Decoupled evolution of the *Sex Peptide* gene family and *Sex Peptide Receptor* in *Drosophilidae*

Ben R. Hopkins[1*], Aidan Angus-Henry[1], Bernard Y. Kim[2], Jolie A. Carlisle[3], Ammon Thompson[1], Artyom Kopp[1]

[1] Department of Evolution and Ecology, University of California – Davis, CA, USA
[2] Department of Biology, Stanford University, CA, USA
[3] Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

*Corresponding author: brhopkins@ucdavis.edu

Abstract
Across internally fertilising species, males transfer proteins in their ejaculate that trigger wide-ranging changes in female behaviour and physiology. Much theory has been developed to explore the drivers of ejaculate protein evolution. The accelerating availability of high-quality genomes now allows us to test how these proteins are evolving at fine taxonomic scales. Here, we use genomes from 199, mostly drosophilid, species to chart the evolutionary history of Sex Peptide (SP), a potent regulator of female post-mating responses. We infer that *SP* has followed markedly different evolutionary trajectories in different lineages. Outside of the *Sophophora-Lordiphosa* radiation, *SP* exists largely as a single copy gene that has been independently lost in several lineages. In contrast, within the *Sophophora-Lordiphosa* radiation *SP* has repeatedly and independently duplicated. Up to seven copies, collectively displaying extensive variation in sequence, are present in some species. We use cross-species RNA-seq data to provide evidence that this lineage-specific burst in evolutionary activity did not follow a significant shift in the sex- or tissue-specificity of *SP's* expression. We also document considerable interspecific variation in accessory gland microcarriers that appears to be independent of *SP* presence or sequence. We end by showing that the mode of SP's evolution is decoupled from that of its receptor, *SPR*, in which we detect no evidence of correlated diversifying selection in its coding sequence. Collectively, our work describes the divergent evolutionary trajectories that an apparently novel drosophilid gene has followed in different branches of the phylogeny and finds a surprisingly weak coevolutionary signal between a supposedly sexually antagonistic protein and its receptor.

Introduction
Female post-mating changes are a taxonomically widespread – if not general – phenomenon in internal fertilisers. Often mediated by non-sperm components of the male ejaculate, such as seminal fluid proteins, the female traits subject to post-mating plasticity are numerous and diverse: immune systems can be modified (*Drosophila melanogaster*, Schwenke & Lazzaro, 2017; humans: Abu-Raya *et al.*, 2020), ovulation stimulated (camelids, Berland *et al.*, 2016), and dietary preferences shifted following copulation (crickets, Tsukamoto *et al.*, 2014; *D. melanogaster*, Walker, Corrales-Carvajal, & Ribeiro,

2015). Evolutionary biologists have a long-standing interest in post-mating changes as they bear intimate connections to reproductive success, can form barriers to hybridisation (e.g. Patterson, 1946; Ahmed-Braimah, Wolfner, & Clark, 2021), and, through the involvement of males in their induction, act as a point of evolutionary tension between the fitness interests of males and females (Rice, 1996; Holland & Rice, 1998; Civetta & Clark, 2000; Hollis *et al.*, 2019). Indeed, post-mating changes have provided one of the centrepieces around which much of the discussion on interlocus sexual conflict has revolved (see Hopkins & Perry, 2022 for historical details). However, while theory on the evolution of ejaculate-female reproductive interactions is well developed (reviewed in Arnqvist & Rowe, 2005; Rice & Gavrilets, 2014), relatively few empirical studies have mapped the origin and diversification of seminal regulators of post-mating responses at fine taxonomic scales.

Available data suggest that different taxa can use non-homologous proteins to induce common – or at least overlapping – phenotypic endpoints in mated females. To reduce female sexual receptivity, for example, the moth *Helicoverpa zea* uses pheromonostatic peptide (PSP, Kingan *et al.*, 1995), the mosquito *Aedes aegypti* uses Head Protein 1 (HP-1, Duvall *et al.*, 2017), and *D. melanogaster* uses Sex Peptide (SP, Chen *et al.*, 1988). Not only are these three proteins non-homologous to one another, but none have clear homologs in either of the other species' genomes. This pattern suggests that regulators of female post-mating change might experience a high degree of evolutionary turnover, with new regulators evolving and old regulators being lost from populations through time. If so, the questions then are what evolutionary forces drive this turnover, how quickly does this process occur, and how are new regulators born? More fundamentally, it makes seminal proteins an exceptional model for studying how newly evolved, lineage-specific genes acquire and diversify their functions.

More than 60 years on from its discovery in chromatographic extracts of *D. melanogaster* accessory glands (Fox, 1956; Chen & Diem, 1961), SP remains the best characterised insect seminal protein. Consisting of two exons separated by a 65bp intron, *SP* encodes a 36aa mature protein synthesised via a 55aa signal peptide-containing precursor (Chen *et al.*, 1988; Tsuda *et al.*, 2015). SP is specifically produced in the main cells of the accessory gland, from where it is secreted into the gland's lumen (Styger, 1992; Majane, Cridland, & Begun, 2022). Within the lumen, SP is stored on abundant, lipid-containing microcarriers – structures for which SP acts as a key factor governing assembly and, once inside females, disassembly (Wainwright *et al.*, 2021). Following transfer to females during mating, SP binds to sperm, a process mediated by a suite of additional seminal fluid proteins, and is trafficked into the female sperm storage organs (Peng *et al.*, 2005a; Ravi Ram & Wolfner, 2007, 2009; Gligorov *et al.*, 2013; Findlay *et al.*, 2014; Sitnik *et al.*, 2016; Singh *et al.*, 2018; Misra & Wolfner, 2020). The gradual release of SP from the surface of stored sperm continues to stimulate a wide range of post-mating changes, including shifts in memory formation and sleep patterns, elevating appetite and changing dietary preferences, reducing sexual receptivity, stimulating egg-laying, increasing aggression, and changing gut, metabolic, and immune activity (reviewed in Hopkins & Perry, 2022). At least some of these changes, namely reduced sexual receptivity and stimulated egg-laying, are mediated by SP binding to the Sex Peptide Receptor (SPR) in

a subset of neurons that innervate the female reproductive tract (Yapici *et al.*, 2008; Häsemeyer *et al.*, 2009; Rezával *et al.*, 2012). Different domains of mature SP appear to contribute selectively to different functions in *D. melanogaster*: the tryptophan-rich N-terminus binds to sperm and stimulates juvenile hormone synthesis (Moshitzky *et al.*, 1996; Fan *et al.*, 2000; Peng *et al.*, 2005a), the hydroxyproline-rich mid-section elicits the innate immune response (Domanitskaya *et al.*, 2007), and, through interactions with SPR, the disulphide bridge-containing C-terminus stimulates the core post-mating responses of increased oviposition and reduced sexual receptivity (Schmidt *et al.*, 1993; Ding *et al.*, 2003; Moehle *et al.*, 2011). Consequently, different portions of the *SP* coding sequence are likely to be evolving in response to different selective pressures.

SP is not the only member of its gene family present in *D. melanogaster.* This species also encodes the paralogous *Dup99b,* with which SP shares a high degree of similarity in the amino acid sequence of the C-terminus (Saudan *et al.*, 2002). Collectively, *SP* and *Dup99b* represent the only members of the 'sex-peptide pheromone family' present in *D. melanogaster* (Saudan *et al.*, 2002; Ding *et al.*, 2003). Both stimulate the core post-mating responses of increased oviposition and reduced sexual receptivity, but SP appears to be the 'key player' showing a higher binding affinity for the female reproductive tract and nervous system (Ding *et al.*, 2003) and, in *in vitro* assays, activating SPR at lower concentrations than does Dup99b (Yapici *et al.*, 2008). There are further differences between the paralogs, too. While *SP* is expressed in accessory gland main cells, *Dup99b* is expressed in the ejaculatory duct (Rexhepaj *et al.*, 2003). And unlike SP, the N-terminus of Dup99b does not stimulate juvenile hormone synthesis (Fan *et al.*, 2000). Thus, SP and Dup99b show partial redundancy but different sensitivities within one region of the protein and distinct activities in other regions, suggesting a degree of functional separation between the two paralogs.

SP occupies an important place in contemporary evolutionary biology, having emerged as one of the preeminent systems for experimental work on the genetic basis and fitness effects of sexual conflict (Wigby & Chapman, 2005; Wigby *et al.*, 2009; Fricke, Bretman, & Chapman, 2010; Wensing & Fricke, 2018). However, comparative data on how SP sequence and function has evolved and diversified through time is sparse by comparison. Available data suggest three features of SP's evolution. First, *SP* genes appear to be restricted to the genus *Drosophila: SP* has been traced as far back in the drosophilid phylogeny as *D. virilis,* but appears to be absent from mosquitoes and representative genomes from other insect families (Kim *et al.*, 2010; Tsuda *et al.*, 2015). This contrasts with its receptor *SPR*, which is deeply conserved among members of the Ecdysozoa and Lophotrochozoa. The deep conservation of SPR is partly explained by its interactions with a similarly well conserved class of alternative ligands, the myoinhibitory peptides (MIPs) (Kim *et al.*, 2010). MIP-SPR interactions are known to regulate diverse behaviours across species, including regulating larval settlement behaviour in marine annelids (Conzelmann *et al.*, 2013). In *Drosophila,* MIP-SPR interactions appear to be neither necessary nor sufficient for driving post-mating changes in females (Kim *et al.*, 2010; Poels *et al.*, 2010), but they do fulfil other functions, including regulating sleep behaviour (Oh *et al.*, 2014). Second, *SP* genes have not been detected in genomes from three members of the *Drosophila* genus (*D. grimshawi, D. albomicans,* and *D. mojavensis*), suggesting that *SP*

may have been lost in some drosophilid lineages (Kim *et al.*, 2010; Tsuda *et al.*, 2015; McGeary & Findlay, 2020). Third, robust expression of *SPR* in the female reproductive tract, the ability of SP (derived from the *D. melanogaster* sequence) to bind to female reproductive tract tissue, and a reduction in sexual receptivity upon injection of conspecific SP appear to be restricted to the *melanogaster* group (Tsuda *et al.*, 2015). Thus, despite the presence of SP orthologs beyond the group, many of the defining features of SP in *D. melanogaster* appear to be recently derived.

Here, we take advantage of a surge in the number of high-quality drosophilid genomes to map the origin and diversification of the *Sex Peptide* gene family. Using a comparative genomics approach across >200 Dipteran genomes, we find that *SP* predates the *Drosophilini* tribe and shows lineage-accelerated evolution in sequence, copy number, and genomic position in the *melanogaster* group. The propensity for repeated, independent expansions in *SP* copy number in this lineage contrasts with multiple losses of *SP* elsewhere in the *Drosophilidae*. At the coding level, we show that change in SP's sequence is elevated in the sperm-binding N-terminus and immune-stimulating midsection relative to the post-mating response stimulating C-terminus. Using comparative RNA-seq data from 19 species, we show that despite substantial change in sequence, copy number, and genomic position, *SP* remains a male reproductive tract-biased gene. Staining accessory glands from a diverse set of 18 species, we document considerable inter-specific diversity in the morphology of seminal microcarriers, but fail to find support for the hypothesis that change in SP sequence is a driver of evolutionary change in microcarrier morphology (Wainwright *et al.*, 2021). Finally, we show that the mode of evolutionary change between *SP* and its receptor, *SPR*, is decoupled: expansions of *SP* copy number or branches exhibiting extensive change in SP sequence are associated neither with duplication/loss of *SPR,* nor a significant signature of diversifying selection in its coding sequence. Collectively, our data provides a comprehensive characterisation of the divergent evolutionary trajectories that a lineage-specific master regulator of female post-mating responses has followed across the drosophilid phylogeny.

Results

Sex Peptide predates the *Drosophilini*

Falling within the Brachycera suborder and estimated to have originated at least 30-50 million years ago, the family *Drosophilidae* contains more than 4400 extant species distributed across two sub-families (the *Drosophilinae* Rondani and *Steganinae* Hendel)(Wiegmann *et al.*, 2011; Finet *et al.*, 2021). To identify where *SP* first evolved, we began by searching several non-drosophilid genomes from across the Brachycera, including *Musca domestica* and members of several lineages closely related to *Drosophilidae* (leaf miners: *Liriomyza trifolii*; tsetse flies: *Glossina morsitans*; and shore flies: *Ephydra gracilis* and *Cirrula hians*)(Vicoso & Bachtrog, 2015). We failed to detect *SP* in these species, as well as in two members of the *Drosophilidae* that fall outside of the *Drosophilinae, Phortica variegata* and *Leucophenga varia.* Among three early branching members of the *Drosophilinae*, we detected *SP* in two: *Scaptodrosophila lebanonensis,* which bore 1 copy, and *Scaptodrosophila latifasciaeformis*, which bore 3 tandemly arranged copies (Figure 1). However, we failed to detect *SP* in the third species,

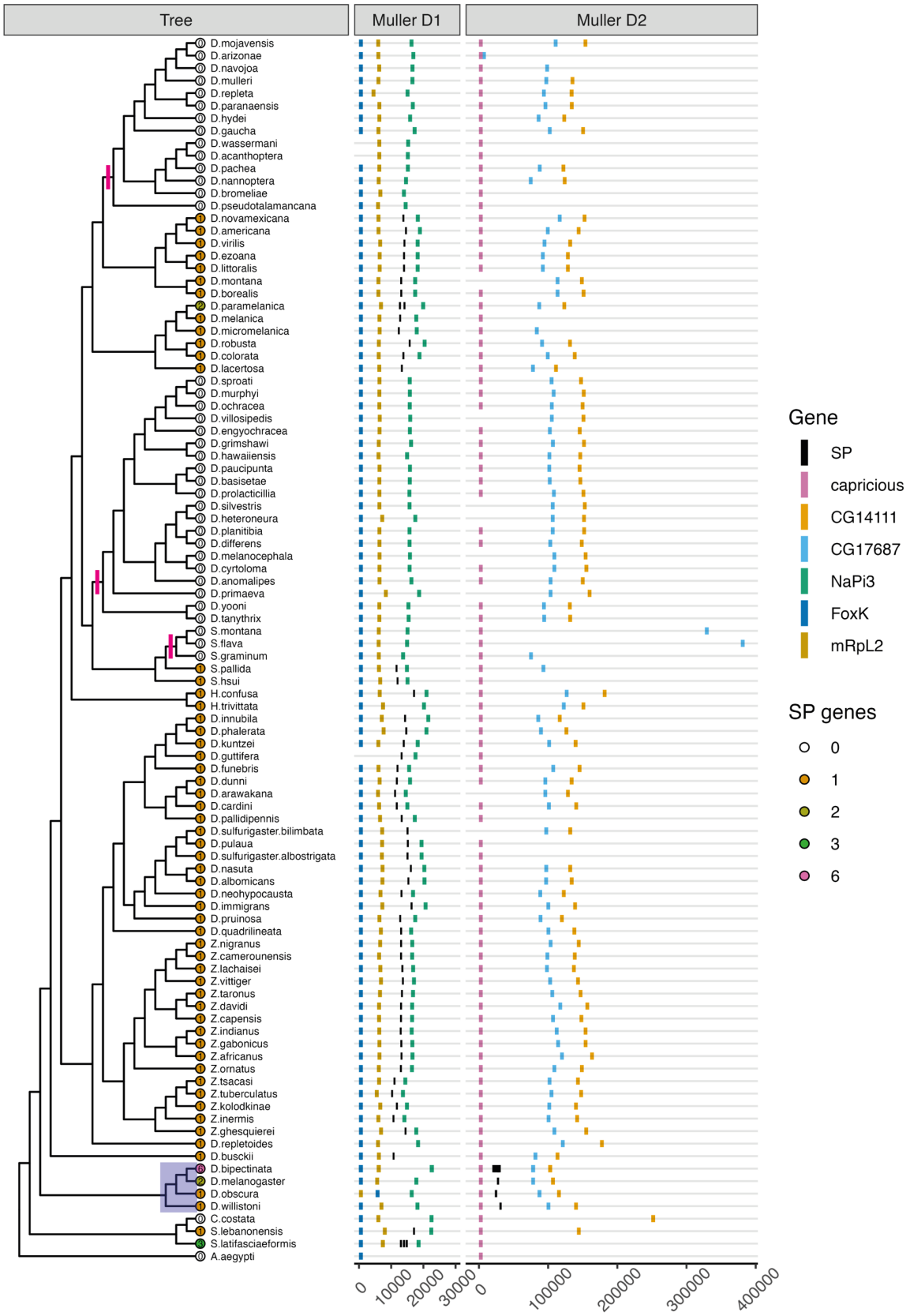

Figure 1. *Sex Peptide* family genes predate the *Drosophilini* and have been repeatedly lost outside of the *Sophophora-Lordiphosa* radiation. This figure focuses on members of the

Drosophilini tribe that are outside of the *Sophophora* subgenus to which *D. melanogaster* belongs. A selection of *Sophophora* species, shaded in the tree in blue, are included for comparison. Also included are four non-*Drosophila* dipterans: *Aedes aegypti,* a representative of the Nematocera suborder, and three members of the *Drosophilinae* subfamily that do not belong to the *Drosophila* genus or *Drosophilini* tribe. These are *Chymomyza costata, Scaptodrosophila lebanonensis,* and *S. latifasciaeformis,* which all belong to the *Colocasiomyini* tribe. The number of *SP* genes detected in a representative of each species' genome is given at the tree tips. Losses are marked with a pink bar. For each species, the structures of two syntenic gene neighbourhoods are plotted. The first, Muller D1, is the canonical neighbourhood in which *SP* genes are detected outside of the *Sophophora.* The second, Muller D2, is the canonical position in the *Sophophora.* Positions of each gene are given relative to the first gene in the neighbourhood (*FoxK* or *capricious*). Absence of a flanking neighbourhood gene (*e.g., FoxK* in *D. wassermani*) doesn't necessarily mean the gene has been lost – it more likely means that a contig breakpoint fell within the neighbourhood. In such cases, the next gene in the sequence was aligned to the position of its ortholog in the closest relative with a complete neighbourhood sequence. Note that *SP* in *Hirtodrosophila trivittata* and, independently, *D. repletoides,* has translocated out of the Muller D1 neighbourhood. While our phylogeny based on 250 orthologs placed *C. costata* within the *Scaptodrosophila* lineage, a published 17-locus phylogeny with denser taxon sampling in this part of the tree has *C. costata* branching earlier than the *Scaptodrosophila* (Finet *et al.*, 2021). Thus, the absence of *SP* in this species may represent ancestral absence rather than secondary loss.

Chymomyza costata. In our phylogeny, which was based on 250 randomly selected orthologs from the genomes, *C. costata* and *Scaptodrosophila* formed a monophyletic sister group to the *Drosophilini* tribe, a radiation of drosophilids that includes the genera *Drosophila*, *Lordiphosa*, *Zaprionus,* and *Scaptomyza.* However, a recent phylogeny with considerably denser taxon sampling at the base of the *Drosophilinae*, but which was based on only 17 genes, positioned *Chymomyza* as an earlier branching lineage than the *Scaptodrosophila* (Finet *et al.*, 2021). Therefore, the absence of *SP* in *C. costata* may be ancestral rather than the secondary loss our phylogeny suggests.

In both *Scaptodrosophila* species, we detected *SP* in a different genomic neighbourhood to where it is found in *D. melanogaster* (Figure 1). In *Scaptodrosophila*, *SP* genes fell within a syntenic neighbourhood that contained orthologs of *FoxK, NaPi-III,* and *mRpL2* – the same position in which we detect *SP* in *D. virilis,* the most distant relative of *D. melanogaster* previously reported to encode *SP* (Kim *et al.*, 2010; Tsuda *et al.*, 2015). In *D. melanogaster* by contrast, *SP* is positioned alongside *capricious, CG14111,* and *CG17687.* Both the *FoxK* and *capricious* neighbourhoods map to Muller element D in *D. melanogaster* but are separated by ~2.1Mb. We therefore name the *FoxK* neighbourhood 'Muller D1' and the *capricious* neighbourhood 'Muller D2'.

Based on the available taxon sample, our data collectively suggest that *SP* originated within the *Drosophilidae*; was present in the most recent common ancestor of *D. melanogaster, D.virilis,* and *Scaptodrosophila*; and has translocated to a new genomic location at some point in the lineage leading to *D. melanogaster*.

Sex Peptide **has been repeatedly and independently lost outside of the** ***Sophophora-Lordiphosa*** **radiation**

After the branching of the *Scaptodrosophila*, the *Drosophilinae* branches into two monophyletic lineages that form part of the *Drosophilini* tribe. The first contains the *Lordiphosa* genus and *Sophophora* subgenus (which we collectively refer to as the '*Sophophora-Lordiphosa* radiation'), the latter of which includes the *melanogaster, obscura, willistoni,* and *saltans* groups (see Supplementary figure 1 for an overview of the *Sophophora* taxonomic terminologies used in this paper). The second lineage includes, among others, the genera *Scaptomyza* and *Zaprionus* and the Hawaiian*, virilis, repleta, immigrans, cardini,* and *quinaria* groups. The *Drosophila* genus is therefore paraphyletic (Yassin, 2013; Finet *et al.*, 2021). Previous work suggested that *SP* is absent in several drosophilids outside of the *Sophophora-Lordiphosa*, namely in the Hawaiian drosophilid *D. grimshawi* and the *repleta* group species *D. mojavensis* (Kim *et al.*, 2010; Tsuda *et al.*, 2015; McGeary & Findlay, 2020). A further study also failed to detect SP in the *immigrans* group species *D. albomicans* (McGeary & Findlay, 2020). To establish whether these putative losses affect single species or broader lineages, we searched the genomes of 91 non-*Sophophora-Lordiphosa* members of the *Drosophilini.* We identified 4 putative loss events (Figure 1). One affected *Hirtodrosophila duncani,* another a monophyletic lineage of 14 species including the *annumilana* (*D. pseudotalamancana*), *bromeliae* (*D. bromeliae*), *nannoptera (e.g. D. nannoptera*), *mesophragmatica* (*D. gaucha*), and *repleta* (*e.g. D. hydei*) groups. We also observed two apparently independent putative losses within the lineage containing the *Scaptomyza* and Hawaiian radiations. One affected all 20 Hawaiian drosophilids in our dataset, while the other affected a monophyletic lineage within the *Scaptomyza* that included *S. montana, S. flava,* and *S. graminum*.

Among the 53 non-*Sophophora-Lordiphosa Drosophilini* species in which we did detect *SP,* all but one had just a single copy (Figure 1). The exception was *D. paramelanica* (*melanica* group) in which we detected a tandem duplication. The two *D. paramelanica* copies shared 100% identity in predicted protein sequence, suggesting against any neofunctionalization. Where *SP* was detected in non-*Sophophora-Lordiphosa* species, in all but two we found it within the same Muller D1 neighbourhood as occupied by *SP* in *Scaptodrosophila*. The exceptions were *Hirtodrosophila trivittata,* in which *SP* mapped to a distinct neighbourhood on *D. melanogaster* Muller element D that contained orthologs of *bruno3, CG3349,* and *CG17173,* and *D. repletoides*, in which *SP* mapped to a neighbourhood on 2L containing orthologs of *halo* and *haf*. Contrary to one previous study, we detected a copy of *SP* in *D. albomicans* (McGeary & Findlay, 2020). Collectively, therefore, we find evidence that outside of the *Sophophora-Lordiphosa* radiation *SP* appears to have been repeatedly and independently lost, with rare translocations and duplications.

