Evolution and adaptations of the seminal proteome in an insect with traumatic insemination

Short title: Bedbug sperm proteome evolution

Martin D. Garlovsky[1*], Oliver Otti[1], Klaus Reinhardt[1], Timothy L. Karr[2,3]

AFFILIATIONS

[1]Applied Zoology, Faculty Biology, Technische Universität Dresden, Dresden, Germany

[2]Biosciences Mass Spectrometry Core Research Facility, Knowledge Enterprise, Arizona State University, USA

[3]Neurodegenerative Disease Research Center, The Biodesign Institute, Arizona State University, USA

*Corresponding author: Martin D. Garlovsky, martin.garlovsky@tu-dresden.de

ABSTRACT

The protein composition of sperm and seminal fluid are key to male fitness. However, we currently lack an understanding of the factors that shape seminal proteome composition. The common bedbug (*Cimex lectularius*) mates by traumatic insemination, subjecting the ejaculate to a unique selective environment as sperm traverse the female genital and paragenital system. We provide the first high-throughput proteomic characterisation of the sperm and seminal fluid proteome in a hemimetabolous insect and the first in-depth proteomic characterisation of the male bedbug reproductive system. Our analysis revealed conserved and unique features of the sperm and seminal fluid proteome with possible links to features of sperm behaviour linked to traumatic insemination. The sperm proteome showed elevated rates of molecular evolution, unlike most other studied species. Conversely, the sperm proteome also contained many conserved proteins. Notably, we found an expansion of Sperm-leucylaminopeptidases (S-Laps) in bedbugs and other hemimetabolous insects, suggesting the origin of S-Laps is perhaps even more ancient than previously thought. Using *in silico* protein-ligand binding predictions, we show that S-Laps have likely retained catalytic activity. Our results provide a list of candidate proteins involved in reproduction and a foundation for future studies of this expanding global pest.

SUMMARY

Proteins are critical components of sperm and seminal fluid needed for fertilisation. Evolutionary novelties in unusual sexual systems represent exciting opportunities to advance our understanding of the evolution and composition of the male seminal proteome. The common bedbug (*Cimex lectularius*) provides an exciting opportunity to advance our understanding of the seminal proteome because unlike in most other species, sperm must cross several tissues and substrates before reaching the egg. Here we use quantitative proteomics to show unusual characteristics of the male bedbug reproductive organs and sperm. Unlike in most species studies so far, bedbug sperm proteins show fast rates of evolution which may relate to unusual reproductive biology. We also identified many proteins characteristic of male seminal proteins found in diverse taxa. Notably, we found a sub-class of proteins first identified in *Drosophila* suggesting an ancient origin of these integral sperm proteins. However, in bedbugs and other insects these proteins retain enzymatic activity, unlike in *Drosophila*. Overall, our findings will help future researchers identify important proteins involved in reproduction for developing strategies to control this expanding global pest. More widely, our results challenge our understanding of the patterns shaping male reproductive trait evolution.

KEYWORDS

Sperm | seminal fluid proteins | bedbugs | leucyl-aminopeptidases | AlphaFold3

INTRODUCTION

The proteomic composition of sperm and seminal fluid is central to fertility and fitness [1,2]. Proteins are a major component of gametes, in large part dictating spermatozoa form and fertilisation capacity [3]. Likewise, seminal fluid proteins transferred along with sperm in the ejaculate aid sperm transport and storage and affect female physiology and morphology after mating [4]. However, until recently the tools needed to study the molecular composition of the ejaculate were lacking [5]. Consequently, our understanding of the factors shaping seminal proteome composition and evolution remain poorly understood.

High-throughput shotgun proteomics studies have revealed the sperm and seminal fluid proteome contain hundreds or thousands of proteins [4,6]. Recent studies have highlighted a nuanced landscape of the seminal proteome. For instance, some seminal fluid proteins may have a testis origin [7,8]. Many reproductive proteins show elevated rates of molecular evolution [9–11]. However, the sperm and seminal fluid proteome also show conservation across taxa [6,12]. For instance, Sperm-leucylaminopeptidases (S-Laps), first identified in the *Drosophila melanogaster* sperm proteome, have been identified in many insects [13–20]. Similarly, peptidases and peptidase inhibitors are common features of seminal fluid proteomes that initiate cascades of downstream interactions between female and male proteins to ensure long-term postmating responses [21–25]. Studying the seminal proteome in non-model organisms allows testing commonly held assumptions about reproductive proteins [26].

The common bedbug (*Cimex lectularius*) presents an attractive model to study the factors shaping the seminal proteome, with unique selective pressures potentially acting on the male ejaculate. First, in species where males are the heterogametic sex (i.e., XY), genes with male-biased expression are expected to be underrepresented on the X chromosome which spends two-thirds of time in females [27]. The karyotype in bedbugs is $2n = 26 + X_1X_2Y$ in males [28–30], which may reduce sex biased selection and alter the landscape of sex biased genes. Second, bedbugs obligately mate by traumatic insemination, where the male intromittent organ (paramere) pierces the female cuticle, bypassing the female reproductive tract, which is used only during oviposition [31]. Insemination takes place in the mesospermalege, a novel organ that has evolved in the Cimicidae [32], which reduces mating costs to females [33,34]. As the site of insemination, sperm first interact with the female in the mesospermalege before sperm exit into the female haemolymph [35,36]. Sperm then travel through the haemolymph and cross into the female reproductive tract to enter the oviduct via an as yet unknown mechanism, where some sperm are stored in a unique storage organ [35]. Other sperm move up the oviduct and, unlike in other insects, fertilise eggs in the ovaries [35]. Sperm must therefore traverse multiple tissues, membranes and substrates to reach the egg [35].

To explore features of the bedbug seminal proteome associated with traumatic insemination, we provide a detailed proteomic investigation of the male bedbug reproductive organs, including the sperm and seminal proteomes. This is the first high-throughput seminal proteome of a hemimetabolous insect, greatly increasing the number of ejaculate proteins in bedbugs [37]. Recently published genomic resources [38,39] allow us to further investigate the evolutionary dynamics of bedbug reproductive proteins, including rates of molecular evolution and chromosomal distribution. Finally, we compare the bedbug seminal proteome to *Drosophila* and mosquitoes (another blood-feeding insect). In doing so, we identified an expansion of the *S-Lap* gene family in the Cimicomorpha, showing retention of catalytic activity based on sequence similarity and *in silico* protein-ligand interactions.

RESULTS

Label-free quantitation using LC-MS/MS of the *Cimex lectularius* male reproductive tract tissues and sperm provide a unique and quantitative view of the bedbug seminal proteome. In total, we identified 4,651 proteins, of which we retained 3,348 (71.7%; Table 1; Fig. S2) after filtering proteins identified by fewer than 2 unique peptides and default filtering steps in *MSstats* [40].

Table 1. Total numbers of proteins identified in each tissue/sperm and the number of tissue/sperm-biased proteins (as a percentage of total in brackets. SFCs, seminal fluid containers.

	Testes	Sperm	Mesadenes	SFCs
No. proteins identified	3244	2494	3017	2272
No. tissue-biased (%)	535 (16.5%)	487 (19.5%)	708 (23.5%)	239 (10.5%)

Unlike most other insects, true bugs have separate organs which are the source of seminal fluid production; the mesadenes, which are physically separated from the seminal fluid storage organ, the seminal fluid containers (Fig. 1A). A principal component analysis (PCA) of normalised protein abundances showed the first two PCs explained more than 83% of variation, separating all four tissues (Fig. 1B). PC1 (47.9% variance explained) separated the sperm and testes from the seminal fluid containers and mesadenes (Fig. 1B). PC2 (35.2% variance explained) further separated the sperm and seminal fluid containers from the testes and mesadenes. This suggests some gross similarities between the organs that are the sources of protein production (testes and mesadenes), separate from the sinks (sperm and seminal fluid containers).

