GSA as rapidly. This is consistent with rapid changes in a factor decreasing sperm competition, such as a shift in mating syndrome and/or the selfing rate, which is further supported by the predicted rapid changes in ancestral states (Figure 4).

Table 1 Results from BBMV analysis uncluding gonadal sex allocation of 120 species. For each model given is the Brownian motion variance (σ^2) and the values of the parameters of the polynomial terms used to shape the macorevolutionary landscape (ax^4 + bx^2 + cx) and the AIC weights. Values in brackets are the bounds of the 95% CI. Note, that b is negative for the BBMV model which results in a landscape with two peaks.

Model	σ^2	a	b	c	AICw
BBMV	0.78 (0.55, 1.33)	3.12 (2.73, 3.61)	-4.12 (-4.69, -3.48)	0.68 (0.1, 1.2)	>0.99
OU	0.72 (0.4, 1.04)	-	0.76 (0.37, 1.19)	0.38 (-0.18, 0.73)	<0.01
BM	0.77 (0.28, 1.25)	-	-	-	<0.01



sp. 68

Trade-off

We find support for a trade-off between investment into the male or female gonad. Across all species, residual testis area was negatively correlated with residual ovary area (Table 2). Increased investment into ovarian tissue thus coincides with reduced investment into testis tissue suggesting that the two gonads are built from a shared resource budget. Further, this indicates that changes in GSA were not only due to changes in one gonad, but instead reflect true adjustment of sex allocation.

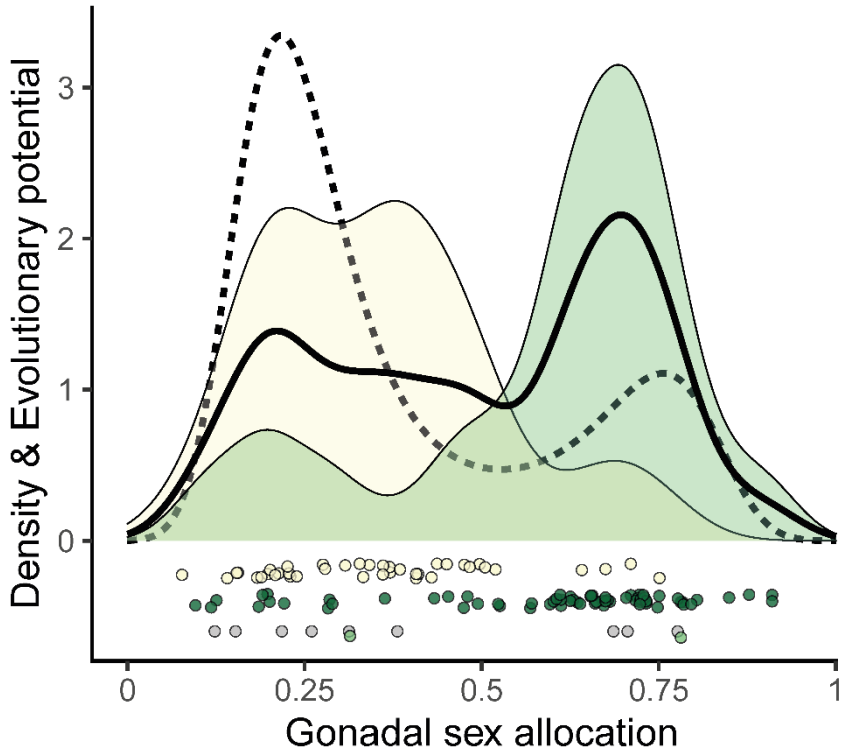


Figure 5 Density plot of GSA across all 120 Macrostomum species. The black solid line represents the empirical density across all species, and coloured curves show the densities of the species showing the hypodermic (yellow) and reciprocal (green) mating syndrome. The points below show the individual data points, jittered in the y-axis direction for visibility. Points coloured in light green and grey represent species that were assigned to the intermediate mating syndrome or could not be assigned, respectively. The dashed line shows the significantly bimodal evolutionary landscape inferred using BBMV.

Predictors of sex allocation

Hypodermic mating, as indicated by the location of received sperm, the state of the sperm bristles, the antrum thickness and the inferred mating syndrome, is clearly associated with a reduction in GSA (Figure 6). We thus performed all correlation analysis with all species and with subsets of the species that were assigned to a mating syndrome to assess if correlations exist within each syndrome or if they are mainly driven by them. Across all species, the univariate PGLS of morphological characters showed a positive correlation of GSA with antrum complexity, stylet length, total sperm length, and a negative correlation with body area and bristle

length (Figure 7 and 8). This was confirmed by the multivariate PGLS, which showed a similar pattern except that bristle length was not significantly correlated (Table 3a). Sperm bristle length is correlated with both stylet and sperm length and it is thus likely that the bristle length effect was masked in the multivariate analysis.

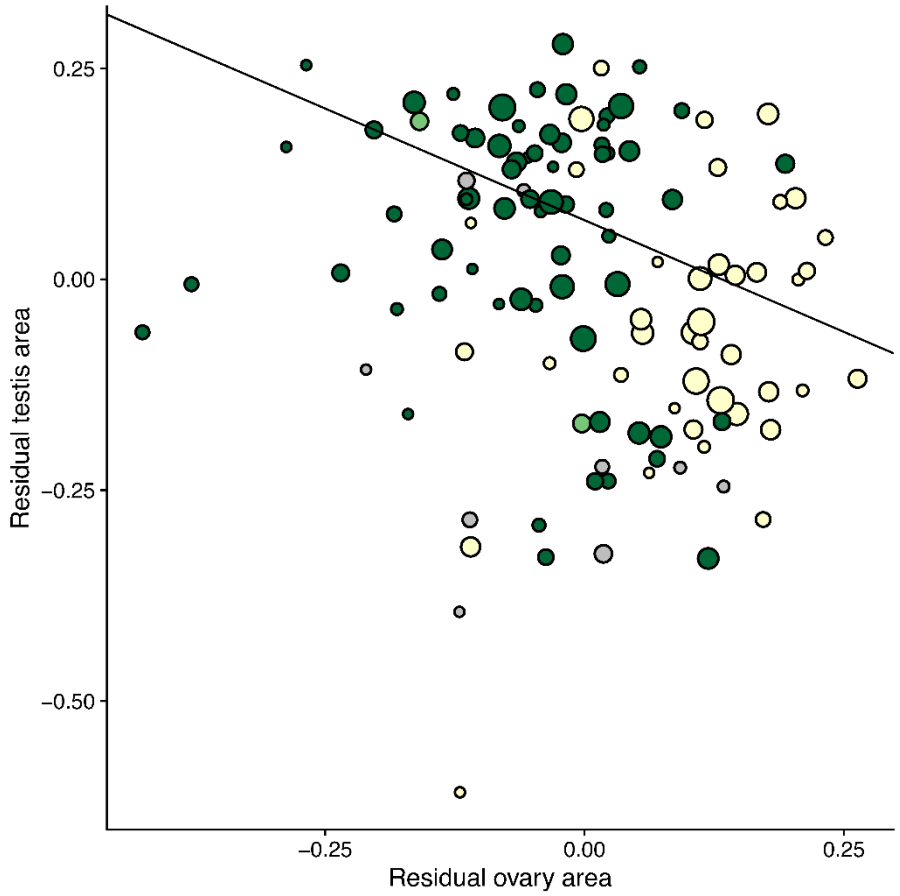


Figure 6 Residual ovary size against residual testis size across all 120 Macrostomum species measured. Colour shows the inferred mating syndrome: hypodermic (yellow), intermediate (light green), reciprocal (green) and unclear (grey). Points are scaled based on sample size to illustrate the weights used for the PGLS model that resulted in the fit line shown. Detailed results are in Table 2.

Table 2 PGLS of the residual testis and residual ovary area across all species. P-values of slopes <0.05 are printed in bold.

Predictor	N	λ	Estimate	SE	*t*	*P*	R^2_{pred}
Intercept	120	0.51	0.07	0.03	2.18	**0.03**	0.11
residual Ovary			-0.53	0.14	-3.70	**<0.001**	

sp. 61

Table 3 Multivariate PGLS between gonadal sex allocation and morphological traits (a-c). Scatterplots for the principal components are in Figure S2. P-values of slopes <0.05 are printed in bold.

	N	λ	Estimate	SE	*t*	*P*	R^2_{pred}
a) All species							
Intercept	111	0.13	1.08	0.32	3.40	<0.001	0.51
Antrum complexity			0.03	0.01	2.30	**0.02**	
Stylet length			0.32	0.10	3.11	**0.002**	
Total sperm length			0.33	0.11	3.11	**0.002**	
Sperm bristle length			-0.05	0.05	-1.14	0.26	
Body Area			-0.68	0.13	-5.15	**<0.001**	
b) Hypodermic syndrome							
Intercept	40	0	0.79	0.47	1.67	0.11	0.12
Antrum complexity			-0.16	0.06	-2.48	**0.02**	
Stylet length			0.20	0.25	0.78	0.44	
Total sperm length			-0.13	0.21	-0.62	0.54	
Sperm bristle length			-0.12	0.13	-0.91	0.37	
Body Area			-0.19	0.28	-0.67	0.51	
c) Reciprocal syndrome							
Intercept	64	0.04	0.81	0.39	2.08	0.04	0.45
Antrum complexity			0.04	0.02	2.55	**0.01**	
Stylet length			0.44	0.11	3.92	**<0.001**	
Total sperm length			0.42	0.12	3.52	**<0.001**	
Sperm bristle length			0.09	0.07	1.41	0.16	
Body Area			-0.81	0.15	-5.56	**<0.001**	

When performing the previous analyses on species classified as hypodermically mating, only the antrum complexity score was still significant in both the univariate and multivariate analysis (Figure 7 and 8, Table 3b-c), but with an opposite sign, indicating species with a more complex female antrum have a relatively ovary-biased GSA. Note, that the complexity score in hypodermic species has only a small range (0,1), with only four species with a score other than 0 and subsequently the predictive power of these models was not high (Figure 7C, $R^2_{pred} = 0$, Table 3b, $R^2_{pred} = 0.12$).

When only looking at species assigned to the reciprocal syndrome, we find similar results to the PGLS across all species. In the univariate and multivariate case, we find that GSA is positively correlated with antrum complexity and sperm length but negatively correlated with body area (Figure 7 and 8, Table 3c). Stylet length was not significant in the univariate analysis, but was positively correlated again in the multivariate analysis, most likely because the multivariate model accounts for the body area. To investigate this, we repeated the stylet length PGLS with body area included as a covariate. In this analysis, stylet length was indeed significantly

positively correlated with GSA (lambda=0.6, beta=0.48, SE=0.13, t-value=3.8, p<0.001), indicating that it is relative, not absolute, stylet length that predicts GSA.

Since the total investment into gonads could influence sex allocation, we also tested whether GSI predicted GSA. Species with a high overall investment into gonads (GSI) had a more testis-biased GSA across all species and this association holds also within the two syndromes (Figure 8B). This could indicate that under intense sperm competition the overall budget allocated to gonads is increased. However, it is also possible that a change in allocation towards the male function is more easily measured, which would lead to an increased GSI in strongly testis-biased species.

Heterozygosity

We found that species assigned to the reciprocal mating syndrome had a higher per-site heterozygosity indicating that hypodermic insemination is associated with a reduction of effective population size, which is consistent with the idea of increased inbreeding due to selfing in the hypodermic mating syndrome (PGLS with OU model, α=20.17, β=0.0005, SE=0.0001, t=3.58, p<0.001, R^2_{pred}=0.24). As expected from these results we also found a significant positive correlation between heterozygosity and GSA across all species (Figure 9). This correlation even persisted within the hypodermic and the reciprocal syndrome, but the effect size was lower in the former (Figure 9). Since the correlation persists in these subsets, it is not simply caused by the syndrome but suggests general higher heterozygosity in species that are likely experiencing more intense sperm competition.

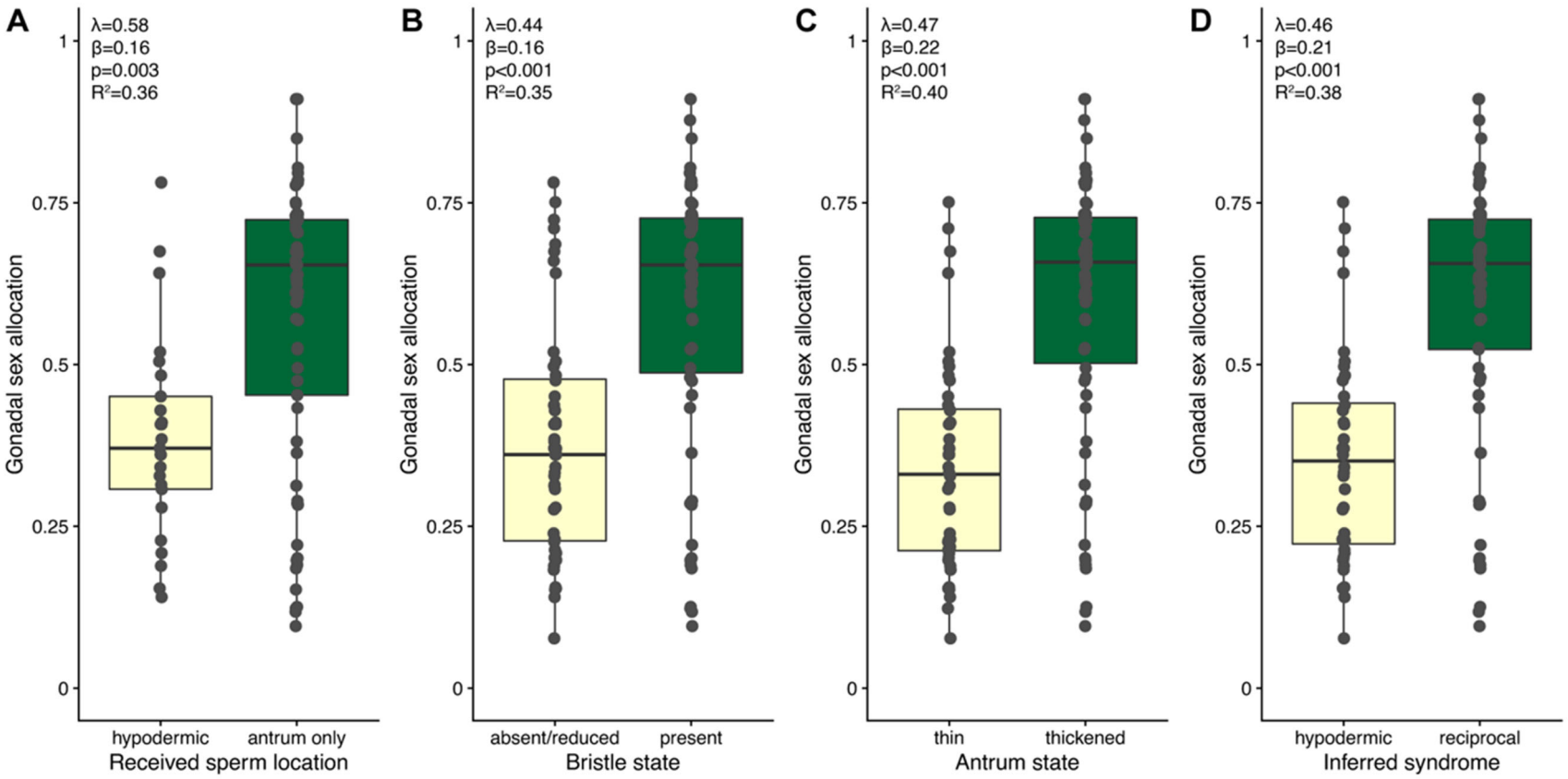


Figure 7 Gonadal sex allocation of species dependent on A: received sperm location, B: sperm bristle state, C: antrum state or D: inferred mating syndrome. Values are slightly jittered in the x direction. Within each panel the main results of PGLS analysis are given. Detailed results are in Table S1.

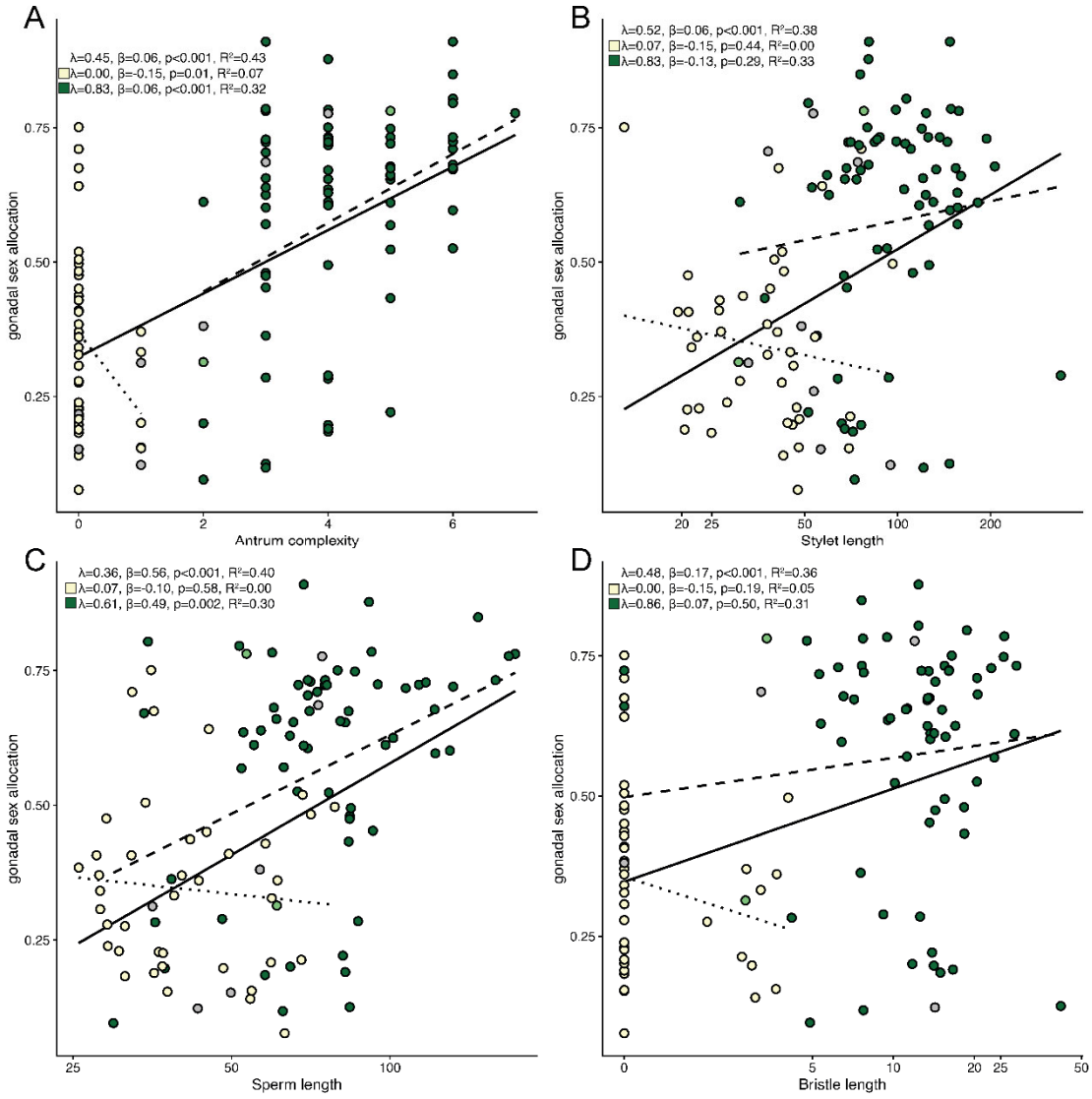


Figure 8 PGLS analysis of gonadal sex allocation predicted by four sexual traits. Colour shows the inferred mating syndrome: hypodermic (yellow), intermediate (light green), reciprocal (green) and unclear (grey). Each panel (A-D) shows the fit and corresponding statistics for PGLS analysis including all species (solid line, top), only species assigned to the hypodermic mating syndrome (dotted line, middle) or only species assigned to the reciprocal mating syndrome (dashed line, bottom). Values are plotted on a logarithmic scale in panels C-D but the axis is labelled with backtransformed values. R values represent R2pred of the full PGLS including the phylogeny. Detailed results are in Table

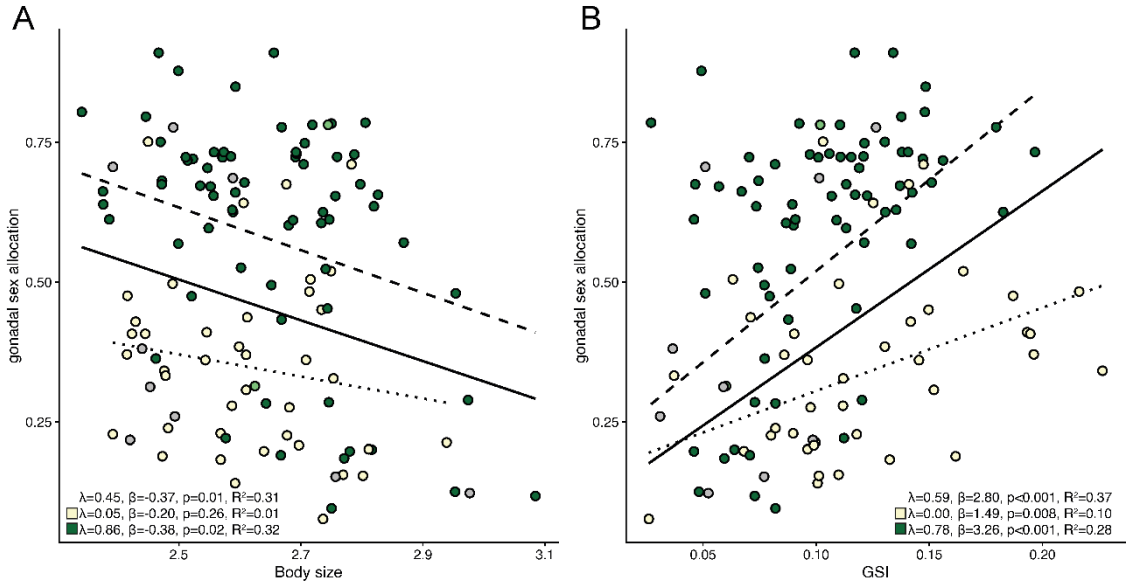


Figure 9 PGLS analysis of gonadal sex allocation predicted by body size (A) and the gonado-somatic index (B). Colour shows the inferred mating syndrome: hypodermic (yellow), intermediate (light green), reciprocal (green) and unclear (grey). Each panel shows the fit and corresponding statistics for PGLS analysis including all species (solid line, top), only species assigned to the hypodermic mating syndrome (dotted line, middle), or only species assigned to the reciprocal mating syndrome (dashed line, bottom). Values are plotted on a logarithmic scale in panel A. R values represent R2pred of the full PGLS including the phylogeny. Detailed results are in Table S2.

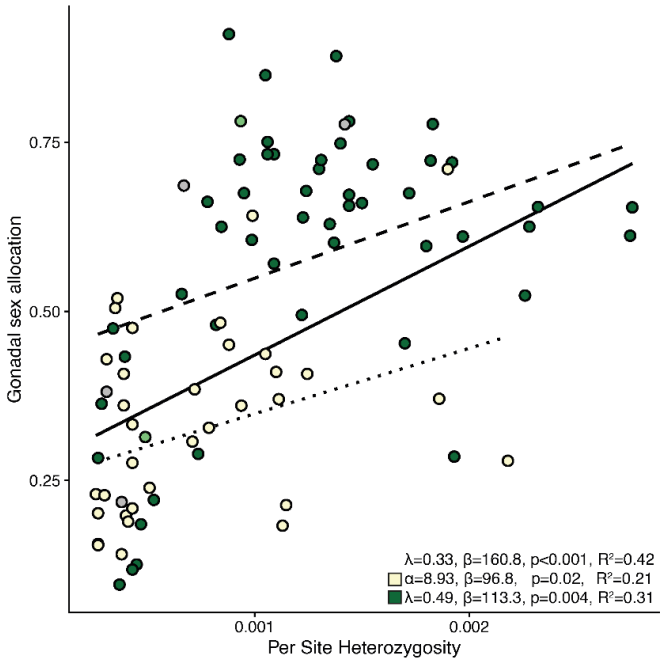


Figure 10 PGLS analysis of gonadal sex allocation predicted by heterozygosity. Colours shows the inferred mating syndrome: hypodermic (yellow), intermediate (light green), reciprocal (green) and unclear (grey). Given is the fit and corresponding statistics for PGLS analysis including all species (solid line, top), only species assigned to the hypodermic mating syndrome (dotted line, middle) or only species assigned to the reciprocal mating syndrome (dashed line, bottom). R values represent R2pred of the full PGLS including the phylogeny. Detailed results are in Table S2.

Discussion

The bimodal distribution of GSA across the genus indicates that species are predominantly either exposed to conditions with low or high sperm competition, with GSA biased towards one gonad in most species. By fitting an evolutionary model of the evolution of GSA, we showed that these empirical peaks likely correspond to two adaptive optima that species evolve towards, with rapid movement to the peak of ovary-biased GSA and a slower change towards a testis-biased allocation. This is in accordance with theoretical predictions since under conditions that drastically reduce sperm competition, such as low population density or selfing, little investment into the male function is predicted (Charlesworth and Charlesworth 1981; Charnov 1982), while species that occur at high densities should invest more into the male function (Charnov 1979). However, these models also predict that absolute allocation towards the male function should not exceed 50% because otherwise SH would be unstable, and the population would be prone to invasion of pure sex male mutants (Charnov et al. 1976). Since many *Macrostomum* species have a GSA >0.5, it is unlikely that one unit of ovary tissue represents the same amount of investment than a unit of testis tissue - making these necessarily relative rather than absolute measures. This is particularly true if different resources are needed for the production of the ovary or the testis, which could be one of the advantages of SH (Charnov et al. 1976). Further to change sex allocation, animals can not only adjust the size of the gonads but also their metabolic rate and a change in testis size can thus underestimate the actual change in allocation (Schärer and Vizoso 2007; Ramm et al. 2014). Evidence for a relationship between testis size and the rate of spermatogenesis or the proportion of spermatogenic tissue has been found across animals (Peirce and Breed 2001; Parapanov et al. 2008; Lüpold et al. 2009; Ramm and Stockley 2010), including *Macrostomum* (Schärer et al. 2004b; Giannakara et al. 2016) while this has not been assessed for ovaries. Energetic cost of the ovaries could thus be much higher than the cost of testis tissue, leading to an underestimation of female allocation. Plastictiy in the rate of oogenesis could also lead to even larger underestimation in relatively female-biased species. However, the rate and plasticity of oogenesis in *Macrostomum* would be interesting targets of future studies, it is unlikely that they would account for the testis-biased GSA. Due to the large absolute difference in size between ovary and testis, the difference in metabolic costs of ovarian tissue would have to be very substantial. Further, the found negative correlation between testis size and the rate of spermatogenesis (Schärer et al. 2004b; Giannakara et al. 2016), will likely skew sex allocation even more towards the male function in species with high GSA. Therefore, it is more likely that testes size is a better representation of allocation than ovary size ant that including other aspects of female allocation would lead to an estimated sex allocation <0.5.

Female allocation can include tissues that produce and provision the developing oocyte in ectolecithal flatworms, like *Macrostomum*, where the yolk is produced by vitellocytes that are external to the ovary (Gremigni 1983; Gremigni et al. 1987). The fully developed egg of most *Macrostomum* species is quite large compared to their body size (up to 10% of body area) and eggs can be laid once every few days (per. obs.). Therefore, it is likely that the production of yolk accounts for a large proportion of the total female allocation. Further, female investment can include the production of egg cocoons (or the mucus surrounding eggs in the case of *Macrostomum*), ovarian fluids and genitalia. Brood care behaviour could also be male or female allocation if it is exclusively directed at eggs produced via one function. Other behaviours have also been interpreted as a part of male or female allocation such as aggressiveness as a male trait in *Ophyotrocha* polychaete worms (Lorenzi et al. 2006), and potentially male behaviours such as mate guarding or mate searching. However, since many SH will act as both donors and recipients either simultaneously or reciprocally, it is difficult to assign such behaviours clearly to one function.

Naturally, GSA is also missing important parts of male allocation such as the production of seminal fluids, other prostate secretions, and the production and maintenance of genitalia and accessory organs (e.g. love darts in pulmonate snails). Investment into these traits has been suggested to confer benefits in the competition between donors in sexual conflict over donated ejaculate, potentially even leading to SA above 0.5, thus making SH prone to invasion by a pure sex female (Michiels et al. 2009). In *M. lignano*, total investment into the production of prostate secretions showed no plasticity to social groups size (Patlar et al. 2019), suggesting that at least in one species investment into prostate remains somewhat constant. But across the genus, it is very well possible that investment into such additional traits contributes significantly to male allocation.

The PGLS analysis of GIS supports the interpretation that testis size is a better proxy for male allocation, than ovary size is for female allocation. GSI was positively correlated with GSA, meaning species that allocated a larger proportion of their body to gonads also had relatively larger testes. No theoretical models predict such a relationship (Schärer 2009) and it is, thus, likely that it is a consequence of our underestimation of female allocation. Species that are relatively more female-biased, likely not only increase allocation to ovary tissue but also to the other traits discussed (e.g. yolk production), while relatively more male-biased species will mostly increase investment into sperm production. This then would directly lead to larger total gonad size in more male-biased species.

Several models assume that aspects of male allocation trade-off with each other (Michiels et al. 2009; Schärer and Pen 2013) and we thus expect a negative correlation between them. We do not find support for this assumption since GSA was positively correlated with stylet length and sperm length both representing

investment into the male function. A positive correlation between sperm length and testis size has been reported in several species (e.g. Gage 1994; Pitnick 1996; Rowley et al. 2019 but see: Gage and Freckleton 2003) and could possibly be linked to slower spermatogenesis of long sperm, requiring a larger testis to product (Pitnick et al. 1995; Pitnick 1996; LaMunyon and Ward 1999; Lüpold et al. 2009; Ramm and Stockley 2010). The correlation of both aspects of male allocation could simply indicate an overall response to intense sexual selection and as such does not imply that the trade-off does not occur within species.

GSA was also correlated with antrum complexity, which could represent female allocation. This is contrary to a second commonly made assumption that male and female allocation should trade-off (Charnov 1979, 1982). However, we do find a negative correlation between residual testis and residual ovary size, which is in support of a trade-off. This suggests that our analysis of GSA does not simply reflect changes in relative testis size but instead represent changes in sex allocation. It should be noted that although we find a significant negative correlation, the variance explained by the model was low (R^2_{pred} = 0.11) suggesting a small effect size. This suggests that investment into testis vs ovaries is to some extent independent. The trade-off assumption has been tested in plants but evidence for it is equivocal (reviewed in Campbell 2000; de Jong and Klinkhamer 2005). This is also the case in animals where evidence for a trade-off is generally lacking and has only been shown in a few cases (De Visser et al. 1994; Yund et al. 1997; Schärer et al. 2005; reviewed in Schärer 2009).

