nature ecology & evolution

Article

https://doi.org/10.1038/s41559-023-02001-3

Convergent and complementary selection shaped gains and losses of eusociality in sweat bees

Received: 9 August 2022

Accepted: 18 January 2023

Published online: 20 March 2023

Check for updates

Beryl M. Jones[1,2,15], Benjamin E. R. Rubin[1,2,15], Olga Dudchenko[3,4,15], Callum J. Kingwell[1,2,5], Ian M. Traniello[1,2], Z. Yan Wang[1,2], Karen M. Kapheim[5,6], Eli S. Wyman[1,2], Per A. Adastra[3], Weijie Liu[1,2], Lance R. Parsons[2], S. RaElle Jackson[2], Katharine Goodwin[2], Shawn M. Davidson[2], Matthew J. McBride[2,7], Andrew E. Webb[1,2], Kennedy S. Omufwoko[1,2], Nikki Van Dorp[1,2], Mauricio Fernández Otárola[8], Melanie Pham[3], Arina D. Omer[3], David Weisz[3], Joshua Schraiber[9,10], Fernando Villanea[9,11], William T. Wcislo[5], Robert J. Paxton[12,13], Brendan G. Hunt[14], Erez Lieberman Aiden[3,4,16] & Sarah D. Kocher[1,2,16] ✉

Sweat bees have repeatedly gained and lost eusociality, a transition from individual to group reproduction. Here we generate chromosome-length genome assemblies for 17 species and identify genomic signatures of evolutionary trade-offs associated with transitions between social and solitary living. Both young genes and regulatory regions show enrichment for these molecular patterns. We also identify loci that show evidence of complementary signals of positive and relaxed selection linked specifically to the convergent gains and losses of eusociality in sweat bees. This includes two pleiotropic proteins that bind and transport juvenile hormone (JH)—a key regulator of insect development and reproduction. We find that one of these proteins is primarily expressed in subperineurial glial cells that form the insect blood–brain barrier and that brain levels of JH vary by sociality. Our findings are consistent with a role of JH in modulating social behaviour and suggest that eusocial evolution was facilitated by alteration of the proteins that bind and transport JH, revealing how an ancestral developmental hormone may have been co-opted during one of life's major transitions. More broadly, our results highlight how evolutionary trade-offs have structured the molecular basis of eusociality in these bees and demonstrate how both directional selection and release from constraint can shape trait evolution.

Organisms situated at the inflection point of life's major evolutionary transitions provide a powerful framework to examine the factors shaping the evolution of traits associated with these transitions[1–4]. Halictid bees ('sweat bees', Hymenoptera: Halictidae) offer a unique opportunity to study the evolution of eusociality (social colonies with overlapping generations and non-reproductive workers), since within this group there have been two independent gains[5] and a dozen losses[6] of eusociality.

The term eusocial was originally coined to describe sweat bee societies[7], whose colonies typically consist of a single reproductive

A full list of affiliations appears at the end of the paper. ✉e-mail: skocher@princeton.edu

https://doi.org/10.1038/s41559-023-02001-3

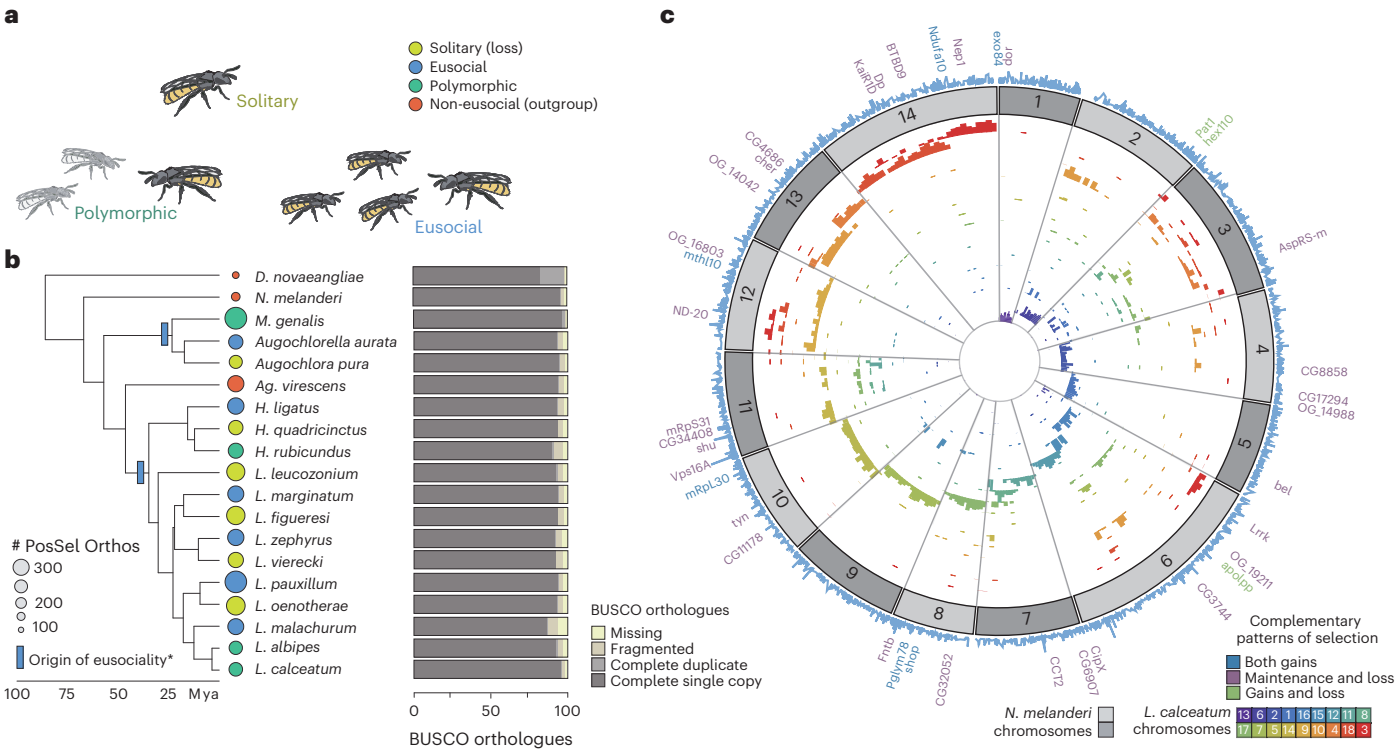


Fig. 1 | Comparative genomic resources for halictid bees. a, Halictid bees encompass a wide range of behaviours, including solitary (yellow), eusocial (blue) and polymorphic species (green) capable of both solitary and eusocial reproduction. Non-eusocial outgroups (red) reproduce independently. **b**, Left: pruned tree showing taxa included in this dataset. Colours at tips indicate species' behaviour and circle sizes are proportional to the number of orthologues under positive selection on each terminal branch (abSREL, HyPhy[134] FDR < 0.05). Light blue rectangles denote gains of eusociality (*inferred from much broader phylogenies[5] than shown here). Right: proportions of complete/fragmented/missing BUSCO orthologues are shown for each genome. **c**, Genomic data aligned to *Nomia melanderi* (NMEL). The 14 NMEL chromosomes are represented as a circular ideogram, with consecutive chromosomes shown in alternating dark/light grey. The inner spiral comprises 18 colour-coded tracks, each corresponding to one *L. calceatum* (LCAL) chromosome; the *y* axis represents the frequency of regions aligning to the corresponding region in NMEL. Most alignments fall into a single 'wedge', indicating that each LCAL chromosome corresponds to just one NMEL chromosome, a pattern typical for sweat bees and unlike that of mammals (Extended Data Figs. 3 and 4). Outer blue line plot indicates number of branches where positive selection was detected at each gene (abSREL, FDR < 0.05), and gene names shown are those with convergent/complementary patterns of selection: positive selection at both origins (blue), intensification of selection in extant eusocial lineages and relaxation of selection in secondarily solitary species (HyPhy RELAX, FDR < 0.1; purple), and complementary patterns of both positive selection on the origin branches and convergent relaxation of selection with losses of eusociality (green). Not all genes in these categories are annotated in NMEL and some are therefore not labelled.

female and a small number of non-reproductive workers (most often ranging between 2 to 12 individuals; Extended Data Fig. 1)[8]. Colony sizes can be quite variable, with up to 400 workers documented in *Lasioglossum marginatum* colonies[9]. Unlike honey bees and ants, sweat bee workers are totipotent—castes are not developmentally determined and all adult females are capable of mating and reproduction in the absence of a queen[10]. Due to the independent gains and losses of eusociality in this group, closely related halictid species encompass a broad spectrum of social behaviour, from solitary individuals that live and reproduce independently to eusocial nests where individuals cooperate to reproduce as a group, and even polymorphic species that produce both solitary and social nests[2]. This evolutionary replication enables a comparative approach to identify the core factors that shape the emergence and breakdown of eusociality.

To identify these factors, we generated 15 de novo genome assemblies and updated 2 additional assemblies in halictids[3,11], all with chromosome-length scaffolds (Fig. 1 and Extended Data Fig. 2). We also included two previously published genomes[12,13]. We selected species with well-characterized social behaviours that encompass both origins and six repeated losses of eusociality (these gains and losses were previously established on the basis of a much broader phylogeny used to reconstruct the origins of eusociality in this group[5]), as well as four socially polymorphic species. Sampling closely related eusocial and solitary species alongside known non-eusocial outgroups provides a powerful framework to examine the molecular mechanisms shaping the evolution of social behaviour in these bees[1–4].

We searched for signatures of positive selection associated with the convergent gains of eusociality as well as signatures of relaxed selection when eusociality is lost. These complementary patterns indicate genomic loci that are associated with costs or trade-offs underlying the maintenance of social traits. We find that some of the targets of selection implicated in the origins and elaborations of eusociality, such as young, taxonomically restricted genes[14–21] and gene regulatory elements[12,20,22,23], also show relaxation of selective pressures when social behaviour is lost. In addition, we uncovered four genes strongly associated with the evolution of eusociality in halictid bees, including two genes that encode primary juvenile hormone binding proteins (JHBPs): *apolipoprotein*[24–26] and *hexamerin110*[27,28]. Using single-cell RNA-sequencing, we localized the expression of *apolipoprotein* in the brain to glial cells involved in forming the insect 'blood–brain barrier' (BBB)[29,30]. In addition, we found evidence that eusocial reproductive females have increased levels of JH III in their brains compared with their solitary counterparts; this could potentially be mediated by changes in the transport of JH. These results provide new insights into how JH signalling may have been modified to shape the evolution of eusociality.

https://doi.org/10.1038/s41559-023-02001-3

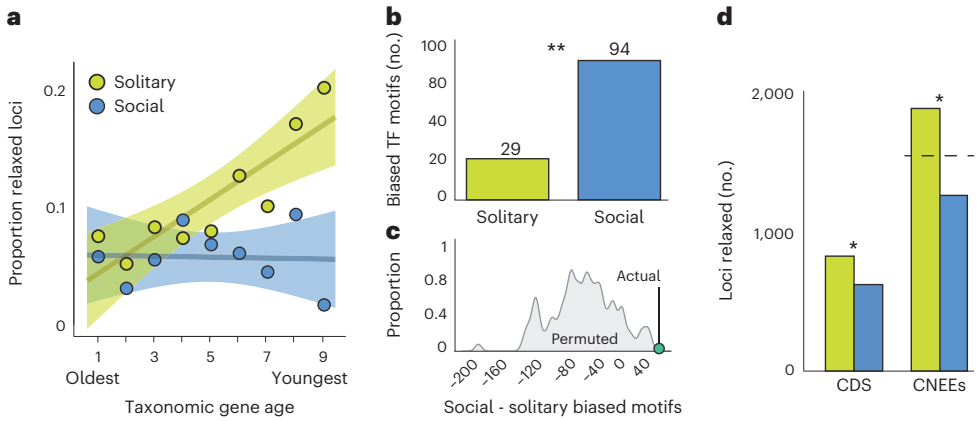


Fig. 2 | The maintenance of eusociality is associated with young genes and gene regulation. a, Younger genes are more likely to show signatures of relaxed selection when social behaviour is lost. Circles show the proportion of genes in each age class that show evidence of relaxed selection in solitary (yellow) or social (blue) lineages, from the oldest Bilaterian group (Age = 1) to the youngest halictid-specific taxonomically restricted genes (Age = 9). For solitary lineages, gene age is significantly correlated with the proportion of those genes with evidence of relaxed selection (Pearson correlation, $r = 0.869$, $P = 0.002$; social lineages, $r = -0.0447$, $P = 0.91$). Shading represents 95% confidence intervals. **b**, Stubb scores[44] were calculated for 223 *Drosophila* transcription factor binding motifs in each genome, and each motif was tested for overrepresentation in solitary/social genomes; 94 motifs were enriched in social genomes compared with 29 motifs enriched in solitary genomes. $^{**}P < 0.01$ for permutation test as shown in **c**. **c**, Permutation tests reveal that the ~3-fold enrichment in **b** is unlikely to occur by chance (empirical $P < 0.01$). **d**, Taxa that have lost eusociality have higher proportions of loci experiencing relaxed selection after phylogenetic correction, both for coding sequences (CDS; Fisher's Exact test, $P = 2.42 \times 10^{-7}$, odds-ratio = 1.48) and for CNEEs. Social lineages have fewer fast-evolving CNEEs than chance (Binomial test, $P < 1 \times 10^{-10}$), while solitary taxa have more fast-evolving CNEEs than chance (Binomial test, $P = 5.27 \times 10^{-9}$). The dashed line indicates the null expectation for CNEEs, and * indicates significant differences in the number of loci between eusocial and solitary species.

Results

We generated new comparative genomic resources for studying the evolution of eusociality in halictid bees. Genome assemblies of 17 species ranged in size from 267 to 470 Mb, with estimated numbers of chromosomes ranging from 9 to 38 (Supplementary Table 1). Broadly, we found that in contrast to mammalian species (Extended Data Fig. 3), genomic rearrangements among the bee species occur disproportionately within rather than between chromosomes (see Fig. 1c and Extended Data Fig. 4). Consequently, loci that are on the same chromosome in one bee species also tend to occur on the same chromosome in other bee species. This observation is similar to previous findings in dipteran genomes[31,32] and may indicate a broader trend during the evolution of insect chromosomes. To increase the quality of genome annotations, we generated tissue-specific transcriptomes for 11 species. In addition, because of new roles emerging for microRNAs in social plasticity and eusocial evolution[33–40], we also sequenced and characterized 1,269 microRNAs expressed in the brains of 15 species (Supplementary Table 2). The number of annotated genes ranged from 11,060 to 14,982, and BUSCO[41] analyses estimated the completeness of our annotations to range from 93.4 to 99.1% (Fig. 1b and Supplementary Table 1). Whole-genome alignments were generated with progressiveCactus[42] (Fig. 1c), which we then used to identify 52,421 conserved, non-exonic elements (CNEEs) present in 10 or more species. All genomes, annotations and alignments can be viewed and downloaded from the Halictid Genome Browser (https://beenomes.princeton.edu).