Sex Peptide **has repeatedly and independently duplicated in the** ***Sophophora-Lordiphosa*** **radiation**

Relative to the other members of the *Drosophilini*, the *Sophophora-Lordiphosa* radiation to which *D. melanogaster* belongs showed a marked uptick in the frequency of duplication (Figure 2). In the earlier branching lineages, we detected apparently independent duplications within a sublineage of the *Lordiphosa,* in *D. subobscura*

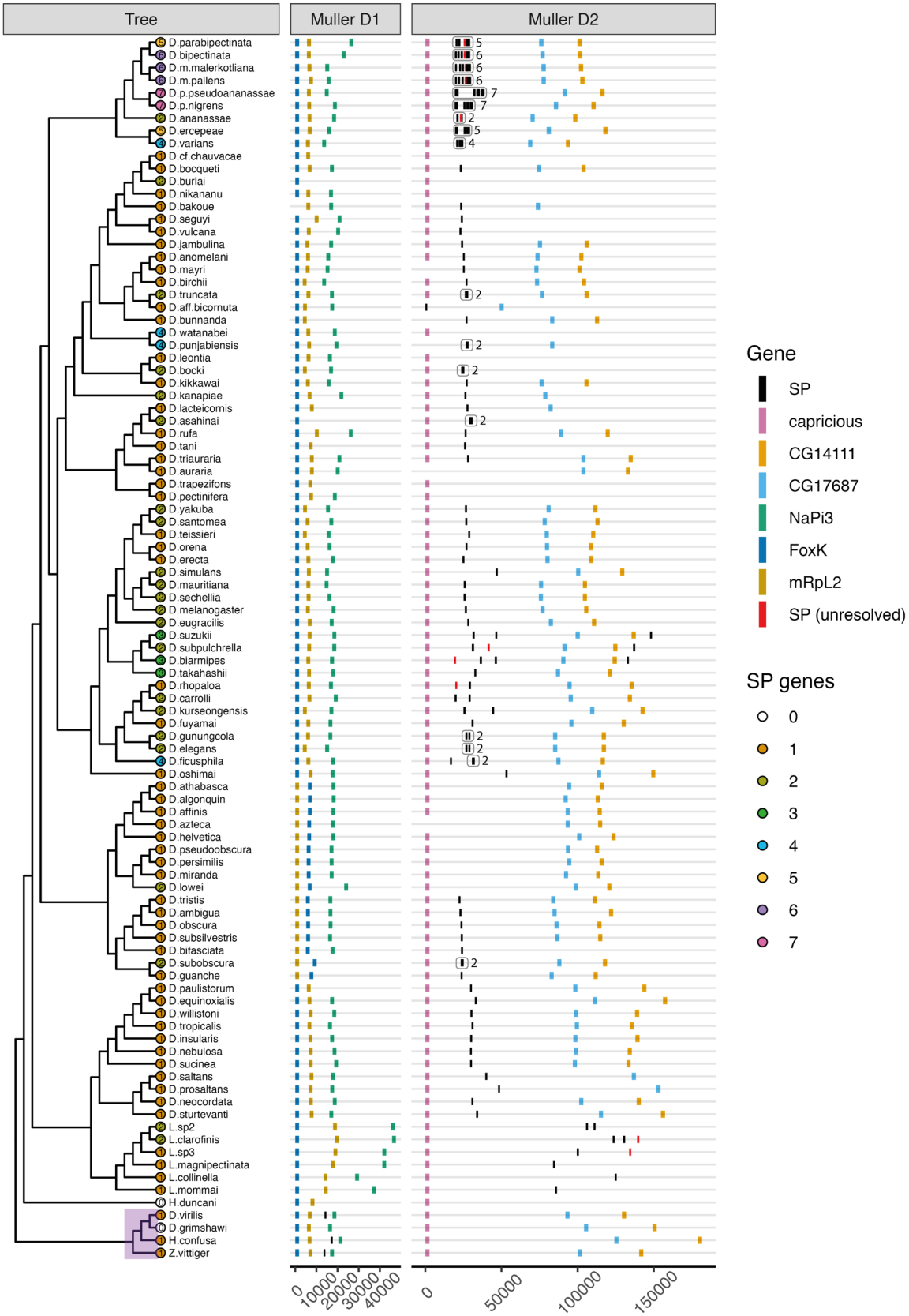

Figure 2. *Sex Peptide* copy number is markedly more variable in the *Sophophora-Lordiphosa* radiation than in other branches of the phylogeny. This figure focuses on the *Sophophora-Lordiphosa* radiation to which *D. melanogaster* belongs. Four non-*Sophophora-Lordiphosa*

drosophilids, shaded in purple, are included as outgroups. The structure of the Muller D1 and Muller D2 neighbourhoods is plotted as in Figure 1. Missing flanking genes are again likely indicative of contig breakpoints falling within the neighbourhood. The exceptions are the *Lordiphosa* species, where substantially elevated intergenic distances meant that the whole neighbourhood would not fit within the limits of the plot. Unresolved SP genes, shown in red, indicate genes that passed the reciprocal blast criteria to be designated as an *SP* gene and fell within one of the conserved *SP*-containing gene neighbourhoods but where a SP-like amino acid sequence couldn't be resolved (due, for example, to a premature stop codon, as in the case of *D. rhopaloa*). Note that all members of the *obscura* group have an inversion that flips the relative positions of *FoxK* and *mRpL2* in the Muller D1 neighbourhood. In a number of cases, some or all copies of *SP* were found to have translocated outside of the Muller D1 and Muller D2 neighbourhoods (an *obscura* group lineage, the *melanogaster* subgroup, *D. kanapiae, D. takahashii*, and *D.eugracilis*; summarised in Figure 4). In the shorter read *montium* subgroup assemblies, short contigs meant that in some species we couldn't identify the neighbourhood in which *SP* was located. This was the case for some *SP* genes in *D. cf. chauvacae, D. burlai, D. leontia, D. nikananu, D. pectinifera, D. punjabiensis,* and *D.watanabe*i. The *SP* genes in *D. auraria* and *D. trapezifons* could be mapped to the Muller D2 neighbourhood based on flanking sequence around the *SP* gene, but the *SP*-containing contigs were too small to include any of the neighbourhood genes.

(*obscura* group; see also Cirera & Aguadé, 1998), and in *D. lowei* (*obscura* group). Within the *melanogaster* group we found much greater variability in *SP* copy number, consistent with repeated, lineage-specific losses and duplications. As many as 7 tandemly arranged copies were present in some *ananassae* subgroup species, including *D. pseudoananassae nigrens* and *D. pseudoananassae pseudoananassae* (Figure 3). Thus, while we detected evidence of several independent loss events of *SP* outside of the *Sophophora-Lordiphosa,* within this clade we detected none, despite more dense taxon sampling.

To resolve the evolutionary relationships between different copies of *SP* within the *Sophophora-Lordiphosa* radiation, we constructed a tree of the predicted SP protein sequences (Supplementary figure 2). The tree supports numerous recent duplications affecting single species or species pairs, including in *D. elegans/D.gunungcola, D. takahashii, D.ficusphila, D. punjabiensis/D. watanabei, D. kanapiae,* and *D. subobscura*. The tree also suggests that there have been three separate expansions of *SP* copy number within the *ananassae* subgroup, one in the *bipectinata* complex and another in each of *D. varians* and *D. ercepeae* (Supplementary figure 2, coloured orange and blue, respectively), likely from an ancestral starting point of the two copies seen in *D. ananassae.* We note, however, that the sequence similarity that we observe between putative paralog pairs may be driven instead by concerted evolution.

Overall, *SP* copy number is considerably more variable among *Sophophora-Lordiphosa* species than in other parts of the *Drosophilinae* and markedly expanded in the "Oriental" lineage, which includes *D. melanogaster,* and *ananassae* subgroup.

Frequent translocations of *Sex Peptide* genes within the *Sophophora* subgenus

At the base of the *Sophophora-Lordiphosa* radiation, *SP* appears to have translocated from the Muller D1 to Muller D2 neighbourhood (Figure 2). Despite translocation, the

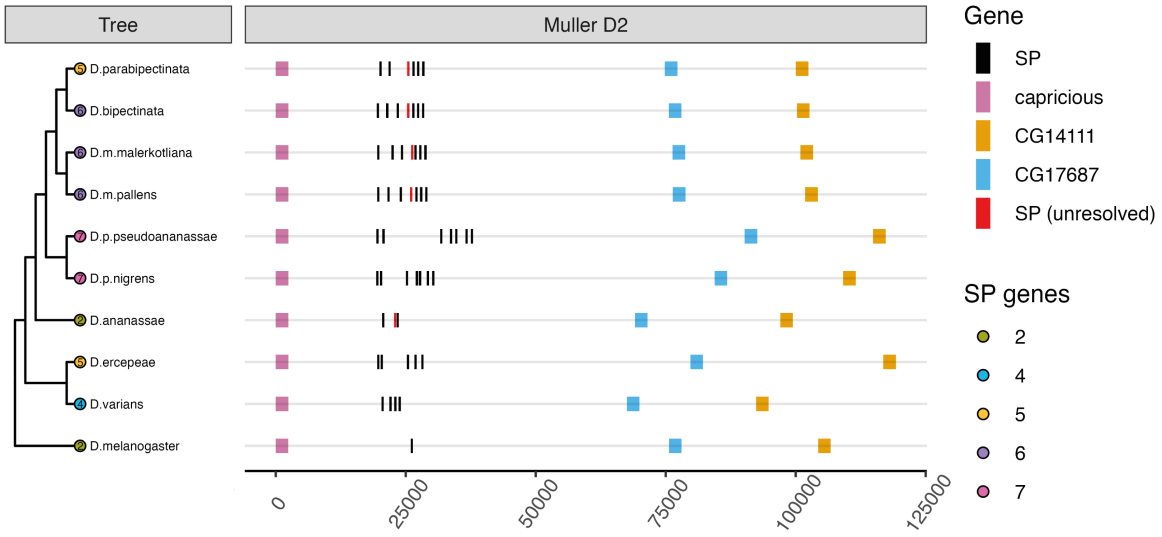

Figure 3. Repeated duplication of *Sex Peptide* genes in the *ananassae* subgroup. A phylogeny of the *ananassae* subgroup species used in this study, with *D. melanogaster* as an outgroup. The number of *SP* genes identified in each species is given at the tip of each branch. 'Unresolved' SP sequences, shown in red, are those which passed the reciprocal blast tests and fell within the syntenic Muller D2 gene neighbourhood but for which we could not resolve an SP-like protein sequence (*e.g.,* due to a premature stop codon). The structure of that neighbourhood is plotted on the right-hand side of the figure. Note that one of *D. melanogaster's SP* copies, the copy corresponding to *Dup99b,* falls outside of the Muller D2 neighbourhood.

configuration of the ancestral Muller D1 gene neighbourhood remains intact in the *Sophophora-Lordiphosa*. Thus, the mechanism of translocation did not lead to the breakup of the Muller D1 neighbourhood via a larger scale rearrangement. Several further translocations are then present. In a sublineage within the *obscura* group that covers the *pseudoobscura* and *affinis* subgroups, *SP* maps to a neighbourhood on *D. melanogaster* Muller element B that contains orthologs of *pgant35A* and *spel1* (Supplementary figure 3). In the *montium* subgroup, *D. kanapiae* has one copy in the canonical Muller D2 position and a translocated copy in a neighbourhood on Muller element D ~300kb away. In the "Oriental" lineage, *D. ficusphila* has three copies present in Muller D2, with one further downstream in the vicinity of *CG11905, CG13023*, and *CG13028.* Each of *D. suzukii, D. subpulchrella,* and *D. biarmipes*, which form a monophyletic clade within the "Oriental" lineage, bear SP copies in the canonical Muller D2 position with an additional copy just the other side of *CG14111* within the same neighbourhood (Figure 2). The protein tree supports a *Dup99b* identity for these copies that have skipped to the other side of *CG14111* (Supplementary figure 2, shown in pink). In *D. takahashii,* the sister species to this clade, two of the three SP genes we detected mapped to a neighbourhood on *D. melanogaster* Muller element C that contained *NLaz, robo2,* and *CG14346* orthologs. The protein tree also supports a *Dup99b* identity for these translocated copies (Supplementary figure 2, shown in pink). Given the low support for internal nodes, we were not able to accurately determine the timing of the initial duplication that gave rise to separate *SP* and *Dup99b* copies.

In *D. melanogaster, SP* falls within the Muller D2 neighbourhood, while *Dup99b* maps to Muller element E in a neighbourhood that contains *dmrt99b, gycalpha99b,* and *CG34296.* This arrangement appears to be ancestral to the *melanogaster* subgroup. The losses of an *SP* gene within a subset of species in this subgroup, namely *D. teissieri* (strain CT02; present in 273.3)*, D. orena,* and *D. erecta,* affect the Muller element E *Dup99b* copy, rather than the Muller D2 *SP* copy (Supplementary figure 4A,B). The Muller element E neighbourhood is not occupied by *SP* in *D. eugracilis,* the *melanogaster* subgroup's sister species. In *D. eugracilis, SP* is present in the Muller D2 neighbourhood, with a second copy in a different position on Muller element E that contains orthologs of *SmD2, CG18048,* and *Hr83*. Despite the lack of synteny, the protein tree supports a *Dup99b* identity for this translocated copy.

In the course of this analysis, we observed that the intergenic distances across both the Muller D1 and Muller D2 neighbourhoods were substantially increased in the *Lordiphosa* species (Figure 2). The genomes of these species are considerably larger than those of *D. melanogaster* (*e.g., L. mommai* = 339.464 Mb; *L. collinella* = 375.097 Mb; *D. melanogaster* = 139.049). Thus, increased intergenic distances may be a widespread feature of the *Lordiphosa* genomes.

Overall, we conclude that SP has undergone frequent translocations within the *Sophophora* and that this is particularly pronounced in the case of the *Dup99b* copy in the *melanogaster* species subgroup (summarised in Figure 4).

Male reproductive tract-biased expression is a conserved feature of *SP* genes

Our identification of *SP* genes was based purely on gene sequence data, leaving open the question of whether and where they are expressed. For 19 species, we were able to test for expression using RNA-seq datasets available through NCBI (https://osf.io/ydpfz). Because many of these datasets were labelled by sex and tissue, we could go further than just asking whether a given *SP* gene is expressed and ask whether it shows sex or tissue expression bias. 38 of the 42 *SP* genes we detected across the 19 species were expressed, although many were un- or incorrectly annotated (*e.g.,* as long non-coding RNAs) in the reference genomes (see Methods). All 4 of those that weren't expressed matched hits for which we weren't able to resolve SP-like protein sequences due to point mutations affecting either the start codon or introducing premature stop codons, suggestive of pseudogenisation. These 4 sequences, which came from a single hit in each of *D. biarmipes, D. ananassae, D. bipectinata*, and *D. rhopaloa* (Supplementary figure 5), therefore did not made it into the final *SP* counts given in Figures 1 and 2. Of the 38 *SP* genes we found to be expressed, all showed strongly male-biased expression, including in the early branching *D. busckii* (Supplementary figure 6). The one exception was detection of appreciable *SP* expression in a single *D. simulans* female sample. However, this sample also showed appreciable expression of the long non-coding RNAs *roX1* and *roX2* (Supplementary figure 7)*,* which among adults are expressed only in males (Meller *et al.*, 1997; Meller, 2003; Lim & Kelley, 2012) and therefore suggests that the sample is contaminated by males or mislabelled.

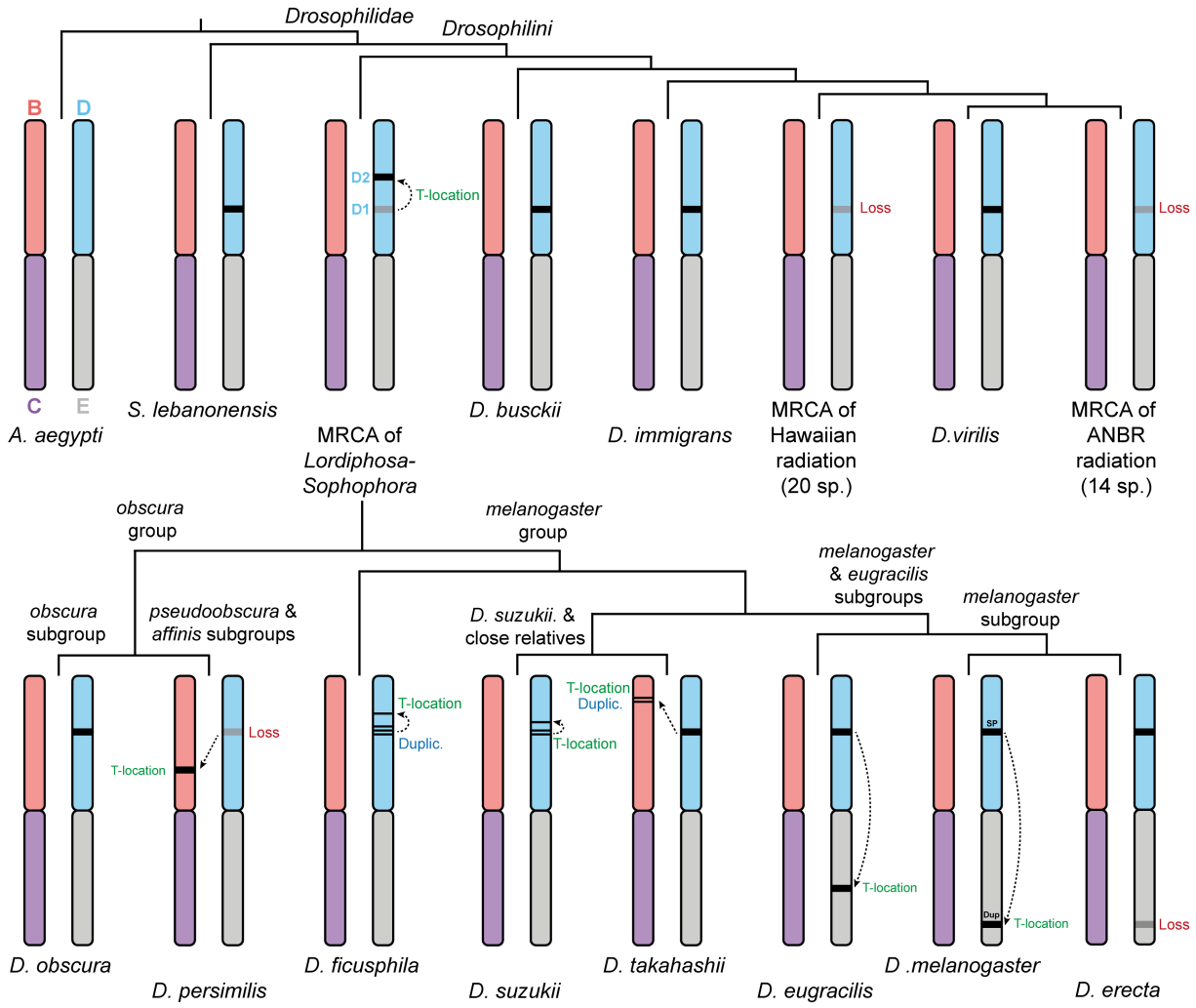

Figure 4. An overview of duplications, losses, and translocations of *Sex Peptide*. This schematic summarises several translocations, duplications, and losses that have occurred within the *Drosophilidae*. Muller elements B to E are depicted for each species. We first detect *Sex Peptide* (*SP*), denoted by a thick black line, in the *Scaptodrosophila,* a *Drosophilidae* genus that forms the outgroup to the *Drosophilini* tribe. In these species, *SP* falls within a gene neighbourhood ('Muller D1') that maps to Muller element D in *D. melanogaster* and contains orthologs of *FoxK, mRpL2,* and *NaPi-III.* The *Drosophilini* then branches into two clades. The first contains the *Sophophora-Lordiphosa* radiation. In the most recent common ancestor (MRCA) of this radiation, *SP* has been lost from the ancestral gene neighbourhood and has translocated to a new position that maps to the same Muller element ('Muller D2'). This Muller D2 neighbourhood contains orthologs of *capricious, CG14111,* and *CG17687.* In the other *Drosophilini* lineage, *SP* is retained in the original genomic location, with the exceptions of *Hirtodrosophila trivittata* and *D. repletoides*, and several independent losses, which include all 20 of the Hawaiian *Drosophila* species we looked at and all 14 of the species we looked at that fell within a monophyletic radiation covering the *annulimana*, *bromeliae*, *nannoptera* and *repleta* ('ANBR') groups. Within the *Sophophora-Lordiphosa* radiation, *SP* has repeatedly duplicated and/or translocated to new locations. In the *obscura* group, *SP* is retained in the Muller D2 position that's ancestral to the *Sophophora-Lordiphosa* radiation in all species except those within the *pseudoobscura* and *affinis* subgroups. In these subgroups, *SP* has been lost from Muller D2 and translocated to a neighbourhood that maps to *D. melanogaster* Muller element B. Within the *melanogaster* group, SP has repeatedly and independently translocated and duplicated. Members of the *melanogaster* subgroup retain *SP* in Muller D2, but show a translocated copy (*Dup99b*) on Muller E. In several species, including *D. erecta* and *D. orena,* the Muller E copy has been lost. *D. eugracilis,* the outgroup to the *melanogaster* subgroup, shows

a Muller D2 copy but its *Dup99b* copy maps to a non-canonical neighbourhood on Muller E. *D. suzukii* and its close relatives (*D. biarmipes* and *D. subpulchrella*) have 1 to 2 copies of *SP* in Muller D2 with an additional copy just the other side of the D2 neighbourhood; *D. takahashii* has a single copy in D2 with a translocation to and subsequent duplication in Muller B. A protein tree supports a *Dup99b* identity for the copies that fall outside of the canonical Muller D2 location in *D. eugracilis, D. suzukii, D.biarmipes, D. subpulchrella,* and *D. takahashii* (Supplementary figure 2). *D. ficusphila* has 3 copies in D2 with an additional copy in a different neighbourhood on Muller element D. Where multiple SP copies are present within a small area, we have thinned the bar signifying an SP gene. Positions of SP genes on each Muller element are approximate and not to scale.

In all 10 species where we had tissue data, including the distantly related *D. virilis,* we observed clearly enriched expression of *SP* family genes in the male reproductive tract (Figure 5). Where datasets were available for sub-portions of the male reproductive tract, expression was generally substantially higher in samples labelled as accessory gland or non-gonadal reproductive tissues than in samples labelled as testes. The extent of testes expression was variable between samples and between species, perhaps reflecting varying degrees of contamination between these closely associated tissues during dissection.