The positive correlation of the morphological character (i.e. sperm length, stylet length and antrum complexity) with GSA suggest that they all are influenced by a common selective regime and that resources allocated to gamete production do not constrain investment into genitalia and sperm design. At least in *Macrostomum,* the energy required to produce and maintain these structures is likely small in comparison to the energy needed to continuously produce sperm and eggs, but this might not be the case in other SH that repeatedly produce genitalia like e. g. intertidal Barnacles (Klepal and Barnes 1974). The morphological changes likely represent adaptations to postcopulatory processes such as cryptic female choice (antrum complexity and sperm design), sperm removal (stylet morphology and sperm design) and control over the mating duration (antrum and stylet morphology). The positive correlation of them with GSA indicates that sperm competition is higher when these processes are at play, despite them potentially introducing random or biased paternity skews that could reduce sperm competition (Charnov 1996; van Velzen et al. 2009; Schärer and Pen 2013).

We found that the bimodal distribution of GSA was largely explained by the mating syndrome with most species in the hypodermic syndrome having an ovary biased GSA. While this could potentially invalidate the previously discussed finding this is

not the case since the correlation of GSA with morphology remains within the reciprocal syndrome and there were a substantial number of species within the reciprocal syndrome that had an ovary biased GSA. Instead, our results suggest that within the reciprocal syndrome, indicators of the intensity of post-copulatory sexual selection are associated with more intense sperm competition. Whereas, hypodermic insemination is associated with an overall reduction of sperm competition, with almost all species assigned to hypodermic syndrome having an ovary biased GSA (<0.5). This finding is contrary to predictions that hypodermic insemination should lead to a more male-biased sex allocation since it could minimize the potential for cryptic female choice and lead to more fair-raffle sperm competition (Schärer and Janicke 2009; Schärer et al. 2011). Hypodermic insemination seems to be linked to the ability to self-fertilise, with some species showing no evidence of inbreeding depression or delayed selfing consistent with these species preferentially selfing (Giannakara and Ramm 2017). We have also established laboratory cultures of many additional species in the hypodermically mating clade (I in Figure 4) from single individuals suggesting selfing could be pervasive (Brand et al. pers. observation).

The link between hypodermic insemination and selfing is further supported by our analysis of heterozygosity, which showed species assigned to the hypodermic mating syndrome had a significantly lower per-site heterozygosity. However, low heterozygosity does not necessitate selfing but could also be cause by other factors leading to small effective populations size, such as low density. Unfortunately, knowledge of the ecology of *Macrostomum* is severely lacking due to the intense effort required in sample extraction. We found in our field collections that the apparent number of animals within a sample can vary from only one to hundreds of individuals in a few cm^3 of sediment (pers. obs.), but our sampling was not spatially explicit or extensive enough to estimate densities of the species collected. Therefore, we cannot exclude the possibility that hypodermically mating species mostly occur in low densities and have an ovary-biased GSA because of this. Another possibility is that injected sperm rarely encounters rival sperm thus leading to local sperm competition. This would, for example be the case if sperm had a very short residence period in the tissue before it becomes non-functional. However, given that *Macrostomum* generally produce an egg every 1-2 days, individuals would most likely need to mate frequently to maintain fertility under these conditions, which would lead to increased investment into the testis.

Conclusion

Our comparative analysis of the sex allocation and their correlates in 120 species of *Macrostomum* flatworms allowed us to evaluate several key aspects of sex allocation theory for SH. We provide evidence for the fundamental assumption that testis and ovary are built from a common resource budget. Contrary to theory we show that investment into various aspects of male allocation are correlated. We find support for the prediction that larger animals should be more female-biased but also find a positive correlation between GSI and GSA suggesting more empirical and theoretical work is needed in this area. We also determined that hypodermic insemination is associated with a reduction in GSA likely because it is associated with increased selfing and/or lower population densities. *Macrostomum* thus is a powerful system for further investigations of sex allocation in SH.

Supporting Figures and Tables

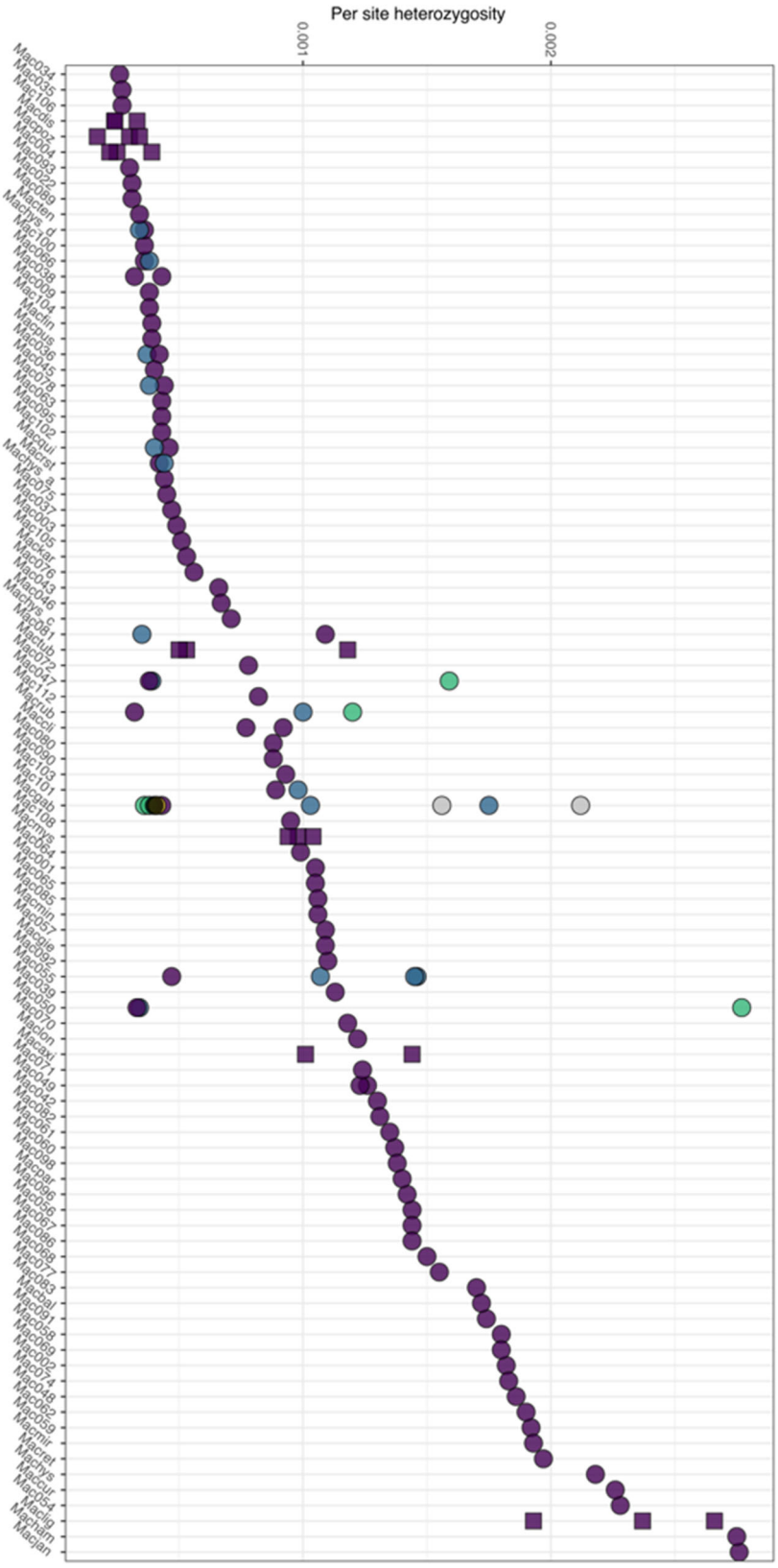


Figure S1 Per site heterozygosity for all evaluated specimens. Specimens are grouped by species and ordered by mean heterozygosity. Colours indicate specimens collected from different locations and likely correspond to populations. Heterozygosity estimates resulting from mapping to a common reference are indicated as squares. Variation within species is largely driven by differences between populations in species with intermediate values, while agreement between populations is high for low values.

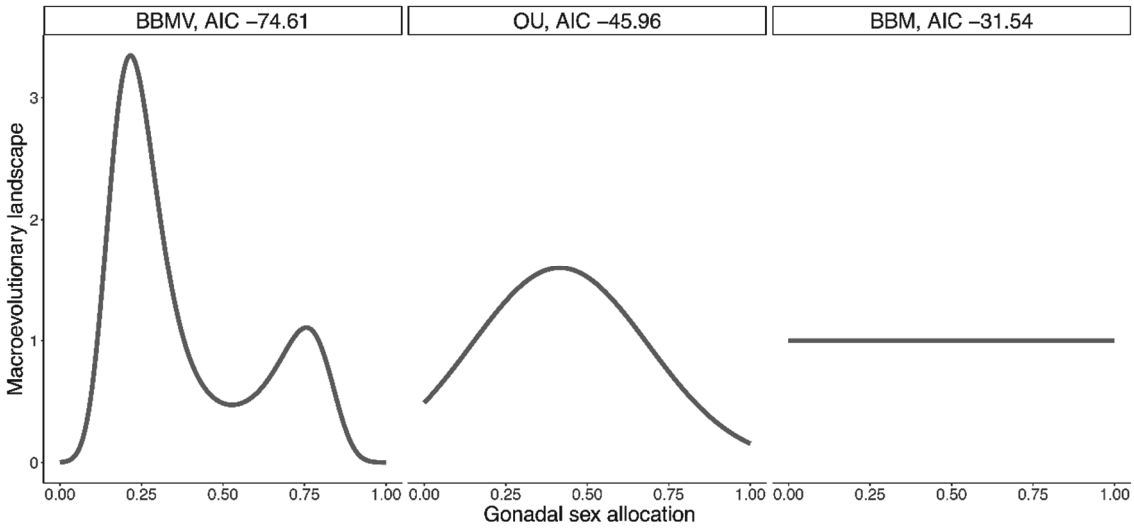


Figure S2 Macroevolutionary landscapes inferred using BBMV. Peaks in the landscape correspond to trait values species are attracted towards. Given are the three models evaluated with their AIC value. The full BBMV model was strongly supported.

Table S1 Results from PGLS analysis of gonadal sex allocation and various indicators of the mating syndrome.

	Predictor	N	λ	Est.	SE	*t*	*P*	R^2_{pred}
Received sperm location (hypodermic vs antrum)	Intercept	94	0.58	0.39	0.05	7.56	<0.001	0.36
	Antrum			0.16	0.05	3.08	**0.003**	
Bristle state (absent/reduced vs present)	Intercept	115	0.44	0.36	0.04	8.64	<0.001	0.35
	Present			0.16	0.04	3.92	**<0.001**	
Syndrome (hypodermic vs reciprocal)	Intercept	109	0.46	0.34	0.04	8.39	<0.001	0.38
	Reciprocal			0.21	0.04	4.93	**<0.001**	
Antrum state (simple vs thickened)	Intercept	118	0.47	0.34	0.04	8.52	<0.001	0.4
	Thickened			0.22	0.04	5.32	**<0.001**	

Table S1 Univariate PGLS with GSA as the dependent variable. Corresponding scatterplots are in Figure S1. P-values of slopes <0.05 are printed in bold.

Variable	Species grouping	Predictor	N	λ	Est.	SE	*t*	*P*	R^2_{pred}
Antrum complexity	*All species*	Intercept	118	0.45	0.32	0.04	8.8	<0.001	0.43
		Complexity			0.06	0.01	6.9	**<0.001**	
	Hypodermic syndrome	Intercept	40	0	0.37	0.02	15.49	<0.001	0.07
		Complexity			-0.15	0.06	-2.72	**0.01**	
	Reciprocal syndrome	Intercept	69	0.83	0.32	0.08	4.11	<0.001	0.32
		Complexity			0.06	0.02	3.9	**<0.001**	
Stylet length	*All species*	Intercept	118	0.52	-0.15	0.15	-0.97	0.33	0.38
		Length			0.34	0.08	3.98	**<0.001**	
	Hypodermic syndrome	Intercept	40	0.07	0.54	0.26	2.1	**0.04**	-0.03
		Length			-0.13	0.16	-0.77	**0.44**	
	Reciprocal syndrome	Intercept	68	0.83	0.34	0.23	1.47	0.15	0.33
		Length			0.12	0.11	1.06	0.29	
Total sperm length	*All species*	Intercept	114	0.36	-0.55	0.18	-3.06	0.003	0.4
		Length			0.56	0.1	5.58	**<0.001**	
	Hypodermic syndrome	Intercept	40	0.07	0.51	0.31	1.68	0.1	-0.01
		Length			-0.1	0.19	-0.56	0.58	
	Reciprocal syndrome	Intercept	66	0.61	-0.34	0.29	-1.19	0.24	0.3
		Length			0.49	0.15	3.19	**0**	
Sperm bristle length	*All species*	Intercept	113	0.48	0.35	0.04	8.04	<0.001	0.36
		Length			0.17	0.04	4.15	**<0.001**	
	Hypodermic syndrome	Intercept	40	0	0.36	0.03	13.83	<0.001	0.05
		Length			-0.15	0.11	-1.35	0.19	
	Reciprocal syndrome	Intercept	66	0.86	0.5	0.12	4.23	<0.001	0.31
		Length			0.07	0.1	0.69	0.5	
Body Area	*All species*	Intercept	120	0.45	1.42	0.38	3.74	<0.001	0.31
		Body Area			-0.37	0.14	-2.58	**0.01**	
	Hypodermic syndrome	Intercept	40	0.05	0.86	0.46	1.88	0.07	0.01
		Body Area			-0.2	0.17	-1.14	0.26	
	Reciprocal syndrome	Intercept	69	0.86	1.59	0.44	3.61	<0.001	0.32
		Body Area			-0.38	0.16	-2.32	**0.02**	
GSI	*All species*	Intercept	120	0.59	0.1	0.07	1.45	0.15	0.37
		GSI			2.8	0.51	5.54	**<0.001**	
	Hypodermic syndrome	Intercept	40	0	0.16	0.07	2.26	0.03	0.1
		GSI			1.49	0.54	2.78	**0.01**	
	Reciprocal syndrome	Intercept	69	0.78	0.19	0.09	2.26	0.03	0.28
		GSI			3.26	0.66	4.94	**<0.001**	
Per site heterozygosity	*All species*	Intercept	90	0.33	0.28	0.05	6.06	<0.001	0.42
		Heterozygosity			160.8	28	5.74	**<0.001**	
	Hypodermic syndrome	Intercept	34	α:8.93	0.25	0.04	7.25	<0.001	0.21
		Heterozygosity			96.8	39.4	2.46	**0.02**	
	Reciprocal syndrome	Intercept	50	0.49	0.44	0.07	6.57	<0.001	0.31
		Heterozygosity			113.3	37.3	3.03	**0**	

References

Ax, P. 2008. Plathelminthes aus Brackgewässern der Nordhalbkugel. Akademie der Wissenschaften und der Literatur ; Steiner, Mainz : Stuttgart.

Boucher, F. C. 2019. BBMV: an R package for the estimation of macroevolutionary landscapes. Ecography 42:558–564.

Boucher, F. C., V. Démery, E. Conti, L. J. Harmon, and J. Uyeda. 2018. A General Model for Estimating Macroevolutionary Landscapes. Systematic Biology 67:304–319.

Brand, J. N., G. Viktorin, C. Beisel, L. J. Harmon, and L. Schärer. in preparation. Frequent shifts to hypodermic insemination and correlated evolution of sperm and genital morphology in a genus of flatworms.

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

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

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

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

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

Charnov, E. L. 1982. The theory of sex allocation. Monogr Popul Biol 18:1–355.

Charnov, E. L., J. J. Bull, and J. Maynard Smith. 1976. Why be an hermaphrodite? Nature 263:125–126.

Charnov, E. L., R. L. Los-den Hartogh, W. T. Jones, and J. van den Assem. 1981. Sex ratio evolution in a variable environment. Nature 289:27–33.

Davidson, N. M., A. D. K. Hawkins, and A. Oshlack. 2017. SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes. Genome Biol 18:148.

de Jong, T. J., and P. G. L. Klinkhamer. 2005. Evolutionary ecology of plant reproductive strategies. Cambridge University Press, Cambridge ; New York.

De Visser, J. A. G. M., A. T. Maat, and C. Zonneveld. 1994. Energy Budgets and Reproductive Allocation in the Simultaneous Hermaphrodite Pond Snail, Lymnaea stagnalis (L.): A Trade-Off between Male and Female Function. The American Naturalist 144:861–867.

Elle, E., and T. R. Meagher. 2000. Sex Allocation and Reproductive Success in the Andromonoecious Perennial *Solanum carolinense* (Solanaceae). II. Paternity and Functional Gender. The American Naturalist 156:622–636.

Felsenstein, J. 1985. Phylogenies and the Comparative Method. The American Naturalist 125:1–15.

Ferguson, F. F. 1954. Monograph of the Macrostomine Worms of Turbellaria. Transactions of the American Microscopical Society 73:137.

Fischer, E. A. 1981. Sexual Allocation in a Simultaneously Hermaphroditic Coral Reef Fish. The American Naturalist 117:64–82.

Frank, S. A. 1985. Hierarchical selection theory and sex ratios. ii. on applying the theory, and a test with fig wasps. Evolution 39:949–964.

Gage, M. J. 1994. Associations between body size, mating pattern, testis size and sperm lengths across butterflies. Proc. R. Soc. Lond. B 258:247–254.

Gage, M. J. G., and R. P. Freckleton. 2003. Relative testis size and sperm morphometry across mammals: no evidence for an association between sperm competition and sperm length. Proceedings of the Royal Society of London. Series B: Biological Sciences 270:625–632.

Garamszegi, L. Z. (ed). 2014. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Springer Berlin Heidelberg, Berlin, Heidelberg.

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

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

Goldman, D. A., and M. F. Willson. 1986. Sex Allocation in Functionally Hermaphroditic Plants: A Review and Critique. Botanical Review 52:157–194.

Gremigni, V. 1983. Platyhelminthes-Turbellaria. Pp. 67–107 *in* Reproductive Biology of Invertebrates, Vol. 1. Oogenesis, oviposition, and oosorption. Wiley, New York, NY.

Gremigni, V., A. Falleni, and P. Lucchesi. 1987. An ultrastructural study of oogenesis in the Turbellarian *Macrostomum*. Acta Embryol. Morphol. Exper. 8:257–262.

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

Ives, A., and D. Li. 2018. rr2: An R package to calculate R^2s for regression models. JOSS 3:1028.

Ives, A. R. 2019. R^2s for Correlated Data: Phylogenetic Models, LMMs, and GLMMs. Systematic Biology 68:234–251.

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

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

Klepal, W., and H. Barnes. 1974. Regeneration of the penis in Balanus balanoides (L.). Journal of Experimental Marine Biology and Ecology 16:205–211.

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

LaMunyon, C. W., and S. Ward. 1999. Evolution of sperm size in nematodes: Sperm competition favours larger sperm. Proceedings of the Royal Society B: Biological Sciences 266:263–267.

Lorenzi, M. C., D. Schleicherová, and G. Sella. 2006. Life history and sex allocation in the simultaneously hermaphroditic polychaete worm Ophryotrocha diadema: the role of sperm competition. Integr Comp Biol 46:381–389. Oxford Academic.

Lüpold, S., G. M. Linz, J. W. Rivers, D. F. Westneat, and T. R. Birkhead. 2009. Sperm competition selects beyond relative testes size in birds. Evolution 63:391–402.

Luther, A. 1960. Die Turbellarien Ostfennoskandiens I. Acoela, Catenulida, Macrostomida, Lecithoepitheliata, Prolecithophora, und Proseriata. Fauna Fennica 7:1–155.

Luther, A. 1947. Untersuchungen an rhabdocoelen Turbellarien VI. Macrostomiden aus Finnland. Fauna Fennica 49:1–38.

Maxfield, J. M., J. L. Van Tassell, C. M. St. Mary, J.-C. Joyeux, and K. D. Crow. 2012. Extreme gender flexibility: Using a phylogenetic framework to infer the evolution of variation in sex allocation, phylogeography, and speciation in a genus of bidirectional sex changing fishes(Lythrypnus, Gobiidae). Molecular Phylogenetics and Evolution 64:416–427.

McCartney, M. A. 1997. Sex Allocation and Male Fitness Gain in a Colonial, Hermaphroditic Marine Invertebrate. Evolution 51:127–140.

Michiels, N. K., P. H. Crowley, and N. Anthes. 2009. Accessory male investment can undermine the evolutionary stability of simultaneous hermaphroditism. Biology Letters 5:709–712. Royal Society.

Parapanov, R., S. Nusslé, J. Hausser, and P. Vogel. 2008. Relationships of basal metabolic rate, relative testis size and cycle length of spermatogenesis in shrews (Mammalia, Soricidae). Reprod. Fertil. Dev. 20:431.

Parker, G. A. 1990. Sperm competition games: raffles and roles. Proceedings of the Royal Society of London. Series B: Biological Sciences 242:120–126. Royal Society.

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

Patlar, B., M. Weber, T. Temizyürek, and S. A. Ramm. 2020. Seminal fluid-mediated manipulation of post-mating behavior in a simultaneous hermaphrodite. Current Biology 30:143–149.

Peirce, E. J., and W. G. Breed. 2001. A comparative study of sperm production in two species of Australian arid zone rodents (Pseudomys australis, Notomys alexis) with marked differences in testis size. Reproduction 121:239–247.

Pen, I., and F. J. Weissing. 1999. Sperm competition and sex allocation in simultaneous hermaphrodites: A new look at Charnov's invariance principle. Evolutionary Ecology Research 1:517–525.

Petersen, C. W. 1991. Sex Allocation in Hermaphroditic Sea Basses. The American Naturalist 138:650–667.

Pitnick, S. 1996. Investment in Testes and the Cost of Making Long Sperm in Drosophila. The American Naturalist 148:57–80.

Pitnick, S., T. A. Markow, and G. S. Spicer. 1995. Delayed male maturity is a cost of producing large sperm in Drosophila. Proceedings of the National Academy of Sciences 92:10614–10618.

Rademaker, M. C. J., and T. J. de Jong. 1998. Effects of flower number on estimated pollen transfer in natural populations of three hermaphroditic species: an experiment with fluorescent dye. Journal of Evolutionary Biology 11:623–641.

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

Ramm, S. A., L. Schärer, J. Ehmcke, and J. Wistuba. 2014. Sperm competition and the evolution of spermatogenesis. Molecular Human Reproduction 20:1169–1179.

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

Ramm, S. A., and P. Stockley. 2010. Sperm competition and sperm length influence the rate of mammalian spermatogenesis. Biology Letters 6:219–221. Royal Society.

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

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

Rowley, A., L. Locatello, A. Kahrl, M. Rego, A. Boussard, E. Garza-Gisholt, R. M. Kempster, S. P. Collin, E. Giacomello, M. C. Follesa, C. Porcu, J. P. Evans, F. Hazin, F. Garcia-Gonzalez, T. Daly-Engel, C. Mazzoldi, and J. L. Fitzpatrick. 2019. Sexual selection and the evolution of sperm morphology in sharks. Journal of Evolutionary Biology 32:1027–1035.

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

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

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

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

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

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

Schärer, L., and I. Pen. 2013. Sex allocation and investment into pre- and post-copulatory traits in simultaneous hermaphrodites: the role of polyandry and local sperm competition. Philosophical Transactions of the Royal Society B: Biological Sciences 368:20120052–20120052.

Schärer, L., P. Sandner, and N. K. Michiels. 2005. Trade-off between male and female allocation in the simultaneously hermaphroditic flatworm Macrostomum sp. Journal of Evolutionary Biology 18:396–404.

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

Sheather, S. J., and M. C. Jones. 1991. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. Journal of the Royal Statistical Society. Series B (Methodological) 53:683–690.

Sicard, A., and M. Lenhard. 2011. The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. Ann Bot 107:1433–1443. Oxford Academic.

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

Singh, P., and L. Schärer. in preparation. Evolution of sex allocation plasticity and its predictors in a flatworm genus.

Singh, P., N. Vellnow, and L. Schärer. 2019. Variation in sex allocation plasticity in three closely related flatworm species. Ecol Evol ece3.5566.

St. Mary, C. M. 2000. Sex Allocation in *Lythrypnus* (Gobiidae): Variations on a Hermaphroditic Theme. Environmental Biology of Fishes 58:321–333.

Taylor, P. D. 1981. Intra-sex and inter-sex sibling interactions as sex ratio determinants. Nature 291:64–66.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 11:11.10.1-11.10.33.

van Velzen, E., L. Schärer, and I. Pen. 2009. The effect of cryptic female choice on sex allocation in simultaneous hermaphrodites. Proc. R. Soc. B 276:3123–3131.

Vizoso, D. B., Gunde Rieger, and Lukas Schärer. 2010. Goings-on inside a worm: functional hypotheses derived from sexual conflict thinking. Biological journal of the Linnean Society 99:370–383.

Weber, M., B. Patlar, and S. A. Ramm. 2020. Effects of two seminal fluid transcripts on post-mating behaviour in the simultaneously hermaphroditic flatworm *Macrostomum lignano*. J Evol Biol jeb.13606.

Weinzierl, R. P., K. Berthold, L. W. Beukeboom, and N. K. Michiels. 1998. Reduced male allocation in the parthenogenetic hermaphrodite *Dugesia polychroa*. Evolution 52:109–115.

Weinzierl, R. P., P. Schmidt, and N. K. Michiels. 1999. High Fecundity and Low Fertility in Parthenogenetic Planarians. Invertebrate Biology 118:87.

West, S. A. 2009. Sex allocation. Princeton University Press, Princeton.

West, S. A., and E. A. Herre. 1998. Stabilizing Selection and Variance in Fig Wasp Sex Ratios. Evolution 52:475.

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

Wolak, M. E., D. J. Fairbairn, and Y. R. Paulsen. 2012. Guidelines for estimating repeatability: Guidelines for estimating repeatability. Methods in Ecology and Evolution 3:129–137.

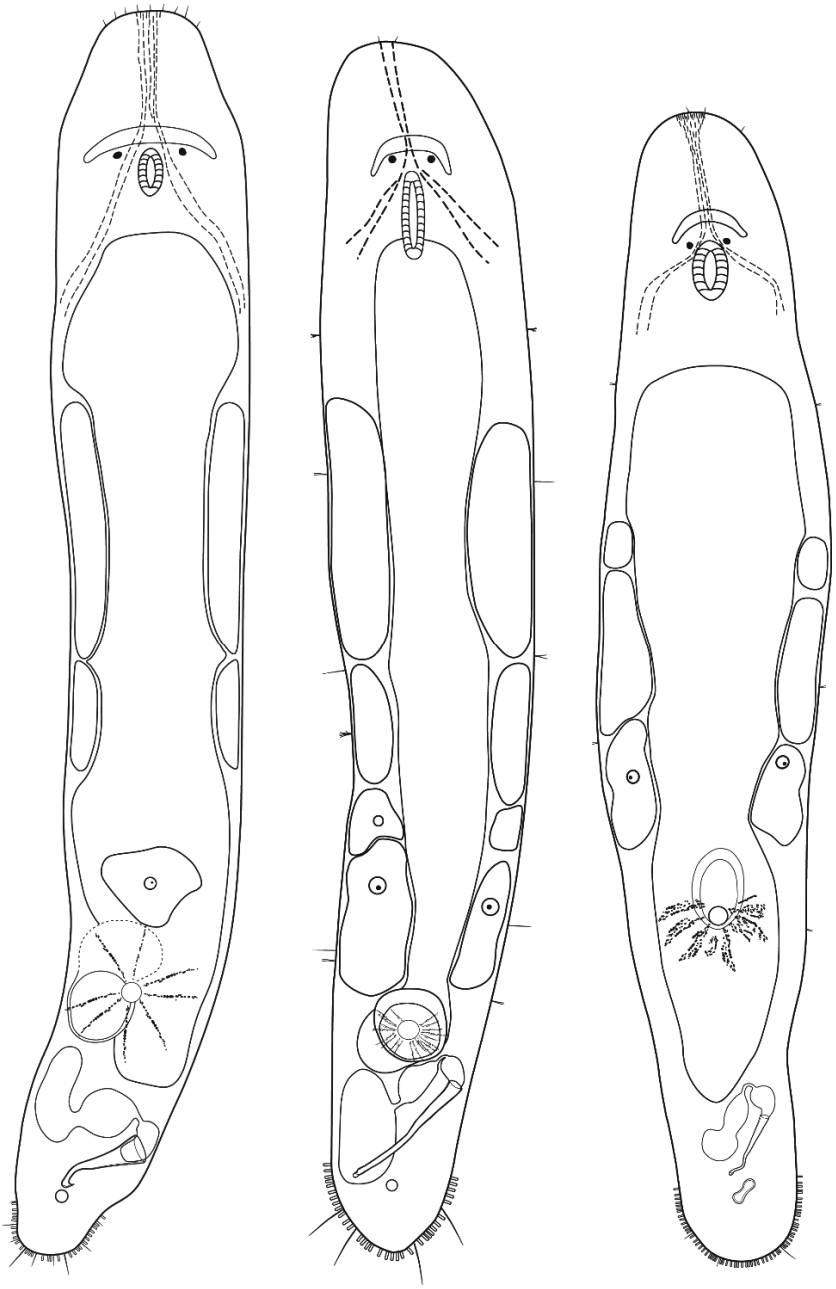
Yund, P. O. 1998. The Effect of Sperm Competition on Male Gain Curves in a Colonial Marine Invertebrate. Ecology 79:328–339.

Yund, P. O., Y. Marcum, and J. Stewart-Savage. 1997. Life-History Variation in a Colonial Ascidian: Broad-Sense Heritabilities and Tradeoffs in Allocation to Asexual Growth and Male and Female Reproduction. Biological Bulletin 192:290–299.

Yund, P. O., and M. A. McCartney. 1994. Male Reproductive Success in Sessile Invertebrates: Competition for Fertilizations. Ecology 75:2151.

Chapter V

A phylogenetically informed search for an alternative *Macrostomum* model species, with notes on taxonomy, mating, behavior, karyology, and genome size



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

Received: 27 March 2019 | Revised: 14 August 2019 | Accepted: 22 August 2019

DOI: 10.1111/jzs.12344

ORIGINAL ARTICLE

JOURNAL OF ZOOLOGICAL SYSTEMATICS AND EVOLUTIONARY RESEARCH

WILEY

A phylogenetically informed search for an alternative *Macrostomum* model species, with notes on taxonomy, mating behavior, karyology, and genome size

Lukas Schärer[1] | Jeremias N. Brand[1] | Pragya Singh[1] | Kira S. Zadesenets[2] | Claus-Peter Stelzer[3] | Gudrun Viktorin[1]

[1]Evolutionary Biology, Zoological Institute, University of Basel, Basel, Switzerland

[2]The Federal Research Center Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

[3]Research Department for Limnology, University of Innsbruck, Mondsee, Austria

Correspondence
Lukas Schärer, University of Basel, Zoological Institute, Evolutionary Biology, Vesalgasse 1, 4051 Basel, Switzerland.
Email: lukas.scharer@unibas.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 31003A-143732 and 31003A-162543; Siberian Branch, Russian Academy of Sciences, Grant/Award Number: 0324-2019-0042

Abstract

The free-living flatworm *Macrostomum lignano* is used as a model in a range of research fields—including aging, bioadhesion, stem cells, and sexual selection—culminating in the establishment of genome assemblies and transgenics. However, the *Macrostomum* community has run into a roadblock following the discovery of an unusual genome organization in *M. lignano*, which could now impair the development of additional resources and tools. Briefly, *M. lignano* has undergone a whole-genome duplication, followed by rediploidization into a 2n = 8 karyotype (distinct from the canonical 2n = 6 karyotype in the genus). Although this karyotype appears visually diploid, it is in fact a hidden tetraploid (with rarer 2n = 9 and 2n = 10 individuals being pentaploid and hexaploid, respectively). Here, we report on a phylogenetically informed search for close relatives of *M. lignano*, aimed at uncovering alternative *Macrostomum* models with the canonical karyotype and a simple genome organization. We taxonomically describe three new species: the first, *Macrostomum janickei* n. sp., is the closest known relative of *M. lignano* and shares its derived genome organization; the second, *Macrostomum mirumnovem* n. sp., has an even more unusual genome organization, with a highly variable karyotype based on a 2n = 9 base pattern; and the third, *Macrostomum cliftonensis* n. sp., does not only show the canonical 2n = 6 karyotype, but also performs well under standard laboratory culture conditions and fulfills many other requirements. *M. cliftonensis* is a viable candidate for replacing *M. lignano* as the primary *Macrostomum* model, being outcrossing and having an estimated haploid genome size of only 231 Mbp.