Signatures of trade-offs on young genes and gene regulation

Previous studies of eusociality have suggested that similar to their importance in the evolution of other novel traits[43], younger or taxonomically restricted genes (TRGs) may play key roles in the evolution of eusocial behaviour[14–21]. To test this hypothesis, we examined the relationship between gene age and selection associated with eusocial origins, maintenance and reversions to solitary life histories. We found that a greater proportion of young genes compared with old genes experience relaxed selection when eusociality is subsequently lost (Fig. 2a; Pearson's $r = 0.869$, $P = 0.002$). This relationship neither holds for orthologues showing evidence of relaxed selection on eusocial branches, nor are younger genes more likely than older genes to experience intensified selection associated with either the origins or maintenance of eusociality (Extended Data Fig. 5 and Supplementary Table 3). We note that we did not include polymorphic species as focal branches in any of these or subsequent analyses of selection because we could not classify these lineages as either eusocial or solitary.

Gene regulatory changes have also been implicated in eusocial evolution[12,20,22,23], including the expansion of transcription factor (TF) motifs in the genomes of eusocial species compared with distantly related solitary species[12]. To assess the degree to which changes in gene regulation may facilitate the evolution of social behaviour in halictids, we characterized TF motifs in putative promoter regions in each halictid genome. For each species, we defined these regions as 5 kb upstream and 2 kb downstream of the transcription start site for each gene[12] and calculated a score for a given TF in each region that reflects the number and/or strength of predicted binding sites[44].

If social species have a greater capacity for gene regulation compared with lineages that have reverted to solitary nesting, then we would expect to find more motifs with scores (reflecting both strength and number of binding sites) that are higher in social taxa compared with secondarily solitary taxa. In support of this hypothesis, we find a greater than 3-fold enrichment of TF binding motifs that are positively correlated with social lineages compared with secondarily solitary lineages after phylogenetic correction (Fig. 2b and Supplementary Table 4), and permutation tests indicate that this difference is highly significant (Fig. 2c). Five of these socially biased motifs were previously identified as associated with eusocial evolution in bees[12], including the motifs for Lola, Hairy, CrebA, CG5180 and the Met/Tai complex, which initiates downstream transcriptional responses to JH.

In addition to changes in TF motif presence, we also tested for signatures of selection on non-coding regions of the genome that potentially play a role in gene regulation[45]. Specifically, we used whole-genome alignments to identify CNEEs, and found that they showed a bias towards faster rates of evolution in secondarily solitary

https://doi.org/10.1038/s41559-023-02001-3

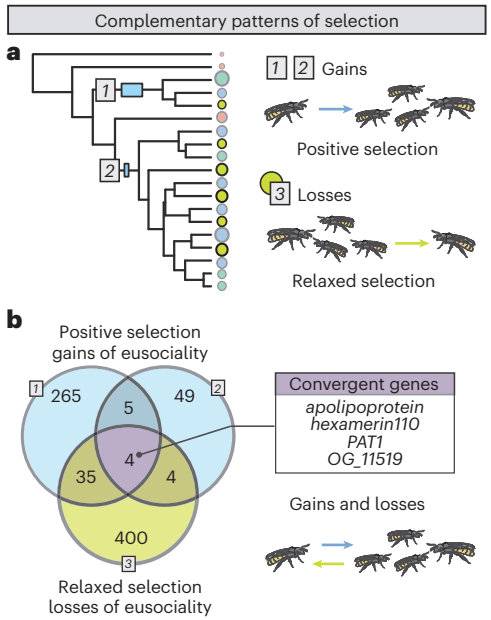


Fig. 3 | Signatures of selection associated with the gains and losses of eusociality in halictids. a, Orthologues were tested for evidence that d*N*/d*S* > 1 at a proportion of sites on focal branches (Gains 1 and 2 denoted with grey squares 1 and 2; abSREL[134] tests in HyPhy, FDR < 0.05) and for evidence of relaxed selection (grey square 3, yellow circles; HyPhy RELAX[54], FDR < 0.1). **b**, Nine loci overlapped both origins (enrichment ratio = 3.97, Fisher's Exact test $P = 0.004$). Four orthologues overlapped all 3 tests (Multi-set Exact test[57]; fold-enrichment = 8.85, $P = 0.001$). Socially polymorphic species were not included as focal branches in either test.

species compared with eusocial lineages (Fig. 2d). This acceleration is likely to be a signature of relaxed constraint in solitary lineages, further supporting the role of gene regulation in the maintenance of eusocial traits.

Convergent, complementary selection across social transitions

To determine whether specific loci show convergent signatures of selection associated with the evolution of eusociality, we identified genes with positive selection on each branch representing a gain of eusociality. We found 309 genes on the Augochlorini origin branch and 62 genes on the Halictini origin branch with evidence of positive selection (Fig. 3 and Supplementary Table 5). On the Augochlorini branch, genes with signatures of positive selection were enriched for cell adhesion (GO:0007155, Fisher's Exact test, $q = 2.01 \times 10^{-4}$; enrichment = 2.92; Supplementary Table 6). Positively selected genes also included *taiman*, which encodes a protein that co-activates the bHLH-PS transcription factor and JH receptor, Met[46–49] (Supplementary Table 5). There was no detectable gene ontology (GO) enrichment among positively selected genes on the Halictini branch (Fisher's Exact test, $q > 0.1$). Nine genes showed signatures of positive selection on both branches (Supplementary Table 5). These genes included *shopper*, which encodes a protein involved in neuronal network function and the ensheathing of glial cells[50], and *ND-42*, encoding a suppressor of Pink1, which is associated with neurocognitive functioning[51].

A unique attribute of halictid bees is that in addition to repeated gains[5], there have also been a number of independent losses of eusociality[4]. These reversals provide a powerful lens to identify key genomic factors needed for the maintenance of social living because organisms are expected to minimize investment in traits when social behaviours are lost or unexpressed. This results in the reduction or removal of selective pressures previously maintaining these costly but essential traits[52,53]. Thus, we predicted that genes associated with trade-offs or costly roles in maintaining eusocial societies should show relaxation of constraint in species that have secondarily reverted to solitary nesting. Consistent with this hypothesis, we found 443 genes showing evidence of convergent relaxed selection on the six branches representing independent losses of eusociality (HyPhy RELAX[54], false discovery rate (FDR) < 0.1; Supplementary Table 5). In contrast, we did not find any genes with evidence of convergent positive selection on all of the solitary loss branches. Genes showing evidence of relaxed selection with eusocial losses were enriched for chromosome condensation (GO:0030261, Fisher's Exact test, $q = 0.067$, enrichment = 4.09), indicating that they may play a role in chromosome accessibility and gene regulation[55]. They were also enriched for vacuolar transport (GO:0007034, Fisher's Exact test, $q = 0.067$, enrichment = 3.11; Supplementary Table 6).

To determine whether this pattern is unique to the loss of eusociality, we ran the same tests for relaxed selection using extant eusocial lineages as the focal branches. We found 305 genes with signatures of relaxation in eusocial species (HyPhy RELAX[54], FDR < 0.1; Supplementary Table 5) enriched for four GO terms related to metabolism (Supplementary Table 6). This is a significantly lower proportion of genes experiencing relaxed selection in eusocial species compared with those experiencing relaxed selection among solitary species (Fisher's Exact test $P = 2.42 \times 10^{-7}$, 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.

We also identified 34 genes that show intensification of selection on extant eusocial lineages and relaxation in secondarily solitary species (HyPhy RELAX, FDR < 0.1 for both tests). The convergent intensification of selection on eusocial lineages suggests that these genes are likely to be particularly relevant to the maintenance or elaboration of eusociality. They were enriched for regulation of SNARE complex assembly (GO:0035542, Fisher's Exact test, $q = 0.074$, enrichment = 80.91; Supplementary Table 6), which is a key component of synaptic transmission that has also been implicated with variation in social behaviour in *L. albipes*[3] and wasps[56].

By comparing genes associated with the emergence of eusociality to those associated with its loss, we have the unique ability to identify some of the most consequential molecular mechanisms shaping social evolution. If a shared set of genes is associated with the emergence and maintenance of social behaviour in this group, then we would expect to find genes experiencing both positive selection when eusociality emerges and relaxed selection when social behaviour is lost (Fig. 3b). Indeed, we find four genes matching these criteria: *OG_11519*, a gene with no known homologues outside of Hymenoptera, *Protein interacting with APP tail-1* (*PAT1*), *apolipoprotein* (*apolpp*) and *hexamerin110* (*hex110*; Fig. 3b). This overlap is significantly more than expected by chance (Multi-set Exact test[57]; fold-enrichment = 8.85, $P = 0.001$). *OG_11519* has no identifiable protein domains but is conserved throughout the Hymenoptera. PAT1 modulates mRNA transport and metabolism[58], and ApoLpp and Hex110 are pleiotropic proteins with roles in storage and lipid transport. Hex110 and ApoLpp have also been established as the primary JHBPs across multiple insect orders[25–27,59]. These complementary patterns of positive selection when eusociality arises and relaxation of selection when it is lost suggest that this small but robust set of genes is associated with costly trade-offs linked to the evolution of eusociality[53].

Selection on juvenile hormone binding and transport

Two of the four genes that show signatures of both positive selection when eusociality is gained and relaxed selection when eusociality is lost (*apolpp* and *hex110*) encode the primary binding proteins for JH, an arthropod-specific hormone with pleiotropic effects on numerous insect life history traits[60]. While ApoLpp and Hex110 have been characterized primarily in non-hymenopteran insects[61,62],

https://doi.org/10.1038/s41559-023-02001-3

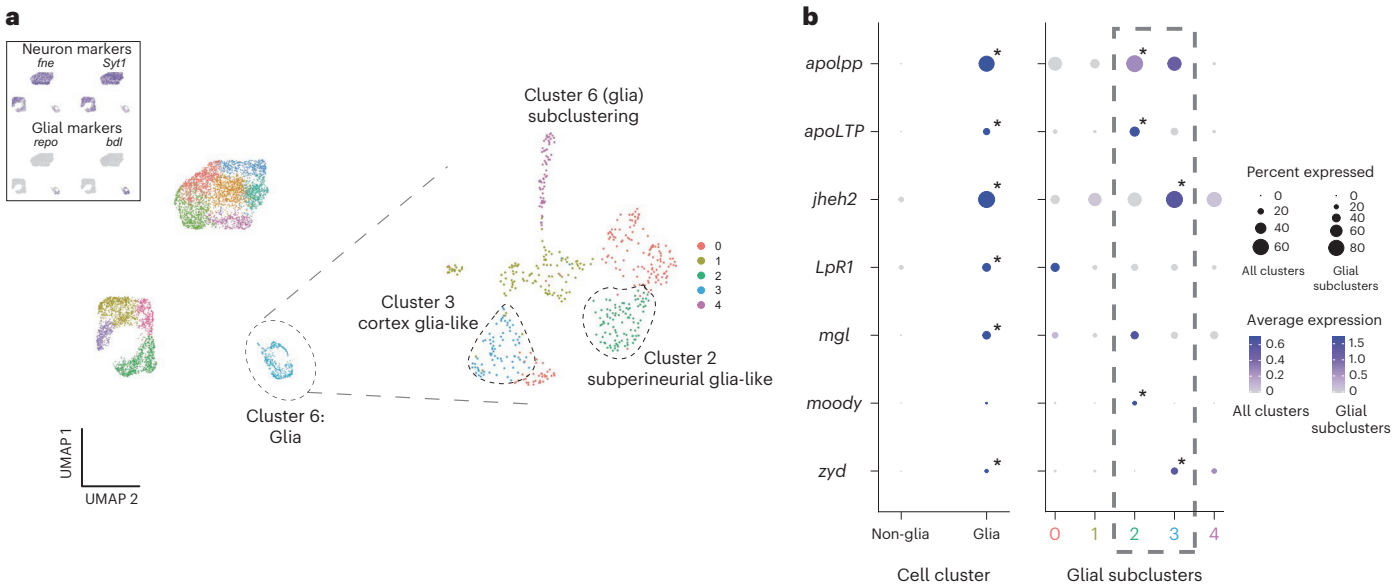


Fig. 4 | *apolpp* and associated lipid transport genes are expressed in glial cells. a, scRNA-seq of the halictid brain. Grouping and visualization of ~7,000 cells from 4 halictid brain samples revealed 11 cell clusters in reduced dimensionality (UMAP) space. Top left: expression of canonical markers of insect brain-cell types showed that only a single cluster, Cluster 6, had low expression levels of neuronal markers *fne* and *Syt1* and high expression levels of glial markers *repo* and *bdl*, suggesting that Cluster 6 is composed of glial cells. Focal subclustering of Cluster 6 identified 5 cell subclusters, 2 of which contained known markers of insect cortex or subperineurial glia. Purple colouration in top left panel corresponds to normalized expression levels of the gene listed above the plot, with darker colour representing higher expression. **b**, Four genes associated with lipid binding, including *apolpp*, *apoLTP, LpR1 and mgl*, as well as *jheh2*, were co-expressed in Cluster 6 ('glia') compared with remaining cell clusters ('non-glia'). Focal subclustering of Cluster 6 into 5 subclusters (coloured numbers 0–4 correspond to subclusters in **a**) revealed that *apolpp* and *Apoltp* are co-expressed with *moody*, a marker of subperineurial glia (glia Subcluster 2), and *jheh2* is co-expressed with *zyd*, a marker of cortex glia (glia Subcluster 3). Circle diameter corresponds to the percentage of cells within a given cell type/subcluster (column) that express a given gene (row), and circle colour corresponds to that gene's average expression within a given cell type/subcluster following sequencing depth normalization. Asterisks in the top-right of some circles indicate that a specific gene is also significantly upregulated in a given cell type/subcluster compared with others following differential expression testing with the MAST algorithm.

they function as JH binding proteins in a wide range of species[25–27,59], suggesting that their function is conserved. In addition to binding JH, these proteins have additional pleiotropic functions related to nutrient storage[63], lipid transport[62] and cuticular hydrocarbon transport[64–67], all of which may also play important roles in the evolution of social traits[27,68–73].

To further investigate the evidence for selection on these genes, we implemented a mixed-effects maximum-likelihood approach[74] and found region-specific, faster rates of evolution on eusocial branches compared with non-eusocial outgroups for both proteins. Sites with evidence of positive selection are present in the functional regions of both proteins (Extended Data Fig. 6), including the receptor binding domain and predicted binding pocket for ApoLpp, as well as in all three Haemocyanin domains of Hex110 (associated with storage functions[61]) and its predicted binding pocket. More broadly, our analyses and previous studies[61] demonstrate that these proteins are rapidly evolving within the Hymenoptera (Extended Data Fig. 7). Although both of these proteins are highly pleiotropic[61,62], their shared role in JH binding and transport suggests that positive selection may have shaped JHBP function as eusociality emerged in two independent lineages of halictids and that some of these changes may also be associated with costs when eusociality is lost.

apolpp expression patterns and JH III levels in the brain

While associations between circulating JH levels and division of labour are well established in the social insects[71,75–79], we still do not understand which components of JH signalling pathways are targeted by natural selection to decouple JH from its ancestral role in development and reproductive maturation[80] and generate new links between JH and social traits. Because ApoLpp delivers cargo to target tissues and has been shown to cross the BBB[81], we hypothesized that differences in ApoLpp transport and the availability of JH in the brain can help generate novel relationships between JH and behaviour. To examine the potential role of cell-type-specific expression of *apolpp* in modulating JH signalling in the brain, we generated a single-cell RNA-sequencing (scRNA-seq) brain dataset using two sweat bee species: *L. albipes* and *L. zephyrus* (Fig. 4a). We identified one cluster, characterized by markers of glial cells, to be the primary location of *apolpp* brain expression (Fig. 4b). In addition, three related lipid transfer-associated genes, *apolipoprotein lipid transfer particle* (*apoLTP*), *megalin* (*mgl*; experiencing relaxed selection in solitary lineages; Supplementary Table 5) and *Lipophorin receptor 1 (LpR1)*, as well the gene encoding the JH degradation enzyme, Jheh2, are expressed primarily in these glial cells (Fig. 4b). Further subclustering demonstrates that both *apolpp* and *apoLTP* are primarily expressed in a subperineurial glia-like cluster (Fig. 4b); subperineurial glia, along with perineurial glia, form and regulate the permeability of the BBB in *D. melanogaster*[29,30].

