# Extended Data:

A picture containing umbrella, accessory

Description automatically generated

Extended data Fig. 1. Developmental transcriptomes in ants.

**a**, Transcriptomic developmental trajectories in *A. echinatior*, based on Spearman rank correlation similarity in gene expression across individual transcriptomes. Trajectories were constructed and visualized with a similar procedure as in **Fig. 1a**, except for *A. echinatior* having three worker sub-castes, which exhibited increasing transcriptomic divergence across pupal and adult stages.

Graphical user interface

Description automatically generated

Extended Data Fig. 2. Prediction of caste identities with BPA.

**a,** The *Backward Progressives Algorithm* (BPA) predicts caste identities in previous developmental stages, using non-validated transcriptomes at the target stage (*S*t) to construct PCAs and then projects these onto the subsequent stage (*S*t+1) where caste identities are known. The BPA then uses the known caste labels at *S*t+1 to identify PC axes that are associated with confirmed caste identity and uses linear discriminant analysis to train a predictive model, assuming that the PC axes at *S*t+1 are also associated with caste identities at *S*t as expected under developmental continuity. The trained model then predicts caste identities at *S*t, after which it assumes these predicted caste identities to be real and initiates a next round to predict caste identities at stage *S*t-1. This process continues until the prediction likelihoods at *S*t-n become too low to be informative.

**b**, BPA predicted the sex of sampled individuals among 1st instar (44 hour) larvae in *D. melanogaster* (left panel). While sex can be distinguished in 2nd instar (50 hour) larvae (right panel) via genotyping after simultaneous DNA and RNA extraction, the biomass of 1st instar larvae was too small to perform such simultaneous extractions. By examining the proportion of reads that mapped to the Y chromosome of *Drosophila*, we found that predicted males in 1st instar larvae had a significantly higher proportion of reads mapped to the Y chromosome, confirming our prediction. Individual samples are colored according to their predicted probability to be female or male in the 1st-instar and symbols were sized according to the number of reads mapped to Y chromosome per million reads (YPM).

**c**, Testing BPA on samples of developmental stages with known caste identities used individual *M. pharaonis* transcriptomes of individuals with distinct morphology. The table presents prediction accuracies for each target stage, calculated as the ratio of the number of correctly assigned individuals and the total number of individuals sampled at each stage, using two alternative approaches: 1. *Independent*: Predicting caste at each targeted stage (Sn) using the observed (true) caste labels as training stage (Sn+1) to examine the prediction accuracy when the training caste identities are in fact known from morphological information. Here, the ratio in each stage reflects the accuracy of BPA in each stage. *Progressive*: Predicting caste starts from the late pupal stage using adults of known caste identity as training data. Here, BPA is then performed progressively using the predicted caste labels in late pupae to predict the caste identity in early pupae. This process was repeated recurrently until the 2nd larval instar. As the first step of BPA constructs a PCA from target stage data, we also compared the accuracies between PCAs obtained from whole transcriptomes and PCAs obtained from caste DEGs at the subsequent stage (training data). We achieved a higher prediction accuracy when PCAs were constructed with caste DEGs at the subsequent stage compared to using whole transcriptome PCAs, probably because the DEG method excluded uninformative housekeeping genes.

**d,** Anti-body staining (VASA protein, red. RRID: AB\_2893405) and in-situ hybridisation (*vasa* RNA, green) in 192-hour old embryos of *M. pharaonis*, showing that germline differentiation has already occurred at this stage. Among the 67 examined embryos, 18 (27%) could be documented to have no germline, indicating that it should be possible to match these presence/absence results among 192-hour embryos with BPA predictions based on 1st instar larval transcriptomes.

**e**, Second instar larvae of *A. echinatior* lack the full-body curly hairs that distinguish gynes from workers in the 3rd larval instar, which means caste cannot be identified morphologically. We applied BPA to predict caste identities among 2nd-instar larvae (**Fig. 2a**). A closer inspection showed that suspected gynes have in fact some gyne-like curly hairs, which are thicker than those in suspected workers, on their ventral thorax (arrows). These observations indicated these individuals are future gynes and were consistent with our BPA predictions.

