www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

Exploring the genetics of social behaviour in *C. calcarata*

Dova B. Brenman-Suttner, Sandra M. Rehan & Amro Zayed

Studies investigating social evolution often focus on species that are obligately eusocial, where presumably all of the adaptive genetic changes associated with sociality have already been completed. To fully understand eusociality, we must study species with facultative social behaviour. The small carpenter bee *Ceratina calcarata* is an ideal model for studying the genetics and molecular biology of eusocial evolution as it can exhibit both subsocial behaviour with parental care and social behaviour facilitated by the altruistic dwarf eldest daughter. Here, we sequenced the genomes of subsocial and social *C. calcarata* to identify mutations and genes associated with social behaviour and used these data to test several hypotheses related to the evolution of eusociality. Many single nucleotide polymorphisms that had high levels of genetic differentiation (F_{st}) between social and subsocial *C. calcarata* were in or near genes or regions important for regulating gene expression. These results are consistent with the Genetic Toolkit Hypothesis of eusocial evolution. Our findings suggest that the low behavioural complexity observed in *C. calcarata* may involve modulation of existing regulatory genes and gene networks to generate phenotypes associated with social behaviour.

Keywords *Ceratina calcarata*, Population genetics, F_{st}, Regulation of gene expression, Genomics

Eusociality, defined by overlapping generations, cooperative brood care, and a reproductive division of labour, has evolved 11 different times within Hymenoptera, and four of which have occurred within bees[1,2]. While it is generally accepted that eusociality occurs within kin groups, where the cost of altruism to an individual is smaller than the benefits to the recipient scaled by their relatedness coefficient[3], the precise molecular and genomic changes involved in eusocial evolution are not well understood. Much of the research has focused on obligately eusocial species as a model for eusocial evolution[4,5]. However, studying species with less complex forms of social behaviour, such as subsocial or facultatively social species, provides an excellent opportunity to study different forms of social organization, which may represent the earliest stages of eusocial evolution[6]. Subsocial species exhibit various forms of maternal care, while incipiently social species can exhibit maternal care, remain totipotent during adulthood, build their own nests and live cooperatively with simple labour divisions to help rear a single brood[7–9]. Species capable of exhibiting both social and subsocial behaviour are known as facultative incipiently social. These species provide important insights into the genomic and ecological factors shaping social behaviour by allowing comparisons of different social strategies within a single evolutionary lineage while controlling for phylogenetic distance and their environment.

The small carpenter bee, *Ceratina calcarata*, is an example of a facultative incipiently social bee that has been studied to understand less-complex forms of social behaviour[10]. *C. calcarata* can act subsocially in the spring by providing prolonged parental care, including mass provisioning and tending to brood cells. In some nests, the reproductive female will produce a small, sterile worker daughter called the dwarf eldest daughter (DED) in autumn through differential mass provisioning[11]. The DED will forego her own reproduction to assist with guarding and foraging tasks and, because she will not survive the overwintering period, will act altruistically to her siblings[12,13]. This period is considered the social phase where behaviours associated with a reproductive division of labour are emerging among the offspring. These behaviours provide a valuable opportunity to investigate factors that may contribute to the evolution of social complexity[14–16].

Several theories address how solitary species can evolve to exhibit eusocial behaviour. The first group of hypotheses discuss how the regulation of gene expression in various contexts is important for this evolutionary transition. The Genetic Toolkit Hypothesis focuses on the regulation of conserved genes in terms of timing and location of their expression, to facilitate the emergence of novel social phenotypes[6,17,18]. Support for the genetic toolkit hypothesis comes from comparative studies of *Apis, Bombus, Dufourea, Euglossa, Eulaema, Eufriesea, Exoneura, Frieseomelitta, Habropoda, Lasioglossum, Melipona*, and *Megachile*[4,19], as well as other studies in *Apis*[20,21] and in fire ants of the genus *Solenopsis*[22,23]. Other gene regulatory hypotheses such as the Ovarian Ground-plan Hypothesis[24] suggest that sociality evolves through the uncoupling of solitary reproduction and foraging behaviour that ultimately results in caste-like differences and is supported by studies from social

Department of Biology, York University, Toronto, ON, Canada. email: zayed@yorku.ca

wasps[25] and honey bees[26–28], with evidence to the contrary found thus far in carpenter bees[12]. The Maternal Heterochrony Hypothesis, with support from social wasps[25], bumble bees[29] and *Ceratina calcarata*[12], suggests that social evolution evolves through reorganization of the timing and location of gene expression related to parental care[16]. While the previously mentioned hypotheses propose that novel behaviours emerge through the differential regulation of genes and networks present in solitary ancestors, the Novel Gene Hypothesis posits that new genes are essential for creating the phenotypic novelty observed in eusocial insects[30]. Evidence supporting this hypothesis is seen in the advanced eusocial honeybee, *Apis mellifera*, which possesses a large number of novel, taxonomically restricted genes (TRGs)[30]. These TRGs are predominantly expressed in workers and exhibit much higher rates of positive selection compared to non-TRGs[31]. These hypotheses are also not mutually exclusive and understanding which processes are occurring during the different stages of eusociality is an active area of research in the field.

Hymenoptera with incipient and facultative behaviour such as *C. calcarata* aid in our understanding of the evolution of sociality. Several studies provide evidence for the genomic, transcriptomic, and behavioural differences associated with social behaviour in *C. calcarata*. Shell and Rehan (2019) investigated differences between guarding and foraging behaviour among mothers and daughters at the transcriptomic level and found similar patterns of gene expression in the fall despite their one-year age gap[32]. Differential gene expression was also found among transcription factors associated with foraging and guarding behaviours which is suggested to be important for the broadening of caste-specific behaviours[32]. A study examining the effects of maternally cared for and orphaned offspring found differential alternatively spliced genes between the two groups[33–35]. Comparative studies of insects across the social spectrum have also shed light on the evolution of social behaviour. One study conducted by Imrit et al.[36] examined 10 insects with varying degrees of sociality which revealed relaxed negative selection with increased social complexity, likely due to changes in effective population size. These changes reflected the strength of genetic drift in more complex societies, which hinders the ability of natural selection to eliminate slightly deleterious mutations[36]. Moreover, advanced eusocial insects exhibit some of the highest recombination rates known, likely serving the purpose of increasing genotypic diversity within the worker population[37–40]. By studying *C. calcarata* and other insects, researchers have gained valuable insights into the genomic changes associated with different forms of social complexity, shedding light on the genetic mechanisms underlying social evolution.