KEYWORDS
genome evolution, karyology, phylogeny, taxonomy, Turbellaria

1 | INTRODUCTION

The free-living flatworm *Macrostomum lignano* Ladurner, Schärer, Salvenmoser, Rieger 2005 (Rhabditophora, Macrostomorpha) has emerged as a model for a broad range of research topics (Ladurner, Schärer, Salvenmoser, & Rieger, 2005), including the biology of aging (Mouton, Grudniewska, Glazenburg, Guryev, & Berezikov, 2018; Mouton et al., 2009), bioadhesion (Lengerer, Hennebert, Flammang, Salvenmoser, & Ladurner, 2016; Lengerer et al., 2014; Wunderer et al., 2019), regeneration (Egger, Ladurner, Nimeth, Gschwentner, & Rieger, 2006; Lengerer et al., 2018), stem cell biology (Grudniewska

Jeremias N. Brand (jeremias.brand@unibas.ch); Pragya Singh (pragya.singh@unibas.ch); Kira S. Zadesenets (kira_z@bionet.nsc.ru); Claus-Peter Stelzer (claus-peter.stelzer@uibk.ac.at); Gudrun Viktorin (gudrun.viktorin@unibas.ch)

TABLE 1 Specimens analyzed for five *Macrostomum* species, *M. lignano* (Maclig), *M. janickei* n. sp. (Macjan), *M. cliftonensis* n. sp. (Maccli), *M. mirumnovem* n. sp. (Macmir), and *M. hystrix* (Machtx)

Species	Location	Origin	ID/worms	*28S* sequence	reads	*COI* sequence	Whole mount
Maclig	Lignano	LS1 (1)	MTP LS 244	FJ715326 (2)	–	KP730568 (3)	–
			6	–	3	MK690037	–
				–	3	MK690041	–
		DV1 (4)	2	–	2	MK690040	–
		DV4 (4)	2	–	2	MK690043	–
		DV6 (4)	2	–	2	MK690033	–
		DV16 (4)	2	–	2	MK690039	–
		DV18 (4)	2	–	2	MK690047 (a)	–
		DV22 (4)	2	–	2	MK690044 (a)	–
		DV47 (4)	2	–	2	MK690042	–
	Sithonia	LS3 (5)	MTP LS 2426	MK684173	–	–	–
			MTP LS 2427	MK684174	–	–	–
			MTP LS 2429	MK684175	–	–	–
			24	–	13 (6)	MK690035	–
				–	7 (7)	MK690034	–
Macjan	Palavas-les-Flots	B14 (8)	MTP LS 536	–	–	–	NMB-RHAB 00085b
			MTP LS 537	MK684168	2	MK690018	NMB-RHAB 00085a
		A1 (5)	Karyo 1	–	–	–	–
			Karyo 18	–	–	–	–
		Culture (5)	MTP LS 3212	MK684169	–	–	–
			4	–	4	MK690038	–
Maccli (9)	Lake Clifton	G7 (10)	**MTP LS 2896**	MK684170 (b)	2	MK690022	WAM V9393
			MTP LS 2906	same as (b)	2	MK690023	–
			MTP LS 2907	same as (b)	2	MK690031	WAM V9398
			MTP LS 2908	same as (b)	2	MK690029	WAM V9399
			MTP LS 2909	MK684171 (c)	2	MK690032	WAM V9400
			MTP LS 2910	–	–	–	–
			MTP LS 2911	same as (c)	0	–	–
			MTP LS 2931	same as (b)	2	MK690024	–
		G8 (11)	MTP LS 2901	same as (b)	2	MK690045	–
			MTP LS 2920	same as (c)	2	MK690030	WAM V9404
		G9 (12)	MTP LS 2903	same as (b)	2	MK690036	–
			MTP LS 2945	(13)	2	MK690021	WAM V9409
		G11 (14)	MTP LS 2900	same as (c)	2	MK690046	WAM V9396
			MTP LS 2930	same as (b)	0	–	–
Macmir (15)	Port Phillip Bay	A8 (16)	MTP LS 2993	–	–	–	NMW F258455
			MTP LS 2994	MK684172 (d)	2	MK690019	–
			MTP LS 2995	–	–	–	NMW F258456
			MTP LS 2996	–	–	–	–
		A13 (16)	MTP LS 3012	–	–	–	NMW F258461
			MTP LS 3014	–	–	–	NMW F258462
			MTP LS 3015	–	–	–	–
			MTP LS 3016	–	–	–	–
		C21 (17)	MTP LS 3019	same as (d)	2	MK690026	–
		D19 (17)	MTP LS 3147	same as (d)	2	MK690027	NMW F258509
		F11 (18)	MTP LS 3168	same as (d)	2	MK690025	NMW F258524
		E6 (19)	MTP LS 3498	same as (d)	2	MK690028	–

(Continues)

TABLE 1 (Continued)

Species	Location	Origin	ID/worms	*28S* sequence	reads	*COI* sequence	Whole mount
M. sp.	Lake Charra	–	MCZ DNA106151	KC869843 (20)	–	–	–
Machtx	Bibione	C12-14 (2)	MTP LS 68	FJ715323 (2)	–	KP730561 (3)	–
	San Rossore	SR1 (21)	MTP LS T8	–	2	MK690020	–

Notes: The table lists short species name, collection location, the laboratory culture or field sample that the worms originate from, specimen ID (bold if HOLOTYPE) or number of worms analyzed from laboratory cultures, GenBank accession number of partial *28S rRNA* gene sequences, the number of reads contributing to the *COI* gene sequence, GenBank accession number of partial *COI* gene sequences, and museum accession number of the stylet whole-mount permanent preparations (NMB, Natural History Museum Basel; WAM, Western Australian Museum; and NMW, Museum Victoria). Note that some specimens were not sequenced and that identical sequences of the same species are marked by the same bracketed letters and represented by the underlined specimen in the molecular analysis. Also note that the bracketed numbers refer to the footnotes. All newly deposited specimens are available on Zenodo at https://zenodo.org/record/2602479.

(1) Marie-Orleach et al. (2013).
(2) Schärer et al. (2011); specimens also available on Zenodo at https://zenodo.org/record/2581116.
(3) Janssen et al. (2015); specimens also available on Zenodo at https://zenodo.org/record/2580820; note that the originally deposited sequence KP730561 has now been corrected (https://www.ncbi.nlm.nih.gov/nuccore/KP730561.2), that is, the previously inserted N at the site of the frameshift deletion has been removed, based on the results presented here.
(4) Vellnow et al. (2017).
(5) Zadesenets et al. (2016).
(6) Two sequences differ in one base each (with both being silent third site mutations).
(7) Two sequences differ in one base each (with both being silent third site mutations).
(8) **Type locality:** sample B14 collected on 17 January 2009 from moist sand in the upper intertidal of a sheltered beach of a brackish water lagoon (at 10‰ salinity) near Palavas-les-Flots, France (N 43.50079, E 3.87226).
(9) PCR with adapted primers and sequencing with M13 primers; see also Table S1.
(10) **Type locality:** sample G7 collected on 17 January 2017 from sandy shore with reduced salinity water (35‰) seeping into a puddle on the shores of the increasingly hypersaline Lake Clifton (97‰ at time of sampling), Western Australia (S 32.76125, E 115.66019).
(11) Sample G8 collected on 17 January 2017 from sandy substrate with reduced salinity water (32‰) seeping out next to vegetation close to the type locality (S 32.76127, E115.66033)
(12) Sample G9 collected on 17 January 2017 from sandy substrate with reduced salinity water (65‰) seeping out next to vegetation close to the type locality (S 32.76111, E 115.66028)
(13) This sequence was not considered for analysis because it has ambiguities at the polymorphic sites
(14) Sample G11 collected on 17 January 2017 from sandy substrate with reduced salinity water (23‰) seeping out in between vegetation close to the type locality (S 32.76063, E 115.66037)
(15) PCR with adapted primers and sequencing with M13 primers; see also Table S1.
(16) **Type locality:** sample A1 collected on 28 January 2017 from the upper intertidal on a sheltered beach (at 35‰ salinity) in front of the Victorian Marine Science Consortium, Queenscliff, Port Phillip Bay, Victoria (S 38.27007, E 144.63894), and sample A13 collected from same site on 29 January 2017.
(17) Sample C21 collected on 30 January 2017 and sample D19 collected on 1 February 2017 from the upper intertidal under vegetation on a sheltered beach (at 35‰ salinity) in Edwards Point Wildlife Reserve, Port Phillip Bay, Victoria (S 38.21880, E 144.70047).
(18) Sample F11 collected on 8 February 2017 from the upper intertidal on a sheltered and drying pond (at 47‰ salinity) on Mud Island, Port Phillip Bay, Victoria (S 38.26964, E 144.76346).
(19) From a laboratory culture established from specimens collected in sample E6 on 6 February 2017 from a drainage ditch of Swan Bay Salt marsh, Queenscliff, Victoria (S 38.26503, E 144.62098).
(20) Laumer & Giribet, (2014).
(21) Ramm et al. (2015), Zadesenets et al. (2016).

et al., 2016; Ladurner et al., 2008), and sexual selection (Marie-Orleach, Janicke, Vizoso, David, & Schärer, 2016; Schärer, Littlewood, Waeschenbach, Yoshida, & Vizoso, 2011; Sekii et al., 2013). This research has led to the establishment of many resources and tools that are crucial for a modern genetic and genomic model, including in situ hybridization (Pfister et al., 2008), RNA interference (Kuales et al., 2011; Sekii et al., 2013), gene expression (Arbore et al., 2015; Grudniewska et al., 2016; Lengerer et al., 2018), genome and transcriptome assemblies (Wasik et al., 2015; Wudarski et al., 2017), transgenesis (Marie-Orleach, Janicke, Vizoso, Eichmann, & Schärer, 2014; Wudarski et al., 2017), and a clarification of the phylogenetic context (Janssen et al., 2015; Schärer et al., 2011). Jointly, these achievements make *M. lignano* an excellent model to complement the, in several aspects, more established planarian flatworm models (Newmark & Sánchez Alvarado, 2002; Pellettieri & Sánchez Alvarado, 2007; Rink, 2013; Rouhana et al., 2013). This is in part linked to the high transparency of *Macrostomum* flatworms, which is particularly powerful in combination with transgenesis tools (Marie-Orleach et al., 2016; Wudarski et al., 2017), which so far have proven difficult to establish in planarian flatworms, possibly due to their unusual embryology (Cardona, Hartenstein, & Romero, 2005).

In spite of its general appeal as a flatworm model, the further establishment of *M. lignano* as a broadly employed genetic and genomic model has recently run into a roadblock. Such models should ideally have small and stable genomes, facilitating the establishment of (a) highly contiguous genome assemblies, (b)

high-resolution genetic maps for forward genetics using quantitative trait loci (QTL) and genome-wide association studies (GWAS; Bazakos, Hanemian, Trontin, Jiménez-Gómez, & Loudet, 2017), and (c) efficient genome editing for reverse genetics (e.g., CRISPR/Cas9; Brooks & Gaj, 2018). In this light, the recent discovery that *M. lignano* has an unusual genome organization is problematic. Earlier work had suggested that *M. lignano* has a 2n = 8 karyotype (with two large and six small chromosomes), while most other *Macrostomum* species are 2n = 6 (with six small chromosomes), or more rarely 2n = 12 (with twelve small chromosomes) (Egger & Ishida, 2005). Subsequent karyological analyses have confirmed this unusual karyotype and also revealed that individual *M. lignano* can deviate from the 2n = 8 karyotype and that such variation segregates within freshly field-collected worms, outbred laboratory cultures, and also inbred lines (Zadesenets et al., 2016). Most of this variation involves the presence of either 2, 3, or 4 copies of the large chromosome (yielding 2n = 8, 2n = 9, and 2n = 10 individuals, respectively), but individuals with additional karyotypes (including some with chromosomal rearrangements) also occur at low frequency (Zadesenets et al., 2016). More recent analyses have further shown that the large chromosome represents a fusion product of one full haploid chromosome set, making the 2n = 8, 2n = 9, and 2n = 10 karyotypes effectively hidden tetra-, penta-, and hexaploids, respectively. It is likely that this situation emerged from a recent whole-genome duplication that was followed by a rediploidization into a 2n = 8 karyotype (Zadesenets, Ershov, Berezikov, & Rubtsov, 2017a; Zadesenets, Schärer, & Rubtsov, 2017b). These recent insights—and the complex genome organization they reveal—clearly dampen the prospects of *M. lignano* becoming a full-fledged genetic and genomic model.

Here, we present efforts to find alternatives within the species-rich *Macrostomum* genus, which could supplement or even replace *M. lignano* as the primary model. We have collected several new and currently undescribed *Macrostomum* species, which—as we show below—are closely related to *M. lignano*. The new species include *Macrostomum janickei* Schärer n. sp., *Macrostomum cliftonensis* Schärer and Brand n. sp., and *Macrostomum mirumnovem* Schärer and Brand n. sp. We describe these species taxonomically, covering both anatomical and behavioral aspects, place them phylogenetically using molecular analyses, and—by determining their karyotypes and genome sizes—evaluate their suitability as genetic and genomic models.

2 | MATERIALS AND METHODS

2.1 | Collection, documentation, preservation, and specimen deposition

Details on species and specimens included are given in Table 1. Briefly, *M. lignano* was collected in the Northern Adriatic Sea (including the type locality) and the Aegean Sea (Ladurner et al., 2005; Zadesenets et al., 2016), and *M. janickei* was collected in the Gulf of Lion (Zadesenets et al., 2016; called *Macrostomum* sp. 8 therein). Long-term laboratory cultures have been established from these species, including genetically outbred cultures (of both *M. lignano* and *M. janickei*) and inbred lines (of *M. lignano*) (Vellnow, Vizoso, Viktorin, & Schärer, 2017; Zadesenets et al., 2016). Furthermore, we have collected *M. cliftonensis* from the shores of Lake Clifton, an increasingly hypersaline coastal lake South of Perth (Western Australia; Lane, Clarke, & Winchcombe, 2017), as well as *M. mirumnovem* from several sites in Port Philipp Bay near Queenscliff (Victoria). Long-term outbred laboratory cultures of both species have been established. As the outgroup for the molecular analyses (see below) we use *Macrostomum hystrix* Ørsted 1843, collected from two locations in Italy, a species previously shown to be genetically close to *M. lignano* (Schärer et al., 2011).

Specimens were either collected directly from the field or taken from laboratory cultures, and many specimens were extensively documented morphologically using digital photomicrography; when possible, we also prepared whole-mount permanent preparations of the stylet, the male intromittent organ (see Janssen et al., 2015; Schärer et al., 2011). Whole worms or fragments were then preserved in either absolute ethanol (stored cool for at most a few weeks and then at −20°C) or RNAlater (stored at 4°C for at most a few weeks and then at −80°C). The studied specimens were then deposited on the Macrostomorpha Taxonomy and Phylogeny website (at http://macrostomorpha.info), the images and videos were deposited in the open access repository Zenodo (at https://doi.org/10.5281/zenodo.2602479), and the whole-mount permanent preparations were deposited in various museums (Table 1).

2.2 | Image processing

The images used in the species description panels were processed in Photoshop (version 2017.1.1), including rotation and cropping from the original digital images (filling in blank spaces with the approximate background color), and contrast enhancement using the "Levels" function. Moreover, overview images were stitched manually (including slight adjustments of the aspect ratios to compensate for worm movement) or using the "Photomerge" function (using the reposition function), and these are labeled "stitched" in the panels. Finally, some structures are visualized using the "Auto-Blend Layers" function, where different focal planes (in our case a video focusing through the structure of interest) are synthesized into a single compound image, and these are labeled "merged" in the panels. While such images facilitate visualization of structures in different focal planes, they can be somewhat misleading in that they can hide curvature in the z-axis.

2.3 | Morphometrics

Morphometric analyses (see Table 2) were performed using the software ImageJ (version 1.51w) and the plugin ObjectJ (version 1.04r), which allows marking structures in the original images in a non-destructive manner. The pixel length of structures was converted into µm by calibration using a stage micrometer. Body length was measured by placing a segmented line along the central body axis. Body width was measured by placing a line perpendicular to the body axis

at the height of the testes. Testis and ovary lengths were measured as the longest distance of the structure parallel to the body axis. Eye diameter, sensory cilia length, rhabdite granule length, straight stylet length, the width of the stylet openings, sperm bristle length, and sperm brush length were measured by placing a straight line along the structures. To measure segmented stylet length, we placed a segmented line along both sides of the stylet and averaged the length. For sperm length, we placed a segmented line along the central axis of the sperm. We defined sperm feeler length and sperm body length as previously described (Janicke & Schärer, 2010).

2.4 | Mating behavior

We analyzed mating behavior using established techniques (Marie-Orleach, Janicke, & Schärer, 2013; Schärer, Joss, & Sandner, 2004). Briefly, worms were paired and placed in small drops between two glass slides with spacers, then filmed in the resulting mating chamber using time-lapse video (1 frame per second), and behaviors scored from the resulting video sequences. Here, we present only a brief characterization of the behaviors of these species based on a single mating pair and compare these observations to the behaviors observed in *M. lignano* (Schärer et al., 2004). Briefly, *M. lignano* shows a reciprocal mating behavior, where partners first reciprocally engage in a precopulatory behavior called *circling*, during which the worms crawl around on each other. This then often leads to a characteristic copulatory posture that resembles two interlocking "G" characters forming a tight disk, during which the worms reciprocally insert their stylet into the partner's female antrum via the female gonopore and during which the anterior ventral surface of each worm is in contact with the partner's posterior dorsal surface. Copulations on average last for 8.8 s (range: 5–16 s). Once the copulation ends, the worms often show a stereotypical behavior called *sucking*, during which a worm places its mouth opening over its own female gonopore, presumably in an attempt to remove ejaculate components from the female antrum (Schärer et al., 2004, 2011). This *sucking* behavior on average lasts 4.9 s (range: 4–7 s) and is performed, generally within 5 s, by either none, one, or both worms after about a third of the copulations, respectively.

The paired *M. janickei* adults were isolated as hatchlings from the laboratory culture (and were thus virgins), the paired *M. cliftonensis* were field-collected worms from the type locality, and the paired *M. mirumnovem* were isolated for over a week after being taken from the laboratory culture (isolating worms prior to pairing can increase the likelihood of observing matings). More detailed analyses of the mating behaviors of the new species are underway (P. Singh, D. Ballmer, M. Laubscher, & L. Schärer, unpublished data).

2.5 | Karyology and genome size

Metaphase plates were prepared from individual worms of *M. janickei*, *M. cliftonensis*, *M. mirumnovem*, and *M. hystrix*, using the single-worm karyotyping technique described previously (Zadesenets et al., 2016). Chromosomes were counterstained with VECTASHIELD Antifade Mounting Medium with DAPI (Vector Laboratories). Specimens of *M. janickei* were taken from natural populations, while specimens of *M. cliftonensis*, *M. mirumnovem*, and *M. hystrix* were taken from outbred laboratory cultures (Table 1). In total, we always analyzed ≥10 metaphase spreads per individual. Microscopic analysis of chromosome slides was done using a CCD-camera installed on an Axioplan 2 compound microscope (Carl Zeiss) equipped with filter cube #49 (ZEISS) using ISIS4 software (MetaSystems GmbH) at the Inter-institutional Shared Center for Microscopic Analysis of Biological Objects (Institute of Cytology and Genetics SB RAS).

To obtain genome size estimates from the studied *Macrostomum* species, we used a detergent–trypsin method with propidium iodide (PI) staining for flow cytometric analysis of genome size (Stelzer, Riss, & Stadler, 2011). Briefly, for each replicate 7–20 starved worms were washed in few mL of stock solution (3.4 mM Trisodium citrate dihydrate, Nonidet P-40 at 0.1% v/v, 1.5 mM Spermine tetrahydrochloride, 0.5 mM Tris(hydroxymethyl)-aminomethane, pH 7.6) and then transferred to 750 µl of stock solution in a 1 ml Dounce tissue homogenizer. Worms were homogenized on ice with 20 strokes using the "tight" pestle of the homogenizer. As an internal standard of known genome size, we used the fruit fly, *Drosophila melanogaster* (strain ISO-1, diploid nuclear DNA content: 0.35 pg; Gregory, 2019). After adding two female *Drosophila* heads, the sample was further homogenized with ten strokes. Large debris was removed by filtration through a 40 µm mesh nylon sieve. After addition of 100 µl of 0.021% Trypsin (dissolved in stock solution), the sample was incubated for exactly 10 min at 37°C. To prevent further degradation, 75 µl of 0.25% trypsin inhibitor was added (this solution also included 0.05% RNase A) and the samples were incubated for another 10 min at 37°C. Finally, samples were stained with propidium iodide at a concentration of 50µg/ml and kept overnight on ice in the dark. Flow cytometric analysis was performed on the next day on an Attune NxT acoustic focusing cytometer (ThermoFisher) with an excitation wavelength of 561 nm (yellow) and a custom-made 590–650 nm bandpass filter for detection of PI fluorescence. Flow cytometric data were analyzed using FlowJo software version 10.0.7r2 (FlowJo LLC). Flow cytometry events were gated by YL2A versus YL2H for doublet exclusion. Coefficients of variance (CVs) of individual peaks typically ranged between 1.5% and 6% for both *Drosophila* and worms. Very few measurements had CVs higher than 6%, and those replicates were discarded. Conversion from picograms DNA to base pairs were made with the factor: 1 pg = 978 Mbp (Gregory, 2019).

2.6 | Molecular phylogenetic placement

To determine the molecular phylogenetic placement, we initially used a partial (nuclear) *28S ribosomal RNA* (*28S rRNA*) gene sequence, a marker previously used in *Macrostomum* (Schärer et al., 2011). However, as outlined below, this marker was too conserved to successfully resolve interrelationships between *M. lignano* and *M. janickei*. We therefore also used a partial (mitochondrial) *cytochrome c oxidase I* (*COI*) gene sequence, a more rapidly evolving marker

TABLE 2 Morphometric measurements based on image and video material from the deposited specimens of the three new *Macrostomum* species, *M. janickei* n. sp., *M. cliftonensis* n. sp. and *M. mirumnovem* n. sp.; L: length, W: width

	M. janickei n. sp.				*M. cliftonensis* n. sp.				*M. mirumnovem* n. sp.			
Parameter	Mean (µm)	*SD* (µm)	range (µm)	*n*	Mean (µm)	*SD* (µm)	range (µm)	*n*	Mean (µm)	*SD* (µm)	range (µm)	*n*
Body L	1,153	430	701–1731	4	1,220	312	709–1762	14	1,170	333	670–1667	12
Body W	339	151	143–496	4	324	81	214–456	14	329	94.6	169–441	12
Eye ø	8.3	1.9	6.9–9.6	2	10.1	1.7	7.2–12.6	11	9.2	1.2	7.2–10.1	5
Testis L	224	8.7	217–230	2	221	50.8	167–328	13	64.8	16.1	40.3–94.1	11
Ovary L	126	64.2	80.8–172	2	120	46.6	58.3–221	12	140	51.9	73.6–227	9
Short sensory cilia L	10.2	1.7	8.8–12.2	3	10	1.4	7.3–12.3	8	7.5	1.9	4.4–9.3	6
Long sensory cilia L	14.6	-	-	1	26	8.6	15.4–40.7	7	11.2	0.4	11–11.6	3
Rhabdite L	7.2	2.1	4.8–9.4	4	6.9	1.2	5.3–10.1	12	6.1	0.4	5.2–6.6	10
Stylet traits												
Straight L	72	5.9	67.8–76.2	2	123	10.7	98.5–132	10	83.3	9.6	60.4–93.4	9
Segmented L	69.1	10.3	57.5–77.0	3	126	8.5	111–140	11	89.8	7.2	78–101	10
Proximal opening W	21.9	8.3	13.9–30.5	3	19.4	4.7	13.9–30	11	13.3	2.7	10.4–19.9	10
Distal opening W	6.1	1.9	4.9–8.4	3	7.9	2.2	3.9–11.5	11	4.1	1.4	1.8–6.6	9
Sperm traits												
Total L	81.1	2.1	79.6–82.6	2	103	2.8	98.1–106	5	86	5.8	79.9–92.5	6
Feeler L	22.7	5.7	18.7–26.8	2	24.4	3	20.1–28.1	5	32.5	5.7	26.2–41	5
Body L	11.6	4.1	8.7–14.5	2	14.5	0.8	13.8–15.6	5	13	1.1	11.1–14.2	5
Bristle L	13.6	0.9	13–14.3	2	12.3	0.5	11.8–12.8	5	11.5	0.6	10.5–12.2	6
Brush L	4.8	0.2	4.6–4.9	2	3.8	0.4	3.3–4.1	4	3.4	1.2	2.1–4.6	5

previously used in the Macrostomorpha (Janssen et al., 2015). Moreover, to better understand the intraspecific phylogenetic structure, we sequenced the partial *COI* gene in multiple specimens per species, including several *M. lignano* laboratory lines and cultures, as well as field-collected specimens. As an outgroup, we used *M. hystrix*, a species that is closely related to *M. lignano* (Schärer et al., 2011).

DNA was extracted either, for worms stored in ethanol, using the DNeasy Blood and Tissue kit (Qiagen, Germany) following evaporation of ethanol, or, for worms stored in RNAlater, using the NucleoSpin RNA XS kit with the combined RNA/DNA buffer set (Macherey-Nagel, Germany). Extracted DNA samples were stored at −80°C until used.

We used the ZX-1 and 1500R primers to amplify the partial *28S rRNA* gene sequence (Table S1; Schärer et al., 2011), yielding PCR fragments of 1,177 bp (with respect to the *M. lignano* sequence) and Sanger sequenced the resulting fragments in both directions using the same primers (Microsynth, Switzerland) (and in one case additional internal sequencing primers, 300F and ECD2; Schärer et al., 2011). Cycling conditions were either as in Schärer et al. (2011), or using Q5 Hot Start polymerase (NEB), 98°C for 30 s, 7 touchdown cycles with denaturation for 7 s at 98°C, annealing for 30 s with temperatures starting at 70°C and decreasing 1°C each cycle, elongation at 72°C for 20 s, followed by 28 more cycles with 64°C annealing temperature and a final elongation for 2 min at 72°C. All reads were subjected to BLAST searches (http://www.ncbi.nlm.nih.gov) to check for possible contaminations, then de novo assembled into consensus sequences in Geneious (11.1.4, Biomatters) using the built-in assembler (with the highest sensitivity). The resulting assemblies were visually checked, manually trimmed and edited where necessary. For all in-group species, multiple specimens were sequenced and one representative specimen was chosen for each unique consensus sequence (Table 1). The resulting sequences were aligned in Geneious using MAFFT (v7.388), with default settings (using FFT-NS-i x1000). The alignment was trimmed to a uniform length of 1,040 bp (see Alignment S1), and then, a consensus tree was generated using Geneious Tree Builder (using Tamura-Nei Neighbor Joining, with 1,000 bootstrap samples), and *M. hystrix* was used to root the tree. As outlined below, the tree topology suggested that this partial *28S rRNA* gene sequence is not variable enough to resolve the interrelationships between *M. lignano* and *M. janickei*. More elaborate analyses were therefore not considered and effort instead focused on the *COI* gene, which we describe next.

We used a range of primers to amplify the partial *COI* gene (Table S1). Unless otherwise stated, we used Mac_COIF and Mac_COIR (Janssen et al., 2015), optimized using a *M. lignano* mitochondrial genome (NCBI Reference Sequence: NC_035255.1; Egger, Bachmann, & Fromm, 2017), yielding a PCR fragment of 709 bp. Alternatively, we used PCR primers optimized using whole-body transcriptome data of *M. cliftonensis* and *M. mirumnovem* (Brand et al., unpublished data), respectively, yielding PCR fragments of 1,125 bp (excluding the M13 tails). Cycling conditions for these primers were analogous to the *28S rRNA* gene, except a duration of 40 s was used for all elongations. Moreover, we either used 7 touchdown cycles with annealing temperatures decreasing from 70 to 59°C, followed by 28 cycles using 59°C as annealing temperature, giving a total of 35 cycles. Or we used 11 touchdown cycles with 2 min annealing time and annealing temperatures decreasing from 68 to 56°C followed by nine cycles with annealing for 1 min at 56°C and 20 cycles with annealing for 30 s at 64°C, giving a total of 40 cycles. The resulting fragments were generally Sanger sequenced in both directions using the PCR primers (or using the M13 tails of the primers for *M. cliftonensis* and *M. mirumnovem*). For the Mac_COIF/Mac_COIR primers, the quality of the resulting sequences was somewhat variable, particularly for the reverse primer. For these sequences, we therefore, in some cases, decided to de novo assemble multiple specimens, which in most cases yielded well-supported consensus sequences (Table 1). Note that the different DV lines of *M. lignano* represent inbred lines initiated from a single mother, which are therefore expected to have just one mitochondrial cytotype within each line (see also Vellnow et al., 2017), which was also confirmed by our *COI* gene sequences.

For alignment, these protein-coding sequences were translated with the echinoderm and flatworm mitochondrial code (transl_table = 9; Telford, Herniou, Russell, & Littlewood, 2000), using the reading frame of the annotated *COI* gene in the *M. lignano* mitochondrial genome (Egger et al., 2017). This revealed that both of the analyzed *M. hystrix* specimens showed a single-base deletion at position 336 (counting from the ATT start codon of the 1,548 bp *COI* gene), leading to a frameshift mutation and a TAG stop codon only 10 bases on (and 8 more stop codons in the sequenced fragment alone), which one would expect to lead to truncation the resulting COI protein. Given the importance of the COI protein in the electron transport chain of mitochondrial oxidative phosphorylation, this finding was surprising and required further validation (see Section 3).