Based on these data, we conclude that male reproductive tract-biased expression is a conserved feature of the *SP* gene family and is therefore likely the ancestral expression pattern within the *Drosophilini.*

The domains of Sex Peptide homologs show markedly different degrees of evolutionary change

The C-terminus portion of SP is responsible for stimulating post-mating responses in *D. melanogaster* (Schmidt *et al.*, 1993; Peng *et al.*, 2005a). In previous studies, it was also the C-terminus that displayed greatest sequence conservation across 10 species' SP proteins (Tsuda *et al.*, 2015) and, along with the signal peptide, between *D. melanogaster* Dup99b and SP (Saudan *et al.*, 2002; Ding *et al.*, 2003). Aligning 233 SP sequences from 148 genomes, we find that the C-terminus is highly conserved and the best conserved part of the protein (Figure 6A). Several residues in this region are present in almost all SP sequences in our dataset: within the consensus sequence KWCRLNLGPAWGGRGKC, W_2, C_3, G_8, P_9, G_{12}, G_{13}, and C_{17} are each conserved in >97% of sequences (Figure 6A). Thus, the disulphide bond between the two C-terminus cysteine residues known from *D. melanogaster* SP and Dup99b appears to be a general feature of SP family proteins.

Previous work has shown that substitution of W_2 with alanine is associated with a steep drop in the efficiency of SP in eliciting a receptor-mediated response. Combined substitution of both W_2 and W_{11} with alanine is associated with almost total loss of activity with SPR in *in vitro* assays (Poels *et al.*, 2010). While W_2 was present in >97% of sequences, substituted only in *D. quadrilineata* and a copy in each of *D. biarmipes* and *D. suzukii*, it was surprising to find that W_{11} was widely substituted, being conserved in only 68.7% of sequences. We detected substitution of W_{11} in a broad range of species, including in members of the *virilis* and *willistoni* groups and *Zaprionus* and *Scaptomyza*

genera, as well as in *D. melanogaster* Dup99b. In a species with high copy numbers, *D. bipectinata,* 4 of its 6 copies maintained both W_2 and W_{11}, while the remaining two showed substitution at W_{11}.

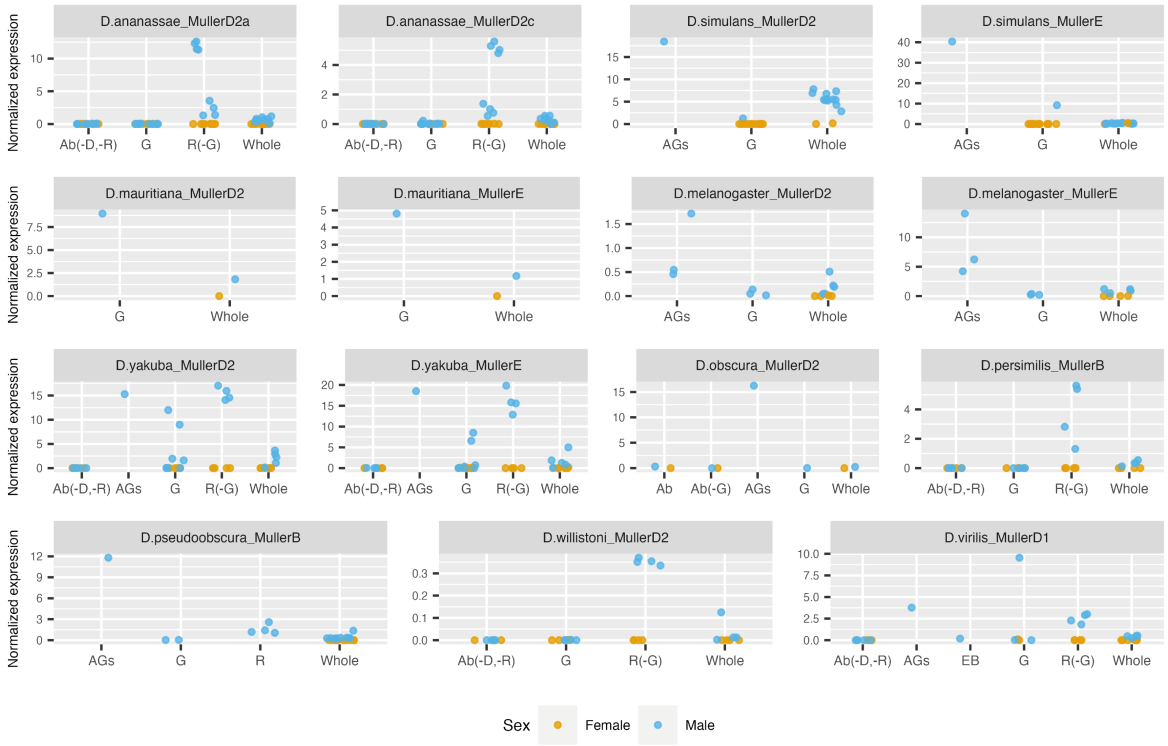

Figure 5. Male reproductive tract biased expression is a conserved feature of *SP* genes. We obtained tissue-specific RNA-seq datasets from 10 different species hosted on NCBI. We extracted the maximum exonic read coverage value for each *SP* locus and divided it by the equivalent value extracted from the housekeeping gene *RpL32* to normalise expression across samples. Each plot depicts data from a different *SP* homolog. Genes are named by species, the chromosome arm and neighbourhood they map to, and, in the case of tandemly arranged copies in the Muller D2 neighbourhood, in their order from *capricious.* Note that there is no D.ananassae_MullerD2b as this is the 'unresolved' copy that can be seen in Figure 3, for which we failed to find expression in any examined tissue. Ab (-D, -R) = Abdomen, digestive and reproductive tracts removed; R = Reproductive tract; G = Gonad (ovaries/testes); R (-G) = Reproductive tract without gonad; Whole = Whole flies; AGs = Accessory glands; Ab = Abdomen; Ab (-G) = Abdomen without gonad; EB = Ejaculatory bulb. The raw data, including dataset accessions, is available at https://osf.io/ydpfz.

Strong conservation of the C-terminus was clear in separate alignments of *Sophophora-Lordiphosa* and non-*Sophophora-Lordiphosa* sequences, suggesting that it is not an artifact of the high number of paralogs derived from *Sophophora* species (Figure 6B,C). In contrast, the mid-section, which has been implicated in stimulating innate immune responses (Domanitskaya *et al.*, 2007), and the N-terminus (following cleavage of the signal peptide), which is responsible for binding to sperm and stimulating juvenile hormone synthesis (Peng *et al.*, 2005a), showed quite limited sequence conservation.

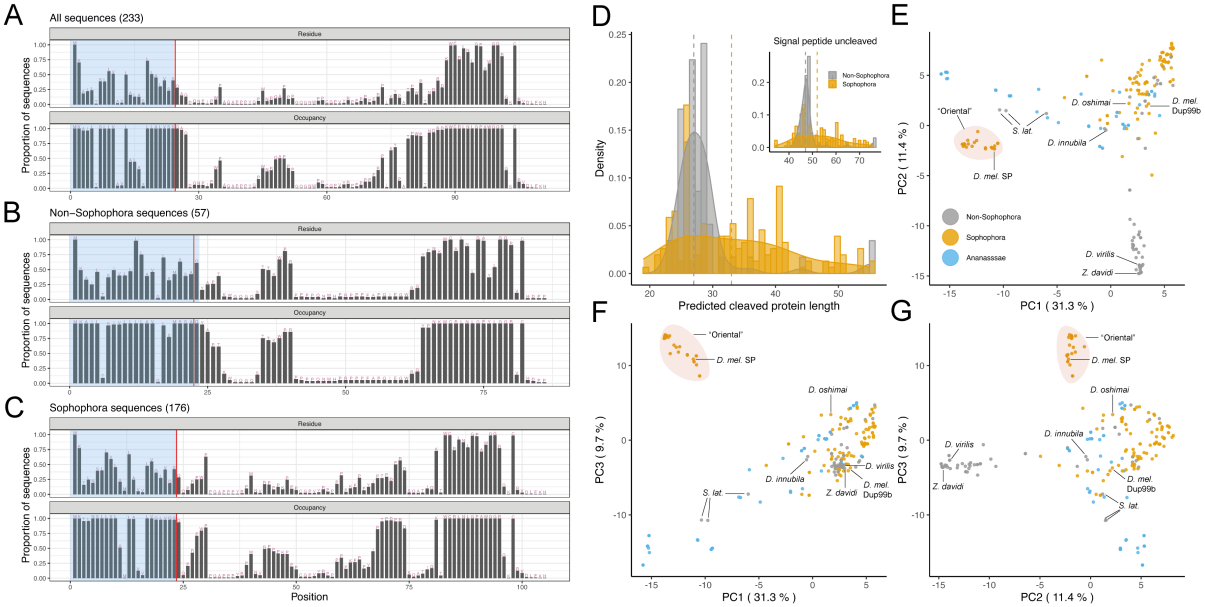

Figure 6. Sex Peptide has undergone domain-specific evolutionary change. (A-C) Consensus sequences based on MAFFT alignment of the resolvable amino acid sequences of (A) all 233 SP genes, (B) all 57 from the non-*Sophophora-Lordiphosa* species, and (C) all 176 from the *Sophophora-Lordiphosa* species. In each panel, the top plot gives the proportion of sequences with the consensus amino acid in the same position, while the bottom plot gives the proportion of sequences in which each position is occupied in the alignment. A red line indicates the position of a predicted signal peptide cleavage site determined using SignalP 6.0. The probability of a cleavage site in the given position was 0.980 in A, 0.973 in B, and 0.982 in C. (D) The distribution of predicted protein lengths for each resolved SP sequence plotted separately for *Sophophora-Lordiphosa* and non-*Sophophora-Lordiphosa* sequences. The main panel shows the protein lengths after removing the signal peptide at the predicted site; the inset shows the full lengths prior to cleavage of the predicted peptide. Four sequences, two from each of the *montium* subgroup species *D. punjabiensis* and *D. watanabei,* did not contain a predicted signal peptide and are therefore omitted from the post-cleavage plot. Dashed lines indicate median lengths for the *Sophophora-Lordiphosa* and non-*Sophophora-Lordiphosa* species, separately. The smoothed curve shows a kernel density estimate, a smoothed version of the histogram, derived using the geom_density function in ggplot2. (E-G) PCA plots based on BLOSUM62 substitution scores from the MAFFT-aligned resolved SP sequences. (E) PC1 vs PC2. (F) PC1 vs PC3. (G) PC2 vs PC3. The percentage values in the axis titles reflect the proportion of variance explained by a given PC. Points are coloured based on whether they correspond to *Sophophora-Lordiphosa,* non-*Sophophora-Lordiphosa,* or *ananassae* subgroup species.

These results should be considered alongside a caveat. The first is that we used the *D. melanogaster* SP C-terminus sequence (KWCRLNLGPAWGGRC) as the search query for identifying *SP* genes in other species (see Methods). Therefore, while the consensus sequence we generated may suggest that this C-terminus sequence is the defining characteristic of SP family proteins, the high degree of conservation we detect might be an artefact of us ourselves having used this sequence as the defining characteristic of an *SP* gene. Because of this search design, while we can readily pick up SP sequences with highly diverged N-termini, we would not be able to detect an SP sequence with a conserved N-terminus and highly diverged C-terminus. What we can conclude, however,

is that the C-terminus appears highly conserved, that certain residues within the C-terminus are almost invariable, and that where *SP* genes are detected based on the C-terminus sequence the N-terminus and midsection regions appear to be considerably more variable in sequence.

Interspecific variation in Sex Peptide protein length is elevated in the *Sophophora-Lordiphosa radiation*

In our analyses, change in SP family proteins wasn't restricted to the identities of individual amino acid residues. We observed that the length of SP family proteins was variable. The median length of predicted SP proteins in the *Sophophora-Lordiphosa* radiation was longer than that of the species outside of this lineage (signal peptide uncleaved protein length: 52aa vs 47aa; cleaved protein: 33aa vs 27aa; in *D. melanogaster,* the length of uncleaved Dup99B is 54aa while SP is 55aa). What was particularly notable, however, was the markedly greater degree of variability in SP length among *Sophophora-Lordiphosa* species (Figure 6D). These differences held after *in silico* cleavage of predicted signal peptides, indicating that the variation is driven by change in domains with post-secretion functions (Figure 6D). Generally, variation in length was due to the introduction of additional amino acids upstream of the conserved C-terminus region, although in a small number of cases additional amino acids were present beyond the terminal cysteine of *D. melanogaster's* SP sequence. The longest lengths we detected were a pair of 76aa paralogs (pre-cleavage) in the non-*Sophophora* species *S. latifasciaeformis*; the shortest were a pair of 34aa paralogs in each of the *montium* subgroup species *D. watanabei* and *D. punjabiensis.* Interestingly, these short sequences were the only among our 233 for which we did not find evidence of a high-confidence signal peptide, suggesting they may have lost the capacity to be secreted.

A caveat to this length analysis is that our sequence predictions were based on alignment to the *D. melanogaster* gene structure (as used in Tsuda *et al.*, 2015; McGeary & Findlay, 2020). When it comes to estimating protein length, erroneous annotation of exon/intron boundaries could lead us to include/exclude additional amino acids in the predicted amino acid sequence. Fortunately, we could return to the RNA-seq data available for 38 *SP* genes in 19 species as a source of validation. In all 38 cases, our predicted exon/intron boundaries matched those derived from the RNA-seq data. The expression data alone therefore supports a change in pre-cleavage SP sequence length between *e.g., D. virilis* (47aa) and *D. rhopaloa* (60aa), as well as *D. bipectinata* expressing a set of 6 SP proteins of variable length (46aa, 49aa, 54aa, 62aa, 68aa, 72aa).

Evolutionary shifts in Sex Peptide protein sequence

To explore evolutionary shifts in SP protein sequence we generated a pairwise similarity score matrix using BLOSUM62 and used a PCA to create a spatial representation of the similarities between the 229 post-signal peptide cleavage sequences (*i.e.,* all but the pair of short paralogs in *D. watanabei* and *D. punjabiensis*). Three patterns were apparent (Figure 6E-G). The first was that *D. melanogaster* SP – but not Dup99b – and it's orthologs in the "Oriental" lineage (with the exception of the most basal species we sampled from this lineage, *D. oshimai*) showed clear separation from the remaining sequences across PC1, 2, and 3. With a range of 36-41aa, the sequences in this cluster

were all above the median post-cleavage length for *Sophophora-Lordiphosa* (33aa), but far from the longer sequences of *S. latifasciaeformis* (55aa), suggesting that length alone wasn't responsible for their separation. Instead, their distinct clustering may be driven by their N-terminus and midsection sequences, which showed limited conservation with those of other SP proteins. The second pattern was that *ananassae* subgroup sequences were highly dispersive, spreading over a large portion of PC space in each plot. This pattern suggests a high degree of sequence diversity between paralogs within this lineage. The third was that a paraphyletic combination of *Hirtodrosophila, Scaptomyza, Zaprionus, virilis, robusta,* and *melanica* sequences clustered tightly together, suggesting limited interspecific variation. This did not represent the totality of non-*Sophophora-Lordiphosa* sequences however: sequences from the *immigrans, quinaria, guttifera, cardini* and *dunni* groups, along with those of *D. busckii* and *S. lebanonensis,* were interspersed among many of those from the *Sophophora-Lordiphosa* radiation, including the Dup99b copies. This separation between non-*Sophophora-Lordiphosa* sequences seemed to largely stem from shifts in the protein sequence of the post-cleavage N-terminus and midsection. Collectively, this PCA points to an increase in variation among SP protein sequences within the *Sophophora*, driven largely by particularly divergent sequences in the "Oriental" lineage and *ananassae* subgroup.

Microcarrier morphology is not clearly linked to the copy number of *Sex Peptide*

We next wanted to explore functional consequences of the diversity in the phylogenetic distribution and sequence of *SP* genes. Recently, SP was shown in *D. melanogaster* to be a key assembly and disassembly factor for microcarriers – large, lipid-based structures that appear to store and traffic seminal fluid proteins – and that SP was required to maintain their distinctive morphology (Wainwright *et al.*, 2021). Because of this relationship between SP and microcarrier structure, the authors (including the first author of this paper) suggested that variation in SP sequence might be associated with inter-specific variation in microcarrier morphology. Stainings performed on accessory glands from several species revealed changes in microcarrier shape, size, and density, providing some support for this hypothesis (Wainwright *et al.*, 2021).

Using the neutral lipid-specific dye LipidTox, which has previously been used to stain microcarriers (Wainwright *et al.*, 2021), we sought to examine the relationship between SP and microcarrier structure on two levels. The first was to ask whether variation in *SP* copy number is associated with a shift in microcarrier morphology. For this, the *ananassae* subgroup provides an ideal system, given that its constituent species encode between 2 and 7 *SP* copies. Within the *bipectinata* species complex, all of the species that we looked at (*D. parabipectinata, D. bipectinata, D. m. malerkotliana, D. m. pallens, D. p. nigrens*), which each encode between 5 and 7 *SP* copies, showed small, globular microcarriers (Figure 7A-F). Curiously, these microcarriers appeared almost doughnut-like in shape, with a central indentation reminiscent of the biconcave disk shape of human red blood cells. This morphology was clearly distinct from that of *D. melanogaster* microcarriers, which appear as a heterogeneous mix of fusiform, ellipsoid, and thread-like structures (Figure 7L). The *bipectinata* complex morphology was also distinct from those of three other *ananassae* subgroup species: *D. ananassae* (2 copies), which had thread-like and spiral or doughnut shaped microcarriers (Figure 7G); *D. ercepeae* (5

copies), which had thread-like microcarriers (Figure 7H); and *D. varians* (4 copies), which displayed a highly divergent organization of the lumen's contents (Figure 7I,J). In *D. varians*, LipidTox appeared to be excluded from vacuolar structures that were filled with small, weakly stained droplets. The vacuoles appeared larger in the proximal region of the gland, suggesting that they may fuse as they move towards the ejaculatory duct (Figure 7J). We observed a similar pattern in the three-copy-encoding *D. takahashii*, a non-*ananassae* subgroup species, although here the LipidTox staining was negligible (Figure 7N). This contrasted with *D. takahashii*'s close relative *D. biarmipes,* which despite also encoding three copies showed a unique staining pattern of strongly stained, tiny microcarriers that appeared to aggregate (Figure 7M). The microcarriers of *D. biarmipes* adopted a conformation reminiscent of *D. melanogaster's* following transfer to the female reproductive tract as they begin to break down into smaller puncta (Wainwright *et al.*, 2021). The conformation observed in *D. biarmipes* appears to be a derived state as the more distantly related *D. carrolli* (Figure 7O) bears microcarriers that more closely resemble those of *D. melanogaster, D. sechellia,* and *D. simulans* (Figure 7L; Wainwright *et al.*, 2021).

Consequently, while the *melanogaster* group shows remarkable diversity in both microcarrier morphology and *SP* copy number, there appears to be no clear relationship between them. Indeed, while high *SP* copy numbers appear to be associated with the small, spherical microcarriers in the *bipectinata* complex, similar shapes (albeit without the doughnut/biconcave component) were previously observed in the single-copy-encoding *obscura* group species *D. persimilis* and *D. pseudoobscura* (Wainwright *et al.*, 2021).

Detection of LipidTox$^-$ microcarrier-like structures both within and beyond the *Sophophora*

Previous staining of accessory glands from the single-copy-encoding *D. virilis* demonstrated that a copy of *SP* is not sufficient for LipidTox-stained microcarriers (Wainwright *et al.*, 2021). *D. virilis* instead displayed a more uniform 'flocculence' within the gland's lumen that showed little evidence of LipidTox staining. We observed a similar flocculent arrangement in the single-copy-encoding *D. immigrans* (Figure 7Q). We also observed an essentially microcarrier-free glandular lumen in the four-copy-encoding *D. ficusphila* (Figure 7P). In this species we observed only a handful of weakly stained structures per gland, the rarity and structural inconsistency of which renders their classification as microcarriers doubtful. Alongside these cases, we detected instances of microcarrier-like, ellipsoid structures that failed to take up LipidTox in several species from diverse parts of the drosophilid tree, namely the *montium* subgroup species *D. kikkawai* (Figure 7K) and the non-*Sophophora* species *Scaptomyza hsui* (Figure 7S). All four of these species – *D.immigrans, D. ficusphila, D.kikkawai,* and *S. hsui* – each encode at least one *SP* copy, providing further support for the claim that a copy of *SP* is not sufficient for LipidTox$^+$ microcarriers. Moreover, the detection of LipidTox$^-$, ellipsoid, microcarrier-like structures and LipidTox$^-$ flocculence raises the question of whether these represent fundamentally different structures to microcarriers or whether taking up LipidTox – an indicator that they contain large quantities of triglycerides and other nonpolar lipids (Wainwright *et al.*, 2021) – is a feature of some, but not all, microcarriers.

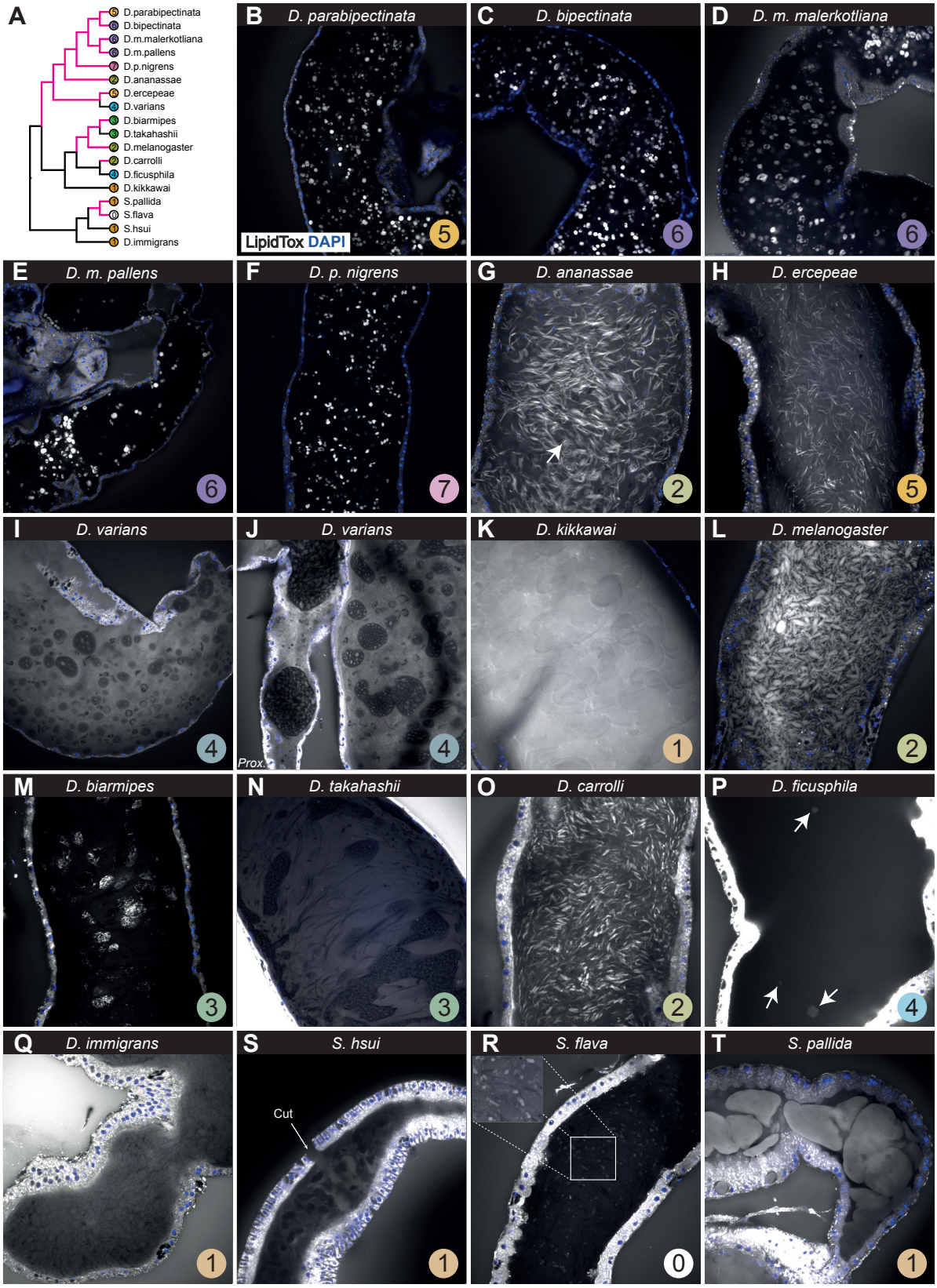

Figure 7. *Sex Peptide* is neither necessary nor sufficient for microcarriers. (A) A phylogeny of all species included in this figure. Branches coloured pink indicate species shown in B-T that

demonstrate canonical (*i.e., D. melanogaster*-like) staining with LipidTox, a neutral lipid-specific dye used to selectively stain microcarriers (Wainwright *et al.*, 2021). (B-T) Drosophilid accessory glands stained with LipidTox and the nuclear stain DAPI (blue). The circled number in the bottom right-hand corner of each panel indicates the number of *Sex Peptide* copies we detect in each species. (G) The arrow is highlighting a spiral/doughnut shaped microcarrier, a shape which is rare in comparison to the more common thread-like conformation in this species. (J) Prox. refers to the proximal region of the gland, *i.e.* the region that connects to the ejaculatory duct. (P) Arrows point to the ambiguous, sparse, and weakly stained material we observed in stained *D. ficusphila* glands. (S) An arrow points to a cut in the glandular epithelium, which was made to enhance dye penetration. In each case, glands were co-stained in the same well with those from *D. melanogaster* to act as a control for assessing whether the staining was successful.