Researchers have found a correlation between the evolution and complexity of social behaviour in bees and their enhanced ability to regulate gene expression. This is evident from a study of non-coding alignable regions (NCARs) within 11 bee species, each representing independent origins of eusociality, which revealed that certain NCARs evolved more slowly than others[41]. These slow-evolving NCARs were associated with genes linked to eusocial traits and were enriched for gene ontology terms related to the regulation of gene expression[41]. The previous study also identified the largest number of clade-specific NCARs in corbiculate bees, suggesting that the evolution of eusociality in this group was accompanied by an increase in regulatory capacity mediated by NCARs[41]. Another study examined 16 bee species within a monophyletic clade with members ranging from solitary to incipiently social (including *C. calcarata*) to investigate the genomic effect of sociality on a broad scale[34]. Shell et al. (2021) found that the evolution of complex social behaviours in bees is strongly correlated with the expansion and specialization of regulatory networks. They found that eusocial species exhibited significantly higher numbers of transcription factors (TFs), with many uniquely enriched in individuals performing specific social roles, such as foraging or guarding[34]. The expansion in regulatory networks, alongside conserved patterns of gene expression across lineages, emphasizes the role of enhanced gene regulation in the evolution of sociality. Other facultatively social species, such as *Megalopta genalis, Lasioglossum albipes*, and *Ceratina australensis* have each been studied to understand the genetic and molecular basis of social behaviour[42–45]. For example, *M. genalis* is a facultatively social sweat bee that can exhibit both social and solitary behaviour and one study found genetic differences between social and solitary nests involving high genetic differentiation at many single nucleotide polymorphisms (SNPs)[46]. A similar study in the facultatively social *C. australensis* identified specific loci that were associated with social nesting that showed signs of positive selection[45]. This suggests that alleles in genes associated with social nesting behaviour may provide fitness benefits to the bees. In *L. albipes*, a socially polymorphic halictid bee that can exhibit both solitary and eusocial behaviour, studies have uncovered genetic differences in genes involved in chemical signaling, pheromone production, and neurotransmitter regulation between social and solitary nests[42,43]. Overall, the research indicates that developmental plasticity in these species may have contributed to the evolution of diverse social behaviours, including eusociality. This was possibly achieved through genetic mutations that influence transcription factor binding, especially those related to neurogenesis and hormone-induced gene expression[4,44]. In other species, the genetic architecture of social traits often exhibit considerable variation, ranging from supergenes in ants that govern colony type formation[47], a single locus (and likely a single SNP) controlling parthenogenesis in the Cape honey bee, *Apis mellifera capensis*[48], to multiple genes that may control hygienic behaviour[49] and *Varroa*-specific defense behaviour in honey bees[50]. This diversity illustrates that social behaviour can be governed by a single gene, multiple genes, or pre-existing genetic pathways through regulation. To date, no genetic studies have explored the genetic basis of sociality in *C. calcarata*. This study is part of a broader effort to understand the complexities of sociality, similar to research on other bee species with comparable social behaviour[42,43,45,46].

Previous studies on *C. calcarata* have helped us understand the transcriptomic changes associated with social behaviours among social and subsocial groups[12,32,34,51–53]. However, it remains unclear whether these behavioural differences are linked to genetic variation, especially because social and subsocial nests coexist within the same population. While no common-garden or brood-swap studies have been carried out in this species at this time, our work represents the first effort to quantify genetic differences between solitary and social nests in *C. calcarata,* which are known to also exhibit transcriptional differences[32–35]. Here, we investigate the genomic changes associated with incipient sociality by sequencing and comparing the genomes of subsocial and social *C. calcarata*. After identifying mutations and genes associated with behavioural variation in this species, we used our dataset to test some of the mechanistic hypotheses for the evolution of social behaviour described above.

Results

Genetic differences between social and subsocial *Ceratina calcarata*

ADMIXTURE analysis revealed that the sympatric social and subsocial *Ceratina calcarata* nests form a single population (K = 1, average cross validation (CV) error = 0.852; Figure S1). While these groups do not represent distinct populations, we were still interested in identifying specific loci that exhibit difference in allele frequencies that *may* underlie the different social phenotypes. To achieve this, we examined pairwise differences at single nucleotide polymorphisms (SNPs), either individually, or averaged across genes, to understand the genetic differences between these groups. We used the fixation index (F_{st}) – a standard measure to assess genetic differences between populations[31,45,54–57]- to estimate genetic differentiation between the two behavioural groups. We were able to estimate F_{st} values between social and subsocial nests for 3,106,635 single nucleotide polymorphisms (SNPs) with an average F_{st} of 0.019 ± 0.05 SD (Fig. 1A). Outlier analysis has been widely used in many studies to identify SNPs and genes with high F_{st} values (in the top 5% or 1% of a distribution) between groups, highlighting regions of the genome with the most significant differences between populations[31,45,54,58,59]. This approach stems from the assumption that the majority of loci are evolving neutrally and that the extreme tails of the F_{st} distribution are likely enriched for loci experiencing positive selection[60]. We identified 157,934 SNPs in the top 5% and 31,356 SNPs in the top 1% of the empirical distribution of F_{st} between social and subsocial bees to be used for further analysis. We then estimated the average F_{st} value (0.021 ± 0.02) for all of *C. calcarata*'s 10,545 genes (Fig. 1B). We found 528 genes within the top 5% ($F_{st} \geq 0.052$) and 106 genes within the top 1% ($F_{st} \geq 0.091$) of the empirical distribution of F_{st} values averaged along a gene.

Evidence for hitchhiking near outlier SNPs

We investigated if regions of the genome with outlier SNPs have nearby loci that also have above average levels of genetic differentiation; this pattern would be expected if natural selection was acting to maintain specific mutations associated with social behaviour, thereby also affecting the allele frequency of nearby neutral mutations that are at linkage disequilibrium with causal loci[61]. We tested this hypothesis by comparing the average F_{st} of non-outlier SNPs in 500-base pair windows containing at least one outlier SNP (100,394 500-base pair windows) with those in windows without outlier SNPs (228,026 500-base pair windows). The average F_{st} of non-outlier SNPs in windows with at least one outlier SNP (0.011 ± 0.010) was significantly higher ($t = 44.918$, $p < 0.0001$, Cohen's D = 0.169, Figure S2) than that of SNPs in windows with no outliers (mean $F_{st} = 0.009 \pm 0.011$).

Outlier SNPs are proportionally found more often in intergenic, intragenic and exonic regions

We conducted a G test of independence to determine the association between the total distribution of SNPs within different regions of the genome compared to the outlier distribution of SNPs with high F_{st} between social and subsocial *C. calcarata*. We assessed the proportional difference of SNPs within different regions along a gene including intronic, exonic, intergenic, upstream, and downstream regions (Fig. 2; Table S1A). We compared the SNP distribution within the entire dataset to the distribution within two subsets of outliers: the top 5% and the top 1%. Overall, the results remained qualitatively consistent, indicating that expanding the threshold does not substantially alter the overall patterns observed. When we compare the distribution of SNPs across the different regions (Fig. 2 compared to Figure S3) we see very similar results, suggesting that the patterns reflect