**f**, Among 2nd instar larvae of *M. pharaonis*, PCA with whole transcriptomes showed that the overall transcriptomic difference between gynes and males was not significant (*P* = 0.28) while reproductive larvae of both sexes were always separated from worker larvae (*P* < 5e-5 and *P* < 5e-6 for gynes and males, respectively). This is consistent with images of 2nd instar gyne and male larvae being indistinguishable after we used microsatellite genotyping to determine whether individuals were haploid (male) or diploid (female). Numbers of differentially expressed genes (adjusted *P* value < 0.05, detected with a generalized linear model that accounted for body size differences) also support this conclusion: 152 genes were differentially expressed between gyne andworker larvae, while 50 genes were differentially expressed between gyne andmale larvae.

A picture containing diagram

Description automatically generated

Extended Data Fig. 3. Transcriptomic canalization during caste differentiation in ants.

**a**, Within-stage transcriptome variation in *M. pharaonis* (upper panel) from 0–12 h old embryos to late pupae, plotted separately for gynes (red), workers (blue) and all individuals within each stage (black) depending on available information. The lower panel gives the same information for *A. echinatior*, where embryonic data were not available and 1st instar caste phenotypes (grey) were inseparable with BPA. Transcriptome variation was quantified as 1 – r, the extent of imperfection of transcriptome-level Spearman’s correlations between a target individual and all other same-stage and same-caste individuals. Caste identities of 1st instar individuals of *M. pharaonis* and 2nd instar individuals of *A. echinatior* were predicted by BPA. In *M. pharaonis*, transcriptome variation for all individuals peaked in 36–48 h old embryos (equivalent to the gastrulation stage, 6 – 7, in *Drosophila* larvae). For both species, transcriptome variation among gynes was consistently lower than among workers. In pupal stages of *M. pharaonis*, transcriptome variation across all individuals exceeded transcriptome variation for the gyne and worker subsets, indicating that transcriptome differences primarily reflected realized caste differentiation, in contrast to the pattern observed across the larval stages, where the black curve was intermediate between the red and blue curves. Fourth larval instar and prepupal gyne samples of *A. echinatior* were excluded from this analysis, because these samples were sequenced in a different technical batch, making their transcriptome variation incomparable with the other samples.

**b**, Developmental potential (∆) for individual gynes and workers in *A. echinatior*, measured as the transcriptomic distance between a focal individual and an average gyne or worker (pooling all three worker subcastes) phenotype in the next developmental stage. Developmental potential was quantified and presented as in *M. pharaonis* (**Fig. 3a**), except that all three worker subcastes were included. Caste identities for gynes and (pooled) workers in 2nd instar larvae were predicted by BPA. As in the **panel a**, fourth larval instar and prepupal gyne samples were excluded to avoid a batch effect.

**c,** PCAs of early and late pupal stage transcriptomes in *M. pharaonis* (left) and *A. echinatior* (right). For both species, the first PC axis (PC1) separates individual transcriptomes by developmental stage (early pupae to the left and late pupae to the right) while PC2 captures the caste-related transcriptomic variation. The overall transcriptomic difference between gynes (red) and workers (blue) increases from the early to the late pupal stage (upper panels), and the absolute values of the PC2 residuals (lower panels), representing the variation within each caste, were always lower among gynes than among workers (*P* < 1x10-3 for both species, two-sided *t*-tests). This is consistent with the mean extent of canalization being stronger in gynes than in workers. In *A. echinatior*, the absolute residual differences increase for the workers, consistent with *A. echinatior* having worker sub-castes that differentiate rather late in development.

Graphical user interface

Description automatically generated

**Extended Data Fig. 4. Early larval caste differentiation in ants.**

**a,** Tissue-specific relative expression levels for the conserved caste-biased DEGs in early larvae, shown separately for gyne-biased (rows marked in red) and worker-biased (blue) genes. Heatmap brightness of cells reflects tissue specificity, the percentage of transcripts from targeted tissues (columns), ranging from 0% (black) to 100% (yellow). These relative abundances, based on the larval gene expression atlas of *Drosophila*, show that the gyne-biased DEGs in the early larval stages were mainly expressed in the midgut, fat body, and tracheal tissues, while the worker-biased DEGs were mainly expressed in the brain and central nervous system.

**b**, Expression profiles of *circadian clock-controlled protein* (*daywake*), *juvenile hormone acid O-methyltransferase-like* (*jhamt-like*) and *hexamerin* among gynes and workers of the two ant species as larvae grow. All three genes are associated with the juvenile hormone signalling pathway and are significantly differentially expressed between castes in 2nd and 3rd instar larvae. Expression profiles are plotted against body length (log scale) to show expression dynamics as larvae grow in body length.