The actions of JH on the brain are sometimes assumed to be indirect[82], and while JH has previously been quantified in whole-insect heads[83], it has not yet been quantified directly in the insect brain. We used liquid chromatography–mass spectrometry (LC–MS) to measure JH III titres in dissected brains (see Methods and Extended Data Figs. 8–10) and found higher concentrations of JH III in social (*A. aurata*) foundresses compared with solitary (*A. pura*) foundresses; social *A. aurata* workers appear to have intermediate levels of JH III (Fig. 5a). Next we used topical abdominal applications of isotopically labelled JH III-d3 to demonstrate that the bee brain is permeable to JH III (Fig. 5b). Taken together, our results suggest differential responses to JH in the brain and other tissues could be associated with behavioural polyphenisms among the social insects[80].

https://doi.org/10.1038/s41559-023-02001-3

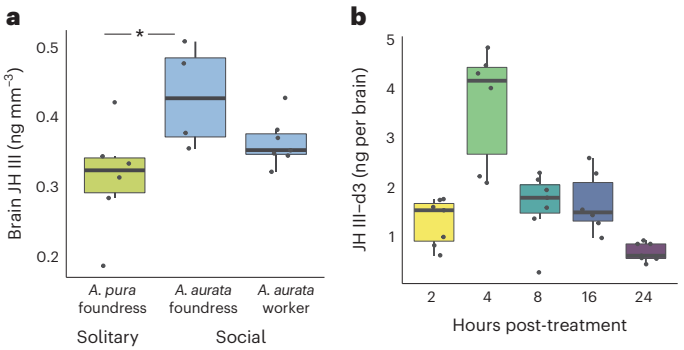


Fig. 5 | JH is higher in brains of eusocial foundresses and crosses the BBB. **a**, LC–MS quantification of JH III in the brains of solitary *A. pura* ($n = 6$; green boxplot) and eusocial A. aurata (blue boxplots) foundresses ($n = 4$) and workers ($n = 7$) reveals higher endogenous levels of JH III in social foundresses (Wilcoxon rank sum test, $P = 0.029$; * indicates post-hoc pairwise comparison, $P < 0.05$). **b**, Isotopically labelled JH III-d3 applied to the abdomens of *A. aurata* is detected in the brain as soon as 2 h later and peaks at 4 h post-treatment ($n = 7$ for 2, 8 and 24 h; $n = 6$ for 4 and 16 h; colours denote sampling timepoint). Boxes show median, 25th and 75th percentiles. Whiskers show minimum and maximum values without outliers.

Discussion

We leveraged the powerful natural variation in social behaviour of sweat bees and developed new comparative genomic resources to characterize the mechanisms that shape transitions in social evolution. In addition to multiple gains of eusociality, halictids provide an excellent opportunity to study genomic signatures of eusocial loss, with repeated recent reversions from eusocial to solitary life history strategies[1,2,4]. By studying both gains and losses within this group, we have uncovered multiple genome-wide patterns as well as specific targets of selection associated with eusociality.

First, we tested for broad patterns in genome evolution that have previously been implicated in the elaboration or maintenance of eusociality, including a role for younger or taxonomically restricted genes (TRGs)[14–21] and changes in gene regulation[12,20]. We found several lines of evidence suggesting that these patterns extend to social evolution in sweat bees. Our finding that TRGs disproportionately experience a relaxation of selection pressure when social behaviour is lost suggests that in addition to influencing the maintenance of eusociality, younger genes are associated with costs or trade-offs linked to eusociality. These findings support studies of complex eusocial hymenopterans, including honey bees, ants and wasps, in which young TRGs have been linked to the evolution of non-reproductive workers[14–21]. Our work extends this body of evidence to suggest that evolutionary changes in these TRGs may also incur costs when lineages revert to solitary strategies. We also find evidence that changes in gene regulation are associated with the elaboration of eusocial traits in halictids, similar to other eusocial lineages[12,20,84]. TF motifs are expanded in social halictid genomes compared with secondarily solitary lineages, implicating more complex gene regulatory networks associated with eusociality. Further, many putative regulatory regions of halictid genomes (CNEEs[45]) show faster rates of evolution in secondarily solitary lineages. This acceleration is likely to be a signature of relaxed constraint in solitary lineages, further supporting the role of gene regulation in the maintenance of eusocial traits.

In addition to the genome-wide patterns associated with eusocial transitions, the repeated gains and losses of eusociality among halictids enabled us to probe the most consequential molecular mechanisms associated with social evolution. We identified genes experiencing positive selection when eusociality is gained and relaxation of selection when eusociality is lost. Our results suggest that the loss of eusociality is associated with a release of selective constraint on a shared set of genes in halictid bees. A similar comparative genomic analysis of social spiders also linked relaxation of selection with social evolution, although in spiders, there is greater relaxed selection in social lineages, where social species have elevated genome-wide rates of molecular evolution. These elevated rates are likely to be driven largely by demographic factors, including a reduced effective population size and increased inbreeding linked to a social life history in this group[85].

We identified a small but robust set of genes with complementary signatures of selection associated with both the gains and losses of eusociality in sweat bees. These patterns highlight the importance of these genes in eusocial lineages: complementary signatures of positive selection when eusociality is gained and relaxed selection when eusociality is lost suggest that they are also associated with costly trade-offs[53]. Strikingly, 2 of these 4 genes (*apolpp* and *hex110*) encode the primary binding proteins for juvenile hormone (JH), a hormone that regulates many aspects of insect life history including development, reproduction, diapause and polyphenism[75,86]. Together, ApoLpp and Hex110 are thought to bind nearly all JH in insect haemolymph[59]. We identified positive selection on sites present in the functional regions of both ApoLpp and Hex110, with faster rates of evolution on eusocial branches compared with non-eusocial lineages. Evolutionary changes to ApoLpp and Hex110 that modify binding affinity and/or cellular uptake could alter levels of JH[87], potentially leading to discrete behavioural phenotypes[80].

In *Drosophila*, ApoLpp forms a complex with other lipoprotein particles and can cross the BBB[81]. Our single-cell transcriptomics dataset reveals that *apolpp* is expressed in the brain (in addition to the fat body; Supplementary Table 7) and enriched in the cell cluster expressing markers of subperineurial glia, a glial subtype that contributes to the formation and permeability of the BBB. Thus, our data suggest that ApoLpp may have a similar role in mediating transport of cargo to the brain in bees. Whether glial-expressed ApoLpp has unique isoforms or different functions compared with ApoLpp expressed in other tissue types is unknown, but the interaction between ApoLpp and other glial-expressed proteins could facilitate brain-specific responses to JH signalling.

These findings suggest a model where glial expression of ApoLpp and other lipid transport proteins may work in concert with JH degradation enzymes to modulate the uptake and availability of JH to the brain in a way that could differentiate social behaviours in halictids. This model is consistent with theory suggesting that conditional expression of a trait (that is eusociality) leads to independent selection pressures and evolutionary divergence[88]; evolution of JHBPs may be one example of such divergence associated with conditional expression of social behaviour. Lending support to this hypothesis, we found that endogenous levels of JH III are higher in the brains of social compared with closely related solitary female foundresses. Because the effects of JH on the brain are sometimes assumed to be indirect[82], we also used isotopically labelled JH to demonstrate that JH is able to cross the BBB in bees, including halictids.

Similar hormonal gatekeeping mechanisms have also been recently proposed in ants[89], suggesting that the regulation of JH in the brain may be a convergently evolved feature of caste differentiation in the social insects. In addition, modifications to JH response-elements may have also helped to fine-tune JH signalling in this group of bees. For example, following transport to relevant tissues, JH binds to the JH receptor, Met/Gce, which when coupled to a co-receptor, Taiman, initiates downstream effects[47]. We found evidence for positive selection on *taiman* linked to the origin of eusociality in the Augochlorini, as well as expansion of the Met/Tai TF motif in eusocial lineages. Future studies are needed to test each of these hypotheses and elucidate the relative influences of modulating JH availability versus refining downstream responses to JH in the origins and elaboration of social traits.

JH is essential to reproductive maturation in solitary insects, but this signalling system has also been frequently co-opted during major life history transitions[60,86,90,91], including eusociality[75]. In relatively small eusocial societies such as halictids[76,92], paper wasps[77] and bumblebees[93,94], JH has maintained its association with reproductive maturation, but has also gained a new role in mediating aggression and dominance. In the more elaborate social colonies of honey bees, modifications to JH pathways have also resulted in novel relationships between JH and vitellogenin[95–97], and the reproductive ground plan hypothesis[95–97] has been proposed to explain this secondary decoupling of JH in workers, independent of its ancestral reproductive role[75]. More generally, it has been hypothesized that the origins of queen and worker castes in social insects can be linked to an evolutionary decoupling of reproductive and non-reproductive behavioural programmes plastically expressed throughout the life cycle of a solitary ancestor[98–100]. This hypothesis, known as the ovarian ground plan hypothesis, specifically implicates modification to hormone signalling as a major route for the evolution of social insect castes. Our finding of convergent selection associated with social transitions on genes that bind and transport JH adds another layer of support to these predictions.

We identified a shared set of genes experiencing selection associated with the origins of eusociality in sweat bees, supporting the hypothesis that a shared genetic toolkit could facilitate the evolution of social behaviour[101–106]. Moreover, the sweat bee system provides a unique opportunity to identify convergent and complementary patterns of selection associated with both gains and losses of this trait. Our findings demonstrate a role for both directional selection and release from constraint in social evolution, and reveal how evolutionary trade-offs can structure the molecular underpinnings of eusociality.

Conclusions

Sweat bees repeatedly traversed an evolutionary inflection point between a solitary lifestyle and a caste-based eusocial one with multiple gains and losses of this trait. We developed new comparative genomic resources for this group and identified complementary signatures of convergent selection associated with the emergence and breakdown of eusociality. Factors associated with the origins or maintenance of eusociality are also associated with its loss, indicating that there may be trade-offs, constraints or costs associated with these genomic changes. Strikingly, we find that the functional domains of two proteins implicated in juvenile hormone binding and transport show convergent and complementary signatures of selection as eusociality has been gained and lost in halictids. Coupled with our finding that JH is present in the insect brain, our results help to explain how novel linkages between social behaviours and endocrine signalling could convergently shape the evolution of eusociality.

Methods

Detailed methods are provided in the Supplementary Information.

Sequencing datasets

We built 10X Genomics linked-read libraries for 15 species and updated two additional assemblies with Hi-C (Supplementary Table 1). Bulk mRNA transcriptome sequencing of four tissues and miRNA sequencing were also conducted for most species. Four scRNA-seq libraries were also prepared from whole brains of *Lasioglossum zephyrus* and *L. albipes*. Detailed information on all 194 sequencing libraries generated is provided in Supplementary Table 8.

Genome assembly

Genomes were assembled with Supernova v1.0.0 (ref. [107]) and evaluated for completeness using BUSCO2 (refs. [41,108,109]), with *Apis mellifera* as the seed species and the set of 4,415 Hymenoptera genes from OrthoDB v9. Gaps were closed with ABySS 2.0 (ref. [110]), scaffolds of likely bacterial origin were filtered, and repetitive elements were masked. Hi-C scaffolding was used to error-correct, order, orient and anchor the draft genomic sequences to chromosomes[32,111]. Candidate chromosome-length genome assemblies were generated using 3D-DNA[32], followed by additional finishing using Juicebox Assembly Tools[32,112].

Coding sequence annotation

We generated gene predictions for each genome using BRAKER v2.1.0 (refs. [113,114]), with RNA-seq reads mapped to repeat-masked genomes using HISAT v.2.0.5 (ref. [115]). MAKER v3.00.0 (refs. [116,117]) was run on the repeat-masked genomes. The GFF files of aligned expressed sequence tags (ESTs) from the PASA (Program to Assemble Spliced Alignments) bioinformatic pipeline[118,119] were used as EST evidence. All high-quality protein predictions from Transdecoder[120] from all species were combined and used as protein evidence for each genome. In addition, OGSv3.2 from *Apis mellifera*, OGSv5.42 from *L. albipes* and all UniProt proteins were included as protein evidence. An Official Gene Set v2.1 was created for each genome. These gene sets are relatively complete (as measured by BUSCO2) when compared with the 4,415 genes expected to be present in all Hymenoptera species on the basis of OrthoDB v9 (ref. [121]) (Fig. 1 and Supplementary Table 1).

Orthology and gene ages

We used OrthoFinder v2.3.2 (refs. [122,123]) to identify orthologous groups of genes across the 19 species analysed. Gene names, orthologous *D. melanogaster* genes and orthologous *A. mellifera* genes were assigned on the basis of OrthoDB groups. Gene Ontology terms were assigned to orthogroups using Trinotate v3.0.1 (ref. [124]) as well as GO terms of both *D. melanogaster* and *A. mellifera* orthologues determined by OrthoDB mapping. GOATOOLS (v1.0.3)[125] was used with our custom orthogroup-to-GO mapping table for GO enrichment analysis. TFs in halictids were identified from orthology to *A. mellifera* genes identified as TFs in the Regulator database[126].

We used the Phylostratigraphy pipeline (https://github.com/AlexGa/Phylostratigraphy)[127,128] to identify the approximate evolutionary age of origin for orthogroups in our study, with proteins from 11 species as a reference set. To assign an age to an orthogroup, we required that representative genes from at least 5 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[129–131].

Coding sequence evolution

Coding sequences were aligned using the codon-aware version of Fast Statistical Aligner v1.15.9 (ref. [132]), then extensively filtered to exclude poorly aligned regions following ref. [133]. We used HyPhy RELAX v2.3.11 (ref. [54]) to identify genes experiencing relaxed selection in secondarily solitary species, requiring that at least 12 taxa be present and at least 1 closely related pair of social and solitary species from each of the 3 social clades was included. 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 also used HyPhy aBSREL[134] tests to identify signatures of selection on individual branches.