LipidTox$^+$ microcarriers predate the *Sophophora,* but copies of *Sex Peptide* are neither necessary nor sufficient for their presence

We have provided evidence that a copy of *SP* is not sufficient for the presence of LipidTox$^+$ microcarriers. But is a copy of *SP* necessary? To test this, we stained glands from a species that we identified as having lost a copy*, S. flava.* We observed small, globular microcarriers reminiscent of those from the *obscura* group, albeit weaker in their staining (Figure 7R). Thus, *SP* is not necessary for microcarriers. Moreover, the previous complement of species that had been stained suggested that LipidTox$^+$ microcarriers were confined to the *obscura* and *melanogaster* groups. We now show that they are present outside of the *Sophophora.*

To better understand the distribution of microcarriers within the *Scaptomyza,* we also looked at *S. flava*'s single-copy-encoding sister species, *S. pallida*. This species showed strong LipidTox staining, but the pattern was unlike any other species we looked at (Figure 7T). Rather than the lumen being filled with large numbers of small microcarriers with well-defined shapes, the *S. pallida* lumen was filled with substantial clouds of stained secretion that in many cases spanned the full diameter of the gland's internal space. This pattern was reminiscent of that observed in repeatedly mated *SP* null – but not wild-type – *D. melanogaster* males (Wainwright *et al.*, 2021). Thus, in the presence of an *SP* ortholog we observe in *S. pallida* an apparent phenocopying of an *SP* null conformation, further evidence that microcarrier morphology may, at a broad taxonomic scale, be largely decoupled from evolutionary change in *SP.*

The mode of evolutionary change is decoupled between *Sex Peptide* and its receptor, *Sex Peptide Receptor*

Our analyses suggest that *SP* has been repeatedly lost and duplicated during the course of *Drosophila* evolution. But what about its receptor, *SPR*? Does this show correlated instances of copy number change? The answer appears to be no. We failed to detect a single instance of duplication across 195 *Drosophilidae* genomes (and *A. aegypti*). At the level of loss, we were unable to resolve an SPR coding sequence for five species (Supplementary figure 8). These were *L. magnipectinata, Z. taronus,* and three *Scaptomyza* species (*S. hsui, S. montana,* and *S. flava*). Thus, the losses of *SP* in the Hawaiian and *repleta* groups do not appear to have been accompanied by loss of *SPR.* Conversely, the apparent losses of *SPR* in *Z. taronus, S. hsui,* and *L. magnipectinata* were not accompanied by loss of *SP*. However, we did record two species, *S. flava* and

S. montana, in which we were unable to resolve neither *SP* nor *SPR* sequences, suggesting that although there is no phylogenetically repeatable association in the copy number of the two genes, the loss of one can be accompanied by the loss of the other.

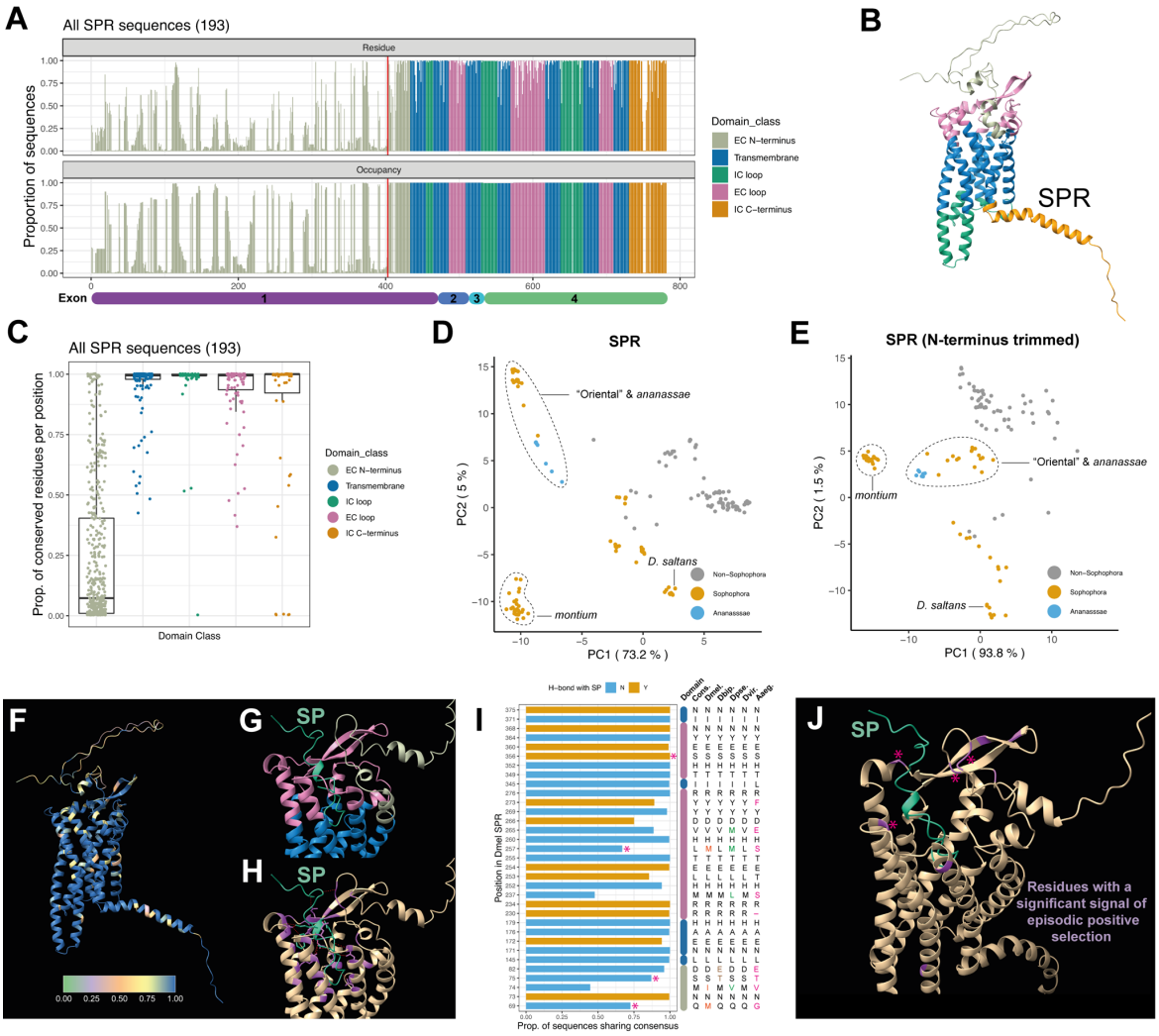

Figure 8. SPR residues showing evidence of episodic positive selection are enriched in the ligand-facing domains. (A) A consensus sequence based on MAFFT alignment of the resolvable amino acid sequences of all *SPR* coding sequences. The top plot gives the proportion of sequences with the consensus amino acid in the same position, while the bottom plot gives the proportion of sequences in which each position is occupied in the alignment. Each residue is coloured based on the functional domain to which it belongs based on the UniProt annotations. The red line indicates the start of the conserved region we use in the molecular evolution analyses. Beneath the plot, we show the corresponding exon that encodes each consensus residue. EC = extracellular, IC = intracellular (B) The AlphaFold prediction of the structure of *D. melanogaster* SPR as downloaded from UniProt (AF-Q8SWR3-F1) and coloured by the domain each residue belongs to based on the positions listed in the UniProt 'Features' table. (C) A boxplot showing the proportion of residues at each position that matched the consensus residue (*i.e.,* the degree of conservation at each position). Residues are plotted separately according to their domain class of origin. (D) PCA plot based on BLOSUM62 substitution scores from the MAFFT-aligned SPR protein sequences. The percentage values in the axis titles reflect the proportion of variance explained by a given PC. Points are coloured based on whether they

correspond to *Sophophora-Lordiphosa,* non-*Sophophora-Lordiphosa,* or *ananassae* subgroup species. (E) As (D) but after removing the non-conserved region of the N-terminus (*i.e.* the region preceding the red line in (A)). (F) The same prediction shown in (B) but with residues coloured by the proportion of conserved residues per position (see also Supplementary movie 1). High values indicate high conservation. (G) The ColabFold top-ranked prediction of the interactions between SP (shown in green) and SPR (residues coloured by domain). (H) As in (G) but with predicted contact residues coloured purple and predicted hydrogen bonds between SP and SPR residues shown with red-dotted lines (See also Supplementary movies 2 and 3). (I) A bar chart showing the proportion of sequences sharing the consensus residue for each predicted contact residue. Bars are coloured by whether the residue is also predicted to form a hydrogen bond with SP. Asterisks denote predicted contact residues for which we detected evidence of episodic positive selection using *MEME.* Alongside the plot, coloured bars, using the same colour scale as in (A-C), denote the functional domain the residue falls within. The two adjacent blue bars denote separate, consecutive transmembrane domains. Alongside are the corresponding amino acid residues in each position for each of the consensus ('Cons.'), *D. melanogaster* ('Dmel.'), *D. bipectinata* ('Dbip.'), *D. pseudoobscura* ('Dpse.'), *D. virilis* ('Dvir.'), and *A. aegypti* ('Aaeg.') sequences. Residues that depart from the mode among these plotted sequences are coloured. (J) As (G) but colouring only the 10 residues in SPR for which we detected evidence of positive selection using *MEME*. Asterisks denote predicted contact residues that show evidence of positive selection.

Evolutionary change in the protein coding sequence of *SPR* is concentrated in the extracellular N-terminus domain

With no clear evidence of copy number evolution in *SPR* among our species study set, we next sought to examine changes in *SPR* sequence along the drosophilid phylogeny. Our first analysis asked whether particular SPR domains show evidence of greater evolutionary change. Comparing the blast scores extracted from the top hit returned from querying each *D. melanogaster* SPR exon, we observed that the amino acid sequence of exon 1, which encodes the extracellular N-terminus through to the second transmembrane domain, showed much greater variability between species than the remaining exons 2,3, and 4 (Supplementary figure 8). Specifically, species in the "Oriental" lineage showed notably higher exon 1 blast scores than the remaining species sequences, which were otherwise generally tightly clustered. In exons 2 through 4, no such gap between this lineage and the remaining species was observed. The portion of the SPR amino acid sequence encoded by exon 1 also showed the weakest alignment in an alignment of all SPR sequences we extracted. The highly variable region of exon 1 accounted for approximately the first 66aa of the 435aa long *D. melanogaster* SPR sequence, all of which were contained within the extracellular N-terminus domain (Figure 8A-C). Beyond the N-terminus region, the remainder of the protein sequence showed much stronger conservation (Figure 8A-C).

No clear signal of episodic diversifying selection in SPR sequence among drosophilids

If SP and SPR are coevolving, perhaps through a sexually antagonistic coevolutionary arms race, then we might predict that the bursts of evolutionary change in *SP* copy number and protein coding sequence in the *ananassae* subgroup and "Oriental" lineage correlate with similar bursts in evolutionary change in SPR protein sequence. However, a PCA of the full alignment of 193 SPR sequences, which we generated using the same approach as for SP, showed that while the degree of diversity among SPR sequences

was apparently higher in the *ananassae* subgroup and "Oriental" lineages relative to the *montium* subgroup, the overall diversity did not appear markedly elevated compared to the full spectrum of SPR sequences (Figure 8D,E). We complemented our protein sequence analysis with branch- and gene-level tests for selection. For this analysis, we removed the poorly conserved region of the N-terminus. We found no evidence for episodic diversifying selection in *SPR* sequences using *aBSREL* (adaptive Branch-Site Random Effects Likelihood) or *BUSTED* (branch-site unrestricted statistical test of episodic diversification)(Murrell *et al.*, 2015; Smith *et al.*, 2015). This held whether we formally tested all branches, only those in the *Sophophora-Lordiphosa* radiation, or only those outside of the *Sophophora-Lordiphosa* radiation (see Methods for how to view the full output of these analyses). Thus, we fail to find evidence of a burst in evolutionary change in *SPR* that correlates with the lineage-accelerated evolutionary changes we record for *SP*.

SPR sites with evidence of episodic positive selection are disproportionately located in predicted extracellular facing domains

Although the dominant signature of variation between SPR sequences fell within the extracellular N-terminus, several residues in the extracellular loops and extracellular facing transmembrane domains – regions likely to be critical for ligand-binding – showed reduced conservation (Figure 8A,F; Supplementary movie 1). To assess whether these sites are under selection, we used *FUBAR* (Fast, Unconstrained, Bayesian Approximation)(Murrell *et al.*, 2013) to test for evidence of pervasive diversifying selection at individual sites in the N-terminus trimmed SPR sequence. We detected evidence of *pervasive* diversifying selection at 0/377 sites and purifying selection at 357/377. We followed this analysis with a test for *episodic* positive selection at individual sites under a proportion of branches, implemented through MEME (Mixed Effects Models of Evolution)(Murrell *et al.*, 2012). In this analysis, we detected evidence of episodic positive selection at 10/377 sites. Running the same analysis separately for the *Sophophora-Lordiphosa* (91 species) and non-*Sophophora-Lordiphosa* species (89 species), we identified 4/371 and 5/376 sites respectively as showing evidence of episodic positive selection (the identity of these positively selected sites did not overlap between the two analyses). Therefore, we found no evidence that the proportion of sites experiencing episodic positive selection was elevated in the *Sophophora*, which includes the lineages in which *SP* showed greatest evolutionary change.

Intriguingly, of the 10 positively selected sites identified in the phylogeny-wide analysis, 9 fell within extracellular domains: 3 in the N-terminus region close to the start of the first transmembrane domain, 6 across the three extracellular loops, and then one in the third intracellular loop. We therefore asked whether these residues are important in binding SP. To the best of our knowledge, the binding site of SP remains unknown within the SPR protein sequence. We therefore used ColabFold (Mirdita *et al.*, 2017; Mirdita, Steinegger, & Söding, 2019; Mirdita *et al.*, 2022; Mitchell *et al.*, 2019), which combines the fast homology search of MMseqs2 with AlphaFold-multimer, to predict the contact residues between *D. melanogaster* SP and SPR. The top ranked model had an interface pTM (ipTM) score of 0.812 (additional metrics: local structural accuracy, pLDDT = 69.5; overall topological accuracy, pTM = 0.711; Supplementary figure 9A-C). Recent work has used

a stringent cutoff of ipTM = 0.85 for calling high-confidence protein-protein interactions, observing that values between 0.55-0.85 perform better than random, with increasing accuracy at higher values (O'Reilly *et al.*, 2023). Thus, the prediction falls a little short of the high-confidence cutoff, but still suggests a reasonable level of confidence. As a further test, we compared the ipTM scores between *D. melanogaster* SP and each of the 44 members listed on Flybase of the Class A/rhodopsin-like G-Protein Coupled Receptors family to which SPR belongs. SPR gave the highest score of all (Supplementary figure 9D). Using the interfaces and H-bond functions in ChimeraX (Pettersen *et al.*, 2021), we detected 33 residues in SPR that were predicted to interface with SP, of which 12 were additionally predicted to form hydrogen bonds with SP residues (Figure 8G-I; Supplementary movies 2, 3). Most of these residues were highly conserved across the 193 sequences: 12/33 were 100% conserved and 20/33 were >99% conserved (Figure 8I). But this level of conservation was not atypical among the extracellular facing residues: those residues that neighboured the predicted contact residues were similarly well conserved (Supplementary figure 9E). Of the 13 less well conserved residues, none showed clear evidence of concerted change among *ananassae* subgroup, Oriental lineage, or *Sophophora* species. However, there was a significant enrichment of sites showing significant evidence of episodic positive selection among the 33 predicted contact residues (4/33; χ^2=12.56, *df* =1, *p*=0.0004; marked by asterisks in Figure 8I,J).

Overall therefore, while we do detect evidence that sites in the putative SP-binding pocket of SPR have undergone episodic positive selection, the number of changes does not appear to be elevated in the *Sophophora-Lordiphosa* radiation, where the major genomic and functional changes (e.g., Tsuda *et al.*, 2015) in SP have occurred.

A validated mutational route that SPR could take to decouple responses to SP and MIPs remains unexploited in drosophilids

If receipt of SP is associated with a net reduction in female fitness, the potentially deleterious effects of disrupting MIP-SPR interactions may constrain SPR's ability to evolve to defend against SP binding. However, we know from previous work that substitution of certain residues in SPR can have decoupled effects on the receptor's sensitivity to its different ligands. Specifically, replacing the QRY motif at the boundary between the second intracellular loop and third transmembrane domain with the DRY motif more widely found in class A GPCRs is associated with a decrease in the responsiveness of SPR to SP, but not the ancestral MIP ligands, in *in vitro* assays (Poels *et al.*, 2010). Yet in no drosophilid did we detect change at this position. Therefore, this potential avenue through which a single amino acid sequence change could reduce sensitivity to SP without affecting pre-existing ligand interactions remains unexploited.

Discussion

Over the past few decades, we've built up a detailed understanding of the function of SP in *D. melanogaster*. We know that it is required for the normal assembly and disassembly of seminal storage and trafficking structures ('microcarriers', Wainwright *et al.*, 2021); that it triggers an extensive range of physiological and behavioural changes in females, at least some of which are mediated by its interactions with SPR in female reproductive tract neurons (Yapici *et al.*, 2008; Häsemeyer *et al.*, 2009; Rezával *et al.*, 2012); and that SP's

effects in females are extended via its binding to the surface of sperm, a process facilitated by a network of other male-derived proteins (Peng *et al.*, 2005a; Ram & Wolfner, 2007; Ravi Ram & Wolfner, 2009; Findlay *et al.*, 2014; Singh *et al.*, 2018). And yet, previous work has suggested that despite its integral roles in *D. melanogaster* reproduction, and despite the complex sperm-binding machinery with which it interacts, *SP* is restricted to drosophilids and perhaps, therefore, a drosophilid innovation. Consequently, *SP* represents a powerful system in which to chart the origin and diversification of function in a novel gene across different lineages. The cross-species analysis of *SP* and *SPR* genes that we report here makes several contributions to this.

The first relates to the phylogenetic distribution of *SP* genes. Previous work traced *SP* as far back as the split between *D. melanogaster* and *D. virilis* and identified three apparently independent loss events in the non-*Sophophora* species *D. grimshawi, D. mojavensis,* and *D. albomicans* (Kim *et al.*, 2010; Tsuda *et al.*, 2015; McGeary & Findlay, 2020). The data we present here pushes the origin of *SP* back deeper into the drosophilid tree. Based on the available taxon sample, *SP* appears to have originated at least as far back as the split between the *Scaptodrosophila* and *Drosophilini.* We also showed that *SP* is present in a genome of *D. albomicans* and that the losses in *D. grimshawi* and *D. mojavensis* are not species-specific, but instead cover larger radiations: one, a monophyletic lineage of the 20 Hawaiian species that we sampled; the other, a monophyletic lineage of 14 sampled species that includes members of the *annulimana*, *bromeliae*, *nannoptera,* and *repleta* groups. Alongside these, we detected evidence of several more losses, namely in *H. duncani* and in a lineage of *Scaptomyza*. Given the critical role of *SP* in many aspects of *D. melanogaster* reproduction, these losses are intriguing. However, there is data that suggests the functions of SP vary between species, despite the apparent conservation in expression pattern we show. Although the taxon sampling has been somewhat sparse, SP injection experiments suggest that the ability of SP to reduce female sexual receptivity is restricted to the *melanogaster* group (Tsuda *et al.*, 2015). It's therefore possible that outside of this lineage, SP performs more limited functions that are easier to dispense of – hence the losses. Alternatively, where *SP* has been lost it may be that its functions have been taken over by non-homologous proteins. Indeed, there is a clear precedent for non-homologous reproductive proteins to be used to achieve similar phenotypic endpoints in different species (moths, Kingan *et al.*, 1995; e.g. mosquitoes, Duvall *et al.*, 2017).

The second is our detection of substantial variation in copy number and sequence in the *Sophophora-Lordiphosa* radiation*,* a feature that was particularly pronounced in the "Oriental" lineage and *ananassae* subgroup. Outside of the *Sophophora-Lordiphosa, SP* is almost invariably a single (or 0) copy gene. But inside*,* where we detect repeated, independent duplication of *SP* across lineages, the story is quite different. Repeated duplication is at its most extreme in the *ananassae* subgroup, where we see as many as 7 copies present in *D. pseudoananassae nigrens* and 6 copies, all of which we found to be specifically expressed in males, in *D. bipectinata*. Importantly in this subgroup, the intraspecific paralogs are not identical in sequence. Instead, they generally showed considerable variation in length and amino acid composition. What, then, are the functional consequences of this intraspecific diversity? If these paralogs are all interacting

with SPR and varying to different degrees in their C-terminal sequences, do they vary in the efficiency with which they bind SPR? If so, is there a risk that low-efficiency copies block the access of higher efficiency copies to binding pockets? Perhaps, then, different paralogs are specialised for different receptors. We know, for example, that *D. melanogaster* SP can activate another G-protein coupled receptor (GPCR), *Methusaleh (Mth), in vitro.* However, SP-Mth interactions don't seem to be required for the post-mating increase in egg production or reduction in sexual receptivity – at least in *D. melanogaster* (Ja *et al.*, 2009). There is also evidence that SP can induce some of its effects in *D. melanogaster* independently of *SPR,* pointing to the potential existence of additional, unidentified receptors (Haussmann *et al.*, 2013). Thus, it is not unreasonable to suggest that *ananassae* species might be evolving to make use of a wider set of receptors, which techniques such as Evolutionary Rate Covariation analysis (*e.g.* Findlay *et al.*, 2014) may help us to identify.