**c,** DAPI staining of a representative early 3rd instar worker larva and a representative 2nd instar gyne larva of *M. pharaonis*. These animals display similar body size, but wing discs (arrows) were only visible in the gyne larvae, indicating that caste determination and differentiation has already been initiated well before this early larval stage.

A picture containing calendar

Description automatically generated

Extended Data Fig. 5. JH and E20 signalling pathways play key role in the regulation of canalized caste phenotypes

**a**, Expression profiles of eight key regulators for insect metamorphosis that are part of the juvenile hormone and ecdysone signalling pathways (**Fig. 3b**), plotted against body length (log scale) of 2nd and 3rd instar *M. pharaonis* larvae. The expression levels of half of these genes (*jheh2*, *jhamt, usp* and *E93*) showed caste-specific body length thresholds in the 3rd larval instar. This pattern indicates gyne and worker individuals are gated by different critical masses for entering the metamorphic molt.

**b**, Compared to the control group (3rd instar worker larvae fed with 10% EtOH PBS), feeding JH analog (JHA) to 3rd instar worker larvae delayed achieving pupation.

**c**, JHA fed 3rd instar worker larvae induced inter-caste with phenotype intermediate between gyne and worker. JHA fed workers have larger body size and developed wing buds (arrowed), however, they never developed ovaries (not show), indicating early bifurcation between colony germ-soma phenotypes.

**d**, Compared to the control group (3rd instar gyne larvae fed with 10% EtOH PBS), precocene I fed 3rd instar gyne larvae were smaller and developed abnormal wings.

Graphical user interface

Description automatically generated

Extended Data Fig. 6. Canalized genes play important roles in producing adaptive caste phenotypes.

**a**, Tissue-specific relative expression abundances in *Drosophila* for gene orthologs that show caste-specific canalized expression in *M. pharaonis*, plotted separately for gyne-biased and worker-biased genes. Compared to whole-genome background, gyne-biased canalized genes have a higher relative transcript abundance in ovaries, midgut, and fat body, whereas worker-biased canalized genes have a higher relative expression in the brain, eyes, and thoracic ganglia (two-sided *t*-tests; \* *P* < 0.001).

**b**,Diagrammatic illustration of gyne-biased canalized genes being associated with traits in ovaries andwing muscles, whereas worker-biased canalized genes are associated with brain function and behaviour (see Supplementary table 4 for full list of canalized genes).Image courtesies: Anna Mosegaard Schmidt (adult gyne) and Luigi Pontieri.

**c**, Canalization scores for flight related (left) and ovary specific (right) genes in the two ant species. Flight related genes were identified based on their *D. melanogaster* homologs associated either with flight performance itself or with striated muscle functionality (the crucial wing muscles tissue in insects). Ovary specific genes were genes having > 30% expression abundance in ovaries of *D. melanogaster* females compared to the sum of their expression in all tissues (see Methods). Colours of cells represents the canalization score, ranging from -3 (blue, canalized in worker biased direction) to 3 (red, canalized in gyne biased direction). Canalization scores in *A. echinatior* were calculated by comparing gyne and small worker transcriptomes.

**d**, Developmental expression dynamics of *ATP- dependent RNA helicase vasa* (*vas*), *vitellogenin receptor* (*yl*) and *serine/threonine-protein kinase Chk2* (*lok*) in the two ant species. All three genes are ovary specific with high expression abundance in *D. melanogaster* ovaries. Although these three genes showed increasing gyne-biased canalization in *M. pharaonis* as development proceeds, there was little expression difference between gyne and worker individuals in *A. echinatior*, except for *yl* in prepupae and to a lesser extent also in 3rd instar larvae.

A picture containing graphical user interface

Description automatically generated

**Extended Data Fig. 7. *Freja’s* functional role in canalizing gyne phenotypes.**

**a**, Predicted functional domains of the protein encoded by *Freja* (LOC10587931), annotated with InterPro (see Methods). *Freja* contains a signal peptide domain at the N terminus, indicating secretion or membrane insertion. In addition, Freja contains a leucine-rich-repeat domain, suggesting its role in protein binding.