Transcriptomic analyses

For each bulk RNA-seq sample, transcript per million (TPM) values were calculated for each gene using Salmon v0.9.1 (ref. [135]) with quasi-mapping and controlling for GC bias, with quantile-normalized TPM used for subsequent analyses. We calculated the specificity of expression of each orthogroup across four tissues using these normalized expression levels[136], requiring data from at least eight species and all four tissues to calculate a specificity index for each gene. We also performed phylogenetic generalized least squares (PGLS) analyses using the R package geiger v2.0 (ref. [137]) to identify correlations

between sociality and expression level for each orthogroup, excluding species with polymorphic social behaviours.

Whole brain expressed miRNAs were identified from each species using miRDeep2 (v2.0.0.8)[138]. We detected known and novel miRNAs in each species using a reference set of miRNAs from 10 other insect species[33,34,139]. Expression was quantified and novel miRNAs were filtered on the basis of no similarity to rRNA or tRNA, minimum of five reads each on mature and star strands of the hairpin sequence, and a randfold $P < 0.05$. To identify homologous miRNAs across species, we added this filtered set to the list of known miRNAs and repeated the miRDeep2 detection step. We predicted gene targets of each miRNA using a combination of miRanda (v3.3)[140] and RNAhybrid (v2.12)[141], keeping only those miRNA-target pairs that were predicted by both programmes.

Single-cell RNA-seq libraries of *L. albipes* and *L. zephyrus* were aligned to the respective genomes with CellRanger V6 (10X Genomics). We performed normalization and integration to combine scRNA-seq data across species using the SCTransform pipeline with default settings in Seurat V4 (refs. [142,143]). Dimensionality reduction was performed with the uniform manifold approximation and projection (UMAP) technique, and cell clusters were identified using shared nearest-neighbour Louvain modularity optimization, with cell-type-specific genes identified using the MAST algorithm[144] (Supplementary Tables 9 and 10).

Associations between promoter TF motifs and sociality

For each of 223 TF motifs (identified in *Drosophila* and filtered for presence in bee genomes following ref. [4]), a 'stubb' score[44] was calculated across windows in each genome, and the highest score for each motif within 5 kb upstream and 2 kb downstream of the transcription start site was assigned to each gene within a species. We then examined correlations between rank-normalized stubb scores for a given motif and social behaviour using PGLS as implemented in geiger v2.0 (ref. [137]) after excluding all socially polymorphic taxa. We excluded motifs for a given orthogroup that did not have a normalized rank of at least 0.95 in one species and required that the number of genes with significant correlations of a motif be at least five. We then counted the numbers of significant (PGLS $P < 0.01$) correlations detected with higher stubb scores in social taxa and with higher stubb scores in solitary taxa for each motif. Those motifs for which there were differences in numbers of significant correlations between social and solitary taxa of at least 20% were considered associated with the behaviour with which they more often correlated.

CNEEs

Repeat-masked genomes of each of the 19 species included in our study were aligned using Cactus[42,145]. Fourfold degenerate sites were extracted from the alignment using halTools v2.1. phyloFit of the PHAST package[146] and phastCons[147] were used to identify CNEEs. The coordinates of the resulting CNEEs were identified in each species using halLiftover. Coding sequences were removed from the CNEEs in each species individually and sequences of less than 100 bases were discarded. CNEEs were aligned with FSA[132] and alignments were filtered using TrimAl v1.4.rev22 (ref. [148]). 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 (refs. [149,150]). Resulting branch lengths were standardized to average rates of evolution in each genome using the branch lengths estimated by RAxML[151] and fourfold degenerate sites extracted using CODEML v4.9 (refs. [149,150]). We compared the standardized branch lengths of the five pairs of closely related social and solitary species in our study. Loci for which the branch lengths were longer in all social taxa than in the most closely related solitary taxa were considered fast-evolving in social species and vice versa.

JH extraction, LC–MS and stable isotope tracing

Bee brains and other tissues were flash frozen in liquid nitrogen, then ground and centrifuged before adding extraction solvent supplemented with 0.2 μg ml^{-1} valine-D8 as an internal standard. LC–MS was carried out following previously described protocols[152] using an Xbridge BEH amide HILIC column (Waters) with a Vanquish UHPLC system (Thermo Fisher). The chemical structure of JH III, parent and fragment m/z values for labelled and unlabelled forms, chromatograms and absolute quantification of JH III and JH III-D3 are shown in Extended Data Fig. 10. JH III-D3 treatments were applied to abdominal sternites of both *Bombus impatiens* workers as well as *Augochlorella aurata* foundresses. For LC–MS of *B. impatiens*, bees were chill anaesthetized and then decapitated. For LC–MS of *A. aurata*, heads were lyophilised at −80 °C and 300 mTorr for 30 min, after which brains were dissected and stored at −80 °C until analysis. Data were analysed using El-MAVEN software (Elucidata). 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 the valine-D8 (Cambridge Isotopes Laboratories) internal standard. Absolute quantification of JH III and JH III-D3 was carried out using a standard curve of purified JH III (Sigma, J2000) and JH III-D3 (Toronto Research Chemicals, E589402). For social *A. aurata* and solitary *A. pura* comparisons, JH III quantities were standardized to ng mm^{-3} using brain volumes previously quantified for each species[153].

ApoLpp evolution

We examined sliding windows across the *apolpp* alignment for our 19 halictid species and inferred branch lengths on the species tree using RaxML v7.3.0 (ref. [151]) with the PROTGAMMAWAG model of evolution. We then calculated the branch length between each tip species and the outgroup species, *Dufourea novaeangliae*. To identify changes in evolutionary rates associated with transitions in social behaviour, we compared the 16 species with shifts in social behaviour (either gains or losses) to *N. melanderi* and *Agapostemon virescens*, two species without such shifts. We also used fpocket2 (ref. [154]), as implemented in the Phyre2 (refs. [155,156]) Investigator tool, to predict the likely active sites of ApoLpp (and Hex110) in halictids. We next explored ApoLpp evolution across insects to place patterns of evolution within Halictidae into a broader context. We collected *apolpp* sequences from 78 additional Neopteran insect species (Supplementary Table 11), aligned them with MUSCLE v3.8.31 (ref. [157]), and filtered and backtranslated them to nucleotides with trimAl v1.4.rev22 (ref. [148]). We ran aBSREL, as implemented in HyPhy v2.3.11, on the resulting nucleotide alignment, testing all 191 branches within the phylogeny to detect branches experiencing selection in each major clade examined. We also examined shifts in rates of protein evolution during the evolution of Halictidae using PAML and likelihood ratio tests between five pairs of sister taxa on both the entire protein sequence as well as just functional pockets predicted by fpocket2. Finally, we used the branch model of PAML v4.9 (refs. [149,150]) to estimate dN/dS ratios for each of the major clades examined.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw sequencing data are available at NCBI under the following accession numbers: PRJNA613468, PRJNA629833, PRJNA718331 and PRJNA512907 (Hi-C). Hi-C data and genome assemblies are also available at www.dnazoo.org. Genomes and browsers can be accessed at beenomes.princeton.edu and on NCBI: PRJNA613468. Please address inquiries or material requests to skocher@princeton.edu.

Code availability

Repository with code is on GitHub at https://github.com/kocherlab/HalictidCompGen.

References

1. Kocher, S. D. & Paxton, R. J. Comparative methods offer powerful insights into social evolution in bees. *Apidologie* **45**, 289–305 (2014).
2. Schwarz, M. P., Richards, M. H. & Danforth, B. N. Changing paradigms in insect social evolution: insights from halictine and allodapine bees. *Annu. Rev. Entomol.* **52**, 127–150 (2007).
3. Kocher, S. D. et al. The genetic basis of a social polymorphism in halictid bees. *Nat. Commun.* **9**, 4338 (2018).
4. Wcislo, W. T. & Danforth, B. N. Secondarily solitary: the evolutionary loss of social behavior. *Trends Ecol. Evol.* **12**, 468–474 (1997).
5. 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).
6. Danforth, B. N., Conway, L. & Ji, S. Phylogeny of eusocial *Lasioglossum* reveals multiple losses of eusociality within a primitively eusocial clade of bees (Hymenoptera: Halictidae). *Syst. Biol.* **52**, 23–36 (2003).
7. Batra, S. W. T. Nests and social behavior of halictine bees of India (Hymenoptera: Halictidae). *Indian J. Entomol.* **28**, 375–393 (1966).
8. Michener, C. D. *The Social Behavior of the Bees: A Comparative Study* (Belknap Press of Harvard University Press, 1974).
9. Plateaux-Quénu, C. Un nouveau type de société d'Insectes: *Halictus marginatus* Brullé (Hym., Apoidea). *Ann. Biol.* **35**, 325–455 (1959).
10. Michener, C. D. Comparative social behavior of bees. *Annu. Rev. Entomol.* **14**, 299–342 (1969).
11. Kapheim, K. M. et al. Draft genome assembly and population genetics of an agricultural pollinator, the solitary alkali bee (Halictidae: *Nomia melanderi*). *G3* **9**, 625–634 (2019).
12. Kapheim, K. M. et al. Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 1139–1143 (2015).
13. Kapheim, K. M. et al. Developmental plasticity shapes social traits and selection in a facultatively eusocial bee. *Proc. Natl Acad. Sci. USA* **117**, 13615–13625 (2020).
14. Johnson, B. R. Taxonomically restricted genes are fundamental to biology and evolution. *Front. Genet.* **9**, 407 (2018).
15. Jasper, W. C. et al. Large-scale coding sequence change underlies the evolution of postdevelopmental novelty in honey bees. *Mol. Biol. Evol.* **32**, 334–346 (2015).
16. Johnson, B. R. & Linksvayer, T. A. Deconstructing the super-organism: social physiology, groundplans, and sociogenomics. *Q. Rev. Biol.* **85**, 57–79 (2010).
17. Feldmeyer, B., Elsner, D. & Foitzik, S. Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Mol. Ecol.* **23**, 151–161 (2014).
18. Ferreira, P. G. et al. Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol.* **14**, R20 (2013).
19. Johnson, B. R. & Tsutsui, N. D. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* **12**, 164 (2011).
20. Simola, D. F. et al. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Res.* **23**, 1235–1247 (2013).
21. Warner, M. R., Qiu, L., Holmes, M. J., Mikheyev, A. S. & Linksvayer, T. A. Convergent eusocial evolution is based on a shared reproductive groundplan plus lineage-specific plastic genes. *Nat. Commun.* **10**, 2651 (2019).
22. Molodtsova, D., Harpur, B. A., Kent, C. F., Seevananthan, K. & Zayed, A. Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviors. *Front. Genet.* **5**, 431 (2014).
23. Søvik, E., Bloch, G. & Ben-Shahar, Y. Function and evolution of microRNAs in eusocial Hymenoptera. *Front. Genet.* **6**, 193 (2015).
24. Engelmann, F. & Mala, J. The interactions between juvenile hormone (JH), lipophorin, vitellogenin, and JH esterases in two cockroach species. *Insect Biochem. Mol. Biol.* **30**, 793–803 (2000).
25. de Kort, C. A. D. & Koopmanschap, A. B. Molecular characteristics of lipophorin, the juvenile hormone-binding protein in the hemolymph of the Colorado potato beetle. *Arch. Insect Biochem. Physiol.* **5**, 255–269 (1987).
26. Sevala, V. L., Bachmann, J. A. S. & Schal, C. Lipophorin: a hemolymph juvenile hormone binding protein in the german cockroach, *Blattella germanica*. *Insect Biochem. Mol. Biol.* **27**, 663–670 (1997).
27. Martins, J. R., Nunes, F. M. F., Cristino, A. S., Simões, Z. L. P. & Bitondi, M. M. G. The four hexamerin genes in the honey bee: structure, molecular evolution and function deduced from expression patterns in queens, workers and drones. *BMC Mol. Biol.* **11**, 23 (2010).
28. Ismail, S. M. & Gillott, C. Identification, characterization, and developmental profile of a high molecular weight, juvenile hormone-binding protein in the hemolymph of the migratory grasshopper, *Melanoplus sanguinipes*. *Arch. Insect Biochem. Physiol.* **29**, 415–430 (1995).
29. Daneman, R. & Barres, B. A. The blood-brain barrier–lessons from moody flies. *Cell* **123**, 9–12 (2005).
30. Stork, T. et al. Organization and function of the blood-brain barrier in *Drosophila*. *J. Neurosci.* **28**, 587–597 (2008).
31. Nene, V. et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* **316**, 1718–1723 (2007).
32. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
33. Ashby, R., Forêt, S., Searle, I. & Maleszka, R. MicroRNAs in honey bee caste determination. *Sci. Rep.* **6**, 18794 (2016).
34. Kapheim, K. M. et al. Brain microRNAs among social and solitary bees. *R. Soc. Open Sci.* **7**, 200517 (2020).
35. Collins, D. H. et al. MicroRNAs associated with caste determination and differentiation in a primitively eusocial insect. *Sci. Rep.* **7**, 45674 (2017).
36. Behura, S. K. & Whitfield, C. W. Correlated expression patterns of microRNA genes with age-dependent behavioural changes in honeybee. *Insect Mol. Biol.* **19**, 431–439 (2010).
37. Liu, F. et al. Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: comparison between nurses and foragers. *Insect Mol. Biol.* **21**, 297–303 (2012).
38. Greenberg, J. K. et al. Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. *Genes Brain Behav.* **11**, 660–670 (2012).
39. Nunes, F. M. F., Ihle, K. E., Mutti, N. S., Simões, Z. L. P. & Amdam, G. V. The gene *vitellogenin* affects microRNA regulation in honey bee (*Apis mellifera*) fat body and brain. *J. Exp. Biol.* **216**, 3724–3732 (2013).
40. Bendena, W. G., Hui, J. H. L., Chin-Sang, I. & Tobe, S. S. Neuropeptide and microRNA regulators of juvenile hormone production. *Gen. Comp. Endocrinol.* **295**, 113507 (2020).