For phylogenetic analysis, we inserted an "N" character in place of the deletion, as was also done for the originally deposited *COI* gene sequence of *M. hystrix* (KP730561 in Janssen et al., 2015), so that the alignment, which was trimmed to a uniform length of 573 bp (Alignment S2), generated a correct protein-coding sequence in all specimens. From this alignment, a maximum likelihood tree was calculated with IQ-TREE (multicore version 1.6.10; Nguyen, Schmidt, Haeseler, & Minh, 2015). We first ran ModelFinder (Kalyaanamoorthy, Minh, Wong, von Haeseler, & Jermiin, L. S., 2017), which identified HKY + F + G4 as the best fitting substitution model (BIC = 4,906.3), and we then ran a tree reconstruction with 500 non-parametric bootstrap replicates.

2.7 | Nomenclatural acts

The electronic edition of this article conforms to the requirements of the amended International Code of Zoological Nomenclature (ICZN), and hence, the new names contained herein are available under that code from the electronic 'Early View' edition of this article (published in 2019).

This published work and the nomenclatural acts it contains have been registered in ZooBank, the online registration system for the ICZN. The ZooBank LSIDs (Life Science Identifiers) can be resolved and the associated information viewed through any standard web browser by appending the LSID to the prefix "http://zoobank.org/". The LSID for this publication is: urn:lsid:zoobank.org:pub:4CDD54C7-B44E-4070-8D8F-92C170C3EF4E. The journal's Online ISSN is 1439-0469 and it is archived in CLOCKSS. Note that these names have also appeared in the electronic 'Early View' edition of another article from our group (Singh, Vellnow, & Schärer, 2019), but since that work was not intended to conform to the ICZN requirements, having an experimental rather than taxonomic focus, these names cannot be considered available from that work.

3 | RESULTS AND DISCUSSION

Species descriptions

Family Macrostomidae Van Beneden, 1870

Genus *Macrostomum* Schmidt, 1848

Macrostomum janickei Schärer, n. sp.

urn:lsid:zoobank.org:act:56A75401-E27A-4814-8176-B34F5CC3B714 (Figures 1–3; Figure 10a, b; Figure 11a, b; Table 1, Table 2)

Material examined. Live observations on four field-collected specimens and several specimens from laboratory cultures. Holotype:

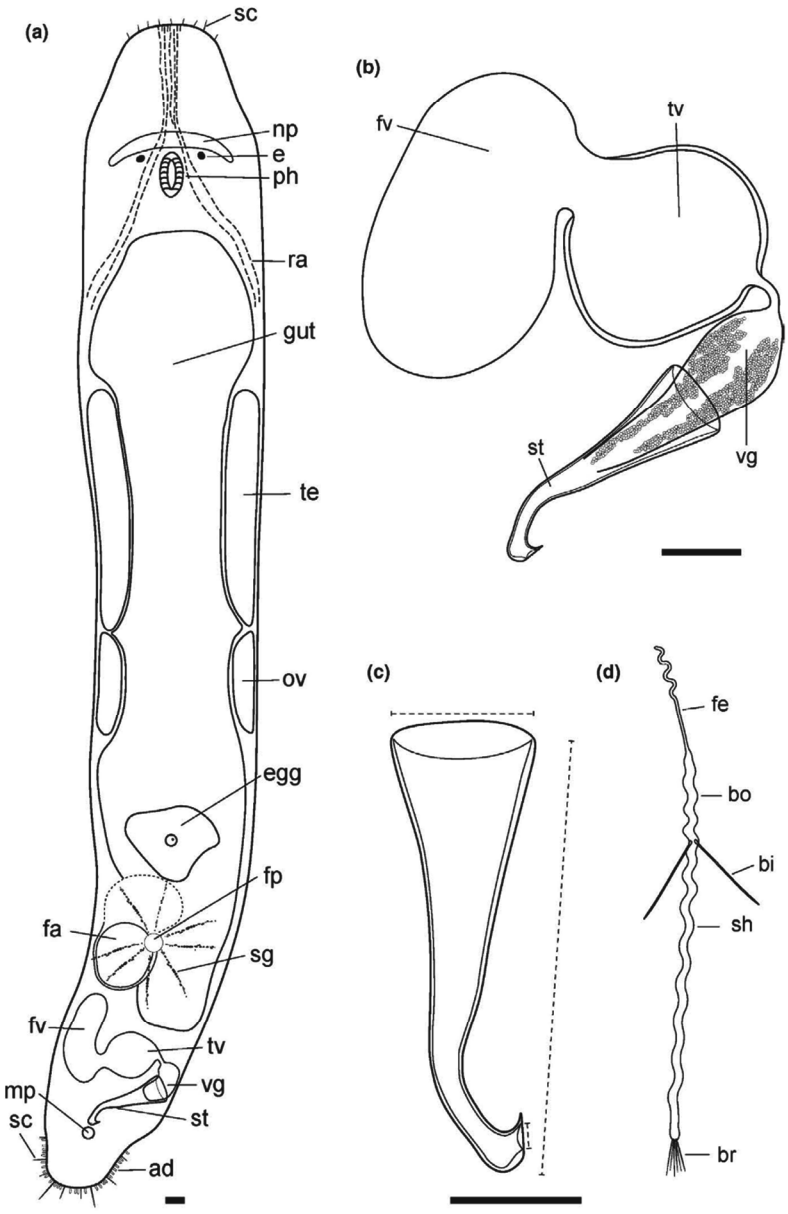


FIGURE 1 Line drawings of *Macrostomum janickei* n. sp.: (a) Habitus (viewed from dorsal and lightly squeezed), with sensory cilia (sc), neuropile (np), eyes (e), pharynx (ph), rhammites (ra), testis (te), ovary (ov), female gonopore (fp), female antrum (fa), shell glands (sg), male gonopore (mp), and adhesive glands (ad). (b) Male genital system with (non-muscular) false seminal vesicle (fv), muscular true seminal vesicle (tv), vesicula granulorum (vg), and stylet (st). (c) Detailed drawing of the stylet (indicated are the measurements for the straight stylet length and the width of the proximal and distal stylet openings, while the segmented stylet length corresponds to the average length of both sides of the stylet). (d) Mature sperm cell with feeler (fe), body (bo), bristles (bi), shaft (sh), and brush (br)

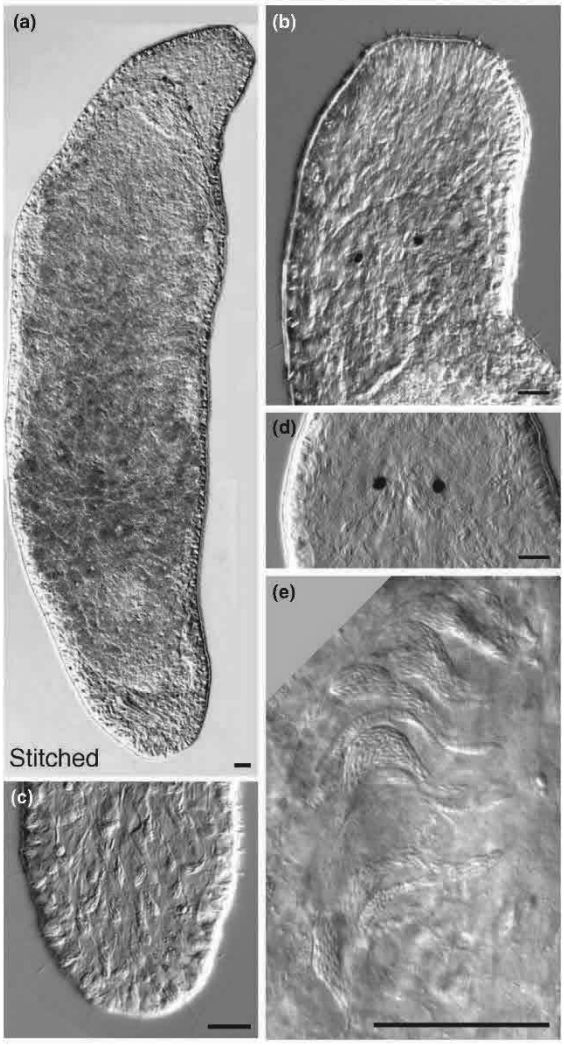


FIGURE 2 Micrographs of somatic structures of *Macrostomum janickei* n. sp. collected from the field (the bracket denotes the code of the deposited specimen, with MTP LS 537 being the HOLOTYPE; note that all specimens, except Karyo 18, were imaged from the dorsal side). (a) Overview of an adult, lightly squeezed, worm (Karyo 18, manually stitched from 3 images); (b) head region with neuropile, eyes, and pharynx; note the short sensory hairs on the rostrum (MTP LS 537); (c) tail region with dense rhabdite bundles and several long sensory hairs (MTP LS 537); (d) eyes with rhammite gland secretions penetrating the neuropile (MTP LS 536); (e) details of the pharynx gland necks showing two distinct pharyngeal gland secretions (MTP LS 537). Scale bars 25 µm

one extensively documented worm (MTP LS 537, 60 images and videos) from the type locality (i.e., sample B14 collected on 17 January 2009 from moist sand in the upper intertidal of a sheltered beach of a brackish water lagoon near Palavas-les-Flots, France; N 43.50079, E 3.87226), including a whole mount (NMB-RHAB 00085a) and a partial *28S rRNA* gene sequence. Paratypes: one extensively documented worm from same sample as holotype (MTP LS 536, 54 images and videos), including a whole mount (NMB-RHAB 00085b); two partly documented specimens (Karyo 1 and Karyo18; karyology failed, specimens not deposited), collected in 2014 from near the type locality; one worm from laboratory culture, established from the 2014 collections, extensively documented in 2017 (MTP LS 3212, 30 images and videos; see also Table 2); and 4 undocumented worms from laboratory culture for sequencing of the partial *COI* gene.

Etymology. Species name in honor of Tim Janicke, who was the first to culture this species during his PhD with LS, and who has assisted LS during multiple subsequent collections. To facilitate backwards compatibility with previous records and reports, note that this species has previously been referred to as *Macrostomum* sp. 8 (or Mac008 for short).

Diagnosis. *Macrostomum* with spindle-shaped body and trapezoid rostrum (Figure 1a; Figure 2a). Body lengths of field caught worms range from 701 to 1731 µm (see also Table 2). Sensory cilia are generally short and restricted to rostrum and tail plate. Two small pigment cup eyes. Gut extends caudally beyond female antrum. Testes clearly larger than ovaries. The highly distinctive stylet (~70 µm), is a long and gradually narrowing funnel that includes first a slight turn (of ~40°) and then a sharp turn (of >90°) toward the distal end (Figure 1c), leading to the distal opening being oriented laterally and obliquely cut off (Figure 1c), giving the stylet tip a hook-like appearance. Prostate granules of vesicula granulorum reach into about half the length of the stylet. The vagina is central and the female antrum—often containing received sperm—appears relatively complex and may be displaced slightly toward the left side of the body. Sperm (81 µm) show general morphology of reciprocally mating species (Schärer et al., 2011).

Description (see also Table 2)

General morphology. The body is spindle-shaped (width/length ratio: 1:3.4) and widest at the level of the anterior testes (Figure 1a; Figure 2a). The rostrum is trapezoid (Figure 2b), and the tail plate is relatively indistinct. The pigment cup eyes are small and fairly round (Figure 2d). The body is covered homogeneously with cilia, and sensory cilia are largely restricted to the rostrum and tail region, being very short and stiff on the rostrum and longer on the tail (Figure 2c). The mouth and pharynx are unremarkable (Figure 2e), and the gut extends posteriorly beyond the female antrum, ending at the seminal vesicles at 80% BL (% of body length). Rhabdite bundles are sparse, and most abundant on the dorsal side of the head and tail (Figure 2c), forming bundles of 6–12 granules. The rhammite glands originate dorsally anterior of the testes, and their granules are secreted via the neuropile into the anterior rostrum, where prominent rhammite bundles can be observed. The adhesive glands are arranged horseshoe-like in a single line along the edge of the tail plate.

Male system. The prominent testes extend from 27% to 40% BL and show developing sperm aligned in the center (Figure 3a). While the vas deferens is not easily seen, in some specimens one can see how the developing sperm appear to be aligned to enter the vas deferens, in a position where the distal side of the testis is in contact

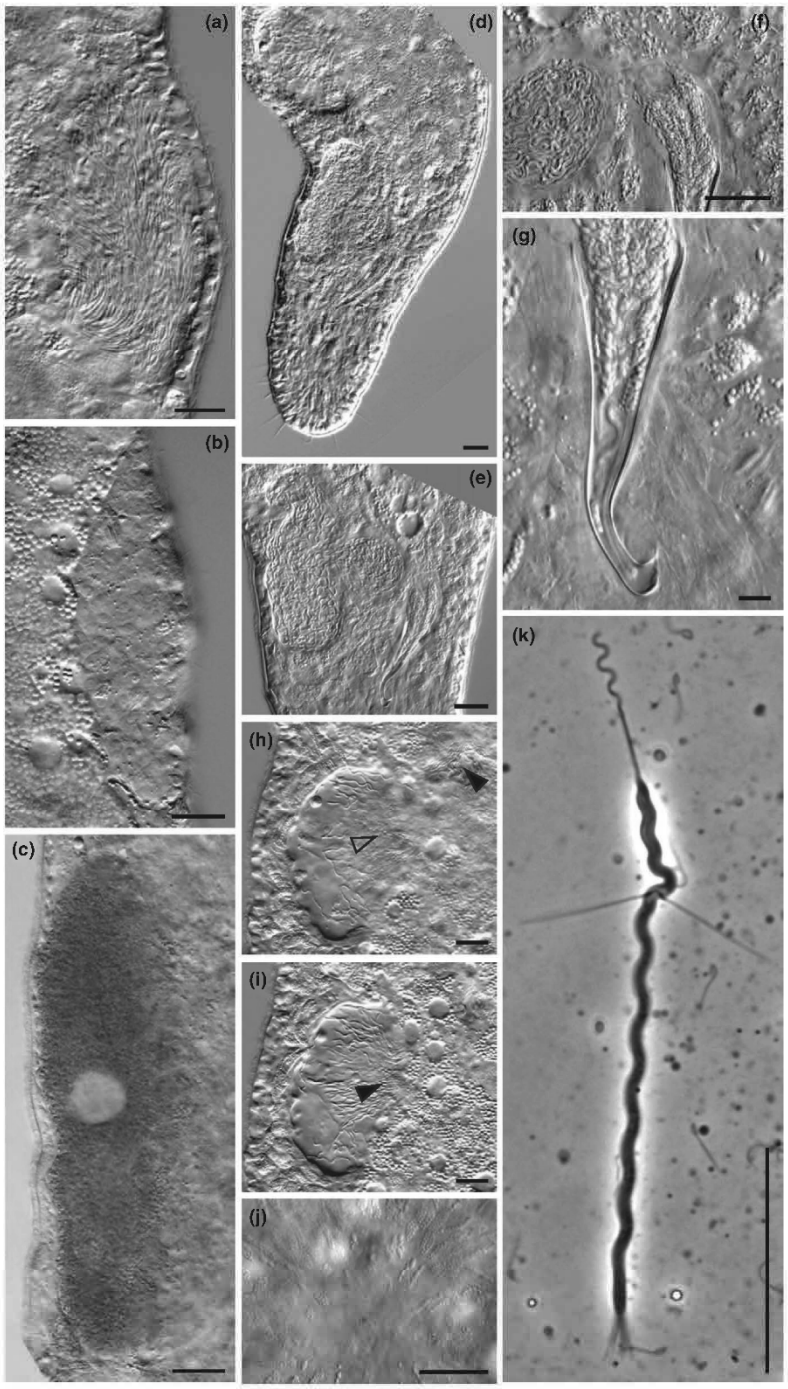


FIGURE 3 Micrographs of reproductive structures of *Macrostomum janickei* n. sp. collected from the field (the bracket denotes the code of the deposited specimen, with MTP LS 537 being the HOLOTYPE; note that all specimens, except MTP LS 2908, were imaged from the ventral side). (a) ripe testis with elongating spermatids aligned in the central region (MTP LS 536); (b) ripe ovary with oocytes that are beginning to form yolk and shell granules (MTP LS 537); (c) forming egg with the clearly visible nucleus, and many yolk and shell granules (Karyo 1); (d) body region containing the female antrum, the false and true seminal vesicle, vesicula granulorum, and stylet; note the long sensory hairs on the tail plate (MTP LS 537); (e) false and true seminal vesicle, vesicula granulorum, and stylet (MTP LS 537); (f) muscular true seminal vesicle, ductus intervesicularis, vesicula granulorum, and base of the stylet (MTP LS 537, frame extracted from a movie); (g) stylet showing the sharply turning distal end and prostate gland necks reaching far into the stylet (MTP LS 537, frame extracted from a movie); (h) female antrum with received sperm and coagulated seminal fluid; note the ciliary tuft near the center (open arrowhead) and the anchored bundle of sperm reaching the upper right corner (closed arrowhead) (MTP LS 537, frame extracted from a movie). (i) Female antrum with received sperm and coagulated seminal fluid; note the anchored bundle of sperm right of center (closed arrowhead) in the region of the ciliary tuft (MTP LS 537, frame extracted from a movie). (j) Vagina region surrounded by shell glands (MTP LS 537, frame extracted from a movie). (k) Sperm with typical feeler, body, bristles, shaft, and brush (MTP LS 536). Scale bars 25 μm

with the gut region. There is a prominent (non-muscular) false seminal vesicle that lies on the left of the tail plate (Figures 1b, 3d), connecting to a muscular true seminal vesicle that lies quite centrally (Figure 3e). The true seminal vesicle is connected via a short ductus intervesicularis to the prominent and muscular vesicula granulorum (Figure 3f) that sits on the proximal stylet opening. The tail plate contains many prostate gland cells that send out their necks into the vesicula granulorum, from which prostate glands necks reach to about half the length of the stylet (Figures 1b, 3g). The stylet is a gradually narrowing funnel that includes a slight turn of about 40° at about 80% of its length, then turning sharply by more than 90° in a knee-like fashion in the last 5% of its length (Figure 1c). The

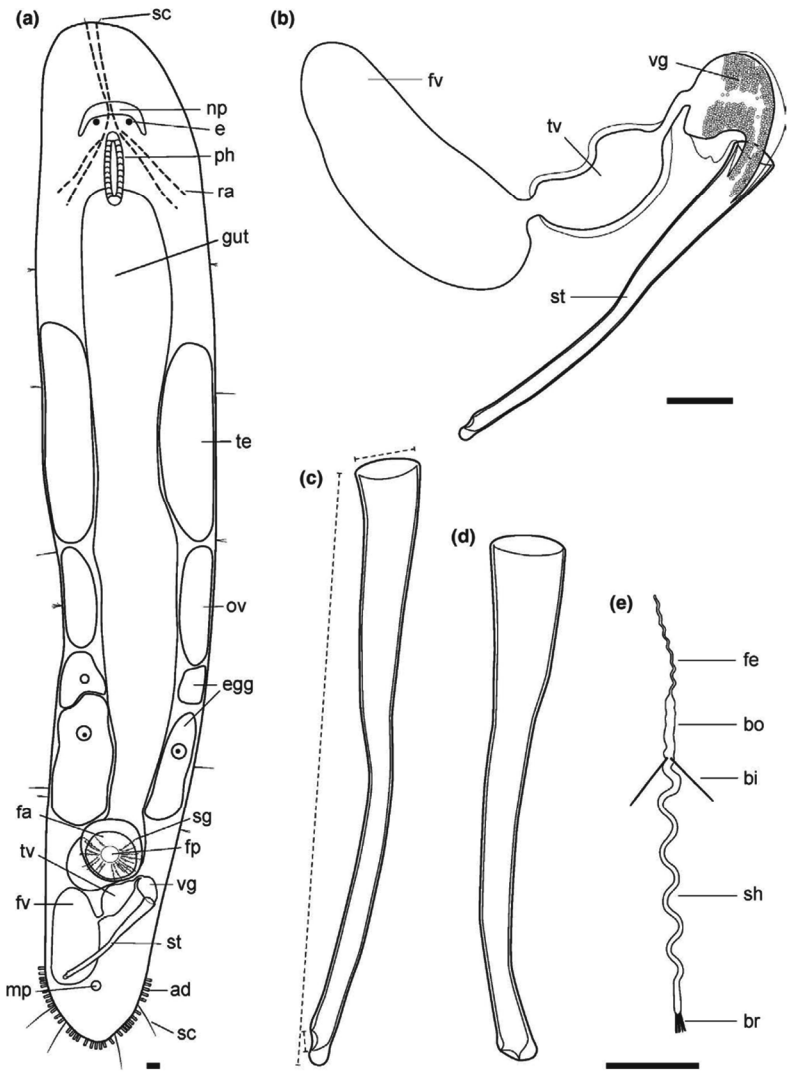


FIGURE 4 Line drawings of *Macrostomum cliftonensis* n. sp.: (a) Habitus (viewed from dorsal and lightly squeezed). (b) Male genital system with false seminal vesicle, muscular true seminal vesicle, vesicula granulorum, and stylet. (c) Detailed drawing of the most common view of the stylet. (d) Detailed drawing of an alternate view of the stylet (observed in MTP LS 2945). (e) Mature sperm cell with feeler, body, bristles, shaft, and brush (see Figure 1 for abbreviations)

outside turn of the knee is slightly thickened, while the inside turn ends in a point, leading to the distal opening being oriented laterally and being obliquely cut off (Figure 1c). Together, this gives the stylet a resemblance of a hook. The male canal ends in an unremarkable male antrum, and the male gonopore (at 95% BL) is slightly ciliated. The sperm (Figures 1d; 3k) have the usual morphology for a species in the reciprocally mating clade (Schärer et al., 2011).

Female system. The relatively small ovaries lie directly behind the testes, extending from 45% to 54% BL (Figures 1a, 3b). The developing eggs show ample yolk and shell granules (Figure 3c). Eggs in progressive stages of development propagate posterior from the ovaries and are frequently seen in the central body region before they enter the female antrum. The female antrum is relatively complex, and parts of it appear to be displaced somewhat to the left side of the body (Figures 1a, 3d, h–i), while the vagina appears to lie centrally (at 79% BL). Received sperm can often be observed in multiple locations, on the one hand opposite of the vagina (closed arrowhead in Figure 3i), but in many specimens also in more anterior regions (closed arrowhead in Figure 3h), which suggest that there may be a second chamber to the antrum, or a relatively loose cellular valve that sperm can penetrate deeply. Where the vagina enters the female antrum, there is a dense ciliary tuft with long cilia that reach into the antrum lumen (open arrowhead in Figure 3h). Relatively sparse shell glands surround the vagina extending to approximately the width of the animal (Figure 3j).

Mating behaviour. In Video S1, the worms engage in precopulatory behavior from 13 to 49 s, during which they circle while trying to get into the copulatory posture. The copulation starts at 50 s and ends at 147 s, lasting for ~97 s. During the copulation, the worms are interlinked in a disk-like posture and rotate intermittently about their

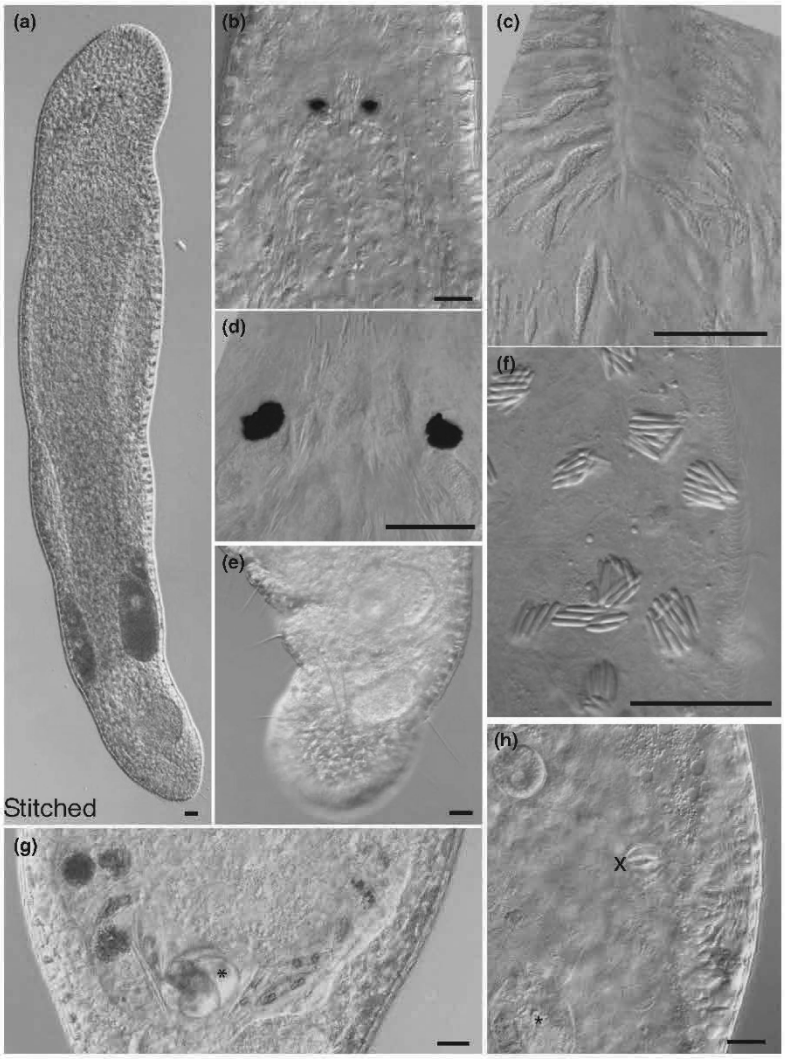


FIGURE 5 Micrographs of somatic structures of *Macrostomum cliftonensis* collected from the field (the bracket denotes the code of the deposited specimen, with MTP LS 2986 being the HOLOTYPE; note that all these specimens were imaged from the ventral side). (a) Overview of an adult, lightly squeezed, worm (MTP LS 2920, stitched from 3 images); (b) head region with neuropile, eyes, and pharynx (MTP LS 2930); (c) mouth opening with two distinct pharyngeal gland secretions (MTP LS 2911); (d) eyes with rhammite gland secretions penetrating the neuropile (MTP LS 2911); (e) tail region with numerous long sensory hairs (MTP LS 2906); (f) rhabdite bundles and ciliated epidermis (MTP LS 2930); (g) gut content with green algae, diatoms, and a foraminifer (*) (MTP LS 2910); (h) details of gut contents with a partially digested rotifer (*) and a rotifer jaw (x) (MTP LS 2930; for more detail see also 2017-01-18_19-37-08.074.avi). Scale bars 25 μm

center of mass (see also Figure 11a drawn from a more detailed recording). At the end of the copulation, one worm swims away, while the other remains at the same spot. At 189 s, the worms again come into contact and engage in circling, resulting in an unsuccessful copulation attempt at 212 s. This is followed by the bottom worm swimming away (at 248 s) and sucking (at 267 s) for a duration of 11 s (see also Figure 11b drawn from a more detailed recording).

Karyology and genome size. As already shown by Zadesenets et al. (2016), the most common karyotype of *M. janickei* (called *Macrostomum* sp. 8 therein) is 2n = 10 (18/22 specimens or 81.2%), with six small and four large metacentric chromosomes (Figure 11a), but there were also two specimens each that had either three (2n = 9) or five (2n = 11) large chromosomes (Figure 11b). The flow cytometric genome size estimates of *M. janickei* showed a complex pattern that approximately matched our karyotype results (Figure S1), with the most prominent peak having a relative fluorescence that was on average 4.55× larger than that of *D. melanogaster*, which would therefore correspond to a haploid genome size in *M. janickei* of 779 Mbp. However, note that these large chromosomes appear to represent near identical copies (Zadesenets, Schärer, et al., 2017b; Zadesenets et al., 2016), so that the actual amount of unique genome sequence might be lower. More detailed analyses of the genome organization of *M. janickei* are underway (K. S. Zadesenets, I. E. Jetybayev, L. Schärer, & N. B. Rubtsov, unpublished data).

Discussion

Given its characteristic stylet, *M. janickei* is clearly distinct from all other species in the genus *Macrostomum* published to date. The only species whose stylets bear any resemblance are *M. hamatum* Luther, 1947, described from the Baltic Sea at Tvärminne, Finland (Luther, 1947), one of the drawings of *M. balticum meridionalis* Papi, 1953, described from the San Rossore park near Pisa, Italy (Papi, 1953),

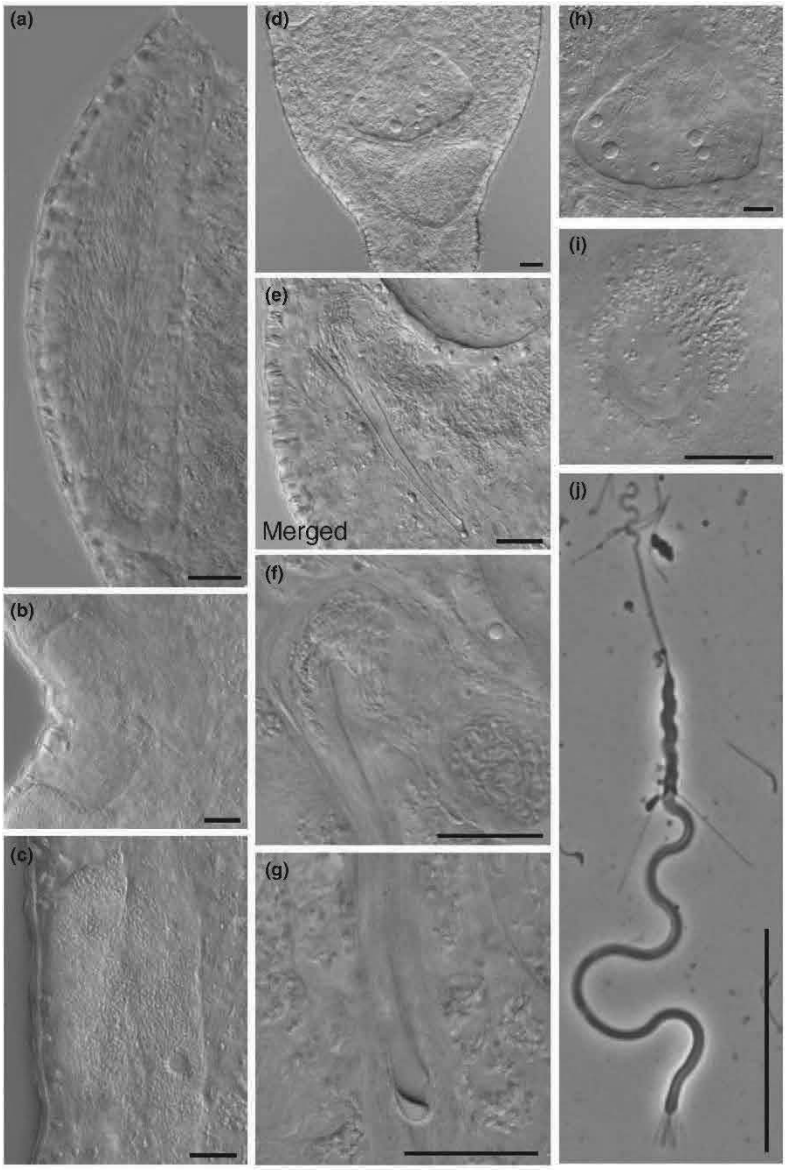


FIGURE 6 Micrographs of reproductive structures of *Macrostomum cliftonensis* collected from the field (the bracket denotes the code of the deposited specimen, with MTP LS 2986 being the HOLOTYPE; note that all specimens, except MTP LS 2908, were imaged from the ventral side). (a) Ripe testis with elongating spermatids aligned in the central region (MTP LS 2907); (b) ripe ovary with oocytes that are beginning to form yolk and shell granules (MTP LS 2896); (c) forming egg with the clearly visible nucleus, and many yolk and shell granules (MTP LS 2896); (d) body region containing the female antrum, the false and true seminal vesicle, vesicula granulorum, and stylet (MTP LS 2896); (e) false and true seminal vesicle, vesicula granulorum, and stylet (MTP LS 2907, merged from a movie); (f) muscular true seminal vesicle, ductus intervesicularis, vesicula granulorum, and base of the stylet (MTP LS 2907; see also the 2017-01-17_18-40-30.180.avi for additional detail); (g) tip of the stylet showing asymmetrical distal thickening (MTP LS 2911; for more detail see also 2017-01-18_10-53-55.330.avi); (h) female antrum with received sperm and coagulated seminal fluid; note the ciliary tuft in the center (MTP LS 2896). (i) Vagina surrounded by shell glands; note that the anterior shell glands appear first when focusing into the antrum (MTP LS 2896). (j) Sperm with typical feeler, body, bristles, shaft, and brush (MTP LS 2908, frame extracted from a movie). Scale bars 25 μm

specifically Figure 11 of Papi, 1953, and to a lesser degree *M. reynoldsoni* Young, 1976, described from the river Mwena near Mombasa, Kenya. We discuss these in turn in the following.