But what about the other regions of SP proteins, the regions beyond the post-mating response stimulating C-terminus? After all, it's the N-terminus and midsection regions that we show to be most variable between homologs. In *D. melanogaster*, we know that SP binds to sperm at its N-terminus; the N-terminal WEWPWNR motif then remains bound to SP after the rest of the peptide is cleaved (Peng *et al.*, 2005a). Our data suggest that clear variants of this motif are restricted to at least one SP copy in the Muller D2 neighbourhood of each member of the "Oriental" lineage (with the exception of *D. oshimai*). This raises two possibilities: (1) the ability of SP to bind sperm, which has to the best of our knowledge only been tested and reported in *D. melanogaster,* is restricted to SP copies carrying variants of this motif in this lineage; (2) the protein network that underlies SP-sperm binding is able to facilitate attachment using a wide set of N-terminus sequences. Such flexibility may stem from rapid evolution of the sequence or identity of SP network proteins, or of any sperm surface proteins that SP might interact with. Indeed, there is evidence that several sex peptide network proteins have experienced recurrent positive selection in the *melanogaster* group (McGeary & Findlay, 2020). Flexible use of N-terminus sequences may also be due to some inherent, accommodating property of the sperm binding apparatus. Consistent with this, no such WEWPWNR motif is present in *D. melanogaster* Dup99b, but it nevertheless binds to sperm, albeit only to the sperm head and only during the first few hours after mating, unlike SP (Peng, Zipperlen, & Kubli, 2005b). But whether Dup99b is relying on the same network of proteins as SP to bind to sperm, or perhaps interacting with distinct proteins on the sperm surface, remains untested. Where multiple, divergent SP copies are present, as in the *ananassae* subgroup for example, we may be seeing specialisation in the N-terminus region that relates to a given peptide's mechanism of interacting with the female: while some may be adapted to bind sperm, others might be adapted to enter into the hemolymph, as *D. melanogaster* SP has been shown to do (Pilpel *et al.*, 2008).

The third relates to the genomic distribution of *SP* genes. We observed that *SP* genes have frequently translocated to new genomic locations, a feature that's particularly pervasive in the *Sophophora.* A key question here is to what extent these translocations have shaped the evolution of SP function. Did the 'Muller D1' to 'Muller D2' translocation at the base of the *Sophophora-Lordiphosa* lineage, for example, open the door to the

novel evolutionary trajectories taken within this lineage? We could imagine that this translocation placed *SP* within a new *cis*-regulatory environment that changed either the strength, timing, or tissue-specificity of its expression, thereby exposing it to new selective forces. Our data suggest against this at a global level, at least in relation to tissue-specificity, as we observed accessory gland biased expression of *SP* in *D. virilis*, a species that pre-dates the translocation. However, it's possible that subsequent translocations within some branches of the *Sophophora-Lordiphosa* radiation are associated with shifts in expression pattern. After all, the translocated Muller element E copy of *SP* in *D. melanogaster* (*Dup99b*) is expressed not in the accessory glands, like *SP,* but in the ejaculatory duct (Rexhepaj *et al.*, 2003).

The fourth is our failure to detect a clear association between evolutionary change in *SP* and microcarriers. That a relationship between the two might exist was previously suggested on the basis that (a) microcarrier morphology is markedly changed following the loss or knockdown of *SP,* and (b) substantial variation exists between species in both SP sequence and microcarrier morphology (Wainwright *et al.*, 2021). However, we found multiple examples that argue against this association. For one, the presence of LipidTox$^+$ microcarriers in *S. flava,* which lacks a copy of *SP*, suggests that *SP* isn't necessary for microcarriers. Moreover, the absence of canonical LipidTox$^+$ microcarriers in *SP*-encoding species, such as *D. immigrans*, suggests that a copy of *SP* isn't sufficient. Thus, SP playing a role in structuring microcarriers might itself be a derived trait, as might its association with microcarriers more broadly (*e.g.,* whether it acts as a microcarrier cargo), thereby casting doubt on the idea that this could be SP's ancestral function (Hopkins & Perry, 2022). That association might be relatively recent, as even among many *SP*-bearing species within the *Sophophora*, variation in *SP* copy number and sequence diversity didn't seem to be obviously connected to variation in microcarrier morphology. The microcarriers of *D. bipectinata,* for instance, bore close resemblance to those of *D. pseudoobscura* and *D. persimilis* (the latter were reported by Wainwright *et al.*, 2021). That is despite *D. bipectinata* encoding 6 copies of *SP*, all of which varied in sequence to the single copies of the two *obscura* group species. Our stainings also raise the question of what exactly defines a 'microcarrier'. While we recorded canonical microcarriers in many species – *i.e.,* those conforming to a stringy, spherical, or ellipsoid conformation and clearly stained by LipidTox – the luminal contents of others were more ambiguous. For example, the LipidTox$^-$ 'flocculence' of *D. immigrans* (reminiscent of *D. virilis*; Wainwright *et al.*, 2021), the LipidTox$^-$ ellipsoids of *S. hsui,* and the vacuolar structures of *D. varians.* To what extent these all represent homologous structures remains unclear. Regardless, the ability of structures in the accessory gland lumen to take up LipidTox appears to have been gained and lost repeatedly, a feature that may point to changes in their architecture or lipid content, the latter of which may derive from higher-level dietary or metabolic changes.

Our data point to lineage-accelerated evolution of *SP* within the *Sophophora-Lordiphosa* radiation, marked by repeated, independent rounds of gene family expansion. What forces are driving this trend? And why does it appear to be so much more pervasive in these lineages? A tempting response to the latter question is that it coincides with some evolutionary shift in the activity of SP. As previously mentioned, an injection study of 11

drosophilid species found that the ability of conspecific SP to reduce female sexual receptivity is confined to the *melanogaster* group (Tsuda *et al.*, 2015). Whether the many other effects of SP on female physiology and behaviour that have been documented in *D. melanogaster* are similarly phylogenetically restricted remains, to the best of our knowledge, untested. Intriguingly, this gain in responsiveness to SP also appears to coincide with both the gain of robust expression of *SPR* within the female reproductive tract and the ability of SP (derived from *D. melanogaster*) to bind to female reproductive tract tissue (Tsuda *et al.*, 2015). Viewed from a sexual conflict perspective, therefore, we might be detecting the effects of a sexually antagonistic coevolutionary arms race that was initiated after the acquisition of new functions by SP. If responding to SP is deleterious to female fitness (*e.g.* Wigby & Chapman, 2005), then females might evolve resistance, which in turn selects for structurally divergent copies of *SP* through which males can overcome that resistance. Indeed, the multiple copies of *SP* of varying sequence that are exhibited by individual members of the *ananassae* subgroup are consistent with theory suggesting that males might gain from transferring numerous, diverse sexually antagonistic seminal fluid products simultaneously as a 'combination' strategy to overcome the evolution of resistance (Chapman, 2018).

But if females are evolving resistance, then we fail to find strong evidence that it is occurring through *SPR*: we find no clear and obvious burst in evolutionary change that coincides with the lineage-accelerated evolution of *SP*. The only region of SPR where we see extensive evolutionary change is in the N-terminus. But there are reasons to believe that any role this region plays in ligand binding is relatively minor. For one, the only predicted contact residues within the N-terminus that we identified were all directly adjacent to the transmembrane domain and fell within the region that showed highest conservation. A counter to our failure to detect contact residues in the faster evolving regions of the N-terminus would be to point out that the 3D structure of the N-terminus was predicted with the lowest confidence of any part of the protein, and we may therefore be missing potential interaction sites with SP. That being said, the low confidence of this region's structure is also consistent with it being intrinsically disordered, a feature known to be present in the N-termini of other GPCRs such as human beta-2 adrenergic receptor (Cherezov *et al.*, 2007). This disorder may reflect a lack of functional importance for ligand binding and evolution under relaxed selection. Indeed, the N-termini of other GPCRs appear to be considerably more mutationally tolerant than other parts of the protein (Jones *et al.*, 2020). However, there are examples, such as rhodopsin, where the N-terminus plays a key role in structuring the region surrounding the ligand-binding site (Cherezov *et al.*, 2007). These broader features of GPCRs aside, the best evidence we have against the N-terminus of SPR being a strong driver of responsiveness to SP comes from previous domain swapping work, where different domains of *D. melanogaster* SPR were swapped with the homologous regions from *A. aegypti* SPR and the effects on SP binding tested using *in vitro* assays. The N-terminus swap, which extended deep into the first transmembrane domain, conducted as part of this previous work had a much more modest effect on SP binding relative to swapping regions containing transmembrane and extracellular loop domains (Lee *et al.*, 2020).

Beyond the N-terminus, we did detect evidence of episodic, but not pervasive, positive selection along branches at 10 different sites in SPR. All but one of these sites fell within the extracellular facing, and therefore potentially ligand-binding, domains and 4 were predicted contact sites with SP in *D. melanogaster.* However, the number of sites for which we detected evidence of positive selection was similar between sperate analyses of the *Sophophora-Lordiphosa* (4 sites) and non-*Sophophora-Lordiphosa* (5 sites) radiations. Therefore, while our data suggests that the ligand-binding domains of SPR are evolving under episodic positive selection, that change is distributed throughout the phylogeny, rather than being restricted to the lineages experiencing accelerated evolution of *SP.*

Why, then, are we not seeing SPR rapidly evolving in concert with SP? The classic explanation here is that the evolution of SPR is constrained by the need to maintain interactions with additional ligands, such as MIPs. But we know from *in vitro* assays that there is at least one mutational route for SPR that has decoupled effects on MIP and SP binding, negatively affecting the latter more strongly than the former (Poels *et al.*, 2010). And yet across 193 drosophilid SPR sequences we failed to find any instance where this route has been exploited. Another possibility is that any resistance to SP-mediated sexual antagonism might be mediated up- or down-stream of SP-SPR binding, such as via mechanisms that degrade SP, block it from binding, or in how the neural circuitry that SPR feeds into responds to SP. Genes involved in these processes may represent better targets for selection if they don't carry the same pleiotropic baggage as SPR, a receptor that we know is deeply integrated into pathways with additional ligands. One such potential action point is sperm cleavage: if females could block the cleavage of SP from the surface of the sperm – the mechanism for which remains uncharacterised – then they could markedly reduce the timeframe over which SP's effects are active (~10-14 days in D. melanogaster; Peng *et al.*, 2005a).

But limited evolutionary change in *SPR* is also consistent with another possibility: the role of sexual conflict in driving the evolution of SP is relatively weak. While the functions of SP are often framed in terms of male manipulation of female reproductive decision-making or collateral damage in the pursuit of improved performance in sperm competition (*e.g.* Bussiégre *et al.*, 2006; Pröschel, Zhang, & Parsch, 2006; Price *et al.*, 2010; Slatyer *et al.*, 2012), the evidence base for this is not strong (reviewed in Hopkins & Perry, 2022). There is theoretical support for antagonistic effects of seminal proteins in general (*e.g.* Johnstone & Keller, 2000) and empirical support in *D. melanogaster* for the antagonistic effects of seminal proteins (Chapman *et al.*, 1995) and SP specifically (Wigby & Chapman, 2005; Tower *et al.*, 2017), but there is also empirical support for positive, neutral, and context-dependent (*i.e.,* in relation to female nutritional state) effects of SP on female fitness (Rogina, 2009; Fricke *et al.*, 2010; Wensing & Fricke, 2018), as well as considerable uncertainty over the extent to which fitness measurements made in these laboratory studies reflect those experienced by wild-living populations (Hopkins & Perry, 2022). Indeed, the recently demonstrated gain of both robust *SPR* expression in a subset of female reproductive tract neurons and the ability of SP to bind to reproductive tract tissue in the *melanogaster* group (Tsuda *et al.*, 2015) suggests that females did, at least at some point, benefit from responding to SP, perhaps as part of a mechanism through

which the receipt of sperm could be aligned with the induction of reproductive processes. Whether that may have subsequently initiated conflict – and what the associated genomic consequences might have been – remains to be resolved. But by charting the origin and diversification of SP-female interactions we may be able to draw more robust inferences regarding the evolutionary forces that are shaping the distinct trajectories of SP across *Drosophilidae*.

Methods:

Genomes. 206 genome sequences for 199 species (with 2-3 genomes from different strains for 6 species) were obtained in the same manner as described in previous work (Kim *et al.*, 2021; Suvorov *et al.*, 2022). We sequenced and *de novo* assembled genomes for 43 species (Kim et al., *in* prep), downloaded 155 publicly available assemblies from NCBI databases (Scott *et al.*, 2014; Dos Santos *et al.*, 2015; Vicoso & Bachtrog, 2015; Sanchez-Flores *et al.*, 2016; Mahajan *et al.*, 2018; Puerma *et al.*, 2018; Matthews *et al.*, 2018; Vanderlinde *et al.*, 2019; Ellison & Bachtrog, 2019a, 2019b; Hill, Koseva, & Unckless, 2019; Bracewell *et al.*, 2019, 2020; Renschler *et al.*, 2019; Mai, Nalley, & Bachtrog, 2020; Bronski *et al.*, 2020; Paris *et al.*, 2020; Torosin *et al.*, 2020; Liao *et al.*, 2021; Chakraborty *et al.*, 2021; Conner *et al.*, 2021; Tvedte *et al.*, 2021; Kim *et al.*, 2021; Wei *et al.*, 2022; Sottolano *et al.*, 2022; Pelaez *et al.*, 2023), and downloaded unassembled reads from the NCBI Sequence Read Archive for 8 species (Mai *et al.*, 2020; Suvorov *et al.*, 2022). Detailed sample information and accession numbers are provided at https://osf.io/z3a8k. We sequenced on Oxford Nanopore MinION and Illumina NovaSeq instruments and assembled *de novo* genomes with the hybrid approach described in Kim *et al.* (2021) and Kim *et al.,* (in prep). Genomes of species for which only short read data were available were assembled with SPAdes 3.13.1 (Bankevich *et al.*, 2012).

Phylogeny. A species tree was generated from 250 orthologs, again following the workflows established in previous work (Kim *et al.*, 2021; Suvorov *et al.*, 2022). Briefly, we identified single-copy dipteran orthologs using BUSCO v5 (Manni *et al.*, 2021) and randomly selected 250 genes identified as single-copy in the assemblies and missing for no more than 10 species. Each ortholog was individually aligned with MAFFT (Katoh & Standley, 2013) and gene trees were inferred with IQTREE 2.2.0.7 (Minh *et al.*, 2020). A species tree was inferred from the gene trees with ASTRAL-III (Zhang *et al.*, 2018).

Identifying SP genes. To identify SP orthologs, we first searched each genome by tBLASTn using a liberal E-value cut-off of 100. As the query, we used the *D. melanogaster* C-terminus sequence (KWCRLNLGPAWGGRC), which is entirely encoded by the second exon. This is the best conserved domain between SP and Dup99b in *D. melanogaster* (Saudan *et al.*, 2002) and which previous work on SP sequences from 10 species had shown to be the only strongly conserved region (Tsuda *et al.*, 2015). We extracted 350nts either side of the start and end points of each returned hit and reciprocally queried the extracted sequence by BLASTx against the *D. melanogaster* reference protein list, retaining only those sequences that gave an SP paralog (SP or

Dup99B) as the top hit. The extracted sequence length is more than sufficient to capture the full *D. melanogaster SP* (288nts) and *Dup99b* (310nts) genes. To be included as an *SP* gene in the total species counts (*i.e.,* those given in Figures 1 and 2), a sequence had to meet these BLAST criteria and give a recognisable SP-like protein sequence when aligned to the *D. melanogaster SP* gene structure (*e.g.,* no premature stop codons). For species that returned no hits through this approach, we used two follow up approaches (within each of the Hawaiian and *annulimana*, *bromeliae*, *nannoptera* and *repleta* ('ANBR') radiations, we chose 5 representative species). First, we used tBLASTn to query the full-length, including N-terminus, resolved SP protein sequences from a selection of 11-14 species from across our phylogeny against each apparently *SP*-less genome. In only one of 13 species tested using this approach (*D. neohypocausta*) did we detect a *SP* homolog. In the remaining 12 species we detected nothing that resembled a *SP* homolog, nor any individual hit that fell within the Muller D1 neighbourhood despite using a liberal E-value cut-off. Second, we repeated this approach using tBLASTn to query a selection of 6-9 full-length SP sequences from a phylogenetically varied set of species against only the Muller D2 neighbourhood of each of the remaining 12 species. In no case did we detect evidence of any remaining fragment of *SP.*

Mapping the syntenic neighbourhoods of SP genes. To position *SP* hits within their chromosomal context, we sought to identify and reconstruct syntenic gene neighbourhoods surrounding each *SP* hit. To do this, we first used the NCBI genome viewer to identify and extract the protein sequences of genes immediately up and downstream of *SP* in the *D. melanogaster* reference genome (DmelRS6). In chromosomal order, these genes were *tartan, CG33262, snky, capricious, Sfp70A4, CG42481, CG43147, (SP), CG14113, CG17687, Nplp2,* and *CG14111.* We then searched each of these sequences against each species' genome using tBLASTn, extracting the coordinates for the top returned hit in each case. Of these genes, we focused on *capricious, CG17687,* and *CG14111* as these were particularly well-conserved. In a minority of cases, contig breakpoints fell within a neighbourhood preventing us from fully reconstructing the region. In one case (*D. ananassae*), we were able to stitch two contigs together that spanned an *SP*-containing neighbourhood by cross referencing to the *D. ananassae* reference genome (DanaRS2.1). In another (*L. mommai*), we used the other *Lordiphosa* species to stitch two Muller D2 contigs together that separated *capricious* from *SP, CG14111,* and *CG17687,* allowing us to plot the position of *SP* in Figure 2. In a small number of cases, this search revealed the presence of *SP* genes within the syntenic gene neighbourhood that hadn't passed our reciprocal BLASTx test. These were *D. nasuta, D. immigrans, D. immigrans (kari17)*, *D. guttifera, D. sulfurigaster bilimbata, D. seguyi,* and *D. oshimai*. We included these for further analysis based on their shared chromosomal location with *SP* genes that did pass the reciprocal blast test in other species and the SP-like protein sequences that we were able to resolve for them.

For SP hits left unlocated by this approach, we first repeated this same process for the gene neighbourhood surrounding *Dup99b* (*CG34296, dmrt99b, gycalpha99b*). If that failed to identify their position, we extracted variable, larger regions of sequence surrounding the hit on each *SP*-containing contig and used a combination of BLASTn and

tBLASTn searches against the *D. melanogaster* reference genome. From this, we identified syntenic gene neighbourhoods, extracted the protein sequences for these *D. melanogaster* genes from Flybase (Gramates *et al.*, 2022), and repeated the approach used for *SP* and *Dup99b* with these new neighbourhood genes. This analysis showed that in almost all species SP genes were located within a small number of highly conserved syntenic neighbourhoods. Neighbourhood plots were constructed in R (V4.1.1) using bespoke scripts that used ape (Paradis & Schliep, 2019), ggtree(Yu, 2020), and ggplot2 (Wickham, 2016).

Reconstructing SP gene structure and protein sequence. To annotate *SP* gene structure, we generated a MAFFT alignment (Madeira *et al.*, 2022) of all our identified *SP* sequences and identified landmarks (start and stop codons, exon/intron boundaries) that were conserved with the *D. melanogaster SP* and *Dup99b* sequences. This approach was previously used by Tsuda *et al.* (2015) on a smaller bank of 10 *SP* sequences. As in Tsuda *et al.* (2015), our alignment revealed broad conservation of both the start and stop codons and the intron donor (consensus: GTAAGT) and acceptor (AG) sequences across species. In the small number of sequences where these landmarks were not present at the conserved position in our alignment, we selected the closest plausible option to the consensus sequence position that was conserved among similar species. We further required that the selected option had to maintain the KWCRLNLGPAWGGRC-like reading frame in exon 2. Of 266 identified SP-like nucleotide sequences, we could resolve SP-like protein sequences for all but 34, most of which fell outside syntenic gene neighbourhoods that housed *SP* in other species, suggesting that they are unlikely to be real *SP* genes. Signal peptides were removed from the translated protein sequences using SignalP-6.0 run on 'fast' mode and specifying 'Eukarya' as the organism (Teufel *et al.*, 2022). A signal peptide sequence was detected with a probability of >0.9992 for all but four sequences, two of which came from the *montium* subgroup species *D. punjabiensis* (0.0584 and 0.0055) and the other two from its sister species in our dataset *D. watanabei* (0.1284 and 0). In these cases, it might be that these represent non-homologous sequences that bear a resemblance to SP in the C-terminus or that these SP proteins are no longer secreted into the extracellular environment.

Expression analysis. For 19 species, we identified RNA-seq datasets on NCBI that we could use to test for expression of our predicted SP orthologs. For each of these species, we used BLASTn to search each ortholog against a given species' genome. We then used the graphics option in the NCBI blast suite to view the hit mapped to the genome, extracted the associated GeneID and gene class identity, and loaded available RNA-seq exon coverage tracks. For each loaded track, we recorded the associated metadata (tissue, sex, dataset accession, *etc.)* and the peak exon coverage within the SP ortholog coding sequence. For each RNA-seq sample, we extracted the equivalent value at the housekeeping gene *RpL32* locus. We calculated relative expression values by adding 0.1 to the peak *SP* exon coverage value and dividing it by the equivalent *RpL32* value. We further used each *SP* hit in each species for which we had RNA-seq data to confirm our correct annotation of exon/intron boundaries.

During this analysis, we observed that many of the *SP* genes that we detect are incorrectly annotated in the species' reference genomes. In the case of *D. bipectinata,* 4 of its *SP* genes are designated as 'uncharacterised long non-coding RNAs' (LOC108127578, LOC108127576, LOC108127577, and LOC108127621). This was a common designation for *SP* genes, one that we also saw for *D. kikkawai* (LOC108075137), *D. takahashii* (LOC108063118 and LOC108063123, but LOC108062332 was listed as a *SP* homolog), *D. ananassae* (LOC116655869 and LOC26513946), *D. willistoni* (LOC124460714), *D. virilis* (LOC26531923), and *D. ficusphila* (LOC108095416 and LOC108100235**,** but LOC108095478 was listed as a *SP* homolog). Some *SP* homologs that we detected in each of *D. biarmipes and D. ficusphila* showed clear male-biased expression but were unaccompanied by any gene annotation, not even a long non-coding RNA (Supplementary figure 10). Thus, automated genome annotation pipelines appear to often fail to recognise the homology of *SP* genes and, in some cases, fail to recognise them as genes entirely. These issues likely stem from a combination of the small size of *SP* family genes – a feature known to cause problems for gene finding tools (Warren *et al.*, 2010; Belkorchia *et al.*, 2015) – and it's low conservation beyond the C-terminus.