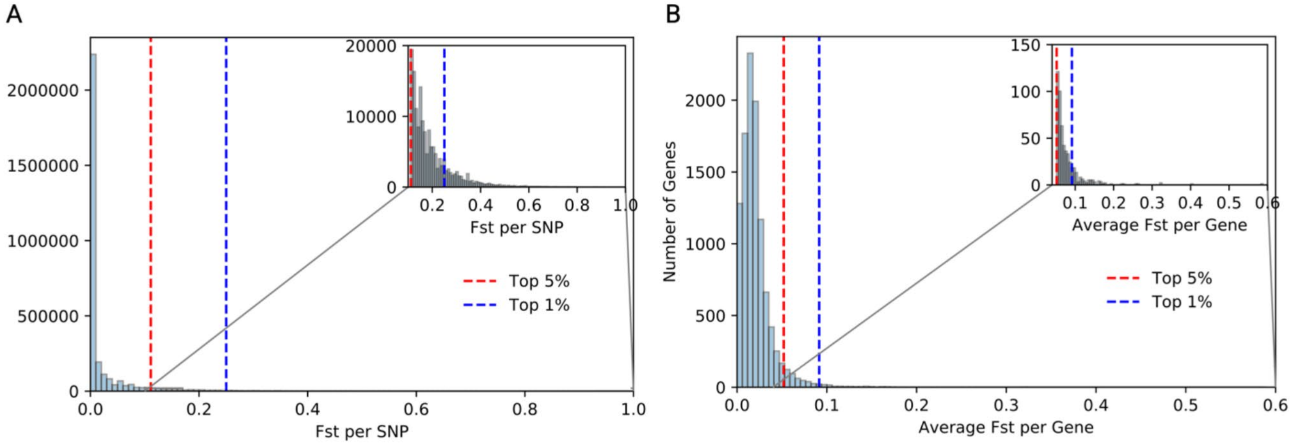

Fig. 1. The vast majority of SNPs (**A**) and genes (**B**) show very low levels of genetic differentiation (pairwise $F_{st} \sim 0$) between social and subsocial *C. calcarata*. Loci with high levels of genetic differentiation are depicted in the insets of each panel. Red dashed lines mark the top 5% F_{st} thresholds (SNPs ≥ 0.111, genes ≥ 0.052), while blue dashed lines mark the top 1% thresholds (SNPs ≥ 0.250, genes ≥ 0.091).

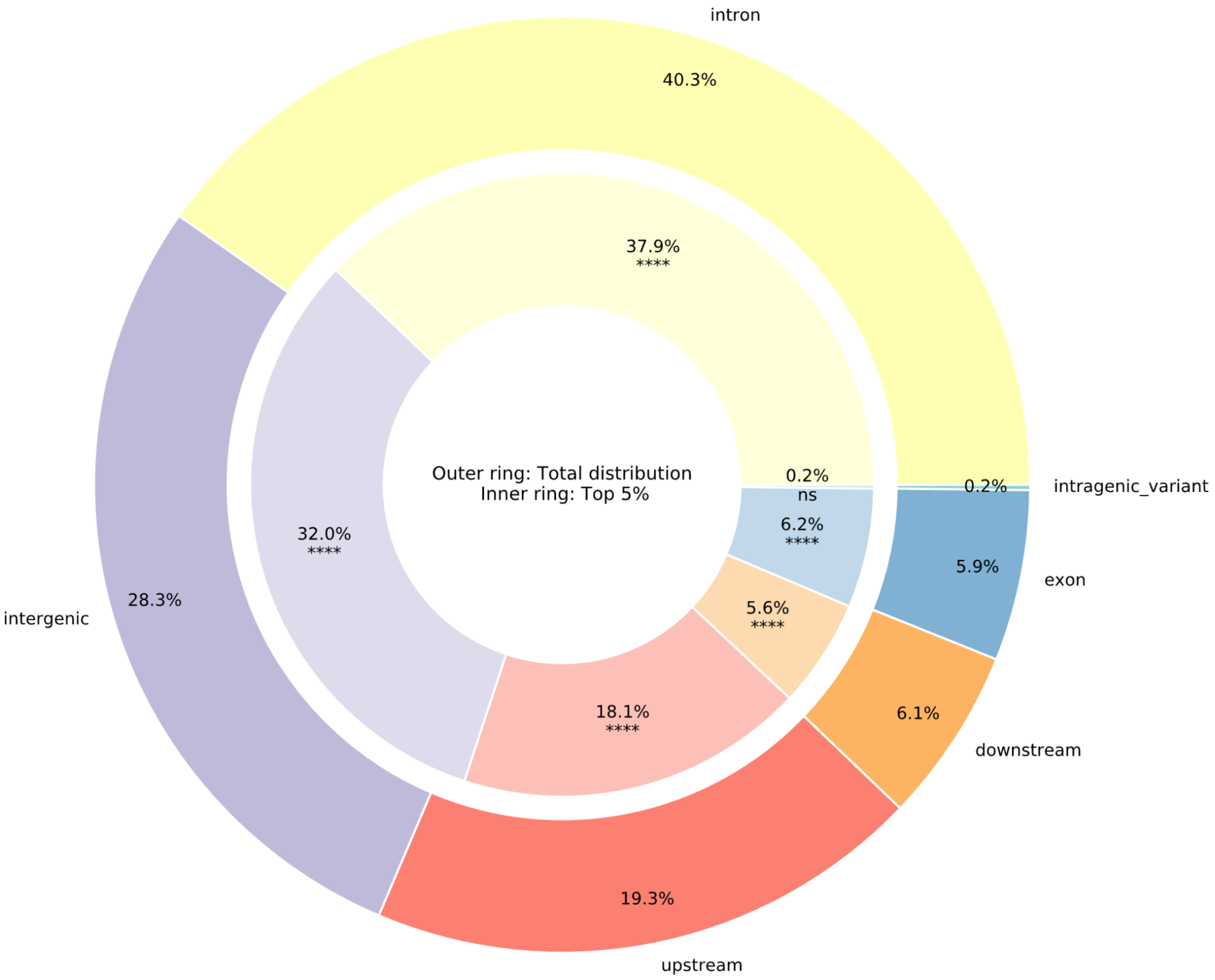

Fig. 2. Enrichment of outlier SNPs across different mutation types in *C. calcarata.* The rings show the relative distribution of SNPs among gene introns, exons, upstream and downstream regulatory sequences as we as intergenic and intragenic regions for all SNPs (outer ring) and SNPs with outlier F_{st} (inner ring). We found that intergenic, exonic and intragenic SNPs were enriched for outlier F_{st}. (ns = not significant, *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p < 0.0001$).

meaningful genomic variation rather than being driven solely by the stringency of the threshold. As a result, we continue with the top 5% from now on (Top 1% can be found in Figure S3 and Table S1B, C, D, Figure S4B). Our findings revealed significant differences in the proportion of SNPs with high F_{st} values between behaviour groups across the *C. calcarata* genome. These SNPs may therefore be causal or linked to causal SNPs that influence the expression of subsocial or social behaviour. Specifically, we observed a significantly lower proportion of intronic SNPs in outlier SNPs (G = 349.024, p = 6.914e-78) compared to the overall SNP distribution. Likewise, we identified a significantly lower proportion of SNPs in upstream regions (G = 135.167, p = 3.035e-31), as well as in downstream regions of the outliers (G = 51.976, p = 5.616e-13), compared to the overall distribution. In contrast, we observed a significantly higher proportion of SNPs within intergenic regions in the top 5% (G = 1009.145, p = 1.847e-221), as well as a higher proportion of exonic SNPs (G = 12.281, p = 4.577e-4) and intragenic SNPs (G = 19.621, p = 9.443e-6) compared to the total distribution. When we looked at the proportions of different types of exonic SNPs, we found more synonymous variants (G = 82.030, p = 1.34e-19), splice region variants with a synonymous variant (G = 7.149, p = 0.007) or a gained stop codon (G = 6.518, p = 0.011) in the top 5% than expected whereas fewer missense variants were found (G = 77.593, p = 1.265e-18) than expected (Figure S4A, Table S1E).

Predicted function of genes with high F_{st} values in *Ceratina calcarata*

Most of our genes had an ortholog in *Drosophila melanogaster* (6,773, 64.23%; Table S2). We identified 306 orthologs within the top 5% of the F_{st} distribution. We used these genes to perform gene ontology enrichment analysis and identified significant enrichment of annotations related to metabolic processes (Figure S5, Table S5) which was also found in the highest enrichment cluster (enrichment score = 2.245; Table S6), followed by

DNA repair and response to stress (enrichment score = 1.665). A word cloud of the most common annotations was also generated to visualize common processes and annotation terms found among enriched gene ontology clusters (Figure S6).