**b**, *Freja* is the most strongly canalized gene in *M. pharaonis,* showing an increasing between-caste expression difference and a decreasing within-caste expression variance as development proceeds. The caste identities in 1st instar larvae are based on BPA prediction.

**c**, Tissue-specific RT-PCR quantification of *Freja* transcript abundance in adult gynes, showing *Freja*’s expression is restricted to the abdomen and especially highly abundant in the ovaries.

**d,** RT-PCR quantification of the efficiency of *Freja* RNAi in 3rd instar larvae (left) and adult gynes (right). Compared to the *GFP* RNAi control group, *Freja* RNAi significantly reduced the expression level of *Freja* (p < 1e-3 in one-way ANOVAs in each age group).

**e**, Compared with the control group, *Freja* RNAi significantly reduced the number of yolky oocytes in adult gynes (p = 0.004 in one-way ANOVA).

Extended Data Table 1. Experimental design.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Developmental stage | | Number of individual replicates | | | RNA extraction method | Average RNA concentration  (ng/ul) | Average amount of RNA sequencing data  (Gb per replicate) |
| Unknown caste | Gyne | Worker  (Small/Medium/Large in *A. echinatior*) |
| ***M. pharaonis*** | | | | | | | |
| Embryo | 0 – 12 h | 27 |  | | PicoPure RNA kit | 0.1 | 8 |
| 12 – 24 h | 21 | 0.2 | 7 |
| 36 – 48 h | 14 | 0.2 | 6 |
| 60 – 72 h | 21 | 0.4 | 9 |
| 84 – 96 h | 19 | 0.2 | 9 |
| 108 – 120 h | 14 | 0.4 | 11 |
| 132 – 144 h | 20 | 0.4 | 5 |
| 156 – 168 h | 19 | 0.7 | 6 |
| 180 – 192 h | 35 | 0.4 | 8 |
| Larva | 1st instar | 53 | 1.9 | 10 |
| 2nd instar |  | 49 | 26 | RNeasy micro kit | 2.5 | 10 |
| 3rd instar | 37 | 17 | 5.1 | 9 |
| Pre-pupa | 40 | 38 | 13.0 | 10 |
| Pupa | Early | 20 | 29 | 24.8 | 10 |
| Late | 30 | 25 | 26.5 | 11 |
| Adult | Imago | 24 | 16 | 3.5 \* | 9 |
| ***A. echinatior*** | | | | | | | |
| Larva | 1st instar | 40\*\* |  | | RNeasy micro kit | 3.5 | 10 |
| 2nd instar | 62\*\* | 23.5 | 9 |
| 3rd instar |  | 24 | 41/10/27 | 123.5 | 7 |
| 4th instar | 26 |  | 673.7 | 6 |
| Pre-pupa | 29 | 33/16/36 | 358.3 | 7 |
| Pupa | Early | 18 | 14/14/21 | 380.1 | 12 |
| Late | 18 | 17/16/16 | 261.8 | 11 |
| Adult | Imago | 16 | 24/1/15 | 294.2 | 11 |
| ***D. melanogaster*** | | | | | | | |
|  |  | Unknown sex | Female | |  |  |  |
| Embryo | 1 h | 28 |  | | PicoPure RNA kit | 2.7 | 7 |
| 3 h | 24 | 7.3 | 8 |
| 6 h | 21 | 4.4 | 7 |
| 9 h | 18 | 4.0 | 6 |
| 12 h | 16 | 3.6 | 7 |
| 15 h | 18 | 4.0 | 7 |
| 18 h | 22 | 3.4 | 7 |
| 21 h | 21 | 3.7 | 8 |
| Larva | 1st instar  (24 – 44 h) | 79 | 7 | 10 |
| 2nd instar  (50 – 68 h) |  | 21 | | RNeasy micro kit | 8.8 | 10 |
| 3rd instar  (Pre-pupa)  (80 – 116 h) | 36 | | 19.8 | 11 |
| Pupa | Early  (128 h) | 18 | | 26.2 | 8 |
| Late  (140 h) | 14 | | 23.9 | 8 |
| Adult | Imago  (180 h) | 14 | | 16.6 | 9 |

\* The lower RNA concentrations in *M. pharaonis* adults may be due to the harder cuticles in adults, which made lysis of these animals more difficult during the RNA extraction procedure. While the transcriptome variation in adults may not be comparable with that in other stages, we retained these samples to represent the final state of irreversible caste differentiation, because the mean transcriptomes for all samples of the same stage remain meaningful for appreciating the overall caste differentiation process.