41. 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).
42. Armstrong, J. et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
43. Chen, S., Krinsky, B. H. & Long, M. New genes as drivers of phenotypic evolution. *Nat. Rev. Genet.* **14**, 645–660 (2013).
44. Sinha, S., Liang, Y. & Siggia, E. Stubb: a program for discovery and analysis of cis-regulatory modules. *Nucleic Acids Res.* **34**, W555–W559 (2006).
45. Lowe, C. B. et al. Three periods of regulatory innovation during vertebrate evolution. *Science* **333**, 1019–1024 (2011).
46. Miura, K., Oda, M., Makita, S. & Chinzei, Y. Characterization of the *Drosophila* Methoprene-tolerant gene product: juvenile hormone binding and ligand-dependent gene regulation. *FEBS J.* **272**, 1169–1178 (2005).
47. Charles, J.-P. et al. Ligand-binding properties of a juvenile hormone receptor, Methoprene-tolerant. *Proc. Natl Acad. Sci. USA* **108**, 21128–21133 (2011).
48. Li, M., Mead, E. A. & Zhu, J. Heterodimer of two bHLH-PAS proteins mediates juvenile hormone-induced gene expression. *Proc. Natl Acad. Sci. USA* **108**, 638–643 (2011).
49. Zhang, Z., Xu, J., Sheng, Z., Sui, Y. & Palli, S. R. Steroid receptor co-activator is required for juvenile hormone signal transduction through a bHLH-PAS transcription factor, Methoprene tolerant. *J. Biol. Chem.* **286**, 8437–8447 (2011).
50. Otto, N. et al. The sulfite oxidase Shopper controls neuronal activity by regulating glutamate homeostasis in *Drosophila* ensheathing glia. *Nat. Commun.* **9**, 3514 (2018).
51. Quinn, P. M. J., Moreira, P. I., Ambrósio, A. F. & Alves, C. H. PINK1/PARKIN signalling in neurodegeneration and neuroinflammation. *Acta Neuropathol. Commun.* **8**, 189 (2020).
52. Wittwer, B. et al. Solitary bees reduce investment in communication compared with their social relatives. *Proc. Natl Acad. Sci. USA* **114**, 6569–6574 (2017).
53. Lahti, D. C. et al. Relaxed selection in the wild. *Trends Ecol. Evol.* **24**, 487–496 (2009).
54. 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).
55. Martin, R. M. & Cardoso, M. C. Chromatin condensation modulates access and binding of nuclear proteins. *FASEB J.* **24**, 1066–1072 (2010).
56. Wyatt, C. D. R. et al. Genetic toolkit for sociality predicts castes across the spectrum of social complexity in wasps. Preprint at *bioRxiv* https://doi.org/10.1101/2020.12.08.407056 (2020).
57. Wang, M., Zhao, Y. & Zhang, B. Efficient test and visualization of multi-set intersections. *Sci. Rep.* **5**, 16923 (2015).
58. Loiseau, P., Davies, T., Williams, L. S., Mishima, M. & Palacios, I. M. *Drosophila* PAT1 is required for Kinesin-1 to transport cargo and to maximize its motility. *Development* **137**, 2763–2772 (2010).
59. Hidayat, P. & Goodman, W. G. Juvenile hormone and hemolymph juvenile hormone binding protein titers and their interaction in the hemolymph of fourth stadium *Manduca sexta*. *Insect Biochem. Mol. Biol.* **24**, 709–715 (1994).
60. Hartfelder, K. & Emlen, D. J. in *Insect Endocrinology* (ed. Gilbert, L. I.) 464–522 (Academic Press, 2012).
61. Burmester, T. Evolution and function of the insect hexamerins. *Eur. J. Entomol.* **96**, 213–226 (1999).
62. Smolenaars, M. M. W., Madsen, O., Rodenburg, K. W. & Van der Horst, D. J. Molecular diversity and evolution of the large lipid transfer protein superfamily. *J. Lipid Res.* **48**, 489–502 (2007).
63. Telfer, W. H. & Kunkel, J. G. The function and evolution of insect storage hexamers. *Annu. Rev. Entomol.* **36**, 205–228 (1991).
64. Fan, Y., Schal, C., Vargo, E. L. & Bagnères, A.-G. Characterization of termite lipophorin and its involvement in hydrocarbon transport. *J. Insect Physiol.* **50**, 609–620 (2004).
65. Gu, X., Quilici, D., Juarez, P., Blomquist, G. J. & Schal, C. Biosynthesis of hydrocarbons and contact sex pheromone and their transport by lipophorin in females of the German cockroach (*Blattella germanica*). *J. Insect Physiol.* **41**, 257–267 (1995).
66. Fan, Y., Chase, J., Sevala, V. L. & Schal, C. Lipophorin-facilitated hydrocarbon uptake by oocytes in the German cockroach *Blattella germanica* (L.). *J. Exp. Biol.* **205**, 781–790 (2002).
67. Schal, C. et al. Tissue distribution and lipophorin transport of hydrocarbons and sex pheromones in the house fly, *Musca domestica*. *J. Insect Sci.* **1**, 12 (2001).
68. Zhou, X., Oi, F. M. & Scharf, M. E. Social exploitation of hexamerin: RNAi reveals a major caste-regulatory factor in termites. *Proc. Natl Acad. Sci. USA* **103**, 4499–4504 (2006).
69. Scharf, M. E., Buckspan, C. E., Grzymala, T. L. & Zhou, X. Regulation of polyphenic caste differentiation in the termite *Reticulitermes flavipes* by interaction of intrinsic and extrinsic factors. *J. Exp. Biol.* **210**, 4390–4398 (2007).
70. Hunt, J. H. et al. A diapause pathway underlies the gyne phenotype in *Polistes* wasps, revealing an evolutionary route to caste-containing insect societies. *Proc. Natl Acad. Sci. USA* **104**, 14020–14025 (2007).
71. Hawkings, C., Calkins, T. L., Pietrantonio, P. V. & Tamborindeguy, C. Caste-based differential transcriptional expression of hexamerins in response to a juvenile hormone analog in the red imported fire ant (*Solenopsis invicta*). *PLoS ONE* **14**, e0216800 (2019).
72. Hunt, J. H. et al. Differential gene expression and protein abundance evince ontogenetic bias toward castes in a primitively eusocial wasp. *PLoS ONE* **5**, e10674 (2010).
73. Zhou, X., Tarver, M. R. & Scharf, M. E. Hexamerin-based regulation of juvenile hormone-dependent gene expression underlies phenotypic plasticity in a social insect. *Development* **134**, 601–610 (2007).
74. Murrell, B. et al. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764 (2012).
75. Hartfelder, K. Insect juvenile hormone: from "status quo" to high society. *Braz. J. Med. Biol. Res.* **33**, 157–177 (2000).
76. Smith, A. R., Kapheim, K. M., Pérez-Ortega, B., Brent, C. S. & Wcislo, W. T. Juvenile hormone levels reflect social opportunities in the facultatively eusocial sweat bee *Megalopta genalis* (Hymenoptera: Halictidae). *Horm. Behav.* **63**, 1–4 (2013).
77. Tibbetts, E. A. & Izzo, A. S. Endocrine mediated phenotypic plasticity: condition-dependent effects of juvenile hormone on dominance and fertility of wasp queens. *Horm. Behav.* **56**, 527–531 (2009).
78. Frederik Nijhout, H. *Insect Hormones* (Princeton Univ. Press, 1998).
79. Bloch, G., Wheeler, D. E. & Robinson, G. E. in *Hormones, Brain and Behavior* (eds Pfaff, D. W. et al.) 195–235 (Academic Press, 2002).
80. West-Eberhard, M. J. in *Natural History and Evolution of Paper Wasps* (eds Mary Jane West-Eberhard, M. J. & Turillazzi, S.) 290–317 (Oxford Univ. Press, 1996).
81. Brankatschk, M. & Eaton, S. Lipoprotein particles cross the blood-brain barrier in *Drosophila*. *J. Neurosci.* **30**, 10441–10447 (2010).
82. Pandey, A. & Bloch, G. Juvenile hormone and ecdysteroids as major regulators of brain and behavior in bees. *Curr. Opin. Insect Sci.* **12**, 26–37 (2015).
83. Glastad, K. M. et al. Epigenetic regulator CoREST controls social behavior in ants. *Mol. Cell* **77**, 338–351.e6 (2020).
84. Rubin, B. E. R., Jones, B. M., Hunt, B. G. & Kocher, S. D. Rate variation in conserved noncoding DNA reveals regulatory pathways associated with social evolution. *Phil. Trans. R. Soc. B* **374**, 20180247 (2019).

85. Tong, C., Avilés, L., Rayor, L.S., Mikheyev, A. S. & Linksvayer, T. A. Genomic signatures of recent convergent transitions to social life in spiders. *Nat. Commun.* **13**, 6967 (2022).
86. Nijhout, H. F. & Wheeler, D. E. Juvenile hormone and the physiological basis of insect polymorphisms. *Q. Rev. Biol.* **57**, 109–133 (1982).
87. Nijhout, H. F. & Reed, M. C. A mathematical model for the regulation of juvenile hormone titers. *J. Insect Physiol.* **54**, 255–264 (2008).
88. West-Eberhard, M. J. Phenotypic plasticity and the origins of diversity. *Annu. Rev. Ecol. Syst.* **20**, 249–278 (1989).
89. Ju, L. et al. Hormonal gatekeeping via the blood brain barrier governs behavior. Preprint at *bioRxiv* https://doi.org/10.1101/2022.12.01.518733 (2022).
90. Corbitt, T. S. & Hardie, J. Juvenile hormone effects on polymorphism in the pea aphid, *Acyrthosiphon pisum*. *Entomol. Exp. Appl.* **38**, 131–135 (1985).
91. Zera, A. J. & Denno, R. F. Physiology and ecology of dispersal polymorphism in insects. *Annu. Rev. Entomol.* **42**, 207–230 (1997).
92. Bell, W. J. Factors controlling initiation of vitellogenesis in a primitively social bee, *Lasioglossum zephyrum* (Hymenoptera: Halictidae). *Insectes Soc.* **20**, 253–260 (1973).
93. Shpigler, H. Y. et al. Juvenile hormone regulates brain-reproduction tradeoff in bumble bees but not in honey bees. *Horm. Behav.* **126**, 104844 (2020).
94. Pandey, A., Motro, U. & Bloch, G. Juvenile hormone interacts with multiple factors to modulate aggression and dominance in groups of orphan bumble bee (*Bombus terrestris*) workers. *Horm. Behav.* **117**, 104602 (2020).
95. Amdam, G. V., Csondes, A., Fondrk, M. K. & Page, R. E. Jr. Complex social behaviour derived from maternal reproductive traits. *Nature* **439**, 76–78 (2006).
96. Amdam, G. V., Norberg, K., Fondrk, M. K. & Page, R. E. Jr. Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. *Proc. Natl Acad. Sci. USA* **101**, 11350–11355 (2004).
97. Page, R. E., Scheiner, R., Erber, J. & Amdam, G. V. The development and evolution of division of labor and foraging specialization in a social insect (*Apis mellifera* L.). *Curr. Top. Dev. Biol.* **74**, 253–286 (2006).
98. Turillazzi, S. & West-Eberhard, M. J. *Natural History and Evolution of Paper Wasps* (Oxford Univ. Press, 1996).
99. West-Eberhard, M. J. Flexible strategy and social evolution. In *Animal Societies. Theories and Facts* (eds Ito, Y., Brown, J. L. and Kikkawa, J.) 35–51 (Japan Scientific Societies Press, Ltd., Tokyo, 1987).
100. West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford Univ. Press, 2003).
101. Toth, A. L. & Robinson, G. E. Evo-devo and the evolution of social behavior. *Trends Genet.* **23**, 334–341 (2007).
102. Toth, A. L. et al. Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc. Biol. Sci.* **277**, 2139–2148 (2010).
103. Berens, A. J., Hunt, J. H. & Toth, A. L. Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Mol. Biol. Evol.* **32**, 690–703 (2015).
104. Rittschof, C. C. & Robinson, G. E. Behavioral genetic toolkits: toward the evolutionary origins of complex phenotypes. *Curr. Top. Dev. Biol.* **119**, 157–204 (2016).
105. Qiu, B. et al. Towards reconstructing the ancestral brain gene-network regulating caste differentiation in ants. *Nat. Ecol. Evol.* **2**, 1782–1791 (2018).
106. Feldmeyer, B. et al. Evidence for a conserved queen-worker genetic toolkit across slave-making ants and their ant hosts. *Mol. Ecol.* **31**, 4991–5004 (2022).
107. 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).
108. Waterhouse, R. M. et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).
109. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019).
110. Jackman, S. D. et al. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* **27**, 768–777 (2017).
111. Burton, J. N. et al. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
112. Dudchenko, O. et al. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. Preprint at https://www.biorxiv.org/content/10.1101/254797v1 (2018).
113. 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).
114. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. in *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) 65–95 (Springer, 2019).
115. 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).
116. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
117. Cantarel, B. L. et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
118. Haas, B. J. et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
119. Haas, B. J., Zeng, Q., Pearson, M. D., Cuomo, C. A. & Wortman, J. R. Approaches to fungal genome annotation. *Mycology* **2**, 118–141 (2011).
120. 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).
121. 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).
122. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
123. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
124. Bryant, D. M. et al. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. *Cell Rep.* **18**, 762–776 (2017).
125. Klopfenstein, D. V. et al. GOATOOLS: a Python library for gene ontology analyses. *Sci. Rep.* **8**, 10872 (2018).
126. Wang, K. & Nishida, H. REGULATOR: a database of metazoan transcription factors and maternal factors for developmental studies. *BMC Bioinformatics* **16**, 114 (2015).
127. 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).

128. 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).
129. Misof, B. et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014).
130. 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).
131. Dohrmann, M. & Wörheide, G. Dating early animal evolution using phylogenomic data. *Sci. Rep.* **7**, 3599 (2017).
132. Bradley, R. K. et al. Fast statistical alignment. *PLoS Comput. Biol.* **5**, e1000392 (2009).
133. Sackton, T. B. et al. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* **364**, 74–78 (2019).
134. 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).
135. 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).
136. Yanai, I. et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
137. 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).
138. 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).
139. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68–D73 (2014).
140. Enright, A. J. et al. MicroRNA targets in *Drosophila*. *Genome Biol.* **5**, R1 (2003).
141. Krüger, J. & Rehmsmeier, M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* **34**, W451–W454 (2006).
142. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
143. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
144. 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).
145. Paten, B. et al. Cactus: algorithms for genome multiple sequence alignment. *Genome Res.* **21**, 1512–1528 (2011).
146. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
147. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
148. 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).
149. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
150. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
151. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
152. Wang, L. et al. Peak annotation and verification engine (PAVE) for untargeted LC-MS metabolomics. *Anal. Chem.* **91**, 1838–1846 (2019).
153. 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).
154. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 168 (2009).
155. 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).
156. 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).
157. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).

Acknowledgements

We thank our many colleagues who contributed samples and field support to this dataset, including: J. Gibbs, S. Droege, R. Jeanson, J. Milam, J. Straka, M. Podolak, J. Cane and M. Hagadorn; M. Richards, C. Plateaux-Quenu, J. Gibbs and L. Packer for discussion and insights on halictid life history and behaviour; T. Sackton, R. Corbett-Detig, N. Clark, A. Siepel and X. Xue for providing discussion and advice on data analysis; W. Tong for the bee drawings and M. Sheehan for assistance with RNA library preparation; and J. Rabinowitz and lab for support and access to LC–MS equipment. This work was supported by NSF-DEB1754476 awarded to S.D.K. and B.G.H., NIH 1DP2GM137424-01 to S.D.K., USDA NIFA postdoctoral fellowship 2018-67012-28085 to B.E.R.R., DFG PA632/9 to R.J.P., a Smithsonian Global Genome Initiative award GGI-Peer-2016-100 to W.T.W. and C.J.K., a Smithsonian Institution Competitive Grants Program for Biogenomics (W.T.W., K.M.K., B.M.J.), a Smithsonian Tropical Research Institute fellowship to C.J.K., and a gift from Jennifer and Greg Johnson to W.T.W. M.F.O. was supported by Vicerrectoría de Investigación, UCR, project B7287. E.L.A. was supported by an NSF Physics Frontiers Center Award (PHY1427654), the Welch Foundation (Q-1866), a USDA Agriculture and Food Research Initiative Grant (2017-05741), and an NIH Encyclopedia of DNA Elements Mapping Center Award (UM1HG009375). Sampling permit details: S.D.K., E.S.W. and M.F.O. (R-055-2017-OT-CONAGEBIO), S.D.K. (P526P-15-04026) and R.J.P. (Belfast City Council, Parks and Leisure Dept).

