Science
AAAS

Supplementary Materials for

Social regulation of insulin signaling and the evolution of eusociality in ants

Vikram Chandra[#,*], Ingrid Fetter-Pruneda[#,*], Peter R. Oxley, Amelia L. Ritger, Sean K. McKenzie, Romain Libbrecht, Daniel J. C. Kronauer[*]

[#]These authors contributed equally

*Correspondence to vchandra@rockefeller.edu (V.C.), ifetter@rockefeller.edu (I.F-P.), dkronauer@rockefeller.edu (D.J.C.K.)

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

MATERIALS AND METHODS

RNA sequencing (RNA-Seq) comparative screen

For this experiment, we compared RNA-Seq data from seven focal ant species: *Dinoponera quadriceps, Ooceraea biroi, Camponotus planatus, Odontomachus ruginodis, Solenopsis invicta, Harpegnathos saltator, and Acromyrmex echinatior*. *C. planatus*, *O. ruginodis*, and *S. invicta* samples were collected specifically for this study, while the data from *D. quadriceps*, *A. echinatior*, *H. saltator* and *O. biroi* were from previously published studies (*7–10*). Details of the life history and experimental sampling and procedures for each species can be found in Table S1. Code for the comparative analysis is available at (*31*).

Sample collection

We collected multiple queenright colonies of *Odontomachus ruginodis* and *Camponotus planatus* at the Archbold Biological Station near Lake Placid, Florida, in February 2016. Each colony contained a single queen or gyne, and 20-30 workers. We maintained these colonies in the lab on a diet of sugar water *ad libitum*. After five days all queens and workers were snap-frozen on dry ice and maintained at -80°C until dissected.

Queenright *Solenopsis invicta* colonies were collected in Gainesville, Florida, in November 2014. We maintained the *S. invicta* colonies in the lab for several months on the Bhatkar diet (without eggs) (*32*) and mealworms. From each of 5 colonies, we collected a reproductively active queen, five reproductively inactive gynes, 20 foraging workers (collected directly from the mealworm and flan feeding trays), and 20 workers from within the nest (collected directly from the brood pile). All queens and workers were snap-frozen on dry ice and maintained at -80°C until dissected. For the RNA-Seq analysis presented here (described in detail below), we treated both gynes and queens as reproductives, and both nurses and foragers as non-reproductives.

O. biroi colonies were maintained in the lab at 25°C in boxes with a plaster of Paris floor. During the brood care phase colonies were fed thrice weekly with frozen *S. invicta* brood. Experiments were performed with ants from clonal line B.

Dissection and RNA-Seq

Brains were dissected in 1x PBS at 4°C. Dissected brains were immediately transferred to TRIzol (Invitrogen) and placed on dry ice. We also dissected the ovaries of each ant to assess its reproductive status (see Table S1). RNA was extracted in TRIzol using RNeasy (Qiagen) purification with DNAse I (Qiagen) on-column digestion, using a previously reported protocol (*7*).

Library preparation and RNA-Seq were performed at the Rockefeller University Genomics Resource Center as follows. For the *O. biroi* dataset (Fig. 2F), 2 ng of total RNA was used to generate full length cDNA (of mRNA) using Clontech's SMART-Seq v4 Ultra Low Input RNA Kit (Cat # 634888). 1 ng of cDNA was then used to prepare libraries using Illumina Nextera XT DNA sample preparation kit (Cat # FC-131-1024).

16 libraries were prepared with unique barcodes and pooled at equal molar ratios. The pool was denatured and sequenced on an Illumina NextSeq 500 sequencer to generate 150 bp single-end reads following the manufacturer's protocol. For *S. invicta*, 1 ng of total RNA was used to generate full-length cDNA using Clontech's SMART-Seq v4 Ultra Low Input Kit (Cat # 634888), 1 ng of which was then used to prepare libraries using Illumina Nextera XT DNA sample preparation kit (Cat # FC-131-1024). Libraries with unique barcodes were pooled at equal molar ratios. Each pool was sequenced on two lanes on an Illumina HiSeq 2500 sequencer to generate 100 bp single-end reads, following the manufacturer's protocol (Cat #15050107 V03). For *O. ruginodis* and *C. planatus,* 100 ng of total RNA was used to generate libraries using the Illumina TruSeq stranded mRNA LT kit (Cat # RS-122-2101). Libraries prepared with unique barcodes were pooled at equal molar ratios. The pool was denatured and sequenced on an Illumina NextSeq 500 sequencer using high output V2 reagents and NextSeq Control Software v1.4 to generate 150 bp paired-end reads, following the manufacturer's protocol (Cat # 15048776 Rev.E).

Transcriptome Assembly of *Odontomachus ruginodis* and *Camponotus planatus*

Sequences were trimmed using Trimmomatic (*33*), followed by quality checking with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmed sequences were then assembled using Trinity (*34*) with the default settings; i.e. including a minimum contig length of 200 bp and a normalized maximum read coverage of 100.

Prediction and translation of peptide sequences was performed using Transdecoder (2014; http://transdecoder.github.io/). Peptide predictions were trained on the annotated set of *O. biroi* peptides. To assess transcriptome quality, we used BUSCO assessment (*35*), comparing against the BUSCO arthropod database.

***O. ruginodis* TransDecoder Peptide BUSCO assessment:**

2347 (87.7%)	Complete BUSCOs
2111 (78.9%)	Complete and single-copy BUSCOs
236 (8.8%)	Complete and duplicated BUSCOs
180 (6.7%)	Fragmented BUSCOs
148 (5.5%)	Missing BUSCOs
2675	Total BUSCO groups searched

***C. planatus* TransDecoder Peptide BUSCO assessment:**

2413 (90.2%)	Complete BUSCOs
2155 (80.6%)	Complete and single-copy BUSCOs
258 (9.6%)	Complete and duplicated BUSCOs
108 (4.0%)	Fragmented BUSCOs
154 (5.8%)	Missing BUSCOs
2675	Total BUSCO groups searched

Because the *O. ruginodis* and *C. planatus* libraries were sequenced on the same lane, we found low levels of index-switching-like contamination. Specifically, it appears that during demultiplexing some highly expressed genes in *O. ruginodis* were misidentified as

belonging to a *C. planatus* sample and *vice versa*. This is a known issue that adds low levels of noise to multiplexed sequencing experiments (*36*). We estimate that roughly 0.03% - 0.04% of reads in each library in *C. planatus* and *O. ruginodis* are contaminants, in line with previous estimates of contamination caused by index misassignment. Contaminated reads are normally filtered out when they are aligned to a reference genome. In our case, however, we use the reads to generate *de novo* transcriptomes for *O. ruginodis* and *C. planatus*, and this means that many contaminants are annotated as unique genes. Because we aimed to analyze single-copy orthologs that are conserved across the ants, this constituted a potential issue that could have interfered with our analysis. Specifically, such contamination could have caused some genes to be wrongly classified as having undergone gene duplication in *O. ruginodis* or *C. planatus*. These ortholog groups would thus have been excluded from our analysis. To overcome this potential problem, we identified and removed contaminant transcripts from the annotated transcriptomes before proceeding with our analyses.

To identify specific contaminants in each transcriptome, we selected genes/transcripts that had >90% blastp (*37*) identity to a gene in the other transcriptome. We then found the best blastp hit to the focal gene in a set of ant and bee species. This set contained the seven ant species in our screen (including the query species: *O. ruginodis* or *C. planatus*, respectively), *C. floridanus*, *Lasius niger*, *Apis mellifera*, *Megachile rotunda*, and *Bombus impatiens*. We then aligned these sequences and constructed maximum likelihood protein phylogenies (using a JTT + gamma substitution model and an automated version of the procedure described in the section on phylogenetic analysis below) for the ortholog groups that these genes were part of. We rooted the resulting trees on the ancestor of all bees, excluding trees that failed to produce a nearly-monophyletic bee grouping, or that did not contain a gene from any of the ponerine or formicine ant species in our alignment. When the alignments contained no bee sequences, we rooted trees on *O. biroi* instead. We used the resulting tree topologies to classify genes as contaminated, contaminating, or clean. Contaminated and contaminating genes were defined as those that produced clades of *O. ruginodis* and *C. planatus*, in distinct contradiction to the expected species topology (seen in Fig. 1). Contaminated genes fell within the wrong ant subfamily on the phylogeny, while contaminating genes were in the right ant subfamily but had a monophyletic relationship between *O. ruginodis* and *C. planatus*. We also used secondary features of the phylogenies (i.e. the genetic distances between the *O. ruginodis* and *C. planatus* genes in each tree, and the Robinson-Foulds (*38*) distance between each gene tree and the expected species tree), as well as features of each gene (including the length of its predicted amino acid, its average expression level, etc.) to identify genes that our phylogenetic classification missed. Preliminary analysis (not shown) found that, as expected, contaminated genes (as classified by our phylogenetic analysis) strongly tended to have low expression and short peptide lengths relative to their best hit (i.e. a contaminating gene) in the other species (i.e. in *C. planatus* when the focal gene was from *O. ruginodis*, and *vice versa*). These genes also tended to have extremely low genetic distance (calculated from the ML phylogeny) to their best hit in the other species, and their phylogenies tended to have high Robinson-Foulds distances to the hypothetical species phylogeny. Based on these data, we also classified *O. ruginodis* (or *C. planatus*) genes as contaminated if (a) the gene fell inside a formicine (or

ponerine in the case of *C. planatus*) clade and it was >20% shorter and >20% less expressed than its best hit in the other species, or (b) if the gene wasn't in the ponerine (or formicine) clade, had ≥0.6 Robinson-Foulds distance, and had either extremely low genetic distance to their best hit in the other species or was both >20% shorter and >20% less expressed.