While the stylet of *M. hamatum* also has a hook-like appearance on the distal end, the stylet is more slender and has a fairly steady crescent-like turn of the stylet shaft. While Luther does not give any stylet length measurements, Ax (2008) reports on specimens collected in Sylt, Germany, that reach lengths of up to 150 μm, thus greatly exceeding the values of *M. janickei*. Moreover, we collected *M. hamatum* several times close to the type locality in Finland and find similar stylet lengths in our specimens. Finally, ongoing molecular phylogenetic analyses clearly show that *M. hamatum* is distinct from *M. janickei* (J. N. Brand, & L. Schärer, unpublished data). It is tempting to speculate that the stylet morphology of *M. janickei* might be involved in accessing difficult to reach parts of the female antrum, in order to remove rival sperm or lodge own sperm into places where they can efficiently anchor themselves with their feelers.

In the description of *M. balticum meridionalis*, Papi (1953) presents several drawings of the stylet, only one of which bears any resemblance to *M. janickei*, namely his Figure 11 (a reconstruction for serial sections). While all the other drawings of this species (Figures 6–9) clearly show a subterminal opening and a large unilateral distal thickening that exits at an angle <90° (a character state shared with *M. balticum* Luther, 1947, the species Papi considers his to be a subspecies of), the distal tip in Figure 11 is not drawn in much detail, and could possibly show a >90° turn, but leaving the placement of

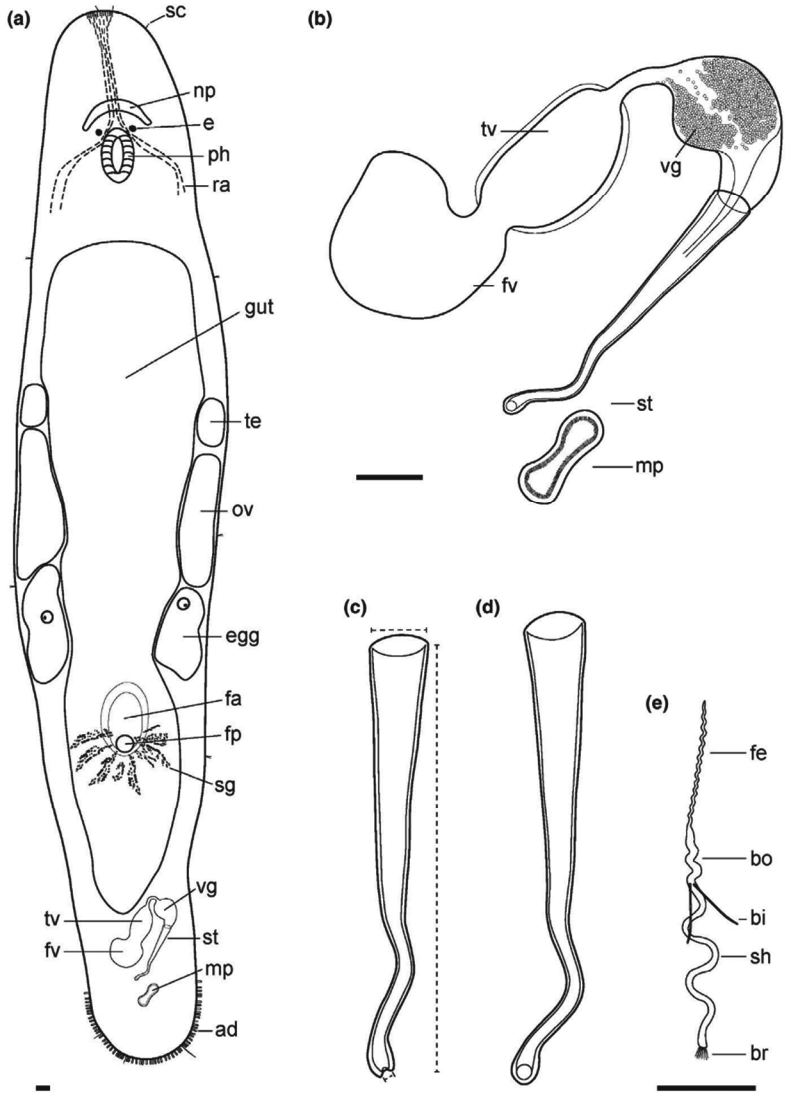


FIGURE 7 Line drawings of *Macrostomum mirumnovem* n. sp.: (a) Habitus (viewed from dorsal and lightly squeezed). (b) Male genital system with false seminal vesicle, muscular true seminal vesicle, vesicula granulorum, and stylet. (c) Detailed drawing of a common view of the stylet. (d) Detailed drawing of an alternate view of the stylet. (e) Mature sperm cell with feeler, body, bristles, shaft, and brush (see Figure 1 for abbreviations)

the opening unclear. Given the difficulties in reconstructing complex shapes of small (5–10 µm) structures from relatively thick paraffin sections (those of Papi are 3–4 µm, L. Schärer, pers. obs.), and given the fairly consistent stylet tip morphology in the other drawings, Figure 11 should clearly not be considered diagnostic for the stylet tip. Moreover, *M. balticum* has previously been phylogenetically placed close to *M. spirale* Ax, 1956, forming a clade that is quite diverged from the clade containing *M. janickei* (and *M. lignano* and *M. hystrix*), and we have ourselves recently collected *M. balticum meridionalis* close to the type locality, based on which we will show that, while being close to *M. balticum*, it is clearly a separate species (J. N. Brand, & L. Schärer, unpublished data).

While the stylet of *M. reynoldsoni* also turns >90° close to the distal tip, it does so in a much more gradual way than what we see in *M. janickei* and *M. hamatum*, leading to a nearly complete semicircle, rather than a sharp point. Moreover, the stylet of *M. reynoldsoni* lacks the "slight turn of about 40° at about 80% of its length" that we consider diagnostic for *M. janickei*.

Given that *M. janickei* is the closest relative of *M. lignano* found to date, we compared their mating behavior in pure versus mixed pairs and found that while both the copulation and suck durations were significantly longer in *M. janickei* than in *M. lignano*, these species are capable of cross-species mating, leading to low levels of successful hybridization (P. Singh, D. Ballmer, M. Laubscher, & L. Schärer, unpublished data), which further supports that they are phylogenetically closely related. Intriguingly, the resulting hybrid offspring showed an intermediate stylet morphology (P. Singh, D. Ballmer, M. Laubscher, & L. Schärer, unpublished data). Given what we currently know about the distribution of these species, it is unclear if there might be a naturally occurring hybrid zone.

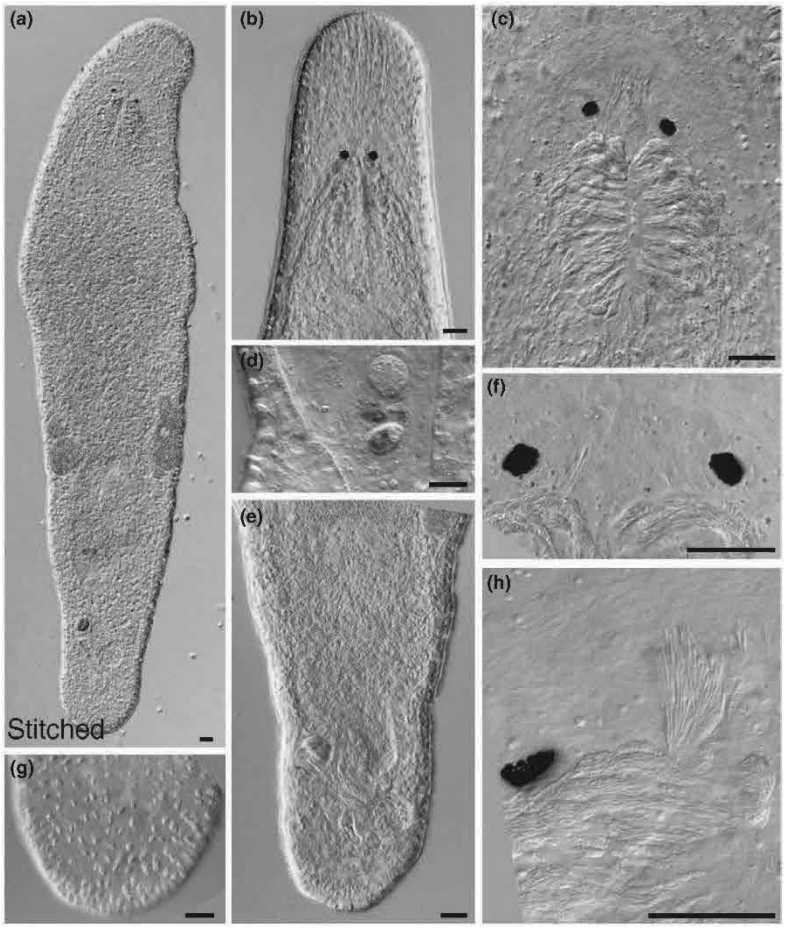


FIGURE 8 Micrographs of somatic structures of *Macrostomum mirumnovem* collected from the field (the bracket denotes the code of the deposited specimen, with MTP LS 3014 being the HOLOTYPE; note that all specimens, except MTP LS 3012, were imaged from the ventral side). (a) Overview of an adult, lightly squeezed, worm (MTP LS 3015, stitched from three images); (b) head region with rhammite glands, neuropile, eyes, and pharynx (MTP LS 3012); (c) pharynx region with neuropile, eyes, pharynx gland necks, and mouth opening (MTP LS 3015, frame extracted from a movie); (d) gut content with diatoms, and an unknown round cellular structure (MTP LS 2995, frame extracted from a movie); (e) tail region with adhesive glands (MTP LS 2993); (f) eyes with a few rhammite gland secretions penetrating the neuropile (MTP LS 3014); (g) rhabdite bundles on the tail plate (MTP LS 2996); (h) details of the pharynx gland necks showing two distinct pharyngeal gland secretions, as well as one pigment cup and some rhammite gland secretions (MTP LS 2994). Scale bars 25 μm

Macrostomum cliftonensis Schärer & Brand, n. sp.

urn:lsid:zoobank.org:act:18D35DE3-DE39-4E2A-9931-01499ABECA59 (Figures 4–6, Figure 10c, d, Figure 11c, d; Table 1, Table 2)

Material examined. Live observations on 14 field-collected specimens. Holotype: one extensively documented worm (MTP LS 2896, 94 images and videos) from the type locality (i.e., sample G7 collected on 17 January 2017 from sandy shore with reduced salinity water of 35‰ salinity seeping into a puddle on the shores of the increasingly hypersaline Lake Clifton with 97‰ salinity at time of sampling, Western Australia; S 32.76125, E 115.66019), including a whole mount (WAM V9393) and both partial *28S rRNA* and *COI* gene sequences. Paratypes: 13 extensively documented specimens (for a total of 724 images and videos), which complement the characterization of the holotype (see also Table 2), and many of which were also sequenced and/or include whole mounts.

Etymology. Species name refers to the type locality. To facilitate backward compatibility with previous records and reports, we note that this species has previously been referred to as *Macrostomum* sp. 84 (or Mac084 for short).

Diagnosis. *Macrostomum* with slightly dorsoventrally flattened and tongue-shaped body, and rounded rostrum (Figure 4a; Figure 4a). Body lengths of field caught worms range from 709 to 1,762 μm (see also Table 2). Short and long sensory cilia, with former mainly on anterior body and latter particularly pronounced on tail plate. Two small pigment cup eyes that sometimes appear kidney-shaped. Mouth relatively long, and gut extending caudally beyond female antrum, but posterior part relatively inconspicuous. Plentiful rhabdite bundles all over body, both on the dorsal and ventral side. Testes clearly larger than ovaries. The stylet is a long (~120 μm), slender and gradually narrowing funnel that, when looked at laterally, includes a slight turn (of ~15°) in about the center (although the stylet is also slightly curved at a right angle to this axis). Stylet ends in an oblique and almost lateral distal opening, with outside showing drop-shaped thickening and inside barely thickened (Figures 4c, 6g). Vesicula granulorum stands at a >90° angle to stylet axis. Vagina central, and received sperm can often be observed in female antrum, being anchored with feelers inside cellular valve. Shell glands granules surrounding vagina first appear in the anterior region when one focusses into antrum. Sperm relatively long (103 μm) showing general morphology of reciprocally mating species (Schärer et al., 2011).

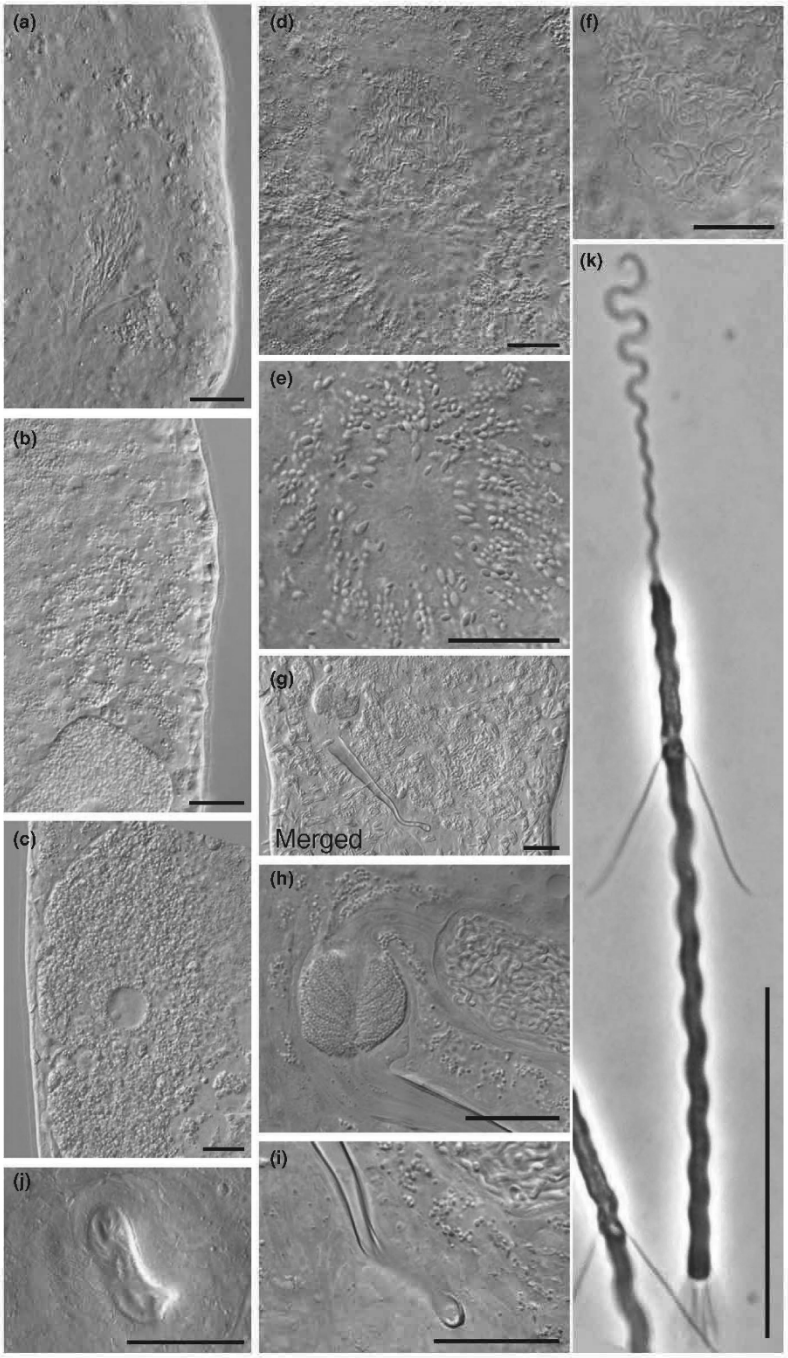


FIGURE 9 Micrographs of reproductive structures of *Macrostomum mirumnovem* collected from the field (the bracket denotes the code of the deposited specimen, with MTP LS 3014 being the HOLOTYPE; note that all these specimens were imaged from the ventral side). (a) ripe testis with some elongating spermatids aligned in the central region and the start of the vas deferens (MTP LS 3016); (b) ripe ovary with oocytes that are beginning to form yolk and shell granules, and a forming eggs (MTP LS 2994); (c) forming egg with the clearly visible nucleus and nucleolus, and many yolk and shell granules (MTP LS 3014); (d) body region containing the female antrum, the vagina and the shell glands (MTP LS 3015, frame extracted from a movie); (e) vagina surrounded by shell glands (MTP LS 3015, frame extracted from a movie). (f) Posterior region of the female antrum with received sperm, some coagulated seminal fluid; note the ciliary tuft left of the center (MTP LS 3015, frame extracted from a movie). (g) False and true seminal vesicle, vesicula granulorum, and stylet (MTP LS 3014, merged from a movie); (h) muscular true seminal vesicle, ductus intervesicularis, vesicula granulorum, and base of the stylet (MTP LS 3015); (i) tip of the stylet showing considerable s-shaped curvature and weak distal thickenings (MTP LS 3015, frame extracted from a movie); (j) the unusually large opening of the male antrum, lined by many cilia (MTP LS 2993, frame extracted from a movie). (k) Sperm with typical feeler, body, bristles, shaft, and brush (MTP LS 2993, frame extracted from a movie). Scale bars 25 μm

Description (see also Table 2)

General morphology. The body is slightly dorsoventrally flattened and tongue-shaped (width/length ratio: 1:3.8), and widest at the level of testes or ovaries (Figures 4a, 5a). The rostrum is rounded and the tail plate barely set apart from the rest of the body. The pigment cup eyes are small and often slightly kidney-shaped (Figure 5d). The body is covered homogeneously with cilia, and short sensory cilia are present on the rostrum, while there are sparse long sensory cilia all along the body, particularly on the tail plate (Figure 5e). The pharynx is associated with a relatively long mouth opening (Figure 5b, c), and the gut (often containing food items, Figure 5g, h) extends posteriorly beyond the female antrum, ending at the seminal vesicles (at 85% BL). There are many rhabdite bundles all over the body, both on

the dorsal and ventral side (with the exception of the regions around the mouth and the gonopores), forming bundles of 6–13 granules (Figure 5f). The rhammite glands originate dorsally anterior of the testes, and their granules are secreted via the neuropile (Figure 5b) into the anterior rostrum, where very prominent rhammite bundles can be observed. The adhesive glands are arranged horseshoe-like in multiple rows along the edge of the tail plate.

Male system. The prominent testes extend from 30% to 48% BL, with developing sperm aligned in the center (Figure 6a). The vas deferens can be seen in some specimens (e.g., MTP LS 2909) and appears to leave the testis at the posterior end. There is a prominent and often oval-shaped false seminal vesicle that lies on the left of the tail plate and which often extends quite far posterior (Figure 6d). It connects anterolaterally to a muscular true seminal vesicle that lies in a more central position. The true seminal vesicle is connected anterolaterally via a short ductus intervesicularis to the prominent and muscular vesicula granulorum (Figures 4b; 6e, f), whose main axis is positioned on the proximal stylet opening at a >90° angle to the main axis of the stylet. The tail plate contains many prostate gland cells that send out their necks into the vesicula granulorum, from which prostate glands necks only reach into the first ~15% of the stylet length (Figure 6f). The stylet is a very slender and gradually narrowing funnel that, when looked at laterally, includes a very slight turn of about 15° at about 60% of its length, before then ending in an oblique and almost lateral distal opening (Figures 5c, 6f); note, however, that the stylet is also slightly curved at a right angle to this axis, which can be seen in the second drawing (Figure 4d). The outside of the distal end shows a drop-shaped thickening, while the inside is barely thickened (Figures 4c, 6g); note that stylet and tip can appear quite differently when somewhat rotated (see Figure 4d). The male canal ends in a distinct male antrum, and the male gonopore (at 95% BL) is ciliated. The sperm (Figures 4e, 6j) have the usual morphology for a species in the reciprocally mating clade (Schärer et al., 2011).

Female system. The ovaries lie directly behind the testes (from 49% to 62% BL) and show the usual morphology (Figures 4a, 6b), and the developing eggs also show the usual structure, with ample yolk and shell granules (Figure 6c). The female antrum lies centrally and appears somewhat variable in shape, often having a distinct lumen and a prominent anterior thickening of the epithelium (Figure 6d + h). The vagina opens directly into the antrum lumen (at 80% BL). Received sperm can often be observed being anchored with their feelers inside of the cellular valve; note that one can often observe coagulated prostate secretions (Figure 6d + h). Where the vagina enters the female antrum, there is a distinct ciliary tuft with long cilia that reach into the antrum lumen (Figure 6h). Shell glands surround the vagina, with the granules first appearing in the anterior region when one focusses into the antrum (Figure 6j). The shell gland secretions are remarkably dense, but do not extend far from the female opening, rarely reaching the sides of the animal.

Mating behaviour. In Video S2, the worms get into contact at 6 s, leading to the initiation of precopulatory behavior at 47 s, in which the worms circle and crawl on each other for 23 s. From 70 s, the worms are firmly interlinked in a disk-like posture and start copulating. The copulation lasts for ~34 s and ends at 105 s, during which the worms rotate about their center of mass, followed by both worms sucking (at 105 s) for a duration of 10 and 14 s, respectively.

Karyology and genome size. All 100 analyzed specimens of *M. cliftonensis* showed a 2n = 6 karyotype with six similar-sized metacentric chromosomes (Figure 11c-d) and that simple karyotype was also supported by our flow cytometric genome size estimates (Figure S1). These showed a single peak having a relative fluorescence that was on average 1.35× larger than that of *D. melanogaster*, which would therefore correspond to a haploid genome size in *M. cliftonensis* of 231 Mbp. More detailed analyses of the genome organization of *M. cliftonensis* are underway (K. S. Zadesenets, I. E. Jetybayev, L. Schärer, & N. B. Rubtsov, unpublished data).

Discussion

As shown in Figure 4c, d, the stylet of *M. cliftonensis* can appear somewhat variable in different specimens, which is, at least in part, due to the angle at which these stylets are viewed. Moreover, in one deposited specimen (MTP LS 2920) the stylet appears malformed and shorter than usual, possibly because the initial elongation during stylet formation was hampered (stylets start forming from the distal tip; Egger et al., 2009), leading to a mass of accumulated stylet material at the distal tip.

While the stylet of *M. cliftonensis* is unique, there are five *Macrostomum* species with stylets that are similar, all of which match by inhabiting coastal brackish habitats. First, *M. curvituba* Luther, 1947—described from the Hanko peninsula, Finland—differs by being blind and having a stylet that is smaller (75–103 μm), more evenly curved and ending in bilateral distal thickenings (Ax, 1994; Luther, 1947). Second, *M. magnacurvituba* Ax, 1994—described from coastal waters of Greenland and Iceland—is also blind and has a stylet that is somewhat similar but much larger (140–175 μm) than either *M. curvituba* (hence the name; Ax, 1994) or *M. cliftonensis*. Third, *M. mediterraneum* Ax, 1956—described from Etang de Sigean, Southern France—has a stylet that matches *M. cliftonensis* by being somewhat (although fairly evenly) curved and having a unilateral distal thickening. However, Ax actually draws this distal thickening of this species somewhat differently in different papers, being smaller and more similar to *M. cliftonensis* in the initial description (Ax, 1956), and larger and more rounded (and thus less similar) in a later account from the Marmara Sea, Turkey (Ax, 1959). Moreover, in both cases these stylets are either shorter (88–90 μm) or in the lower range (103 μm), respectively, than that of *M. cliftonensis*. Fourth, *M. longituba* Papi, 1953—described from a brackish water canal in the San Rossore Reserve, Pisa—has a stylet that is not only more evenly curved and longer, but also ends in a much narrower tip that lacks a very distinct distal thickening and includes a slight turn at the very end (Papi, 1953).

Finally, *M. greenwoodi* Faubel & Cameron, 2001—described from the Coomera salt marshes, Queensland—is superficially the most similar species, and since it was described from coastal habitats in

Queensland, Australia (Faubel & Cameron, 2001), we discuss it in some more depth. In overall appearance, the two species are similar, with the reported body length of 980 µm being within the range of *M. cliftonensis*, and with the distribution of sensory cilia and rhabdites approximately matching as well. The stylet of *M. greenwoodi* is more evenly and more strongly curved, and reported to be shorter than that of *M. cliftonensis*, with values in the text given as 98.3 µm, although Figure 3a, b of Faubel (2001) suggest conflicting values of 105 µm and 53 µm, respectively. For the following comparison, we assume that this discrepancy stems from the scale bar of Figure 3b being wrong, suggesting a stylet length of ~100 µm, which is also supported by a re-examination of the serially sectioned holotype of *M. greenwoodi* (Queensland Museum, QMG217363), which Ladurner et al. (2005) carried out as part of their species description of *M. lignano*. This stylet size is in the lower range of what we have found in *M. cliftonensis* (though note that our shortest stylet may have been malformed). Moreover, in *M. greenwoodi* the size of the proximal opening (13 µm) as well as the distal opening (4 µm) is, respectively, just outside and just inside the range of what we observe in *M. cliftonensis*, suggesting that the stylet of *M. greenwoodi* is overall smaller and narrower, as well as more evenly and strongly curved. And while Faubel does not mention distal thickenings of the stylet tip, the re-examination of the holotype had actually suggested that this species does carry thickenings (Ladurner et al., 2005). Other aspects that do not match well are the positions of the female and male gonopores, which are both more anterior in *M. greenwoodi* (70% and 91%) compared to *M. cliftonensis* (82% and 95%). And finally, Faubel (2001) does not draw cilia that surround the region where the vagina enters the female antrum (and this absence was confirmed by the re-examination of the holotype; Ladurner et al., 2005), while we clearly see such cilia in *M. cliftonensis*. We therefore conclude that *M. cliftonensis* is distinct from *M. greenwoodi*, and all the other species discussed above.

Macrostomum mirumnovem Schärer & Brand, n. sp.

urn:lsid:zoobank.org:act:740C2A0B-8288-4CB6-9E09-0959A4338120
(Figures 7–9, Figure 10e, f, Figure 11e, f; Table 1, Table 2)

Material examined. Live observations on 11 field-collected specimens and one specimen from laboratory culture. Holotype: one extensively documented worm (MTP LS 3014, 68 images and videos) from the type locality (i.e., sample A13 collected on 29 January 2017 from the upper intertidal on a sheltered beach, at 35‰ salinity, in front of the Victorian Marine Science Consortium, Queenscliff, Port Phillip Bay, Victoria; S 38.27007, E 144.63894), including a whole mount (NMW F258462). Paratypes: 11 extensively documented specimens, some from the type locality and others from other locations in the general area (for a total of 579 images and videos), which complement the characterization of the holotype (see also Table 2), and many of which were sequenced and/or include whole mounts.

Etymology. Species name refers to unusual karyotype that generally consists of nine chromosomes (Lat. *mirum* = strange and *novem* = nine). To facilitate backward compatibility with previous records and reports, we note that this species has previously been referred to as *Macrostomum* sp. 94 (or Mac094 for short).

Diagnosis. *Macrostomum* species with slightly dorsoventrally flattened and lancet-shaped body, conical rostrum with frequently rounded tip, and tail plate being slightly set aside (Figures 7a, 8a). Body lengths of field caught worms range from 670 to 1,667 µm (see also Table 2). Short and slightly longer sensory cilia, with former mainly on anterior body and latter only on tail plate. Two small pigment cup eyes that sometimes appear crescent-shaped. Gut extending caudally far beyond female antrum. Testes clearly smaller than ovaries. The distinctive stylet (~80 µm) is a slender and gradually narrowing funnel that then stays relatively constant in diameter in distal half, widening again near tip (Figures 7c, 9g), multiple turns in s-shaped mid region, occurring in more than one plane (compare views in Figure 7c, d). Depending on viewing angle, the distal stylet opening ends terminally or somewhat subterminally, both sides carrying slight distal thickenings (Figure 9i). Prostate gland cells send out necks into vesicula granulorum, but granules do not reach into stylet, while a set of strong muscles do. Vagina central and received sperm can often be observed in the female antrum, anchored with feelers inside cellular valve. Dense shell glands surround vagina, with majority radiating laterally, none directly anterior and only few posterior. Sperm (86 µm) showing the general morphology of reciprocally mating species (Schärer et al., 2011), having a relatively long feeler (32 µm).

Description (see also Table 2)

General morphology. The body is slightly dorsoventrally flattened and lancet-shaped, and widest at the level of the testes (Figures 7a, 8a). The rostrum is conical and the tip often rounded (Figure 8b), and the tail plate is slightly set aside (Figure 8e). The pigment cup eyes are small and often crescent-shaped (Figure 8c + f + h). The body is covered homogeneously with cilia, and sparse sensory cilia are largely restricted to the rostrum, with a few slightly longer ones on the tail plate. The mouth and pharynx are unremarkable (Figure 8b, c), and the gut extends posteriorly far beyond the female antrum (sometimes containing food items, Figure 8d + e), ending directly at the seminal vesicles (at 83% BL). Rhabdite bundles are relatively sparse, and most abundant on the dorsal side of the head and tail, forming bundles of 4 to 12 granules (Figure 8g). The rhammite glands originate dorsally anterior of the testes, and their granules are secreted via the neuropile into the anterior rostrum, where prominent rhammite bundles can be observed (Figure 8b, c). The adhesive glands are arranged horseshoe-like in multiple rows along the edge of the tail plate.