SP *protein sequence analysis using PCA.* PCA was performed using Jalview (v2.11.2.5) on a BLOSUM62 substitution matrix of all resolved SP protein sequences (Waterhouse *et al.*, 2009). The transformed values were extracted from JalView and plotted using the gpplot2 package in R (v4.1.1).

Constructing a* SP *protein tree. We inferred the evolutionary history of Sex Peptide proteins applying the Maximum Likelihood method with a Jones-Taylor-Thornton (JTT) matrix-based model (Jones, Taylor, & Thornton, 1992) to a MUSCLE alignment of 233 extracted SP amino acid sequences. The tree with the highest log likelihood (-10372.14) is presented in Supplementary figure 2. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. This analysis was conducted in MEGA X (Kumar *et al.*, 2018; Stecher, Tamura, & Kumar, 2020).

Fly rearing. The following strains were used for microcarrier stainings: *D. melanogaster* (RAL-517; Mackay *et al.*, 2012), *D. malerkotliana malerkotliana* (mal0-isoC), *D. malerkotliana pallens* (Q120-isoG), *D. pseudoananassae pseudoananassae* (wau125), *D. pseudoananassae nigrens* (VT04-31), *D. varians* (CKM15-L1), *D. ercepea* (aag001 copy2), *D. bipectinata* (14024-038.07 nanopore), *D. ananassae* (OGS-98K1), *D. biarmipes* (361.3-iso11 1-11) *D. ficusphila* (iso1 L.10), *D. kikkawai* (4_2_3_2_1_2_1_1), *D. immigrans* (15111.1731.12), *D. rufa* (EH091 isoC L-3), *D. carrolli* (KB866), and *D. takahashii* (14022-0311.14). For most species, virgin males were collected within 8 hours of eclosion, held in groups of 7-10 individuals, and aged 8-10 days before their dissection and staining. The exceptions were *D. melanogaster*, which were used at 3-5 days (as in Wainwright *et al.*, 2021); *D. carrolli*, whose virgin status was unknown but which had been isolated from females 48h prior to dissection to allow accessory gland replenishment; and the *Scaptomyza* species, adults of which were generously provided by Julianne Peláez and Noah Whiteman. *S. pallida* and *S. hsui* were maintained on standard Bloomington

media with thawed frozen chopped spinach. Non-virgin males from these species aged 4-6 days in *S. pallida*, 4-8 days in *S. hsui*, and 1-4 days for *S. flava* were used.

Microcarrier stainings. Male reproductive tracts were dissected in 1x PBS in glass wells. Prior to staining, the PBS was removed from the wells and the tracts fixed for 30 mins in 200µl of 4% PFA (in PBS) solution on a rocker. The PFA solution was then removed, and the tracts subjected to three, 2-minute washes in 1x PBS on a rocker. After washing, we removed all excess tissue (including testes and seminal vesicles) except the accessory glands and pierced the glands with forceps to facilitate dye penetration (as in Wainwright *et al.*, 2021). Accessory glands were then stained in 100µl of 1:50 LipidTox Deep Red Neutral Lipid Stain (Invitrogen, H34477) in PBS for 1 hour with 1µl 1:100 DAPI added for the last 15 mins. During staining, the accessory glands were placed onto a rotator and covered to prevent exposure to light. After staining, the glands were washed 3 times in 1x PBS for 2 minutes each on the rotator. Finally, glands were mounted in 40µl Fluoromount 50 (Southernbiotech) on polylysine-coated slides and refrigerated until imaging. For each species, accessory glands from *D. melanogaster* were co-stained in the same well as a control to ensure successful staining. The species identity of mounted accessory glands was readily identifiable by morphological features of the accessory glands and microcarriers. At least 5 glands were mounted per species. Confocal images were taken using an Olympus FV1000 laser scanning confocal microscope.

Identifying SPR orthologs. To identify *SPR* orthologs, we used tBLASTn to separately query the coding sequence of each of the four *D. melanogaster SPR* exons against each species' genome. Each returned hit was then reciprocally blasted against the *D. melanogaster* reference protein list and all sequences that gave SPR as their top hit were extracted along with 1500 nucleotides up- and down- stream. The coordinates of each retained hit for each exon in each species were then extracted. We removed any sequences from one exon that were not contiguous with sequences from another and where all four contiguous exons were present (*e.g.,* an exon 1 hit on contig_A would be dropped if exons 1 through 4 were tandemly arranged on contig_B). All sequences from a given exon were then aligned using MAFFT and manually annotated based on features conserved with the *D. melanogaster SPR* sequence. The start codon of exon 1 aligned poorly at a phylogeny-wide level. In these cases, we selected the start codon closest to one that was conserved among closely related species that gave an SPR-like coding sequence. For 5 species, we were unable to resolve SPR sequences. In *Z. taronus*, a premature stop codon was present within the coding sequence of exon 4. For *S. montana* and *S. flava,* we failed to detect SPR-like sequences across either of its 4 exons when each *D. melanogaster SPR* exon was queried against the two species' genomes. In these species, the top hit for each exon gave a markedly lower score than the median across all species studied, suggesting loss of the whole protein-coding sequence (Supplementary figure 8). In the cases of *S. hsui* and *L. magnipectinata*, we found only 2 of the 4 exons gave a clear SPR match (*L. magnipectinata*: exons 1 and 4; *S. hsui:* exons 3 and 4), suggestive of a partial loss of coding sequence. Specific domains and their corresponding positions within the SPR protein sequence were extracted from the 'Features' section of the UniProt entry for SPR (Q8SWR3).

Molecular evolution analysis. To test for selection in SPR coding sequence, we used a set of methods available through the 'Datamonkey' adaptive evolution server (Delport *et al.*, 2010; Weaver *et al.*, 2018). We used aBSREL (adaptive Branch-Site Random Effects Likelihood) to test if positive selection has occurred on a proportion of branches (Kosakovsky Pond *et al.*, 2011; Smith *et al.*, 2015), BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification) to test whether *SPR* has experienced positive selection at one or more sites on one more branches (Murrell *et al.*, 2015), FUBAR (Fast, Unconstrained Bayesian AppRoximation) to detect sites evolving under pervasive positive selection in a subset of branches (Murrell *et al.*, 2013), and MEME (Mixed Effects Model of Evolution) to detect sites evolving under episodic positive selection in a subset of branches (Murrell *et al.*, 2012). Each analysis was performed using default parameters on the conserved region (*i.e.,* N-terminus removed) of SPR sequences, aligned using a codon alignment generated using PAL2NAL (Suyama, Torrents, & Bork, 2006) on the MAFFT-aligned protein sequences. For these analyses, the 193 sequences used to generate the consensus sequence plot was reduced to 182 to drop 3 sequences that were not in our phylogeny (*D. chauvacae*-like*, Leucophenga varia,* and *D. sp st01m*) and 8 where we had multiple lines per strain. A single strain was used for each of these species: *Z. indianus* (BS02)*, D. paulistorum* (14030-0771.06)*, D.immigrans* (15111.1731-12)*, D. robusta* (Kim *et al., in prep*)*, D. teissieri* (CT02)*,* and *D. willistoni* (14030-0811.00). Alongside the full species set, we repeated these analyses separately on alignments of the *Sophophora-Lordiphosa* and non-*Sophophora-Lordiphosa* species. A .json file for the results of each analysis is available at https://osf.io/tzu6v along with instructions of how to view the full output using https://vision.hyphy.org/.

Binding site prediction. To predict the SP binding site, we ran the *D. melanogaster* SPR and cleaved SP sequences through ColabFold (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb). The cleaved form of SP was missing both the signal peptide and the first 7 residues of the mature peptide. Previous work has shown that SP is cleaved from the surface of sperm at a putative R_7K_8 trypsin cleavage site, with the remainder of the peptide left bound to the surface of the sperm (Peng *et al.*, 2005a). Additional work has shown that both the full length (SP_{1-36}) and N-terminally truncated form of SP (SP_{8-36}) are equally efficient in eliciting a receptor-mediated response in *in vitro* assays (Poels *et al.*, 2010). However, we opted to use the cleaved form to match *in vivo* conditions. The .pdb file of the top ranked model was then loaded in ChimeraX (v1.5)(Pettersen *et al.*, 2021) and the 'interface residues' and 'hbond' commands used with default settings to identify putative contact and hydrogen bond forming residues in SPR, respectively. To compare ipTM scores between cleaved SP and other *D. melanogaster* Class A GPCRs, we extracted protein sequences from Flybase for each gene listed in the 'Class A GPCR neuropeptide and protein hormone receptor' gene group (FBgg0000195) and ran them through ColabFold (Mirdita *et al.*, 2017, 2019, 2022; Mitchell *et al.*, 2019) using default settings, as for SPR.

Data availability
All code, extracted coding and protein sequences, expression data, and protein models are available at https://osf.io/tzu6v/.

Acknowledgements
We thank Clive Wilson and Mark Wainwright for helpful feedback on the microcarrier part of this work; Julianne Peláez and Noah Whiteman for providing us with *Scaptomyza* stocks; and Barbara Parker, Kathy Darragh, Pavitra Muralidhar, and Carl Veller for valuable discussion. This work was supported by a long-term fellowship from the Human Frontier Science Program Organization (LT000123/2020-L to B.R.H.), a UC Davis College of Biological Sciences Dean's Circle Summer Undergraduate Research Program Fellowship to A. A-H., and National Institutes of Health F32 GM135998 to B.Y.K., R37-HD038921 to Mariana Wolfner, and R35 GM122592 to AK.

Contributions
B.R.H.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft preparation, Writing – review and editing, Visualization, Funding acquisition.

A.A-H.: Methodology, Validation, Investigation.

B.Y.K.: Resources, Formal analysis, Data curation

J.A.C.: Methodology

A.T.: Software

A.K.: Methodology, Writing – review and editing, Funding acquisition, Resources, Project administration, Project supervision.

Figure legends:

Figure 1. *Sex Peptide* family genes predate the *Drosophilini* and have been repeatedly lost outside of the *Sophophora-Lordiphosa* radiation. This figure focuses on members of the *Drosophilini* tribe that are outside of the *Sophophora* subgenus to which *D. melanogaster* belongs. A selection of *Sophophora* species, shaded in the tree in blue, are included for comparison. Also included are four non-*Drosophila* dipterans: *Aedes aegypti,* a representative of the Nematocera suborder, and three members of the *Drosophilinae* subfamily that do not belong to the *Drosophila* genus or *Drosophilini* tribe. These are *Chymomyza costata, Scaptodrosophila lebanonensis,* and *S. latifasciaeformis,* which all belong to the *Colocasiomyini* tribe. The number of *SP* genes detected in a representative of each species' genome is given at the tree tips. Losses are marked with a pink bar. For each species, the structures of two syntenic gene neighbourhoods are plotted. The first, Muller D1, is the canonical neighbourhood in which *SP* genes are detected outside of the *Sophophora.* The second, Muller D2, is the canonical position in

the *Sophophora.* Positions of each gene are given relative to the first gene in the neighbourhood (*FoxK* or *capricious*). Absence of a flanking neighbourhood gene (*e.g., FoxK* in *D. wassermani*) doesn't necessarily mean the gene has been lost – it more likely means that a contig breakpoint fell within the neighbourhood. In such cases, the next gene in the sequence was aligned to the position of its ortholog in the closest relative with a complete neighbourhood sequence. Note that *SP* in *Hirtodrosophila trivittata* and, independently, *D. repletoides,* has translocated out of the Muller D1 neighbourhood. While our phylogeny based on 250 orthologs placed *C. costata* within the *Scaptodrosophila* lineage, a published 17-locus phylogeny with denser taxon sampling in this part of the tree has *C. costata* branching earlier than the *Scaptodrosophila* (Finet *et al.*, 2021). Thus, the absence of *SP* in this species may represent ancestral absence rather than secondary loss.

Figure 2. *Sex Peptide* copy number is markedly more variable in the *Sophophora-Lordiphosa* radiation than in other branches of the phylogeny. This figure focuses on the *Sophophora-Lordiphosa* radiation to which *D. melanogaster* belongs. Four non-*Sophophora-Lordiphosa* drosophilids, shaded in purple, are included as outgroups. The structure of the Muller D1 and Muller D2 neighbourhoods is plotted as in Figure 1. Missing flanking genes are again likely indicative of contig breakpoints falling within the neighbourhood. The exceptions are the *Lordiphosa* species, where substantially elevated intergenic distances meant that the whole neighbourhood would not fit within the limits of the plot. Unresolved SP genes, shown in red, indicate genes that passed the reciprocal blast criteria to be designated as an *SP* gene and fell within one of the conserved *SP*-containing gene neighbourhoods but where a SP-like amino acid sequence couldn't be resolved (due, for example, to a premature stop codon, as in the case of *D. rhopaloa*). Note that all members of the *obscura* group have an inversion that flips the relative positions of *FoxK* and *mRpL2* in the Muller D1 neighbourhood. In a number of cases, some or all copies of *SP* were found to have translocated outside of the Muller D1 and Muller D2 neighbourhoods (an *obscura* group lineage, the *melanogaster* subgroup, *D. kanapiae, D. takahashii*, and *D.eugracilis*; summarised in Figure 4). In the shorter read *montium* subgroup assemblies, short contigs meant that in some species we couldn't identify the neighbourhood in which *SP* was located. This was the case for some *SP* genes in *D. cf. chauvacae, D. burlai, D. leontia, D. nikananu, D. pectinifera, D. punjabiensis,* and *D.watanabe*i. The *SP* genes in *D. auraria* and *D. trapezifons* could be mapped to the Muller D2 neighbourhood based on flanking sequence around the *SP* gene, but the *SP*-containing contigs were too small to include any of the neighbourhood genes.

Figure 3. Repeated duplication of *Sex Peptide* genes in the *ananassae* subgroup. A phylogeny of the *ananassae* subgroup species used in this study, with *D. melanogaster* as an outgroup. The number of *SP* genes identified in each species is given at the tip of each branch. 'Unresolved' SP sequences, shown in red, are those which passed the reciprocal blast tests and fell within the syntenic Muller D2 gene neighbourhood but for which we could not resolve an SP-like protein sequence (*e.g.,* due to a premature stop codon). The structure of that neighbourhood is plotted on the right-hand side of the figure.

Note that one of *D. melanogaster's SP* copies, the copy corresponding to *Dup99b,* falls outside of the Muller D2 neighbourhood.

Figure 4. An overview of duplications, losses, and translocations of *Sex Peptide*. This schematic summarises several translocations, duplications, and losses that have occurred within the *Drosophilidae*. Muller elements B to E are depicted for each species. We first detect *Sex Peptide* (*SP*), denoted by a thick black line, in the *Scaptodrosophila,* a *Drosophilidae* genus that forms the outgroup to the *Drosophilini* tribe. In these species, *SP* falls within a gene neighbourhood ('Muller D1') that maps to Muller element D in *D. melanogaster* and contains orthologs of *FoxK, mRpL2,* and *NaPi-III.* The *Drosophilini* then branches into two clades. The first contains the *Sophophora-Lordiphosa* radiation. In the most recent common ancestor (MRCA) of this radiation, *SP* has been lost from the ancestral gene neighbourhood and has translocated to a new position that maps to the same Muller element ('Muller D2'). This Muller D2 neighbourhood contains orthologs of *capricious, CG14111,* and *CG17687.* In the other *Drosophilini* lineage, *SP* is retained in the original genomic location, with the exceptions of *Hirtodrosophila trivittata* and *D. repletoides*, and several independent losses, which include all 20 of the Hawaiian *Drosophila* species we looked at and all 14 of the species we looked at that fell within a monophyletic radiation covering the *annulimana*, *bromeliae*, *nannoptera* and *repleta* ('ANBR') groups. Within the *Sophophora-Lordiphosa* radiation, *SP* has repeatedly duplicated and/or translocated to new locations. In the *obscura* group, *SP* is retained in the Muller D2 position that's ancestral to the *Sophophora-Lordiphosa* radiation in all species except those within the *pseudoobscura* and *affinis* subgroups. In these subgroups, *SP* has been lost from Muller D2 and translocated to a neighbourhood that maps to *D. melanogaster* Muller element B. Within the *melanogaster* group, SP has repeatedly and independently translocated and duplicated. Members of the *melanogaster* subgroup retain *SP* in Muller D2, but show a translocated copy (*Dup99b*) on Muller E. In several species, including *D. erecta* and *D. orena,* the Muller E copy has been lost. *D. eugracilis,* the outgroup to the *melanogaster* subgroup, shows a Muller D2 copy but its *Dup99b* copy maps to a non-canonical neighbourhood on Muller E. *D. suzukii* and its close relatives (*D. biarmipes* and *D. subpulchrella*) have 1 to 2 copies of *SP* in Muller D2 with an additional copy just the other side of the D2 neighbourhood; *D. takahashii* has a single copy in D2 with a translocation to and subsequent duplication in Muller B. A protein tree supports a *Dup99b* identity for the copies that fall outside of the canonical Muller D2 location in *D. eugracilis, D. suzukii, D.biarmipes, D. subpulchrella,* and *D. takahashii* (Supplementary figure 2). *D. ficusphila* has 3 copies in D2 with an additional copy in a different neighbourhood on Muller element D. Where multiple SP copies are present within a small area, we have thinned the bar signifying an SP gene. Positions of SP genes on each Muller element are approximate and not to scale.

Figure 5. Male reproductive tract biased expression is a conserved feature of *SP* genes. We obtained tissue-specific RNA-seq datasets from 10 different species hosted on NCBI. We extracted the maximum exonic read coverage value for each *SP* locus and divided it by the equivalent value extracted from the housekeeping gene *RpL32* to normalise expression across samples. Each plot depicts data from a different *SP* homolog. Genes are named by species, the chromosome arm and neighbourhood they

map to, and, in the case of tandemly arranged copies in the Muller D2 neighbourhood, in their order from *capricious.* Note that there is no D.ananassae_MullerD2b as this is the 'unresolved' copy that can be seen in Figure 3, for which we failed to find expression in any examined tissue. Ab (-D, -R) = Abdomen, digestive and reproductive tracts removed; R = Reproductive tract; G = Gonad (ovaries/testes); R (-G) = Reproductive tract without gonad; Whole = Whole flies; AGs = Accessory glands; Ab = Abdomen; Ab (-G) = Abdomen without gonad; EB = Ejaculatory bulb. The raw data, including dataset accessions, is available in https://osf.io/ydpfz.

Figure 6. Sex Peptide has undergone domain-specific evolutionary change. (A-C) Consensus sequences based on MAFFT alignment of the resolvable amino acid sequences of (A) all 233 SP genes, (B) all 57 from the non-*Sophophora-Lordiphosa* species, and (C) all 176 from the *Sophophora-Lordiphosa* species. In each panel, the top plot gives the proportion of sequences with the consensus amino acid in the same position, while the bottom plot gives the proportion of sequences in which each position is occupied in the alignment. A red line indicates the position of a predicted signal peptide cleavage site determined using SignalP 6.0. The probability of a cleavage site in the given position was 0.980 in A, 0.973 in B, and 0.982 in C. (D) The distribution of predicted protein lengths for each resolved SP sequence plotted separately for *Sophophora-Lordiphosa* and non-*Sophophora-Lordiphosa* sequences. The main panel shows the protein lengths after removing the signal peptide at the predicted site; the inset shows the full lengths prior to cleavage of the predicted peptide. Four sequences, two from each of the *montium* subgroup species *D. punjabiensis* and *D. watanabei,* did not contain a predicted signal peptide and are therefore omitted from the post-cleavage plot. Dashed lines indicate median lengths for the *Sophophora-Lordiphosa* and non-*Sophophora-Lordiphosa* species, separately. The smoothed curve shows a kernel density estimate, a smoothed version of the histogram, derived using the geom_density function in ggplot2. (E-G) PCA plots based on BLOSUM62 substitution scores from the MAFFT-aligned resolved SP sequences. (E) PC1 vs PC2. (F) PC1 vs PC3. (G) PC2 vs PC3. The percentage values in the axis titles reflect the proportion of variance explained by a given PC. Points are coloured based on whether they correspond to *Sophophora-Lordiphosa,* non-*Sophophora-Lordiphosa,* or *ananassae* subgroup species.

Figure 7. *Sex Peptide* is neither necessary nor sufficient for microcarriers. (A) A phylogeny of all species included in this figure. Branches coloured pink indicate species shown in B-T that demonstrate canonical (*i.e., D. melanogaster*-like) staining with LipidTox, a neutral lipid-specific dye used to selectively stain microcarriers (Wainwright *et al.*, 2021). (B-T) Drosophilid accessory glands stained with LipidTox and the nuclear stain DAPI (blue). The circled number in the bottom right-hand corner of each panel indicates the number of *Sex Peptide* copies we detect in each species. (G) The arrow is highlighting a spiral/doughnut shaped microcarrier, a shape which is rare in comparison to the more common thread-like conformation in this species. (J) Prox. refers to the proximal region of the gland, *i.e.* the region that connects to the ejaculatory duct. (P) Arrows point to the ambiguous, sparse, and weakly stained material we observed in stained *D. ficusphila* glands. (S) An arrow points to a cut in the glandular epithelium, which was made to enhance dye penetration. In each case, glands were co-stained in the same

well with those from *D. melanogaster* to act as a control for assessing whether the staining was successful.

Figure 8. SPR residues showing evidence of episodic positive selection are enriched in the ligand-facing domains. (A) A consensus sequence based on MAFFT alignment of the resolvable amino acid sequences of all *SPR* coding sequences. The top plot gives the proportion of sequences with the consensus amino acid in the same position, while the bottom plot gives the proportion of sequences in which each position is occupied in the alignment. Each residue is coloured based on the functional domain to which it belongs based on the UniProt annotations. The red line indicates the start of the conserved region we use in the molecular evolution analyses. Beneath the plot, we show the corresponding exon that encodes each consensus residue. EC = extracellular, IC = intracellular (B) The AlphaFold prediction of the structure of *D. melanogaster* SPR as downloaded from UniProt (AF-Q8SWR3-F1) and coloured by the domain each residue belongs to based on the positions listed in the UniProt 'Features' table. (C) A boxplot showing the proportion of residues at each position that matched the consensus residue (*i.e.,* the degree of conservation at each position). Residues are plotted separately according to their domain class of origin. (D) PCA plot based on BLOSUM62 substitution scores from the MAFFT-aligned SPR protein sequences. The percentage values in the axis titles reflect the proportion of variance explained by a given PC. Points are coloured based on whether they correspond to *Sophophora-Lordiphosa,* non-*Sophophora-Lordiphosa,* or *ananassae* subgroup species. (E) As (D) but after removing the non-conserved region of the N-terminus (*i.e.* the region preceding the red line in (A)). (F) The same prediction shown in (B) but with residues coloured by the proportion of conserved residues per position (see also Supplementary movie 1). High values indicate high conservation. (G) The ColabFold top-ranked prediction of the interactions between SP (shown in green) and SPR (residues coloured by domain). (H) As in (G) but with predicted contact residues coloured purple and predicted hydrogen bonds between SP and SPR residues shown with red-dotted lines (See also Supplementary movies 2 and 3). (I) A bar chart showing the proportion of sequences sharing the consensus residue for each predicted contact residue. Bars are coloured by whether the residue is also predicted to form a hydrogen bond with SP. Asterisks denote predicted contact residues for which we detected evidence of episodic positive selection using *MEME.* Alongside the plot, coloured bars, using the same colour scale as in (A-C), denote the functional domain the residue falls within. The two adjacent blue bars denote separate, consecutive transmembrane domains. Alongside are the corresponding amino acid residues in each position for each of the consensus ('Cons.'), *D. melanogaster* ('Dmel.'), *D. bipectinata* ('Dbip.'), *D. pseudoobscura* ('Dpse.'), *D. virilis* ('Dvir.'), and *A. aegypti* ('Aaeg.') sequences. Residues that depart from the mode among these plotted sequences are coloured. (J) As (G) but colouring only the 10 residues in SPR for which we detected evidence of positive selection using *MEME*. Asterisks denote predicted contact residues that show evidence of positive selection.