Overall, this analysis identified roughly 700 contaminated genes in each transcriptome; i.e. about 3% of annotated transcripts in each transcriptome were contaminated. We removed these genes from the transcriptomes before differential expression analysis. Consequently, they were also excluded from our single-copy ortholog identification. Orthology identification with contaminations excluded found 40 more single-copy orthologs across the ants. If we had not removed contaminated genes from the transcriptomes, our comparative analysis (described in detail below) would be slightly more conservative. This is because we only analyze single-copy orthologs. The presence of a contaminant causes an ortholog group to appear as though one of its member species possesses two paralogs of the gene, and would thus exclude it from further analysis. Thus, even if our contamination cleanup and transcriptome curation had missed a few contaminants, this would serve to increase the probability that our comparative analysis has false negatives, but not the rate of false positives. Further, we presume that this index misassignment affects all our samples equally. Although we are able to detect and remove contaminated genes when they are found in the transcriptome of a species different from that of their origin, we are unable to identify contaminated reads that hop across two samples of the same species. In principle, this could contribute to the low statistical power of our differential expression analysis, and it could be one reason for the low number of common differentially expressed genes we detect across our seven focal ant species. Despite this, *O. ruginodis* and *C. planatus* have higher numbers of differentially expressed genes than any other species in our screen (Table S1), while *H. saltator* and *D. quadriceps* have very low numbers of differentially expressed genes. Thus, it appears that the latter species explain the low numbers of common differentially expressed genes we detect, and possible contamination in *O. ruginodis* and *C. planatus* does not limit our statistical power.

Differential gene expression analysis

RNA-Seq fastq files for *D. quadriceps*, *A. echinatior*, *H. saltator*, and *O. biroi* were obtained from the NCBI Short Read Archive (PRJNA255520, PRJNA223531, PRJNA327090, and PRJNA304722, respectively). RNA data from all seven species were hereafter analyzed in the same manner. RNA-Seq reads were adapter trimmed with Trimmomatic, quality checked with FastQC, and aligned to their respective genomes using STAR aligner (*39*). For *O. ruginodis* and *C. planatus*, we used STAR to align reads to their respective curated transcriptomes. Aligned read counts for each gene were calculated using HTSeq (*40*). We used a modified version of HTSeq (https://github.com/oxpeter/htseq-transcriptome) to produce read counts for *O. ruginodis* and *C. planatus*; this was done to ensure that HTSeq did not exclude counts for reads mapping to multiple isoforms of the same predicted gene. Differentially expressed genes were identified using DESeq2 (*41*), using a Wald Test to contrast the queen and worker castes, and an FDR q-value cutoff of 0.05 (although we varied this cutoff for the analysis

presented in Fig. S1). See Table S2 for the results of our differential expression analysis for all single-copy orthologs in all seven focal ant species.

Single copy ortholog identification

For orthology detection, we downloaded 16 hymenopteran genome annotations from NCBI, including the five focal ant species in our screen for which genome sequences have been published (Table S4). For each gene, the longest transcript was chosen as the representative transcript for that gene. The inferred peptide sequences corresponding to the longest transcripts for each of the 16 species, as well as the inferred peptide sequences of *O. ruginodis and C. planatus,* were then used as input for OrthoMCL (*42*). 17,601 groups of orthologs were identified by OrthoMCL. From these, we identified 5,581 groups in which all seven focal species had one, and only one, gene present. These groups were considered single-copy orthologs across our focal taxa. See Table S2 for single-copy ortholog groups.

Brain Staining

ILP2 antibody production

Custom rabbit polyclonal anti-ILP2 designed to recognize an epitope in the B chain was generated and affinity purified by YenZym. YenZym also used ELISA to assess the specificity of the antibody response to the immunizing peptide in vitro. To assess peptide specificity in situ, we pre-incubated the antibody with ILP2 peptide. The subsequent immunostain showed no ILP2 staining, as we would expect if the antibody was bound to synthetic ILP2 and therefore unable to bind to endogenous peptide (Fig. S7E,F).

Immunohistochemistry

Whole mount brain stainings as well as ovary and fat body stainings for ILP2 were performed following a similar protocol as in (*43*). Briefly, *O. biroi* tissues were dissected in cold phosphate-buffered saline (PBS), pH 7.4. The tissues were then fixed by incubation in 4% (wt/vol) paraformaldehyde solution in PBS overnight at 4°C. Tissues were washed in PBS once, followed by three 20 min washes in PBS containing 0.5% Triton-X (PBT) at room temperature on a shaker. Samples were blocked in 1% Bovine Serum Albumin (BSA) in PBS for 30 min, washed in PBS 0.01% Tween for 5 min and incubated for either 24 hours at RT or for 48 hours at 4°C (this varied across experiments) with our anti-ILP2 antibody (1:500) containing solution in 1% BSA and 0.5% Triton-X in PBS solution. The next day, samples were washed 3 times for 10 min with PBS Tween (0.01%), incubated with a secondary antibody Alexa Fluor 488 goat anti-rabbit solution (1:250) containing Alexa Fluor 555 or 647 phalloidin (Thermo Fisher Scientific) 1:50 µL of stock solution 6.6 µM) and 4',6-diamidino-2-phenylindole (DAPI) (1:1000) in 1% BSA and 0.5% Triton-X in PBS solution for 2 h and washed five times in PBS. Tissues were mounted with Dako mounting medium between two cover slips separated by a stack of two reinforcement labels (Avery 5720), mounted on a frosted slide, and sealed using clear nail polish. Tissues were imaged using a Zeiss LSM 880 NLO laser scanning confocal microscope. Images were acquired at a pixel resolution of 1024 x 1024 keeping configuration settings equal within experiments.

For the comparison of workers between the reproductive- and brood care phase, we split an age- and phase- matched colony of ants into two, and desynchronized them by taking the larvae away from one colony and allowing the ants to complete a full reproductive cycle. When the two colonies were in the peak of their reproductive and brood care phases, respectively, we dissected all ant brains, stained them in parallel with anti-ILP2 antibody as described above, and analyzed them as described below.

For each comparison of workers and intercastes in both phases of the reproductive cycle, we established a colony of newly-eclosed callow workers and intercastes from a large stock colony. For the experiment in the reproductive phase, we waited approximately four weeks, until the ants were in the middle of their reproductive phase. For the experiment in the brood care phase, we waited roughly five weeks, until the ants had progressed through a reproductive phase and were in the peak of their subsequent brood care phase. We dissected and stained all ant brains in parallel for each experiment.

Image analysis

We imaged immunostained ant brains with a confocal microscope as described above. We then used IMARIS (Bitplane) to semi-automatically segment clusters of insulin-producing cells in each brain image stack (Fig. S6D,E). We then used the total intensity of all fluorescent voxels in each image to quantify insulin-producing cell fluorescence intensity (Fig. 2D, Fig. 4B,C). We used the imaging software Zen (Zeiss) to construct 3D projections of the ovaries (Fig. 4C,D) and FIJI (ImageJ) to construct 3D projections of the brains (Fig. 2A-C, Fig. S6, Fig. S8) and fat bodies (Fig. S7C,D).

ILP sequence analysis and peptide synthesis

Predicted ILP1 and ILP2 structures in *Ooceraea biroi*

Insulin ortholog and paralog sequences are highly divergent. All members in the insulin superfamily have six characteristic conserved cysteines that form three disulfide bonds (Fig. S5A). The ILP1 sequence resembles mammalian insulin-like growth factor (IGF) in that it has a short C chain with flanking regions. Furthermore, ILP1 lacks recognizable dibasic amino acid cleavage sites flanking the C chain (Fig. S5B,C). Short uncleaved C chains are a common feature of IGF-like peptides (*44*, *45*). The ILP2 sequence has a longer C chain, as in canonical insulins, that is flanked by dibasic and/or monobasic peptide cleavage sites (*46*, *47*) (Fig. S5B). This suggests that, as in canonical insulins, the C chain is cleaved from the propeptide to form a mature peptide consisting of A and B chains held together by disulfide bonds (Fig. S5C).

Phylogenetic Analysis

Peptide sequences for ILP genes from representative Hymenoptera taxa were retrieved from NCBI (Table S4) as well as our transcriptome assemblies of *O. ruginodis* and *C. planatus*. The sequences were aligned using the linsi settings of the MAFFT alignment algorithm (*48*). A maximum likelihood tree was then constructed using RAxML (*49*) under the LG+I+G model of protein evolution. Support for this tree was calculated using 100 bootstrap replicates. We also used MrBayes (*50*) to construct a Bayesian phylogeny, with a proportion of invariant sites and gamma-distributed rate variation (LG + I +

Gamma). Two runs were conducted for 200,000 iterations, with three hot chains and one cold chain each. The average standard deviation of split frequencies was <0.01. As the Bayesian and ML topologies were nearly identical, we show the ML phylogeny overlaid with ML bootstrap scores and Bayesian posterior probabilities (Fig. S4).

ILP2 synthesis

ILP2 peptide synthesis was performed by Phoenix Pharmaceuticals based on the predicted peptide sequence in *O. biroi*:
A Chain: GIHEECCVNACTISELSSYCGP,
B Chain: SSISAPQRYCGKKLSNALQIVCDGVYNSMF,
[Disulfide bonds: A6-A11, A7-B10, A20-B22].
Phoenix Pharmaceuticals performed a mass spectrophotometric analysis that showed that the synthetic peptide had the expected molecular weight. The peptide was >95% pure. The lyophilized peptide was first reconstituted in a minimal volume of water (pH 2.6) and further diluted to a concentration of 100 mM in PBS (pH 7.4). Aliquots were frozen at -80°C until used.

