Supplementary Methods

Data generation

<u>Sample collections</u>
Details of all samples are provided in Supplementary Table 8.

<u>Genomic DNA extraction</u>
We used Qiagen Genomic Tip (Qiagen, USA, Catalog #10223) kits to extract HMW DNA from bee samples with minor modifications. First, halictids exceed the maximum suggested mass for the kit so individual bees were split into two digestions, one that included both the head and thorax and the other with just the abdomen. For each digestion, bee parts were gently ground with a pestle over dry ice before adding 350 ul of buffer ATL and 4 ul of RNAse A followed by incubation at 37°C for 30 minutes. 50ul of proteinase K and 1 mL of G2 buffer were then added and the samples incubated overnight at 50 C with occasional mixing. The standard genomic tip protocol was subsequently followed, combining multiple digestions from the same sample onto the same column when applicable. Centrifugations were done at 12,000g for 30 and 15 minutes and DNA was eluted in 50 ul of TE.

<u>HiC sequencing</u>
For Hi-C scaffolding of the 10X drafts, *in situ* Hi-C libraries were generated as previously described[1]. Specifically, head and thorax tissue from a single bee for each species (with the exception of *H. quadricinctus,* for which a single abdomen was used) was crosslinked and then lysed with nuclei permeabilized but still intact. DNA was then restricted and the overhangs filled in incorporating a biotinylated base. Free ends were then ligated together *in situ*. Crosslinks were reversed, the DNA was sheared to 300-500 bp and then biotinylated ligation junctions were recovered with streptavidin beads. This was followed by a standard Illumina library construction protocol. Briefly, DNA was end-repaired using a combination of T4 DNA polymerase, *E. coli* DNA Pol I large fragment (Klenow polymerase) and T4 polynucleotide kinase. The blunt, phosphorylated ends were treated with Klenow fragment (3' to 5' exo minus) and dATP to yield a protruding 3- 'A' base for ligation of Illumina's adapters which have a single 'T' base overhang at the 3' end. After adapter ligation DNA was PCR amplified with Illumina primers for 8-12 cycles and library fragments of 400-600 bp (insert plus adaptor and PCR primer sequences) were purified using SPRI beads. The purified DNA was captured on an Illumina flow cell for cluster generation. Libraries were sequenced on the Illumina sequencer following the manufacturer's protocols.

<u>Whole-genome resequencing</u>
For whole genome resequencing, we used the standard Qiagen DNEASY (Qiagen, USA, catalog #69504) protocol to extract DNA, eluting in 100 ul of buffer AE. Libraries were constructed using the PCR-free NuGEN Ovation Rapid DNA (TECAN, USA, catalog #0328-96) library prep kit and sequenced on the Illumina HiSeq4000 or HiSeq2500 with 100 base paired-end reads.

<u>Genome assembly</u>
We built 10x Genomics linked-read libraries for a single individual from each species and, in general, each library was sequenced on approximately half a lane of the Illumina HiSeq X platform with 150 base paired-end reads. *Lasioglossum vierecki* is quite small so to obtain sufficient DNA, we pooled two males together for extraction. For *Augochlorella aurata*, two separate libraries were constructed, each from a different individual. One of these libraries was sequenced on a half-lane and the other was sequenced on a full lane. *Lasioglossum figueresi* was sequenced on two, half-lanes, and *Halictus rubicundus* and *Agapostemon virescens* were each sequenced on a single full lane. While most samples used were diploid females, we used single haploid males for *A. aurata, L. oenotherae, L. pauxillum,* and *L. leucozonium.* A single additional individual from each species was also collected for Hi-C library preparation except for *H. quadricinctus,* which is large enough that the same bee was split for Hi-C (abdomen) and 10x (head).

<u>Transcriptome sequencing for genome annotation</u>

For RNA-sequencing, heads, abdomens, and most larvae and pupae were extracted using Zymo Quick-RNA extraction kits (Zymo, USA, cat. #R1057). For antennae, Dufour's glands, and very small larvae, the ThermoFisher PicoPure RNA Isolation Kit (ThermoFisher, USA, cat. #KIT0204) was used. Libraries were made using the NEBNext Ultra Directional RNA Library Prep Kit (NEB, USA, cat. #E7760) with Poly-A bead enrichment and were sequenced on two lanes of the Illumina HiSeq4000 with 100 base paired-end reads. Antennal libraries from *L. zephyrus* and *L. oenotherae* as well as all Dufour's gland libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit (NEB, USA, cat. #E7770) with Poly-A bead enrichment and were sequenced on one Illumina HiSeq4000 lane with 150 base paired-end reads.

We collected transcriptomic data from 11 species (*Ag. virescens, Augochlora pura, Augochlorella aurata, H. ligatus, L. calceatum, L. figueresi, L. leucozonium, L. malachurum, L. marginatum, L. pauxillum,* and *L. vierecki*). For each species, we constructed separate RNA-seq libraries from whole abdomens, whole heads, and antennae, and each library was sequenced to a depth of approximately 10 million reads. For *L. pauxillum* and *L. calceatum* we also sequenced larvae and pupae. Antennae were additionally sequenced from *L. zephyrus* and *L. oenotherae*, and additional antennal libraries were prepared for *A. pura* and *L. marginatum.* Dufour's glands were extracted and sequenced from a total of 12 species (*A. aurata, A. pura, L. leucozonium, L. marginatum, L. vierecki, L. zephyrus, L. calceatum, L. albipes, L. malachurum, L. oenotherae, L. pauxillum,* and *L. figueresi*). These four antennal and all Dufour's gland libraries were sequenced to a depth of approximately 17 million reads.

Brain single-cell RNA-Sequencing: Brain dissection and cell recovery

Four single-cell RNA-Sequencing (scRNAseq) experiments were prepared from whole brains of *Lasioglossum zephyrus* and *Lasiogossum albipes*. For *L. zephyrus*, three queen or worker brains were dissected and pooled per sample, and for *L. albipes* a single worker brain was used for each of two samples. We followed the "short dissociation protocol" from[2]. Briefly, brains were dissected in ice-cold Dulbecco's PBS (DPBS) and transferred to 500 μl DPBS containing 0.05% Trypsin-EDTA. Manual cell dissociation was performed via shaking at 25°C, 1000 rpm for 20 min with gentle trituration every 5 min. Cells were then washed with cold DPBS and resuspended in 400 μl of DPBS with 0.04% BSA and passed through a 40 um Flowmi® cell strainer. Cell counting was performed using propidium iodide-based sorting on the MoxiGO system (Orflo Technologies, Ketchum, ID USA).

miRNA isolation and sequencing

miRNAs were isolated from flash-frozen bee brains using the mirVana miRNA isolation kit (AM1560, ThermoFisher). Libraries were constructed using NEBNext Small RNA Library Prep Kit (E7330S, NEB) and sequenced on an Illumina HiSeq, single-end 50 bp reads.

JH extraction and LC-MS

Bee brains and other tissues were flash frozen in liquid nitrogen, then ground and centrifuged before adding 200 µl of extraction solvent (40:40:20 methanol:acetonitrile:H2O with 0.5% formic acid[3]) supplemented with 0.2 µg/ml valine-D8 as an internal standard and 16.8 µl of 15% ammonium bicarbonate. LC-MS was carried out following previously described protocols[4]. LC was performed using a Xbridge BEH amide HILIC column (Waters) with a Vanquish UHPLC system (Thermo Fisher). Solvent A was 95:5 water: acetonitrile with 20 mM ammonium acetate and 20 mM ammonium hydroxide at pH 9.4. Solvent B was acetonitrile. The gradient used for metabolite separation was 0 min, 90% B; 2 min, 90% B; 3 min, 75%; 7 min, 75% B; 8 min, 70%, 9 min, 70% B; 10 min, 50% B; 12 min, 50% B; 13 min, 25% B; 14 min, 25% B; 16 min, 0% B, 21 min, 0% B; 21 min, 90% B; 25 min, 90% B. MS analysis was performed using a Q-Exactive Plus mass spectrometer (Thermo Fisher) in positive ionization mode. An m/z range of 70 to 1000 was scanned. Measured JH in *Bombus impatiens* is presented in Extended Data Figs. 8-9. Chemical structure of JH III, parent and fragment m/z values for labeled and unlabeled forms, chromatograms and absolute quantification of JH III and JH III-D3 are shown in Extended Data Fig. 10. Note that the corpora allata are not known to store JH[5].

Stable isotope tracing

Bumblebee (*Bombus impatiens*) workers were collected from a queenright colony (Koppert Biological Systems, Natupol Excel) and isolated for 24 hours. *Augochlorella aurata* foundresses were collected in May 2022 from Princeton, NJ and were fed sucrose solution overnight in isolation. Following isolation, bees were

chill anesthetized until immobile, and then treated with 50ug (*B. impatiens*) or 25ug (*A. aurata*) of deuterated JH III (JH III-D3) (Toronto Research Chemicals, E589402) dissolved in acetone or acetone alone as a control. Treatments were applied by pipetting 1-2 ul of 25 ug/ul JH III-D3 onto the abdominal sternites. Bees remained on an ice pack until the treatments were absorbed and dry. Then, they were returned to isolated tubes for 2, 4, 8, 16, 24, or 72 hours until frozen. For LC-MS of *B. impatiens*, bees were chill anesthetized and then decapitated. For LC-MS of *A. aurata,* heads were lyophilized at -80°C and 300 mTorr for 30 minutes, after which brains were dissected and stored at -80°C until analysis.

Data Analysis

<u>Genome</u> <u>assembly</u>
Initial assemblies showed the presence of substantial numbers of Illumina adapter sequences. We therefore performed stringent adapter removal. Adapter trimming was performed by running cutadapt v1.11 [6] twice, both times using the adapter sequences AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT, AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT, and TGTAGATCTCGGTGGTCGCCGTATCATT. The first time, the defaults were used to only remove three base matches to adapters from the ends of both the forward and reverse reads (-a, -A). The second time, adapters were removed from within reads (-b, -B), requiring matches of at least four bases. Reads with any sign of adapter sequence were removed from the dataset rather than being trimmed.

Assemblies on the resulting adapter-filtered data were then performed using Supernova version 1.0.0 [7]. Note that we attempted to use later versions but, while successful on some species, the program crashed on others so we settled on 1.0.0 for the sake of consistency. Each species was assembled three times using random subsets of 105 million read pairs. Initial explorations of the coverage needed using *L. albipes* suggested that ~105 million reads maximized assembly quality metrics. Genome files were then generated using the mkfasta command with style pseudohap and minimum scaffold size of 1,000 bases. These three assemblies were then sequentially merged using quickmerge[8] in every possible order of all three and also subsets of just two assemblies. This yielded a complete set of 27 assemblies, including the original three. Each assembly was then evaluated based on continuity statistics (e.g. N50 lengths and numbers), as well as completeness as determined by BUSCO2[9–11] using *Apis mellifera* as the seed species and the set of 4,415 Hymenoptera genes from OrthoDB v9. Gaps were then closed with the sealer module of ABySS 2.0[12] using the settings -B 3000, -G 1000, -P 10, though for certain genomes, these parameters had to be decreased to keep the runtime under reasonable limits.

L. figueresi has a larger genome than other *Lasioglossum* so the 10x library was sequenced on two half lanes. The data were filtered as for other genomes and then all data were used to make the assembly. No random subsets were used and no merging was performed as it did not improve the assembly. Similarly, gapclosing with sealer was not performed for this genome because it did not substantially alter the resulting assembly.