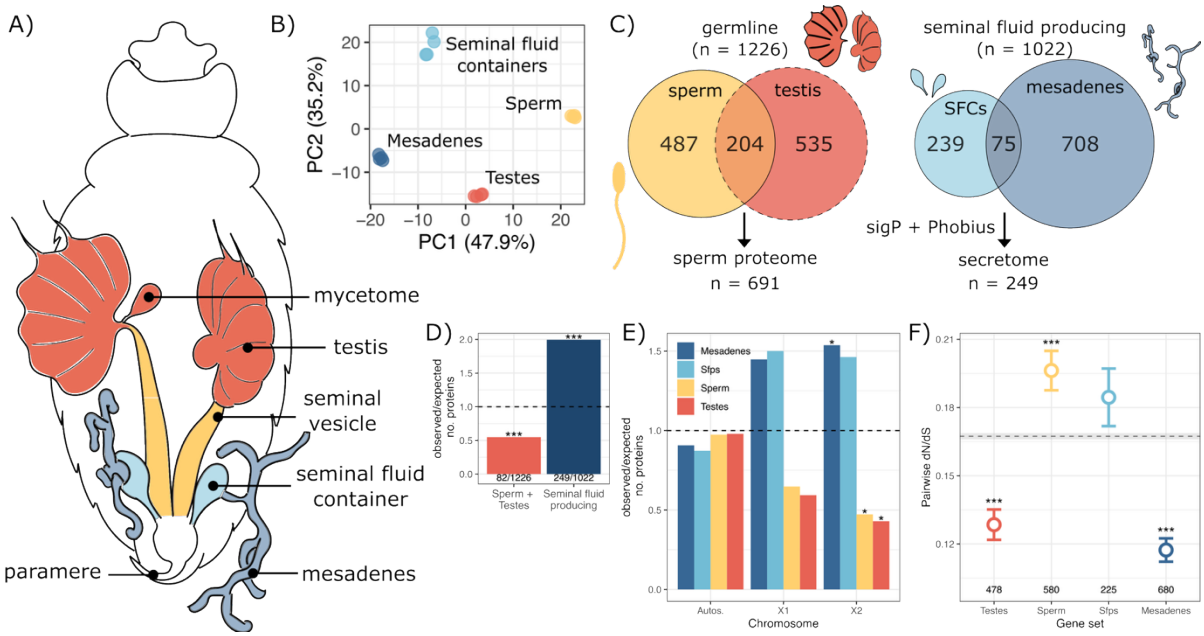

Figure 1. (A) *Dorsal view of male* Cimex lectularius *reproductive system highlighting organs dissected for proteomics (adapted from* [35]*). Sperm were isolated from the seminal vesicles. Bacteriomes (aka mycetomes) were included with the testes for LC-MS/MS.* ***B)*** *Principal component analysis of protein abundances for the 500 most*

variable proteins. ***C)*** *Reproductive protein classification workflow. Numbers in Venn diagrams show number of proteins biased towards each tissue (log2FC > 1 & adjusted p-value < 0.05).* ***D)*** *Enrichment of proteins containing a signal peptide sequence in the bedbug reproductive proteome. Numbers below bars are the observed and expected number of proteins in each group, respectively. The dashed line indicates the null expectation, i.e., the expected proportion of proteins in the whole proteome. P-values obtained from* χ^2 *tests.* ***E)*** *Chromosomal distribution of male-biased proteins on the autosomes and each of the two* C. lectularius *X chromosomes. Asterisks represent results from comparing the observed to expected number of proteins on each chromosome after multiple testing correction using* χ^2 *tests. The dashed line indicates the null expectation.* ***F)*** *Pairwise nonsynonymous (dN) to synonymous (dS) nucleotide substitution rate, dN/dS (mean ± standard error) between C. lectularius and C. hemipterus. Numbers below points indicate numbers of genes in each category. Asterisks represent results from Mann-Whitney U tests comparing each gene set to the genome average shown by dashed line (dN/dS = 0.167 ± 0.001, n = 17580). * p < 0.05, ** p < 0.01, *** p < 0.001. Sfp, seminal fluid proteins. SFCs, seminal fluid containers.*

Characterisation of the male reproductive proteome

We defined proteomes for each tissue/sperm based on proteins having higher abundance (or the exclusive presence) compared to all other tissues (log2FC > 1 & adjusted *p*-value < 0.05; Fig. 1C; Table 1; Fig. S2&S3).

Testis and sperm proteome. The combined testis and sperm proteome comprised 1,226 proteins and showed underrepresentation of proteins containing a signal peptide sequence (χ^2 = 30.67, df = 1, $p < 0.001$; Fig. 1D). Sperm- and testes-biased proteins were underrepresented on chromosome X_2 (sperm: χ^2 = 10.0, df = 1, $p = 0.046$; testes: χ^2 = 9.14, df = 1, $p = 0.05$; Fig. 1E & Fig. S5).

The putative sperm proteome, comprising proteins biased towards sperm or sperm and testes (n = 691; Fig. 1C) was evolving at an elevated rate compared to the genome average (mean pairwise dN/dS ± standard error = 0.20 ± 0.01, n = 580; pairwise Wilcoxon test, $p < 0.001$; Fig. 1F). The sperm proteome showed GO enrichment of biological processes involving metabolic and glycolytic processes, proteolysis, cilium assembly and microtubule-based process; cellular components typically associated with sperm cells, such as mitochondria, cilium, axoneme, dynein complex, cytoplasm and microtubules; and molecular functions including ion binding and serine-type endopeptidase activity (Table S3).

Proteins exclusively biased towards testes (n = 535) showed GO enrichment of biological processes including protein folding, cellular transport and organisation and translation; cellular components including proteasomes, cytoplasm, microtubules and mitochondria; molecular functions including protein folding, ATP binding and heat shock protein binding (Table S4). Endosymbiotic bacteria (e.g., *Wolbachia*) [41] are localised in bacteriomes (aka mycetomes) attached to the *vas deferens* at the base