B chain peptide synthesis

Another batch of the B chain of *O. biroi* ILP2 was synthesized separately as an injection control by The Rockefeller University Proteomics Resource Center and verified through liquid chromatography and mass spectrometry. Lyophilized peptides were first reconstituted in a small volume of water (pH 2.6), diluted to a concentration of 100 mM in PBS (pH 7.4), and stored in frozen aliquots at -80°C until used.

Pharmacology

ILP2 injections

For each of the experiments involving ILP2 injections, all *O. biroi* ants were collected as callows (i.e. newly eclosed adults) from a single large colony entering the brood care phase. This was to ensure that all ants were of the same age and genotype. We then allowed the ants to progress through a natural colony cycle. We injected ILP2 or the B chain control into one month-old ants in peak brood care phase (i.e. 5 days after the larvae had hatched; Fig. 3, Fig. S10) and in early reproductive phase (i.e. 2 days after the larvae had pupated; Fig. S11B,C). ILP2 and the B chain stock solution were diluted in PBS to reach concentrations of 10 mM and 100 mM. We selected these concentrations based on data from *Aedes aegypti* (*19*), which suggest that the 10-100 mM range is physiologically relevant. We injected approximately 0.1 - 0.2 mL of ILP2 or the ILP2 B chain into each ant. We used a 36 gauge beveled needle attached to a nanofil syringe (World Precision Instruments, Inc.) for the injections. Each ant was immobilized between silicone pads on the sides of a modified crescent wrench. The needle was inserted dorsally between the first and second tergites of each ant's gaster. In each experiment, we injected 50 ants with ILP2 and 50 ants with the control. Roughly 50% of injected ants died within the first two days after injection. This high initial mortality reflects the fact that injections are technically challenging given the small size of the ants, and that a large proportion of ants are physically damaged by the injection process. However, we found no difference in mortality rates between treatments, and did not detect differences in

long-term survival between uninjected control ants and injected ants that had survived the first two days after injection. For the reported experiments, all ants that were still alive three days after injection were dissected and assayed.

Ovary dissection, staining and mounting

Ants were briefly immersed in 95% ethanol and then transferred to PBS. Ant ovaries were dissected in cold PBS at pH 7.4. They were then fixed by incubation in 4% (wt/vol) paraformaldehyde solution in PBS overnight, washed in PBS and stained with DAPI, and then washed in PBS 5x and mounted in DAKO mounting medium. Ovary images for analysis were acquired with an epifluorescence Olympus BX53 microscope. We measured the largest cross-sectional area of the largest oocyte (or second-largest oocyte) in any ovariole for each ant using the image-processing program Cell Sens Standard. We also counted the total number of follicles in each ant's ovaries. Representative images were acquired using a Zeiss LSM 880 NLO confocal microscope. 3D projections were constructed using Zen imaging software (Zeiss).

Starvation experiment

We established ten colonies of 40 one-month-old ants entering the reproductive phase. Each colony of ants laid roughly 20 eggs. Once these eggs had hatched, we fed half the colonies *ad libitum* (i.e. ca. 30 frozen fire ant (*S. invicta*) pupae every two days) and we limited how much food the other colonies received (i.e. ca. 5 fire ant pupae every four days). Once the larvae had all developed into adults, we dissected all of them to assess the proportions of callow intercastes and regular workers in each colony (Fig. S13).

Statistical analyses

Statistics for the immunostaining and oocyte size quantifications were performed using GraphPad Prism 7. Normality was determined by D'Agostino-Pearson normality tests. Where appropriate, we corrected for multiple testing using a Bonferroni correction. Statistics for the differential expression analysis are described above.

Table S1: Summary statistics for RNA-Seq screen, including information about all seven focal ant species (see attached file).

Table S2: Orthology groups and differentially expressed gene lists for all single-copy orthologs in all seven ant species in our screen (see attached file).

Table S3: List of genes that are concordantly and differentially expressed in at least 5 species, with references to their putative role in insulin signaling.

Obir LOC	NCBI gene name (name in this study)	Involved in insulin signaling?	Select references
LOC105274728	diacylglycerol kinase theta 3	Yes	(*51*)
LOC105274915	60S ribosomal protein L26	Probably relevant	(*52*)
LOC105275104	zinc finger protein 350 (*kruppel-homolog1*)	Yes	(*53–56*)
LOC105275581	uncharacterized LOC105275581 3	Unknown	
LOC105276358	serine/threonine-protein kinase MARK2-like 2	Probably relevant	(*57*)
LOC105276512	S-methyl-5'-thioadenosine phosphorylase 1	Unknown	
LOC105276918	14-3-3 protein epsilon 4	Yes	(*58*, *59*)
LOC105278508	semaphorin-2A 3	Unknown	
LOC105278524	LIRP-like (*ilp2*)	Yes	(*16*, *18*, *24*, *60*)
LOC105278591	diacylglycerol kinase 1	Probably relevant	(*61*)
LOC105279346	uncharacterized LOC105279346 3	Unknown	
LOC105280324	neo-calmodulin-like 4	Probably relevant	(*62*)
LOC105280489	inactive dipeptidyl peptidase 10 3	Unknown	
LOC105281606	fas-associated death domain protein 1	Yes	(*63*, *64*)
LOC105281630	dual specificity protein phosphatase 15 4	Unknown	
LOC105281763	dopamine receptor 1 1	Yes	(*21*, *65*)
LOC105282438	uncharacterized LOC105282438 2 (*adenosine receptor*)	Yes	(*66*, *67*)
LOC105282526	60S ribosomal protein L17	Unknown	
LOC105282705	cationic amino acid transporter 2 2	Yes	(*68*)
LOC105283606	insulin-like growth factor 1 (*ilp1*)	Yes	(*16*, *18*, *24*, *60*)
LOC105284100	probable muscarinic acetylcholine receptor gar-2	Unknown	
LOC105284454	neurexin-3-like 7	Probably relevant	(*69*)
LOC105286698	GTP-binding protein RAD	Probably relevant	(*70*)
LOC105288016	synaptic vesicular amine transporter	Probably relevant	(*71*)

Table S4: Species used for orthology identification. The seven species used for RNA-Seq and candidate gene identification are indicated with an asterisk.

Species	**Assembly version**	**Annotation release**
Atta cephalotes	Attacep1.0	100
**Acromyrmex echinatior*	Aech_3.9	100
Apis mellifera	Amel_4.5	103
**Ooceraea biroi*	CerBir1.0	101
Camponotus floridanus	CamFlo_1.0	101
**Camponotus planatus*	transcriptome	n/a
Drosophila melanogaster	Release 6 plus ISO1 MT	n/a
**Dinoponera quadriceps*	ASM131382v1	100
**Harpegnathos saltator*	HarSal_1.0	101
Linepithema humile	Lhum_UMD_V04	100
Monomorium pharaonis	M.pharaonis_V2.0	100
Nasonia vitripennis	Nvit_2.1	102
**Odontomachus ruginodis*	transcriptome	n/a
Pogonomyrmex barbatus	Pbar_UMD_V03	100
Polistes canadensis	ASM131383v1	100
**Solenopsis invicta*	Si_gnG	100
Vollenhovia emeryi	V.emery_V1.0	100
Wasmannia auropunctata	wasmannia.A_1.0	100

Table S5: Differential expression (or lack thereof) of *ilp2* in non-ant social Hymenoptera. Included are microarray, RNA-Seq, and qPCR data from the heads or brains of queens and workers.

Species	***ilp2* differential expression**	**Method**	**Tissue**	**Reference**
Polistes canadensis	Not differentially expressed	RNA-Seq	Brain	(*8*)
Polistes canadensis	Upregulated in queens	RNA-Seq	Brain	(*72*)
Polistes metricus	Downregulated in queens	Microarray	Brain	(*73*)
Polistes dominula	Not differentially expressed	RNA-Seq	Half head	(*74*)
Megalopta genalis	Not differentially expressed	RNA-Seq	Brain	(*75*)
Bombus terrestris	Not differentially expressed	qPCR	Brain	(*76*)

Table S6: Species abbreviations and names for phylogeny in Fig. S4.