Augochlorella aurata also has a much larger genome than most other species examined. Two 10x libraries from one haploid male and one diploid female were constructed. One was sequenced on a half lane of Illumina HiSeqX and the other on a full lane of HiSeqX. All data was pooled for analyses. Assemblies using the approach outlined above were of poor quality. FASTQC v0.11[13] analysis of the raw data revealed that the libraries were highly GC-biased with overrepresentation of tandem duplications of the motif ATGGGCT. We therefore filtered 99% of reads containing two tandem sequences of this motif. We kept 1% of these reads assuming that they were being overamplified in the library but still represented real sequence in the genome. This procedure substantially improved the contiguity of the assembly but we suspected that additional data may be present in the 10x libraries that may have been filtered out before assembly. We therefore also ran fragScaff[14] using a half lane of one of the sequencing libraries. We aligned the data with LongRanger[15] and then performed fragScaff analysis with the parameters -b 1 -E 3000 and -G R which approximately doubled the N90 size of the assembly.

A similar procedure removing overrepresented motifs was also required for *H. rubicundus* (motif TATCCTATCC)*, H. quadricinctus* (motif TATCCTATCC)*,* and *Ag. virescens* (motif TAACCCAGACC). Additional scaffolding with fragScaff was performed for *H. rubicundus* but not the other two species as it was not found to improve the assembly.

Hi-C scaffolding

We sought to perform Hi-C scaffolding to error-correct, order and orient as well as anchor the draft genomic sequences to chromosomes[16,17]. For this, candidate chromosome-length genome assemblies were generated using 3D-DNA[17] followed by additional finishing using Juicebox Assembly Tools[17]. The resulting contact matrices as well as matrices corresponding to the draft assembly, before the Hi-C scaffolding, are available on www.dnazoo.org.

Transcriptome assembly

We generated three directional RNAseq libraries from a single individual per species. Tissues were: antennae, heads, and abdomens from *A. virescens, L. figueresi, A. pura, L. marginatum, A. aurata, L. leucozonium, H. ligatus, L. pauxillum, L. vierecki, L. malachurum,* and *L. calceatum.* For each species, multiple libraries were combined and assembled using Trinity[18] with the –trimmomatic, --SS_lib_type RF and –normalize_reads options. These libraries were also assembled using the genome-guided approach by first mapping to the genome assemblies using HISAT2 v2.0.5[19] and setting a maximum intron length of 100,000 bases.

The non-directional libraries from the antennae of additional individuals from *A. pura, L. marginatum, L. zephyrus,* and *L. oenotherae* and the Dufuor's glands from *A. aurata, A. pura, L. leucozonium, L. marginatum, L. vierecki, L. zephyrus, L. calceatum, L. albipes, L. malachurum, L. oenotherae, L. pauxillum,* and *L. figueresi* were assembled separately from the directional libraries using both *de novo* and genome-guided Trinity and the –trimmomatic and –normalize_reads options.

For *L. albipes*, we assembled two datasets in addition to the Dufuor's gland. First, we took the RNAseq data from an existing *L. albipes* paper[20] and reassembled it using *de novo* and genome-guided Trinity. We also single-end sequenced two directional libraries constructed from the head and antennae of a single individual.

Normalized gene expression data for each species and tissue type is provided in Supplementary Table 7.

Filtering bacterial scaffolds

Initial genome assemblies included varying numbers of scaffolds of likely bacterial origin. We filtered these by comparing scaffolds to previously sequenced genomes. First, the database of complete bacterial genomes from NCBI was downloaded on February 7, 2017. We then used BLASTN to compare all assembled scaffolds to the bacterial genome database using default parameters. Significant hits to bacterial genomes were merged if they were within 10 bases of each other. If more than 25% of the length of a scaffold had a significant hit to a bacterial genome or more than 10% of the length of the scaffold in more than 10 segments had a significant hit to a bacterial genome then the scaffold was extracted. These sequences were then compared with a separately downloaded full set of animal genomes from NCBI using BLASTN, again with default parameters. If a region had a higher percent similarity to bacterial genomes than animal genomes then that scaffold was considered of bacterial origin and was removed from the assembly prior to HiC scaffolding and annotation.

Repeat masking

In order to obtain high quality genome annotations, we first characterized the repetitive elements present in each individual genome using RepeatModeler v1.0.7[21]. We removed redundancy from the resulting set of repetitive elements by clustering those with 80% or greater similarity using CD-HIT v4.6[22]. We compared the resulting repeat elements to all UniProt proteins and *Drosophila melanogaster* proteins using BLASTX. Putative repetitive elements with bitscores of at least 100 or more than 50% similarity over 50% of the length of either the protein or repeat sequence were removed from the set of repetitive elements so as to avoid masking protein sequences in the genome. Only repetitive elements longer than 80 bases were retained. The resulting set of repeat elements was combined with the repetitive elements for Arthropods available from RepBase (downloaded from https://www.girinst.org/ on March 8, 2017) and then used to mask each genome.

Identification of conserved non-exonic elements (CNEEs)

Repeat-masked genomes of each of the 19 species included in our study were aligned using Cactus[23,24]. We followed the methods outlined in Sackton et al.[25] (https://github.com/tsackton/ratite-

genomics/tree/master/04_wga) to identify conserved non-exonic elements. First, fourfold degenerate sites were extracted from the alignment based on the *Nomia melanderi* annotation using halTools v2.1. phyloFit, part of the PHAST package[26], was used to generate a model of neutral evolution across these species using these data. We generated five of these neutral models using different random starting points and all models were nearly identical so one was chosen arbitrarily. This model was updated to reflect the average GC content in extant genomes using modFreqs.

We used phastCons[27] to identify conserved non-exonic elements. First, MAFs (Mutation Annotation Format files) were generated from the Cactus HAL files using *Nomia melanderi* as a reference sequence. Self-alignments from other regions of the *N. melanderi* assembly were filtered out of each aligned block and the MAFs were filtered using mafDuplicateFilter from mafTools[28]. mafStrander was then used to force all alignment blocks to a positive strand for the *N. melanderi* reference. Finally, alignment blocks were sorted based on coordinates using mafSorter and split into non-overlapping 1Mb blocks with at least 1,000 informative sites.

We ran phastCons using a range of parameters to identify optimal settings for identifying CNEEs. In total, we ran the procedure nine times with every combination of expected conserved element length of 20, 45, and 70 combined with target coverage proportions of 0.2, 0.3, and 0.4. Based on these results we chose a conserved element length of 45 and target coverage of 0.3 for identifying conserved elements. *N. melanderi* coding sequences were removed from the resulting conserved elements. We required that a conserved element be at least 100 bases in length in *N. melanderi* to be included in our dataset. The coordinates of the resulting CNEEs were identified in each species using halLiftover. Coding sequences were then removed from the CNEEs in each species individually and sequences of less than 100 bases were discarded. CNEEs were aligned with FSA[29] and alignments were filtered using the "-automated1" option of TrimAl v1.4.rev22[30].

Analysis of CNEEs

In order to identify CNEEs associated with social evolution, we examined differences in the rates of evolution in CNEEs among extant taxa. For each locus, branch lengths were calculated on the species topology using BASEML v4.9[31,32]. The resulting branch lengths were standardized to average rates of evolution in each genome using the branch lengths estimated by RAxML[33] using fourfold degenerate sites extracted using CODEML v4.9[31,32]. We then compared the standardized branch lengths of the five pairs of closely related social and solitary species in our study (*Augochlorella aurata* vs. *Augochlora pura*, *Halictus ligatus* vs. *Halictus quadricinctus*, *Lasioglossum marginatum* vs. *L. figueresi*, *L. zephyrus* vs. *L. vierecki*, and *L. pauxillum* vs. *L. oenotherae*). Loci for which the branch lengths were longer in all social taxa than the most closely related solitary taxa were considered fast-evolving in social species and vice versa.

miRNA annotation

miRNA was extracted from whole brains of individual females using the *mir*Vana™ miRNA Isolation Kit (Invitrogen cat. no. AM1560), and libraries were prepared with the NEBNext Small RNA Library Prep Set for Illumina (NEB cat. no. E7330S) (sample details in Supplementary Table 8). We used miRDeep2 (v. 2.0.0.8)[34] to identify miRNAs expressed in the brain of each species. We removed adapters, removed reads shorter than 18 nucleotides or with noncanonical letters prior to mapping reads to the genome assembly using the mapper.pl function. We then used miRDeep2.pl to detect known and novel miRNAs in each species. Known miRNAs were mature miRNA sequences previously described in other species (*Apis mellifera*, *Drosophila melanogaster*, *Nasonia vitripennis*, *Tribolium castenum, Bombyx mori, Bombus impatiens, Bombus terrestris, Megalopta genalis, Megachile rotundata, Nomia melanderi*)[35–37]. We quantified expression with the quantifier.pl script. We then filtered novel miRNAs based on the following criteria: no rRNA/tRNA similarities, minimum of five reads each on mature and star strands of the hairpin sequence, and a randfold p-value < 0.05; randfold describes the RNA secondary structure of potential precursor miR (pre-miRs). To identify homologous miRNAs across species, we added this filtered set to the list of known miRNAs and repeated the miRDeep2.pl detection step, followed by filtering. We characterized the genome location of each miRNA in the genome using bedtools (v. 2.27.0)[38] intersect to find overlap between miRNAs and predicted gene models.

We used computational methods to predict the gene targets of each miRNA in each species. First, we extracted potential target sites 500 bp downstream from each gene model using bedtools flank and getfasta. We used miRanda (v. 3.3)[39] with a minimum energy threshold of -20 (-en -20), minimum score of

140 (-sc 140), and strict alignment to seed region (-strict). We also used RNAhybrid (v. 2.12)[40] with a minimum free energy threshold -20 (-e -20). We kept only those miRNA-target pairs that were predicted by both programs at p < 0.001.

Coding sequence annotation
We generated gene predictions for each genome using BRAKER v2.1.0[41,42]. First, all RNAseq reads for each genome were mapped to the repeat-masked genomes using HISAT v.2.0.5[19] with a maximum intron length of 100,000. BRAKER was run on these mapped reads and the repeat-masked genome (softmasked). For *Halictus quadricinctus* (HQUA), for which we were unable to obtain RNAseq data, we used the RNAseq data from *Halictus rubicundus* (HRUB) for BRAKER prediction.

MAKER v3.00.0[43,44] was run on the repeat-masked genomes. The GFF files of aligned ESTs from PASA[45,46] were used as EST evidence. All high-quality protein predictions from Transdecoder[47] from all species were combined and used as protein evidence for each genome. In addition, OGSv3.2 from *Apis mellifera*, OGSv5.42 from *Lasioglossum albipes*, and all UniProt proteins downloaded on December 2, 2016 were included as protein evidence.

All gene predictions from BRAKER were fed to MAKER as pred_gff and we also used the MAKER implementation of EvidenceModeler. We used both the always_complete and correct_est_fusion options in MAKER.

Included in the output from MAKER is a set of gene predictions for which protein and transcript evidence did not overlap (*.non_overlapping_ab_initio.*). We used InterProScan v5.21-60.0[48] to identify those sequences within this set with conserved protein domains. We considered anything that InterProScan assigned an InterPro family to as a true gene and reran MAKER, incorporating these predictions into the annotation. The resulting annotations were used as the basis for the OGS. The only subsequent processing that occurred was the removal of any annotations on scaffolds deemed to be spurious duplications in the assembly step.

An Official Gene Set v2.1 was created for each genome. These gene sets are relatively complete as measured by BUSCO2 when comparing with the 4,415 genes expected to be present in all Hymenoptera species based on OrthoDB v9[49] (Fig. 1). The average percent of complete BUSCOs present in these Official Gene Sets is 93.6%. On average, 2.8% of genes are missing and 3.5% are fragmented.

Orthology
We used OrthoFinder v2.3.2[50,51] to identify orthologous groups of genes within the 19 species of Halictidae. We also uploaded protein sets to OrthoDB v10[52]. Gene names, orthologous *D. melanogaster* genes, and orthologous *A. mellifera* genes were assigned based on the OrthoDB groups. If more than 50% of the genes in a particular orthogroup were mapped to the same OrthoDB group, that orthogroup was assigned the gene name from the OrthoDB group.