183 of the testes [35,42]. As expected, *Wolbachia* proteins showed the highest
184 abundance in the testes/bacteriome samples (Fig. S4).
185
186 ***Mesadene and seminal fluid proteome.*** The seminal fluid producing proteome,
187 including proteins biased towards mesadenes and/or seminal fluid containers
188 comprised 1022 proteins (Fig. 1C). Proteins containing a signal peptide sequence
189 were overrepresented in the mesadenes and seminal fluid containers (X^2 = 123.45,
190 df = 1, $p < 0.001$; Fig. 1D). We classified the 249 proteins containing a signal peptide
191 sequence from the seminal fluid producing organs as our putative seminal fluid
192 proteome (Fig. 1C); a number similar to the number of "high confidence" Sfps
193 identified in *Drosophila melanogaster* [4]. While the presence of a signal peptide
194 sequence is indicative of a protein being secreted [43] and a hallmark of seminal
195 fluid proteins, not all secreted proteins contain signal peptides. We found a subset of
196 putative seminal fluid proteins expressed highly (in the top 20% most abundant
197 proteins) in sperm (3.6%; 9/249) and testes (8%; 20/249), in line with recent studies
198 suggesting a more complex landscape of seminal fluid biology [7,8].
199
200 Seminal fluid proteins were evolving at a similar rate to sperm (p = 0.106) but were
201 not evolving faster than the genome average (dN/dS = 0.18 ± 0.01, n = 224; p =
202 0.135; Fig. 1F). The seminal fluid proteome showed GO enrichment for biological
203 processes including protein folding and proteolysis, glycosylation and negative
204 regulation of peptidase activity; cellular components including endoplasmic reticulum
205 and extracellular space; and molecular functions including ion and molecule binding,
206 and (serin-type) endopeptidase inhibitor activity (Table S5).
207
208 The remaining 773 proteins classified as the seminal fluid producing mesadenal
209 category were overrepresented on X_2 (X^2 = 11.8, df = 1, p = 0.035; Fig. 1E).
210 Mesdadenal proteins showed GO enrichments often associated with accessory
211 gland/secretory functions; biological processes such as translation, protein transport
212 and Golgi/vesicle transport; cellular components including ribosome, endoplasmic
213 reticulum and Golgi apparatus; molecular functions including RNA- and GTP- binding
214 (Table S6).
215
216 ***Comparison with other seminal proteomes***
217 In the bedbug sperm proteome (n = 691), 189 orthologs of *D. melanogaster* sperm
218 proteins were identified [7,8] and 58 *Aedes aegypti* sperm proteins [17]. Among the
219 most abundant proteins in the bedbug sperm proteome were orthologs of *D.*
220 *melanogaster* sperm proteins, including proteins involved in spermatid development
221 (e.g., betaTub56D, bb8), S-Laps, mitochondrial proteins (e.g., P5CDh1, Cs1, Ant2,
222 mAcon, Mitofilin) and Y-linked fertility factors (e.g., ORY, WDY, Ppr-Y, kl-3, kl-5)
223 (Table 2; Table S7). The bedbug seminal fluid proteome (n = 249) contained fewer
224 orthologs; just 15 *D. melanogaster* seminal fluid proteins [4], and 5 *Ae. aegypti*
225 seminal fluid proteins [17].
226

227 ***Table 2.*** *The most abundant proteins in the bedbug sperm- and seminal fluid-*
228 *proteomes with* D. melanogaster *orthologs (rank ordered by abundance). Bedbug*
229 *proteins matching more than one* Drosophila *gene are separated by a comma. S-*
230 *Laps are labelled based on one-to-one orthology with* D. melanogaster *or*
231 *membership to S-Lap cluster 1 (g1) or 2 (g2)* [13] *followed by a unique ID.* D.
232 melanogaster *Sfps are noted with an asterisk (all found in the Sfp category).*

Sperm proteome			Seminal fluid proteome		
Protein	*Drosophila* name(s)	Rank	Protein	*Drosophila* name(s)	Rank
A0A8I6R8Q4	betaTub56D	100	A0A8I6TE74	Pdi	99.8
A0A8I6S6N3	S-Lap3	99.9	A0A8I6S6T0	Gp93	99.6
A0A8I6THS3	S-Lap-g1_2	99.7	A0A8I6R8G9	OstDelta	99.1
A0A8I6RD41	S-Lap4	99.6	A0A8I6S4A5	Grp170	98.6
A0A8I6RDX4	bb8	99.5	A0A8I6SFV2	mfas	95.0
A0A8I6R8J1	S-Lap-g2_4	99.1	A0A8I6SCN7	Swim	84.4
A0A8I6RZZ4	P5CDh1	99.0	A0A8I6RVV7	NUCB1	81.0
A0A8I6TEZ7	S-Lap-g1_5	98.9	A0A8I6RXP7	GCS2alpha	88.6
A0A8I6TD95	S-Lap-g2_6	98.7	A0A8I6SQD0	Prx4	85.3
A0A8I6TJT9	Cs1	98.5	A0A8I6RJB5	Spn88Ea, Spn88Eb	65.4
A0A8I6RI95	Ant2	98.4	A0A8I6RPW4	CG17271	74.1
A0A8I6TIM8	mAcon1, mAcon2	98.4	A0A8I6RND8	Phm	46.7
A0A8I6RBN0	Mitofilin	98.0	A0A8I6RWL8	CG8420	67.3
A0A8I6RL88	Alat	97.6	A0A8I6RAJ9	CG12093	23.0
A0A8I6SB96	S-Lap-g1_7	97.5	A0A8I6RFN7	Qsox1, Qsox3, Qsox2, Qsox4	33.9
A0A8I6SDI3	CG3213, CG13405	97.3			
A0A8I6RRZ2	CG3339, kl-5	97.2			
A0A8I6TFY4	SdhA	96.6			
A0A8I6S5H6	kl-3	96.4			
A0A8I6SIB1	AIF	96.3			

233

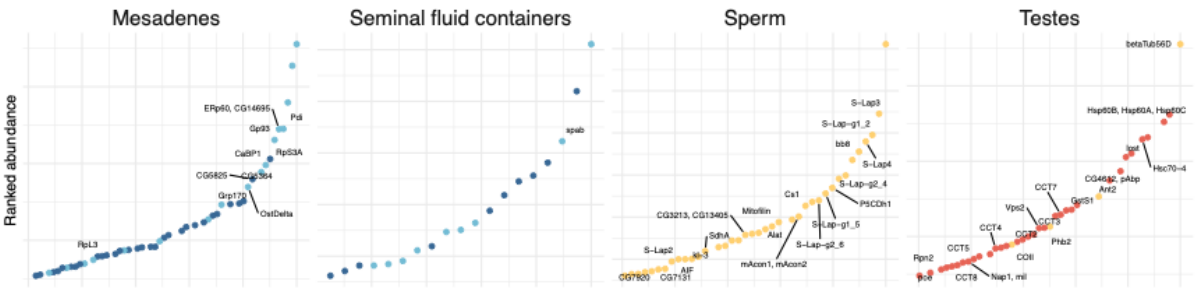

Figure 2. *Ranked abundance of the top 5% most abundant proteins in each tissue/sperm. Points are coloured by proteome category (mesadenes, dark blue; seminal fluid proteins, light blue; sperm, yellow; testes, orange). Orthologs of* Drosophila melanogaster *sperm* [7] *and seminal fluid proteins* [4] *are labelled. S-Laps are labelled based on one-to-one orthology with* D. melanogaster *or membership to S-Lap cluster 1 or 2* [13] *followed by a unique ID.*

Sperm-leucylaminopeptidase evolution

Notably, 6 of the 10 most abundant bedbug sperm proteins with *D. melanogaster* orthologs were S-Laps (Fig. 2; Table 2) [13]. We identified 13 *C. lectularius* S-Lap orthologs in total; five belonging to S-Lap cluster 1, eight in cluster 2 [13], all of which were most abundant in sperm (Fig. S6). To further explore orthology between the bedbug and *Drosophila* S-Laps we used BLASTp [44] which showed matches to all 8 *Drosophila* S-Laps. A more stringent reciprocal best BLAST search (e-value = 1e-2, min. ident = 30, min. coverage = 50) returned four one-to-one S-Lap orthologs: S-Laps 2, 3, 4 and 7 (Fig. 3B).

There appear to be no S-Lap orthologs in more distantly related Hemiptera (*Acyrthosiphon pisum*) or Orthoptera (*Gryllus bimaculatus*), but S-Lap orthologs were present in the Coleoptera (*Callosobruchus maculatus*) (see also [20] (Fig. 3A). However, orthologs of *granny-smith*, the closest S-Lap paralog [13], were identified in all species studied (Fig. 3A & Fig. S7). Across the 73 S-Lap orthologs from 7 species studied (including *D. melanogaster*), proteins first cluster by S-Lap cluster (1 or 2) rather than species based on sequence similarity, suggesting the origin of S-Laps predates the separation of the Holometabola from the Hemimetabola [45].