Abbreviation	**Species name**
Oabi	*Orossus abietinus*
Nvit	*Nasonia vitripennis*
Tpre	*Trichogramma pretiosum*
Pdom	*Polistes dominula*
Mrot	*Megachile rotunda*
Aflo	*Apis florea*
Dqua	*Dinoponera quadriceps*
Orug	*Odontomachus ruginodis*
Aech	*Acromyrmex echinatior*
Sinv	*Solenopsis invicta*
Obir	*Ooceraea biroi*
Cpla	*Camponotus planatus*
Hsal	*Harpegnathos saltator*

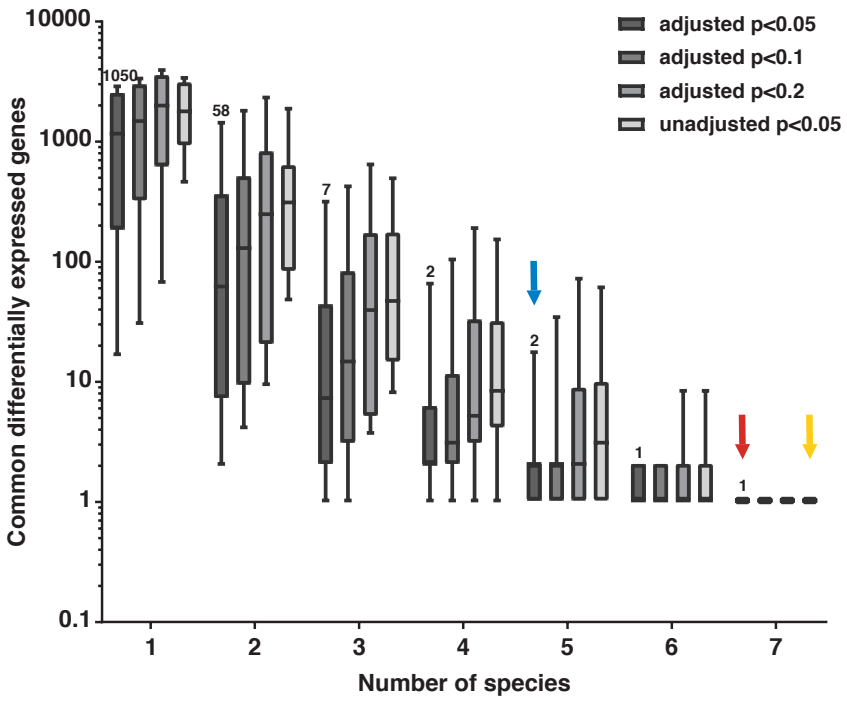


Figure S1: *ilp2* is the single most robust concordant differentially expressed gene. A single gene (*ilp2*) is concordantly differentially expressed across all seven species (red arrow). This finding is robust to relaxing significance thresholds, including foregoing FDR correction entirely (yellow arrow). The screen recovers additional genes only when the number of included species is restricted to certain subsets. For subsets of five species, a total of 24 concordant differentially expressed genes are recovered across comparisons (blue arrow; Fig. S2; Table S3). The median number of differentially expressed genes for the comparison with adjusted $p < 0.05$ is shown above each box in black.

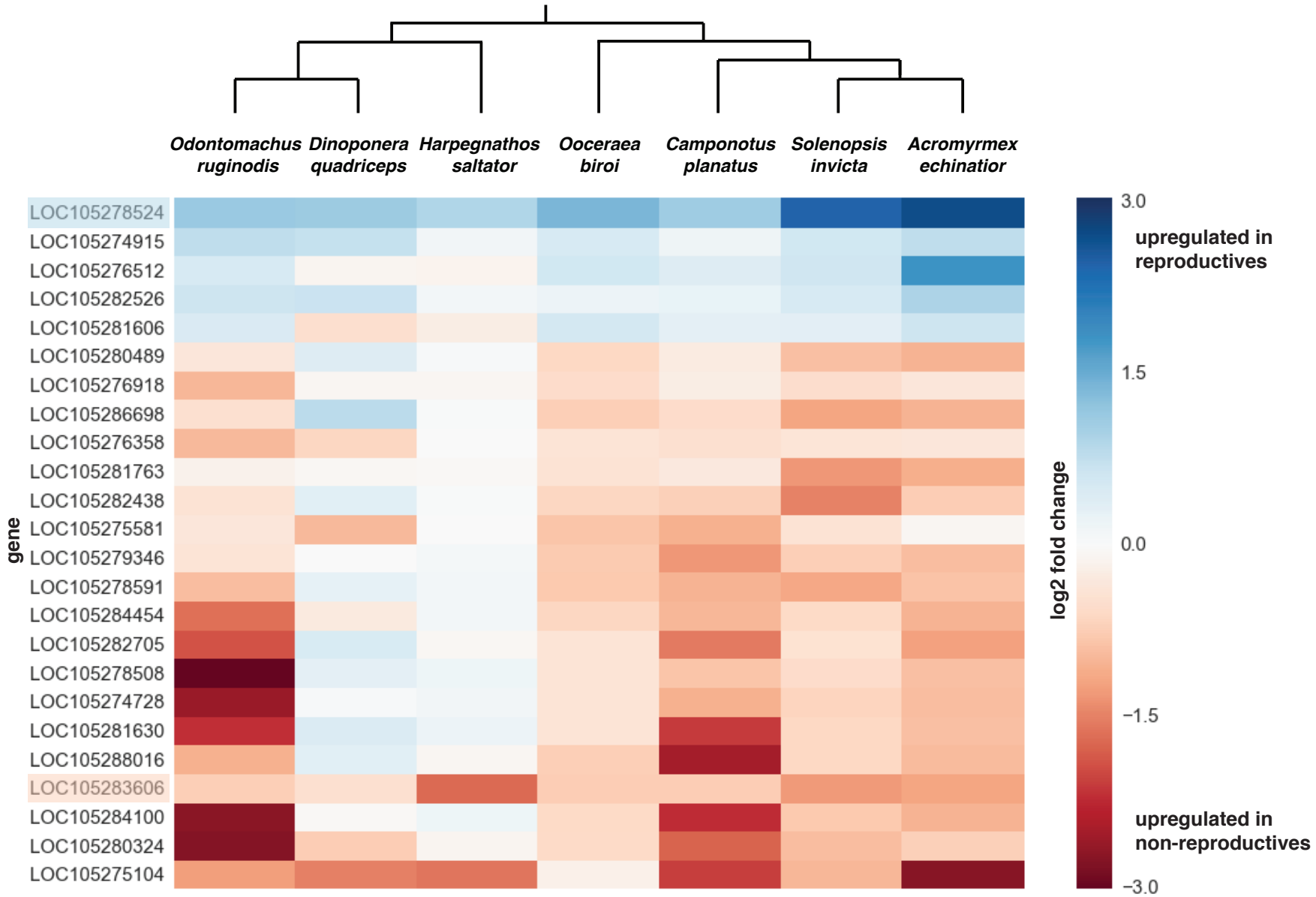


Figure S2: Heatmap showing log2 fold changes of the 24 genes that are concordantly differentially expressed in at least five ant species. *O. biroi* gene identifiers are given on the left (see Table S3). *ilp2* is highlighted in blue and *ilp1* is highlighted in orange.

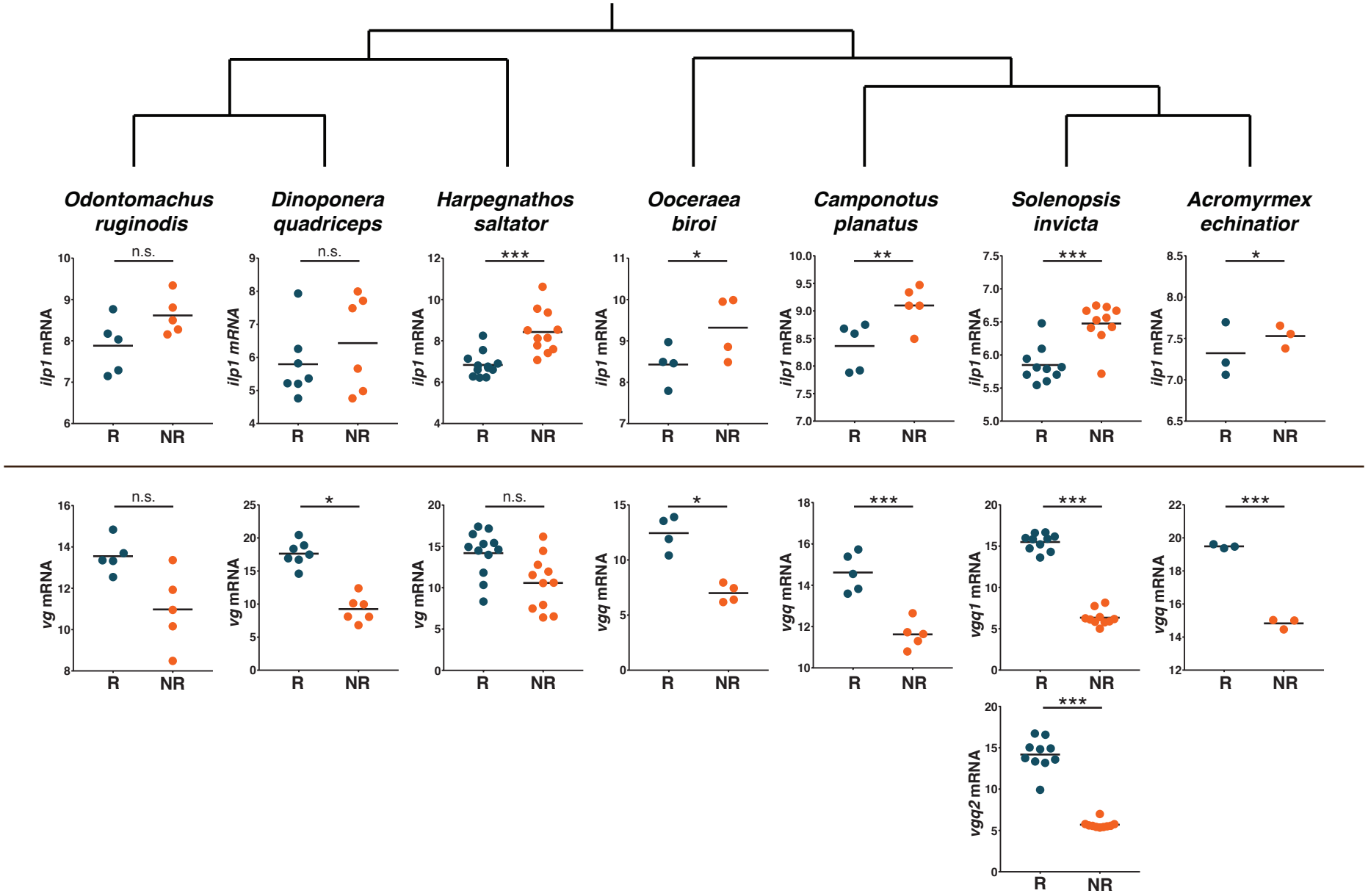


Figure S3: VST read counts for *ilp1* and vitellogenin (*vg*) in the seven ant species in the comparative screen. *ilp1* always tends to have higher expression in the non-reproductives, although it is not significantly differentially expressed in two ponerine species: *O. ruginodis* and *D. quadriceps*. *vg* is duplicated in the formicoids, and here we only show its pro-ortholog, queen vitellogenin (*vgq*), as it is the formicoid *vg* paralog that is associated with reproduction (13). This duplication was likely followed by a loss in the formicines and, as a result, *C. planatus* only appears to have one *vg*. *vgq* is duplicated in *S. invicta*. While the average expression of *vgq* is higher in reproductives of all seven species, this difference is not statistically significant in two ponerine species: *H. saltator* and *O. ruginodis*. Horizontal bars indicate means and asterisks indicate statistically significant differences between groups (Wald test with 5% FDR correction; * $p<0.05$; ** $p<0.01$; *** $p<0.001$; n.s.: not significant).

