

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

|                          |                                                                                                                                                                                                                                                                                                |
|--------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| n/a                      | Confirmed                                                                                                                                                                                                                                                                                      |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement                                                                                                                               |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly                                                                                                                                    |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>                                                               |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested                                                                                                                                                                                                                     |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons                                                                                                                                        |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted<br><i>Give P values as exact values whenever suitable.</i>                     |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings                                                                                                                                                           |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes                                                                                                                                     |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated                                                                                                                                               |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

|                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data collection | Microsoft Excel was used to manipulate basic text files for managing the comparative trait data. Online search platforms including Google Scholar, Semantics Scholar, and Web of Science were used to search the sperm literature. The tree's topology was constructed manually using Mesquite v. 3.61 based on a large set of phylogenetic research articles. Branch length estimates were derived from well-vetted fossils and a node-based dating analysis using BEAST v. 2.6.3. The R package ape v. 5.8.1 was used to generate an initial starting tree for BEAST.                                                                         |
| Data analysis   | All comparative analyses were conducted in the statistical software R v. 4.3.2. Reconstructions of traits across the phylogeny and associated phylogenetic visualizations were performed using R phytools v. 2.4.19. Evolutionary models were fit to the data using R OUwie v. 2.10. Bayesian estimation of character state transitions were conducted using the standalone software BayesTraits v. 4.1.2. Raw spectral data were analyzed using PEAKS Studio X software (Bioinformatics Solutions Inc). Finally, a simple Unix script using base Unix commands was used to generate null distributions for the expected overlap among species. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Specimens examined for novel sperm morphological data are deposited in the lead author's research collection and are available for study upon request. The phenotypic dataset, supertree of 621 arthropods, and associated references are provided online as separate files on DataDryad under DOI: <https://doi.org/10.5061/dryad.79cnp5jbm>. Newly obtained raw reads are archived with the NCBI Short Read Archive (BioProject Accession PRJNA1270706), and novel raw spectra and Mz files are available from the ProteomeXchange database (accession identifier PXD064531). We also analyzed the following published datasets: (1) whirligig beetle RNAseq data archived with the NCBI SRA under PRJNA1063663 [<https://www.ncbi.nlm.nih.gov/bioproject>] and (2) whirligig beetle proteomic data archived with the ProteomeXchange under PXD048928 [<https://proteomecentral.proteomexchange.org>].

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender This information has not been collected

Reporting on race, ethnicity, or other socially relevant groupings See above

Population characteristics See above

Recruitment See above

Ethics oversight See above

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description We report on the results of a comparative evolutionary investigation of sperm conjugation. Our study has two main sources of data and results that come from pairing (1) phylogenetic reconstructions of conjugation for 621 species of arthropods with (2) a comparative proteomic investigation of two independent origins of a unique ejaculate structure.

Research sample The comparative dataset is derived from our mining the sperm ultrastructure literature for 621 arthropods that were chosen because they are adequately studied for sperm morphology, their phylogenetic position is supported by modern phylogenomic or multigene studies, and their inclusion serves to aid in the reconstruction of character state transitions in conjugation. Selection of the four species chosen for novel proteomic data was informed by the phylogenetic results, which revealed that the true bugs (Order Hemiptera) and beetles (Order Coleoptera) both have independently derived spermatostyles, a unique example of conjugation that features a distinct structure. These four bugs were chosen because they represent different major lineages (families) of true bugs and they show variation in spermatostyle form among species.

Sampling strategy Species chosen for comparative analysis were scored as to the whether conjugation was present or absent; other categorical sperm traits such as conjugate type were also recorded. Sample sizes for the broader comparative study are not reported as they vary per species based upon the original literature. For comparative proteomics of bug and beetle spermatostyles, we solubilized spermatostyles from adult male seminal vesicles and prepared two biological replicates per species. Our high-resolution spectral data shows very little variation between replicates suggesting that two replicates are sufficient for an initial characterization of the proteomes of spermatostyles in bugs beetles.