Functional annotation analysis of high F_{st} genes in *C. calcarata*

We identified distinct patterns in genes with high levels of genetic differentiation between social and subsocial *Ceratina calcarata*, based on gene functions assigned via HymenopteraMine (Table S2). We found several genes associated with initiating or regulating transcription including *Transcription activator MSS11* (F_{st} = 0.053), *Transcription and mRNA export factor ENY2* (F_{st} = 0.086), *Transcription factor 25* (F_{st} = 0.092), *Transcription factor BTF3 homolog 4* (F_{st} = 0.078), *Transcription initiation factor TFIID subunit 4*-like (F_{st} = 0.057), *Transcription initiation protein SPT3 homolog* (F_{st} = 0.067), *Transcriptional adapter 1*-like (F_{st} = 0.060), *Transcriptional regulator ATRX homolog* (F_{st} = 0.095). We also found several genes associated with chromatin organization and remodelling including *histone chaperone asf1* (F_{st} = 0.103), *histone deacetylase complex subunit SAP18* (F_{st} = 0.076), *histone H2A* (F_{st} = 0.228) and *H2A*-like (F_{st} = 0.082), *histone H3* (F_{st} = 0.137), several SNPs within *histone H4* (F_{st} = 0.176, F_{st} = 0.160), a *histone H4 transcription factor* (F_{st} = 0.077), several histone related proteins such as *histone RNA hairpin-binding protein* (F_{st} = 0.093), *histone-arginine methyltransferase CARMER* (F_{st} = 0.062), and two SNPs within *histone-lysine N-methyltransferase SETMAR*-like (F_{st} = 0.114, F_{st} = 0.071). We also identified two antioxidant enzymes p450 4C1-like (two genes annotated; F_{st} = 0.105, F_{st} = 0.111) and p450 6K1-like (F_{st} = 0.054) as well as Cyp9E2 (F_{st} = 0.053) and Cyp6A2 (F_{st} = 0.125). Among the *Drosophila melanogaster* orthologs with the highest genetic differentiation between behaviour groups, several noteworthy genes are involved in the regulation of transcription and chromatin, including genes such as *calypso* (F_{st} = 0.092), *Nucleolar protein 66* (F_{st} = 0.054) and *No child left behind* (F_{st} = 0.051) as well as genes involved in splicing like *Small ribonucleoprotein particle protein SmD2* (F_{st} = 0.101; Table S3).

Ceratina calcarata genes tend to be found in older taxonomic groups

We assessed the evolutionary history of all genes in the *Ceratina calcarata* genome to determine whether genes with very high levels of F_{st} tended to be enriched in specific taxonomic groups. Since taxonomically restricted genes have been shown to play a role in the elaboration of sociality[62], we aimed to investigate whether genes that were highly differentiated between behavioural groups were also restricted to particular taxonomic levels, such as Aculeata and Apoidea. Our orthology analysis revealed that the majority of *C. calcarata* genes with any F_{st} value between groups are of ancient evolutionary origin (Fig. 3, Supplemental Table S7A). Additionally, when we examined enrichment for specific taxonomic classifications (e.g., Eukaryota, Metazoa) among genes with outlier F_{st} values (top 5%) compared to all genes in the *C. calcarata* genome, we did not find significant enrichment (Figure S7, Table S7B, C). When we grouped genes based on their evolutionary age (old vs. new) we also did not find any significant enrichment (P = 0.930, Table S7D). These findings show that although we identified some genes with very high levels of genetic differentiation associated with vastly different behaviours in this species, both ancient and recently evolved genes play a similar role in behaviour.

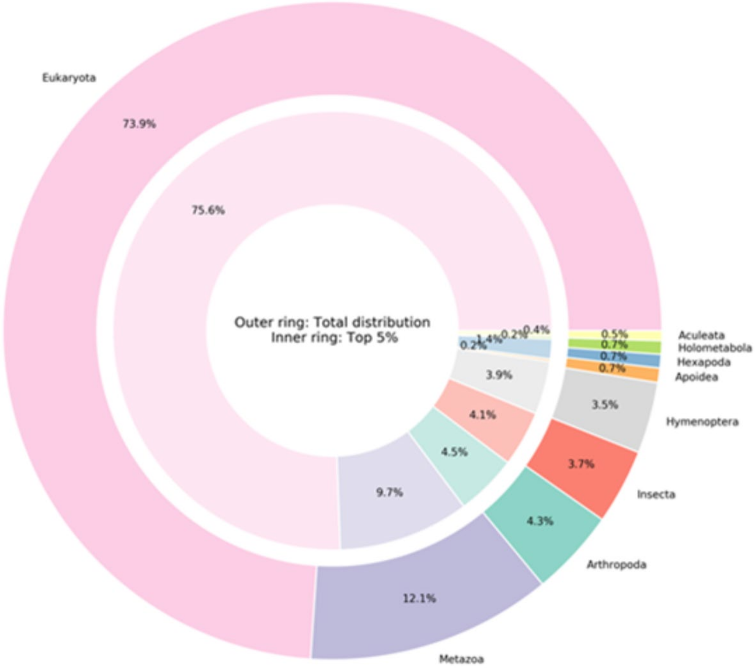

Fig. 3. Genes associated with social behaviour among social and subsocial groups in *C. calcarata* are predominantly evolutionarily ancient and the outlier list of genes are not enriched for any group (ns = not significant, *$p \le 0.05$, **$p \le 0.01$, ***$p \le 0.001$, ****$p < 0.0001$).

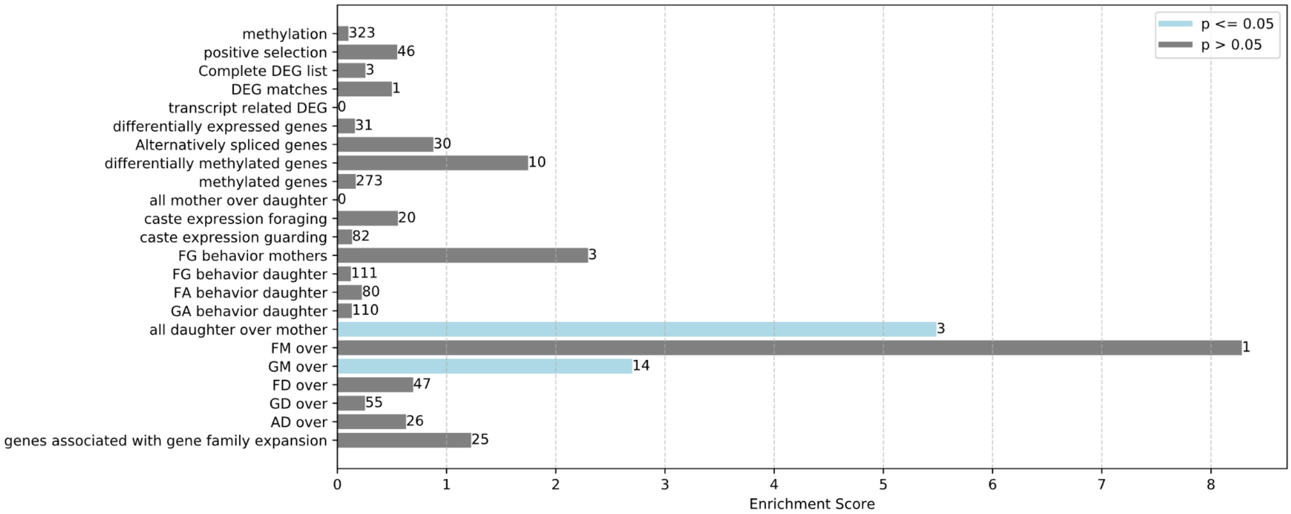

Fig. 4. Enrichment for genes with outlier levels of F_{st} (top 5%) associated with social behaviours among social and subsocial groups and previous genomic and transcriptomic studies in *C. calcarata*. The numbers at the end of each bar represent the number of genes with outlier levels of F_{st} associated with social or subsocial behaviour that were also found in the specific transcriptomic studies. We found significant overlap between genes with outlier levels of F_{st} and differentially expressed genes among several different categories of daughters that were either significantly upregulated over all others and among guarding mothers that are upregulated over all others. See Table S7B for full list of acronyms.