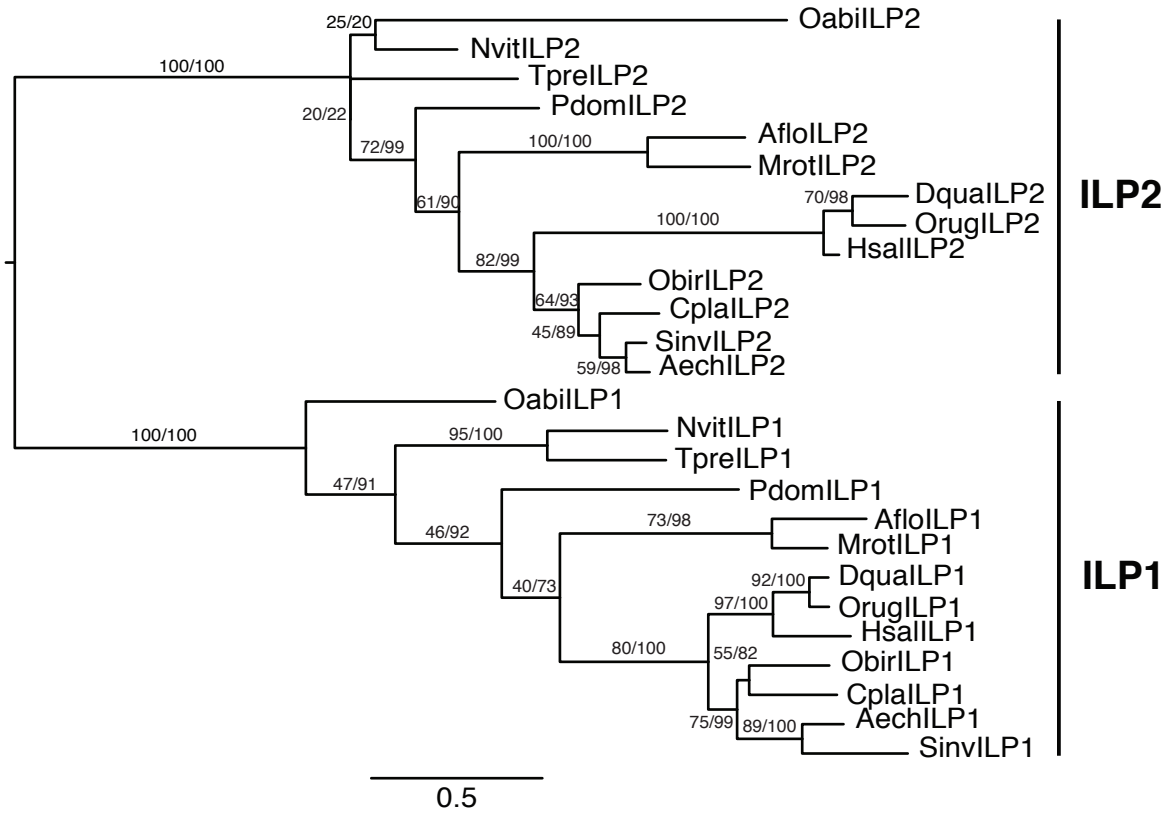
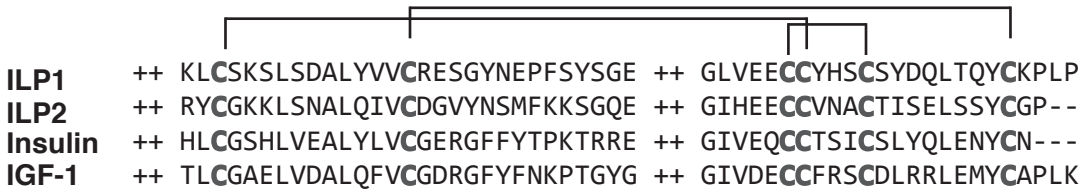
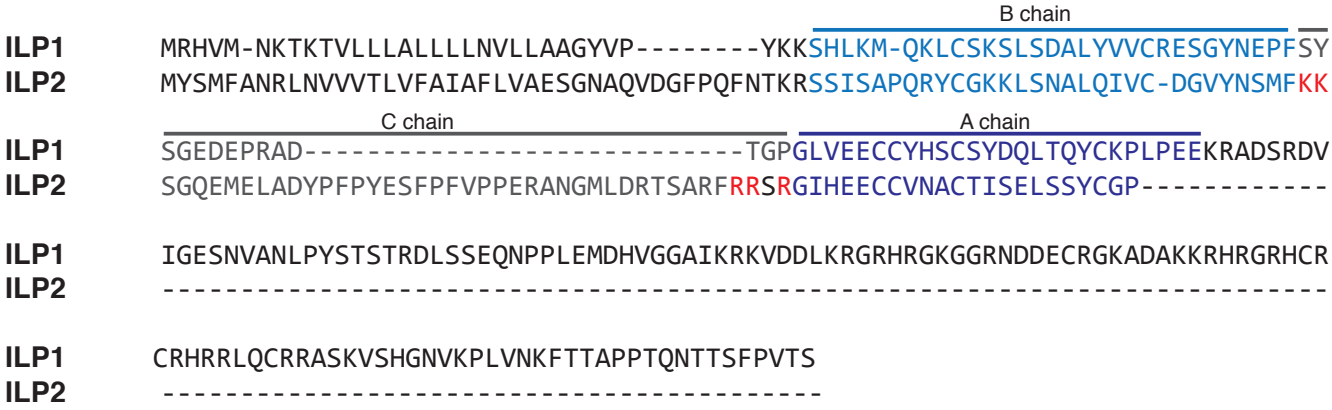


Figure S4: Maximum likelihood phylogeny of insulin-like peptides across the Hymenoptera identifies two clades representing the two ant ILPs, ILP1 and ILP2. Bootstrap support / Bayesian posterior probabilities are shown next to corresponding nodes. See Table S3 for species abbreviations. The scale bar depicts the expected number of substitutions per site.

A



B



C

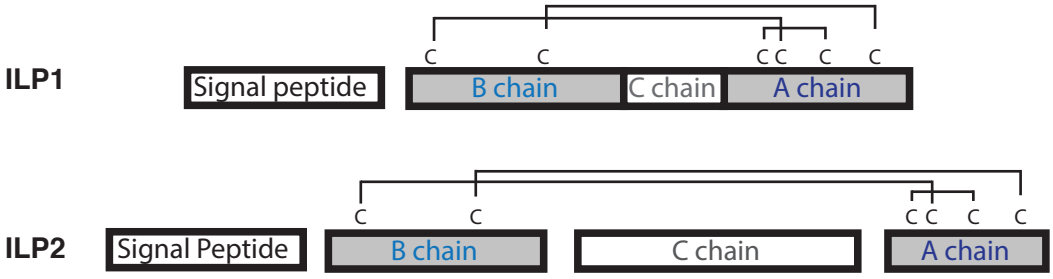


Figure S5: Predicted sequences and structures of ILP1 and ILP2. (**A**) Sequence alignments of *O. biroi* ILPs with human insulin and IGF-1, with the conserved cysteine residues colored in gray. Connecting lines represent predicted disulfide bonds. (**B**) Full sequence alignment of *O. biroi* ILPs. Potential cleavage sites at dibasic and/or monobasic residues are colored in red. The predicted B chains are colored in light blue, A chains in dark blue, and C chains in gray. (**C**) Diagram of predicted ILP1 and ILP2 structures that show a single chain ILP1 peptide with a short C chain and a cleaved ILP2 peptide with no C chain in the mature peptide.