Gene Ontology terms were assigned to orthogroups using information from three sources. First, Trinotate v3.0.1[53] was used to assign GO terms to every gene in each of the halictid genomes. In addition, GO terms of both the *D. melanogaster* and *A. mellifera* orthologs determined by OrthoDB mapping were compiled for each species. Orthogroups for which at least 30% of the genes in the group had been assigned a particular GO term from at least one of these sources were assigned those GO terms. For GO enrichment results, the find_enrichment.py script from GOATOOLS (v. 1.0.3)[54] was used with our custom orthogroup to GO mapping table. Only Biological Process terms were considered, and Benjamini Hochberg multiple testing correction was used to calculate q-values for GO terms with at least two representative genes in the target set tested for enrichment.

We used the 318 genes identified as transcription factors (TFs) in *Apis mellifera* in the Regulator database[55] to identify TFs in halictids. The *A. mellifera* orthologs determined by OrthoDB and assigned to orthogroups as above were used as the TF set in our study. Using this approach, we were able to identify 295 TFs in halictids.

Previous studies have identified a conserved set of genes associated with social behavior in insects and vertebrates, including autism spectrum disorder (ASD) in humans[56,57]. We thus looked for overlap between our candidate gene lists and the SFARI gene list, which is a curated list of genes associated with ASD. In order to identify human orthologs associated with ASD, we identified reciprocal best BLAST hits between the set of human UniProt sequences and the protein sequences from each halictid genome. If at least eight genes in an orthogroup had a reciprocal best hit to humans and greater than half of the genes

in an orthogroup were reciprocal best hits to the same human gene, then the orthogroup was assigned to that human gene. Gene associations with ASD were taken from the SFARI Gene web portal (https://gene.sfari.org/) on Oct. 15, 2017. This yielded human gene assignments to 5,787 orthogroups. The SFARI genes are ranked according to the amount of evidence associating them with autism. S is "syndromic" and then the scale is 0 to 6 with 0 having the highest degree of evidence. We included tests of all groups of genes of a certain degree of evidence or greater. We found Halictid orthologs of 30 SFARI "S" genes, 68 of "0", 13 of "1", 29 of "2", 55 of "3", 84 of "4", 41 of "5", and 5 of "6".

Gene ages

We used the Phylostratigraphy pipeline (https://github.com/AlexGa/Phylostratigraphy) developed in previous studies[58,59] to identify the approximate evolutionary age of origin for orthogroups in our study. The proteins used as a reference set for identifying gene ages included *Acyrthosiphon pisum* NCBI annotation v2.1, *Athalia rosae* i5k OGS v1.0, *Blattella germanica* i5k OGS v1.0, *Bombyx mori* NCBI Annotation Release 102, *Danio rerio* NCBI Annotation Release 106, *Drosophila melanogaster* FlyBase annotation r6.16, *Helobdella robusta* JGI Helro1 annotation release 46, *Homo sapiens* UniProt Proteome UP000005640, *Nasonia vitripennis* NasoniaBase OGS v1.2, *Octopus bimaculoides* Metazome v3.0 release, and *Polistes dominula* NCBI Annotation Release 100.

The createPSmap.pl script from Phylostratigraphy (https://github.com/AlexGa/Phylostratigraphy) was run on each of the 19 species included in this study with a sequence offset of 50 and an e-value of 1×10^{-5}. To assign an age to an orthogroup, we required that representative genes from at least five species be assigned an age by Phylostratigraphy and that the majority of the genes with an assigned age be assigned to the same age.

To test for correlations with ages, we extracted the crown age of taxonomic levels from the literature[60–62]. The ages used were 684 Mya for Bilateria (Age=1), 632 Mya for Protostomia, 373 Mya for Neoptera, 345 Mya for Holometabola, 240 Mya for Hymenoptera, 192 Mya for Apocrita, 162 Mya for Aculeata, 134 Mya for Apoidea, and 71 Mya for Halictidae (Age=9).

Gene expression

For each RNAseq sample, we calculated TPM (transcript per million) values for each gene using Salmon v0.9.1[63] with quasi-mapping and controlling for GC bias. Dufour's gland libraries were of the IU form; all other libraries were directional for which we instead used ISR. For all subsequent analyses, we used quantile normalization (normalize.quantiles from the preprocessCore library in R[64]) to normalize TPMs across species for each tissue. We then performed PGLS (phylogenetic generalized least squares) analyses using the R package geiger v2.0[65] to identify correlations between sociality and expression level for each orthogroup, excluding those species with polymorphic social behaviors.

Brain single-cell RNA-Sequencing

Raw sequencing reads from two *L. zephyrus* and two *L. albipes* brain scRNAseq experiments were trimmed with fastp[66] using default settings and aligned to respective genomes with CellRanger V6 (10X Genomics). Prior to analysis, we converted *L. albipes* gene IDs to one-to-one orthologs in *L. zephyrus* such that each species could be analyzed in a common transcriptomic landscape. We performed normalization and integration to combine scRNAseq data across species using the SCTransform pipeline with default settings in Seurat V4[67,68]. This yielded a final dataset of 6,716 cells for downstream analysis. Genes without one-to-one orthologs were retained in the Seurat data object and used for differential expression analysis but not for integration or clustering. This allowed us to identify cell types based on genes shared between the two species while also permitting genes that were both species- and cluster-specific to be identified.

The top 3,000 most variable genes were used to approximate the true dimensionality of the dataset, and a ranking of principle components (PCs) by percentage variance explained indicated that ten dimensions were suitable for cell clustering and visualization. Dimensionality reduction was performed with the Uniform Manifold Approximation and Projection (UMAP) technique, and cell clusters were identified with shared nearest-neighbor Louvain modularity optimization using a resolution of 0.6. Finally, cell-type-specific genes were identified using the MAST algorithm[69] in the "FindAllMarkers" function with "only.pos=T" and a Bonferroni-corrected *P* value of 0.05 (Supplementary Table 9).

Our analysis identified 11 cell clusters that could be broadly categorized as "neurons" or "glia" based on the mutually exclusive expression of either canonical neuronal markers *fne* and *Syt1* or glial

markers *repo* and *bdl*, in accordance with previous sc analyses of insect brains[2,70,71]. Upon closer examination, we found that *apolpp* was specifically upregulated in the only glia cell cluster, Cluster 6, along with two lipoprotein receptors, *LpR1* and *mgl*, as well as *apoltp*, all of which have been shown to play a role in neuron-glia lipoprotein shuttling and sequestration[72,73]. The juvenile hormone (JH) epoxide hydrolase (*jheh2*) was also significantly upregulated in Cluster 6 relative to other cell clusters.

To gain deeper insight into the specific expression patterns of *apolpp* and related genes in glia, we subsetted and clustered glial cells using six PCs and a resolution of 0.3 to compensate for the smaller cell population, yielding five subclusters within the glial population. We then identified glial cell-type-specific DEGs following the method described for the total cell population, and this finer-grained analysis revealed that *apoltp* and *apolpp* were co-expressed in a glial subcluster that also expressed *moody,* a marker of subperineurial glia, a subtype of surface glia that form the blood brain barrier (BBB) and regulate its permeability in the *D. melanogaster* brain[74,75]. In a separate subcluster, we found that *jheh2* was significantly upregulated along with *zyd*, a marker of cortex glia[71,76,77] (Supplementary Table 10). Unlike surface glia, cortex glia encase and support neuronal somata and are thought to play a role in juvenile hormone sensitivity[78,79].

Tissue specificity

We calculated the specificity of expression of each orthogroup across the four tissues examined using the normalized expression levels[80]. We required data from at least eight species and all four tissues in order to calculate a specificity index for each gene. When multiple genes were included in an individual orthogroup, a single representative was randomly chosen for the specificity analysis. The specificity metric ranges between 0 and 1. The closer the value is to 1, the more tissue-specific the gene.

Coding sequence evolution

Alignment. Coding sequences were aligned using the codon-aware version of Fast Statistical Aligner v1.15.9[29]. For several analyses, we then performed an extensive filtering procedure to exclude poorly aligned regions. This filtering followed Sackton *et al.* [25] (https://github.com/tsackton/ratite-genomics/tree/master/03_homology/03_protein_coding_alignment). Each site in an alignment was only retained if at least eight sequences and 30% of sequences were a non-gap base at that site and were represented by at least four species. We then masked poorly aligned regions using the scripts of Jarvis *et al.* [81]. Following this masking, the first filter of alignment columns was performed again on the masked data. We required that 50% of the length of the original sequences not be masked by any of these steps and that less than 50% of the aligned sequence be gap characters. All sequences had to have at least 300 non-gap or masked bases.

Relaxation using HyPhy RELAX. We used HyPhy RELAX v2.3.11[82] to identify genes experiencing relaxed selection in solitary species that evolved from ancestrally social lineages. For an orthogroup to be tested, we required that at least 12 taxa be present and that *Halictus ligatus, H. quadricinctus, Augochlorella aurata, Augochlora pura,* and at least one pair of *Lasioglossim marginatum* / *L. figueresi, L. zephyrus* / *L. vierecki*, and *L. pauxillum* / *L. oenotherae* be represented. This requirement ensured that at least one closely related pair of social and solitary species from each of the three social clades was included. Socially polymorphic species were excluded for this analysis. 6,904 loci met these requirements. In addition to examining signatures of relaxation in solitary species, we performed a parallel test on the extant lineages of social species to represent a null baseline. We do not expect an elevation of relaxed selection in social species that evolved from social ancestors. FDR-correction was used to account for multiple testing at FDR<0.1.

Selection using aBS-REL. We also used HyPhy aBSREL[83] tests to identify signatures of selection on individual branches. To include an orthogroup in this test, we required that at least two *Halictus* species, at least two of *Augochlorella aurata, Augochlora pura,* and *Megalopta genalis*, at least two *Lasioglossum*, and at least two of *Dufourea novaeangliae, Nomia melanderi,* and *Agapostemon virescens* be represented. A total of at least 12 species had to be represented. This test yielded p-values for every branch across every locus. This full set of p-values was FDR-corrected for multiple testing at FDR<0.05.

Because we did not require that all taxa be represented to include a locus in this test, identifying the same branches across loci was not straightforward. Therefore, we stringently defined branches based on the extant taxa that make up the descendants of those branches. We required that a locus include sequences for all descendants of a branch as represented in the full dataset as well as at least one member

of the sister clade. In addition to identifying specific branches under selection, we also counted the number of branches identified as under selection in a locus regardless of taxonomic composition.

Phylogeny inference. Fourfold-degenerate sites were extracted from the "4fold.nuc" files produced by PAML during inference of ancestral sequences. For those orthogroups with single representative sequences in each of the 19 halictid species, these sequences were concatenated, producing a 561,041 base matrix. The phylogeny of the group was inferred using RAxML with a GTRGAMMA model of evolution. The resulting topology matched previous results[84]. This phylogeny (with the inferred branch lengths) was used in all of the evolutionary analyses outlined above.

Associations between promoter TFBSs and sociality

Previous studies examining associations of genetic elements with social evolution in bees found that transcription factor motif binding strength in promoters was more often positively correlated with social behavior in 223 motifs identified in *Drosophila*[85]. We replicated those analyses here with a few minor differences.

For each of the motifs used in that previous study[4], a "stubb" score[86] was calculated for each 500 base window in every genome, with a step size of 250 bases. For each motif for every species represented in each orthologous group, we then identified the window with the greatest stubb score within 5kb upstream and 2kb downstream of the transcription start site and assigned that score to that gene for that species. Then, for a given species and a given motif, the motif scores were rank-normalized (best score=1, worst score=0). We did not perform the GC-normalization and p-value calculation used previously[85], instead analyzing these rank scores directly.