Author contributions

B.M.J., B.E.R.R. and S.D.K. designed the study. B.M.J., B.E.R.R. and S.D.K. wrote the initial draft of the manuscript. All authors revised and commented on the manuscript. B.M.J., B.E.R.R., E.S.W., B.M.J., S.D.K., M.F.O., K.M.K., R.J.P., W.T.W., C.J.K. and K.S.O. collected samples. B.E.R.R., C.J.K., Z.Y.W. and N.V.D. generated genomic libraries. O.D., P.A.A., M.P., A.D.O. and D.W. performed Hi-C sequencing and genome assembly. B.E.R.R., L.R.P. and A.E.W. conducted genome annotation/alignment. B.M.J., B.E.R.R., S.D.K., O.D., K.M.K., B.G.H., J.S., F.V. and I.M.T. conducted genomic analyses. A.E.W. curated the database. W.L., E.S.W., S.R.J., K.G., S.M.D., C.J.K., Z.Y.W., M.J.M., K.S.O. and N.V.D. performed laboratory experiments. E.L.A. supervised Hi-C sequencing and genome assembly. S.D.K. provided overall project supervision.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-023-02001-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-023-02001-3.

Correspondence and requests for materials should be addressed to Sarah D. Kocher.

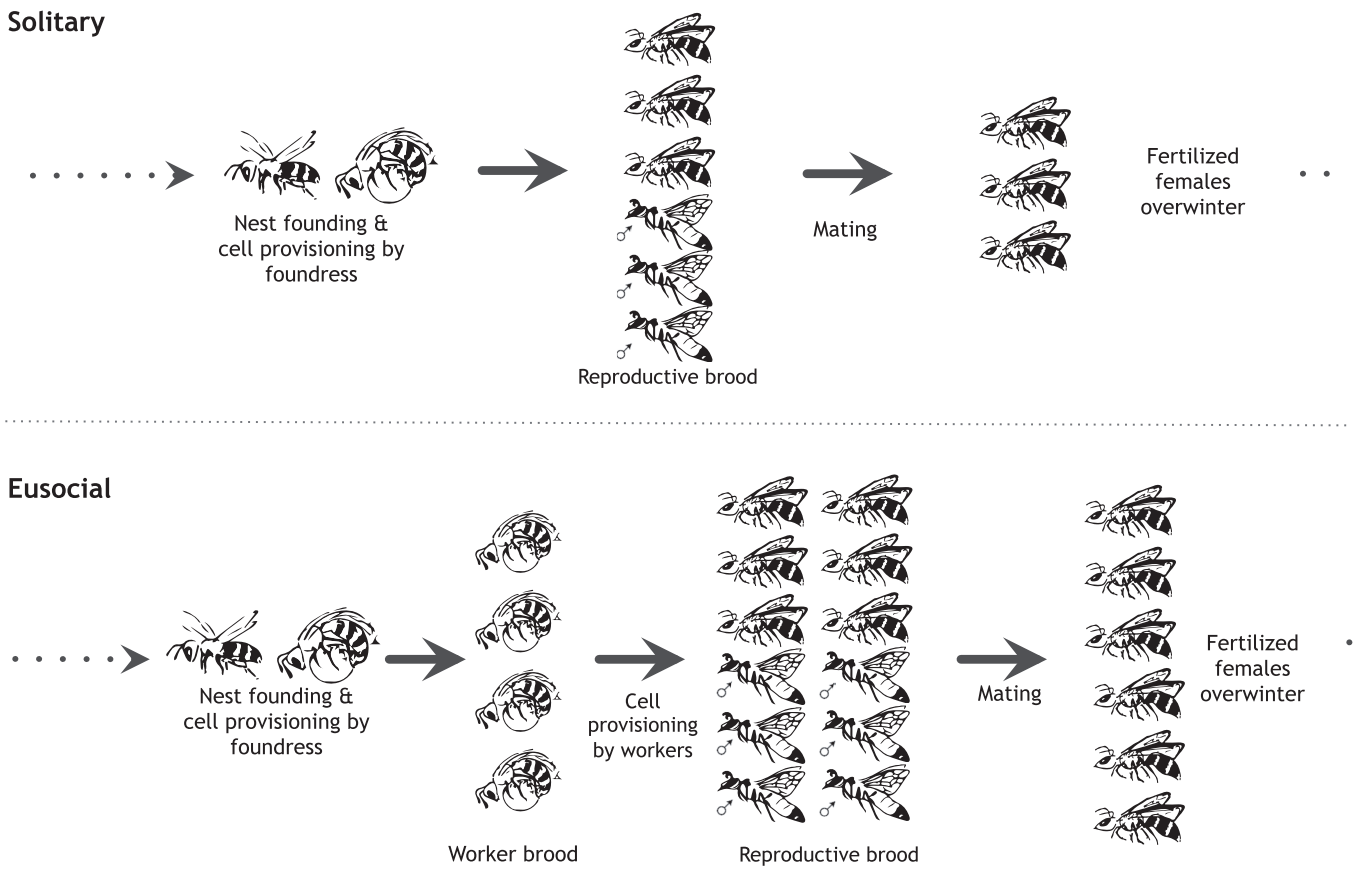
Peer review information *Nature Ecology & Evolution* thanks Guojie Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

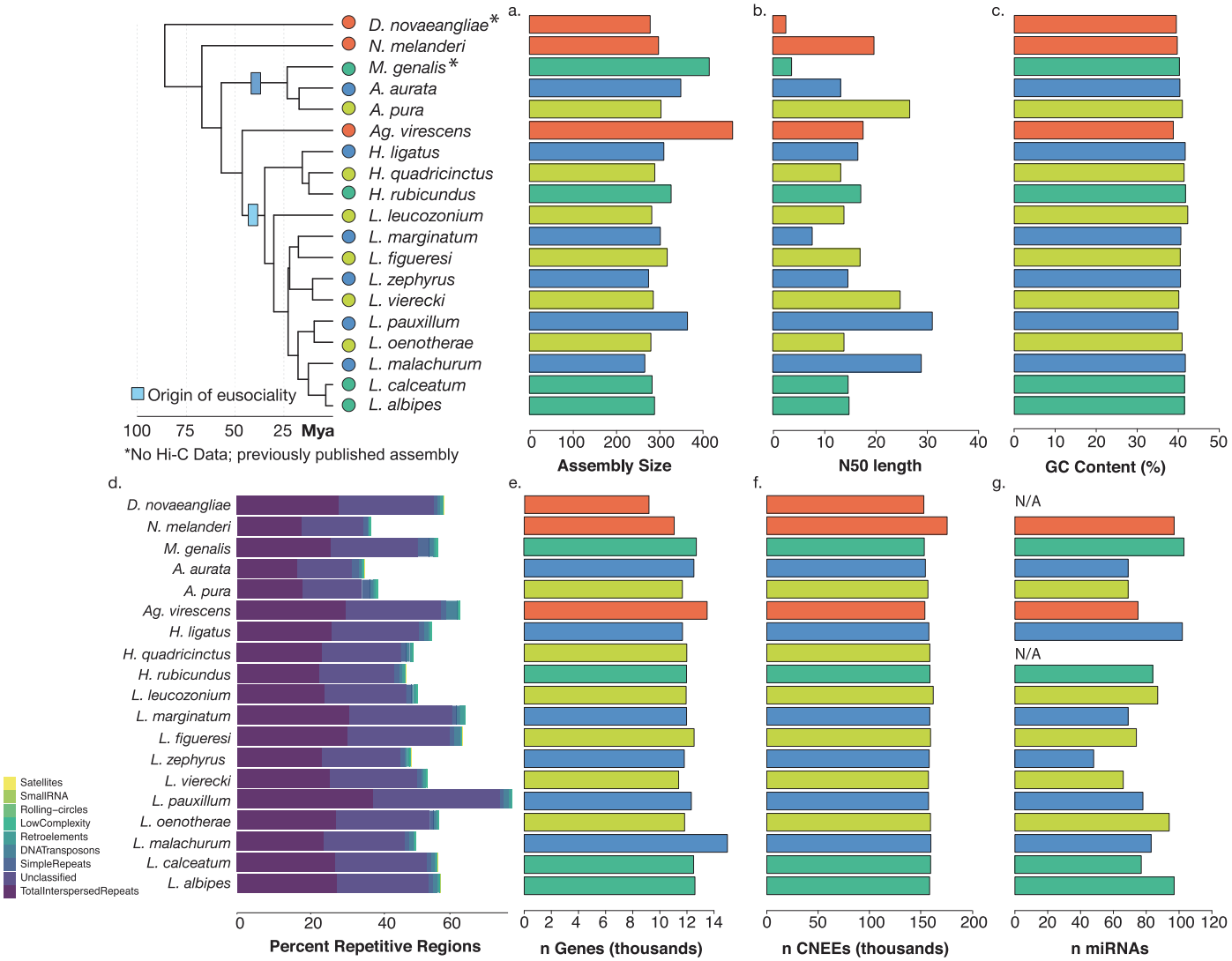
[1]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. [2]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. [3]The Center for Genome Architecture, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. [4]Center for Theoretical Biological Physics, Rice University, Houston, TX, USA. [5]Smithsonian Tropical Research Institute, Panama City, Republic of Panama. [6]Department of Biology, Utah State University, Logan, UT, USA. [7]Department of Chemistry, Princeton University, Princeton, NJ, USA. [8]Biodiversity and Tropical Ecology Research Center (CIBET) and School of Biology, University of Costa Rica, San José, Costa Rica. [9]Department of Biology, Temple University, Philadelphia, PA, USA. [10]Illumina Artificial Intelligence Laboratory, Illumina Inc, San Diego, CA, USA. [11]Department of Anthropology, University of Colorado Boulder, Boulder, CO, USA. [12]Institute of Biology, Martin-Luther University Halle-Wittenberg, Halle, Germany. [13]German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany. [14]Department of Entomology, University of Georgia, Athens, GA, USA. [15]These authors contributed equally: Beryl M. Jones, Benjamin E. R. Rubin, Olga Dudchenko. [16]These authors contributed equally: Erez Lieberman Aiden, Sarah D. Kocher. ✉e-mail: skocher@princeton.edu

https://doi.org/10.1038/s41559-023-02001-3



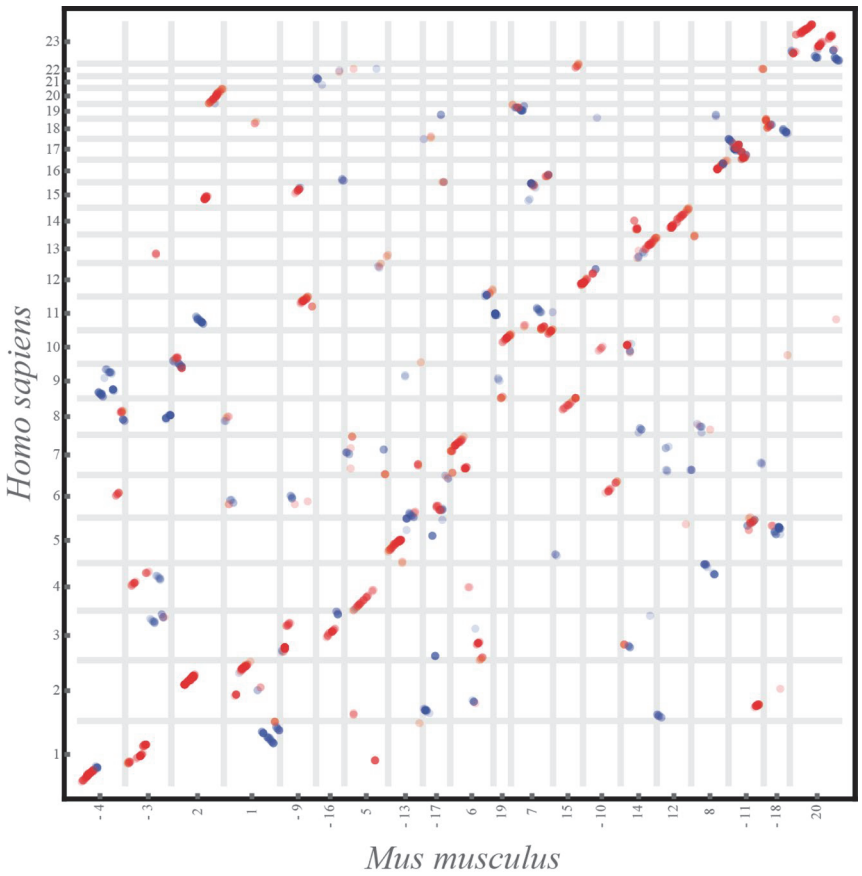
Extended Data Fig. 1 | Typical life cycle for solitary and eusocial sweat bees. In a typical solitary species, reproductive females produce a single brood that is a mix of males and females. However, some solitary species are multivoltine and can produce multiple reproductive broods in a year. In typical eusocial halictid species, females produce two broods: first workers, then reproductives. At the end of the season, females mate and overwinter as adults.