\*\* 1st and 2nd instar larvae in *A. echinatior* were all females, verified by five highly polymorphic nuclear microsatellite loci (see Methods for details).

Extended Data Table 2. Determination of developmental stages and caste identities in the two ant species.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Stage | | Caste | | Head capsule width range | Body length range | Morphological characters |
|  | |  | | ***M. pharaonis*** |  |  |
| Embryos | | U | |  | Not measured |  |
| Larva | 1st instar | U | | 0.11 – 0.15 mm | < 0.4 mm |  |
| 2nd instar | S | | 0.16 – 0.23 mm | 0.5 – 1.2 mm | No body hairs |
| W | | 0.15 – 0.20 mm | 0.4 – 0.7 mm | With body hairs |
| 3rd instar | S | | 0.23 – 0.32 mm | 1.2 – 3.1 mm | No body hairs |
| W | | 0.20 – 0.28 mm | 0.7 – 2.1mm | With body hairs |
| Pre-pupa | S | |  | 1.9 – 2.6 mm | Gut empty & No body hairs |
| W | | 1.3 – 1.7 mm | Gut empty & With body hairs |
| Pupa | Early | G | | 2.6 – 2.9 mm | White cuticle with gyne morphology |
| W | | 1.4 – 1.8 mm | White cuticle with worker morphology |
|  | Late | G | | 2.6 – 2.9 mm | Dark cuticle with gyne morphology |
| W | | 1.4 – 1.8 mm | Dark cuticle with worker morphology |
| Adult | Imago | G | | Not measured |  |
| W | | Not measured |  |
|  |  | |  | ***A. echinatior*** |  |  |
| Larva | 1st instar | F | | 0.18 – 0.31 mm | 0.5 – 0.9 mm |  |
| 2nd instar | F | | 0.30 – 0.60 mm | 0.8 – 1.9 mm |  |
| 3rd instar | G | | 0.31 – 0.57 mm | 2.0 – 5.0 mm | Curly hairs cover the whole body; Gyne type body shape (Abdomen is larger than the frontal body). |
| SW | | 0.41 – 0.54 mm | 1.4 – 2.9 mm | No curly hairs; Few Y-shaped hairs on ventral thorax. No gyne type body shape. |
| MW | | 0.46 – 0.51 mm | 3.1 – 3.9 mm | Same as above. |
| LW | | 0.47 – 0.52 mm | 4.0 – 5.3 mm | Same as above. |
| 4th instar | G | | 0.49 – 0.56 mm | 5.0 – 7.2 mm | Curly hairs cover the whole body; Gut visible. |
| Pre-pupa | G | |  | 6.3 – 7.2 mm | Curly hairs cover the whole body; Gyne type body shape. Developing legs and eye pigmentation are visible. Gut empty |
| SW | | 2.0 – 3.0 mm | Developing legs and eye pigmentation are visible. Gut empty |
| MW | | 3.0 – 4.0 mm | Same as above. |
| LW | | 4.0 – 5.2 mm | Same as above. |
| Pupa | Early | G | | 7.5 – 8.2 mm |  |
| SW | | 2.1 – 3.0 mm |  |
| MW | | 3.1 – 3.9 mm |  |
| LW | | 4.1 – 6.1 mm |  |
|  | Late | G | | 7.3 – 8.1 mm |  |
| SW | | 2.2 – 3.0 mm |  |
| MW | | 3.1 – 4.0 mm |  |
| LW | | 4.0 – 5.7 mm |  |
| Adult | Imago | G | | 6.9 – 8.2 mm |  |
| SW | | 2.0 – 3.9 mm |  |
| MW | | 4.0 mm (only one sample) |  |
| LW | | 4.1 – 7.0 mm |  |

\*\*\* S: Sexual, including male and female reproductives, which are morphological indistinguishable in larvae and pre-pupae. We later used microsatellite genotyping to separate males and females. F: Female. G: Gyne. W: Worker. SW: Small worker. MW: Medium worker. LW: Large worker. U: Unknown.

For morphological characterization of caste identities and developmental stages in *A. echinatior*, see: <https://megalomyrmex.osu.edu/temp/acro-larva-key/>