nature portfolio

Corresponding author(s): Satoshi Miyazaki

Last updated by author(s): Jun 17, 2026

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 custom software was used for data collection.
Data analysis	Analyses were performed with publicly available software (referred in manuscript).

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

Raw sequencing reads generated in this study (ddRAD-seq, de novo genome sequencing, and RNA-seq) have been deposited in the DDBJ Sequence Read Archive under BioProject accession number [PRJDBXXXXX / PRJNAXXXXX] (available upon publication). All other datasets supporting the findings of this study, (e.g., GWAS-SNP profiles, genome sequencing summaries, results of outlier and RNA-seq analysis) are provided within Supplementary Information files. Any remaining source data are available from the corresponding author upon reasonable request.

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	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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	This study investigates the genetic basis and evolutionary history of wing polymorphism (alate vs. ergatoid queens) in the ant Myrmecina nipponica. We integrated de novo genome assembly, population genomics via ddRAD-seq and genome-wide association study (GWAS), linkage disequilibrium analysis, and transcriptomic analysis (RNA-seq) to identify a putative supergene associated with queen phenotypes and characterize its structural features.
Research sample	The research samples consisted of queens and a male of the ant Myrmecina nipponica. Specifically, 96 queens (39 alate and 57 ergatoid) from four distinct natural populations (Chitose, Mt. Kurai, Mt. Takanawa, and Toyama) were used for ddRAD-seq and subsequent population genomics. For RNA-seq, 23 individuals (11 pupae and 12 newly eclosed adults) were obtained from laboratory-maintained colonies originally collected from the Chitose population. For de novo genome sequencing, three individuals (one male, one alate queen, and one ergatoid queen) were used. All samples originated from colonies collected from the natural populations and were maintained under laboratory conditions prior to nucleic acid extraction.
Sampling strategy	Natural ant colonies were sampled from multiple geographic locations in Japan. No statistical methods were used to predetermine sample sizes. Sample sizes (n = 96 for ddRAD-seq and n = 23 for RNA-seq) were chosen based on the natural availability of colonies during the field-sampling season. Consistent with standards in ant evolutionary genomics, these sample sizes were demonstrated to be sufficient for population structure inference, GWAS, and differential expression analyses, as evidenced by the identification of clear genetic associations and robust statistical power.
Data collection	Biological specimens and phenotype data (alate vs. ergatoid) were collected and recorded by S.M. Genomic DNA and total RNA extractions were performed by S.M. For ddRAD-seq, sequencing libraries were generated and sequenced on an Illumina platform by K.Y. and S.S. Genome sequencing (Illumina platform) and RNA-seq (DNBSeq platform) were outsourced to commercial providers. Bioinformatic raw data processing was performed by K.Y. and S.S., while downstream statistical and genomic analyses were performed by S.M. and Y.H., using the standard pipelines described in the Methods.
Timing and spatial scale	Field sampling was conducted between July and September 2017 across four geographic regions in Japan. The spatial scale spanned four distinct locations across Japan to capture regional genetic diversity: Chitose (northern lowland), Mt. Kurai (central highland), Toyama (central lowland), and Mt. Takanawa (southern highland).
Data exclusions	No data or samples were excluded from the analyses.
Reproducibility	All attempts at replication were successful. To ensure the reproducibility and robustness of our findings, statistical and bioinformatic analyses were validated using independent biological replicates (n = 96 queens for ddRAD-seq and n = 23 ants for RNA-seq) across multiple natural populations.
Randomization	Randomization was not relevant to this study because it does not involve manipulative experimental interventions; instead, it is a field-based genomic and transcriptomic analysis of naturally occurring phenotypic variants in wild populations. Samples were grouped strictly based on their predetermined biological features: geographic origin, developmental stage (pupa vs. adult), and discrete wing phenotype (alate vs. ergatoid queens). To control for potential confounding effects of population structure and

geographic imbalance of wing phenotypes in the GWAS, the first two principal components (PC1 and PC2) from the genetic PCA were included as fixed covariates in the rrBLUP model, as described in methods.

Blinding

Blinding was not relevant or possible during data acquisition because the wing phenotypes (alate vs. ergatoid queens) are morphologically distinct and easily distinguishable upon collection. However, bioinformatic workflows and statistical criteria (e.g., FDR thresholds, quality filtering in PLINK and edgeR) were strictly predetermined and executed programmatically via command-line and R scripts, eliminating potential investigator bias during data analysis.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions

Field sampling was conducted during the active season of the ants (typically from June to September) to collect natural colonies for subsequent analyses.

Location

Colonies were collected from four distinct geographic locations across Japan (Figure 1) to capture genetic diversity among populations.
1. Chitose, Hokkaido (Northern lowland, Lat: 42°47'N, Lon: 141°27'E, Elevation: 200 m)
2. Mt. Kurai, Gifu (Central highland, Lat: 36°00'N, Lon: 137°13'E, Elevation: 1000-1200 m)
3. Toyama, Toyama (Central lowland, Lat: 36°42'N, Lon: 137°10'E, Elevation: 74 m)
4. Mt. Takanawa, Ehime (Southern highland, Lat: 33°56'N, Lon: 132°50'E, Elevation: 980 m)

Access & import/export

Field sampling was conducted in strict compliance with local and national regulations in Japan. Sampling at Mt. Kurai were carried out with permission from Gifu University. No special permits were required for sampling at the other populations. All samples remained within Japan: hence, no international import/export permits were applicable.

Disturbance

Disturbance to the natural habitat was kept to a minimum. During colony sampling, mosses, fern roots, and decaying logs were carefully replaced to restore the original microenvironment. Only a limited number of colonies were collected from each population to ensure the long-term sustainability of the local natural populations.

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

Plants

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A