For every orthogroup and every motif, we examined correlations between stubb score and social behavior using PGLS as implemented in geiger v2.0[65] after first removing all socially polymorphic taxa. These analyses were performed on the phylogeny generated by RAxML on a concatenated matrix of all fourfold degenerate nucleotide sites present across all 19 halictid species. In order to consider the correlation of a particular motif and orthogroup with social behavior, we required that the normalized rank of that motif be at least 0.95 in at least one species. Ranks of 0.95 or greater indicate that the stubb score for that particular motif is in the highest 5% among the 7,102 genes considered across the genome, indicating that these sequences are the most likely to actually be recognized and used by TF's to modulate expression. We also required that the number of genes with significant correlations of a motif be at least five. For each motif, we then counted the numbers of significant (PGLS p-value < 0.01) correlations detected with higher stubb scores in social taxa and with higher stubb scores in solitary taxa. Those motifs for which there were differences in numbers of significant correlations between social and solitary taxa of at least 20% were considered to be associated with the behavior with which they more often correlated.

ApoLpp evolution

Apolpp within Halictidae. We explored the evolution of the gene *apolpp* in detail. If the changes in this gene are associated with social behavior, we expect that these changes would change the function of the protein. We therefore attempted to find particular regions of the apolpp protein sequence associated with social evolution. First, we simply examined 100 AA sliding windows with a 50 AA step size across the *apolpp* alignment. For each window, we inferred branch lengths on the species tree using RaxML v7.3.0[33] with the PROTGAMMAWAG model of evolution. We then calculated the branch length between each tip species and the outgroup species, *D. novaeangliae.* Most of the species in our analysis, regardless of social behavior, evolved from a social ancestor. However, three species, *D. novaeangliae, N. melanderi* and *Ag. virescens,* evolved from strictly solitary ancestors. Transitions in social behavior, rather than sociality itself, might be associated with shifts in evolutionary rates, thus we compared the 16 species with any shift in social behavior to *N. melanderi* and *Ag. virescens.* Specifically, we looked for regions with longer branches in all 16 species with at least one ancestral shift in social behavior relative to the two species without such shifts. We also used fpocket2[87], as implemented in the Phyre2[88,89] Investigator tool, to predict the likely active sites of *apolpp* in halictids, using the *H. ligatus* protein sequence as a reference.

Apolpp across insects. We also explored *apolpp* evolution across insects as a whole in order to place the patterns of evolution within Halictidae into a broader context. We collected *apolpp* sequences from 78 additional Neopteran insect species. In particular, we included an additional 12 species of bees, 18 species of ants, eight other Hymenoptera, 10 Coleoptera, 12 Diptera, five Hemipterans, and 11 Lepidoptera (Supplementary Table 11).

Protein sequences were aligned using MUSCLE v3.8.31[90] and filtered (with the -automated1 option) and backtranslated to nucleotides with trimAl v1.4.rev22[30]. We ran aBSREL as implemented in HyPhy v2.3.11 on the resulting nucleotide alignment, testing all 191 branches within the phylogeny. We used FDR-correction to adjust the resulting p-values and counted the number of branches detected as experiencing selection in each of the major clades examined.

We also examined shifts in rates of protein evolution during the evolution of Halictidae. We performed likelihood ratio tests assessing the possibility of rate shifts between five pairs of sister taxa: Holometabola versus Hemiptera, Hymenoptera versus all other Holometabola, ants and bees versus other Hymenotpera, bees versus ants, and Halictidae versus other bees. For each test, we used PAML to calculate the likelihoods of models of evolution that fixed the rates of evolution in the two clades being compared, or allowed them to differ. We also repeated the same set of LRTs as above using just the 51 amino acids present in the pockets predicted by fpocket2.

Finally, we used the branch model of PAML v4.9[31,32] to estimate dN/dS ratios for each of the major clades examined.

<u>LC-MS Data Analysis</u>
Data were analyzed using El-MAVEN Software (Elucidata, LLC., elucidata.io). Chromatographs and mass spectra were exported from El-MAVEN and plotted using MATLAB, and peak areas were exported, processed and plotted using R. Peak areas were all normalized to a valine-D8 (Cambridge Isotopes Laboratories Cambridge, MA) internal standard. Absolute quantification of JH III and JH III-d3 was carried out using a standard curve of purified JH III (Sigma, Cat. #J2000) and JH III-d3 (Toronto Research Chemicals, Cat. # E589402), respectively. For social *A. aurata* and solitary *A. pura* comparisons, JH III quantities were standardized to ng/mm3 using brain volumes previously-quantified for each species[91]. Raw data available upon request.

Supplementary Text

<u>Sweat bee life history</u>
Most sweat bees exhibit an annual life history, with adult females initiating nests in early spring (Extended Data Fig. 1). Halictid females begin their reproductive cycle by excavating a nest typically composed of a central tunnel and several adjacent brood cells that they line with waxy secretions from their Dufour's glands. During this time, reproductive females forage for nectar and pollen, and they construct a pollen ball as a mass provision for offspring. Then, females deposit eggs on top of this pollen ball, seal the brood cell, and wait for the first brood to eclose. Some species open and inspect these brood cells while others do not. The next generation of reproductives will mate and overwinter as adults in hibernacula (cells that they excavate in their nesting substrate).

In eusocial nests, females typically produce two broods: first a female-biased brood that remain in the nest as workers followed by a second brood of males and females that become the next reproductive generation in the following year. Some eusocial species can produce multiple worker broods per season, and many adjust the number of broods produced depending on local environmental conditions (presumably due to variation in season length).

In this study, we sampled species for which social behaviors have been relatively well characterized and focused on species that allowed us to sample the phylogenetic breadth of gains and losses of eusociality in this group. Below, we describe some aspects of the life histories of each of these species.

***Dufourea novaeangliae* (Robertson, 1897)**
Solitary
Specialist
Ground nesting

Life history

D. novaeangliae is a solitary sweat bee that specializes on the Pickerelweed (*Pontederia cordata*). This species is found in the northeastern North America, with nesting sites typically in sandy soils near freshwater ponds and streams (close to where Pickerel weed grows)[92].

Nomia (Acunomia) melanderi Cockerell, 1906
Solitary
Generalist
Ground nesting

N. melanderi is a solitary, ground-nesting sweat bee that is found in arid or semi-dry environments in the Western United States. These bees nest in dense aggregations, and are often agriculturally managed as key pollinators of alfalfa (*Medicago sativa*)[93,94].

Megalopta (Megalopta) genalis Meade-Waldo, 1916
Polymorphic
Generalist
Wood nesting

M. genalis is a wood-nesting sweat bee found in Central and South America. Populations in Panama are socially-polymorphic[95], with a mix of eusocial and solitary nests found at the same site. Solitary nests consist primarily of an adult female and male offspring that disperse from their natal nest shortly after eclosion. Eusocial nests contain a reproductive female and 1-2 workers. Foundresses that establish solitary nests are smaller and take longer to initiate egg-laying than those that establish eusocial nests, though their respective size ranges overlap[95]. Females construct their nests primarily in the pith of twigs, sticks, and lianas. *M. genalis* is unusual amongst the bees because it is crepuscular and forages only at dawn and dusk[96].

Augochlorella aurata (Smith, 1853)
Eusocial
Generalist
Ground nesting

A. aurata (=A. striata) is a ground-nesting eusocial bee with a wide range throughout the eastern US and Canada. Nests are eusocial and typically contain 3-5 workers[97]. A study of this population in Cape Breton, Canada (the northern limit of its range) identified a population containing a mix of eusocial and solitary nests[98], but this appears to be a notable exception for this species; eusocial nests are common throughout the rest of its range. *A. aurata* is a generalist, and mated, female gynes overwinter as adults.

Augochlora (Augochlora) pura (Say, 1837)
Solitary
Generalist
Wood nesting

A. pura is a wood-nesting, solitary bee found in the eastern US. Females overwinter as adults and found nests in early spring in rotting wood. There are several reproductive broods each season (n=2-4 broods in Kansas[99] and in New Jersey (Kingwell, Omufwoko, Wyman and Kocher, *personal observation*), and foraging continues through the late fall. Females mate in the fall and overwinter in small groups in hibernacula. Individual females found new nests in the spring, and the first brood of adult females mate and produce a second round of nests about a month later. Nests typically contain between 6-12 cells. This is a generalist species[99,100].

Agapostemon (Agapostemon) virescens (Fabricius, 1775)
Non-eusocial/communal
Generalist
Ground nesting

A. virescens is a large, communal sweat bee that nests in the ground. Females often share and guard nest entrances but provision cells and lay eggs independently[101]. Communal nesting may be a mechanism to protect nests against cleptoparasites; nest switching is also commonly observed in this species[102]. Nests of similar species typically contain between 5-10 females and 50-100 brood cells, and observations made in *A. virescens* suggests nests of this species are typically similar in size[101]. This is a common, generalist species in North America.

Halictus (Odontalictus) ligatus Say, 1837
Eusocial
Generalist
Ground nesting (primarily)

H. ligatus is an obligate eusocial sweat bee that nests primarily in the ground. *H. ligatus* is a commonly occuring bee across North and Central America. Nests range in size across this cline, with nests found in warmer, tropical regions producing multivoltine broods and larger colonies[103–106]; nests found in northern areas of the range with cooler, temperate climates are smaller and typically only produce a single worker brood followed by the production of reproductives. The social structure of *H. ligatus* varies across this cline with "classic" eusocial colonies being produced in more temperate, harsher environments. As climates warm in the southern regions of the range and colony sizes increase, several individuals within a nest can act as reproductive females[103–106].

Halictus (Halictus) quadricinctus (Fabricius, 1776)
Solitary
Generalist
Ground nesting

H. quadricinctus is a solitary sweat bee species found throughout Eurasia[84,107,108]. Nests are constructed in the ground and consist of a single brood of reproductive offspring. This is the largest sweat bee species in the world, with female body lengths up to 20mm.

Halictus (Protohalictus) rubicundus (Christ, 1791)
Polymorphic
Generalist
Ground nesting

H. rubicundus is a socially polymorphic sweat bee that nests in the ground[109]. It is found across the Holarctic. Solitary populations of *H. rubicundus* have been documented at high elevations in the Rocky Mountains[110], as well as at high latitudes in Scotland[111]. Reciprocal transplant experiments in the UK have demonstrated that this variation is likely to be the result of behavioral plasticity, but phylogenetic studies in the western US suggest that the solitary and social populations found in Utah may be genetically distinct. In New York, the first brood can contain a mix of worker and reproductive females as well as males[112].

Lasioglossum (Leuchalictus) leucozonium (Schrank, 1781)
Solitary
Generalist
Ground nesting

L. leucozonium is a solitary sweat bee species found throughout the Holarctic (as an introduced species in North America). This species can produce multiple broods per year (e.g. can be multivoltine[113]), but it appears to be solitary at all locations[109,114,115].

Lasioglossum (Evylaeus) marginatum (Brullé, 1832)
Eusocial

Generalist
Ground nesting

L. marginatum is a eusocial halictid species with a unique life history. This ground nesting species undergoes a perennial life cycle, where a queen produces subsequent worker broods each season until the final year, when hundreds of reproductive offspring are produced[116,117]. Colony sizes for this species are the largest among halictids, and can contain >400 workers and with thousands of brood cells within a single nest. Mating takes place within the nests, and there is no queen-worker dimorphism observed. *L. marginatum* has a Palearctic distribution.

Lasioglossum (Dialictus) figueresi Wcislo, 1990
Solitary
Generalist
Soil nesting

L. figueresi is a non-eusocial sweat bee species found in Central America. Though this is a neotropical species, it appears to undergo a temperate-like annual life cycle. Adult females initiate nests in vertical soil banks, typically alone but sometimes small numbers of same-generation females can communally found a nest. Reproductive females do not survive the wet season, and the next dry season, new adult females found the next generation of nests[118–121].