Connecting genetic differentiation and candidate genes for sociality

Several studies have examined various aspects of gene expression and genomic features in *C. calcarata*. Research thus far has focused on transcriptomic differences among behavioural groups, such as differential expression in foraging and guarding mothers, daughters, and dwarf eldest daughters[32]. Other studies have identified differentially expressed genes following aggression and those involved in gene family expansion or contraction in a comparison with 16 other bee genomes[34]. Additionally, genomic features of *C. calcarata* have been explored, including methylation patterns, alternatively spliced genes and differentially methylated genes[33,63]. In our study, we identified genes with high levels of F_{st} that were also differentially expressed or present in previous research to determine if any studies had an overrepresentation of genes with high genetic differentiation between social and subsocial *C. calcarata*. This approach helps us understand the characteristics of genes with high genetic differentiation based on their expression across various contexts (Fig. 4, Table S8A). We found significantly more overlap than expected among differentially expressed genes (DEGs) in different behaviours of *C. calcarata* including all daughter DEGs upregulated over mother genes ($p = 0.017$; Table S9) and DEGs among guarding mothers upregulated over all others ($p = 5.33\text{e-}4$), highlighting the connection between genetic differentiation among our behaviour groups and their gene expression outcomes. Some notable genes and patterns we identified that had both differential gene expression and high genetic differentiation in our study included Odorant receptor 13a-like with several SNPs with outlier F_{st} values between behaviour groups as well as differential expression between mothers and daughters ($F_{st} = 0.077$, 0.062, 0.054, 0.405). This is the first evidence linking genetic differences associated with social and subsocial behaviour in *C. calcarata* to gene expression differences.

Discussion

In this study, we investigated the genetic underpinnings of social behaviour in *Ceratina calcarata*. *C. calcarata* exhibits a stable form of a behavioural polymorphism with either subsocial or social nests, making it an excellent model organism for studying the genetics and evolution of social behaviour[12,64]. It has been previously argued that species with simpler forms of social behaviour can offer insights into the evolutionary steps toward more complex social organization, a hypothesis known as the 'Social Ladder'[6]. However, it is also possible that transitions to complex sociality do not involve step-wise evolution of social organization[65]. Regardless, our study of *C. calcarata* provides an opportunity to study the evolution of a very simple form of social organization and test if some of the mechanistic hypotheses for the evolution of eusociality are at play in this lineage. We applied a commonly used paradigm for identifying mutations and genes that *may be* associated with eusociality by highlighting genes with extreme levels of genetic differentiation between social and subsocial *C. calcarata* populations[31,45,54–59]. The outlier approach we used posits that loci that underly phenotypic differences between the subsocial and social populations of *C. calcarata* have outlier population genetic parameters relative to the vast majority of neutral markers in the genome that do not affect behavioural phenotypes[60]. While this analysis on its own does not prove that any specific marker causally influences behaviour, it does allow us to identify putative candidates that *may* be influencing behaviour. Overall, our findings indicate some association between genetic differences and social behaviour; while there may be a genetic component, it may not be the sole determinant. Social behaviour in this species may be influenced by a complex interplay of genetic, environmental, and possibly epigenetic factors[2,66–68]. Thus, while our study provides some insight into the genetic basis of social behaviour in this species, it also highlights the need for further research to fully elucidate these relationships.

The small carpenter bee is an excellent model for studying the incipient stages of social evolution without any confounds associated with comparing groups or species with different social behaviours that have been separated by millions of years of independent evolution or from species that inhabit different environments. Our genomic analysis confirms that social and subsocial *C. calcarata* belong to a single, randomly mating population, supporting the idea that social behaviour in this species is not strictly genetically determined. The average genetic differentiation between subsocial and social *C. calcarata* ($F_{st} = 0.019 \pm 0.05$, Fig. 1) is comparable to F_{st} observed between solitary and social nests of the facultatively eusocial *Lasioglossum albipes* and *Megalopta genalis* (average F_{st} of 0.06 and 0.03, respectively)[43,46]. This amount of genetic variation suggests that there has been recent gene flow between the behavioural groups of *C. calcarata.* These observations underscore the complexity of genetic influences on social behaviour, suggesting that while broad genetic differentiation is minimal, specific genetic factors may still have significant effects. We also found that non-outlier SNPs in 500-bp windows containing at least one outlier SNP had significantly higher average F_{st} than those found in windows with no outliers – a pattern that is consistent with the hypothesis that outlier SNPs may be experiencing directional selection.

Our study revealed a higher frequency of single nucleotide polymorphisms (SNPs) with elevated F_{st} values between social and subsocial behaviours within intergenic regions, intragenic regions, and exons. SNPs within exons have obvious fitness consequences, and exonic regions tend to be enriched for outlier levels of F_{st} in other bees, including the advanced eusocial *Apis mellifera*[31,69,70]. In contrast, SNPs with outlier levels of F_{st} between social and subsocial *C. calcarata* were less frequently found in introns and regions directly upstream and downstream (within 5 kb on either side) of genes (Fig. 2). The observed enrichment of SNPs in intergenic regions may indicate that the behavioural differences observed in *C. calcarata* are due to variations in regulatory sequences within these regions. For example, regulatory elements like enhancers can act at considerable distances from the genes they control[71], while silencers and insulators are typically located in intergenic regions[72].

The Genetic Toolkit Hypothesis proposes that eusociality evolved through the repurposing of highly conserved genes and molecular pathways for various social functions[19]. Under this hypothesis, a key mechanism for the evolution of novel behaviours is the co-option of existing genes and pathways, coupled with changes in their expression dynamics—such as timing, spatial localization, and duration—resulting in distinct behaviours[17]. The regulation of these patterns of expression is important for the transition to social behaviour as seen in genomic studies of ants[73–75], the bumblebee *Bombus terrestris*[19,29], the halictid bee *Lasioglossum albipes*[43], in transcriptomic studies of *C. calcarata*[32] and in a comparative study of *Apis mellifera*, *Solenopsis invicta*, and *Polistes metricus*[23]. In our study, we discovered that genes that contain mutations with high genetic differences in allele frequencies between social and subsocial groups in *C. calcarata* showed enrichment in annotation terms associated with the regulation of gene expression, particularly in pathways related to RNA metabolism, which is closely tied to gene regulation[76]. For instance, genes displaying significant genetic differentiation among behaviour groups were found to be associated with annotations related to non-coding RNA metabolism (see Figure S5), highlighting the crucial role of non-coding RNA as regulators of gene expression, which is pivotal for social evolution[75]. We also found many genes with outlier levels of F_{st} associated with behaviour related to activating, initiating or regulating transcription (Table S1). Histone modifications in regulatory regions have also been shown to be important for establishing and maintaining caste-differences in honey bees[77] and caste-specific foraging behaviour in the carpenter ant *Camponotus floridanus*[78]. We identified several genes associated with behaviour in *C. calcarata* that had high genetic differentiation in histones or histone-related genes. Several notable *Drosophila melanogaster* orthologs with the highest values of F_{st} associated with behaviour include genes related to the regulation of transcription and chromatin. Our work suggests that some genetic variants associated with social behaviour in *C. calcarata* may be involved in gene regulation. However, due to the lack of strong genetic differentiation, these findings may reflect regulatory variation that facilitates behavioural plasticity rather than being due to a genetic determinant of sociality, which still aligns with the Genetic Toolkit hypothesis[79].