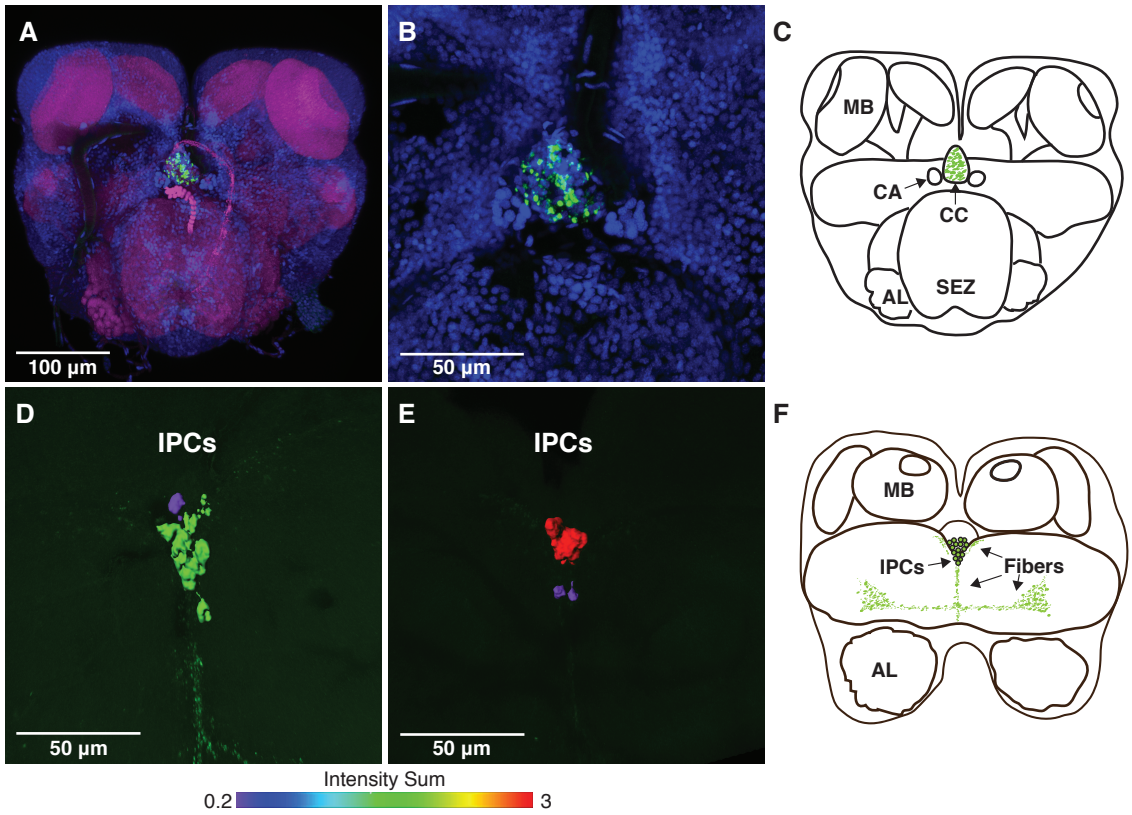


Figure S6: ILP2 localization and quantification in *O. biroi* brains. (**A-C**) Immunohistochemistry with anti-ILP2 antibody on an *O. biroi* brain shows that ILP2 is sequestered in the corpora cardiaca, a neurosecretory organ that lies adjacent to the esophagus (body-axis ventral view). SEZ: sub-esophageal zone; CA: corpora allata; CC: corpora cardiaca. Colors in (**A**) and (**B**) are as described in Fig. 2. (**D-E**) Semi-automatic segmentation of the insulin-producing cells. Volumes are colored by ILP2 fluorescence intensity (see scale bar below images). These images represent data points closest to the mean values for ants in the (**D**) reproductive and (**E**) brood care phase, respectively. (**F**) Schematic of the brain (reproduced from **Fig. 2C**) showing the localization of the insulin-producing cells (body-axis dorsal view).

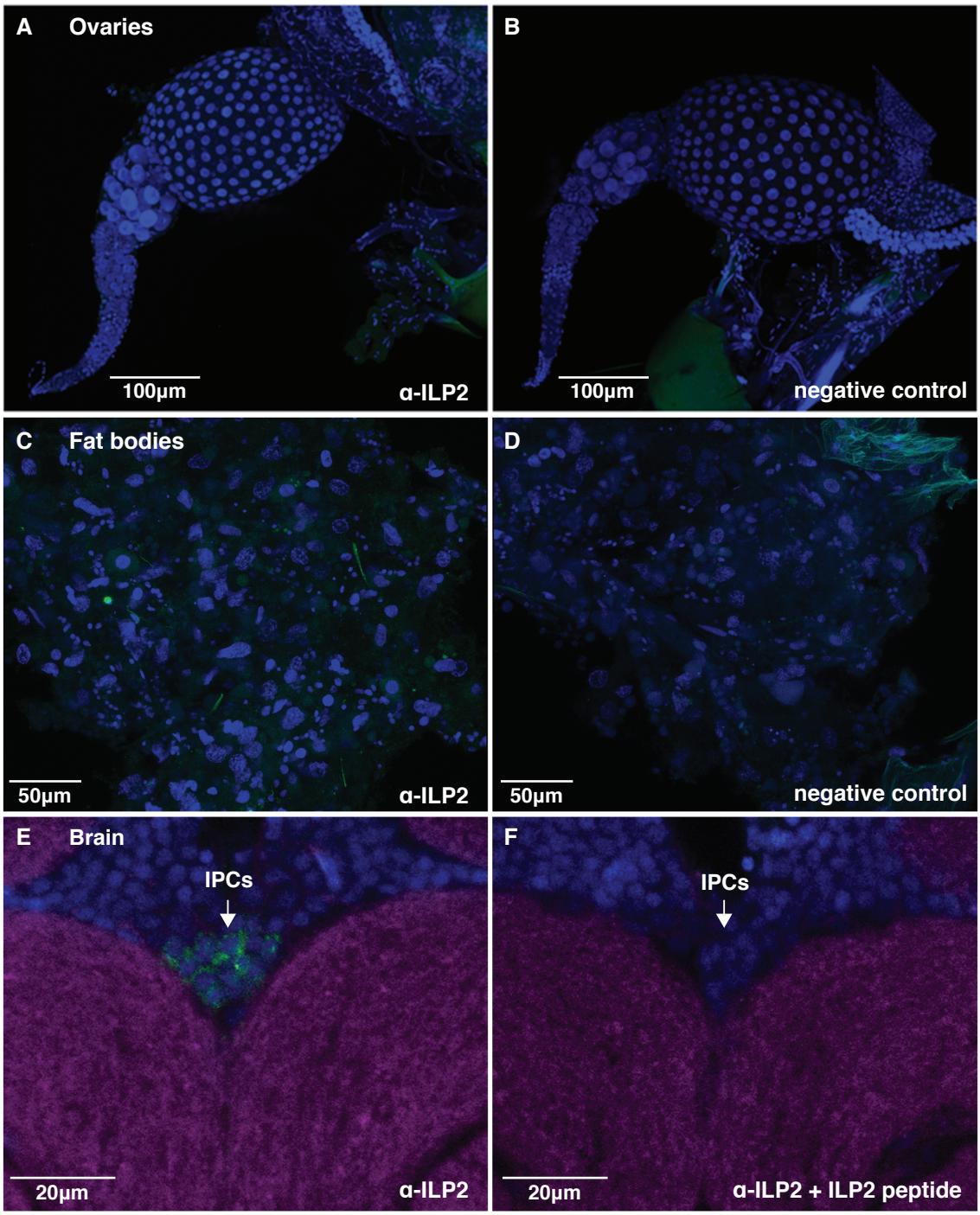


Figure S7: Additional ILP2 and control staining. (**A**) No ILP2 staining is detected in ovaries. (**B**) Negative control (no primary antibody). The green fluorescence in these images is cuticular autofluorescence.(**C**) ILP2 staining is also absent in fat bodies. (**D**) Negative control (no primary antibody). Note that the green speckles are background noise present in preparations with (**C**) and without (**D**) primary antibody. ILP2 antibody specificity in the brain was confirmed by preincubating the antibody with the ILP2 peptide. (**E**) ILP2 staining in the insulin-producing cells in the ant brain. (**F**) Lack of staining in the insulin-producing cells following preincubation of the ILP2 antibody with ILP2 peptide demonstrates that the antibody binds to ILP2 specifically. The colors in the image indicate DAPI (blue), anti-ILP2 (green), and phalloidin (magenta).

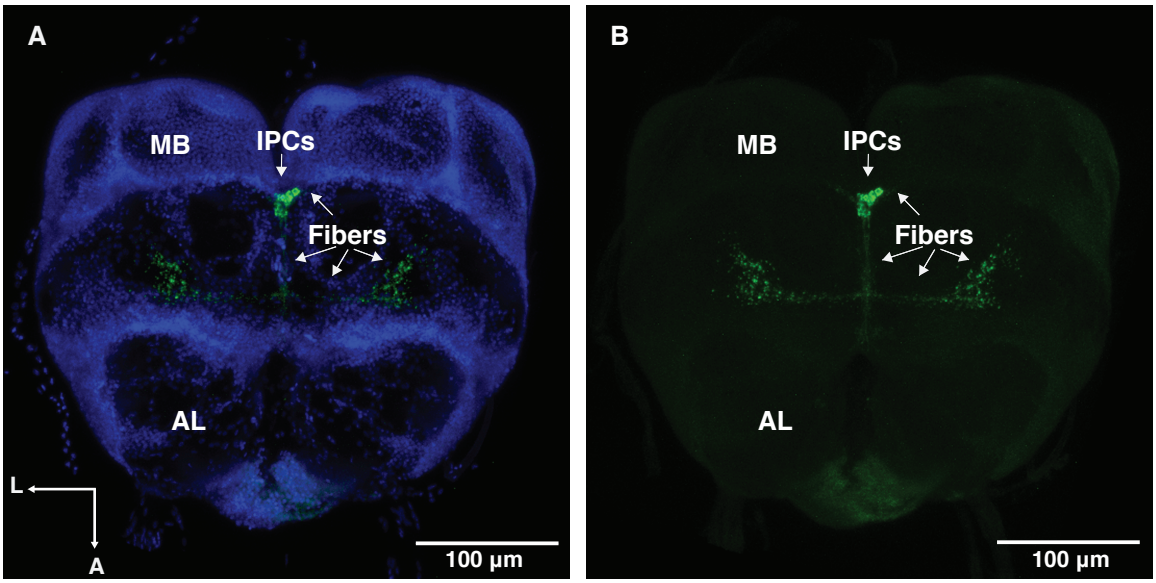


Figure S8: Projections of insulin-producing cells. (**A-B**) Anti-ILP2 immunostain shows that the insulin-producing cells project laterally and anteriorly to the body axis. The colors in the images indicate DAPI (blue) and anti-ILP2 (green). In the axis (left panel, bottom left), A: body-axis anterior, L: lateral. (**B**) Same image as in (**A**) but with DAPI staining removed to highlight the projections more clearly.