Lasioglossum (Dialictus) zephyrus (Smith, 1853)
Eusocial
Generalist
Ground nesting

L. zephyrus (=L. zephyrum) is an obligate eusocial sweat bee species that ranges throughout the US and Canada. This species nests in soil, often in vertical banks but also on flat ground. Colonies are initiated by single reproductive females, and the first brood of workers is female-biased[122,123]. These females remain in the nest as daughter workers and forage, provision, excavate and clean brood cells. The mother queen is the primary egg layer. Division of labor is maintained through physical interactions between queens and workers in the colony[124]. If the queen dies, the oldest sister is typically the female to inherit the nest and maintains her dominance through the same sets of physical interactions[125]. Colonies tend to range in size from 3-24 workers.

Lasioglossum (Dialictus) vierecki (Crawford, 1904)
Solitary
Generalist
Ground nesting

L. vierecki is a solitary sweat bee species that prefers to nest in sandy soils[84,126,127], and is distributed across the mid- and eastern United States. This species is very small, with females only about 4mm in size. Nest excavations have revealed nests with only single females inside with activated ovaries, and dissections of multiple females in New Jersey revealed all individuals had activated ovaries, further affirming the solitary life history of this species (Wyman and Kocher, personal observations).

Lasioglossum (Sphecodogastra) pauxillum (Schenck, 1853)
Eusocial
Generalist
Ground nesting

L. pauxillum is a eusocial, ground nesting species found throughout Europe and the Palearctic [128,129]. They follow a typical eusocial life history with the production of a single worker brood followed by a reproductive brood at the end of the season. Adults mate in the fall and then females overwinter to initiate new eusocial nests in the spring.

***Lasioglossum (Sphecodogastra) oenotherae* (Stevens, 1920)**
Solitary
Specialist
Ground nesting

L. oenotherae is a solitary sweat bee that specializes on pollen collection from primroses (*Oenothera*)[115,130]. It is distributed throughout eastern North America.

***Lasioglossum (Sphecodogastra) malachurum* (Kirby, 1802)**
Eusocial
Generalist
Ground nesting

L. malachurum is an abundant, obligate eusocial sweat bee found throughout Europe[128,131–134]. This species nests in the ground, and can produce 1-3 worker broods within a year. In Greece, females have been documented to be multivoltine and produce multiple broods each season[135]. In more temperate climates, this species follows the typical halictine eusocial life history, with the production of 1 worker brood followed by reproductives. Adult females mate in the fall, overwinter as adults, and initiate new nests in the spring. Queens and workers are different sizes, with queens being approximately 10-55% larger than workers[135]. Colony sizes range from 10-15 workers in a nest.

***Lasioglossum (Sphecodogastra) calceatum* (Scopoli, 1763)**
Polymorphic
Generalist
Ground nesting

L. calceatum is a socially polymorphic sweat bee with a Palearctic distribution. *L. calceatum* is eusocial throughout France[136], but social and solitary colonies have been documented in the UK[137]. Reciprocal transplants from Scotland (solitary) to southern England (social) suggest that these behavioral differences have a primarily genetic underpinning and that the high latitude solitary species are losing the ability to produce eusocial nests[138]. Colony sizes range from 2-6 workers in a nest, and queens are only slightly larger than workers (3-5%).

***Lasioglossum (Sphecodogastra) albipes* (Fabricius, 1781)**
Polymorphic
Generalist
Ground nesting

L. albipes is a socially polymorphic sweat bee that is eusocial in western populations of France and solitary in eastern populations in France[136,139,140]. The behavioral differences studied in France have been shown to have a genetic underpinning, and populations appear to be locally adapted to shorter seasons found at higher elevation, eastern populations[56,140]. This is a ground nesting species, with colonies ranging in size from 2-24 workers[141]. Queens and workers have a slight size dimorphism with a difference of about 3-5% in overall body size.

Coding sequence evolution
Signatures of selection across halictids. We tested 7,769 genes for signatures of positive selection within Halictidae using HyPhy aBS-REL. We obtained p-values for every branch present in the phylogeny for every gene. Thus, we only examined those p-values that passed FDR correction, combining all p-values across all loci and branches for correction. Most basically, we examined those genes with signatures of selection on at least a given number of branches ranging from one to 10. For this analysis we were agnostic to the taxonomic composition of those nodes identified as under selection (i.e., clades did not have to be fully represented taxonomically to be included in these gene sets). Finally, we also examined all genes identified with signatures of selection on all 35 branches present in the complete phylogeny of all 19 species.

4,261 genes show signatures of selection on at least one branch in the phylogeny (FDR-corrected p < 0.05; Supplementary Table 5) which are enriched for three GO terms (Fisher's Exact q < 0.1; Supplementary Table 6). These GO terms are associated with structural functions (i.e., "cell adhesion", "biological adhesion", and "cell junction assembly"). Genes with larger numbers of branches identified as experiencing differential selection are enriched for larger numbers of GO terms, with the greatest number of GO terms (230) identified in the set of 40 genes with at least nine branches identified as under selection. These terms include "synaptic transmission, glutamatergic" as well as a number of others related to gene regulation and synapse function (Supplementary Table 6). Because previous studies have identified a conserved set of genes associated with social behavior in insects and vertebrates, including ASD in humans[56,57], we looked for overlap between our candidate gene lists and genes associated with ASD (the SFARI gene list). Many of the genes experiencing positive selection in halictids are also significantly enriched for SFARI genes (Hypergeometric test p < 0.05; Supplementary Table 12). While these significant overlaps do not reveal anything about the differences between social and solitary species specifically, they do indicate that these genes have experienced changing selection pressures within Halictidae as a whole, a highly socially variable family. Interestingly, the gene *Syx1A*, which has been previously implicated in social evolution in Halictidae[56], is identified by aBSREL on six branches within the family (*N. melanderi, L. vierecki, L. marginatum, L. albipes*, the branch ancestral to *L. albipes* and *L. calceatum,* and the branch ancestral to *L. oenotherae* and *L. pauxillum*), supporting the possibility that changes in this gene are fundamental to social behavior within this group.

There is a weak but significant correlation between the number of branches identified with signatures of selection for a gene and the age of that gene (Pearson correlation p=4.3x10^{-5}, r=0.047), indicating that younger genes tend to be more malleable. There are also significant positive correlations between the proportion of genes of a given age with at least three (p=0.023, r=0.78) and four (p=0.044, r=0.72) branches under selection and the age of those genes, again indicating that younger genes are more likely to be undergoing selective changes (Supplementary Table 3). Concordantly, genes with any number of branches identified by aBSREL have significantly higher tissue-specificity than other genes (Wilcoxon p < 0.005; Supplementary Table 13).

Positive selection associated with the origins of eusociality. We selected the two branches on which social behavior convergently originated within halictids[84]: at the root of the Augochlorini and at the root of the genera *Halictus* and *Lasioglossum*. The aBSREL tests identified 309/6312 genes under positive selection on the Augochlorini origin branch and 62/5058 genes under positive selection on the *Halictus + Lasioglossum* origin branch. Of the 4,341 genes that were tested by aBSREL on both of these branches, 191 were significant on the Augochlorini branch and 53 were significant on the *Halictus + Lasioglossum* branch; nine genes were detected as experiencing positive selection on both, significantly more than expected by chance (Fisher's Exact, p-value=0.004, 3.97-fold enrichment). Perhaps due to the small number of genes identified, no discernable GO enrichment was found for this gene set, although we do detect significant enrichment for genes that differ significantly in expression in the antennae of social and solitary bees (Supplementary Table 12).

Positive selection associated with the loss of eusociality. We did not find any genes that showed a signature of positive selection on all of the branches associated with the loss of eusociality, but there were 95 genes with evidence of positive selection on at least 1 secondarily solitary lineage in both the Augochlorini and the Halictini *s.s*. These genes are not enriched for any GO terms, but include the JH co-receptor, *taiman* (Supplementary Table 5).

Intensification of selection on extant eusocial branches.
To search for signatures of selection associated with the maintenance or elaboration of eusociality, we identified genes experiencing intensified purifying or positive selection on terminal eusocial branches in our phylogeny. We found 393 genes with intensified selection on terminal eusocial branches (HyPhy RELAX, FDR<0.1), 15 of which also experienced positive selection on at least one eusocial origin branch (Supplementary Table 5). These 393 genes are enriched for cell adhesion (GO:0007155, Fisher's Exact q=3.86e-3, enrichment=2.49) and the non-canonical Wnt signaling pathway (GO:0035567, Fisher's Exact q=8.87e-3, enrichment=8.09), among others (Supplementary Table 6). In addition, genes under intensified selection on eusocial branches included 16 TFs, including *ultraspiracle* (*usp*) and *Ecdysone Receptor* (*EcR*), both of which are associated with the *usp* binding motif previously linked to eusocial evolution in stingless bees[85] and are known transducers in the JH signaling pathway.

Relaxed selection in lineages that have lost social behavior.

Our species sampling includes six species that are solitary but that evolved from a social ancestor (*Augochlora pura, H. quadricinctus, L. leucozonium, L. figueresi, L. oenotherae,* and *L. vierecki*). We assumed that the evolutionary loss of social behavior would lead to relaxed selection in those genes most constrained among social species and for which the fewest changes can be tolerated. We used HyPhy RELAX to identify those genes experiencing relaxed selection across all six of these species, testing 6,890 loci and identifying significant signatures of relaxation in 443 loci (FDR-corrected LRT p < 0.1). Four GO terms were significantly enriched in this gene set, including "chromosome condensation", "DNA packaging", and "DNA conformation change" (Supplementary Table 6), all three of which may be related to the regulatory functions identified in the correlations between social behavior and constraint.

In order to determine whether the loss of social behavior did, indeed, lead to additional relaxed selection, we also used HyPhy RELAX to identify those genes experiencing relaxed selection in the six extant social lineages represented in our dataset, testing a total of 6,887 loci, 305 of which showed significant signatures of relaxation (FDR-corrected LRT p < 0.1). These 305 genes were enriched for antibiotic metabolic process (GO:0016999, Fisher's Exact, q=0.053, enrichment=4.36), drug catabolic/metabolic processes (GO:0042737, GO:0017144, Fisher's Exact, q=0.093, enrichment=4.03,2.25), and malate metabolic process (GO:0006108, Fisher's Exact, q=0.095, enrichment=11.29) (Supplementary Table 6). This is a significantly lower proportion than the 443 of 6,890 loci identified to be experiencing relaxed selection among the solitary species (Fisher's exact test p=2.42x10^{-7}, odds-ratio=1.48), suggesting that the loss of eusociality is more often associated with a release of constraint compared with eusocial maintenance or elaboration.

The proportion of genes undergoing relaxed selection in solitary species is correlated with gene age; a larger fraction of younger genes is experiencing relaxed selection (Pearson correlation p=0.002, r=0.869; Supplementary Table 3; Fig. 2). Genes experiencing relaxed selection in solitary species also have significantly greater tissue specificity than other genes (Wilcoxon test p=0.00167; Supplementary Table 13). Therefore, those genes associated with social evolution tend to be more recently evolved with more tissue-specific function making them less likely to be essential for overall organism function when expressed in a solitary phenotype.