The Novel Gene Hypothesis proposes that new genes that are uniquely found within a lineage-specific group are important for the elaboration of eusociality[30]. Comparative genomic studies have shown that the honey bee genome contains around 700 taxonomically restricted genes, which may have played a crucial role in the elaboration of eusociality[30]. If our data supported this theory, we would have seen enrichment for taxonomically restricted genes within newer taxa such as Apoidea and Aculeata. However, we found that most of our genes with high genetic differentiation between social and subsocial behaviour in *C. calcarata* had both an ortholog in *D. melanogaster* and were not enriched for any specific taxonomic category. Therefore, our data do not support the Novel Gene Hypothesis. Instead, this analysis is consistent with the Genetic Toolkit hypothesis, as the lack of enrichment for taxonomically restricted genes emphasizes the importance of ancient genes, which exhibit the highest genetic differentiation between social and subsocial groups in this species. Taken together along with the gene ontology enrichment analysis, we provide strong evidence for the Genetic Toolkit Hypothesis.

The Maternal Heterochrony Hypothesis proposes that the evolution of eusocial behaviour in bees is driven by changes in the timing and expression of maternal care behaviours[16]. According to this hypothesis, ancestrally solitary bees exhibited maternal care behaviours towards their offspring and then throughout the transition to social behaviour, these behaviours became expressed earlier in development and were geared towards siblings rather than offspring[16]. This shift in timing of maternal care is thought to be a key step in the transition from solitary to eusocial behaviour. Transcriptomic analysis of *C. calcarata* have revealed similar patterns of gene expression between sibling care and maternal care[12]. Our study focused on identifying genes with high genetic differentiation between *C. calcarata* that displayed either maternal care alone (subsocial) or maternal and sibling care (social). Identifying the function of these highly differentiated genes was done by overlapping our genes with genes that were found to be differentially expressed at the transcription level in previous studies. We found significant overlap with differentially expressed genes where all daughter genes were upregulated over mothers, which is expected given that we sequenced the daughters of social and subsocial colonies. We also found significant overlap between our genes and differentially expressed genes among guarding mother genes that were upregulated over all others. This is particularly interesting because we only sequenced daughters from both social and subsocial colonies yet found significant overlap with genes related to maternal behaviours, especially those important for offspring care. This suggests that we identified genetic components of a key maternal behaviour in daughters (Fig. 4)[32]. These findings provide some of the first genetic evidence supporting the observed transcriptomic differences that align with the Maternal Heterochrony Hypothesis.

One of the genes identified with outlier levels of F_{st} associated with social or subsocial behaviour that was also differentially expressed among mothers and daughters was odorant receptor 13a-like that had several SNPs with outlier levels of F_{st} within the gene. Social individuals require enhanced communication with nestmates to organize task allocation, offspring care, nest defense and to display their reproductive status via pheromones[80]. Plasticity in these chemical signals is regulated at the transcriptional level to alter composition, proportion, and timing of display of these chemicals[81]. In a study of the communication system during the transition from solitary to social behaviour in the socially polymorphic halictid bee *Lasioglossum albipes*, Wittwer et al.[80] found that solitary individuals had fewer sensilla, altered antennal morphology and changes in signaling chemistry compared to their social counterparts. With fewer opportunities for communication, the solitary individuals have reduced investment in communication and have almost lost their ability for this type of complex communication[80,82]. Furthermore, specific sensory receptors such as olfactory receptors have also been shown to be under positive selection in *L. albipes* as demonstrated by genetic differences among social and solitary groups[42].

Antioxidant enzymes are important for bees who regularly face environmental stimuli, such as foragers who frequently encounter different flowers[83]. This need has driven the expansion of certain gene families such as the Cyp6A family in the genome of *Apis mellifera*[84]. Research has shown a link between antioxidants and behaviour; for example, the expression of Cytochrome p450 4C1 is upregulated in solitary *Ceratina japonica* females, while cytochrome p450 6K1 is more prominent in worker-class genes[51]. In our study, we identified these two genes, p450 4C1-like and p450 6K1-like, which exhibited significant genetic differentiation between social and subsocial *C. calcarata*. Lago, et al.[85] focused on differentially expressed genes in larval gonads to understand the link between polyandrous mating and ovariole number in developing reproductive individuals, including the gene Cyp9E2, which we also identified as having outlier levels of genetic differentiation here. Additionally, another study identified Cyp6A2 as being under positive selection in *Apis*[84], which we also found in our study. These findings further reinforce the connection between transcriptomic changes associated with developmental and behavioural differences and their genomic foundations.

In this study, our investigation into the genetic underpinnings of social behaviour in *Ceratina calcarata* sheds light on the intricate dynamics at play during the early stages of social evolution. While our findings indicate minimal genetic differentiation between social and subsocial groups overall, the presence of a few loci with higher differentiation suggests potential targets for further investigation. Given that *C. calcarata* forms a single genetic population, these differences may not be the primary drivers of social behaviour but rather part of a broader regulatory or environmental mechanism influencing behavioural plasticity. These observations highlight the complex nature of social behaviour, suggesting that while genetic factors contribute to the observed behavioural differences, they do not act alone in shaping behavioural outcomes. Our research highlights *C. calcarata* as a valuable model for studying the early stages of social evolution, offering insights into the complex interplay of genetic, environmental, and epigenetic factors. Therefore, further studies are needed to fully understand the mechanisms driving social behaviour in this species. Because this is the first study to explore a genetic link to social behaviour, additional validation experiments are necessary to support our findings and determine whether the observed genetic differences have a functional significance or if social behaviour is primarily shaped by environmental and regulatory factors.