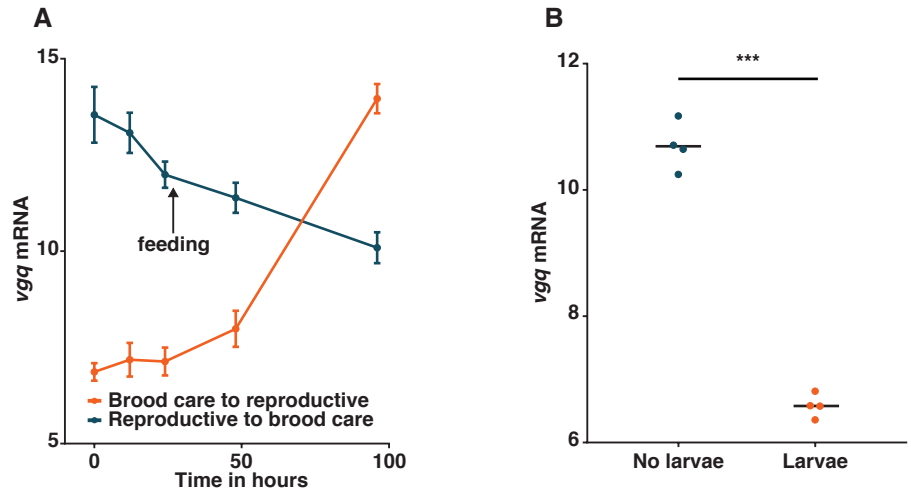


Figure S9: *vgq* expression is suppressed by larval signals. (**A**) Brain RNA-Seq time course shows that the addition of larvae downregulates *vgq*, while the removal of larvae upregulates *vgq* (time:transition interaction, Likelihood Ratio Test with 5% FDR correction; $p<10^{-15}$). (**B**) Brain RNA-Seq shows that *vgq* is upregulated eight days after larvae are removed from *O. biroi* workers in the brood care phase under nutritionally controlled conditions (n=4, Wald test with 5% FDR correction; $p<10^{-13}$). The data are variance-stabilized and transformed read counts. Horizontal bars in (**B**) indicate means.

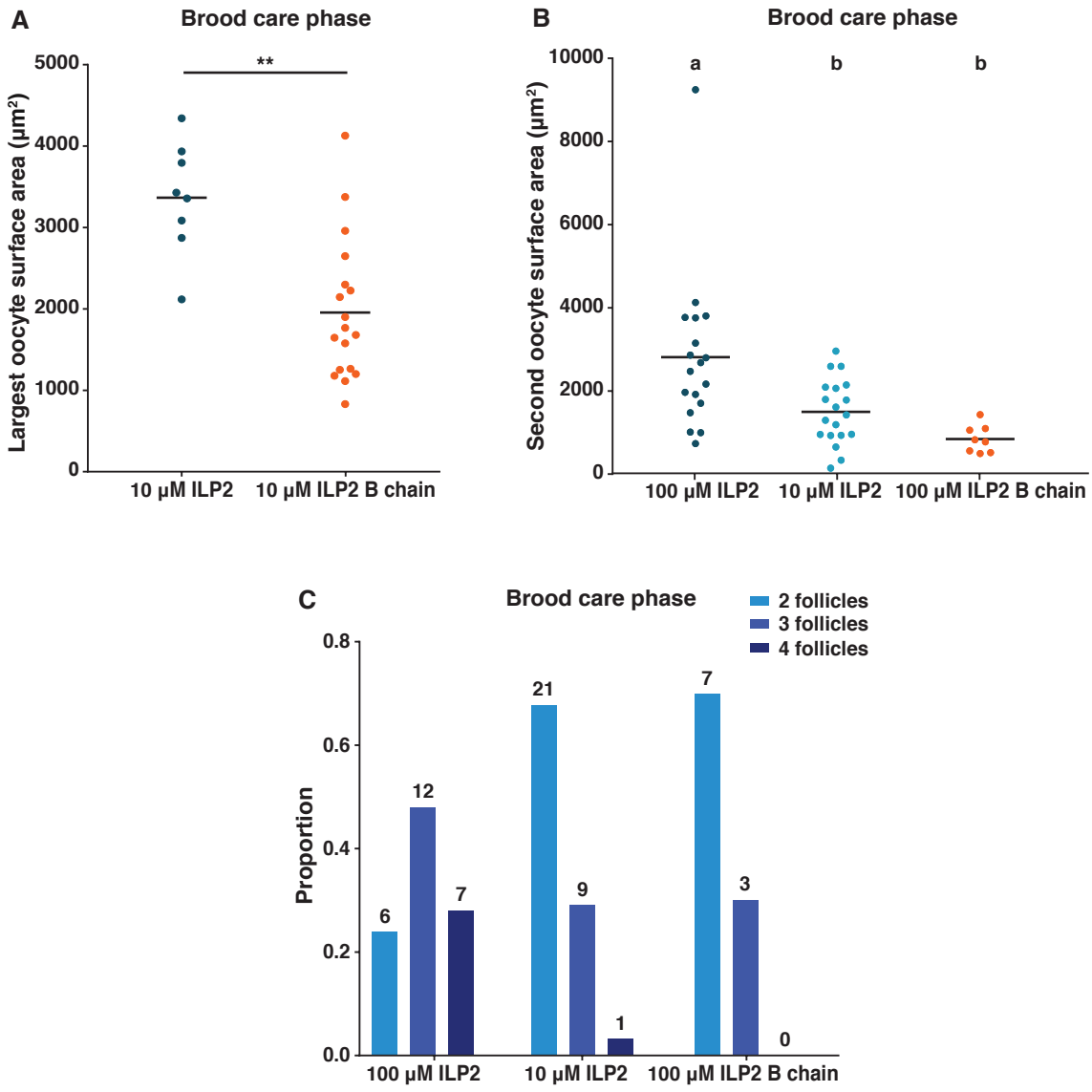


Figure S10: ILP2 supplementation overrides larval suppression of adult reproduction. (**A**) 10µM ILP2 causes ants in the brood care phase to activate their ovaries relative to control workers injected with 10µM B chain, despite being in the presence of larvae ($n \geq 8$, Welch's t-test with Bonferroni correction (related data in Fig. S11A); $p=0.002$). (**B**) 100µM ILP2 injections also cause the smaller of two ovarioles to have more developed oocytes than injections with 10µM ILP2 or B chain control ($n \geq 8$, letters above the columns indicate significant differences at $p<0.05$ after a Kruskal-Wallis test with Dunn's correction on post-hoc pairwise comparisons). Horizontal bars indicate means on all dot plots. (**C**) The amount of ILP2 injected alters the number of eggs an ant can develop simultaneously (chi-squared test; $p=0.0032$). Y-axis shows proportion of ants with 2, 3, or 4 follicles in each treatment. Numbers above each bar indicate sample size.