Gene ages and patterns of selection

We examined the relationship between gene age and selection associated with eusocial origins, maintenance, and reversions to solitary life histories to determine whether or not there was any relationship between selection and gene age (Extended Data Fig. 5). There is a greater proportion of young genes experiencing relaxed selection when eusociality is lost (Pearson's r=0.869, p=0.002); we found no significant association with relaxation on extant, eusocial branches. Next, we looked at the sets of genes that show intensification of selection pressures (HyPhy RELAX, FDR<0.1), but neither of these sets showed any significant association with gene age. Finally, we looked at genes that showed evidence of positive selection (HyPhy aBSREL, FDR<0.05). We found no relationship between gene age and the proportion of genes with evidence of positive selection on at least 1 branch representing the origins of eusociality (Positive, gains). Likewise, there was no relationship between gene age and the proportion of genes with evidence for positive selection (HyPhy aBSREL, FDR<0.05) on at least 1 loss branch in the Halictini and on the 1 loss branch in the Augochlorini (Positive, losses). Interestingly, when we combine the lists of genes with signatures of positive selection on the Augochlorini (309) and Halictini (62) branches, we find a trend suggesting a higher proportion of younger genes experience positive selection (linear regression, R^2=0.415, p=0.0846). These results add some support to our observation that younger genes are more likely to experience relaxed selection when eusociality is lost.

Complementary patterns of selection associated with emergence and breakdown of eusociality

There were 4,334 loci for which we tested the two social origin branches for signatures of positive selection and that we tested for signatures of relaxation on extant solitary species with social ancestors using HyPhy RELAX. Of the 443 genes detected as experiencing relaxed selection, four of them overlap both sets of genes with signatures of positive selection on both origin branches. This is a significant enrichment of overlapping signals between these three tests (Multi-Set Enrichment, enrichment=8.85, p-value=0.001). Given the overlap of these signals, these genes (*apolpp, Pat1, Hex110*, and a Hymenoptera-specific gene,

OG_11519) are likely involved in the evolution of sociality. We were unable to assign any gene ontology terms to OG_11519, indicating that it is either completely novel or has diverged substantially in sequence from previously characterized protein domains.

Positive selection associated with origins and intensification of selection on extant eusocial lineages

We find 15 genes that show a signature of positive selection on at least one of the branches associated with the origins of eusociality and intensification of selection on extant eusocial branches (14 Augochlorini, 1 Halictini, 0 shared), suggesting that these 15 genes may be particularly relevant to the emergence and elaboration of eusociality (Supplementary Table 5). 14/15 of these genes are conserved across Bilateria, suggesting that this set is highly conserved and consistent with regression results on the 393 genes experiencing intensification of selection above. Some of the more interesting genes in this set include: *Nop60B* (RNA metabolism), *Bruce* (an inhibitor of apoptosis), and *ReepA* (ER stress prevention in the presynapse). These genes are also enriched for embryo implantation (GO:0007566, Fisher's Exact, q=9.92e-3, enrichment=153; Supplementary Table 6).

Intensification in social lineages, Relaxation in solitary lineages

34 genes show intensification of selection on extant eusocial lineages and relaxation in secondarily solitary species (HyPhy RELAX, FDR<0.1 for both tests; Supplementary Table 5). These genes likely represent trade-offs associated with the maintenance or elaboration of eusociality, and they are enriched for regulation of SNARE complex assembly (GO:0035542, Fisher's Exact, q=0.074, enrichment=80.91; Supplementary Table 6), which is a key component of synaptic transmission that has also been implicated with variation in social behavior in *L. albipes*[56] and wasps[142]. None of the 34 genes showed a signature of positive selection on either of the branches representing the origins of eusociality, nor do they show a correlation with gene age.

Overlap with previous comparative genomic studies in bees

We did not find a significant overlap between our gene lists and those presented in [85]. We used *L. albipes* reciprocal blasts to translate from the list of genes identified as under selection for Halictidae origins[85] to our new *L. albipes* annotation, which results in some loss of genes. Of the 105 genes remaining after translation (167 in [85]), none overlapped with the genes we identified as under selection in Halictini (not different from random expectation; p=0.487), and 5 overlapped with genes we identified under selection in Augochlorini (not different from random expectation; p=0.224). Five genes (1 of which is also under selection in Augochlorini) that are experiencing relaxed selection on solitary branches overlapped with the genes identified in [85], but this is also not significantly more than expected by chance (p=0.494). This suggests that our study design and approach is identifying novel mechanisms associated with the evolution of eusociality relative to those previously identified using more distantly related species[85]. We also compared the genes we identified as experiencing selection with either gains or losses of sociality with genes previously identified to be associated with sociality in the socially polymorphic *L. albipes*[56]. After converting *L. albipes* gene IDs from this assembly[56] to our orthogroup annotations, we found no genes overlapping with our lists of genes associated with social origins (not different from random expectation; p=0.146 for Halictini origin and p=0.463 for Augochlorini origin). Three genes that are experiencing relaxed selection on solitary branches overlapped with the genes identified in the previous study[56], but this is also not significantly more than expected by chance (p=0.136).

Hex110 and ApoLpp show evidence for selection on functional domains

Both Hex110 and ApoLpp show evidence for domain-specific positive selection associated with the origins of eusociality. Domain predictions were made with PHYRE[88,89]. Predicted binding pockets by PHYRE are shown with pink rectangles, glycosylation sites in orange squares. Branch-site specific tests for diversifying positive selection were conducted using the two branches associated with the origins of eusociality as focal branches (MEME[143], p<0.05). We also identified sites with a mutational effect score (SuSpect[144]) >=1. Conservation scores were estimated with AL2CO[145]. In both JHBPs, MEME identifies sites in functional regions of the protein, including the receptor binding domain and predicted binding pocket for ApoLpp as well as in all three Hemocyanin domains (associated with storage functions in these proteins) and in the predicted binding pocket of Hex110. Moreover, for both proteins, we also find region-specific, faster rates of evolution on eusocial branches compared to non-eusocial outgroups in this phylogeny (data not shown):

the predicted binding pocket for ApoLpp and all Hemocyanin domains for Hex110. Taken together, these results suggest that positive selection shaped protein function as eusociality emerged in this group of bees.

ApoLpp evolution within halictids

We identified seven 100-residue regions evolving faster in lineages with at least one ancestral transition in social behavior than in *Ag. virescens* and *N. melanderi.* Five of these regions occurred within the first 900 amino acids of *apolpp,* comprising a total of 400 amino acids. The vitellogenin-N domain of *apolpp* makes up residues 41-573 and is likely the receptor-interacting site[146]. This region encompasses two overlapping windows that have evolved faster in lineages with social transitions from residues 200-300 and 250-350, indicating that these windows are, indeed, likely to be functionally significant. 51 sites were predicted to be part of the protein pockets of this gene, all of which are within the first 950 residues and 17 of which are also within those windows found to be evolving faster on lineages with social transitions. We also performed the same rate shift analysis using just the residues identified as a part of the pocket of *apolpp* and, indeed, rates of change are higher in lineages with social transitions.

ApoLpp evolution across insects

97 coding sequences were used to generate an insect-wide apolpp phylogeny (Supplementary Table 11). We tested five nodes in the phylogeny of *apolpp* across insects for rate shifts and four of these branches showed significant differences in rates (LRT p < 0.01; Extended Data Fig. 7). Rates appear to have increased significantly in Holometabola as a whole, to have decreased in Hymenoptera relative to the rest of Holometabola, and then increased in ants and bees relative to other Hymenoptera, and again in Halictidae relative to other bees. Rates did not shift significantly between bees and ants (p=0.979). Locations of rate changes are indicated in Extended Data Fig. 7. Tests of the same clades but with an alignment limited to only those 51 amino acids predicted to be a part of the protein pocket are all nonsignificant (LRT p > 0.05) except for the test between ants and bees and the rest of Hymenoptera, which shows a significant rate increase within the ants and bees (p=0.0051).

We also used HyPhy aBSREL to identify those branches with signatures of positive selection for *apolpp*. Branch-specific rate shifts in amino acid evolution were identified using likelihood ratio tests in PAML[31,32]; methods as in[147]. Remarkably, 32% of branches within Halictidae have experienced positive selection, a larger proportion than in any other clade examined (Supplementary Table 14). dN/dS ratios were also highest for halictids (0.19) compared with all other clades examined (Coleoptera (0.021), Hemiptera (0.062), Lepidoptera (0.088), Diptera (0.096), ants (0.17), and non-halictid bees (0.14)). *Apolpp* appears to change particularly quickly within Halictidae. Results are summarized in Extended Data Fig. 7.

TFBS evolution in promoters

To assess the degree to which changes in gene regulation may facilitate the evolution of social behavior in halictids, we characterized TF motifs in putative promoter regions in each halictid genome. For each species, we defined these regions as 5kb upstream and 2kb downstream of the transcription start site for each gene[85] and calculated a score for a given TF in each region that reflects the number and/or strength of predicted binding sites[86]. For each of the 223 motifs examined, we counted the numbers of significant (PGLS p-value < 0.01) correlations between social behavior and stubb scores in gene promoters. We split this set of correlations into those with positive correlations with social species and those with positive correlations with solitary species. If social species have a greater capacity for gene regulation compared to lineages that have reverted to solitary nesting, then we would expect to find more motifs with scores that are higher in social taxa compared to secondarily solitary taxa. Only counting those motifs for which there were differences in numbers of significant correlations between social and solitary taxa of at least 20%, there were 94 motifs with more positive correlations with social taxa and 29 motifs with more positive correlations with solitary taxa. Five of the socially-biased motifs were previously associated with eusocial evolution in bees[85], including the motifs for *lola, hairy, CrebA, CG5180,* and the *met/tai* complex (Supplementary Table 4).

This appears to be a large difference in the number of motifs positively correlated with social behavior and positively correlated with solitary behavior. We determined the significance of this result using a permutation test. We reran the PGLS analyses 100 times using six (of 15) randomly chosen species as focal (social) taxa. Only two of these permutations yielded at least 94 motifs with more positive correlations in these taxa. None yielded differences between the number of motifs with more positive correlations with

social behavior and more positive correlations with solitary behavior as or more extreme than our result (94-29=65). Therefore, the pattern found appears to be significantly more extreme than expected by chance (Permutation test p-value < 0.01).

Greater constraint on noncoding elements in social species

Interestingly, genes showing relaxed selection associated with the loss of eusociality are enriched for chromosome condensation (Supplementary Table 6), suggesting there may be trade-offs or costs associated with this increased regulatory potential. To determine whether or not we see a signature of these trade-offs in non-coding sequences, we used phastCons[27] to identify conserved non-exonic elements (CNEEs) in each species. We identified CNEEs that were either slow- or fast- evolving in eusocial and solitary lineages. Using the logic outlined above, we would predict that CNEEs essential to the maintenance of social behavior should be more highly conserved and evolving more slowly in extant eusocial taxa vs. secondarily solitary taxa. After phylogenetic correction, we find 1,876 CNEEs that meet these criteria, with faster rates of evolution in secondarily solitary species (Fig. 2c). This is significantly more than expected by chance (Binomial test, $p=5.27e^{-9}$); these regions are proximal to genes enriched for several GO terms, including serotonin receptor signaling (Permutation tests, $p=0.007$) and adherens junction maintenance (Permutation tests, $p=0.009$; see Supplementary Table 15 for complete list). We also find 1,255 CNEEs that are more highly conserved in secondarily solitary lineages (fast in eusocial; slower in solitary). This number is smaller than expected by chance (Binomial test, $p<1e^{-10}$), but these regions are proximal to genes enriched for cytoskeletal organization and processes related to synapse and neurotransmitter functions, among other GO terms (Permutation tests, $p<0.05$; Supplementary Table 15), potentially providing molecular support to hypotheses that the evolution of eusociality involves altered investments in cognitive function[148,149].

microRNA evolution

We identified a total of 1269 microRNAs (miRs) expressed in the brains of 15 species (including two populations of *L. albipes,* one solitary and one social) (Supplementary Table 2). Most miRs were located in intergenic regions, and target predictions from Miranda and RNAhybrid identified gene targets for all but five sequenced miRs. Predicted targets of miRs were enriched for many GO terms (1857 terms for all miR targets; 1352 terms for targets in social species; 1148 terms for targets in solitary species). Social state of species did not predict the number of miRs expressed, nor the number of targets present across miRs (PGLS, $p>0.05$ in all comparisons).