Methods

Sample collection and sequencing

We collected diploid female *Ceratina calcarata* bees from social (n = 10) or subsocial (n = 10) nests in New Hampshire, USA. The sample sizes here (10 diploid individuals, which is 20 haploid genomes per group) reflect several other studies with a similar number of participants[31,51,86–88] and these sample sizes have been recognized as suitable for this type of population genomic study[89]. The high coverage of these diploid individuals, combined with strict filtering criteria outlined below, minimize the likelihood of genotyping error, thus providing us with high-confidence single nucleotide polymorphism calls. Following established methods[90], social nests were classified as those with a female in the first brood cell, whereas subsocial nests had a male present in the first brood cell. We removed the thorax of each bee, ground it up in liquid nitrogen with a pestle (FisherBrand™ RNase-Free Disposable Pellet Pestles) and extracted DNA using the Mag-Bind® Blood & Tissue DNA HDQ 96 Kit (Omega Bio-tek Inc., USA) optimized for KingFisher™ Flex Purification System (Thermo Fisher Scientific Inc., USA) to a final elution volume of 70-100 µl using established methods[91]. Tissues were lysed using 350 µl of Tissue Lysis Buffer, 20 µl (20 g/mL) of Proteinase K and heating samples overnight at 55 °C. The DNA was quantified using NanoDropTM 2000 Spectrophotometer (Thermo Fisher Scientific Inc., USA) and assessed for quality using 2.0% agarose gel electrophoresis. Library preparation and sequencing were performed at the Genome Quebec Innovation Centre (McGill University). Genome sequencing was performed using the Illumina HiSeq 2500 system (paired end reads, 2 × 150 reads) using four lanes with six bees per lane to yield a total coverage of 167X total (27X per bee) and a genome size of 364 Mb. We performed quality control on all samples to check the quality, read length, GC content and other standard parameters using Qualimap 2[92].

Alignment and SNP calling

We used previously published pipelines for alignment and SNP calling[36,45,49]. Briefly, adapter sequences were trimmed from raw reads using TRIMMOMATIC v0.32[93] and then aligned to the *Ceratina calcarata* ASM165200v1 genome[63] using NextGenMap v0.4.12[94]. Duplicate reads were removed with Picard v1.8 MarkDuplicates (http://broadinstitute.github.io/picard/) and we realigned reads around indels with Genome Analysis Toolkit (GATK)[95]. We then identified single nucleotide polymorphisms (SNPs) using GATK HaplotypeCaller via an established pipeline[96].

Filtering SNPs

We filtered samples on several criteria as per GATK recommendations (MappingQuality < 20, QualitybyDepth < 2.0, FisherStrand > 60, StrandOddsRatio > 3.0, ReadPosRankSum < -8.0, MappingQualityRankSum < -12.5, MappingQualityRankSum > 12.5)[95]. Following previous methods, we then filtered SNPs based on abnormally high or low read depth. The upper limit was determined by calculating 1.5 × the interquartile range of the cumulative depth of each sample[36]. The lower limit for depth was determined by allowing an average depth of coverage of 5 reads per base pair per individual (5 reads × 20 samples lower pooled threshold of 100)[97]. Low frequency SNPs using a minor allele frequency threshold of less than 0.05 were also removed from the sample set[97].

Population genetic analysis

We determined the population structure of social and subsocial *C. calcarata* (n = 20) using ADMIXTURE v.1.30[98]. We ran an unsupervised model with K = 1–4 populations from 10,000 randomly selected SNPs (shuf -n 10000) from the filtered variant call format (VCF) file, in accordance with other studies[45]. This process was repeated 3 times with 20 cross validation steps each (–cv = 20). A stacked bar graph for each replicate and each K value was generated using ggplot2 for R[99].

To prepare the VCF file for further analysis, we first removed sites with more than 50% missing data using VCFtools with the parameter –max-missing 0.5, following the approach used in similar studies[100,101]. We measured genetic differentiation (F_{st}) between social (n = 10) and subsocial (n = 10) *C. calcarata* using VCFtools with the –weir-fst-pop parameter[100]. The F_{st} values associated with each SNP form the basis for all downstream analyses. A sample pool of this size has also been shown to be robust enough for population genetics studies in both a simulation setting[88] and empirical study[31,51,86]. To calculate the gene-wise F_{st}, which is the average F_{st} value across a gene, we identified the location of each SNP within a gene by matching the SNP's position to the gene's start and end points as listed in the GFF file (https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/156304/101/GCF_001652005.1_ASM165200v1/GCF_001652005.1_ASM165200v1_genomic.gff) using a custom Python script. We then averaged the F_{st} values of SNPs located within that gene to determine the gene-wise F_{st} score.

We also looked for evidence of genetic linkage around SNPs with high F_{st} by dividing the entire genome into 500-base-pair non-overlapping genomic windows (VCF tools with the flags –window-pi 500). We then selected windows that contained at least one SNP with outlier levels of F_{st} (which we refer to as Group 1) and windows that did not contain any SNPs with outlier levels of F_{st} (which we refer to as Group 2). We counted the average number of SNPs in each window in both groups and also averaged F_{st} across windows to identify the differences SNPs in Group 1 and Group 2. We used python matplotlib[102] to perform all statistics including a t-test and Cohen's D to quantify the effect size.

SnpEff was used to identify the physical location of all SNPs along a gene[103]. Attributes (intron, exon, intergenic, etc.) were extracted from the INFO column of the output VCF file and were simplified according to information from the SnpEff website (http://pcingola.github.io/SnpEff/se_inputoutput/; Table S1A). This information was used to create a contingency table of attributes for all SNPs compared to those found in outlier SNPs in the top 5% and 1%. We performed a G test of independence on this table using chi2_contingency for scipy.stats (1.10.1) in python followed by a calculation of the standardized residuals to test for directional significance (significantly more or less than expected). We also used the original classification of SNPs within exons (not simplified) to further understand the distribution of our SNPs within coding regions (synonymous variant, missense variant etc.). Donut plots depicting the separation of total SNPs into different categories in the total (outer ring) and outlier (inner ring) distributions were made using Plotly for python (https://plot.ly). To find the gene names associated with each SNP, we used bedtools intersect with the GFF file and the final VCF file[104].

FlyBase search and gene ontology analysis

We used *Drosophila melanogaster* orthologs to *Ceratina calcarata* genes to study the function of genes with high genetic differentiation between behavioural groups. First, we converted *C. calcarata* genes to *Drosophila melanogaster* genes using HymenopteraMine QueryBuilder to identify matching GeneIDs[105]. There was a total of 58% overlap between the *Ceratina calcarata* genome and the *Drosophila melanogaster* genome (7,069 Drosophila genes were matched out of a total of 12,144 *C. calcarata* genes provided on HymenopteraMine; https://hymenopteramine.rnet.missouri.edu/hymenopteramine/begin.do). We then converted the *D. melanogaster* Gene ID's to *Drosophila melanogaster* genes (in the form "FBgn") using Flybase ID validator (https://flybase.org/convert/id)[106]. To explore the biological and molecular function of these orthologous genes we downloaded all gene summaries from the Flybase FTP page for "best gene summaries" (http://ftp.flybase.net/releases/current/precomputed_files/genes/) and merged it with our file based on common column criteria (Table S2).

Gene ontology annotation enrichment clustering analysis was carried out using DAVID 2022[107] to explore the biological process (GOTERM_BP_FAT), molecular function (GOTERM_MF_FAT), and cellular components (GOTERM_CC_FAT) associated with genes with outlier levels of genetic differentiation between subsocial and social bees (within the top 5%, $F_{st} > = 0.052$). We used WordCloud in Python to visualize the most frequently occurring words in enriched Gene Ontology (GO) terms identified through gene ontology enrichment analysis, presented in a functional annotation chart (amueller.github.io/word_cloud; python 3.5.2).