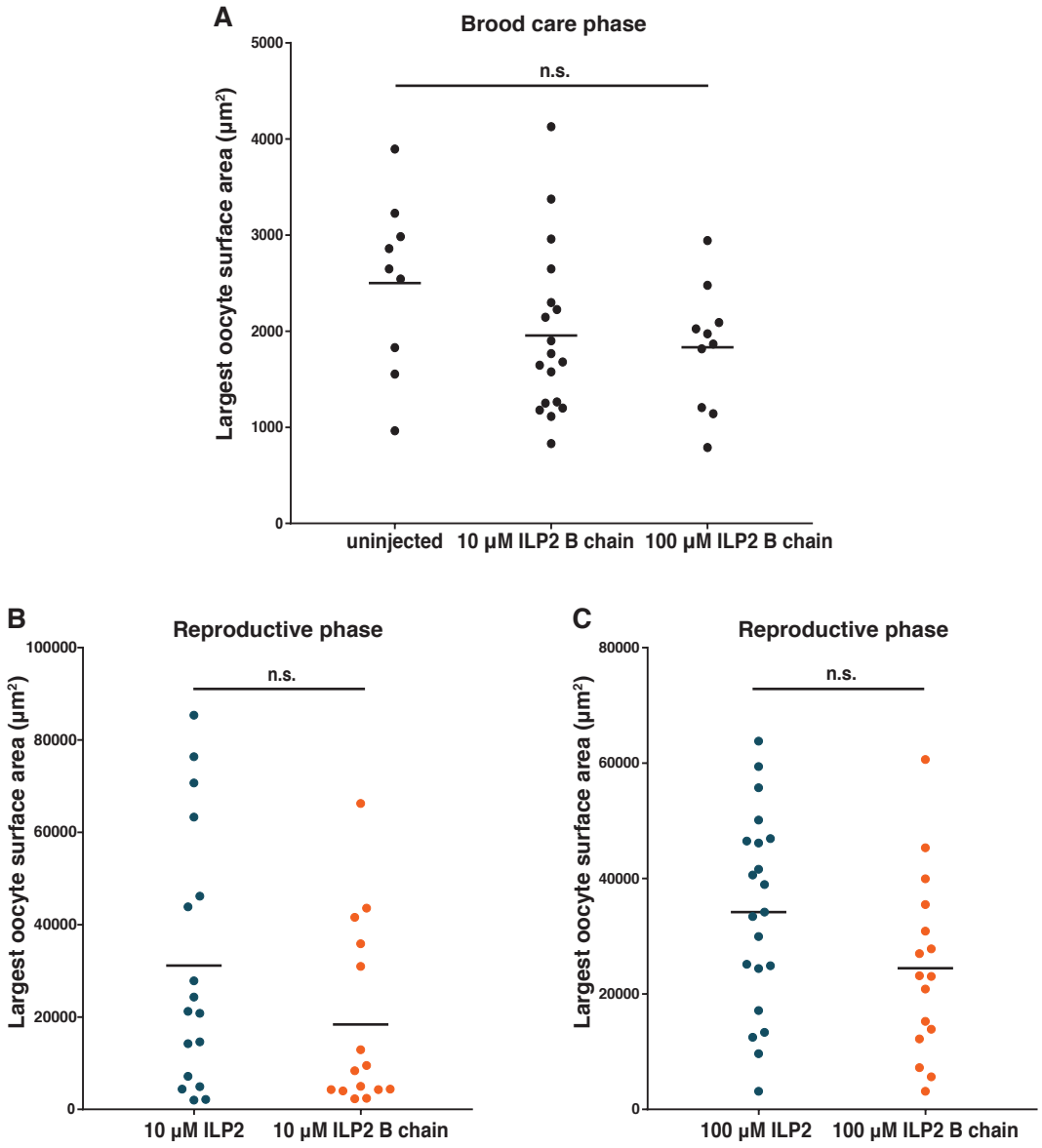


Figure S11: ILP2 injection controls. (**A**) Ants in the brood care phase injected with ILP2 B chain have oocytes that are indistinguishable from uninjected ants (n≥9, pairwise t-tests followed by Bonferroni correction: p>0.05), showing that the B chain is an appropriate negative control. The data for the B chain injections are the same as those in Fig. 4A and Fig. S10A. Injecting ILP2 at 10 µM (n≥15, Welch's t-test; p=0.14) (**B**) and 100 µM (n≥16, Welch's t-test; p=0.08) (**C**) does not significantly further activate ovaries in the reproductive phase, likely because these ants already have levels of ILP2 permissive to ovary activation and their ovaries are already active.

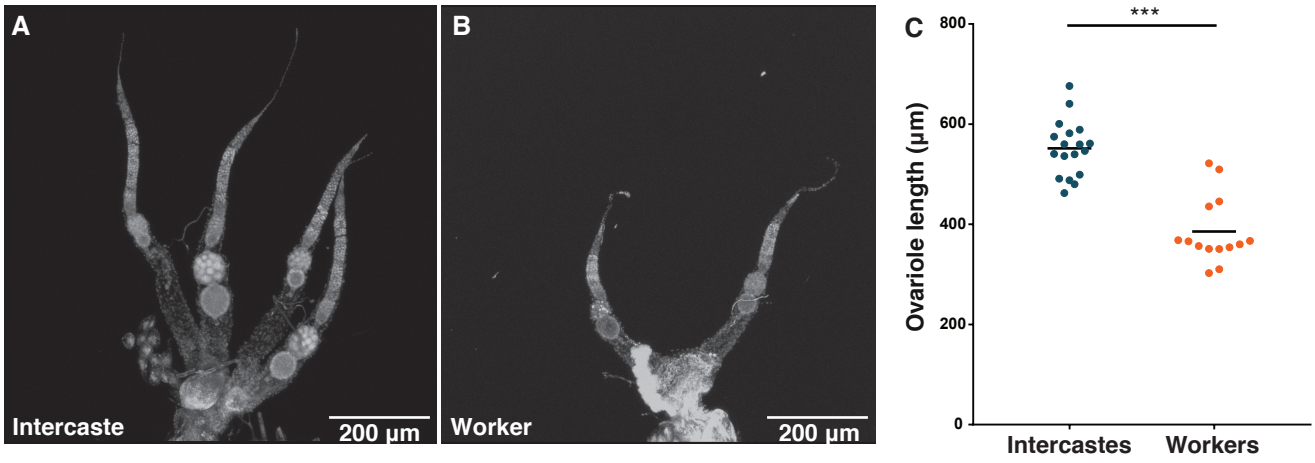


Figure S12: Intercastes have larger ovaries than regular workers. (**A**) Intercastes tend to have four ovarioles, while (**B**) regular workers tend to have two. In the brood care phase, intercaste ovarioles are also more developed (Fig. 4A). (**C**) Quantification of ovariole length between young phase-matched intercastes and regular workers in the brood care phase shows that intercastes have longer ovarioles ($n \geq 14$, Welch's t-test; $p < 0.0001$).

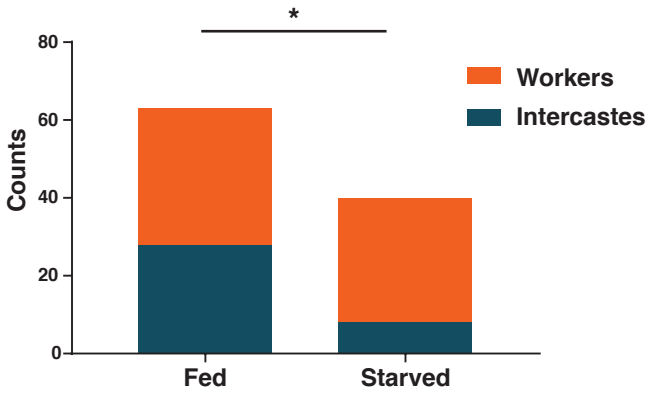


Figure S13: When developing larvae are starved, they are less likely to grow into intercastes. The y-axis here shows the number of intercastes and workers per treatment, summed across five replicate colonies. Fisher's exact test p=0.0120.

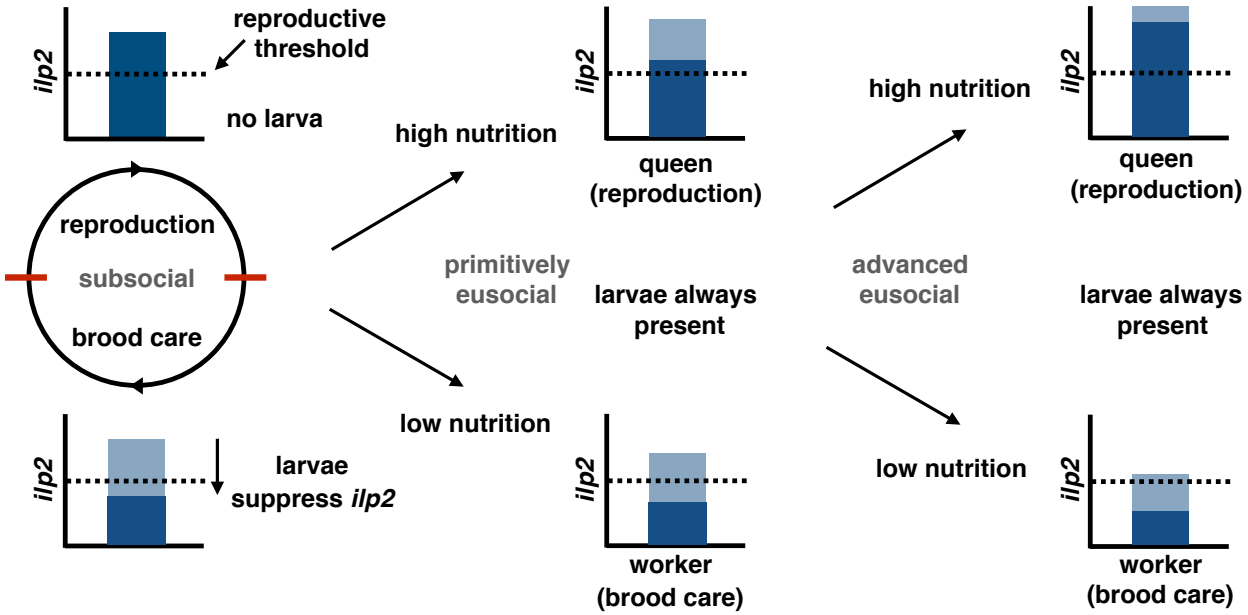


Figure S14: Schematic illustrating the putative origin and elaboration of eusociality from a subsocial cycle in ants. The dark bar in each bar graph illustrates realized *ilp2* expression, whereas the light bar shows maximum potential *ilp2* expression. *ilp2* expression in adults is reduced in the presence of larvae, thereby explaining the ancestral subsocial cycle. In the primitively eusocial state, queens have slightly higher maximum potential *ilp2* levels than workers do, and larval signals (or similar inhibitors of reproduction such as aggression or queen pheromones) amplify this developmentally determined reproductive asymmetry. This allows queens to reproduce permanently instead of cyclically, and workers are permanently inhibited from reproducing. Over evolutionary time, this reproductive DOL may be further elaborated, as seen in the advanced eusocial condition.