Supplementary figure 1. Taxonomic nomenclature used in this study. This figure shows the phylogeny and *SP* copy numbers (given in coloured circles at each tip) from Figure 2 with branches coloured according to higher order groupings we use in this study.

The classification of the "Oriental" lineage follows the convention of Kopp and True (2002). The positioning of *D. helvetica* within the *affinis* subgroup follows Gao *et al.* (2007).

Supplementary figure 2. A protein tree of Sex Peptide sequences. The protein tree was constructed using the Maximum Likelihood method on a MUSCLE alignment of 233 extracted SP amino acid sequences. The sequences include the signal peptide. Several clades of interest that are discussed in the text are coloured.

Supplementary figure 3. *Sex Peptide* translocated at the base of the clade that includes the *affinis* and *pseudoobscura* subgroups. The number of *SP* genes identified in each species is given at the tip of each branch. *SP* genes are plotted in relation to one of two gene neighbourhoods, one of which mapped to Muller element B in *D. melanogaster* and the other of which mapped to Muller element D. This 'Muller D2' neighbourhood represents the canonical location of *SP* within the *Sophophora-Lordiphosa* radiation. Note that the absence of neighbourhood genes in a species doesn't necessarily indicate their absence from that species' genome. It may reflect the presence of a contig breakpoint within the neighbourhood that prevented us from resolving the neighbourhood's structure. Note also the apparently independent duplications of *SP* in *D. lowei* and *D. subobscura.* In the latter case, the two bars are hard to visually separate due to their close proximity. A pink line indicates the branch where the translocation occurred.

Supplementary figure 4. The localisation of a *Sex Peptide* gene to the *Dup99b* position is restricted to the *melanogaster* subgroup. (A) A phylogeny of the "Oriental" lineage species used in this study. Species in the *melanogaster* subgroup are highlighted with a blue square. The number of *SP* genes identified in each species is given at the tip of each branch. 'Unresolved' SP sequences, shown in red, are those which passed the reciprocal blast tests and fell within the syntenic Muller D2 gene neighbourhood but for which we could not resolve an SP-like protein structure (*e.g.* due to a premature stop codon). The structure of the Muller E and Muller D2 gene neighbourhoods are plotted on the right-hand side of the figure. Several sequences are not plotted as they fall outside of these neighbourhoods: one from *D. ficusphila*, which falls in a separate neighbourhood that maps to Muller element D; one from *D. eugracilis,* which maps to a region on Muller element E near to, but distinct from, the canonical *Dup99b* position; and two from *D. takahashii,* which map to Muller element B. (B) In genomes for two strains of *D. teissieri* (CT02 and 273.3), we were unable to resolve an SP-like sequence for the copy in the *Dup99b* Muller element E neighbourhood. Alongside *SP* genes in the canonical Muller D2, both strains also encoded an additional copy in a different gene neighbourhood that mapped to Muller element E in *D. melanogaster* and contained orthologs of *poly, Lip3,* and *mthl12*. This neighbourhood was ~16.3Mb away from *Dup99b* on *D. melanogaster* Muller element E. However, we could only resolve an SP-like protein sequence for this copy in one of the two strains (273.3). CT02 contained a 37bp deletion covering the start codon, as shown in this alignment. E/I and I/E represent the exon/intron boundaries.

Supplementary figure 5. Pseudogenisation of a *Sex Peptide* paralog in *D. rhopaloa*. As in its two closest relatives, *D. carrolli* and *D. kurseongensis,* we detected two *SP* copies in the Muller D2 neighbourhood of *D. rhopaloa.* (A) In this screenshot from the NCBI genome data browser we can see the position of the two *SP* genes (depicted by red arrows) in the genome we searched. One copy, which we refer to as 'MullerD2b', corresponds to the gene LOC108051015. RNA-seq tracks from male and female samples show clear male-specific expression spanning the two exons that correspond to this gene. The other *SP* gene we identify, the copy closest to *capricious* and which we refer to as 'MullerD2a', has neither evidence of expression nor an associated annotated feature. Interestingly, LOC108051016, which separates the two *SP* hits, showed male-specific expression and encoded a protein with no clear homologous sequence in the *D. melanogaster* reference protein list, suggesting it may encode a lineage- and sex-specific short peptide. (B) Along with the absence of expression, we couldn't resolve a SP-like protein sequence for *D. rhopaloa*_MullerD2a. This was due to a G to A point mutation that introduced a premature stop codon into the SP reading frame, the position of which is indicated by a red arrow. E/I and I/E represent the exon/intron boundaries.

Supplementary figure 6. Male-biased expression is a conserved feature of *SP* genes. We obtained sex-labelled RNAseq datasets from whole body samples of 19 different species hosted on NCBI. We extracted the maximum exonic read coverage value for each *SP* locus and divided it by the equivalent value extracted from the housekeeping gene *RpL32* to normalise expression across samples. Each plot depicts data from a different *SP* gene, with rows corresponding to copies within the same species. Genes are named by species, the chromosome arm and neighbourhood they map to, and, in the case of tandemly arranged copies in the Muller D2 neighbourhood, in their order from *capricious.* The one *SP* copy that does not show clear male-biased expression mapped to the Muller element E *Dup99b* location in *D. simulans.* However, this sample (SAMN02713493) was almost certainly contaminated with males or its sex mislabelled as, unlike the other female sample, it showed appreciable expression of the male-specific long non-coding RNAs *roX1* and *roX2*, which function in X chromosome dosage compensation (Supplementary figure 7). The raw data, including dataset accessions, is available in Supplementary data 1.

Supplementary figure 7. Expression of *SP* in a whole-body *D. simulans* female sample is likely due to contamination. In Supplementary figure 6, we showed that a single RNAseq sample from a female of one of the 19 species examined showed appreciable *SP* expression. This sample, however, was almost certainly mislabelled or contaminated with males. To establish this, we used BLAST to query the *D. melanogaster* transcript sequences of *roX1* (FBgn0019661) and *roX2* (FBgn0019660) against the *D. simulans* reference genome (Prin_D.sim_3.1) and then overlaid over the returned hit the RNAseq expression tracks from the whole-body male and female samples plotted for *D. simulans* in Supplementary figure 6. For both genes, we detect clear, male-like expression in the *SP*-expressing female sample SAMN02713493, but not the other female sample. Among adults, *roX1* and *roX2* are exclusively expressed in males (Meller *et al.*, 1997; Meller, 2003; Lim & Kelley, 2012). M = male, F = female. F? = the female-labelled SAMN02713493 sample.

Supplementary figure 8. The protein sequence encoded by exon 1 shows the greatest between-species variation among the exons of *SPR*. The tBLASTn scores plotted from the top hit returned when the amino acid sequence encoded by each of *D. melanogaster's* four *SPR* exons was queried against each species in our dataset. *A. aegypti* is the most phylogenetically distant species in the dataset and its position is labelled in each plot. The circled region in the exon 1 plot contains all species in our dataset from the "Oriental" lineage. Note the sharp, exon-specific drop in score that occurs outside of this lineage. All other labelled points correspond to species in which we couldn't resolve SPR sequences. Note how some species have low scores recorded for each exon (*e.g. S. montana* and *S. flava*), indicative of loss of the whole gene, others show a drop in a subset of exons, such as *Z. taronus*, which recorded a drop in score only in exon 4 (due to a premature stop codon).

Supplementary figure 9. Summary statistics for the ColabFold-predicted interactions between SPR and SP. (A-C) Summary plot outputs from the ColabFold prediction of SP-SPR interactions. (A) The multiple sequence alignment coverage per position. The vertical black line on the right-hand side of the image marks the separation between SP and SPR. (B) The AlphaFold confidence measure lDDT (local Distance Difference Test), which provides a measure of the model confidence at each position. (C) The predicted alignment error per position for SPR ('A') and SP ('B') for each of the 5 ranked structures (rank 1 is the highest, *i.e.,* best performing). (D) The ipTM scores, which describes the confidence in the interface structure, plotted for each of the 5 ranked structures predicted between SP and each *D. melanogaster* Class A G-protein coupled receptor. (E) A boxplot showing the proportion of sequences at each position in an alignment of 193 N-terminus trimmed SPR sequences that share the consensus residue. Residues are plotted separately based on whether they are a predicted SP-contacting residue, a residue that flanks (either side) a predicted contact residue, or neither. Points are coloured based on the SPR domain they belong to. EC = extracellular; IC = intracellular.

Supplementary figure 10. Unannotated *Sex Peptide* genes in *D. biarmipes* and *D. ficusphila.* Screenshot from the NCBI genome data browser showing the positions of *SP* hits (red arrows) in genomes of *D. biarmipes* (A) and *D. ficusphila* (B). In each case, the *SP* genes we identify are accompanied by male-specific expression in the RNA-seq exon coverage tracks, but only one of each species' two shown hits has an accompanying feature annotation in the 'Genes' window. The 'BLAST' window shows the position of the unannotated SP gene in each species

Supplementary movie 1. SPR coloured by sequence conservation at each residue (high values = high conservation).

Supplementary movie 2. The interface between SPR (beige) and SP (green). Pink residues in SPR indicate predicted contact residues. Red dotted lines indicate predicted H-bond forming residues between SP and SPR.

Supplementary movie 3. The interface between SPR and SP. Interfacing residues are coloured.

Supplementary movie 4. The interface between SPR (beige) and SP (green). Pink residues in SPR indicate the 10 residues identified using MEME as showing evidence of episodic positive selection.

References:

ABU-RAYA, B., MICHALSKI, C., SADARANGANI, M. & LAVOIE, P.M. (2020) Maternal Immunological Adaptation During Normal Pregnancy. *Frontiers in Immunology* **11**, 2627. Frontiers Media S.A.

AHMED-BRAIMAH, Y.H., WOLFNER, M.F. & CLARK, A.G. (2021) Differences in Postmating Transcriptional Responses between Conspecific and Heterospecific Matings in Drosophila. *Molecular Biology and Evolution* **38**, 986–999. Oxford Academic.

ARNQVIST, G. & ROWE, L. (2005) *Sexual Conflict*. Princeton University Press.

BANKEVICH, A., NURK, S., ANTIPOV, D., GUREVICH, A.A., DVORKIN, M., KULIKOV, A.S., LESIN, V.M., NIKOLENKO, S.I., PHAM, S., PRJIBELSKI, A.D., PYSHKIN, A. V., SIROTKIN, A. V., VYAHHI, N., TESLER, G., ALEKSEYEV, M.A., ET AL. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**, 455–477.

BELKORCHIA, A., GASC, C., POLONAIS, V., PARISOT, N., GALLOIS, N., RIBIÈRE, C., LERAT, E., GASPIN, C., POMBERT, J.F., PEYRET, P. & PEYRETAILLADE, E. (2015) The prediction and validation of small CDSs expand the gene repertoire of the smallest known eukaryotic genomes. *PLoS ONE* **10**. Public Library of Science.

BERLAND, M.A., ULLOA-LEAL, C., BARRÍA, M., WRIGHT, H., DISSEN, G.A., SILVA, M.E., OJEDA, S.R. & RATTO, M.H. (2016) Seminal Plasma Induces Ovulation in Llamas in the Absence of a Copulatory Stimulus: Role of Nerve Growth Factor as an Ovulation-Inducing Factor. *Endocrinology* **157**, 3224–3232.

BRACEWELL, R., CHATLA, K., NALLEY, M.J. & BACHTROG, D. (2019) Dynamic turnover of centromeres drives karyotype evolution in drosophila. *eLife* **8**.

BRACEWELL, R., TRAN, A., CHATLA, K. & BACHTROG, D. (2020) Chromosome-level assembly of drosophila bifasciata reveals important karyotypic transition of the X chromosome. *G3: Genes, Genomes, Genetics* **10**, 891–897.

BRONSKI, M.J., MARTINEZ, C.C., WELD, H.A. & EISEN, M.B. (2020) Whole genome sequences of 23 species from the drosophila montium species group (Diptera: Drosophilidae): A resource for testing evolutionary hypotheses. *G3: Genes, Genomes, Genetics* **10**, 1443–1455.

BUSSIÉGRE, L.F., HUNT, J., JENNIONS, M.D. & BROOKS, R. (2006) Sexual Conflict and Cryptic Female Choice in the Black Field Cricket, Teleogryllus Commodus. *Evolution* **60**, 792–800.

CHAKRABORTY, M., CHANG, C.H., KHOST, D.E., VEDANAYAGAM, J., ADRION, J.R., LIAO, Y., MONTOOTH, K.L., MEIKLEJOHN, C.D., LARRACUENTE, A.M. & EMERSON, J.J. (2021) Evolution of genome structure in the Drosophila simulans species complex. *Genome Research* **31**, 380–396.

CHAPMAN, T. (2018) Sexual Conflict: Mechanisms and Emerging Themes in Resistance Biology. *The American Naturalist* **192**, 217–229.

CHAPMAN, T., LIDDLE, L.F., KALB, J.M., WOLFNER, M.F. & PARTRIDGE, L. (1995) Cost of mating in Drosophila melanogaster females is mediated by male accessory gland products. *Nature* **373**, 241–244.

CHEN, P.S. & DIEM, C. (1961) A sex-specific ninhydrin-positive substance found in the paragonia of adult males of Drosophila melanogaster. *Journal of Insect Physiology* **7**, 289–298.

CHEN, P.S., STUMM-ZOLLINGER, E., AIGAKI, T., BALMER, J., BIENZ, M. & BÖHLEN, P. (1988) A male accessory gland peptide that regulates reproductive behavior of female D. melanogaster. *Cell* **54**, 291–298.

CHEREZOV, V., ROSENBAUM, D.M., HANSON, M.A., RASMUSSEN, S.G.F., FOON, S.T., KOBILKA, T.S., CHOI, H.J., KUHN, P., WEIS, W.I., KOBILKA, B.K. & STEVENS, R.C. (2007) High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor. *Science* **318**, 1258–1265. American Association for the Advancement of Science .

CIRERA, S. & AGUADÉ, M. (1998) The sex-peptide gene (Acp70A) is duplicated in Drosophila subobscura. *Gene* **210**, 247–254. Gene.

CIVETTA, A. & CLARK, A.G. (2000) Correlated effects of sperm competition and postmating female mortality. *Proceedings of the National Academy of Sciences* **97**, 13162–13165.

CONNER, W.R., DELANEY, E.K., BRONSKI, M.J., GINSBERG, P.S., WHEELER, T.B., RICHARDSON, K.M., PECKENPAUGH, B., KIM, K.J., WATADA, M., HOFFMANN, A.A., EISEN, M.B., KOPP, A., COOPER, B.S. & TURELLI, M. (2021) A phylogeny for the Drosophila montium species group: A model clade for comparative analyses. *Molecular Phylogenetics and Evolution* **158**, 107061.

CONZELMANN, M., WILLIAMS, E.A., TUNARU, S., RANDEL, N., SHAHIDI, R., ASADULINA, A., BERGER, J., OFFERMANNS, S. & JEKELY, G. (2013) Conserved MIP receptor-ligand pair regulates Platynereis larval settlement. *Proceedings of the National Academy of Sciences* **110**, 8224–8229.

DELPORT, W., POON, A.F.Y., FROST, S.D.W. & KOSAKOVSKY POND, S.L. (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**, 2455–2457. Oxford Academic.

DING, Z., HAUSSMANN, I., OTTIGER, M. & KUBLI, E. (2003) Sex-peptides bind to two molecularly different targets inDrosophila melanogaster females. *Journal of Neurobiology* **55**, 372–384.

DOMANITSKAYA, E. V., LIU, H., CHEN, S. & KUBLI, E. (2007) The hydroxyproline motif of male sex peptide elicits the innate immune response in Drosophila females. *FEBS Journal* **274**, 5659–5668.

DUVALL, L.B., BASRUR, N.S., MOLINA, H., MCMENIMAN, C.J. & VOSSHALL, L.B. (2017) A Peptide Signaling System that Rapidly Enforces Paternity in the Aedes aegypti Mosquito. *Current Biology* **27**, 3734-3742.e5.

ELLISON, C. & BACHTROG, D. (2019a) Contingency in the convergent evolution of a regulatory network: Dosage compensation in drosophila. *PLoS Biology* **17**, e3000094.

ELLISON, C. & BACHTROG, D. (2019b) Recurrent gene co-amplification on Drosophila X

and Y chromosomes. *PLoS Genetics* **15**, e1008251.
FAN, Y., RAFAELI, A., MOSHITZKY, P., KUBLI, E., CHOFFAT, Y. & APPLEBAUM, S.W. (2000) Common functional elements of Drosophila melanogaster seminal peptides involved in reproduction of Drosophila melanogaster and Helicoverpa armigera females. *Insect Biochemistry and Molecular Biology* **30**, 805–812.
FINDLAY, G.D., SITNIK, J.L., WANG, W., AQUADRO, C.F., CLARK, N.L. & WOLFNER, M.F. (2014) Evolutionary rate covariation identifies new members of a protein network required for Drosophila melanogaster female post-mating responses. *PLoS genetics* **10**, e1004108.
FINET, C., KASSNER, V.A., CARVALHO, A.B., CHUNG, H., DAY, J.P., DAY, S., DELANEY, E.K., DE RÉ, F.C., DUFOUR, H.D., DUPIM, E., IZUMITANI, H.F., GAUTÉRIO, T.B., JUSTEN, J., KATOH, T., KOPP, A., ET AL. (2021) DrosoPhyla: Resources for Drosophilid Phylogeny and Systematics. *Genome Biology and Evolution* **13**. Oxford Academic.
FOX, A.S. (1956) Chromatographic Differences between Males and Females in Drosophila melanogaster and Role of X and Y Chromosomes. *Physiological Zoology* **29**, 288–298. University of Chicago Press .
FRICKE, C., BRETMAN, A. & CHAPMAN, T. (2010) Female nutritional status determines the magnitude and sign of responses to a male ejaculate signal in Drosophila melanogaster. *Journal of Evolutionary Biology* **23**, 157–165.
GAO, J.J., WATABE, H.A., AOTSUKA, T., PANG, J.F. & ZHANG, Y.P. (2007) Molecular phylogeny of the Drosophila obscura species group, with emphasis on the Old World species. *BMC Evolutionary Biology* **7**, 1–12. BioMed Central.
GLIGOROV, D., SITNIK, J.L., MAEDA, R.K., WOLFNER, M.F. & KARCH, F. (2013) A Novel Function for the Hox Gene Abd-B in the Male Accessory Gland Regulates the Long-Term Female Post-Mating Response in Drosophila. *PLoS Genetics* **9**, e1003395.
GRAMATES, L.S., AGAPITE, J., ATTRILL, H., CALVI, B.R., CROSBY, M.A., DOS SANTOS, G., GOODMAN, J.L., GOUTTE-GATTAT, D., JENKINS, V.K., KAUFMAN, T., LARKIN, A., MATTHEWS, B.B., MILLBURN, G. & STRELETS, V.B. (2022) FlyBase: a guided tour of highlighted features. *Genetics* **220**.
HÄSEMEYER, M., YAPICI, N., HEBERLEIN, U. & DICKSON, B.J. (2009) Sensory Neurons in the Drosophila Genital Tract Regulate Female Reproductive Behavior. *Neuron* **61**, 511–518.
HAUSSMANN, I.U., HEMANI, Y., WIJESEKERA, T., DAUWALDER, B. & SOLLER, M. (2013) Multiple pathways mediate the sexpeptide- regulated switch in female Drosophila reproductive behaviours. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20131938.
HILL, T., KOSEVA, B.S. & UNCKLESS, R.L. (2019) The genome of drosophila innubila reveals lineage-specific patterns of selection in immune genes. *Molecular Biology and Evolution* **36**, 1405–1417.
HOLLAND, B. & RICE, W.R. (1998) Perspective: Chase-Away Sexual Selection: Antagonistic Seduction Versus Resistance. *Evolution* **52**, 1.
HOLLIS, B., KOPPIK, M., WENSING, K.U., RUHMANN, H., GENZONI, E., ERKOSAR, B., KAWECKI, T.J., FRICKE, C. & KELLER, L. (2019) Sexual conflict drives male manipulation of female postmating responses in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*