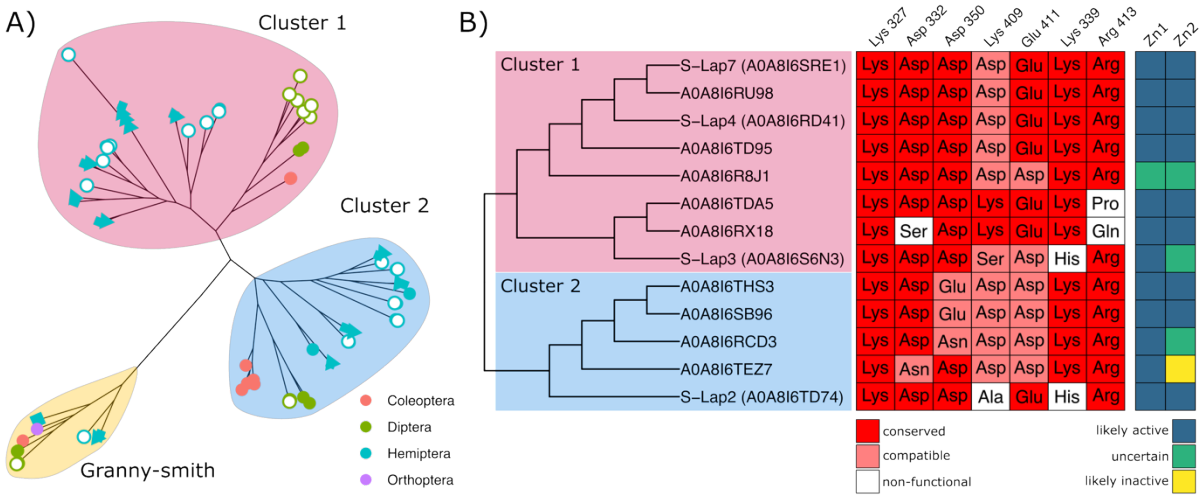

Figure 3. *Sperm-leucylaminopeptidase (S-Lap) enzymatic activity in bedbugs.* ***A)*** *Unrooted gene tree of S-Lap and granny-smith orthologs in insects. Protein sequences were aligned using MAFFT* [46] *and bootstrapped 1,000 times. Tip colour indicates Order and shape indicates species.* Drosophila *and bedbug S-Laps shown with open circles.* ***B)*** *Alignment of bedbug S-Lap orthologs focused on amino acid residues against the M17 amino peptidase metal binding motif indicated above [47]. Amino acids are coloured by alignment to consensus (conserved, red; compatible, pink; non-functional, white). Zn^{2+} ion binding at the "loose" binding site (Zn1) and "tight" binding site (Zn2) is scored as likely active (blue), uncertain (green), or likely inactive (yellow) based on coordinating residue geometry, number and inter-ligand angles measured from AlphaFold3 models in ChimeraX.*

M17 leucyl-aminopeptidases (LAPs) bind two Zn^{2+} ions at specific amino acid residues [48]. S-Laps in *D. melanogaster* belonging to "cluster 1" have mutations at most of the active site residues and sperm display very weak enzymatic activity [13,18]. Most (11 out of 13) bedbug S-Laps, including those in "cluster 1", retain at least 4 out of 5 metal ion binding residues including all the residues coordinating tight Zn^{2+} binding and cofactor binding sites (Fig. 3B).

Structure-based predictions to determine 3D coordination of Zn^{2+} ions also showed that all bedbug S-Laps retain metal binding properties (Fig. 3B & Fig. S7). This pattern of (i) retention of metal binding active site residues, and (ii) 3D coordination of Zn^{2+} was observed across a range of species outside of *Drosophila* including other hemipterans and hemimetabolous insects as well as in Diptera and Coleoptera (Fig. S7). Notably, *C. maculatus* S-Laps in "cluster 1" do not have the conserved Asp residue coordinating site 2 tight zinc binding and lack zinc binding geometry (Fig. S7).

290 **DISCUSSION**
291 Bedbugs are unusual among insects in obligately mating by traumatic insemination.
292 After insemination, sperm move in aggregations out of the mesospermalege towards
293 the ovaries or female sperm storage organs [35,36]. Sperm must therefore cross
294 multiple tissue membranes, likely shaping key adaptations in the bedbug seminal
295 proteome. Our analysis revealed conserved and unique features of the sperm and
296 seminal fluid proteome with possible links to this unusual sperm behaviour.
297
298 Sperm and testes proteins were underrepresented on chromosome X_2, as expected
299 for proteins with male-biased expression [27]. However, the same pattern was not
300 found for chromosome X_1. The two X chromosomes in *Cimex* are thought to be
301 derived from fission of an ancestral X [30,49]. Mesadenal proteins, on the other
302 hand, were overrepresented on X_2, which may result if new mesadenal genes have
303 arisen on the X [50]. The multiple sex chromosomes present interesting opportunities
304 for unusual evolutionary dynamics that warrant further investigation.
305
306 We did not observe the commonly found pattern of rapidly evolving seminal fluid
307 proteins [9], but found few seminal fluid protein orthologs in *Drosophila* or *Aedes*
308 mosquitoes. Surprisingly, elevated rates of molecular evolution were observed in the
309 bedbug sperm proteome. In many taxa, the sperm proteome evolves slowly, thought
310 to result from purifying selection [7]. Elevated rates of sperm evolution might be
311 driven by sexually antagonistic co-evolution in the bedbug associated with traumatic
312 insemination and movement of sperm through the haemolymph. The mating rate in
313 bedbugs is at least ten times above the female optimum [51,52], setting the stage for
314 sexual conflict affecting ejaculate × female interactions [33,34,53]. Sperm traversing
315 the female haemolymph and female reproductive tract might entail novel and
316 unusual selective pressures acting on sperm and seminal fluid proteins. As such, it is
317 unclear if or for how long seminal fluid associates with sperm after insemination, as
318 found in other species [54]. We also found a subset of putative seminal fluid proteins
319 were abundant in sperm and testes, in line with recent studies which point to a
320 possible testis origin of some seminal fluid proteins [7,8].
321
322 Despite overall elevated rate of evolution in the bedbug sperm proteome, many of
323 the most abundant sperm proteins had orthologs in other insects. Most notably, we
324 identified 13 Sperm-leucylaminopeptidases (S-Laps) in bedbugs and other insects
325 including other hemimetabolous insects. The S-Laps in *Drosophila* show evidence of
326 purifying selection [13] and indeed were evolving more slowly than the rest of the
327 genome in bedbugs (Fig. S8). Other S-Lap orthologs have been identified in
328 Lepidoptera, Coleoptera and Diptera, suggesting an ancient origin of these sperm
329 proteins after more than 380 million years since the last common ancestor [7,13–20].
330
331 S-Laps belong to the M17 family of metal-dependent leucyl aminopeptidases, but in
332 *Drosophila* have mostly lost enzymatic activity [13,18]. Our functional analysis
333 revealed that S-Laps in *Cimex* retain the Zn^{2+} binding motif, suggesting that S-Laps

retained enzymatic properties. Additional direct biochemical assay of *Cimex* S-Laps will help support the idea that enzymatic activity has been secondarily lost in *Drosophila*. Our analysis suggests neofunctionalisation of this class of M17 leucyl-aminopeptidases while retaining an active site in bedbugs and other insects. S-Laps have been identified in whirligig beetles and also retain the metal binding motif [20]. Whirligig sperm move in conjugates attached to spermatostyles [55]. Bedbug sperm also move in conjugates, lending support to the notion that S-Laps play a role in sperm conjugation and cooperation [20].