There was weak but significant overlap for genes under positive selection at the origin of eusociality for Augochlorini and predicted miR targets in *Augochlorella aurata* (Hypergeometric test; RF=1.2, $p=0.024$), but not for other social species in Augochlorini. No significant overlap was found between genes under selection at the Halictini origin of eusociality and miR targets of any species. No significant overlap was found between genes under relaxation of selection associated with solitary reversions and miR targets of any species.

Development of LC-MS method for detection of JH III

We used LC-MS to measure JH III titers across tissues in *Bombus impatiens* as a proof of concept. JH III was detected in multiple bumblebee tissue types, including the brain, hemolymph, and ovaries (Extended Data Fig. 8). Very low levels of JH (similar to dissection solution control) were detected in the retrocerebral (RCC) complex, which synthesizes JH (Extended Data Fig. 8). The RCC immediately releases and is not known to store JH[5]. To determine whether we could track the movement of JH III through bee tissues, we applied JH III labeled with deuterium isotopes (JH III-D3) to the ventral abdominal sternites of *B. impatiens* workers. Labelled JH III-D3 levels were high in both bee bodies and the brain after 24 hours, but decayed significantly by 72 hours in both tissues (Extended Data Fig. 9). JH III-D3 accounts for nearly all the total JH III in bee bodies after 24 hours, indicating that the labelled compound is well-absorbed by the bee (Extended Data Fig. 9). Details of JH III and JH III-D3 quantification are provided in Extended Data Fig. 10. Similar methods were used for detection of JH III and JH III-d3 in *A. aurata* and *A. pura,* results of which are presented in the main text.

Legends for Supplementary Tables 1-15

Supplementary Table 1.
Genome annotation statistics. Species name, total assembly length, karyotype, number of genes included in the assembly, the number of isoforms, and BUSCO scores.

Supplementary Table 2.
Characterization of microRNAs expressed in the brains of each species/population.

Supplementary Table 3.
Correlations between gene age and the proportion of gene sets of interest of each age.

Supplementary Table 4.
TF motif results in promoter regions of social vs. solitary lineages.

Supplementary Table 5.
Orthogroups with significant signatures of selection.

Supplementary Table 6.
GO enrichments among sets of coding sequences with signatures of selection. Parameter provides information on which test was conducted (RELAX or aBSREL), selection type denotes which branches were considered, and/or relevant significance thresholds.

Supplementary Table 7.
Gene expression data for each species and tissue type.

Supplementary Table 8.
Samples included in this study. Sample name, species identification, collection date and location, sex of the individual, library preparation method, BioSample and BioProject accession IDs.

Supplementary Table 9.
Cluster-specific DEGs (FDR < 0.05; 10 PCs, Resolution=0.6) for all 6716 cells in halictid scRNAseq experiment.

Supplementary Table 10.
Cluster-specific DEGs (FDR < 0.05;6 PCs, Resolution=0.3) for all cells in Cluster 6 (glia cluster).

Supplementary Table 11.
Sequences included in evolutionary analyses for apolpp across insects.

Supplementary Table 12.
Hypergeometric tests examining enrichment of tested gene sets.

Supplementary Table 13.
Wilcoxon rank-sum tests of differences in tissue specificity between the genes of interest and all other genes.

Supplementary Table 14.
Numbers of branches under selection in *apolpp* across insects.

Supplementary Table 15.

GO enrichments among sets of CNEEs as determined by permutation tests.

Supplementary References

1.   Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
2.   Davie, K. *et al.* A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* **174**, 982-998.e20 (2018).
3.   Lu, W., Wang, L., Chen, L., Hui, S. & Rabinowitz, J. D. Extraction and Quantitation of Nicotinamide Adenine Dinucleotide Redox Cofactors. *Antioxid. Redox Signal.* **28**, 167–179 (2018).
4.   Wang, L. *et al.* Peak annotation and verification engine (PAVE) for untargeted LC-MS metabolomics. *Anal. Chem.* **91**, 1838–1846 (2019).
5.   Frederik Nijhout, H. *Insect Hormones*. (Princeton University Press, 1998).
6.   Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
7.   Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).
8.   Chakraborty, M., Baldwin-Brown, J. G., Long, A. D. & Emerson, J. J. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res.* **44**, e147 (2016).
9.   Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
10.   Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).
11.   Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019).
12.   Jackman, S. D. *et al.* ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* **27**, 768–777 (2017).
13.   Bioinformatics, B. FastQC: a quality control tool for high throughput sequence data. *Cambridge, UK: Babraham Institute* (2011).
14.   Adey, A. *et al.* In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. *Genome Res.* **24**, 2041–2049 (2014).
15.   Zheng, G. X. Y. *et al.* Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nat. Biotechnol.* **34**, 303–311 (2016).
16.   Burton, J. N. *et al.* Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
17.   Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
18.   Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
19.   Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
20.   Kocher, S. D. *et al.* The draft genome of a socially polymorphic halictid bee, Lasioglossum albipes. *Genome Biol.* **14**, R142 (2013).
21.   Smit, A. F. A. & Hubley, R. RepeatModeler Open-1.0. Preprint at (2008).
22.   Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
23.   Paten, B. *et al.* Cactus: Algorithms for genome multiple sequence alignment. *Genome Res.* **21**, 1512–1528 (2011).
24.   Armstrong, J. *et al.* Progressive alignment with Cactus: a multiple-genome aligner for the thousand-genome era. *Cold Spring Harbor Laboratory* 730531 (2019) doi:10.1101/730531.
25.   Sackton, T. B. *et al.* Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**, 74–78 (2019).
26.   Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
27.   Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
28.   Earl, D. *et al.* Alignathon: a competitive assessment of whole-genome alignment methods. *Genome Res.* **24**, 2077–2089 (2014).
29.   Bradley, R. K. *et al.* Fast statistical alignment. *PLoS Comput. Biol.* **5**, e1000392 (2009).

30. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
31. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
32. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
33. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
34. Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* **40**, 37–52 (2012).
35. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68-73 (2014).
36. Ashby, R., Forêt, S., Searle, I. & Maleszka, R. MicroRNAs in Honey Bee Caste Determination. *Sci. Rep.* **6**, 18794 (2016).
37. Kapheim, K. M. *et al.* Brain microRNAs among social and solitary bees. *R Soc Open Sci* **7**, 200517 (2020).
38. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
39. Enright, A. J. *et al.* MicroRNA targets in Drosophila. *Genome Biol.* **5**, R1 (2003).
40. Krüger, J. & Rehmsmeier, M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* **34**, W451-4 (2006).
41. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767–769 (2016).
42. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-Genome Annotation with BRAKER. in *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) 65–95 (Springer New York, 2019).
43. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
44. Cantarel, B. L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
45. Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
46. Haas, B. J., Zeng, Q., Pearson, M. D., Cuomo, C. A. & Wortman, J. R. Approaches to fungal genome annotation. *Mycology* **2**, 118–141 (2011).
47. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
48. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
49. Zdobnov, E. M. *et al.* OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **45**, D744–D749 (2017).
50. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
51. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
52. Kriventseva, E. V. *et al.* OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **47**, D807–D811 (2019).
53. Bryant, D. M. *et al.* A Tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* **18**, 762–776 (2017).
54. Klopfenstein, D. V. *et al.* GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* **8**, 10872 (2018).
55. Wang, K. & Nishida, H. REGULATOR: a database of metazoan transcription factors and maternal factors for developmental studies. *BMC Bioinformatics* **16**, 114 (2015).
56. Kocher, S. D. *et al.* The genetic basis of a social polymorphism in halictid bees. *Nat. Commun.* **9**, 4338 (2018).
57. Shpigler, H. Y. *et al.* Deep evolutionary conservation of autism-related genes. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9653–9658 (2017).
58. Domazet-Lošo, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23**, 533–539 (2007).
59. Drost, H.-G., Gabel, A., Grosse, I. & Quint, M. Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol. Biol. Evol.* **32**, 1221–1231 (2015).
60. Misof, B. *et al.* Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
61. Branstetter, M. G. *et al.* Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. *Curr. Biol.* **27**, 1019–1025 (2017).
62. Dohrmann, M. & Wörheide, G. Dating early animal evolution using phylogenomic data. *Sci. Rep.* **7**, 3599 (2017).
63. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

64. Bolstad, B. M. & Bolstad, M. B. M. Package 'preprocessCore.' (2013).

65. Pennell, M. W. *et al.* geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* **30**, 2216–2218 (2014).

66. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

67. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

68. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

69. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).

70. Traniello, I. M. *et al.* Meta-analysis of honey bee neurogenomic response links Deformed wing virus type A to precocious behavioral maturation. *Sci. Rep.* **10**, 3101 (2020).

71. Sheng, L. *et al.* Social reprogramming in ants induces longevity-associated glia remodeling. *Sci Adv* **6**, eaba9869 (2020).

72. Yin, J. *et al.* Brain-specific lipoprotein receptors interact with astrocyte derived apolipoprotein and mediate neuron-glia lipid shuttling. *Nat. Commun.* **12**, 2408 (2021).

73. Fisher, C. E. & Howie, S. E. M. The role of *megalin* (LRP-2/Gp330) during development. *Dev. Biol.* **296**, 279–297 (2006).

74. Daneman, R. & Barres, B. A. The blood-brain barrier--lessons from moody flies. *Cell* **123**, 9–12 (2005).

75. Stork, T. *et al.* Organization and function of the blood-brain barrier in *Drosophila*. *J. Neurosci.* **28**, 587–597 (2008).

76. DeSalvo, M. K. *et al.* The *Drosophila* surface glia transcriptome: evolutionary conserved blood-brain barrier processes. *Front. Neurosci.* **8**, 346 (2014).

77. Melom, J. E. & Littleton, J. T. Mutation of a NCKX eliminates glial microdomain calcium oscillations and enhances seizure susceptibility. *J. Neurosci.* **33**, 1169–1178 (2013).

78. Stork, T., Bernardos, R. & Freeman, M. R. Analysis of glial cell development and function in *Drosophila*. *Cold Spring Harb. Protoc.* **2012**, 1–17 (2012).

79. Zhang, H. *et al. Drosophila* CG10527 mutants are resistant to juvenile hormone and its analog methoprene. *Biochem. Biophys. Res. Commun.* **401**, 182–187 (2010).

80. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).

81. Jarvis, E. D. *et al.* Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).

82. Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L. & Scheffler, K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832 (2015).

83. Smith, M. D. *et al.* Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).

84. Gibbs, J., Brady, S. G., Kanda, K. & Danforth, B. N. Phylogeny of halictine bees supports a shared origin of eusociality for *Halictus* and *Lasioglossum* (Apoidea: Anthophila: Halictidae). *Mol. Phylogenet. Evol.* **65**, 926–939 (2012).

85. Kapheim, K. M. *et al.* Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 1139–1143 (2015).

86. Sinha, S., Liang, Y. & Siggia, E. Stubb: a program for discovery and analysis of cis-regulatory modules. *Nucleic Acids Res.* **34**, W555-9 (2006).

87. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 168 (2009).

88. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

89. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–371 (2009).

90. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).

91. Pahlke, S., Seid, M. A., Jaumann, S. & Smith, A. The loss of sociality is accompanied by reduced neural investment in mushroom body volume in the sweat bee *Augochlora Pura* (Hymenoptera: Halictidae). *Ann. Entomol. Soc. Am.* **114**, 637–642 (2021).

92. Eickwort, G. C., Kukuk, P. F. & Wesley, F. R. The Nesting biology of *Dufourea novaeangliae* (Hymenoptera: Halictidae) and the systematic position of the Dufoureinae based on behavior and development. *J. Kans. Entomol. Soc.* **59**, 103–120 (1986).

