# nature research

Corresponding author(s):   Guojie Zhang

# Reporting Summary

Nature Research 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 Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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 |
| ☐ | ☒ | An indication of 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<br>*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 statistics 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<br>*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 |
| ☐ | ☒ | Clearly defined error bars<br>*State explicitly what error bars represent (e.g. SD, SE, CI)* |

*Our web collection on statistics for biologists may be useful.*

## Software and code

Policy information about availability of computer code

| Data collection | *Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.* |
|---|---|
| Data analysis | We used GeMoMa (v1.4.2) to perform genome re-annotations,  Proteinortho (v5.16) to identify orthologous genes across ant species, Hisat2 (v2.0.5), StringTie (v1.3.3) and Salmon (v0.8.2) to perform RNAseq alignment and quantification, DESeq2 (R package; v1.16) to identify differentially expressed genes, GAGE (R package; v2.27.3) to do gene set enrichment, and SVA (R package; v3.26.0) to adjust for species or colony identify for gene expression.<br>For gene age determination, Singular Value Decomposition (SVD) and visualization, we wrote our codes with R and python. All codes used in this study are provided in Github: https://github.com/StanQiu/ant_brain_comparative |

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/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about [availability of data](availability_of_data)

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-seq data have been deposited under Bioproject accession number PRJNA427677. The Python and R scripts used to process the data are available on https://github.com/StanQiu/ant_brain_comparative

# Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf](nature.com/authors/policies/ReportingSummary-flat.pdf)

# Life sciences study design

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

| | |
|---|---|
| Sample size | We decided sample size based on the detection power of RNAseq (both https://doi.org/10.1186/s13059-016-0881-8 and our computational simulation result). With minimum of four replicates, we had over 90% detection power for genes with 2-fold expression changes. |
| Data exclusions | We excluded 7 samples from our analyses . For the three sets of RNA samples with ERCC-spike-in, Pearson correlations of spike-in reads (log transformed) across samples of < 0.95 were considered as outliers and removed. Similarities in caste-biased gene expression were also used to check data quality: For each ant species, mapped reads for the genome (log transformed) were compared among samples within the same caste, and samples different from other same-caste samples of the focal species (within-caste Pearson correlations < 0.9) were checked for potential biological or technical deviations, as suggested by the ENCODE RNAseq guideline [goo.gl/d9DvyV]. For example, we excluded potential outlier samples if they were collected from different locations (biological explanation) or were prepared with different experimental procedures (technical explanation). If no potential biological or technical explanations could be identified, samples were retained for downstream analysis. |
| Replication | Our study initially only included four ant species: A.echinatior, L.humile, M.pharaonis and S.invicta, and we found that samples of the same caste can be separated with PCA after normalizing for species identity. We then further included recent published brain transcriptome data from L.niger, and we identified the same pattern. Therefore, we believe our study is reproducible. |
| Randomization | Gynes (future queens) and workers samples were collected from four to five different colonies. To minimize/randomize batch effect of caste and colony identity in preparation, RNA extraction and cDNA library construction were done with the same procedure and within the same run for each species, and sequencing were done with multiplexed bar-coding. |
| Blinding | The investigators were not blinded to phenotype (caste), colony and species information during collection and analysis procedures. But clustering analysis were done with blinded phenotype information. |

# Reporting for specific materials, systems and methods

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☒ Unique biological materials |
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology |
| ☐ | ☒ Animals and other organisms |
| ☒ | ☐ Human research participants |

## Methods

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

# Unique biological materials

Policy information about [availability of materials](#)

| | |
|---|---|
| Obtaining unique materials | Ants are available in labs of Centre for Social evolution, University of Copenhagen or Biodiversity Research Center, Academia Sinica, Taoyuan, Taiwan. |

# Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

| | |
|---|---|
| Laboratory animals | Five colonies of A. echinatior were collected in Panama in 2004 (Ae263), 2014 (Ae704) and 2016 (Ae747, Ae764, Ae767) and reared at the Centre for Social Evolution, University of Copenhagen under a constant temperature of ca. 25ºC and ca. 70% humidity. Gynes and minor workers were collected within the fungus garden, whereas major workers were collected while they were foraging outside, always during daytime (10 am to 4 pm), and isolated in groups of 6-10 individuals of the same caste in small fluon-coated plastic boxes.<br><br>Five colonies of M. pharaonis (3rd Room X1/B, CS10, Donor 3rd, Donor BQ-, 3rd Room X1/A) were reared at the Centre for Social Evolution, University of Copenhagen under constant temperature of 27ºC and 50% humidity from a stock collected in 2008. Gynes were separated from males at the pupal stage and reared with 10–15 workers in fluon-coated petri-dishes to be collected within three days after they hatched as adults, whereas workers were collected directly from the source colonies when they were foraging. All collections were done during daytime (10 am to 4 pm) and adults were kept in groups of 6-10 of the same caste in small fluon-coated boxes.<br><br>Five colonies of L. humile were collected in Caldes d'Estrac, Spain (Catalan3b) and Castell d'Aro, Spain (Main 3a, Main 4a, Main 5b, Main 5d) in 2016 and were reared at the Centre for Social Evolution, University of Copenhagen under constant temperature of 27ºC and 50% humidity. Gyne isolation, worker ant sample collection, brain dissection, and storage proceeded in the same way as for Monomorium pharaonis.<br><br>Four monogynous (single queen) colonies were collected in Taoyuan, Taiwan, two in October 2012 and two in April 2014, and transferred to a fluon-coated plastic box in the lab. Approximately 120 gynes and 200 workers from each colony were randomly selected for dissection. For workers, an equal number of small and large workers were selected.<br><br>Ant sample collection and dissection procedures for L. niger, the clonal and queenless ant O. biroi, and the queenless ant D. quadriceps were comparable to our own procedures and described in original research papers. |
| Wild animals | *Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.* |
| Field-collected samples | *For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.* |