Conclusions

In conclusion, our study provides a substantially expanded database of male reproductive proteins in bedbugs using high-throughput proteomics, increasing coverage of the bedbug seminal proteome by two orders of magnitude [37]. Our analysis provides insights into the landscape of sperm and seminal fluid dynamics in an insect with unique features of reproductive anatomy and physiology. The peculiarities of traumatic insemination and multiple sex chromosomes in the bedbug have led to unique evolutionary characteristics of the seminal proteome. Sperm genes are underrepresented on the second X chromosome and evolve rapidly. S-Laps have retained catalytic activity, which may aid in sperm conjugation. A better understanding of the proteins important for reproduction will aid the development of effective strategies to control and suppress populations of insect pests and disease vectors [56]. Unlike many other blood-feeding insects, bedbugs do not transmit any known human pathogens [56]. Nevertheless, bedbugs are an expanding global pest, causing sleeplessness, skin irritation, severe allergic reactions and growing economic costs [56–58]. Our results provide a useful resource for future studies and a list of candidate proteins involved in reproduction in this economically important insect.

361 **METHODS**

362 ***Bedbug maintenance***

363 We used bedbugs originally collected in London, UK (F4 population) in 2006 and
364 maintained at TU Dresden, Germany, as a large, outbred population with
365 overlapping generations. Bedbugs were maintained at 26ºC, 70% relative humidity
366 on a 12:12 hour light:dark cycle. Bedbugs were fed on human blood once each week
367 during the juvenile stages following long-term experimental protocols [34]. We used
368 males aged 3 weeks post-final moult. Males were housed in groups of 10-15 and fed
369 once a week for 2 weeks after eclosion as adults and left to mature for a further
370 week.
371

372 ***Tissue collection and protein extraction***

373 All dissections were performed in freshly prepared ice-cold phosphate-buffered
374 saline (1X PBS) with 1X protease inhibitors (HALT) and 5mM EDTA (Thermo
375 Fisher). We anaesthetised males on ice and removed reproductive tracts with
376 forceps, insect pins, and dissecting scissors under a stereo dissecting microscope.
377 We had four replicates consisting of 4-6 males each.
378

379 To collect sperm samples, seminal vesicles were separated from the remaining
380 reproductive tract tissue and transferred to a fresh drop of PBS/Halt. We then made
381 a small incision at the distal end of the seminal vesicle and exuded the sperm mass
382 using insect pins. The sperm mass was then transferred to an Eppendorf containing
383 1mL PBS/Halt kept on ice. Seminal fluid containers and mesadenes were moved to
384 a fresh drop of PBS/Halt and separated using insect pins. Mesadenes were
385 transferred to an Eppendorf containing 1mL PBS/Halt kept on ice. Seminal fluid
386 containers were transferred in 2 µL PBS/Halt per gland to a separate Eppendorf.
387 Testes were separated from the remaining tissue using insect pins and transferred to
388 an Eppendorf containing 1mL PBS/Halt kept on ice. Sperm, testis, and mesadene
389 samples were pelleted at 15,000 × G for 2 minutes at 4ºC, supernatant carefully
390 removed, and the pellet resuspended by addition of 1 mL PBS/Halt followed by an
391 additional 2 mins centrifugation at 15,000 × G. The washing procedure was repeated
392 twice, and the final pellet resuspended in 1 mL PBS/Halt and stored at -80ºC.
393 Seminal fluid containers were freeze/thawed and vortexed for 30 seconds three
394 times and then spun at 15,000 × G for 5 minutes at 4ºC and stored at -80ºC
395 suspended in 10 µL of PBS/Halt.
396

397 ***Liquid chromatography tandem mass spectrometry***

398 All LC-MS analyses were performed at the Biosciences Mass Spectrometry Core
399 Facility (https://cores.research.asu.edu/mass-spec/) at Arizona State University. For
400 LC–MS/MS, solubilised proteins were quantified (Thermo Fisher EZQ Protein
401 Quantitation Kit or the Pierce BCA). Proteins were reduced with 50 mM dithiothreitol
402 (Sigma-Aldrich) at 95 °C for 10 min and alkylated for 30 min with 40 mM
403 iodoacetamide (Pierce). Proteins were digested using 2.0 µg of MS-grade porcine
404 trypsin (Pierce), and peptides were recovered using S-trap Micro Columns (Protifi)

405 per manufacturer directions. Recovered peptides were dried via speed vac and
406 resuspended in 30 µL of 0.1% formic acid. All data-dependent mass spectra were
407 collected in positive mode using an Orbitrap Fusion Lumos mass spectrometer
408 (Thermo Scientific) coupled with a Dionex UltiMate 3000 system equipped with an
409 NCS-3500 nano capillary pump and UltiMate 3000 RS autosampler. One µL of the
410 peptide was fractionated using an Easy-Spray LC column (50 cm Å ~75 µm ID,
411 PepMap C18, 2 µm particles, 100 Å pore size, Thermo Scientific). Electrospray
412 potential was set to 1.9 kV and the ion transfer tube temperature to 275 °C. The
413 mass spectra were collected using the method optimised for peptide analysis
414 provided by Thermo Scientific. Full MS scans (350–1500 m/z range) were acquired
415 in profile mode with the following settings: Orbitrap resolution 120,000, cycle time 3 s
416 and mass range "Normal;" RF lens at 35%, and the AGC set to "Standard".
417 Monoisotopic peak determination (MIPS) at "peptide" and included charge states 2–
418 7; dynamic exclusion at 60 s, mass tolerance 10 ppm, intensity threshold at 5.0e4;
419 MS/MS spectra acquired in a centroid mode using quadrupole isolation window at
420 1.6 (m/z); Higher-energy collisional dissociation (HCD) energies at 35%. Spectra
421 were acquired over a 180-min gradient, flow rate 0.250 µL/min as follows: 0–10 min
422 at 0-2%, 10–100 min at 2–8%, 100–110 min at 8–32%, 110-120 min at 32–35%,
423 120–125 min at 35–45%, 125–155 min at 45-98%,155–165 min at 98%, 165 –
424 180 min at 2%,followed by post run wash step.
425
426 ***Protein identification and label-free quantification***
427 We analysed raw files searched against the UniProt (www.uniprot.org) *Cimex*
428 *lectularius* database that included two bacterial predicted proteomes
429 (CIMLE_UP000494040; wCle UP000031663; Rickettsia UP000826664) using
430 Proteome Discover 2.5 (Thermo Scientific). Input data included Trypsin as enzyme,
431 maximum missed cleavage site 3, min/max peptide length 6/144, and precursor ion
432 (MS1) mass tolerance set to 20 ppm and fragment mass tolerance set to 0.5 Da.
433 Carbamidomethyl (C) was specified as fixed modification, and dynamic modifications
434 set to Aceytl and Met-loss at the N-terminus, and oxidation of Met. SequestHT and
435 Percolator ID validation was performed using a minimum of 1 peptide identified using
436 a concatenated target/decoy strategy and false-discovery rate (FDR) set to 1.0%.
437 The Minora feature detector identified between run peak matches and quantification
438 intensities, normalised and scaled to a maximum of 100. Intensity values were
439 averaged across all six runs and ranked from highest to lowest to obtain a relative
440 abundance profile. A combined list of predicted signal peptide sequences from the
441 entire *C. lectularius* proteome was obtained using signal-P 6 [43] and Phobius [59].
442
443 ***Ortholog identification***
444 We downloaded proteomes for *Cimex hemipterus*, assassin bug (*Triatoma*
445 *infestans*), kissing bug (*Rhodnius prolixus*), pea aphid (*Acyrthosiphon pisum*), seed
446 beetle (*Callosobruchus maculatus*), yellow fever mosquito (*Aedes aegypti*) and fruit
447 fly (*Drosophila melanogaster*) from uniprot.org and two-spotted cricket (*Gryllus*
448 *bimaculatus*) from NCBI (https://www.ncbi.nlm.nih.gov/) (Fig. S1; Table S1). Species

were chosen because of their evolutionary relationship to *Cimex lectularius*, i.e., the true bugs and hemimetabolous insects (aphid, cricket), because of similar biology, such as blood feeding (mosquito), or because of a wealth of data on reproductive biology, including the occurrence of copulatory wounding (*D. melanogaster* and seed beetle). Hierarchical orthologs were identified in each species with *OrthoFinder* using default settings [60,61].