93. Batra, S. W. T. Behavior of the Alkali Bee, *Nomia melanderi,* within the Nest (Hymenoptera: Halictidae). *Ann. Entomol. Soc. Am.* **63**, 400–406 (1970).

94. Cane, J. H. A native ground-nesting bee (*Nomia melanderi*) sustainably managed to pollinate alfalfa across an intensively agricultural landscape. *Apidologie* **39**, 315–323 (2008).

95. Kapheim, K. M., Smith, A. R., Nonacs, P., Wcislo, W. T. & Wayne, R. K. Foundress polyphenism and the origins of eusociality in a facultatively eusocial sweat bee, *Megalopta genalis* (Halictidae). *Behav. Ecol. Sociobiol.* **67**, 331–340 (2013).

96. Wcislo, W. T. *et al.* The evolution of nocturnal behaviour in sweat bees, *Megalopta genalis* and *M. ecuadoria* (Hymenoptera: Halictidae): an escape from competitors and enemies? *Biol. J. Linn. Soc. Lond.* **83**, 377–387 (2004).

97. Mueller, U. G. Life history and social evolution of the primitively eusocial bee *Augochlorella striata* (Hymenoptera: Halictidae). *J. Kans. Entomol. Soc.* **69**, 116–138 (1996).

98. Packer, L. Solitary and eusocial nests in a population of *Augochlorella striata* (Provancher) (Hymenoptera; Halictidae) at the northern edge of its range. *Behav. Ecol. Sociobiol.* **27**, 339–344 (1990).

99. Stockhammer, K. A. Nesting habits and life cycle of a sweat bee, *Augochlora pura* (Hymenoptera: Halictidae). *J. Kans. Entomol. Soc.* **39**, 157–192 (1966).

100. Barrows, E. M. Soil nesting by wood-inhabiting halictine bees, *Augochlora pura* and *Lasioglossum coeruleum* (Hymenoptera: Halictidae). *J. Kans. Entomol. Soc.* **46**, 496–499 (1973).

101. Abrams, J. & Eickwort, G. C. Biology of the communal sweat bee *Agapostemon virescen*s (Hymenoptera: Halictidae) in New York State. *Search Agriculture* **1**, 1 (1980).

102. Abrams, J. & Eickwort, G. C. Nest switching and guarding by the communal sweat bee *Agapostemon virescens* (Hymenoptera, Halictidae). *Insectes Soc.* **28**, 105–116 (1981).

103. Richards, M. H. & Packer, L. The Socioecology of body size variation in the primitively eusocial sweat bee, *Halictus ligatus* (Hymenoptera: Halictidae). *Oikos* **77**, 68–76 (1996).

104. Packer, L. & Knerer, G. The Biology of a Subtropical Population of *Halictus ligatus* Say (Hymenoptera: Halictidae): I. Phenology and Social Organisation. *Behav. Ecol. Sociobiol.* **18**, 363–375 (1986).

105. Michener, C. & Bennett, F. D. Graphical variation in nesting biology and social organization of *Halictus ligatus*. https://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1093&context=bee_lab_mi (1977).

106. Richards, M. H. Annual and social variation in foraging effort of the obligately eusocial sweat bee, *Halictus ligatus* (Hymenoptera: Halictidae). *kent* **77**, 484–502 (2004).

107. Plateaux-Quénu, C. *La biologie des abeilles primitives*. vol. 11 (Masson, 1972).

108. Sitdikov, A. A. Nesting of the bee *Halictus quadricinctus* (F.)(Hymenoptera, Halictidae) in the Vdmurt ASSR. (1988).

109. Michener, C. D. The Social Behavior of the Bees. *A Comparative Study* (1974).

110. Soucy, S. L. Nesting biology and socially polymorphic behavior of the sweat bee *Halictus rubicundus* (Hymenoptera: Halictidae). *Ann. Entomol. Soc. Am.* **95**, 57–65 (2002).

111. Field, J., Paxton, R. J., Soro, A. & Bridge, C. Cryptic plasticity underlies a major evolutionary transition. *Curr. Biol.* **20**, 2028–2031 (2010).

112. Yanega, D. Caste determination and differential diapause within the first brood of *Halictus rubicundus* in New York (Hymenoptera: Halictidae). *Behav. Ecol. Sociobiol.* **24**, 97–107 (1989).

113. Proulx, A. Demography and molecular ecology of the solitary halictid *Lasioglossum zonulum*: with observations on Lasioglossum leucozonium. (2020).

114. Plateaux-Quénu, C. & Grassé, P.-P. *La biologie des abeilles primitives*. vol. monographie 11 (Masson, 1972).

115. Zayed, A. Characterization of microsatellite loci from the solitary sweat bees *Lasioglossum leucozonium* and *Lasioglossum oenotherae* (Hymenoptera, Halictidae). *Mol. Ecol. Notes* **6**, 1154–1156 (2006).

116. Plateaux-Quénu, C. Un nouveau type de société d'Insectes: Halictus marginatus Brullé (Hym., Apoidea). in *Annales Biologiques* vol. 35 325–455 (ci.nii.ac.jp, 1959).

117. Plateaux-Quénu, C. Biology of Halictus Marginatus Brullé. *J. Apic. Res.* **1**, 41–51 (1962).

118. Wcislo, W. T., Wille, A. & Orozco, E. Nesting biology of tropical solitary and social sweat bees, *Lasioglossum (Dialictus) figueresi Wcislo.* stri-apps.si.edu.

119. Wcislo, W. T. Social interactions and behavioral context in a largely solitary bee, *Lasioglossum (Dialictus) figueresi* (Hymenoptera, Halictidae). *Insectes Soc.* **44**, 199–208 (1997).

120. Wcislo, W. T., Wille, A. & Orozco, E. Nesting biology of tropical solitary and social sweat bees, *Lasioglossum (Dialictus) figueresi Wcislo* and *L.(D.) aeneiventre* (Friese)(Hymenoptera: Halictidae). *Insectes Soc.* (1993) doi:10.1007/bf01338830?no-access=true&lang=en.

121. Wcislo, W. T. Nest localization and recognition in a solitary bee, *Lasioglossum* (*Dialictus*) *figueresi wcislo* (Hymenoptera: Halictidae), in relation to sociality. *Ethology* **92**, 108–123 (2010).

122. Batra, S. W. T. Behavior of the social bee, *Lasioglossum zephyrum*, within the nest (Hymenoptera: Halictidae). *Insectes Soc.* **11**, 159–185 (1964).

123. Batra, S. W. T. The life cycle and behavior of the primitively social bee, *Lasioglossum zephyrum* (Halictidae). *The University of Kansas Science Bulletin* **46**, 359 (1966).

124. Michener, C. D. & Brothers, D. J. Were workers of eusocial Hymenoptera initially altruistic or oppressed? *Proc. Natl. Acad. Sci. U. S. A.* **71**, 671–674 (1974).

125. Buckle, G. R. Differentiation of queens and nestmate interactions of newly established colonies of *Lasioglossum zephyrum* (Hymenoptera: Halictidae)[Kansas]. *Sociobiology* (1982).

126. Knerer, G. Synergistic evolution of halictine nest architecture and social behavior. *Can. J. Zool.* **47**, 925–930 (1969).
127. Packer, L. Multiple-foundress associations in sweat bees. https://agris.fao.org/agris-search/search.do?recordID=GB9710419 (1993).
128. Westrich, P. Die Wildbienen Baden-Württembergs. 2 Bde., 972 S. *Stuttgart (Ulmer)* (1989).
129. Steitz, I., Brandt, K., Biefel, F., Minat, Ä. & Ayasse, M. Queen recognition signals in two primitively eusocial halictid bees: evolutionary conservation and caste-specific perception. *Insects* **10**, (2019).
130. Knerer, G. & MacKay, P. Bionomic notes on the solitary *Evylaeus oenotherae* (Stevens) (Hymenoptera: Halictinae), a matinal summer bee visiting cultivated Onagraceae. *Can. J. Zool.* **47**, 289–294 (1969).
131. Friedel, A., Paxton, R. J. & Soro, A. Spatial patterns of relatedness within nesting aggregations of the primitively eusocial sweat bee *Lasioglossum malachurum*. *Insectes Soc.* **64**, 465–475 (2017).
132. Zobel, M. U. & Paxton, R. J. Is big the best? Queen size, usurpation and nest closure in a primitively eusocial sweat bee (*Lasioglossum malachurum*). *Behav. Ecol. Sociobiol.* **61**, 435–447 (2007).
133. Soro, A., Ayasse, M., Zobel, M. U. & Paxton, R. J. Complex sociogenetic organization and the origin of unrelated workers in a eusocial sweat bee, *Lasioglossum malachurum*. *Insectes Soc.* **56**, 55–63 (2008).
134. Steitz, I. & Ayasse, M. Macrocyclic lactones act as a queen pheromone in a primitively eusocial sweat bee. *Curr. Biol.* **30**, 1136-1141.e3 (2020).
135. Wyman, L. M. & Richards, M. H. Colony social organization of *Lasioglossum malachurum* Kirby (Hymenoptera, Halictidae) in southern Greece. *Insectes Soc.* **50**, 201–211 (2003).
136. Plateaux-Qu nu, C. Comparative biological data in two closely related eusocial species:*Evylaeus calceatus* (Scop.) and *Evylaeus albipes* (F.) (Hym., Halictinae). *Insectes Soc.* **39**, 351–364 (1992).
137. Davison, P. J. & Field, J. Social polymorphism in the sweat bee *Lasioglossum (Evylaeus) calceatum*. *Insectes Soc.* **63**, 327–338 (2016).
138. Davison, P. J. & Field, J. Limited social plasticity in the socially polymorphic sweat bee *Lasioglossum calceatum*. *Behav. Ecol. Sociobiol.* **72**, 56 (2018).
139. Plateaux-Quénu, C. & Packer, L. A test of the mating limitation hypothesis for caste determination in *Evylaeus albipes* (Hymenoptera: Halictidae), a primitively eusocial halictine bee. *J. Insect Behav.* **11**, 119–128 (1998).
140. Plateaux-Quénu, C., Plateaux, L. & Packer, L. Population-typical behaviours are retained when eusocial and non-eusocial forms of *Evylaeus albipes* (F.) (Hymenoptera, Halictidae) are reared simultaneously in the laboratory. *Insectes Soc.* **47**, 263–270 (2000).
141. Plateaux-Quénu, C. Premières observations sur le caractère social d'*Evylaeus albipes* (F.)(Hymenoptera, Halictinae). *Actes Coll. Ins. Soc* (1989).
142. Wyatt, C. D. R. *et al.* Genetic toolkit for sociality predicts castes across the spectrum of social complexity in wasps. *Cold Spring Harbor Laboratory* 2020.12.08.407056 (2020) doi:10.1101/2020.12.08.407056.
143. Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764 (2012).
144. Yates, C. M., Filippis, I., Kelley, L. A. & Sternberg, M. J. E. SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. *J. Mol. Biol.* **426**, 2692–2701 (2014).
145. Pei, J. & Grishin, N. V. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**, 700–712 (2001).
146. Roth, Z. *et al.* Identification of receptor-interacting regions of vitellogenin within evolutionarily conserved β-sheet structures by using a peptide array. *Chembiochem* **14**, 1116–1122 (2013).
147. Lanfear, R. The local-clock permutation test: a simple test to compare rates of molecular evolution on phylogenetic trees. *Evolution* **65**, 606–611 (2011).
148. Dunbar, R. I. M. The social brain hypothesis. *Evol. Anthropol.* **6**, 178–190 (1998).
149. O'Donnell, S. *et al.* Distributed cognition and social brains: reductions in mushroom body investment accompanied the origins of sociality in wasps (Hymenoptera: Vespidae). *Proceedings of the Royal Society B: Biological Sciences* **282**, 20150791 (2015).