Male system. The small and often difficult to see testes extend from 31% to 38% BL and they usually show relatively few developing sperm that are often curled up in the center (Figure 9a). While the vas deferens is not easily seen, one can sometimes see a few ripe sperm descending toward the seminal vesicles. There is a small false seminal vesicle that lies on the left of the tail plate and which connects anterolaterally to a more central muscular true seminal vesicle (Figures 7b, 9g). The true seminal vesicle is connected via

a ductus intervesicularis that turns sharply into the prominent and muscular vesicula granulorum, which itself sits on the proximal stylet opening at a considerable angle (Figure 9h). The tail contains many prostate gland cells that send out their necks into the vesicula granulorum, but the granules do not reach into the stylet, while a set of strong muscles do (Figure 9h). Besides these distal muscles, the vesicula granulorum does not show signs of strong musculature. The stylet is a slender and gradually narrowing funnel for the first 60% of its length that then stays relatively constant in diameter for the next 35% in the mid region, and finally widening again near the tip (Figures 7c, 9g); there are multiple turns in this s-shaped mid region, the first by about 20° and the second, turning back by about 60°, but the turns occur in more than one plane, and can therefore not be completely captured in 2D (compare the views in Figure 7). Depending on the viewing angle, the distal stylet opening ends terminally or somewhat subterminally (compare Figure 7c, d), and both sides carry slight distal thickenings (Figure 9i). The male canal ends in a remarkably spacious male antrum, and the male gonopore is unusually large, often forming an oblong slit (Figure 9j); both the male antrum and the male pore (located at 92% BL) are strongly ciliated. The sperm (Figures 7e, 9k) have the usual morphology for a species in the reciprocally mating clade (Schärer et al., 2011).

Female system. The prominent ovaries lie directly behind the testes (from 39% to 54% BL) and show the usual morphology (Figure 9b), and the developing eggs show the usual structure, with ample yolk and shell granules (Figure 9c). The female antrum lies centrally, is relatively simple, oblong and has a thick antrum epithelium (Figure 9d). The vagina also lies centrally, entering in the posterior section of the antrum (at 68% BL). Received sperm can often be observed being anchored with their feelers inside the cellular valve. Where the vagina enters the female antrum, there is a small ciliary tuft with cilia that reach into the antrum lumen (Figure 9f). Dense shell glands surround the vagina and they extend asymmetrically, with the majority radiating laterally, none directly anterior and only few posterior (Figure 9e).

Mating behaviour. In Video S3, the worms come into contact at 8 s and begin to circle each other, potentially trying to get into the copulatory posture. At 15 s, they start reciprocally copulating for a duration of 803 s, though by 791 s the copulation seems unilateral, with the worm on the right presumably having removed its stylet from the partner's female antrum. Their copulatory posture appears angular, with the shape resembling an **S** (see e.g. at 529 s). At the beginning of the copulation, the worms vigorously move and elongate their anterior body (at 8–78 s), but this eventually slows down. At 832 s, the worm on the left bends down and starts sucking for a duration of 14 s.

Karyology and genome size. The most common karyotype of *M. mirumnovem* is 2n = 9 (34/52 specimens or 65.4%) with six small and three large metacentrics (Figure 11e). However, there are a large number of specimens, which differ both in the number of small and large chromosomes (K. S. Zadesenets, I. E. Jetybayev, L. Schärer, & N. B. Rubtsov, unpublished data). Surprisingly, the flow cytometric genome size estimates of *M. mirumnovem* showed an apparently simpler pattern than what we could have expected based on our karyotype results (Figure S1), with only a single peak having a relative fluorescence that was on average 2.52× larger than that of *D. melanogaster*, and which would therefore correspond to a haploid genome size in *M. mirumnovem* of 431 Mbp. However, given some caveats regarding the measurements for this species, these values need to be treated with some caution (see Figure S1). More detailed analyses of the genome organization of *M. mirumnovem* are underway (K. S. Zadesenets, I. E. Jetybayev, L. Schärer, & N. B. Rubtsov, unpublished data).

Discussion

While the complex stylet shape of *M. mirumnovem* is unique, there are four described species that have comparably complex stylets, two each inhabiting freshwater and coastal brackish habitats, respectively. The former includes *M. johni* Young, 1972—described from a freshwater lake in Wales (Young, 1972), but also reported from a freshwater coastal lagoon in Brazil (Gamo & Leal-Zanchet, 2004)—has a stylet that is similar in size (78–95 µm) to that of *M. mirumnovem*. Its stylet also shows multiple complex turns, but ends in "a slightly swollen cowl or hood" (Young, 1972), and thus is clearly distinct from our species. Moreover, in *M. johni* the testes are almost twice as long as the ovaries, while the reverse is true for *M. mirumnovem*, and the female antrum is considerably further back in *M. johni* and the gut ends approximately at the level of the antrum. The other freshwater species, *M. retortum* Papi, 1951—described from temporary freshwater ponds in the San Rossore Reserve, Pisa—has a stylet that is not only more complex, but also considerably longer (up to 138 µm) than that of *M. mirumnovem* (Papi, 1951).

Among the coastal species, *M. bellebaruchae* Ax, 2008—described from Winyah Bay, South Carolina in freshwater, but considered to have some brackish influence—has a stylet that also has many turns, but is considerably shorter (56 µm) than that of *M. mirumnovem* and does not carry any distal thickenings on the stylet tip (Ax, 2008). The other brackish species is *M. coomerensis* Faubel & Cameron, 2001—collected from the Coomera salt marshes, Queensland—whose stylet resembles an exaggerated version of that of *M. mirumnovem*, both with respect to the more acute sharpness of the many turns, as well as with respect to its much larger size (125 µm), and the more extreme turn at the end, leading to a clearly subterminal opening (Faubel & Cameron, 2001).

Karyotype and genome size of *Macrostomum lignano* and *Macrostomum hystrix*

As shown previously (and briefly summarized above), *M. lignano* has a complex karyotype, with different inbred lines and outbred cultures showing different levels of karyotype variability (Zadesenets et al., 2016). These observations are well supported by the flow cytometric genome size estimates we obtained here (Figure S1), with the DV1 inbred line and the LS3 outbred culture showing peaks that correspond to the known 2n = 8, 2n = 9, and 2n = 10 karyotypes, and the LS1 outbred culture showing a single peak corresponding

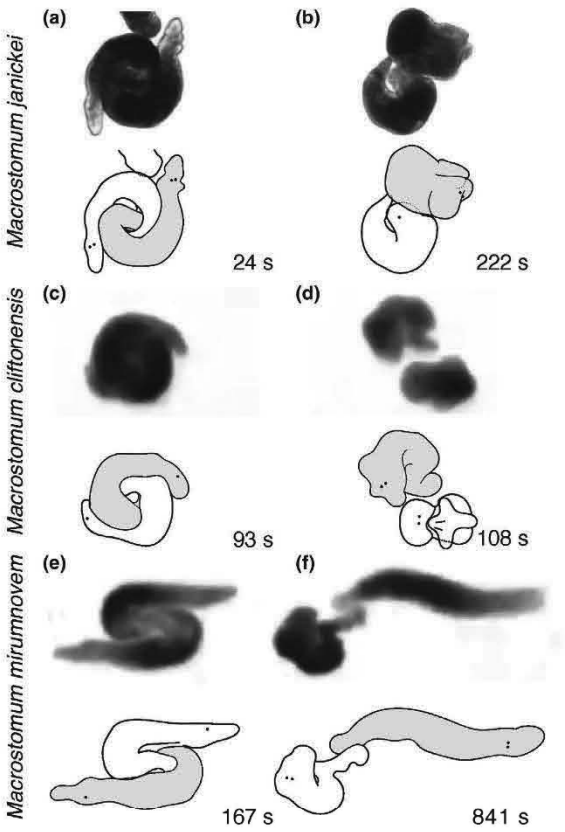


FIGURE 10 Mating behavior of three new *Macrostomum* species, with single frames extracted from the deposited movies and line drawings made from these frames: (a) copulatory posture of *M. janickei* (frame taken from an additional movie that had contained three worms); (b) sucking behavior of *M. janickei* (frame taken from an additional movie), with the upper worm showing the suck posture oriented obliquely; (c) copulatory posture of *M cliftonensis*; (d) sucking behavior of *Macrostomum cliftonensis*, with the upper and lower worms showing the suck posture oriented sideways and toward the camera, respectively; (e) copulatory posture of *M. mirumnovem*; and (f) sucking behavior of *M. mirumnovem*, with the left worm showing the suck posture oriented sideways. Note that the timing information refers to the deposited movies and that the magnifications are not to scale across the species

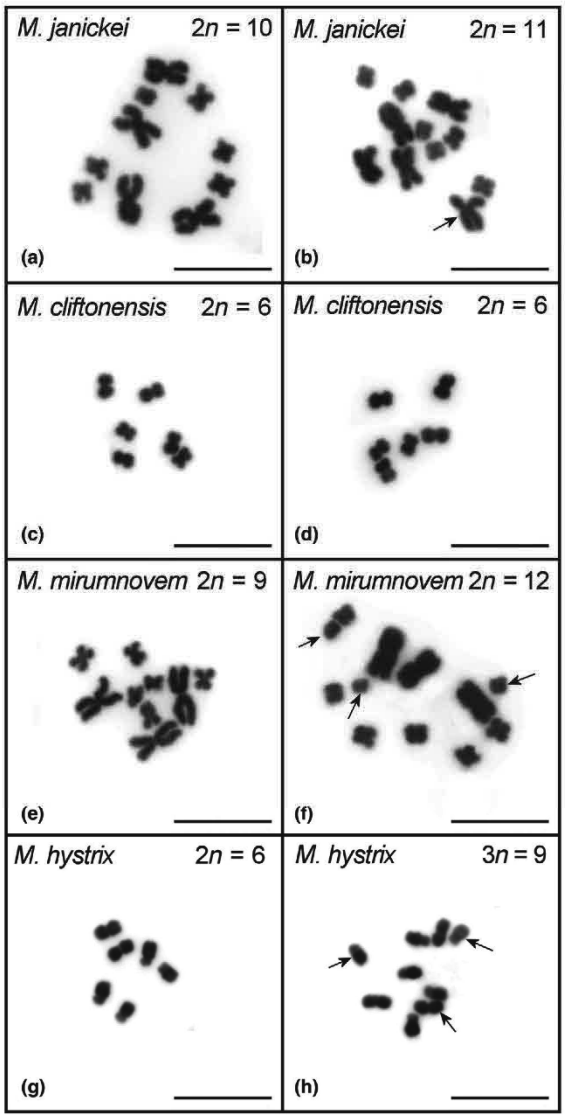


FIGURE 11 Metaphase karyotype spreads of the three new *Macrostomum* species and the outgroup species: (a) the "normal" and most common chromosome set of *M. janickei*; (b) an "abnormal" chromosome set of *M. janickei*; (c-d) two chromosome sets of karyologically uniform *M. cliftonensis*; (e) the "normal" and most common chromosome set of *M. mirumnovem*; (f) one of the many possible "abnormal" chromosome sets of *M. mirumnovem*; (g) the "normal" and almost uniform chromosome set of *M. hystrix*; and (h) a very rare polyploid karyotype variant of *M. hystrix*. Additional chromosomes are marked by arrows. Inverted DAPI image. Scale bar 10 μm

to solely the 2n = 8 karyotype (the other karyotypes are rare in LS1; Zadesenets et al., 2016). These peaks have relative fluorescence values that, respectively, average 3.07x, 3.89x, and 4.69× that of *D. melanogaster*, so that the 2n = 8, 2n = 9, and 2n = 10 karyotypes, corresponding to haploid genome sizes of 525, 666, and 803 Mbp, respectively. However, note that the large chromosomes appear to represent near identical copies (Zadesenets, Schärer, et al., 2017b; Zadesenets et al., 2016), so that the actual amount of unique genome sequence might correspond that observed in the 2n = 8 karyotype.

Earlier results suggested that *M. hystrix* has a stable 2n = 6 karyotype (10/10), with six similar-sized chromosomes (Figure 11g), including two metacentrics and four submetacentrics (Zadesenets et al., 2016). As part of this study, we analyzed an additional 67 worms, which overall confirms this karyotype for almost all specimens (i.e., 75/77, or 97%). However, we also found two specimens with an apparently triploid 3n = 9 karyotype (Figure 11h), with three

metacentrics and six submetacentrics. Shifts from diploidy to triploidy could result from polyspermy (Snook, Hosken, & Karr, 2011; Toda & Okamoto, 2016), which might occur since this species mates by hypodermic insemination (Ramm, Schlatter, Poirier, & Schärer, 2015; Schärer et al., 2011). The notion that *M. hystrix* has a simple karyotype was also supported by our flow cytometric genome size estimates (Figure S1), which showed a single peak having a relative fluorescence that was on average 1.27× that of *D. melanogaster*, corresponding to a haploid genome size in *M. hystrix* of 217 Mbp.

Molecular phylogenetic placements

The topology of the *28S rRNA* gene tree supported three clades among these *M. lignano* relatives (Figure 12), containing (a) *M. janickei* (from France) and *M. lignano* (from Italy and Greece), (b) *M. cliftonensis* (from Western Australia), and (c) *M. mirumnovem* (from Victoria) and an undescribed *Macrostomum* sp. (from South Australia). Note that the latter specimen was collected by Laumer and Giribet (2014) at a site that is ~450 km from our Victoria collection site. Moreover, the tree topology further suggested that this marker is not variable enough to resolve the interrelationships in the first clade, showing a maximum divergence of only 4% among the in-group species. So while the marker is informative for more deeply diverged species (Schärer et al., 2011), more variable markers are needed to resolve these close species.

We therefore used the faster evolving partial *COI* gene, which showed a maximum divergence of close to 20% among the in-group species. The topology of the *COI* gene tree recovered the same three clades, but in addition it also supported a separation between *M. janickei* and the *M. lignano* specimens (Figure 13). Moreover, while all *M. lignano* fall into one supported clade, which we expected given their stylet morphology (Ladurner et al., 2005), there is some substructure among the *M. lignano* specimens from Italy and Greece, with all Italian specimens forming a single well-supported clade, while the Greek specimens from the LS3 line fall into two somewhat separate and poorly supported groups. The LS3 culture was established from collections in two sites on the Sithonia peninsula (i.e., Vourvourou and Porto Koufo; Zadesenets et al., 2016), which lie about 30 km apart. We had not expected population differentiation between those sites, and so the laboratory cultures were established with individuals from both sample sites. It would be interesting to resample these sites to understand if these two *COI* gene sequences map, respectively, onto the two sample sites, or if both populations contain multiple *COI* gene sequences.

Frameshift mutation in the *COI* gene of *Macrostomum hystrix*

The observed frameshift mutation in the *COI* gene was surprising and required further validation. We first confirmed that this deletion was supported by both sequencing directions in both analyzed *M. hystrix* specimens, which stem from two different sampling locations in Italy that are about 300 km apart (Table 1). Next, we confirmed that the deletion was also supported when mapping 10.49 million 101 bp Illumina reads from an RNA-Seq study of the SR1 line of *M. hystrix* (J. N. Brand, & L. Schärer, unpublished data) onto the partial *COI* gene of the SR1 line (using Bowtie2 with local alignment and high sensitivity). In total, 37'921 (0.36%) reads mapped to this fragment, and 3,518 (0.034%) reads mapped to the AAT codon resulting from the deletion, of which 99.9% supported the deletion. This high number of mapped reads clearly suggested that the *COI* gene carrying the deletion is expressed in *M. hystrix* and that it is unlikely that we are dealing with a nuclear pseudogene. To evaluate this further, we used the analogous mapping approach in *M. lignano*, which does not carry the deletion. Specifically, we mapped 12.24 million 100 bp Illumina reads from an RNA-Seq study of the DV1 line of *M. lignano* (Ramm et al., 2019) onto the partial *COI* gene of the DV1 line. In total, 36'710 (0.30%) reads mapped to this fragment, and 5,401 (0.044%) reads mapped to the same codon (which, in the absence of the deletion is an AAA), of which 99.7% supported that codon and thus the absence of a deletion in *M. lignano*.

These results suggest that there are no highly expressed transcripts that lack the deletion in *M. hystrix* (as could have been expected if a version lacking the deletion existed and was expressed, or if RNA editing were to efficiently correct the deletion). Moreover, it seems that the transcript in *M. hystrix* (carrying the deletion) is expressed at a similar level to the transcript in *M. lignano* (not carrying a deletion). Together, this clearly suggests that *M. hystrix* is able to develop, grow and reproduce successfully in spite of the presence of this frameshift mutation. Interestingly, the remainder of the *COI* gene sequence of *M. hystrix* did not show an evidently higher level of divergence, as could have been expected if selection were relaxed after the stop codon. This may suggest that this part of the sequence is still under stabilizing selection, possibly due to residual protein being produced due to translational misreading of the mutant mRNA, as has been shown in other organisms (Andersson, Slechta, & Roth, 1998; Atkins, Elseviers, & Gorini, 1972; Remacle, Gloire, Cardol, & Matagne, 2004). Based upon the new insights presented here, including the above RNA-Seq data, we have now been able to effect a correction of the deposited GenBank sequence to include the deletion (now KP730561.2; see also Table 1).

4 | CONCLUSIONS

We here have identified and taxonomically described three previously undescribed species of the free-living flatworm genus *Macrostomum*, all of which we show to be close relatives of the widely used flatworm model species *M. lignano*. Moreover, all of these newly described species permit to address interesting evolutionary research questions in their own right. First, we show that *M. janickei* is the closest known relative of *M. lignano*, with which it shares an unusual hidden polyploidy. This karyotype organization is likely the result of a whole-genome duplication event via auto-polyploidization, leading to a fusion chromosome that varies considerably in copy number, both within and between these two species (Zadesenets, Ershov, et al., 2017a; Zadesenets, Schärer, et al., 2017b; Zadesenets et al., 2016). This

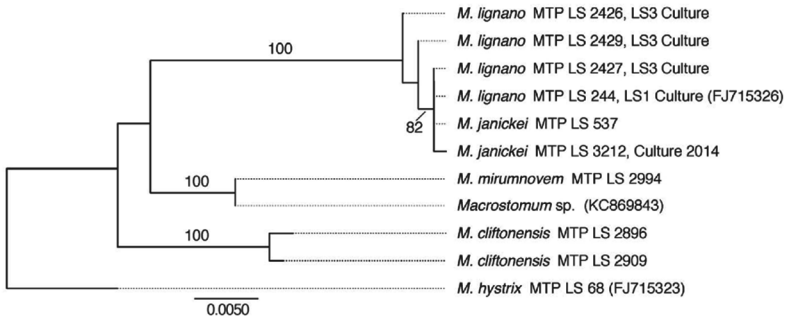


FIGURE 12 Molecular phylogeny of the relatives of *Macrostomum lignano* based on the partial *28S rRNA* gene sequence. The nodal values are bootstrap supports from a Tamura-Nei neighbor joining tree reconstruction (showing only bootstrap values >75%)

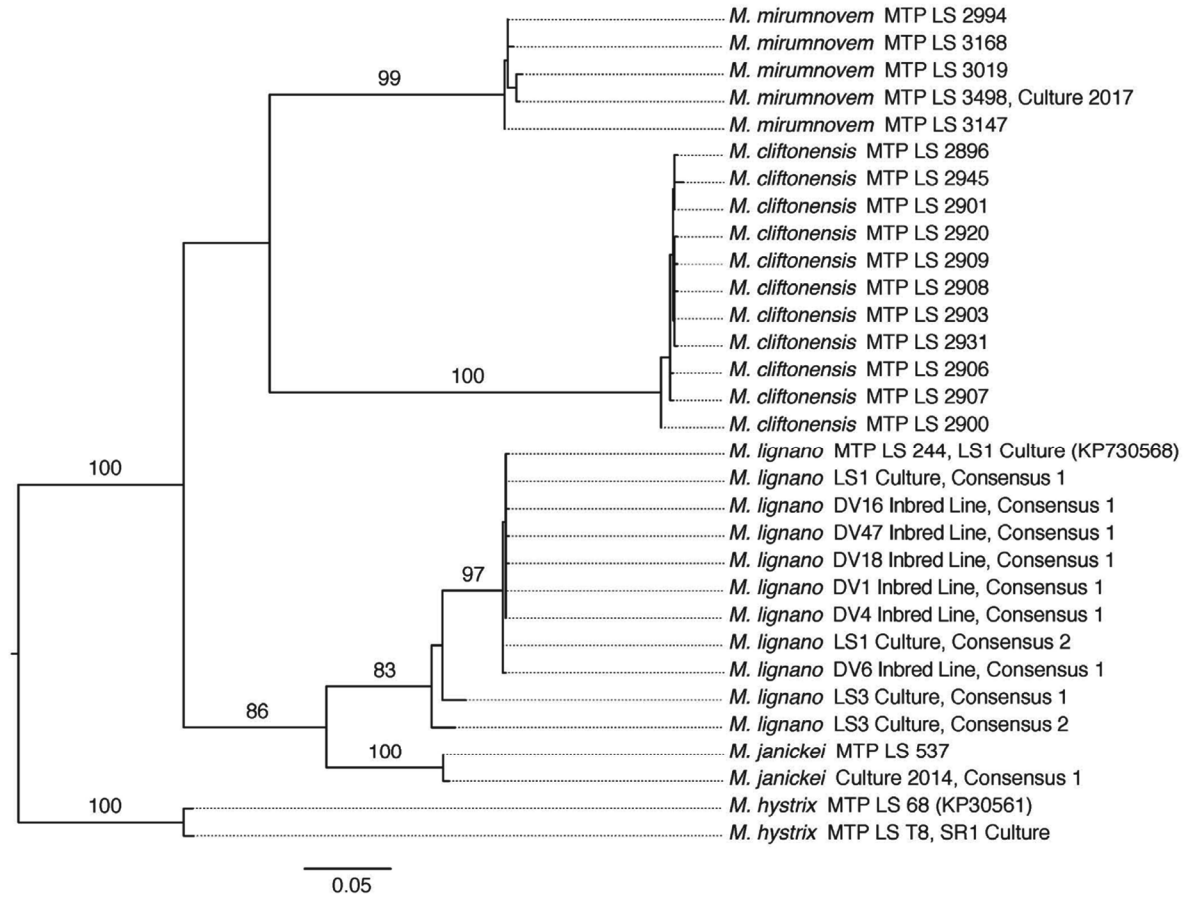


FIGURE 13 Molecular phylogeny of the relatives of *Macrostomum lignano* based on the partial *COI* gene. The nodal values are bootstrap supports from a maximum likelihood tree reconstruction (showing only bootstrap values >75%)

species pair therefore permits to study the evolutionary outcomes of early rediploidization processes following a whole-genome duplication (K. S. Zadesenets, I. E. Jetybayev, L. Schärer, & N. B. Rubtsov, unpublished data), shedding light on processes of genome evolution. Second, *M. mirumnovem* shows an even more unusual genome organization with a 2n = 9 base karyotype pattern, which we think may have resulted via allo-polyploidization (K. S. Zadesenets, I. E. Jetybayev, L. Schärer, & N. B. Rubtsov, unpublished data), and it represents another interesting species to study genome evolution. Moreover, the karyotype variation observed in these *Macrostomum* species calls for studies on gene regulation and dosage compensation. And third, *M. cliftonensis* shows a simple and stable karyotype, and a small haploid genome size of only 231 Mbp, which makes *M. cliftonensis* an attractive species to develop as a genetic and genomic *Macrostomum* model to eventually replace *M. lignano*, particularly since it is also outcrossing and shows the typical reproductive traits of a reciprocally

mating species (Schärer et al., 2011). In addition to describing these new species, we have also shown *M. hystrix* to have a simple and stable karyotype, and a comparably small genome of 217 Mbp, making this a promising model to study the evolution of reproduction in hypodermically mating species (Schärer et al., 2011), particularly since it is currently the closest described relative of *M. lignano* showing that type of mating behavior. Our results showcase the striking biological diversity in *Macrostomum* flatworms, which will permit an even broader range of evolutionary research questions to be addressed in this fascinating genus.

ACKNOWLEDGEMENTS

Tim Janicke and Georgina Rivera-Ingraham; John Evans, Leigh Simmons, and Ben Byrne; and Yvonne Gilbert and Rod Watson, provided support with collecting *M. janickei*, *M. cliftonensis*, and *M. mirumnovem*, respectively. *M. cliftonensis* was collected and exported based on DPAW Reg 17 (01-000135-2) and Reg 18 Licenses (OS002550), respectively. *M. mirumnovem* was collected based on DELWP National Parks Act 1975 Permit (10008144). Dita Vizoso and Anaïs Galli helped with establishing cultures. Eduard Stöckli; Andrew Hosie; and Genefor Walker-Smith and Melanie Mackenzie, kindly received and catalogued specimens for the Natural History Museum Basel, Western Australian Museum and the Museum Victoria, respectively. Manuel Ruedi and Richard Pyle kindly gave advice on questions concerning taxonomy. Maria Pichler and Peter Stadler provided technical assistance with flow cytometry. KZ was supported by project 0324-2019-0042. This research was funded by SNF projects 31003A-143732 and 31003A-162543 to LS.

ORCID

Lukas Schärer https://orcid.org/0000-0002-0209-9871

REFERENCES

Andersson, D. I., Slechta, E. S., & Roth, J. R. (1998). Evidence that gene amplification underlies the adaptive mutability of the *lac* operon. *Science*, *282*, 1133–1135.

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 in *Macrostomum lignano*. *Frontiers in Zoology*, *12*, 14. https://doi.org/10.1186/s12983-015-0106-0

Atkins, J. F., Elseviers, D., & Gorini, L. (1972). Low activity of β-galactosidase in frameshift mutants of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *69*(5), 1192–1195.

Ax, P. (1956). Les turbellariés des étangs côtiers du littoral méditerranéen de la France méridionale. *Vie Et Milieu, Suppl.*, *5*, 1–215.

Ax, P. (1959). Zur Systematik, Ökologie und Tiergeographie der Turbellarienfauna in den ponto-kaspischen Brackwassermeeren. *Zoologische Jahrbücher: Abteilung Für Systematik, Ökologie Und Geographie Der Tiere*, *87*, 43–187.

Ax, P. (1994). *Macrostomum magnacurvituba* n. sp. (Macrostomida, Plathelminthes) replaces *Macrostomum curvituba* in coastal waters of Greenland and Iceland. *Microfauna Marina*, *9*, 335–338.

Ax, P. (2008). *Plathelminthes aus Brackgewässern der Nordhalbkugel*. Stuttgart, Germany: Franz Steiner Verlag.

Bazakos, C., Hanemian, M., Trontin, C., Jiménez-Gómez, J. M., & Loudet, O. (2017). New strategies and tools in quantitative genetics: How to go from the phenotype to the genotype. *Annual Review of Plant Biology*, *68*, 435–455. https://doi.org/10.1146/annurev-arplant-042916-040820

Brooks, A. K., & Gaj, T. (2018). Innovations in CRISPR technology. *Current Opinion in Biotechnology*, *52*, 95–101. https://doi.org/10.1016/j.copbio.2018.03.007

Cardona, A., Hartenstein, V., & Romero, R. (2005). The embryonic development of the triclad *Schmidtea polychroa*. *Development Genes and Evolution*, *215*(3), 109–131.

Egger, B., Bachmann, L., & Fromm, B. (2017). Atp8 is in the ground pattern of flatworm mitochondrial genomes. *BMC Genomics*, *18*, 414. https://doi.org/10.1186/s12864-017-3807-2

Egger, B., Gschwentner, R., Hess, M. W., Nimeth, K. T., Adamski, Z., Willems, M., … Salvenmoser, W. (2009). The caudal regeneration blastema is an accumulation of rapidly proliferating stem cells in the flatworm *Macrostomum lignano*. *BMC Developmental Biology*, *9*, 41. https://doi.org/10.1186/1471-213X-9-41

Egger, B., & Ishida, S. (2005). Chromosome fission or duplication in *Macrostomum lignano* (Macrostomorpha, Plathelminthes) - Remarks on chromosome numbers in 'archoophoran turbellarians'. *Journal of Zoological Systematics and Evolutionary Research*, *43*(2), 127–132. https://doi.org/10.1111/j.1439-0469.2005.00300.x

Egger, B., Ladurner, P., Nimeth, K., Gschwentner, R., & Rieger, R. (2006). The regeneration capacity of the flatworm *Macrostomum lignano*—on repeated regeneration, rejuvenation, and the minimal size needed for regeneration. *Development Genes and Evolution*, *216*(10), 565–577. https ://doi.org/10.1007/s00427-006-0069-4

Faubel, A., & Cameron, B. (2001). Platyhelminthes from salt marshes of Coomera River, Southeastern Queensland, Australia. *Memoirs of the Queensland Museum*, *46*, 511–519.

Gamo, J., & Leal-Zanchet, A. M. (2004). Freshwater microturbellarians (Platyhelminthes) from Rio Grande do Sul, Brazil. *Revista Brasileira De Zoologia*, *21*(4), 897–903. https://doi.org/10.1590/S0101-81752004000400026

Gregory, T. R. (2019). Animal genome size database. http://www.genomesize.com

Grudniewska, M., Mouton, S., Simanov, D., Beltman, F., Grelling, M., de Mulder, K., … Berezikov, E. (2016). Transcriptional signatures of somatic neoblasts and germline cells in Macrostomum lignano. *eLife*, *5*, e20607. https://doi.org/10.7554/eLife.20607

Janicke, T., & Schärer, L. (2010). Sperm competition affects sex allocation but not sperm morphology in a flatworm. *Behavioral Ecology and Sociobiology*, *64*, 1367–1375. https://doi.org/10.1007/s00265-010-0951-y

Janssen, T., Vizoso, D. B., Schulte, G., Littlewood, D. T. J., 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. *Molecular Phylogenetics and Evolution*, *92*, 82–107. https://doi.org/10.1016/j.ympev.2015.06.004

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*, 587–589. https://doi.org/10.1038/nmeth.4285

Kuales, G., De Mulder, K., Glashauser, J., Salvenmoser, W., Takashima, S., Hartenstein, V., … Ladurner, P. (2011). Boule-like genes regulate male and female gametogenesis in the flatworm *Macrostomum lignano*. *Developmental Biology*, *357*, 117–132. https://doi.org/10.1016/j.ydbio.2011.06.030

Ladurner, P., Egger, B., De Mulder, K., Pfister, D., Kuales, G., Salvenmoser, W., & Schärer, L. (2008). The stem cell system of the basal flatworm

Macrostomum lignano. In T. C. G. Bosch (Ed.), *Stem cells: From Hydra to Man* (pp. 75–94). Berlin, Germany: Springer.

Ladurner, P., Schärer, L., Salvenmoser, W., & Rieger, R. M. (2005). A new model organism among the lower Bilateria and the use of digital microscopy in taxonomy of meiobenthic Platyhelminthes: *Macrostomum lignano*, n. sp. (Rhabditophora, Macrostomorpha). *Journal of Zoological Systematics and Evolutionary Research*, *43*(2), 114–126. https://doi.org/10.1111/j.1439-0469.2005.00299.x

Lane, J. A. K., Clarke, A. G., & Winchcombe, Y. C. (2017). *South West Wetlands Monitoring Program 1977–2016*. Perth, WA: Government of Western Australia. Retrieved from https://www.dpaw.wa.gov.au/management/wetlands/wetlands-monitoring-and-research

Laumer, C. E., & Giribet, G. (2014). Inclusive taxon sampling suggests a single, stepwise origin of ectolecithality in Platyhelminthes. *Biological Journal of the Linnean Society*, *111*, 570–588. https://doi.org/10.1111/bij.12236

Lengerer, B., Hennebert, E., Flammang, P., Salvenmoser, W., & Ladurner, P. (2016). Adhesive organ regeneration in *Macrostomum lignano*. *BMC Developmental Biology*, *16*, 20. https://doi.org/10.1186/s12861-016-0121-1

Lengerer, B., Pjeta, R., Wunderer, J., Rodrigues, M., Arbore, R., Schärer, L., … Ladurner, P. (2014). Biological adhesion of the flatworm *Macrostomum lignano* relies on a duo-gland system and is mediated by a cell type-specific intermediate filament protein. *Frontiers in Zoology*, *11*, 12. https://doi.org/10.1186/1742-9994-11-12

Lengerer, B., Wunderer, J., Pjeta, R., Carta, G., Kao, D., Aboobaker, A., … Ladurner, P. (2018). Organ specific gene expression in the regenerating tail of *Macrostomum lignano*. *Developmental Biology*, *433*(2), 448–460. https://doi.org/10.1016/j.ydbio.2017.07.021

Luther, A. (1947). Untersuchungen an rhabdocoelen Turbellarien VI. *Macrostomiden Aus Finnland*. *Acta Zoologica Fennica*, *49*, 1–38.