Data collection Mining the literature for sperm morphological data involved a comprehensive search of terms related to sperm conjugation for diverse arthropod groups on online search platforms including Google Scholar, Web of Science, and Semantics Scholar. RA Gomez recorded the data into a basic spreadsheet including the original citation, and RA Gomez and S Pitnick curated the trait data. The proteomic data were generated using standard proteomic protocols. The solubilized spermatostyle samples were prepared by RA Gomez and shipped to the Cambridge Proteomics Facility for gel electrophoresis, slicing, and analysis with tandem mass

spectrometry.

Timing and spatial scale

The comparative trait data were collected over a period of months from June-November 2024. True bug samples for proteomics were collected between August-October 2022 and were all sexually mature. Proteomic samples were prepared for mass spectrometry in November 2022 and January 2023.

Data exclusions

No data were excluded from the analyses

Reproducibility

Our phylogenetic results are based on a large dataset that includes more than 300 original research articles, which collectively support the phylogenetic distribution of conjugation we model. Our proteomic results are based on robust identifications and include two replicates per species. Our use of an antibody to visualize a sperm protein in samples of true bugs was repeated several times, and the signal was also confirmed with a negative control.

Randomization

We attempted to be comprehensive in our sampling of the variation in conjugation across arthropods, and we did not attempt to randomize species sampling for this purpose. The samples we dissected for proteomics were collected opportunistically and males were assigned to replicates arbitrarily.

Blinding

Blinding is not relevant

Did the study involve field work?



Yes



No

## Field work, collection and transport

Field conditions

Late summer/early Fall days with little rain

Location

Syracuse, NY public lands and University of Syracuse campus

Access &amp; import/export

All samples were collected legally, and none of the studied species are threatened or endangered.

Disturbance

not relevant

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a                                 | Involved in the study                                           |
|-------------------------------------|-----------------------------------------------------------------|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Antibodies                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology          |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                          |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants                                 |

### Methods

| n/a                                 | Involved in the study                           |
|-------------------------------------|-------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

## Antibodies

Antibodies used

custom polyclonal mouse antibody for Sperm-Leucyl Aminopeptidase (S-LAP) 1 in the fruit fly

Validation

We used the antibody in immunohistochemistry of S-LAPs in true bug sperm preparations. The antibody has been previously used on sperm preps of *Drosophila melanogaster* as described in the original article.

Laurinyecz, B., Vedelek, V., Kovács, A. L., Szilasi, K., Lipinski, Z., Slezák, C., ... & Sinka, R. (2019). Sperm-Leucylaminopeptidases are required for male fertility as structural components of mitochondrial paracrystalline material in *Drosophila melanogaster* sperm. *PLoS genetics*, 15(2), e1007987.

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals

The study did not involve laboratory animals

|                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Wild animals            | We collected novel data for 10 species of insects, which necessitated live samples of adults for sperm morphology and spermatostyle proteomics. The species sampled for live specimens include the 4 true bugs sampled for proteomics ( <i>Lepyronia quadrangularis</i> , <i>Philaenus psumarius</i> , <i>Neotibicen lyricen</i> , and <i>Cuernia striata</i> ), four additional bugs ( <i>Lycorma delicatula</i> , <i>Gyponana</i> sp., and <i>Acanalonia bivittata</i> ), and two beetles ( <i>Peltodytes edentulus</i> and <i>Amblycheila baroni</i> ). Insects were transported to lab and either dissected immediately or frozen for future study. |
| Reporting on sex        | Data on spermatostyle proteomes are all derived from spermatostyles isolated from wild-caught mature male true bugs. Several female specimens of various species were also studied for basic sperm morphology and for the spatiotemporal study of spermatostyles in female sperm-storage.                                                                                                                                                                                                                                                                                                                                                               |
| Field-collected samples | Freezing at -20C within labeled empty vials for future proteomics or basic morphology                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Ethics oversight        | No ethical approval or guidance was required as we are working with insects that are not protected                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Plants

|                       |              |
|-----------------------|--------------|
| Seed stocks           | not relevant |
| Novel plant genotypes | not relevant |
| Authentication        | not relevant |