https://doi.org/10.1038/s41559-023-02001-3



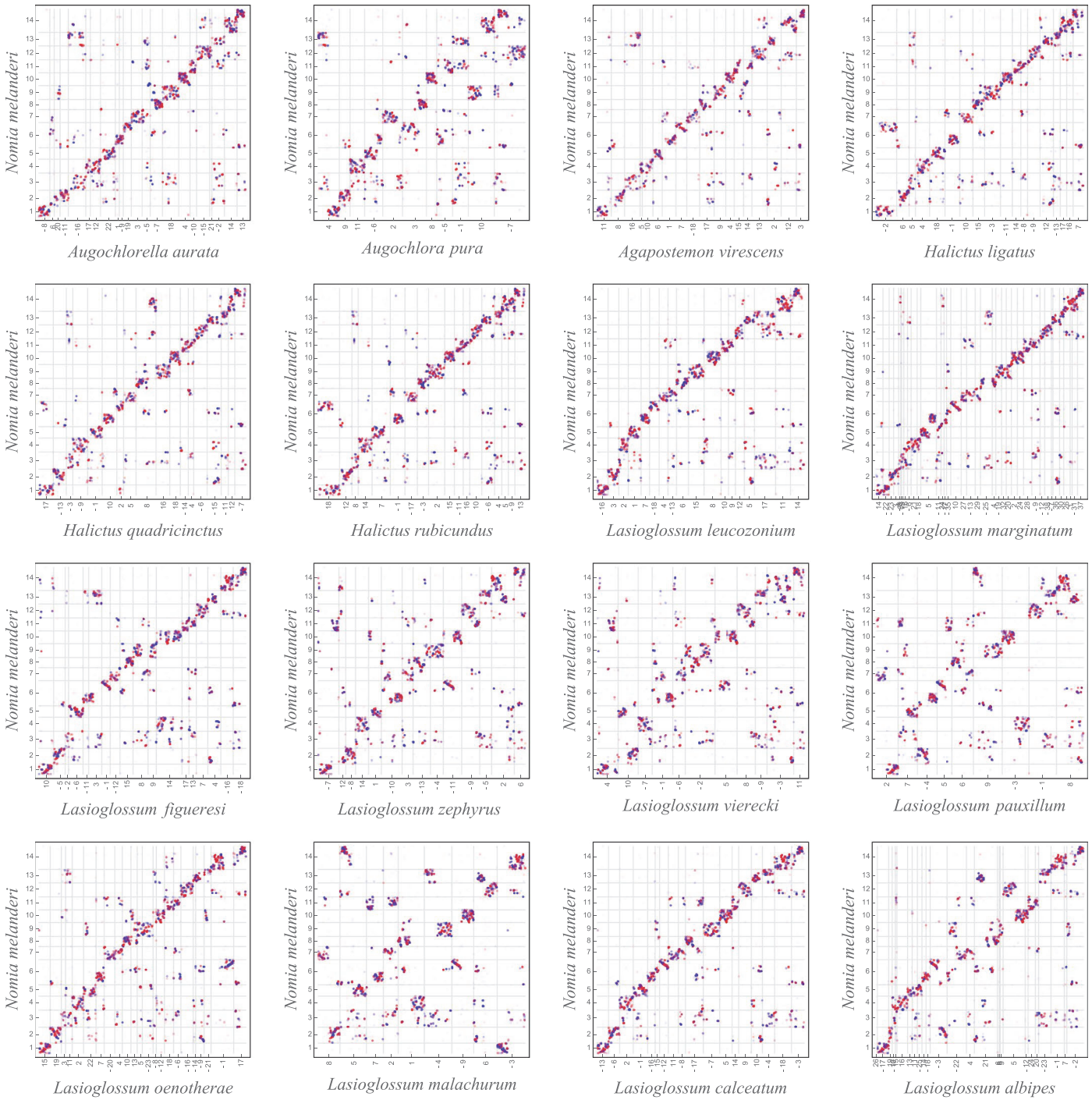
Extended Data Fig. 2 | Genome assembly statistics. Nineteen genomes are included in this comparative dataset. *15 de novo* assemblies were generated using a combination of 10x genomics and/or Hi-C, and 2 previously-published genomes (*N. melanderi*[11], *L. albipes*[3]) were improved by scaffolding with Hi-C data. *M. genalis*[13] and *D. novaeangliae*[12] were used as-is. (**a**) Genome assembly lengths ranged from ~300 to 420 Mb. (**b**) Scaffold N50s for each species following Hi-C scaffolding, and (**c**) GC content was consistent across species. (**d**) RepeatModeler was used to characterize different types of repeats in the halictid assemblies. (**e**) Numbers of genes (in thousands) for each species following individual annotation. (**f**) Conserved non-exonic elements (CNEEs) were called using phastCons on a progressive Cactus alignment. (**g**) microRNAs were also characterized using brain tissue from available species and from[34]. For some species, fresh tissue was not available (N/A).

https://doi.org/10.1038/s41559-023-02001-3



Extended Data Fig. 3 | Human-mouse chromosomal dotplot. Dotplots comparing the chromosomes of two mammalian species, human and mouse, separated by comparable evolutionary distance to the bees examined in this study (~80MY to common ancestor). The vertical and horizontal lines outline the boundaries of chromosomes in respective species, and the numbers on the axes mark the relevant chromosome name and orientation, with '-' in front of the chromosome name representing a reverse complement to the chromosome sequence as reported in the assembly. See Extended Data Fig. 4 for more details. The human-mouse one-to-one alignments file was downloaded from https://github.com/mcfrith/last-genome-alignments.

https://doi.org/10.1038/s41559-023-02001-3

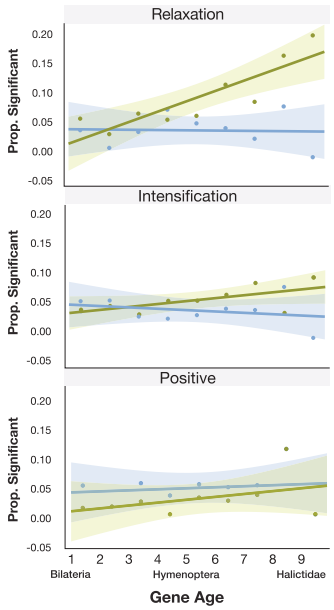


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s41559-023-02001-3

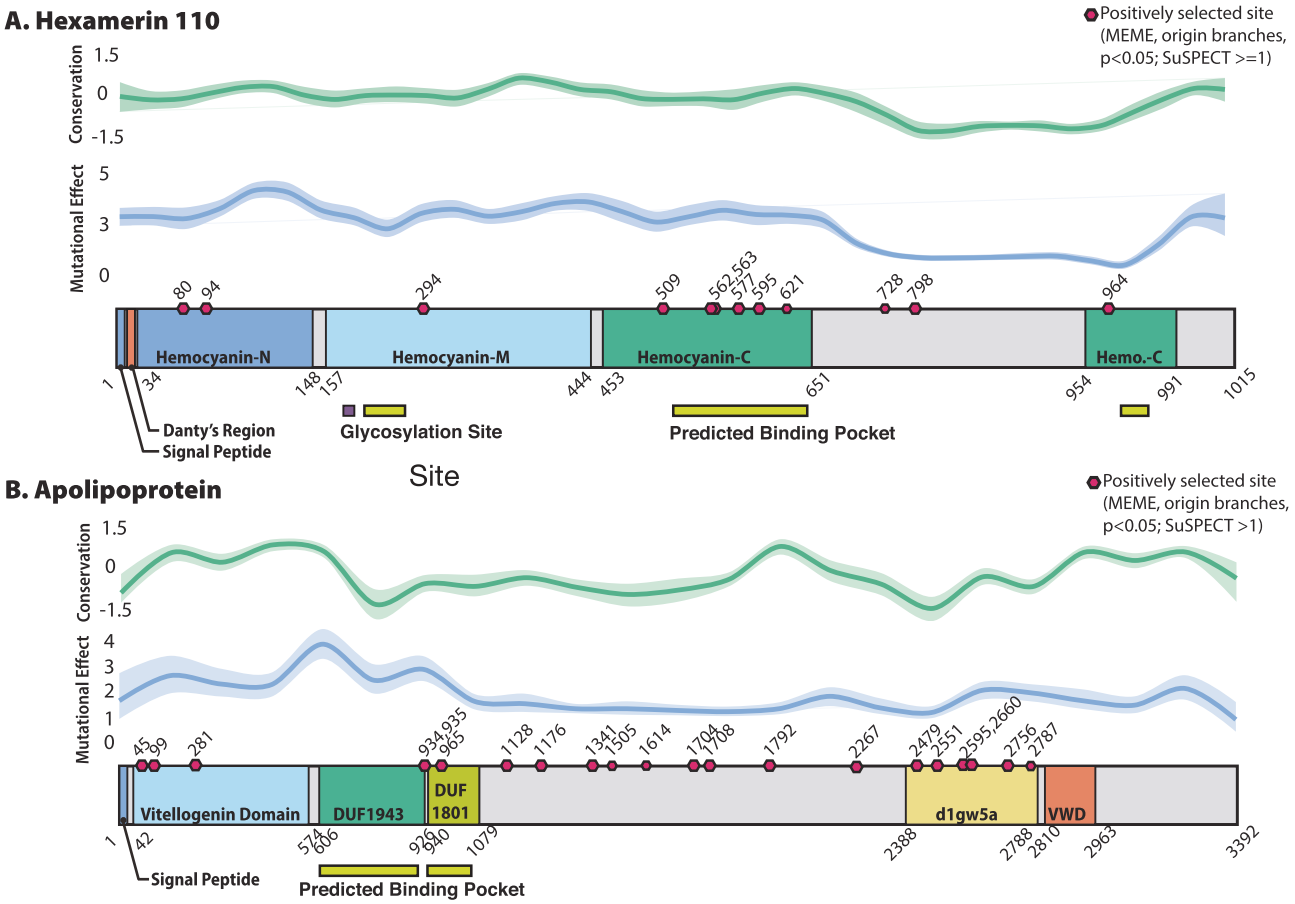
Extended Data Fig. 4 | Pairwise chromosomal dotplots for halictid species. Dotplots showing alignment of chromosome-length scaffolds from 16 bee assemblies against the *N. melanderi* (NMEL) chromosome-length genome assembly. The NMEL reference (generated as part of this study) is shown on the y-axis. The x-axis shows the chromosome-length scaffolds in the respective bee assemblies that have been ordered and oriented to best match the NMEL chromosomes in order to facilitate comparison. The vertical and horizontal lines outline the boundaries of chromosomes in respective species, and the numbers on the axes mark the relevant chromosome name and orientation, with '-' in front of the chromosome name representing a reverse complement to the chromosome sequence as reported in the assembly. Each dot represents the position of a 1000 bp syntenic stretch between the two genomes identified by Progressive Cactus alignments. The colour of the dots reflects the orientation of individual alignments with respect to NMEL (red indicates collinearity, whereas blue indicates inverted orientation). The dotplots illustrate that, with the exception of a few species, highly conserved regions belonging to the same chromosome in one species tend to lie on the same chromosome in other bee species, even though their relative position within a chromosome may change dramatically. This rearrangement pattern accounts for the characteristic appearance of the dotplots with large diffuse blocks of scrambled chromosome-to-chromosome alignments appearing along the diagonal. The pattern is visibly different from those characteristic of mammalian chromosome evolution (for example, see Extended Data Fig. 3). The few exceptions are species with multiple fissions (*L. marginatum*, *L. albipes*) and fusions (*Augochlora pura*, *L. vierecki*, *L. pauxillum*, *L. malachurum*) events. In the fission species, the syntenic regions that belonged to two different chromosomes in one bee species tend to belong to different chromosomes after the fission. The analysis of the fusion species suggests that the regions corresponding to separate chromosomes in the closely related species (and likely the ancestral species) remain separate in the new genome, possibly corresponding to the two chromosome arms. The alignments have been extracted from the hal file using the cactus hal2maf utility with the following parameters: –maxRefGap 500 –maxBlockLen 1000 –refGenome NMEL.

https://doi.org/10.1038/s41559-023-02001-3



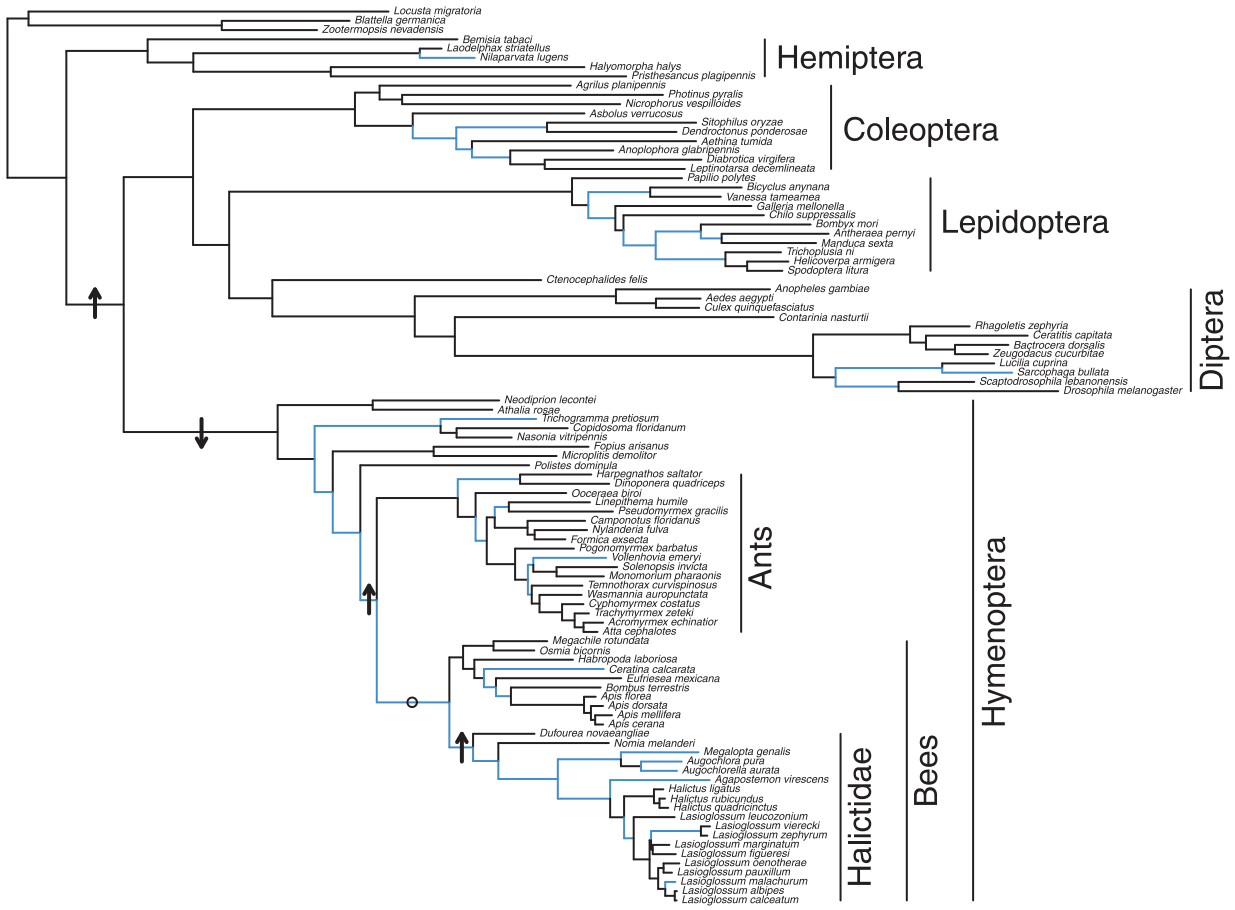
Extended Data Fig. 5 | Relationship between selection and gene age associated with eusocial origins, maintenance, and reversions to solitary life histories. Gene age ranges from the oldest Bilaterian group (Age=1) to the youngest, halictid-specific taxonomically restricted genes (Age=9). The relaxation panel demonstrates that there is a greater proportion of young genes experiencing relaxed selection when eusociality is lost (HyPhy RELAX, FDR < 0.1; Pearson's r = 0.869, p = 0.002); we find no significant association with relaxation on extant, eusocial branches (Pearson correlation, r = −0.044, p = 0.91). 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 (Eusocial: Pearson correlation, r = 0.244, p = 0.492; Solitary: Pearson correlation, r = 0.627, p = 0.071). 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: Pearson correlation, r = 0.156, p = 0.688). 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 1 loss branch in the Augochlorini (Positive, losses: Pearson correlation, r = 0.403, p = 0.283). Shading represents the 95% confidence intervals.