Marie-Orleach, L., Janicke, T., & Schärer, L. (2013). Effects of mating status on copulatory and postcopulatory behaviour in a simultaneous hermaphrodite. *Animal Behaviour*, *85*(2), 453–461. https://doi.org/10.1016/j.anbehav.2012.12.007

Marie-Orleach, L., Janicke, T., Vizoso, D. B., David, P., & Schärer, L. (2016). Quantifying episodes of sexual selection: Insights from a transparent worm with fluorescent sperm. *Evolution*, *70*(2), 314–328. https://doi.org/10.1111/evo.12861

Marie-Orleach, L., Janicke, T., Vizoso, D. B., Eichmann, M., & Schärer, L. (2014). Fluorescent sperm in a transparent worm: Validation of a GFP marker to study sexual selection. *BMC Evolutionary Biology*, *14*, 148. https://doi.org/10.1186/1471-2148-14-148

Mouton, S., Grudniewska, M., Glazenburg, L., Guryev, V., & Berezikov, E. (2018). Resilience to aging in the regeneration-capable flatworm *Macrostomum lignano*. *Aging Cell*, *17*(3), e12739. https://doi.org/10.1111/acel.12739

Mouton, S., Willems, M., Braeckman, B. P., Egger, B., Ladurner, P., Schärer, L., & Borgonie, G. (2009). The free-living flatworm *Macrostomum lignano*: A new model organism for ageing research. *Experimental Gerontology*, *44*, 243–249. https://doi.org/10.1016/j.exger.2008.11.007

Newmark, P. A., & Sánchez Alvarado, A. (2002). Not your father's planarian: A classic model enters the era of functional genomics. *Nature Reviews Genetics*, *3*(3), 210–219. https://doi.org/10.1038/nrg759

Nguyen, L.-T., Schmidt, H. A., Haeseler, A. V., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Papi, F. (1951). Richerche sui Turbellari Macrostomidae. *Archivo Zoologico Italiano*, *36*, 289–341.

Papi, F. (1953). Beiträge zur Kenntnis der Macrostomiden (Turbellarien). *Acta Zoologica Fennica*, *78*, 1–32.

Pellettieri, J., & Sánchez Alvarado, A. (2007). Cell turnover and adult tissue homeostasis: From humans to planarians. *Annual Review of Genetics*, *41*, 83–105. https://doi.org/10.1146/annurev.genet.41.110306.130244

Pfister, D., De Mulder, K., Hartenstein, V., Kuales, G., Borgonie, G., Marx, F., … Ladurner, P. (2008). Flatworm stem cells and the germ line: Developmental and evolutionary implications of macvasa expression in *Macrostomum lignano*. *Developmental Biology*, *319*(1), 146–159. https://doi.org/10.1016/j.ydbio.2008.02.045

Ramm, S. A., Lengerer, B., Arbore, R., Pjeta, R., Wunderer, J., Giannakara, A., … Schärer, L. (2019). Sex allocation plasticity on a transcriptome scale: Socially sensitive gene expression in a simultaneous hermaphrodite. *Molecular Ecology*, *28*(9), 2321–2341. https://doi.org/10.1111/mec.15077

Ramm, S. A., Schlatter, A., Poirier, M., & Schärer, L. (2015). Hypodermic self-insemination as a reproductive assurance strategy. *Proceedings of the Royal Society B-Biological Sciences*, *282*, 20150660. https://doi.org/10.1098/rspb.2015.0660

Remacle, C., Gloire, G., Cardol, P., & Matagne, R. F. (2004). Impact of a mutation in the mitochondrial LSU rRNA gene from Chlamydomonas reinhardtii on the activity and the assembly of respiratory-chain complexes. *Current Genetics*, *45*, 323–330. https://doi.org/10.1007/s00294-004-0490-z

Rink, J. C. (2013). Stem cell systems and regeneration in planaria. *Development Genes and Evolution*, *223*(1–2), 67–84. https://doi.org/10.1007/s00427-012-0426-4

Rouhana, L., Weiss, J. A., Forsthoefel, D. J., Lee, H., King, R. S., Inoue, T., … Newmark, P. A. (2013). RNA interference by feeding in vitro–synthesized double-stranded RNA to planarians: Methodology and dynamics. *Developmental Dynamics*, *242*, 718–730. https://doi.org/10.1002/dvdy.23950

Schärer, L., Joss, G., & Sandner, P. (2004). Mating behaviour of the marine turbellarian *Macrostomum* sp.: These worms *suck*. *Marine Biology*, *145*, 373–380. https://doi.org/10.1007/s00227-004-1314-x

Schärer, L., Littlewood, D. T. J., Waeschenbach, A., Yoshida, W., & Vizoso, D. B. (2011). Mating behaviour and the evolution of sperm design. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1490–1495. https://doi.org/10.1073/pnas.1013892108

Sekii, K., Vizoso, D. B., Kuales, G., De Mulder, K., Ladurner, P., & Schärer, L. (2013). Phenotypic engineering of sperm-production rate confirms evolutionary predictions of sperm competition theory. *Proceedings of the Royal Society B-Biological Sciences*, *280*(1757), 20122711. https://doi.org/10.1098/rspb.2012.2711

Singh, P., Vellnow, N., & Schärer, L. (2019). Variation in sex allocation plasticity in three closely related flatworm species. *Ecology and Evolution*. https://doi.org/10.1002/ece3.5566

Snook, R. R., Hosken, D. J., & Karr, T. L. (2011). The biology and evolution of polyspermy: Insights from cellular and functional studies of sperm and centrosomal behavior in the fertilized egg. *Reproduction*, *142*, 779–792. https://doi.org/10.1530/REP-11-0255

Stelzer, C.-P., Riss, S., & Stadler, P. (2011). Genome size evolution at the speciation level: The cryptic species complex *Brachionus plicatilis* (Rotifera). *BMC Evolutionary Biology*, *11*, 90. https://doi.org/10.1186/1471-2148-11-90

Telford, M. J., Herniou, E. A., Russell, R. B., & Littlewood, D. T. J. (2000). Changes in mitochondrial genetic codes as phylogenetic characters: Two examples from the flatworms. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(21), 11359–11364. https://doi.org/10.1073/pnas.97.21.11359

Toda, E., & Okamoto, T. (2016). Formation of triploid plants via possible polyspermy. *Plant Signaling & Behavior*, *11*(9), e1218107. https://doi.org/10.1080/15592324.2016.1218107

Vellnow, N., Vizoso, D. B., Viktorin, G., & Schärer, L. (2017). No evidence for strong cytonuclear conflict over sex allocation in a simultaneously hermaphroditic flatworm. *BMC Evolutionary Biology*, *17*, 103. https://doi.org/10.1186/s12862-017-0952-9

Wasik, K. A., Gurtowski, J., Zhou, X., Mendivil Ramos, O., Delás, M. J., Battistoni, G., … Schatz, M. C. (2015). Genome and transcriptome of the regeneration-competent flatworm, *Macrostomum lignano*. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(40), 12462–12467. https://doi.org/10.1073/pnas.1516718112

Wudarski, J., Simanov, D., Ustyantsev, K., de Mulder, K., Grelling, M., Grudniewska, M., … Berezikov, E. (2017). Efficient transgenesis and annotated genome sequence of the regenerative flatworm model *Macrostomum lignano*. *Nature Communications*, *8*, 2120. https://doi.org/10.1038/s41467-017-02214-8

Wunderer, J., Lengerer, B., Pjeta, R., Bertemes, P., Kremser, L., Lindner, H., … Ladurner, P. (2019). A mechanism for temporary bioadhesion. *Proceedings of the National Academy of Sciences*, *116*(10), 4297–4306. https://doi.org/10.1073/pnas.1814230116

Young, J. O. (1972). Further studies on the occurrence of freshwater Microturbellaria in the British Isles. I. A description of *Macrostomum johni* sp. nov. *Freshwater Biology*, *2*, 253–258. https://doi.org/10.1111/j.1365-2427.1972.tb00054.x

Young, J. O. (1976). Systematic studies of limnic *Macrostomum* species (Turbellaria, Macrostomida) from East Africa. *Zoologica Scripta*, *5*, 49–60.

Zadesenets, K. S., Ershov, N. I., Berezikov, E., & Rubtsov, N. B. (2017a). Chromosome evolution in the free-living flatworms: First evidence of intrachromosomal rearrangements in karyotype evolution of *Macrostomum lignano* (Platyhelminthes, Macrostomida). *Genes*, *8*, 298. https://doi.org/10.3390/genes8110298

Zadesenets, K. S., Schärer, L., & Rubtsov, N. B. (2017b). New insights into the karyotype evolution of the free-living flatworm *Macrostomum lignano* (Platyhelminthes, Turbellaria). *Scientific Reports*, *7*, 6066. https://doi.org/10.1038/s41598-017-06498-0

Zadesenets, K. S., Vizoso, D. B., Schlatter, A., Konopatskaia, I. D., Berezikov, E., Schärer, L., & Rubtsov, N. B. (2016). Evidence for karyotype polymorphism in the free-living flatworm, *Macrostomum lignano*, a model organism for evolutionary and developmental biology. *PLoS ONE*, *11*(10), e0164915. https://doi.org/10.1371/journal.pone.0164915

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Flow-cytometric measurements of genome size in five *Macrostomum* species (*M. lignano*, *M. janickei*, *M. cliftonensis*, *M. mirumnovem*, and *M. hystrix*), including three different lines/cultures of *M. lignano* (the inbred line DV1, and the outbred cultures LS1 and LS3), with 2–4 replicates per species (rows).

Table S1 Primers used for PCR amplification and sequencing of the analysed partial *28S rRNA* and *COI* gene sequences.

Alignment S1 Alignment of the partial *28S rRNA* gene sequences (FASTA-format).

Alignment S2 Alignment of the partial *COI* gene sequences (FASTA-format).

Video S1 Video sequence of mating behaviour of *Macrostomum janickei* n. sp., showing circling, copulation and suck behaviour.

Video S2 Video sequence of mating behaviour of *Macrostomum cliftonensis* n. sp., showing circling, copulation and suck behaviour.

Video S3 Video sequence of mating behaviour of *Macrostomum mirumnovem* n. sp., showing circling, copulation and suck behaviour.

How to cite this article: Schärer L, Brand JN, Singh P, Zadesenets KS, Stelzer C-P, Viktorin G. A phylogenetically informed search for an alternative *Macrostomum* model species, with notes on taxonomy, mating behavior, karyology, and genome size. *J Zool Syst Evol Res*. 2020;58:41–65. https://doi.org/10.1111/jzs.12344

Supporting Information

A phylogenetically-informed search for an alternative *Macrostomum* model species, with notes on taxonomy, mating behaviour, karyology and genome size

Lukas Schärer, Jeremias N. Brand, Pragya Singh, Kira S. Zadesenets, Claus-Peter Stelzer, and Gudrun Viktorin

Supporting Information Table S1: Primers used for PCR amplification and sequencing of the analysed partial *28S rRNA* and *COI* gene sequences.

Gene	F/R	Sequence	Reference
28S rRNA			
PCR & sequencing primers			
ZX-1	F	ACCCGCTGAATTTAAGCATAT	Schärer et al. 2011
1500R	R	GCTATCCTGAGGGAAACTTCG	Schärer et al. 2011
additional sequencing primers			
300F	F	CAAGTACCGTGAGGGAAAGTTG	Schärer et al. 2011
ECD2	R	CTTGGTCCGTGTTTCAAGACGGG	Schärer et al. 2011
COI			
PCR & sequencing primers			
Mac_COIF	F	GTTCTACAAATCATAAGGATATTGG	Janssen et al. 2015
Mac_COIR	R	TAAACYTCWGGGTGACCAAAAAACCA	Janssen et al. 2015
adapted PCR primers			
M13_COI_084_F	F	M13_GTTAGGTTCTACAAACCACAAGGATATAGG	this study
M13r_COI_084_R	R	M13r_TGGAAATGGGCAACAACATAATATGTGTCGTG	this study
M13_COI_094_F	F	M13_ATTAGGATCTACAAAACACAAGGACATCGG	this study
M13r_COI_094_R	R	M13r_TGAAAATGGGCAACAACATAGTAAGTATCGTG	this study
additional sequencing primers			
M13	F	TGTAAAACGACGGCCAG	from Microsynth
M13r	R	CAGGAAACAGCTATGAC	from Microsynth

Supporting Information Alignment S1. Alignment of the partial *28S rRNA* gene sequences (FASTA-format). (available online)

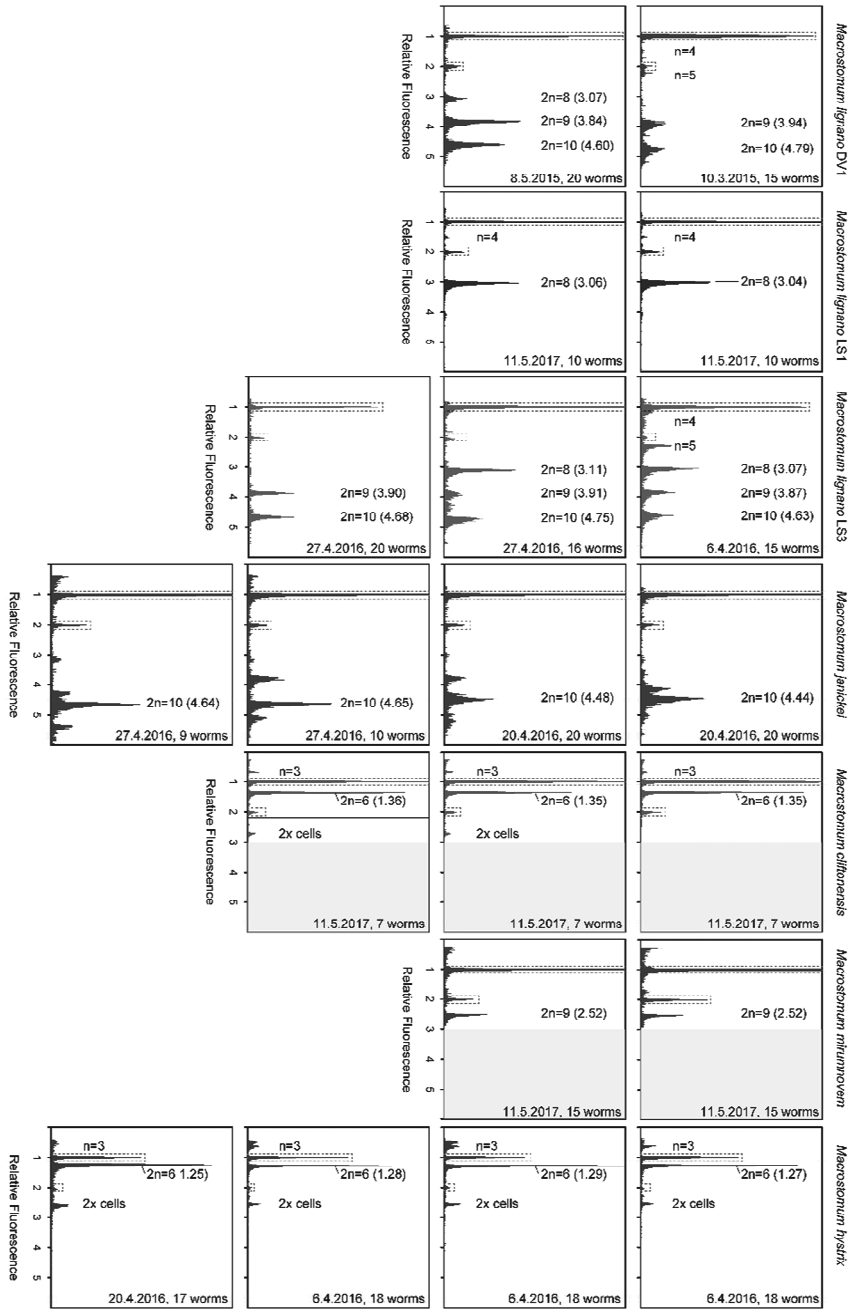
Supporting Information Alignment S2. Alignment of the partial *COI* gene sequences (FASTA-format). (available online)

Supporting Information Video S1. Video sequence of mating behaviour of *Macrostomum janickei* n. sp., showing circling, copulation and suck behaviour. (available online)

Supporting Information Video S2. Video sequence of mating behaviour of *Macrostomum cliftonensis* n. sp., showing circling, copulation and suck behaviour. (available online)

Supporting Information Video S3. Video sequence of mating behaviour of *Macrostomum mirumnovem* n. sp., showing circling, copulation and suck behaviour. (available online)

Supporting Information Figure S1. Flow-cytometric measurements of genome size in five *Macrostomum* species (*M. lignano*, *M. janickei*, *M. cliftonensis*, *M. mirumnovem*, and *M. hystrix*), including three different lines/cultures of *M. lignano* (the inbred line DV1, and the outbred cultures LS1 and LS3), with 2-4 replicates per species (rows). Histograms are based on the YL2-A fluorescence signals (detection bandwidth 590–650 nm) and the y-axis represents cell counts. We rescaled the x-axis, so that the relative fluorescence values of the worm nuclei are displayed as multiples of the fluorescence signals of the diploid *Drosophila melanogaster* nuclei used as the internal standard (=set to 1, and indicated by the tall stippled red box). Note that there are also some "2x cells" of *D. melanogaster* that are likely in the G2-phase of the cell cycle (i.e. after the S-phase DNA replication, indicated by the low stippled red box). For the *Macrostomum* nuclei we labelled the major peaks with the corresponding karyotypes (adding the peak sizes in brackets), and we also identified peaks that match the expected haploid chromosome sets (which, depending on the karyotype can be n=3, n=4, and/or n=5), since some of these worms were likely actively producing sperm when they were sampled, while others might have been somewhat starved and not producing sperm. Moreover, in some cases we also detected "2x cells" in *Macrostomum*, namely in *M. cliftonensis* and *M. hystrix*. Finally, for two species (*M. cliftonensis* and *M. mirumnovem*) the gain had to be increased for the measurement, so that the region >3 cannot be measured reliably, which is indicated by the grey background. While we do not expect values in this range for *M. cliftonensis*, it seems possible that we may have missed peaks for *M. mirumnovem*.


Macrostomum lignano DV1
Macrostomum lignano LS1
Macrostomum lignano LS3
Macrostomum jenickei
Macrostomum cliftonensis
Macrostomum minimovem
Macrostomum hystrix
Relative Fluorescence
n=4
n=5
2n=8 (3.07)
2n=9 (3.84)
2n=10 (4.60)
8.5.2015, 20 worms
2n=9 (3.94)
2n=10 (4.79)
10.3.2015, 15 worms
2n=8 (3.06)
11.5.2017, 10 worms
2n=8 (3.04)
11.5.2017, 10 worms
2n=8 (3.11)
2n=9 (3.91)
2n=10 (4.75)
27.4.2016, 16 worms
2n=8 (3.07)
2n=9 (3.87)
2n=10 (4.63)
6.4.2016, 15 worms
2n=9 (3.90)
2n=10 (4.68)
27.4.2016, 20 worms
2n=10 (4.64)
27.4.2016, 9 worms
2n=10 (4.65)
27.4.2016, 10 worms
2n=10 (4.48)
20.4.2016, 20 worms
2n=10 (4.44)
20.4.2016, 20 worms
n=3
2n=6 (1.36)
2x cells
11.5.2017, 7 worms
2n=6 (1.35)
2x cells
11.5.2017, 7 worms
2n=6 (1.35)
11.5.2017, 7 worms
2n=9 (2.52)
11.5.2017, 15 worms
2n=9 (2.52)
11.5.2017, 15 worms
2n=6 1.25)
2x cells
20.4.2016, 17 worms
2n=6 (1.28)
6.4.2016, 18 worms
2n=6 (1.29)
6.4.2016, 18 worms
2n=6 (1.27)
6.4.2016, 18 worms

Chapter VI
Concluding Remarks

While sexual selection has been studied extensively in *M. lignano* (e.g. Schärer and Ladurner 2003; Janicke et al. 2013; Marie-Orleach et al. 2016; Vellnow et al. 2017) and some experimental work was done on a handful of other species (e.g. Ramm et al. 2012; Giannakara and Ramm 2017; Singh et al. 2019a,b), little large-scale comparative work has been done to date (Singh and Schärer in preparation; Schärer et al. 2011). Thus, our knowledge of the evolution of *Macrostomum* largely relies on extrapolations from this rather restricted subset of species. During my thesis work, I have conducted comparative analysis of sexual traits of 145 *Macrostomum* species, thereby substantially expanding our knowledge on the macroevolutionary dynamics shaping these traits.

In Chapter II, I produced molecular resources for the genus *Macrostomum* by providing *de novo* assembled transcriptomes for three additional species with different mating syndromes. I showed that reproduction-related genes evolve at an accelerated rate. This adds to a growing body of evidence that reproduction-related genes evolve fast across metazoans (Swanson and Vacquier 2002; Wilburn and Swanson 2016). My work represents the first evidence for such rapid evolution in flatworms and in hermaphroditic animals in general. Because we expect hermaphrodites to be under intense sexual selection (Charnov 1979; Schärer and Janicke 2009; Anthes 2010; Schärer et al. 2015), this finding is not surprising, but the genomic resources I generated will now allow detection and detailed investigations of genes that evolve at a particularly fast rate.

To expand on the work of Chapter II, it would therefore be interesting to conduct more detailed molecular investigations of gene functions. The work presented in Chapter II, shows that in-situ hybridization is possible in species other than *M. lignano*, and most likely RNA interference could also be applied to generate knockdown phenotypes of candidate genes to study their function. Particularly interesting in this regard would be the identification of genes involved in spermatogenesis since this could allow manipulation of sperm design (Arbore et al. 2015). Such genetic manipulation could, for example, allow the generation of sperm without bristles in reciprocally mating species. This then will allow experimental tests of the prediction that the bristles are important for postcopulatory sexual selection (Vizoso et al. 2010; Schärer et al. 2011).

The work presented in Chapter III had three important aspects: First, through extensive field sampling in combination with a phylogenomic analysis, I massively expanded our knowledge of the biodiversity of the genus, by documenting 89 species that are new to science. Second, through detailed observations of sexual traits I documented their dynamic history across 145 species of *Macrostomum*. Finally, to construct the phylogeny I generated a large number of transcriptomes, which represent a resource that will be useful for comparative analysis of molecular evolution and detailed investigations of gene function.

Given the, to date, quite restricted sampling of the genus, it is not that surprising that much of the biodiversity of *Macrostomum* (and meiofaunal organisms in general) remains undiscovered (Kennedy and Jacoby 1999; Curini-Galletti et al. 2012). While field collections of flatworms might appear tedious, my findings of Chapter III demonstrate that it can be exceptionally rewarding, in terms of our understanding of biology. Collecting more species has revealed at least nine independent origins of hypodermic insemination and has shown that the true number of convergent events is very likely considerably higher still. Documenting these convergent events, did not only allow me to characterise them more precisely, but also offered insights into the intermediate stages of its evolution. My findings indicate that hypodermic insemination likely evolves through initial internal wounding, leading to traumatic insemination through the antrum wall. Such routes of transition have been previously suggested but, until now, there was little support for them (Lange et al. 2013a). Further, I showed in chapter III that hypodermic insemination is associated with changes in morphospace that tightly correspond to the initially described mating syndromes (Schärer et al. 2011). While comparative methods are necessarily correlational (Garamszegi 2014), documenting convergence across numerous independent evolutionary events makes it likely that the observed changes, in a whole suite of traits, truly represent adaptations to hypodermic insemination.

Chapter III has also demonstrated that the rewards of documenting new species does not only result in "more of the same". Instead, I have discovered modifications to sperm design that expand the morphospace of sperm in *Macrostomum*. For example, I have documented two species that lack visible bristles but mate reciprocally. This introduces complexity and nuance into the interpretation of the adaptive value of the bristles. The interpretation that the sperm bristles are a male persistence trait still seems likely, but apparently the bristles can also be lost due to factors other than the evolution of hypodermic insemination. This emphasises that the experimental studies this work will facilitate, are absolutely needed to further explore and validate the hypothesised function of the sperm bristles. I also discovered several monophyletic species in Lake Tanganyika, Zambia, that had an exaggerated sperm length. Sperm in these species is very long compared to the female antrum, suggesting a novel kind of donor – recipient interaction could be occurring there. Further, my work has revealed that a second female genital opening has evolved convergently at least four time within the genus, suggesting that this might be a trait evolving due to sexual conflict over received ejaculate. Due the robust phylogenomic framework I have created, the stage is now set for more in-depth investigations of these intriguing morphologies.

The phylogenomic work done for Chapter III, indicated that there is substantial gene tree – species tree conflict, possibly due to rapid speciation at the base of the main clades. Since I generated sets of orthologs between the transcriptomes of 98 species,

I will be able to investigate these conflicts in more detail. In particular, it will be possible to apply methods that can distinguish whether the gene tree – species tree conflict is due to incomplete lineage sorting (i.e. rapid speciation) or ancient hybridisation (Kubatko and Chifman 2015; Blischak et al. 2018). My phylogenomic analysis has also revealed that Clade 1—which contains only hypodermically mating species—is much more molecularly diverged than one could assume based on their relatively homogenous morphology alone. Hypodermic insemination seems to canalise the morphology of species and make species difficult to distinguish, which can also be seen by the convergence in morphospace of species with hypodermic insemination. Species in this clade are members of the interstitial meiofauna, occurring primarily on sheltered beaches and this makes them relatively easy to collect because their habitat is easily accessed. I found no strong geographic signal in their distribution, suggesting that many of them are globally distributed. Although, due to their cryptic nature, there is likely more population substructure that remains undiscovered. Since, we have been able to maintain many of these species in the laboratory quite easily, this clade would be a good candidate for population genetic investigations of cryptic speciation. Finally, the inferred phylogeny sets the basis for several changes in the taxonomic nomenclature since several species that were grouped into a different genus are clearly *Macrostomum*. The molecular data generated will thus also support future taxonomic investigations.

In chapter IV, I analysed the sex allocation of 120 *Macrostomum* species. I could show that hypodermic insemination is associated with a shift towards a more female-biased sex allocation. This is contrary to predictions that sex allocation should be more male-biased, in species with hypodermic insemination, because it could lead to a more fair-raffle like sperm competition (Schärer and Janicke 2009; Schärer et al. 2011). While hypodermic insemination likely reduces the recipients ability to control the fate of received sperm (Charnov 1979) and thus potentially removes a factor skewing sperm competition (van Velzen et al. 2009; Schärer and Pen 2013), this does not seem to be a dominating factor affecting sex allocation (but note that there are some hypodermically mating species with relatively more male-biased allocation). Instead, my results indicate that hypodermic insemination is associated with reduced heterozygosity, likely due to selfing. Selfing leads to intense local sperm competition and would thus favour the evolution of a more female-biased sex allocation (Charlesworth and Charlesworth 1981; Charnov 1987; Schärer 2009). A negative correlation between selfing and sex allocation, has been found in plants, where quite a few comparative analysis have been conducted (e.g. Plitmann and Levin 1990; Gallardo et al. 1994; Damgaard and Abbott 1995; Barrett et al. 1996; Jürgens et al. 2002; Galloni et al. 2007). In animals, data on this is rather sparse (Johnston et al. 1998; Winkler and Ramm 2018), although there is evidence that

sperm dependent parthenogenesis in planarian flatworms is associated with low investment into sperm production (Weinzierl et al. 1998, 1999).

My study in chapter IV explicitly modelled the evolution of sex allocation across the phylogeny - inferring the selective optima it likely evolves towards. I am not aware of any other study that has done something similar, most likely because it is necessary to have a robust phylogeny and a large sample size to be able to apply such methods. Quite a few comparative studies of sex allocation have been conducted in plants, but many of them have not taken phylogenetic relatedness into account (reviewed in Jong and Klinkhamer 2005; Sicard and Lenhard 2011). This is problematic since it can lead to pseudoreplication (Felsenstein 1985). With the increasing availability of phylogenetic information for plants, it would be interesting to apply similar methods to sex allocation in plants since there is a rich literature predicting the relationship between various traits and sex allocation, and many of these traits have already been measured (Campbell 2000; Dafni et al. 2000; de Jong and Klinkhamer 2005; Sicard and Lenhard 2011). Large datasets could also allow the fitting of more parameter rich models that go beyond the two optima models I applied in Chapter IV.

In chapter V, I presented three species descriptions that were made with the search for a new model organism in mind. We showed that the closest know relative of *M. lignano, M. janickei* shares its karyotype polymorphisms and is thus also unsuitable for genetic manipulation. We further showed that *M. mirumnovem* has an even less stable karyotype with 16 different karyotypes observed. The karyology of *M. mirumnovem* has been further investigated and is now studied as an example of tetraploidisation due to a hybridisation event (Zadesenets et al. 2020). Finally, we showed that the karyotype of *M. cliftonensis* is stable and that this species has a small genome, making it a good candidate for transgenesis.

In this thesis, I have applied the comparative method to the study of the genomic and morphological consequences of sexual selection in hermaphrodites. My work contributes to a growing body of evidence that sexual selection and in particular sexual conflict are important factors shaping the evolution of hermaphrodites. My work also highlights that it is worthwhile to investigate small invertebrate taxa that are often neglected. Traumatic insemination has been studied most extensively in insects and while great insights have been gained from these systems, they usually represent only one or two independent origins of the behaviour. Hypodermic insemination evolves frequently in *Macrostomum*, resulting in an intriguing level of macroevolutionary replication that will allow in-depth investigations of this mating strategy.

Acknowledgments

First and foremost, I thank my supervisor Lukas Schärer, who gave me the great opportunity to conduct this work. During the time in his group, I have grown as a scientist and learned from him not only about sexual selection but also how to critically evaluate scientific work. Some of my work lies outside his usual expertise and I immensely appreciate that he trusted me to acquire the necessary skills on my own and supported my journey into bioinformatics. It was fun to go field sampling with Lukas and share with him the joy of scientific discovery. These are memories that will always be precious to me. I am also grateful for his additional financial support when my project was taking longer than expected. Finally, I thank Lukas for supporting my decision to visit the laboratory of Luke Harmon, again showing the trust he put in me.

I want to thank Luke J. Harmon for hosting me for a research visit in Moscow, ID. I learned a lot from Luke about comparative methods in biology, but also about kindness and different perspectives in biology.