Evolutionary dynamics

We obtained previously calculated pairwise rates of nonsynonymous (dN) and synonymous (dS) nucleotide substitution rates (dN/dS) and chromosomal positions of genes [36]. Briefly, rates of molecular evolution (dN/dS) were calculated between *C. lectularius* and *C. hemipterius* using PAML [62]. Chromosomal distribution of genes was determined by aligning coding sequences from the *C. lectularius* reference genome Clec_2.1 [63] to the recently published chromosome-level *C. lectularius* genome assembly [39] using BLASTn [44].

Gene ontology enrichment analysis

We performed gene ontology enrichment analysis using the topGO package [64] in R [65] using the GO annotations from uniprot.org. The entire bedbug proteome was used as background for enrichment tests. Statistical analyses in topGO used Fisher's exact tests with topnodes = 10, and corrected p-values for multiple testing using the FDR method [66]. All GO terms reported had an FDR-corrected p-value < 0.05.

Statistical analysis

All statistical analysis was performed in R v4.5.0 [65]. Output from Proteome Discoverer was imported into R and formatted for use in the *MSstats* pipeline [40] using the following options: normalisation = quantile, imputed missing values using the TMP method, MBimpute = TRUE.

To identify proteins biased towards each tissue (seminal fluid containers, mesadenes, testes) or sperm, we performed pairwise differential abundance analysis between each tissue/sperm. We considered a protein to be biased towards a tissue/sperm based on a log2-fold-change > 1 and false discovery corrected p-value < 0.05 compared to all other tissues. Proteins had to be identified in two or more replicates of a tissue/sperm to count for that tissue/sperm. We then classified proteins into 1 of 4 categories: (1) the putative seminal fluid proteome was defined as proteins biased towards the mesadenes and/or seminal fluid containers containing a signal peptide sequence; (2) the remaining proteins from the seminal fluid producing organs were classified as the mesadenal (seminal fluid producing) proteome; (3) the putative *C. lectularius* sperm proteome was defined as proteins more abundant in sperm or sperm and testes; (4) the remaining proteins biased towards testes we classified as the testes proteome (Fig. 1C).

493 ***Predicting protein-interactions***
494 We characterised catalytic activity for S-Laps based on conservation of the M17
495 aminopeptidase metal-binding and catalytic residue motif [48]. Orthologs of S-Laps
496 and granny-smith (UniProt ID: Q9V3D8) identified using *OrthoFinder* and cytosol
497 aminopeptidase (LAP3; UniProt ID: P00727) were aligned using *MUSCLE* [67]. The
498 catalytic domain of each sequence was identified by mapping a ±150 residue
499 window around the active site centre of *granny-smith*. The position equivalent to the
500 start of the motif (Lys327) was pinned in each sequence and the 7 active site
501 positions scored at fixed offsets from the anchor. Each position was classified as (i)
502 conserved, exact match to the M17 LAP consensus; (ii) compatible, a residue with
503 structural or mutagenesis evidence for metal coordination in M17 LAPs [13,47,68]; or
504 (iii) non-functional (Table S2).
505
506 Zinc coordination sites were also assessed using mmCIF results from modelling
507 complexes of each protein (after removing signal peptide sequences identified in
508 *Phobius*) with two Zn^{2+} ions using Alphafold3 [69]. Zinc coordination geometry was
509 assessed with ChimeraX v.1.11.1 [70] in headless mode using a custom Python
510 script. Atoms within 3.5Å of each zinc ion were identified and contacts classified as
511 canonical if they involved the sidechain nitrogen of lysine (NZ) or carboxylate
512 oxygens of aspartate (OD1/OD2) or glutamate (OE1/OE2) consistent with the
513 experimentally derived coordination chemistry of M17 leucyl aminopeptidases (PDB:
514 1LAP) [47]. Where both carboxylate oxygens of a single Asp or Glu residue fell
515 within the distance cutoff (bidentate geometry) only the closer oxygen was retained.
516
517 Each zinc ion was classified as occupying either the exchangeable site (Zn1,
518 characterised by lysine coordination) or the tight site (Zn2, characterised by two
519 aspartate contacts and absence of lysine), based on coordination chemistry of LAP3
520 (P00727). We applied a scoring rubric (from 1 to 10) for coordination environments
521 of Zn1 and Zn2 that incorporated presence of canonical coordinating residues (6
522 points), coordination number (2 points), bond distance (1 point, penalised for
523 contacts > 3.2Å) and mean pLDDT confidence at the binding site (≥ 70, 1 point).
524 Structures were classified as likely active (score ≥ 6), uncertain (score 3-5), or likely
525 inactive (score < 3).
526 Distance thresholds were set to account for the systematic overestimation of metal-
527 ligand bond lengths in AF3 predictions relative to experimental crystal structures.

ACKNOWLEDGEMENTS

We would like to thank Christin Froschauer and Susanne Broschk for help in the lab and Caitlin McDonough-Goldstein for comments on an earlier version of the manuscript. We thank members of the ASU Biosciences Core Mass Spectrometry facility, Kyle Tucker, Sydney Canning and Yishai Gilron for the acquisition of mass spectra and downstream computational analyses. Yan Ge and the ZIH team at TU Dresden provided valuable assistance in implementing AlphaFold3. This work was funded by DFG grants to OO 521/4-1 and KR 1666/4-1. We are grateful for high-performance computing at the NHR Center of TU Dresden, jointly supported by the Federal Ministry of Education and Research and the state governments participating in the NHR (www.nhr-verein.de/unsere-partner).

539 **REFERENCES**

540 1. Snook RR. Sperm in competition: not playing by the numbers. Trends Ecol Evol.
541 2005;20: 46–53. doi:10.1016/j.tree.2004.10.011

542 2. Perry JC, Sirot L, Wigby S. The seminal symphony: how to compose an
543 ejaculate. Trends Ecol Evol. 2013;28: 414–422. doi:10.1016/j.tree.2013.03.005

544 3. Dorus S, Karr TL. Sperm proteomics and genomics. 1st ed. Sperm biology: an
545 evolutionary perspective. 1st ed. Burlington, MA: Academic press; 2009. pp.
546 435–469.