116, 8437–8444.
HOPKINS, B.R. & PERRY, J.C. (2022) The evolution of sex peptide: sexual conflict, cooperation, and coevolution. *Biological Reviews* **97**, 1426–1448. John Wiley & Sons, Ltd.
JA, W.W., CARVALHO, G.B., MADRIGAL, M., ROBERTS, R.W. & BENZER, S. (2009) The Drosophila G protein-coupled receptor, Methuselah, exhibits a promiscuous response to peptides. *Protein Science* **18**, 2203–2208. Wiley-Blackwell.
JOHNSTONE, R.A. & KELLER, L. (2000) How males can gain by harming their mates: Sexual conflict, seminal toxins, and the cost of mating. *American Naturalist* **156**, 368–377.
JONES, D.T., TAYLOR, W.R. & THORNTON, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences : CABIOS* **8**, 275–282. Comput Appl Biosci.
JONES, E.M., LUBOCK, N.B., VENKATAKRISHNAN, A.J., WANG, J., TSENG, A.M., PAGGI, J.M., LATORRACA, N.R., CANCILLA, D., SATYADI, M., DAVIS, J.E., BABU, M.M., DROR, R.O. & KOSURI, S. (2020) Structural and functional characterization of G protein-coupled receptors with deep mutational scanning. *eLife* **9**, 1–52. eLife Sciences Publications Ltd.
KATOH, K. & STANDLEY, D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772–780.
KIM, B.Y., WANG, J.R., MILLER, D.E., BARMINA, O., DELANEY, E., THOMPSON, A., COMEAULT, A.A., PEEDE, D., D'AGOSTINO, E.R.R., PELAEZ, J., AGUILAR, J.M., HAJI, D., MATSUNAGA, T., ARMSTRONG, E.E., ZYCH, M., ET AL. (2021) Highly contiguous assemblies of 101 drosophilid genomes. *eLife* **10**. eLife Sciences Publications Ltd.
KIM, Y.J., BARTALSKA, K., AUDSLEY, N., YAMANAKA, N., YAPICI, N., LEE, J.Y., KIM, Y.C., MARKOVIC, M., ISAAC, E., TANAKA, Y. & DICKSON, B.J. (2010) MIPs are ancestral ligands for the sex peptide receptor. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 6520–6525.
KINGAN, T.G., BODNAR, W.M., RAINA, A.K., SHABANOWITZ, J. & HUNT, D.F. (1995) The loss of female sex pheromone after mating in the corn earworm moth Helicoverpa zea: identification of a male pheromonostatic peptide. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 5082. National Academy of Sciences.
KOPP, A. & TRUE, J.R. (2002) Phylogeny of the oriental Drosophila melanogaster species group: A multilocus reconstruction. *Systematic Biology* **51**, 786–805. Syst Biol.
KOSAKOVSKY POND, S.L., MURRELL, B., FOURMENT, M., FROST, S.D.W., DELPORT, W. & SCHEFFLER, K. (2011) A Random Effects Branch-Site Model for Detecting Episodic Diversifying Selection. *Molecular Biology and Evolution* **28**, 3033–3043. Oxford Academic.
KUMAR, S., STECHER, G., LI, M., KNYAZ, C. & TAMURA, K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution* **35**, 1547–1549. Mol Biol Evol.
LEE, J.-H., LEE, N.-R., KIM, D.-H. & KIM, Y.-J. (2020) Molecular characterization of ligand selectivity of the sex peptide receptors of Drosophila melanogaster and Aedes

aegypti. *Insect Biochemistry and Molecular Biology* **127**, 103472.
LIAO, Y., ZHANG, X., CHAKRABORTY, M. & EMERSON, J.J. (2021) Topologically associating domains and their role in the evolution of genome structure and function in Drosophila. *Genome Research* **31**, 397–410.
LIM, C.K. & KELLEY, R.L. (2012) Autoregulation of the Drosophila Noncoding roX1 RNA Gene. *PLOS Genetics* **8**, e1002564. Public Library of Science.
MACKAY, T.F.C., RICHARDS, S., STONE, E.A., BARBADILLA, A., AYROLES, J.F., ZHU, D., CASILLAS, S., HAN, Y., MAGWIRE, M.M., CRIDLAND, J.M., RICHARDSON, M.F., ANHOLT, R.R.H., BARRÓN, M., BESS, C., BLANKENBURG, K.P., ET AL. (2012) The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–178. Nature Publishing Group.
MADEIRA, F., PEARCE, M., TIVEY, A.R.N., BASUTKAR, P., LEE, J., EDBALI, O., MADHUSOODANAN, N., KOLESNIKOV, A. & LOPEZ, R. (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research* **50**, W276–W279.
MAHAJAN, S., WEI, K.H.C., NALLEY, M.J., GIBILISCO, L. & BACHTROG, D. (2018) De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLoS Biology* **16**, e2006348.
MAI, D., NALLEY, M.J. & BACHTROG, D. (2020) Patterns of Genomic Differentiation in the Drosophila nasuta Species Complex. *Molecular Biology and Evolution* **37**, 208–220.
MAJANE, A.C., CRIDLAND, J.M. & BEGUN, D.J. (2022) Single-nucleus transcriptomes reveal evolutionary and functional properties of cell types in the Drosophila accessory gland. *Genetics* **220**. Oxford Academic.
MANNI, M., BERKELEY, M.R., SEPPEY, M., SIMÃO, F.A. & ZDOBNOV, E.M. (2021) BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution* **38**, 4647–4654.
MATTHEWS, B.J., DUDCHENKO, O., KINGAN, S.B., KOREN, S., ANTOSHECHKIN, I., CRAWFORD, J.E., GLASSFORD, W.J., HERRE, M., REDMOND, S.N., ROSE, N.H., WEEDALL, G.D., WU, Y., BATRA, S.S., BRITO-SIERRA, C.A., BUCKINGHAM, S.D., ET AL. (2018) Improved reference genome of Aedes aegypti informs arbovirus vector control. *Nature* **563**, 501–507.
MCGEARY, M.K. & FINDLAY, G.D. (2020) Molecular evolution of the sex peptide network in Drosophila. *Journal of Evolutionary Biology* **33**, 629–641.
MELLER, V.H. (2003) Initiation of dosage compensation in Drosophila embryos depends on expression of the roX RNAs. *Mechanisms of Development* **120**, 759–767. Elsevier.
MELLER, V.H., WU, K.H., ROMAN, G., KURODA, M.I. & DAVIS, R.L. (1997) roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. *Cell* **88**, 445–457. Elsevier B.V.
MINH, B.Q., SCHMIDT, H.A., CHERNOMOR, O., SCHREMPF, D., WOODHAMS, M.D., VON HAESELER, A., LANFEAR, R. & TEELING, E. (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534.
MIRDITA, M., VON DEN DRIESCH, L., GALIEZ, C., MARTIN, M.J., SODING, J. & STEINEGGER,

M. (2017) Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Research* **45**, D170–D176.

MIRDITA, M., SCHÜTZE, K., MORIWAKI, Y., HEO, L., OVCHINNIKOV, S. & STEINEGGER, M. (2022) ColabFold: making protein folding accessible to all. *Nature Methods* **19**, 679–682.

MIRDITA, M., STEINEGGER, M. & SÖDING, J. (2019) MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **35**, 2856–2858.

MISRA, S. & WOLFNER, M.F. (2020) Drosophila seminal sex peptide associates with rival as well as own sperm, providing SP function in polyandrous females. *eLife* **9**, 1–21.

MITCHELL, A.L., ALMEIDA, A., BERACOCHEA, M., BOLAND, M., BURGIN, J., COCHRANE, G., CRUSOE, M.R., KALE, V., POTTER, S.C., RICHARDSON, L.J., SAKHAROVA, E., SCHEREMETJEW, M., KOROBEYNIKOV, A., SHLEMOV, A., KUNYAVSKAYA, O., ET AL. (2019) MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Res.*

MOEHLE, K., FREUND, A., KUBLI, E. & ROBINSON, J.A. (2011) NMR studies of the solution conformation of the sex peptide from Drosophila melanogaster. *FEBS Letters* **585**, 1197–1202.

MOSHITZKY, P., FLEISCHMANN, I., CHAIMOV, N., SAUDAN, P., KLAUSER, S., KUBLI, E. & APPLEBAUM, S.W. (1996) Sex-peptide activates juvenile hormone biosynthesis in theDrosophila melanogaster corpus allatum. *Archives of Insect Biochemistry and Physiology* **32**, 363–374.

MURRELL, B., MOOLA, S., MABONA, A., WEIGHILL, T., SHEWARD, D., KOSAKOVSKY POND, S.L. & SCHEFFLER, K. (2013) FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Molecular Biology and Evolution* **30**, 1196–1205. Oxford Academic.

MURRELL, B., WEAVER, S., SMITH, M.D., WERTHEIM, J.O., MURRELL, S., AYLWARD, A., EREN, K., POLLNER, T., MARTIN, D.P., SMITH, D.M., SCHEFFLER, K. & KOSAKOVSKY POND, S.L. (2015) Gene-Wide Identification of Episodic Selection. *Molecular Biology and Evolution* **32**, 1365–1371. Oxford Academic.

MURRELL, B., WERTHEIM, J.O., MOOLA, S., WEIGHILL, T., SCHEFFLER, K. & KOSAKOVSKY POND, S.L. (2012) Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLOS Genetics* **8**, e1002764. Public Library of Science.

O'REILLY, F.J., GRAZIADEI, A., FORBRIG, C., BREMENKAMP, R., CHARLES, K., LENZ, S., ELFMANN, C., FISCHER, L., STÜLKE, J. & RAPPSILBER, J. (2023) Protein complexes in cells by AI-assisted structural proteomics. *Molecular Systems Biology*.

OH, Y., YOON, S.-E., ZHANG, Q., CHAE, H.-S., DAUBNEROVÁ, I., SHAFER, O.T., CHOE, J. & KIM, Y.-J. (2014) A Homeostatic Sleep-Stabilizing Pathway in Drosophila Composed of the Sex Peptide Receptor and Its Ligand, the Myoinhibitory Peptide. *PLoS Biology* **12**, e1001974.

PARADIS, E. & SCHLIEP, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in {R}. *Bioinformatics* **35**, 526–528.

PARIS, M., BOYER, R., JAENICHEN, R., WOLF, J., KARAGEORGI, M., GREEN, J., CAGNON, M., PARINELLO, H., ESTOUP, A., GAUTIER, M., GOMPEL, N. & PRUD'HOMME, B. (2020) Near-chromosome level genome assembly of the fruit pest Drosophila suzukii using long-read sequencing. *Scientific Reports* **10**, 11227.

PATTERSON, J.T. (1946) A New Type of Isolating Mechanism in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **32**, 202–208.

PELAEZ, J.N., GLOSS, A.D., GOLDMAN-HUERTAS, B., KIM, B., LAPOINT, R.T., PIMENTEL-SOLORIO, G., VERSTER, K.I., AGUILAR, J.M., DITTRICH, A.C.N., SINGHAL, M., SUZUKI, H.C., MATSUNAGA, T., ARMSTRONG, E.E., CHARBONEAU, J.L.M., GROEN, S.C., ET AL. (2023) Evolution of chemosensory and detoxification gene families across herbivorous Drosophilidae. *bioRxiv*. Cold Spring Harbor Laboratory.

PENG, J., CHEN, S., BÜSSER, S., LIU, H., HONEGGER, T. & KUBLI, E. (2005a) Gradual Release of Sperm Bound Sex-Peptide Controls Female Postmating Behavior in Drosophila. *Current Biology* **15**, 207–213.

PENG, J., ZIPPERLEN, P. & KUBLI, E. (2005b) Drosophila sex-peptide stimulates female innate immune system after mating via the toll and Imd pathways. *Current Biology* **15**, 1690–1694.

PETTERSEN, E.F., GODDARD, T.D., HUANG, C.C., MENG, E.C., COUCH, G.S., CROLL, T.I., MORRIS, J.H. & FERRIN, T.E. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70–82.

PILPEL, N., NEZER, I., APPLEBAUM, S.W. & HEIFETZ, Y. (2008) Mating-increases trypsin in female Drosophila hemolymph. *Insect Biochemistry and Molecular Biology* **38**, 320–330.

POELS, J., VAN LOY, T., VANDERSMISSEN, H.P., VAN HIEL, B., VAN SOEST, S., NACHMAN, R.J. & VANDEN BROECK, J. (2010) Myoinhibiting peptides are the ancestral ligands of the promiscuous Drosophila sex peptide receptor. *Cellular and Molecular Life Sciences* **67**, 3511–3522.

PRICE, T.A.R., LEWIS, Z., SMITH, D.T., HURST, G.D.D. & WEDELL, N. (2010) Sex ratio drive promotes sexual conflict and sexual coevolution in the fly drosophila pseudoobscura. *Evolution* **64**, 1504–1509.

PRÖSCHEL, M., ZHANG, Z. & PARSCH, J. (2006) Widespread Adaptive Evolution of Drosophila Genes With Sex-Biased Expression. *Genetics* **174**, 893–900.

PUERMA, E., ORENGO, D.J., CRUZ, F., GÓMEZ-GARRIDO, J., LIBRADO, P., SALGUERO, D., PAPACEIT, M., GUT, M., SEGARRA, C., ALIOTO, T.S. & AGUAD, M. (2018) The High-Quality Genome Sequence of the Oceanic Island Endemic Species Drosophila guanche Reveals Signals of Adaptive Evolution in Genes Related to Flight and Genome Stability. *Genome Biology and Evolution* **10**, 1956–1969.

RAM, K.R. & WOLFNER, M.F. (2007) Sustained post-mating response in Drosophila melanogaster requires multiple seminal fluid proteins. *PLoS Genetics* **3**, 2428–2438.

RAVI RAM, K. & WOLFNER, M.F. (2007) Seminal influences: Drosophila Acps and the molecular interplay between males and females during reproduction. *Integrative and Comparative Biology* **47**, 427–445.

RAVI RAM, K. & WOLFNER, M.F. (2009) A network of interactions among seminal proteins underlies the long-term postmating response in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 15384–15389.

RENSCHLER, G., RICHARD, G., VALSECCHI, C.I.K., TOSCANO, S., ARRIGONI, L., RAMÍREZ, F. & AKHTAR, A. (2019) Hi-C guided assemblies reveal conserved regulatory topologies on X and autosomes despite extensive genome shuffling. *Genes & development* **33**, 1591–1612.

REXHEPAJ, A., LIU, H., PENG, J., CHOFFAT, Y. & KUBLI, E. (2003) The sex-peptide DUP99B is expressed in the male ejaculatory duct and in the cardia of both sexes.

European Journal of Biochemistry **270**, 4306–4314.
REZÁVAL, C., PAVLOU, H.J., DORNAN, A.J., CHAN, Y.B., KRAVITZ, E. A. & GOODWIN, S.F. (2012) Neural circuitry underlying Drosophila female postmating behavioral responses. *Current Biology* **22**, 1155–1165.
RICE, W.R. (1996) Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. *Nature* **381**, 232–234.
RICE, W.R. & GAVRILETS, S. (2014) *The Genetics and Biology of Sexual Conflict*. Cold Spring Harbor Laboratory Press.
ROGINA, B. (2009) The effect of sex peptide and calorie intake on fecundity in female Drosophila melanogaster. *The Scientific World Journal* **9**, 1178–1189.
SANCHEZ-FLORES, A., PEÑALOZA, F., CARPINTEYRO-PONCE, J., NAZARIO-YEPIZ, N., ABREU-GOODGER, C., MACHADO, C.A. & MARKOW, T.A. (2016) Genome evolution in three species of cactophilic drosophila. *G3: Genes, Genomes, Genetics* **6**, 3097–3105.
DOS SANTOS, G., SCHROEDER, A.J., GOODMAN, J.L., STRELETS, V.B., CROSBY, M.A., THURMOND, J., EMMERT, D.B., GELBART, W.M., BROWN, N.H., KAUFMAN, T., WERNER-WASHBURNE, M., CRIPPS, R., BROLL, K., GRAMATES, L.S., FALLS, K., ET AL. (2015) FlyBase: Introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Research* **43**, D690–D697.
SAUDAN, P., HAUCK, K., SOLLER, M., CHOFFAT, Y., OTTIGER, M., SPÖRRI, M., DING, Z., HESS, D., GEHRIG, P.M., KLAUSER, S., HUNZIKER, P. & KUBLI, E. (2002) Ductus ejaculatorius peptide 99B (DUP99B), a novel Drosophila melanogaster sex-peptide pheromone. *European Journal of Biochemistry* **269**, 989–997.
SCHMIDT, T., CHOFFAT, Y., KLAUSER, S. & KUBLI, E. (1993) The Drosophila melanogaster sex-peptide: A molecular analysis of structure-function relationships. *Journal of Insect Physiology* **39**, 361–368.
SCHWENKE, R.A. & LAZZARO, B.P. (2017) Juvenile Hormone Suppresses Resistance to Infection in Mated Female Drosophila melanogaster. *Current Biology* **27**, 596–601.
SCOTT, J.G., WARREN, W.C., BEUKEBOOM, L.W., BOPP, D., CLARK, A.G., GIERS, S.D., HEDIGER, M., JONES, A.K., KASAI, S., LEICHTER, C.A., LI, M., MEISEL, R.P., MINX, P., MURPHY, T.D., NELSON, D.R., ET AL. (2014) Genome of the house fly, Musca domestica L., a global vector of diseases with adaptations to a septic environment. *Genome Biology* **15**, 466.
SINGH, A., BUEHNER, N.A., LIN, H., BARANOWSKI, K.J., FINDLAY, G.D. & WOLFNER, M.F. (2018) Long-term interaction between Drosophila sperm and sex peptide is mediated by other seminal proteins that bind only transiently to sperm. *Insect Biochemistry and Molecular Biology* **102**, 43–51.
SITNIK, J., GLIGOROV, D., MAEDA, R., KARCH, F. & WOLFNER, M.F. (2016) The female post-mating response requires genes expressed in the secondary cells of the male accessory gland in Drosophila melanogaster. *Genetics* **202**, 1029–1041.
SLATYER, R.A., MAUTZ, B.S., BACKWELL, P.R.Y. & JENNIONS, M.D. (2012) Estimating genetic benefits of polyandry from experimental studies: A meta-analysis. *Biological Reviews* **87**, 1–33.
SMITH, M.D., WERTHEIM, J.O., WEAVER, S., MURRELL, B., SCHEFFLER, K. & KOSAKOVSKY POND, S.L. (2015) Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Molecular Biology and*

Evolution **32**, 1342–1353. Oxford Academic.
SOTTOLANO, C.J., REVAITIS, N.T., GENEVA, A.J. & YAKOBY, N. (2022) Nebulous without white: annotated long-read genome assembly and CRISPR/Cas9 genome engineering in Drosophila nebulosa. *G3: Genes, Genomes, Genetics* **12**.
STECHER, G., TAMURA, K. & KUMAR, S. (2020) Molecular evolutionary genetics analysis (MEGA) for macOS. *Molecular Biology and Evolution* **37**, 1237–1239. Oxford University Press.
STYGER, D. (1992) Molekulare Analyse des Sexpeptidgens aus Drosophila melanogaster. University of Zurich, Zurich, Switzerland.
SUVOROV, A., KIM, B.Y., WANG, J., ARMSTRONG, E.E., PEEDE, D., D'AGOSTINO, E.R.R., PRICE, D.K., WADDELL, P., LANG, M., COURTIER-ORGOGOZO, V., DAVID, J.R., PETROV, D., MATUTE, D.R., SCHRIDER, D.R. & COMEAULT, A.A. (2022) Widespread introgression across a phylogeny of 155 Drosophila genomes. *Current Biology* **32**, 111-123.e5.
SUYAMA, M., TORRENTS, D. & BORK, P. (2006) PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**.
TEUFEL, F., ALMAGRO ARMENTEROS, J.J., JOHANSEN, A.R., GÍSLASON, M.H., PIHL, S.I., TSIRIGOS, K.D., WINTHER, O., BRUNAK, S., VON HEIJNE, G. & NIELSEN, H. (2022) SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology 2022 40:7* **40**, 1023–1025. Nature Publishing Group.
TOROSIN, N.S., ANAND, A., GOLLA, T.R., CAO, W. & ELLISON, C.E. (2020) 3D genome evolution and reorganization in the Drosophila melanogaster species group. *PLoS Genetics* **16**, e1009229.
TOWER, J., LANDIS, G.N., SHEN, J., CHOI, R., FAN, Y., LEE, D. & SONG, J. (2017) Mifepristone/RU486 acts in Drosophila melanogaster females to counteract the life span-shortening and pro-inflammatory effects of male Sex Peptide. *Biogerontology* **18**, 413–427.
TSUDA, M., PEYRE, J.B., ASANO, T. & AIGAKI, T. (2015) Visualizing molecular functions and cross-species activity of sex-peptide in Drosophila. *Genetics* **200**, 1161–1169.
TSUKAMOTO, Y., KATAOKA, H., NAGASAWA, H. & NAGATA, S. (2014) Mating changes the female dietary preference in the two-spotted cricket, Gryllus bimaculatus. *Frontiers in Physiology* **5 MAR**, 1–6.
TVEDTE, E.S., GASSER, M., SPARKLIN, B.C., MICHALSKI, J., HJELMEN, C.E., JOHNSTON, J.S., ZHAO, X., BROMLEY, R., TALLON, L.J., SADZEWICZ, L., RASKO, D.A. & DUNNING HOTOPP, J.C. (2021) Comparison of long-read sequencing technologies in interrogating bacteria and fly genomes. *G3: Genes, Genomes, Genetics* **11**.
VANDERLINDE, T., DUPIM, E.G., NAZARIO-YEPIZ, N.O. & CARVALHO, A.B. (2019) An Improved Genome Assembly for Drosophila navojoa, the Basal Species in the mojavensis Cluster. *Journal of Heredity* **110**, 118–123.
VICOSO, B. & BACHTROG, D. (2015) Numerous Transitions of Sex Chromosomes in Diptera. *PLoS Biology* **13**, e1002078.
WAINWRIGHT, S.M., HOPKINS, B.R., MENDES, C.C., SEKAR, A., KROEGER, B., HELLBERG, J.E.E.U., FAN, S.J., PAVEY, A., MARIE, P.P., LEIBLICH, A., SEPIL, I., CHARLES, P.D., THÉZÉNAS, M.L., FISCHER, R., KESSLER, B.M., ET AL. (2021) Drosophila Sex Peptide controls the assembly of lipid microcarriers in seminal fluid. *Proceedings of the*

National Academy of Sciences of the United States of America **118**, e2019622118.
WALKER, S.J., CORRALES-CARVAJAL, V.M. & RIBEIRO, C. (2015) Postmating Circuitry Modulates Salt Taste Processing to Increase Reproductive Output in Drosophila. *Current Biology* **25**, 2621–2630.
WARREN, A.S., ARCHULETA, J., FENG, W.C. & SETUBAL, J.C. (2010) Missing genes in the annotation of prokaryotic genomes. *BMC Bioinformatics* **11**, 1–12. BioMed Central.
WATERHOUSE, A.M., PROCTER, J.B., MARTIN, D.M.A., CLAMP, M. & BARTON, G.J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191. Oxford Academic.
WEAVER, S., SHANK, S.D., SPIELMAN, S.J., LI, M., MUSE, S. V. & KOSAKOVSKY POND, S.L. (2018) Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Molecular Biology and Evolution* **35**, 773–777. Oxford Academic.
WEI, K.H.C., MAI, D., CHATLA, K. & BACHTROG, D. (2022) Dynamics and Impacts of Transposable Element Proliferation in the Drosophila nasuta Species Group Radiation. *Molecular Biology and Evolution* **39**.
WENSING, K.U. & FRICKE, C. (2018) Divergence in sex peptide-mediated female post-mating responses in Drosophila melanogaster. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20181563.
WICKHAM, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
WIEGMANN, B.M., TRAUTWEIN, M.D., WINKLER, I.S., BARR, N.B., KIM, J.W., LAMBKIN, C., BERTONE, M.A., CASSEL, B.K., BAYLESS, K.M., HEIMBERG, A.M., WHEELER, B.M., PETERSON, K.J., PAPE, T., SINCLAIR, B.J., SKEVINGTON, J.H., ET AL. (2011) Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 5690–5695. National Academy of Sciences.
WIGBY, S. & CHAPMAN, T. (2005) Sex peptide causes mating costs in female Drosophila melanogaster. *Current Biology* **15**, 316–321.
WIGBY, S., SIROT, L.K., LINKLATER, J.R., BUEHNER, N., CALBOLI, F.C.F., BRETMAN, A., WOLFNER, M.F. & CHAPMAN, T. (2009) Seminal fluid protein allocation and male reproductive success. *Current Biology* **19**, 751–757.
YAPICI, N., KIM, Y.-J., RIBEIRO, C. & DICKSON, B.J. (2008) A receptor that mediates the post-mating switch in Drosophila reproductive behaviour. *Nature* **451**, 33–37.
YASSIN, A. (2013) Phylogenetic classification of the Drosophilidae Rondani (Diptera): The role of morphology in the postgenomic era. *Systematic Entomology* **38**, 349–364.
YU, G. (2020) Using ggtree to Visualize Data on Tree-Like Structures. *Current Protocols in Bioinformatics* **69**, e96. John Wiley & Sons, Ltd.
ZHANG, C., RABIEE, M., SAYYARI, E. & MIRARAB, S. (2018) ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153.