nature portfolio

Corresponding author(s): Gang Li, Kun Guo, Yandong Ren

Last updated by author(s): Feb 21, 2025

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	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	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)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), 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	No softwares were used for Data collect1on
Data analysis	All software used to analyze the data are described in detail in the materials and methods section, These include: fastp (v0.23.2), GCE (vl.0.0), Canu (v2.2), BlobTools (vl.1.1), purge_dups (vl.2.5), NextPolish (vl.4.0), Juicer (vl.6.0), 3D-DNA (v180922), juicecebox (vl.9.9), BUSCO (v5.3.2), RepeatModeler (v2.0.l), ReannTE_MergeFasta.pl (https://github.com/4ur-eliek/ReannTE), RepeatMasker (v4.l.2-pl), GeMoMa (vl.8), MMseq2 (v13.45111), tblastn (v2.12.0), genewise (v2-4-1), HISAT2 (v2.2.l), stringTie (v2.2.l), Augustus (v3.4.0), GlimmerHMM (v3.0.4), SNAP (v2017-03-0l), EvidenceModeler (vl.1.1), Orthofinder (v2.5.4), PRANK (v170427), TrimAL (vl.4rev15), RAxML (v8.2.12), R8s (vl.81), blastp (v2.13.0), MCScanX (vl.l), SynVison (https://synvisio.github.io/#/), WGDI (v0.6.5), LAST (v1420), DESCHRAMBLER (https://github.com/jkimlab/DESCHRAMBLER), ANGES (vl.01), BEDtools (v2.30.0), parseRM.pl (https://github.com/4ureliek/ParsingRepeatMasker-Outputs), BWA (v0.7.17-r1188), HiCExplorer (v3.6), hicFindTADs (v3.7.2), DESeq2 (vl.32.0), IGV (v2.12.3), CAFE (v4.2), clusterProfiler (v4.0.5), PAML (v4.5), BITACORA, MUSCLE v5.1, iqtree v2.3.6, meme v5.5.7, pfam_scan.py v1.0, TBtools-II v2.144, PyMOL, v3.1.3, SWISS-MODEL Additional custom-made scripts are available at https://github.com/Mrhuangc/aphids_data_analysis

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

The Whole Genome Shotgun project has been deposited at GenBank under the accession JBEIVQOOOOOOOOO. Raw sequences data generated for this study are available at the NCBI Short Read Archive under the Bioproject accession number PRJNA1061761.

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	No human participants or human data were involved in this study.
Reporting on race, ethnicity, or other socially relevant groupings	No human participants or human data were involved in this study.
Population characteristics	No human participants or human data were involved in this study.
Recruitment	No human participants or human data were involved in this study.
Ethics oversight	No human participants or human data were involved in this study.

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

Ecological, evolutionary & environmental sciences study design

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

Study description	We used PacBio and Hi-C data to assemble and annotate a chromosome-level reference genome for the celery aphid (Semiaph1s heraclei), and applied the same methodology to re-annotate 8 chromosome-level aphid genomes. These were then compared with the genomes of additional 12 insects to obtain insights into the chromosomal evolution of aphids. Additionally, we collected RNAseq data from both female and male celery aphid to analyze the expression patterns of the X chromosome. By generating a chromosomal-level reference genome assembly of the celery aphid (Semiaphis heraclei) and conducting comparative genomic analysis, we revealed that chromosomal evolution rates vary among aphid lineages and positively correlate with species diversity. Throughout the evolutionary history of aphids, their X chromosomes have undergone more frequent intra-chromosomal recombination, whereas autosomes have exhibited an accelerated rate of inter-chromosomal recombination. Moreover, considering both inter- and intra-chromosomal rearrangements, the increased autosomal rearrangement rates may be common across the entire Aphidomorpha lineage. Besides the observed potential driving forces of the expansion of DNA and SINE-type transposable elements, the loss and duplication of genes functionally associated with karyotypic instability (such as RI FI, BROS, DMCl, and TERT genes) may also play crucial roles in aphid chromosomal evolution. Additionally, our analysis revealed that the mutation and expansion of detoxification gene families in S. heraclei may be a key factor in adapting to host plant chemical defenses.
Research sample	The parthenogenetic wingless adult and nymphal aphids of Semiaphis heraclei were collected in May 2019 from a planting base of Lonicera japonica (variety: 'Siji jinyinhua') located in Zhengcheng Town, Pingyi County, Shandong Province, China (35°16′0″N, 117°38′56″E). In the laboratory, the aphids were reared on cuttings of Lonicera japonica for over 10 generations under controlled conditions (temperature: 25±1°C, relative humidity: 70% ± 5%, light cycle: 14 h light : 10 h dark). The specimens are stored at the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences.
Sampling strategy	For DNA sequencing, we cultured a parthenogenetic aphid for at least ten generations, and during the same period, selected 100 aphids of various developmental stages from the population and placed them into cryovials, which were then rapidly frozen in liquid nitrogen. The samples were then sent to a company for PacBio long-read sequencing and short-read sequencing. For RNA sequencing, We selected 20 parthenogenetic aphids from the above-mentioned Parthenohynogenesis line and quickly froze them in cryovials, which were used to assist in gene structure annotation. Subsequently, we collected wingless sexual females and winged

	males from the same experimental field at the Institute of Medicinal Plant Development, with three biological replicates for each sex, each containing 20 adult insects. All samples were collected during the same time period to ensure consistency, and these samples were used for gene expression analysis.
Data collection	The frozen samples were sent to Biomarker Technologies (Beijing, China) for library preparation and sequencing. DNA quantified through 0.75% agarose gel electrophoresis, Nanodrop spectrophotometry (Thermo Fisher), and Qubit 3.0 fluorometry (Invitrogen) PacBio long reads were generated from sequencing on the PacBio Sequel II platform, while Hi-C reads, Illumina short reads and transcriptome data were produced from sequencing on the Illumina NovaSeq 6000 platform
Timing and spatial scale	These celery aphid samples for sequencing (Pacbio, HiC, Illumina, RNA-seq) were collected in May 2019 in Zhengcheng Town, Pingyi County, Shandong Province, China (35°16′0″N, 117°38′56″E). Samples of wingless sexual females and winged males were collected to sequence in December 2022 in Beijing, China.
Data exclusions	No data were excluded.
Reproducibility	We collected samples of three wingless sexual females and three winged males during the same time period, each sample consisting of 20 adult insects. All attempts to repeat the experiments were successful
Randomization	The samples for sequencing were randomly obtained from large colony of celery aphid
Blinding	No blinding was applied in this study, this technique does not apply to our study

Did the study involve field work? ☐ Yes ☒ No

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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

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	celery aphid
Wild animals	none
Reporting on sex	For gene expression analysis, we collected samples of three wingless sexual females and three winged males, and the information of sex and sequence data are listed in Supplementary Table 31.
Field-collected samples	This study did not involve samples collected from the field.
Ethics oversight	No ethical approval or guidance was required.

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

Plants

Seed stocks	No plants sample or data were involved in this study.
Novel plant genotypes	No plants sample or data were involved in this study.
Authentication	No plants sample or data were involved in this study.