547 4. Wigby S, Brown NC, Allen SE, Misra S, Sitnik JL, Sepil I, et al. The *Drosophila*
548 seminal proteome and its role in postcopulatory sexual selection. Philos Trans R
549 Soc B Biol Sci. 2020;375: 20200072. doi:10.1098/rstb.2020.0072

550 5. Karr TL. Reproductive proteomics comes of age. Mol Cell Proteomics. 2019;18:
551 S1–S5. doi:10.1074/mcp.E119.001418

552 6. Pini T, Nixon B, Karr TL, Teperino R, Sanz-Moreno A, Silva-Buttkus P da, et al.
553 Towards a kingdom of reproductive life - the core sperm proteome.
554 Reproduction. 2025;169: e250105. doi:10.1530/REP-25-0105

555 7. Garlovsky MD, Sandler JA, Karr TL. Functional Diversity and Evolution of the
556 Drosophila Sperm Proteome. Mol Cell Proteomics. 2022;21.
557 doi:10.1016/j.mcpro.2022.100281

558 8. McCullough EL, Whittington E, Singh A, Pitnick S, Wolfner MF, Dorus S. The life
559 history of *Drosophila* sperm involves molecular continuity between male and
560 female reproductive tracts. Proc Natl Acad Sci. 2022;119: e2119899119.
561 doi:10.1073/pnas.2119899119

562 9. Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins. Nat Rev
563 Genet. 2002;3: 137–144. doi:10.1038/nrg733

564 10. Dapper AL, Wade MJ. Relaxed selection and the rapid evolution of reproductive
565 genes. Trends Genet. 2020;36: 640–649. doi:10.1016/j.tig.2020.06.014

566 11. Patlar B, Jayaswal V, Ranz JM, Civetta A. Non-adaptive molecular evolution of
567 seminal fluid proteins in Drosophila. Evolution. 2021;75: 2102–2113.
568 doi:10.1111/evo.14297

569 12. Bayram H, Claydon AJ, Brownridge PJ, Hurst JL, Mileham A, Stockley P, et al.
570 Cross-species proteomics in analysis of mammalian sperm proteins. J
571 Proteomics. 2016;135: 38–50. doi:10.1016/j.jprot.2015.12.027

572 13. Dorus S, Wilkin EC, Karr TL. Expansion and functional diversification of a leucyl
573 aminopeptidase family that encodes the major protein constituents of
574 Drosophila sperm. BMC Genomics. 2011;12: 177. doi:10.1186/1471-2164-12-
575 177

576 14. Whittington E, Zhao Q, Borziak K, Walters JR, Dorus S. Characterisation of the
577 Manduca sexta sperm proteome: Genetic novelty underlying sperm composition

578 in Lepidoptera. Insect Biochem Mol Biol. 2015;62: 183–193.
579 doi:10.1016/j.ibmb.2015.02.011

580 15. Whittington E, Forsythe D, Borziak K, Karr TL, Walters JR, Dorus S. Contrasting
581 patterns of evolutionary constraint and novelty revealed by comparative sperm
582 proteomic analysis in Lepidoptera. BMC Genomics. 2017;18: 931.
583 doi:10.1186/s12864-017-4293-2

584 16. Bayram H, Sayadi A, Immonen E, Arnqvist G. Identification of novel ejaculate
585 proteins in a seed beetle and division of labour across male accessory
586 reproductive glands. Insect Biochem Mol Biol. 2019;104: 50–57.
587 doi:10.1016/j.ibmb.2018.12.002

588 17. Degner EC, Ahmed-Braimah YH, Borziak K, Wolfner MF, Harrington LC, Dorus
589 S. Proteins, Transcripts, and Genetic Architecture of Seminal Fluid and Sperm
590 in the Mosquito Aedes aegypti. Mol Cell Proteomics. 2019;18: S6–S22.
591 doi:10.1074/mcp.RA118.001067

592 18. Laurinyecz B, Vedelek V, Kovács AL, Szilasi K, Lipinszki Z, Slezák C, et al.
593 Sperm-Leucylaminopeptidases are required for male fertility as structural
594 components of mitochondrial paracrystalline material in Drosophila
595 melanogaster sperm. PLOS Genet. 2019;15: e1007987.
596 doi:10.1371/journal.pgen.1007987

597 19. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, et al.
598 TimeTree 5: An Expanded Resource for Species Divergence Times. Mol Biol
599 Evol. 2022;39: msac174. doi:10.1093/molbev/msac174

600 20. Gomez RA, Dallai R, Sims-West DJ, Mercati D, Sinka R, Ahmed-Braimah Y, et
601 al. Proteomic diversification of spermatostyles among six species of whirligig
602 beetles. Mol Reprod Dev. 2024;91: e23745. doi:10.1002/mrd.23745

603 21. LaFlamme BA, Avila FW, Michalski K, Wolfner MF. A *Drosophila* protease
604 cascade member, Seminal Metalloprotease-1, is activated stepwise by male
605 factors and requires female factors for full activity. Genetics. 2014;196: 1117–
606 1129. doi:10.1534/genetics.113.160101

607 22. Avila FW, Mattei AL, Wolfner MF. Sex peptide receptor is required for the
608 release of stored sperm by mated *Drosophila melanogaster* females. J Insect
609 Physiol. 2015;76: 1–6. doi:10.1016/j.jinsphys.2015.03.006

610 23. Singh A, Buehner NA, Lin H, Baranowski KJ, Findlay GD, Wolfner MF. Long-
611 term interaction between *Drosophila* sperm and sex peptide is mediated by
612 other seminal proteins that bind only transiently to sperm. Insect Biochem Mol
613 Biol. 2018;102: 43–51. doi:10.1016/j.ibmb.2018.09.004

614 24. Plakke MS, Walker JL, Lombardo JB, Goetz BJ, Pacella GN, Durrant JD, et al.
615 Characterization of female reproductive proteases in a butterfly from functional
616 and evolutionary perspectives. Physiol Biochem Zool. 2019;92: 579–590.
617 doi:10.1086/705722

25. Wainwright SM, Hopkins BR, Mendes CC, Sekar A, Kroeger B, Hellberg JEEU, et al. *Drosophila* Sex Peptide controls the assembly of lipid microcarriers in seminal fluid. Proc Natl Acad Sci. 2021;118: e2019622118. doi:10.1073/pnas.2019622118

26. Zuk M, Garcia-Gonzalez F, Herberstein ME, Simmons LW. Model systems, taxonomic bias, and sexual selection: beyond *Drosophila*. Annu Rev Entomol. 2014;59: 321–338. doi:10.1146/annurev-ento-011613-162014

27. Vibranovski MD, Lopes HF, Karr TL, Long M. Stage-specific expression profiling of Drosophila spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. PLOS Genet. 2009;5: e1000731. doi:10.1371/journal.pgen.1000731

28. Slack HD. Chromosomes of Cimex. Nature. 1939;143: 78–78. doi:10.1038/143078a0

29. Ueshima N. Supernumerary chromosomes in the human bed bug, Cimex lectularius Linn. (Cimicidae:Hemiptera). Chromosoma. 1967;20: 311–331. doi:10.1007/BF00326188

30. Grozeva S, Kuznetsova VG, Anokhin B. Bed bug cytogenetics: karyotype, sex chromosome system, FISH mapping of 18S rDNA, and male meiosis in Cimex lectularius Linnaeus, 1758 (Heteroptera: Cimicidae). Comp Cytogenet. 2010;4: 151–160. doi:10.3897/compcytogen.v4i2.36

31. Siva-Jothy MT. Trauma, disease and collateral damage: conflict in cimicids. Philos Trans R Soc B Biol Sci. 2006;361: 269–275. doi:10.1098/rstb.2005.1789

32. Roth S, Siva-Jothy MT, Balvín O, Morrow EH, Willassen E, Reinhardt K. The evolution of female-biased genital diversity in bedbugs (Cimicidae). Evolution. 2024;78: 329–341. doi:10.1093/evolut/qpad211

33. Morrow EH, Arnqvist G. Costly traumatic insemination and a female counter-adaptation in bed bugs. Proc R Soc Lond B Biol Sci. 2003;270: 2377–2381. doi:10.1098/rspb.2003.2514

34. Reinhardt K, Naylor R, Siva–Jothy MT. Reducing a cost of traumatic insemination: female bedbugs evolve a unique organ. Proc R Soc Lond B Biol Sci. 2003;270: 2371–2375. doi:10.1098/rspb.2003.2515

35. Usinger RL. Monograph of Cimicidae (Hemiptera-Heteroptera). 1966. Available: https://www.jstor.org/stable/3276661?origin=crossref