I thank the members of the biology department of the University of Idaho, for warmly welcoming me into their community. I want to thank especially David C. Tank, John M. Sullivan, Sarah Jacobs, Anahi Espindola, Megan Ruffley, Malia Santos, Sam McCauley and Austin Hanners.

I thank Nele Villabruna, Florian Christ and R. Axel W. Wiberg for helpful comments on parts of this thesis.

I thank Walter Salzburger for leading my defense exam and I thank Dieter Ebert for being the head of my thesis committee.

I thank John Fitzpatrick for being the external examiner for my thesis.

Many people have contribute to the field work here and I want to thank all of them: Peter Ladurner, Gregor Schulte, Christopher Laumer, Gudrun Viktorin, Walter Salzburger, Adrian Indermaur, Bernd Egger, Fabrizia Ronco, Heinz Büscher, Victoria Huwiler, Philipp Kaufmann, Michaela Zwyer, Stefanie von Fumetti, Joe Ryan, Mark Q. Martindale, Marta Chiodin, John Evans, Leigh Simmons, Mauro Tognon, Piero Tognon, Christiano Tognon, R. Axel W. Wiberg, Pragya Singh, Nikolas Vellnow, Christian Felber, Ulf Jondelius, Sarah Atherton, Tim Janicke, Georgina Rivera-Ingraham, Ben Byrne, Yvonne Gilbert and Rod Watson. I apologize, if I forgot anyone, please know that no malice is intended by the omission.

I received helpful logistical support from Gudrun Viktorin, Jürgen Hottinger, Urs Stiefel, Jasmin Picton, Lukas Zimmermann, Daniel Lüscher and Brigitte Aeschbach. Without them none of this work would be possible.

I thank Christel Genoud and Alexandra Graff Meyer from the Electron Microscopy Facility at the Friedrich Miescher Institute for their help and generosity when teaching me how to use their Serial Block-Face Machine.

I am grateful to Peter Ladurner, Julia Wunderer, Birgit Lengerer, Robert Petja and Wili Salvenmoser in Innsbruck for their support and generosity in teaching me about molecular biology and microscopy.

I want to thank all the members of the *Macrostomum* research community and in particular Peter Ladurner, Julia Wunderer, Birgit Lengerer, Robert Pjeta, Philip Berthemes, Wili Salvenmoser, Athina Giannakara, Steven A. Ramm, Michael Weber, Bahar Patlar, Jessica Abbott and Anna Norden, who have made every *Macrostomum* meeting feel like a family reunion.

I want to thank all members of the Schärer group I interacted with at the Institute: Gudrun Viktorin, Axel Wiberg, Pragya Singh, Elena Menéndez, Dita Vizoso, Jussi Lehtonen, Lucas Marie-Orleach, Nikolas Vellnow, Aline Schlatter, David Emde, Christian Felber, Daniel Neuckel and Philipp Kaufmann.

I thank all the members of the Zoological Institute for creating a stimulating work environment.

In thank my dear colleagues and office mates for their support and interesting discussions: Joana Rodrigues Lopes dos Santos, Camille Ameline, Cheng Choon Ang, Andrea Cabalzar, Eric Dexter, Luca Cornetti, Maridel Fredericksen, Peter Fields, Michelle Krebs, Fabrizia Ronco, Nicolas Lichili Ortiz and Michael Matschiner.

I am eternally grateful to have Florian Christ as a friend and flatmate. Not only was he always there when I needed someone to talk too, but he also tolerated my messiness - particularly during the last weeks of writing this thesis.

My Mother and Stepfather, Jeannette and Stefan, have always encouraged me to follow my curiosity and supported me on all steps of this journey. Thank you for always being there for me.

Last but not least, I thank Nele Villabruna for her love and support, her encouragement, and for always believing in me. It was incredibly enjoyable to share this PhD journey with you.

The thesis project was funded by grants of the Swiss National Science Foundation to Lukas Schärer (1003A_162543 and 310030_184916). During my thesis, I was also financially supported by the 'Reisefonds' of the University of Basel.

References

Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc. Natl. Acad. Sci. 106:13410–13414.

Andersson, M. B. 1994. Sexual selection. Princeton University Press, Princeton, N.J.

Angeloni, L. 2003. Sexual selection in a simultaneous hermaphrodite with hypodermic insemination: body size, allocation to sexual roles and paternity. Anim. Behav. 66:417–426.

Anthes, N. 2010. Mate choice and reproductive conflict in simultaneous hermaphrodites. Pp. 329–357 *in* P. Kappeler, ed. Animal Behaviour: Evolution and Mechanisms. Springer Berlin Heidelberg, Berlin, Heidelberg.

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

Anthes, N., A. Putz, and N. K. Michiels. 2006. Sex role preferences, gender conflict and sperm trading in simultaneous hermaphrodites: a new framework. Anim. Behav. 72:1–12.

Anthes, N., H. Schulenburg, and N. K. Michiels. 2008. Evolutionary links between reproductive morphology, ecology and mating behavior in opisthobranch gastropods. Evolution 62:900–916.

Apelt, G. 1969. Fortpflanzungsbiologie, Entwicklungszyklen und vergleichende Frühentwicklung acoeler Turbellarien. Mar Biol 4:59.

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

Arnold, S. J. 1994. Bateman's principles and the measurement of sexual selection in plants and animals. Am. Nat. 144:S126–S149.

Arnqvist, G. 1989. Sexual selection in a water strider: the function, mechanism of selection and heritability of a male grasping apparatus. Oikos 56:344–350.

Arnqvist, G., and L. Rowe. 2002a. Antagonistic coevolution between the sexes in a group of insects. Nature 415:787–789.

Arnqvist, G., and L. Rowe. 2002b. Correlated evolution of male and female morphologles in water striders. Evol. Int. J. Org. Evol. 56:936–947.

Arnqvist, G., and L. Rowe. 2005. Sexual conflict. Princeton University Press, Princeton, N.J.

Arnqvist, G., and L. Rowe. 1995. Sexual conflict and arms races between the sexes: a morphological adaptation for control of mating in a female insect. Proc. R. Soc. Lond. B Biol. Sci. 261:123–127.

Avila, F. W., L. K. Sirot, B. A. LaFlamme, C. D. Rubinstein, and M. F. Wolfner. 2010. Insect seminal fluid proteins: identification and function. Annu. Rev. Entomol. 56:21–40.

Barrett, S. C. H., L. D. Harder, and A. C. Worley. 1996. The comparative biology of pollination and mating in flowering plants. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 351:1271–1280.

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

Baur, B. 1998. Sperm competition in molluscs. Pp. 219–254 *in* Sperm Competition and Sexual Selection. Acad. Press, San Diego.

Beese, K., K. Beier, and B. Baur. 2006. Coevolution of male and female reproductive traits in a simultaneously hermaphroditic land snail. J. Evol. Biol. 19:410–418.

Birkhead, T. R., and A. P. Møller (eds). 1998. Sperm competition and sexual selection. Acad. Press, San Diego.

Birkhead, T. R., and T. Pizzari. 2002. Evolution of sex: postcopulatory sexual selection. Nat. Rev. Genet. 3:262–273.

Blischak, P. D., J. Chifman, A. D. Wolfe, L. S. Kubatko, and D. Posada. 2018. HyDe: a python package for genome-scale hybridization detection. Syst. Biol. 67:821–829.

Borkott, H. 1970. Geschlechtliche Organisation, Fortpflanzungsverhalten und Ursachen der sexuellen Vermehrung von *Stenostomum sthenum* nov. spec. (Turbellaria, Catenulida). Z Morph Tiere 67:183–262.

Campbell, D. R. 2000. Experimental tests of sex-allocation theory in plants. Trends Ecol. Evol. 15:227–232.

Charlesworth, D., and B. Charlesworth. 1981. Allocation of resources to male and female functions in hermaphrodites. Biol. J. Linn. Soc. 15:57–74.

Charnov, E. L. 1987. On sex allocation and selfing in higher plants. Evol. Ecol. 1:30–36.

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

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

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

Charnov, E. L. 1982. The theory of sex allocation. Monogr. Popul. Biol. 18:1–355.

Charnov, E. L., J. J. Bull, and J. Maynard Smith. 1976. Why be an hermaphrodite? Nature 263:125–126.

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

Cooney, C. R., J. A. Bright, E. J. R. Capp, A. M. Chira, E. C. Hughes, C. J. A. Moody, L. O. Nouri, Z. K. Varley, and G. H. Thomas. 2017. Mega-evolutionary dynamics of the adaptive radiation of birds. Nature 542:344–347.

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

Crudgington, H. S., and M. T. Siva-Jothy. 2000. Genital damage, kicking and early death. Nature 407:855–856.

Curini-Galletti, M., T. Artois, V. Delogu, W. H. D. Smet, D. Fontaneto, U. Jondelius, F. Leasi, A. Martínez, I. Meyer-Wachsmuth, K. S. Nilsson, P. Tongiorgi, K. Worsaae, and M. A. Todaro. 2012. Patterns of diversity in soft-bodied meiofauna: dispersal ability and body size matter. PLOS ONE 7:e33801.

Dafni, A., M. Hesse, and E. Pacini (eds). 2000. Pollen and Pollination. Springer Vienna, Vienna.

Damgaard, C., and R. J. Abbott. 1995. Positive correlations between selfing rate and pollen-ovule ratio within plant populations. Evolution 49:214–217.

Darwin, C. 1859. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. John Murray, 1859, London.

de Jong, T. J., and P. G. L. Klinkhamer. 2005. Evolutionary ecology of plant reproductive strategies. Cambridge University Press, Cambridge ; New York.

Eberhard, W. 1996. Female control: sexual selection by cryptic female choice. Princeton University Press, Princeton.

Edvardsson, M., and T. Tregenza. 2005. Why do male *Callosobruchus maculatus* harm their mates? Behav. Ecol. 16:788–793.

Edward, D. A., P. Stockley, and D. J. Hosken. 2015. Sexual conflict and sperm competition. Cold Spring Harb. Perspect. Biol. 7.

Elle, E., and T. R. Meagher. 2000. Sex allocation and reproductive success in the andromonoecious perennial *Solanum carolinense* (Solanaceae). ii. paternity and functional gender. Am. Nat. 156:622–636.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Fischer, E. A. 1981. Sexual allocation in a simultaneously hermaphroditic coral reef fish. Am. Nat. 117:64–82.

Fischer, E. A. 1980. The relationship between mating system and simultaneous hermaphroditism in the coral reef fish, *Hypoplectrus nigricans* (Serranidae). Anim. Behav. 28:620–633.

Fitzpatrick, J. L., M. Almbro, A. Gonzalez-Voyer, N. Kolm, and L. W. Simmons. 2012. Male contest competition and the coevolution of weaponry and testes in pinnipeds: reproductive trait evolution in pinnipeds. Evolution 66:3595–3604.

Fitzpatrick, J. L., R. Montgomerie, J. K. Desjardins, K. A. Stiver, N. Kolm, and S. Balshine. 2009. Female promiscuity promotes the evolution of faster sperm in cichlid fishes. Proc. Natl. Acad. Sci. U. S. A. 106:1128–1132.

Gage, M. J. 1994. Associations between body size, mating pattern, testis size and sperm lengths across butterflies. Proc. R. Soc. Lond. B Biol. Sci. 258:247–254.

Gallardo, R., E. Dominguez, and J. M. Muñoz. 1994. Pollen-ovule ratio, pollen size, and breeding system in *Astragalus* (Fabaceae) subgenus *Epiglottis*: a pollen and seed allocation approach. Am. J. Bot. 81:1611–1619.

Galloni, M., L. Podda, D. Vivarelli, and G. Cristofolini. 2007. Pollen presentation, pollen-ovule ratios, and other reproductive traits in Mediterranean Legumes (Fam. Fabaceae - Subfam. Faboideae). Plant Syst. Evol. 266:147–164.

Garamszegi, L. Z. (ed). 2014. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Springer Berlin Heidelberg, Berlin, Heidelberg.

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

Grudniewska, M., S. Mouton, D. Simanov, F. Beltman, M. Grelling, K. de Mulder, W. Arindrarto, P. M. Weissert, S. van der Elst, and E. Berezikov. 2016. Transcriptional signatures of somatic neoblasts and germline cells in *Macrostomum lignano*. eLife 5:e20607.

Harmon, L. 2019. Phylogenetic Comparative Methods. Open Textbook Library.

Harrison, P. W., A. E. Wright, F. Zimmer, R. Dean, S. H. Montgomery, M. A. Pointer, and J. E. Mank. 2015. Sexual selection drives evolution and rapid turnover of male gene expression. Proc. Natl. Acad. Sci. 112:4393–4398.

Harvey, P. H., and M. D. Pagel. 1991. The Comparative Method in Evolutionary Biology. Oxford University Press, Oxford ; New York.

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

Holland, B., and W. R. Rice. 1998. Perspective: chase-away sexual selection: antagonistic seduction versus resistance. Evolution 52:1–7.

Hosken, D. J., and C. M. House. 2011. Sexual selection. Curr. Biol. 21:R62–R65.

Hosken, D. J., P. Stockley, T. Tregenza, and N. Wedell. 2009. Monogamy and the battle of the sexes. Annu. Rev. Entomol. 54:361–378.

Hotzy, C., and G. Arnqvist. 2009. Sperm competition favors harmful males in seed beetles. Curr. Biol. 19:404–407.

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

Janicke, T., I. K. Häderer, M. J. Lajeunesse, and N. Anthes. 2016. Darwinian sex roles confirmed across the animal kingdom. Sci. Adv. 2:e1500983–e1500983.

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

Jennions, M. D., and H. Kokko. 2010. Sexual Selection. Pp. 343–364 *in* Evolutionary Behavioral Ecology. Oxford University Press, Oxford.

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

Johnson, M. G., and A. J. Shaw. 2016. The effects of quantitative fecundity in the haploid stage on reproductive success and diploid fitness in the aquatic peat moss *Sphagnum macrophyllum*. Heredity 116:523–530.

Johnston, M. O., B. Das, and W. R. Hoeh. 1998. Negative correlation between male allocation and rate of self-fertilization in a hermaphroditic animal. Proc. Natl. Acad. Sci. U. S. A. 95:617–620.

Johnstone, R. A., L. Keller, and A. E. M. Zuk. 2000. How males can gain by harming their mates: sexual conflict, seminal toxins, and the cost of mating. Am. Nat. 156:368–377.

Jürgens, A., Witt T., and Gottsberger G. 2002. Pollen grain numbers, ovule numbers and pollen-ovule ratios in Caryophylloideae: correlation with breeding system, pollination, life form, style number, and sexual system. Sex. Plant Reprod. 14:279–289.

Kamimura, Y. 2007. Twin intromittent organs of *Drosophila* for traumatic insemination. Biol. Lett. 3:401–404.

Kathirithamby, J., M. Hrabar, J. A. Delgado, F. Collantes, S. Dötterl, D. Windsor, and G. Gries. 2015. We do not select, nor are we choosy: reproductive biology of Strepsiptera (Insecta). Biol. J. Linn. Soc. 116:221–238.

Kennedy, A. D., and C. A. Jacoby. 1999. Biological indicators of marine environmental health: meiofauna – a neglected benthic component? Environ. Monit. Assess. 54:47–68.

Koene, J. M., and H. Schulenburg. 2005. Shooting darts: co-evolution and counter-adaptation in hermaphroditic snails. BMC Evol. Biol. 5:25.

Kokko, H., and M. D. Jennions. 2014. The relationship between sexual selection and sexual conflict. Cold Spring Harb. Perspect. Biol. 6:a017517–a017517.

Koyabu, D., H. Endo, C. Mitgutsch, G. Suwa, K. C. Catania, C. P. Zollikofer, S. Oda, K. Koyasu, M. Ando, and M. R. Sánchez-Villagra. 2011. Heterochrony and developmental modularity of cranial osteogenesis in lipotyphlan mammals. EvoDevo 2:21.

Kubatko, L., and J. Chifman. 2015. An invariants-based method for efficient identification of hybrid species from large-scale genomic data. bioRxiv 034348.

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

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

Lange, R., J. Werminghausen, and N. Anthes. 2013b. Cephalo-traumatic secretion transfer in a hermaphrodite sea slug. Proc. R. Soc. B Biol. Sci. 281:20132424–20132424.

Lange, R., Werminghausen Johanna, and Anthes Nils. 2014. Cephalo-traumatic secretion transfer in a hermaphrodite sea slug. Proc. R. Soc. B Biol. Sci. 281:20132424.

Lengerer, B., E. Hennebert, P. Flammang, W. Salvenmoser, and P. Ladurner. 2016. Adhesive organ regeneration in *Macrostomum lignano*. BMC Dev. Biol. 16:20.

Lengerer, B., R. Pjeta, J. Wunderer, M. Rodrigues, R. Arbore, L. Schärer, E. Berezikov, M. W. Hess, K. Pfaller, B. Egger, S. Obwegeser, W. Salvenmoser, and P. Ladurner. 2014. Biological adhesion of the flatworm *Macrostomum lignano* relies on a duo-gland system and is mediated by a cell type-specific intermediate filament protein. Front. Zool. 11:12.

Leonard, J. L. 2005. Bateman's principle and simultaneous hermaphrodites: a paradox. Integr. Comp. Biol. 45:856–873.

Leonard, J. L. 1999. Modern portfolio theory and the prudent hermaphrodite. Invertebr. Reprod. Dev. 36:129–135.

Leonard, J. L. 1993. Sexual conflict in simultaneous hermaphrodites: evidence from serranid fishes. Environ. Biol. Fishes 36:135–148.

Lessells, C. M. 1999. Sexual conflict in animals. Pp. 73–99 *in* L. Keller, ed. Levels of Selection in Evolution. University Press.

Lessells, C. M. 2005. Why are males bad for females? Models for the evolution of damaging male mating behavior. Am. Nat. 165:S46–S63.

Lodi, M., and J. M. Koene. 2016. On the effect specificity of accessory gland products transferred by the love-dart of land snails. BMC Evol. Biol. 16:104.

Losos, J. B. 1998. Contingency and determinism in replicated adaptive radiations of island lizards. Science 279:2115–2118.

Louca, S., and M. W. Pennell. 2020. Extant timetrees are consistent with a myriad of diversification histories. Nature 580:502–505.

Louca, S., P. M. Shih, M. W. Pennell, W. W. Fischer, L. W. Parfrey, and M. Doebeli. 2018. Bacterial diversification through geological time. Nat. Ecol. Evol. 2:1458–1467.

Maddison, W. P., and R. G. FitzJohn. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol. 64:127–136.

Mahler, D. L., T. Ingram, L. J. Revell, and J. B. Losos. 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. Science 341:292–295.

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

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

Martín-Durán, J. M., K. Pang, A. Børve, H. S. Lê, A. Furu, J. T. Cannon, U. Jondelius, and A. Hejnol. 2018. Convergent evolution of bilaterian nerve cords. Nature 553:45–50.

Maurer, G., M. C. Double, O. Milenkaya, M. Süsser, and R. D. Magrath. 2011. Breaking the rules: sex roles and genetic mating system of the pheasant coucal. Oecologia 167:413–425.

McCartney, M. A. 1997. Sex allocation and male fitness gain in a colonial, hermaphroditic marine invertebrate. Evolution 51:127–140.

Meredith, R. W., J. E. Janečka, J. Gatesy, O. A. Ryder, C. A. Fisher, E. C. Teeling, A. Goodbla, E. Eizirik, T. L. L. Simão, T. Stadler, D. L. Rabosky, R. L. Honeycutt, J. J. Flynn, C. M. Ingram, C. Steiner, T. L. Williams, T. J. Robinson, A. Burk-Herrick, M. Westerman, N. A. Ayoub, M. S. Springer, and W. J. Murphy. 2011. Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334:521–524.

Michiels, N. K. 1998. Mating conflicts and sperm competition in simultaneous hermaphrodites. Pp. 219–254 *in* Sperm Competition and Sexual Selection. Elsevier.

Michiels, N. K., and J. M. Koene. 2006. Sexual selection favors harmful mating in hermaphrodites more than in gonochorists. Integr. Comp. Biol. 46:473–480.

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

Morrow, E. H., G. Arnqvist, and S. Pitnick. 2003. Adaptation versus pleiotropy: why do males harm their mates? Behav. Ecol. 14:802–806.

Mouton, S., M. Grudniewska, L. Glazenburg, V. Guryev, and E. Berezikov. 2018. Resilience to aging in the regeneration-capable flatworm *Macrostomum lignano*. Aging Cell 17:e12739.

Mouton, S., M. Willems, P. Back, B. P. Braeckman, and G. Borgonie. 2009. Demographic analysis reveals gradual senescence in the flatworm *Macrostomum lignano*. Front. Zool. 6:15.

Myers, R. J. 1935. Behavior and morphological changes in the leech, *Placobdella parasitica*, during hypodermic insemination. J. Morphol. 57:617–653.

Nakadera, Y., and J. M. Koene. 2013. Reproductive strategies in hermaphroditic gastropods: conceptual and empirical approaches. Can. J. Zool. 91:367–381. NRC Research Press.

Nakadera, Y., E. M. Swart, J. N. A. Hoffer, O. den Boon, J. Ellers, and J. M. Koene. 2014. Receipt of seminal fluid proteins causes reduction of male investment in a simultaneous hermaphrodite. Curr. Biol. 24:859–862.

Parker, G. A. 1979. Sexual selection and sexual conflict. Sex. Sel. Reprod. Compet. Insects 123–166.

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

Parker, G. A. 1998. Sperm competition and the evolution of ejaculates: towards a theory base. Pp. 3–54 *in* Sperm competition and sexual selection. Academic Press, San Diego.

Parker, G. A. 2014. The sexual cascade and the rise of pre-ejaculatory (Darwinian) sexual selection, sex roles, and sexual conflict. Cold Spring Harb. Perspect. Biol. 6:a017509–a017509.

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

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

Picchi, L., G. Cabanes, C. Ricci-Bonot, and M. C. Lorenzi. 2018. Quantitative matching of clutch size in reciprocating hermaphroditic worms. Curr. Biol. 28:3254-3259.e3.

Pitnick, S., and T. A. Markow. 1994. Male Gametic Strategies: Sperm Size, Testes Size, and the Allocation of Ejaculate Among Successive Mates by the Sperm-Limited Fly Drosophila pachea and Its Relatives. Am. Nat. 143:785–819.

Pitnick, S., T. Marrow, and G. S. Spicer. 1999. Evolution of multiple kinds of female sperm-storage organs in *Drosophila*. Evolution 53:1804–1822.

Plitmann, U., and D. A. Levin. 1990. Breeding systems in the *Polemoniaceae*. Plant Syst. Evol. 170:205–214.

Rabosky, D. L., and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Syst. Biol. 64:340–355.

Rademaker, M. C. J., and T. J. de Jong. 1998. Effects of flower number on estimated pollen transfer in natural populations of three hermaphroditic species: an experiment with fluorescent dye. J. Evol. Biol. 11:623–641.

Rademaker, M. C. J., and T. J. Jong. 1999. The shape of the female fitness curve for *Cynoglossum officinale*: quantifying seed dispersal and seedling survival in the field. Plant Biol. 1:351–356.

Ramm, S. A., and P. Stockley. 2010. Sperm competition and sperm length influence the rate of mammalian spermatogenesis. Biol. Lett. 6:219–221.

Ramm, S. A., D. B. Vizoso, and L. Schärer. 2012. Occurrence, costs and heritability of delayed selfing in a free-living flatworm. J. Evol. Biol. 25:2559–2568.

Reinhardt, K., N. Anthes, and R. Lange. 2015. Copulatory wounding and traumatic insemination. Cold Spring Harb. Perspect. Biol. 7:a017582.

Reinhardt, K., R. Naylor, and M. T. Siva-Jothy. 2003. Reducing a cost of traumatic insemination: female bedbugs evolve a unique organ. Proc. R. Soc. B Biol. Sci. 270:2371–2375.

Rodríguez-Martínez, H., U. Kvist, J. Ernerudh, L. Sanz, and J. J. Calvete. 2011. Seminal plasma proteins: what role do they play? Am. J. Reprod. Immunol. 66:11–22.

Rowe, L., G. Arnqvist, A. Sih, and J. J.Krupa. 1994. Sexual conflict and the evolutionary ecology of mating patterns: water striders as a model system. Trends Ecol. Evol. 9:289–293.

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

Schärer, L., J. N. Brand, P. Singh, K. S. Zadesenets, C.-P. Stelzer, and G. Viktorin. 2020. A phylogenetically informed search for an alternative *Macrostomum* model species, with notes on taxonomy, mating behavior, karyology, and genome size. J. Zool. Syst. Evol. Res. 58:41–65.

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

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

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

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

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

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

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

Shine, R., M. Wall, T. Langkilde, and R. T. Mason. 2005. Do female garter snakes evade males to avoid harassment or to enhance mate quality? Am. Nat. 165:660–668.

Sicard, A., and M. Lenhard. 2011. The selfing syndrome: a model for studying the genetic and evolutionary basis of morphological adaptation in plants. Ann. Bot. 107:1433–1443. Oxford Academic.

Simmons, L. W. 2001. Sperm competition and its evolutionary consequences in the insects. Princeton University Press, Princeton, N.J.

Singh, P., D. Ballmer, M. Laubscher, and L. Schärer. 2019a. Successful mating and hybridisation in two closely related flatworm species despite significant differences in reproductive morphology and behaviour. Animal Behavior and Cognition.

Singh, P., and L. Schärer. in preparation. Evolution of sex allocation plasticity and its predictors in a flatworm genus.

Singh, P., N. Vellnow, and L. Schärer. 2019b. Variation in sex allocation plasticity in three closely related flatworm species. Ecol. Evol. ece3.5566.

Sirot, L. K., and M. F. Wolfner. 2015. Who's zooming who? Seminal fluids and cryptic female choice in diptera. Pp. 351–384 *in* A. V. Peretti and A. Aisenberg, eds. Cryptic Female Choice in Arthropods: Patterns, Mechanisms and Prospects. Springer International Publishing, Cham.

Sirot, L. K., A. Wong, T. Chapman, and M. F. Wolfner. 2015. Sexual conflict and seminal fluid proteins: a dynamic landscape of sexual interactions. Cold Spring Harb. Perspect. Biol. 7:a017533.

Siva-Jothy, M. t. 2006. Trauma, disease and collateral damage: conflict in cimicids. Philos. Trans. R. Soc. B Biol. Sci. 361:269–275.

Stewart, M. J., T. Wang, J. M. Koene, K. B. Storey, and S. F. Cummins. 2016. A "love" dart allohormone identified in the mucous glands of hermaphroditic land snails. J. Biol. Chem. 291:7938–7950.

Stockley, P. 2002. Sperm competition risk and male genital anatomy: comparative evidence for reduced duration of female sexual receptivity in primates with penile spines. Evol. Ecol. 16:123–137.

Sturtevant, A. H. 1925. The seminal receptacles and accessory glands of the diptera, with special reference to the acalypterae. J. N. Y. Entomol. Soc. 33:195–215.

Swanson, W. J., and V. D. Vacquier. 2002. The rapid evolution of reproductive proteins. Nat. Rev. Genet. 3:137–144.

Tatarnic, N. J., G. Cassis, and M. T. Siva-Jothy. 2014. Traumatic insemination in terrestrial arthropods. Annu. Rev. Entomol. 59:245–261.

Throckmorton, L. H. 1975. The phylogeny, ecology, and geography of *Drosophila*. Pp. 421–469 *in* Handbook of Genetics, Vol. 3. Plenum Publishing Corporation, New York, NY.

Throckmorton, L. H. 1966. The relationships of the endemic hawaiian Drosophilidae. Pp. 335–396 *in* Studies in Genetics. Univ. Tex. Publ.

Tonnabel, J., P. David, and J. R. Pannell. 2019. Do metrics of sexual selection conform to Bateman's principles in a wind-pollinated plant? Proc. R. Soc. B Biol. Sci. 286:20190532.

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

Uyeda, J. C., M. W. Pennell, E. T. Miller, R. Maia, and C. R. McClain. 2017. The evolution of energetic scaling across the vertebrate tree of life. Am. Nat. 190:000–000.

Uyeda, J. C., R. Zenil-Ferguson, and M. W. Pennell. 2018. Rethinking phylogenetic comparative methods. Syst. Biol. 67:1091–1109.

van Velzen, E., L. Scharer, and I. Pen. 2009. The effect of cryptic female choice on sex allocation in simultaneous hermaphrodites. Proc. R. Soc. B Biol. Sci. 276:3123–3131.

Vellnow, N., D. B. Vizoso, G. Viktorin, and L. Schärer. 2017. No evidence for strong cytonuclear conflict over sex allocation in a simultaneously hermaphroditic flatworm. BMC Evol. Biol. 17.

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

Weber, M., B. Patlar, and S. A. Ramm. 2020. Effects of two seminal fluid transcripts on post-mating behaviour in the simultaneously hermaphroditic flatworm *Macrostomum lignano*. J. Evol. Biol. jeb.13606.

Wedell, N., and D. J. Hosken. 2010. The evolution of male and female internal reproductive organs in insects. P. *in* J. L. Leonard and A. Córdoba-Aguilar, eds. The evolution of primary sexual characters in animals. Oxford University Press, Oxford ; New York.

Weinzierl, R. P., K. Berthold, L. W. Beukeboom, and N. K. Michiels. 1998. Reduced male allocation in the parthenogenetic hermaphrodite *Dugesia polychroa*. Evolution 52:109–115.

Weinzierl, R. P., P. Schmidt, and N. K. Michiels. 1999. High fecundity and low fertility in parthenogenetic planarians. Invertebr. Biol. 118:87.

Whitman, C. O. 1891. Spermatophores as a means of hypodermic impregnation. J. Morphol. 4:361–406.

Wilburn, D. B., and W. J. Swanson. 2016. From molecules to mating: rapid evolution and biochemical studies of reproductive proteins. J. Proteomics 135:12–25.

Wilkinson, G. S., F. Breden, J. E. Mank, M. G. Ritchie, A. D. Higginson, J. Radwan, J. Jaquiery, W. Salzburger, E. Arriero, S. M. Barribeau, P. C. Phillips, S. C. P. Renn, and L. Rowe. 2015. The locus of sexual selection: moving sexual selection studies into the post-genomics era. J. Evol. Biol. 28:739–755.

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

Wunderer, J., B. Lengerer, R. Pjeta, P. Bertemes, L. Kremser, H. Lindner, T. Ederth, M. W. Hess, D. Stock, W. Salvenmoser, and P. Ladurner. 2019. A mechanism for temporary bioadhesion. Proc. Natl. Acad. Sci. 116:4297–4306.

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

Yund, P. O., and M. A. McCartney. 1994. Male reproductive success in sessile invertebrates: competition for fertilizations. Ecology 75:2151.

Zadesenets, K. S., I. Y. Jetybayev, L. Schärer, and N. B. Rubtsov. 2020. Genome and karyotype reorganization after whole genome duplication in free-living flatworms of the genus *Macrostomum*. Int. J. Mol. Sci. 21:680.

Zadesenets, K. S., D. B. Vizoso, A. Schlatter, I. D. Konopatskaia, E. Berezikov, L. Schärer, and N. B. Rubtsov. 2016. Evidence for karyotype polymorphism in the free-living flatworm, *Macrostomum lignano*, a model organism for evolutionary and developmental biology. Plos One 11:e0164915.