https://doi.org/10.1038/s41559-023-02001-3



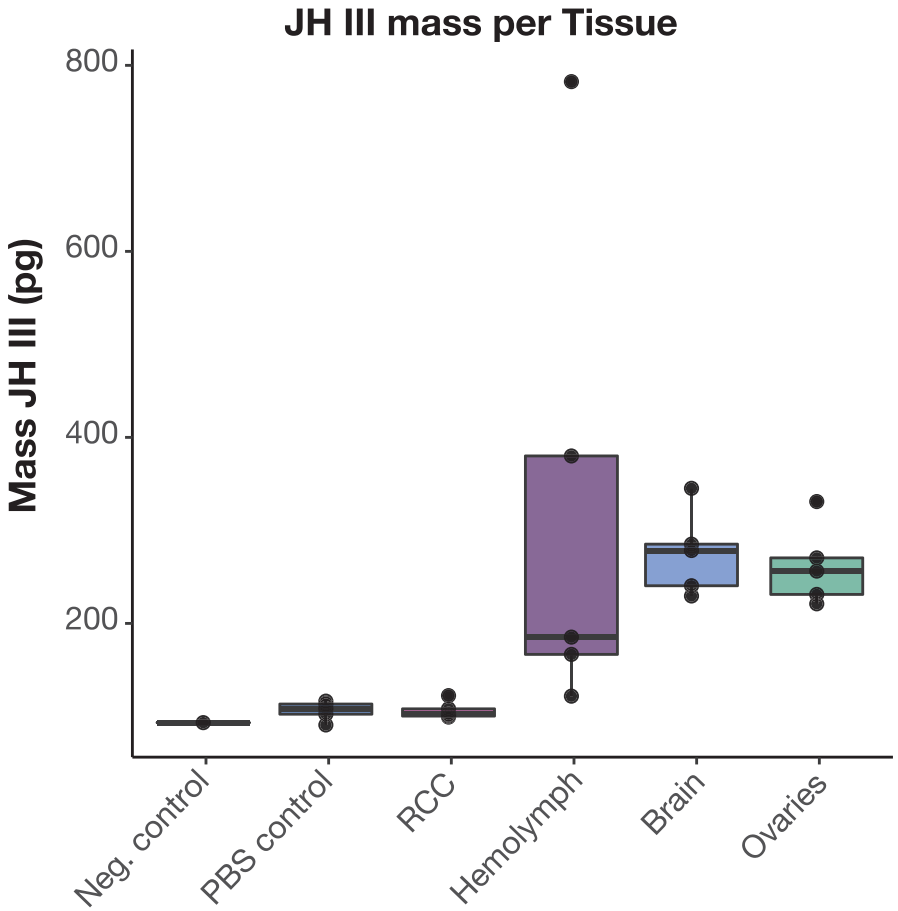
Extended Data Fig. 6 | Evidence of positive selection in domains of Hex110 and ApoLpp. Both Hex110 (**a**) and ApoLpp (**b**) show evidence for domain-specific positive selection associated with the origins of eusociality. Predicted binding pockets by PHYRE are shown with pink rectangles, glycosylation sites in orange squares. Pink hexagons denote MEME-identified sites that also had a mutational effect score > =1 for Hex110 and >1 for ApoLpp. 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: 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.

https://doi.org/10.1038/s41559-023-02001-3



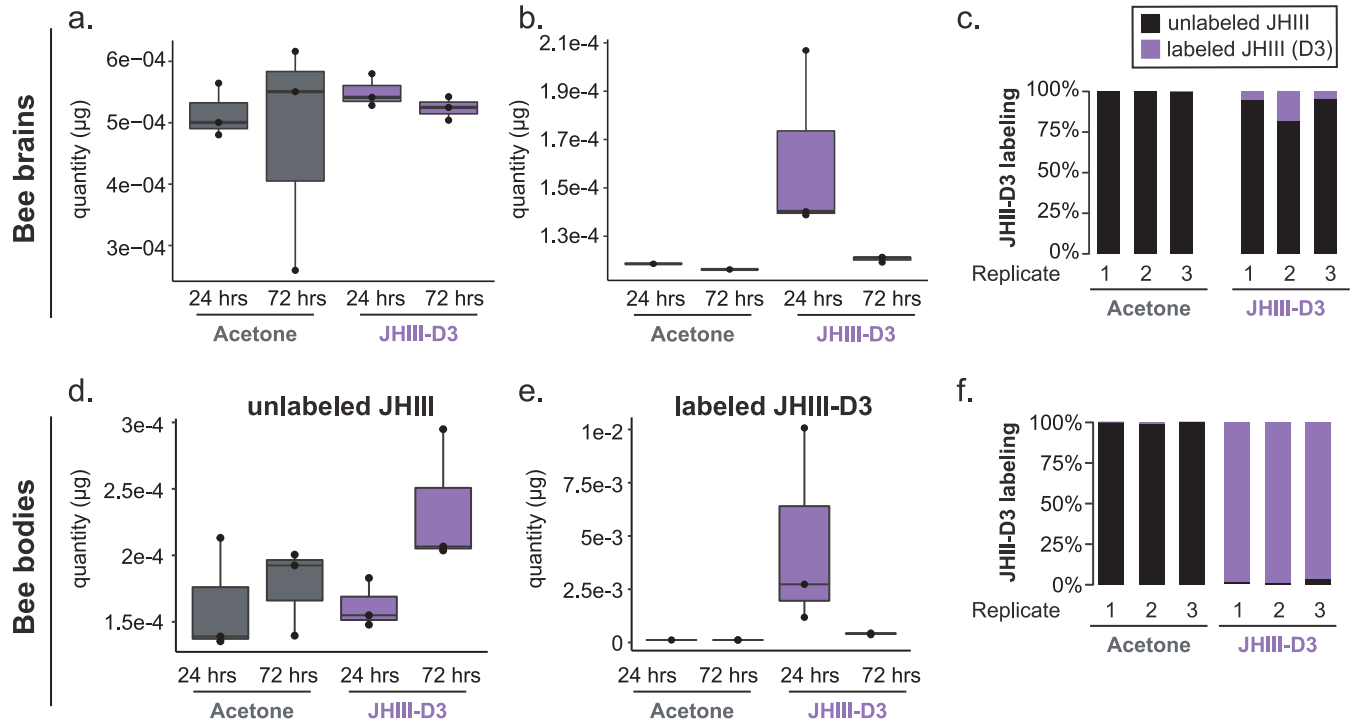
Extended Data Fig. 7 | Apolipoprotein has experienced pervasive positive selection throughout the insects. HyPhy aBSREL was used to identify branches with evidence of positive selection (highlighted in blue). Arrows indicate the direction of significant rate shifts detected on relevant branches. The only branch tested that did now show a significant rate shift is indicated by an 'o'.

https://doi.org/10.1038/s41559-023-02001-3



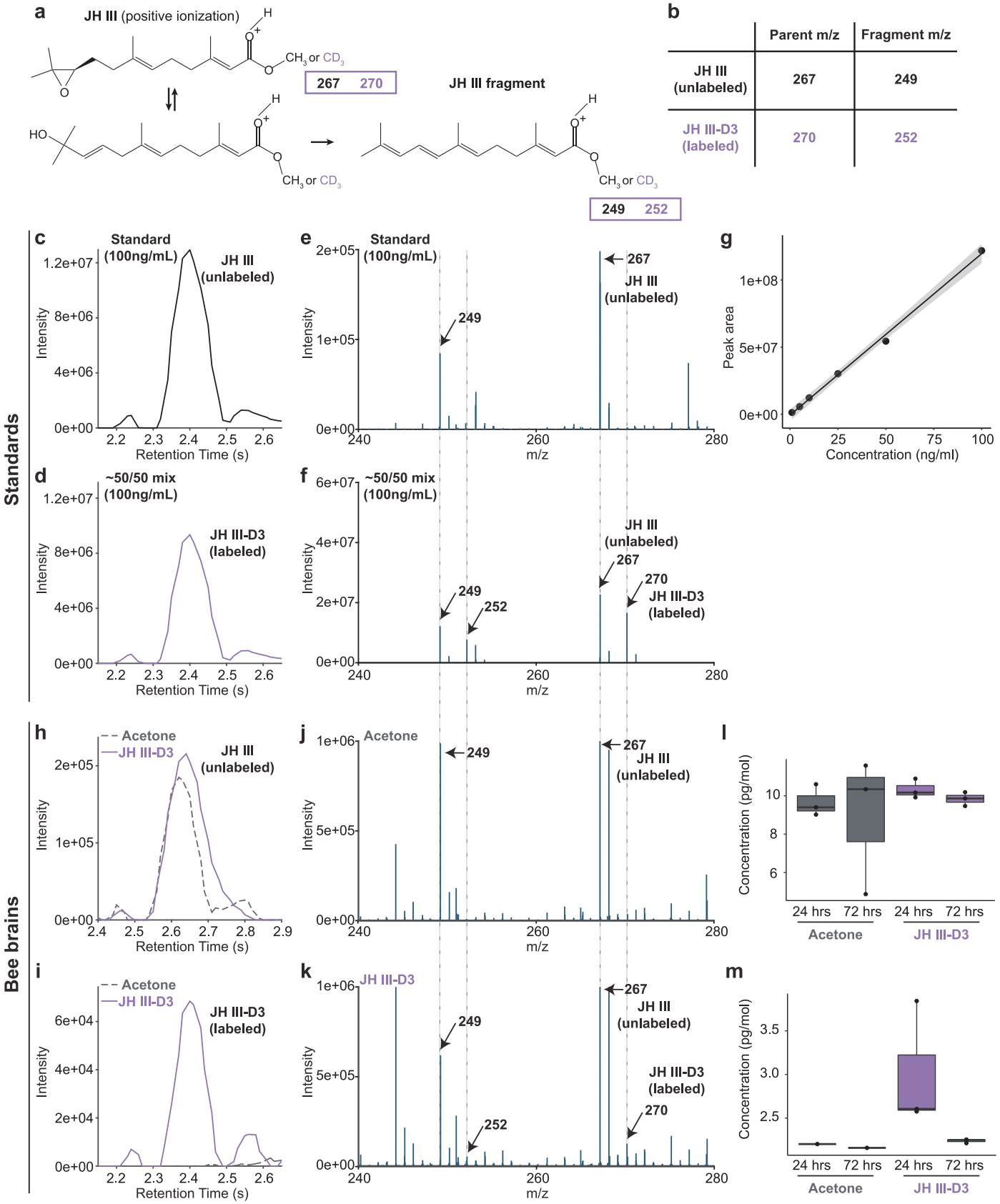
Extended Data Fig. 8 | JH III is present in multiple bumblebee tissue types, including the brain. LC-MS was used to measure JH III levels in dissected brain tissue and hemolymph from worker bumblebees (Koppert). Tissues were dissected in PBS. Negative control=fresh PBS, PBS control=3uL of buffer collected following brain dissections from each sample. RCC = retrocerebral complex. The RCC synthesizes JH and immediately releases it, it is not known to store JH[78]. Hemolymph (3uL) was collected by centrifuging thorax tissue from each sample. Ovaries and brains were dissected in PBS and the entire organ was used for JH III quantification. In all samples, we find detectable levels of JH III in the hemolymph, brain, and ovaries. N = 5 for all sample/tissue types except negative control, where n = 1. Note that because all samples were generated by extraction in a constant volume of buffer from the total input tissue, estimated amounts are not directly comparable across different tissue types. Boxes show median, 25th and 75th percentiles. Whiskers show minimum and maximum values without outliers.

https://doi.org/10.1038/s41559-023-02001-3



Extended Data Fig. 9 | Stable Isotope Tracing of Juvenile Hormone (JH). Absolute quantification of JH III in the brain (**a**) and bodies (**d**) of bumblebees treated with acetone or JH III-D3. Absolute quantification of labelled JH III in the brains (**b**) and bodies (**e**) of bees treated with acetone or JH III-D3. Labelled JH III-D3 levels are high after 24 hours, but decay significantly by 72 hours. (**c**) Labelled JH III-D3 applied to abdomens of bumblebees is detectable in brains 24 h later (acetone is control). n = 3 tissues/condition. (**f**) 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. n = 3 bumblebee workers for each experimental condition. Boxes show median, 25th and 75th percentiles. Whiskers show minimum and maximum values.

https://doi.org/10.1038/s41559-023-02001-3



Extended Data Fig. 10 | See next page for caption.

https://doi.org/10.1038/s41559-023-02001-3

Extended Data Fig. 10 | JH III and JH III-D3 quantification. **(a)** Chemical structure of JH III in positive ionization. JH III can exist in two forms of equal m/z and produces one fragment that retains the D3 label. **(b)** Parent and fragment m/z values for unlabelled and labelled JH III. **(c)** Chromatograph of unlabelled JH III. **(d)** Chromatograph of labelled JH III showing similar retention time to unlabelled JH III. **(e)** Mass spectra of unlabelled JH III showing the expected masses based on B. **(f)** Mass spectra of a 50/50 mix of labelled and unlabelled JH III showing the expected masses based on B. **(g)** Standard curve demonstrating a range of detection of JH III. **(h)** Chromatograph of unlabelled JH III in the brains of bees painted with acetone (grey dotted line) or with JH III-D3 (purple solid line). **(i)** Chromatograph of labelled JH III in the brains of bees painted with acetone (grey dotted line) or with JH III-D3 (purple solid line). **(j)** Mass spectra of JH III in bee brains painted with acetone showing the expected masses based on B. **(k)** Mass spectra of JH III in bee brains painted with JH III-D3 showing the expected masses based on B. **(l)** Absolute quantification of JH III in the brains of bees painted with acetone or JH III-D3. **(m)** Absolute quantification of JH III-D3 in the brains of bees painted with acetone or JH III-D3. n = 3 replicates for each panel. Boxes show median and 25th and 75th percentiles. Whiskers show minimum and maximum values.

nature portfolio

Corresponding author(s): Sarah Kocher

Last updated by author(s): Dec 28, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data and materials availability: NCBI: PRJNA613468, PRJNA629833, PRJNA718331, and PRJNA512907. Code: https://github.com/kocherlab/HalictidCompGen. Genomes and browsers can be accessed at beenomes.princeton.edu. Please address inquiries or material requests to Sarah Kocher (skocher@princeton.edu).
Data analysis	All details are provided at https://github.com/kocherlab/HalictidCompGen. Version numbers are also included in the supplementary materials and in the methods.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data and materials availability: NCBI: PRJNA613468, PRJNA629833, PRJNA718331, and PRJNA512907. Code: https://github.com/kocherlab/HalictidCompGen. Genomes and browsers can be accessed at beenomes.princeton.edu. Please address inquiries or material requests to Sarah Kocher, skocher@princeton.edu.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	All details are presented in the methods and supplementary materials.
Data exclusions	Any excluded data are described in the methods or supplementary materials in the manuscript.
Replication	Multiple lab members wrote, edited, and verified the analyses presented in this manuscript. When possible, experiments were replicated to verify results.
Randomization	Samples from different experimental groups were randomized across trials or replicates.
Blinding	Investigators were blinded to group identity whenever possible.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	N/A
Wild animals	N/A
Reporting on sex	A mix males and females were used to generate genomic data in this study. All details are provided in the supplement as well as in NCBI.
Field-collected samples	Specimens included in this study were field-captured and preserved for downstream analyses. All details are provided in the methods and supplement, as well as in NCBI (PRJNA613468, PRJNA629833, PRJNA718331, and PRJNA512907). Relevant sampling permits were obtained by the authors on this study: R-055-2017-OT-CONAGEBIO, P526P-15-04026, Belfast City Council, Parks & Leisure Dept.
Ethics oversight	No ethical approval is required for research on the invertebrate species included in this study.

Note that full information on the approval of the study protocol must also be provided in the manuscript.