36. Martens BM, McDonough-Goldstein CE, Otti O, Broschk S, Kullmann L, Reinhardt K, et al. Novel female reproductive organ differentiates postmating transcriptional response to insemination versus arrival of sperm in bedbugs. bioRxiv; 2026. p. 2026.03.17.707905. doi:10.64898/2026.03.17.707905

37. Reinhardt K, Wong CH, Georgiou AS. Detection of seminal fluid proteins in the bed bug, Cimex lectularius, using two-dimensional gel electrophoresis and

657 mass spectrometry. Parasitology. 2009;136: 283–292.
658 doi:10.1017/S0031182008005362

659 38. Law STS, Nong W, Li C, Chong TK, Yip HY, Swale T, et al. Genome of tropical
660 bed bug Cimex hemipterus (Cimicidae, Hemiptera) reveals tetraspanin
661 expanded in bed bug ancestor. Insect Sci. 2025;32: 42–54. doi:10.1111/1744-
662 7917.13388

663 39. Miles LS, Adams R, Francioli YZ, Card DC, Castoe TA, Booth W. A
664 chromosome-level reference genome for the common bed bug, Cimex
665 lectularius, with identification of sex chromosomes. J Hered. 2025;116: 382–
666 388. doi:10.1093/jhered/esae071

667 40. Kohler D, Staniak M, Tsai T-H, Huang T, Shulman N, Bernhardt OM, et al.
668 MSstats Version 4.0: Statistical Analyses of Quantitative Mass Spectrometry-
669 Based Proteomic Experiments with Chromatography-Based Quantification at
670 Scale. J Proteome Res. 2023;22: 1466–1482.
671 doi:10.1021/acs.jproteome.2c00834

672 41. Thongprem P, Evison SEF, Hurst GDD, Otti O. Transmission, Tropism, and
673 Biological Impacts of Torix Rickettsia in the Common Bed Bug Cimex lectularius
674 (Hemiptera: Cimicidae). Front Microbiol. 2020;11.
675 doi:10.3389/fmicb.2020.608763

676 42. Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T. Wolbachia as a
677 bacteriocyte-associated nutritional mutualist. Proc Natl Acad Sci. 2010;107:
678 769–774. doi:10.1073/pnas.0911476107

679 43. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos
680 KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein
681 language models. Nat Biotechnol. 2022;40: 1023–1025. doi:10.1038/s41587-
682 021-01156-3

683 44. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al.
684 BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421.
685 doi:10.1186/1471-2105-10-421

686 45. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, et al.
687 Phylogenomics resolves the timing and pattern of insect evolution. Science.
688 2014;346: 763–767. doi:10.1126/science.1257570

689 46. Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, et al.
690 The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024.
691 Nucleic Acids Res. 2024;52: W521–W525. doi:10.1093/nar/gkae241

692 47. Burley SK, David PR, Sweet RM, Taylor A, Lipscomb WN. Structure
693 determination and refinement of bovine lens leucine aminopeptidase and its
694 complex with bestatin. J Mol Biol. 1992;224: 113–140. doi:10.1016/0022-
695 2836(92)90580-D

696 48. Maric S, Donnelly SM, Robinson MW, Skinner-Adams T, Trenholme KR,
697 Gardiner DL, et al. The M17 Leucine Aminopeptidase of the Malaria Parasite

698 Plasmodium falciparum: Importance of Active Site Metal Ions in the Binding of
699 Substrates and Inhibitors. Biochemistry. 2009;48: 5435–5439.
700 doi:10.1021/bi9003638

701 49. Poggio MG, Iorio OD, Turienzo P, Papeschi AG, Bressa MJ. Heterochromatin
702 characterization and ribosomal gene location in two monotypic genera of
703 bloodsucker bugs (Cimicidae, Heteroptera) with holokinetic chromosomes and
704 achiasmatic male meiosis. Bull Entomol Res. 2014;104: 788–793.
705 doi:10.1017/S0007485314000650

706 50. Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. Novel genes derived
707 from noncoding DNA in Drosophila melanogaster are frequently X-linked and
708 exhibit testis-biased expression. Proc Natl Acad Sci. 2006;103: 9935–9939.
709 doi:10.1073/pnas.0509809103

710 51. Stutt AD, Siva-Jothy MT. Traumatic insemination and sexual conflict in the bed
711 bug *Cimex lectularius*. Proc Natl Acad Sci. 2001;98: 5683–5687.
712 doi:10.1073/pnas.101440698

713 52. Reinhardt K, Naylor R, Siva-Jothy MT. Male mating rate is constrained by
714 seminal fluid availability in bedbugs, *Cimex lectularius*. PLOS ONE. 2011;6:
715 e22082. doi:10.1371/journal.pone.0022082

716 53. Sirot LK, Wong A, Chapman T, Wolfner MF. Sexual conflict and seminal fluid
717 proteins: a dynamic landscape of sexual interactions. Cold Spring Harb
718 Perspect Biol. 2015;7: a017533. doi:10.1101/cshperspect.a017533

719 54. Misra S, Wolfner MF. *Drosophila* seminal sex peptide associates with rival as
720 well as own sperm, providing SP function in polyandrous females. eLife. 2020;9.
721 doi:10.7554/eLife.58322

722 55. Breland OP, Simmons E. Preliminary Studies Of The Spermatozoa And The
723 Male Reproductive System Of Some Whirligig Beetles Coleoptera Gyrinidae.
724 Entomol News. 1970;81: 101–110.

725 56. Doggett SL, Dwyer DE, Peñas PF, Russell RC. Bed Bugs: Clinical Relevance
726 and Control Options. Clin Microbiol Rev. 2012;25: 164–192.
727 doi:10.1128/cmr.05015-11

728 57. Reinhardt K, Kempke D, Naylor RA, Siva-Jothy MT. Sensitivity to bites by the
729 bedbug, Cimex lectularius. Med Vet Entomol. 2009;23: 163–166.
730 doi:10.1111/j.1365-2915.2008.00793.x

731 58. Reinhardt K, Siva-Jothy MT. Biology of the Bed Bugs (Cimicidae). Annu Rev
732 Entomol. 2007;52: 351–374. doi:10.1146/annurev.ento.52.040306.133913

733 59. Käll L, Krogh A, Sonnhammer ELL. A Combined Transmembrane Topology and
734 Signal Peptide Prediction Method. J Mol Biol. 2004;338: 1027–1036.
735 doi:10.1016/j.jmb.2004.03.016

60. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16: 157. doi:10.1186/s13059-015-0721-2

61. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20: 238. doi:10.1186/s13059-019-1832-y

62. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 2007;24: 1586–1591. doi:10.1093/molbev/msm088

63. Benoit JB, Adelman ZN, Reinhardt K, Dolan A, Poelchau M, Jennings EC, et al. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. Nat Commun. 2016;7: 10165. doi:10.1038/ncomms10165

64. Rahnenfuhrer AA. topGO: Enrichment Analysis for Gene Ontology. 2022.

65. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024. Available: http://www.R-project.org

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

67. Edgar RC. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. Nat Commun. 2022;13: 6968. doi:10.1038/s41467-022-34630-w

68. Gu Y-Q, Walling LL. Identification of residues critical for activity of the wound-induced leucine aminopeptidase (LAP-A) of tomato. Eur J Biochem. 2002;269: 1630–1640. doi:10.1046/j.1432-1327.2002.02795.x

69. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630: 493–500. doi:10.1038/s41586-024-07487-w

70. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF ChimeraX: Tools for structure building and analysis. Protein Sci. 2023;32: e4792. doi:10.1002/pro.4792

71. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022;50: D543–D552. doi:10.1093/nar/gkab1038