Search for taxonomically restricted genes (OrthoDB)

The OrthoDB v11 database classifies genes and proteins based on their taxonomic levels to identify the ancestry of each gene. The OrthoDB v11 database groups genes into orthologous clusters at different taxonomic levels based on their evolutionary relationships, making it easier to compare genes and their functions across various species[108]. While we recognize that this approach is limited by the quality and quantity of genomes used in the comparison to identify orthologs, OrthoDB is the best resource to identify ancestral and novel genes to date, especially due to the substantial number of hymenopteran genomes within the database[108]. We downloaded flat files from the OrthoDB site, filtered them for annotations involving *Ceratina calcarata* (NCBI taxonomic ID: 156,304), and merged them together based on common criteria within the column (i.e. merged columns with overlapping "NCBI_Tax_ID") to reduce the file size to relevant annotations related to *C. calcarata*. We quantified the total number of orthologs associated with each gene and used the oldest and broadest taxonomic classification to categorize the gene. Our taxonomic classifications included Eukaryota (NCBI Taxonomic ID: 2759), Metazoa (33208), Arthropoda (6656), Hexapoda (6960), Insecta (50557), Holometabola (33392), Hymenoptera (7399), Aculeata (7434) and Apoidea (34735). We then used the F_{st} values associated with social or subsocial behaviour for each gene (see "Population Genomic Analysis") and their oldest taxonomic classification to assess enrichment in a particular category. We performed this analysis for each category in both outlier groups (top 1% and top 5%) and assessed significance with a G test of independence in python. We then grouped genes into "old" (Eukaryota, Metazoa, Arthropoda, Insecta, Holometabola, Hexpoda) vs "new" (Hymenoptera, Aculeata, Apoidea) to assess if our genes with outlier levels of F_{st} are overrepresented in old or new categories and assessed each with a G test of independence.

Comparison to previous gene expression studies

We compared genes associated with social or subsocial behaviour in our study to previously published gene expression datasets in *C. calcarata*. These gene sets used the "Ccalc" gene names, whereas our lists used NCBI's "LOC" gene identifiers. To match these names, we performed a nucleotide BLAST of the genome available for download on the NCBI website and the GFF file to extract LOC names associated with each scaffold. To ensure an exact match between "LOC" genes and "Ccalc" names, we merged these files based on the "start" and "end" positions found in the GFF file to make sure they were a 1:1 match. We were able to match 9,438 "Ccalc" names to 9,916 "LOC" genes, with some genes having multiple exact matching "Ccalc" names (Supplemental Table S3). We then compared genes with outlier levels of F_{st} (top 5%) between social and subsocial *C. calcarata* and compared it to differentially expressed or methylated genes identified in other studies. The significance of the overlap between our outlier genes and other studies was evaluated using a hypergeometric test in python (scipy.stats, stats.hypergeom.sf). We compared our list of genes with outlier levels of F_{st} between social and subsocial *C. calcarata* with:

A. Methylated genes (Rehan et al.[63] supplemental Table 7, fracMeth.CDS.avg > 0.01) and genes showing signs of positive selection (Rehan et al.[63] supplemental Table 10)
B. Differentially expressed genes (DEGs), Alternatively Spliced Genes (ASG), Differentially Methylated Genes (DMG), or Methylated Genes (MG) associated with maternal care (Arsenault et al.[33] supplemental materials 41467_2018_5903_MOESM4_ESM.xslx)
C. Genes associated with aggression (Withee et al.[53] supplemental Table S3; WW > LL)
D. Genes associated with behaviour (foraging, guarding) and social status (mothers, autumn daughters, regular daughters) (Shell and Rehan[32] supplemental Table 8).
E. Genes in gene families that have significantly expanded or contracted with the evolution of social complexity based on a comparative genomic analysis of 16 bee genomes (Shell et al.[34] supplemental Table S5).

Data availability

All original FASTA files can be found uploaded to NCBI upon publication (PRJNA1028991).

Received: 19 June 2024; Accepted: 10 February 2025
Published online: 15 February 2025

References

1. Wilson, E. O. & Holldobler, B. Eusociality: Origin and consequences. *Proc. Natl. Acad. Sci. USA* **102**, 13367–13371 (2005).
2. Groom, S. V. C. & Rehan, S. M. Climate mediated behavioral variability in facultatively social bees. *Biol. J. Linnean Soc.* **125**, 165–170 (2018).
3. Hamilton, W. D. The genetical evolution of social behaviour. *J. Theor. Biol.* **7**, 1–16 (1964).
4. Kapheim, K. M. et al. Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 6239 (2015).
5. Page, R. E. Jr. & Amdam, G. V. The making of a social insect: developmental architectures of social design. *Bioessays* **29**, 334–343 (2007).
6. Rehan, S. M. & Toth, A. L. Climbing the social ladder: the molecular evolution of sociality. *Trends Ecol. Evol.* **30**, 426–433 (2015).
7. Rehan, S.M. & Richards, M.H. The influence of maternal quality on brood sex allocation in the small carpenter bee, *Ceratina calcarata*. *Ethology* (2010).
8. Shell, W. A. & Rehan, S. M. Behavioral and genetic mechanisms of social evolution: insights from incipiently and facultatively social bees. *Apidologie* **49**, 13–30 (2017).
9. West-Eberhard, M.J. Flexible strategy and social evolution. *Anim. Soc. Theor. Fac.* 35–51 (1987).

10. Rehan, S. M. et al. Conserved genes underlie phenotypic plasticity in an incipiently social bee. *Genome Biol. Evolut.* **10**, 2749–2758 (2018).
11. Lawson, S. P., Ciaccio, K. N. & Rehan, S. M. Maternal manipulation of pollen provisions affects worker production in a small carpenter bee. *Behav. Ecol. Sociobiol.* **70**, 1891–1900 (2016).
12. Rehan, S. M., Berens, A. J. & Toth, A. L. At the brink of eusociality: transcriptomic correlates of worker behaviour in a small carpenter bee. *BMC Evolut. Biol.* **14**, 1–11 (2014).
13. Mikát, M., Franchino, C. & Rehan, S.M. Sociodemographic variation in foraging behavior and the adaptive significance of worker production in the facultatively social small carpenter bee, Ceratina calcarata. *Behav. Ecol. Sociobiol.* **71** (2017).
14. Michener, C. D. *The Social Behavior of the Bees* (Harvard University Press, 1974).
15. Wilson, E. O. *The Insect Societies* (Harvard University Press, 1971).
16. Linksvayer, T. A. & Wade, M. J. The evolutionary origin and elaboration of sociality in the aculeate hymenoptera: Maternal effects, sib-social effects, and heterochrony. *Q. Rev. Biol.* **80**, 317–336 (2005).
17. True, J. R. & Carroll, S. B. Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* **18**, 53–80 (2002).
18. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Wiley-Blackwell, 2005).
19. Woodard, S. H. et al. Genes involved in convergent evolution of eusociality in bees. *Proc. Natl. Acad. Sci. USA* **108**, 7472–7477 (2011).
20. Grozinger, C. M., Fan, Y., Hoover, S. E. & Winston, M. L. Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (Apis mellifera). *Mol. Ecol.* **16**, 4837–4848 (2007).
21. Whitfield, C. W., Cziko, A. M. & Robinson, G. E. Gene exprssion profiles in the brain predict behavior in individual honey bees. *Science* **302**, 296–299 (2003).
22. Ometto, L., Shoemaker, D., Ross, K. G. & Keller, L. Evolution of gene expression in fire ants: The effects of developmental stage, caste, and species. *Mol. Biol. Evol.* **28**, 1381–1392 (2011).
23. 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).
24. West-Eberhard, M. & Turillazzi, S. Natural history and evolution of paper wasps (1996).
25. Toth, A. L. et al. Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science* **318**, 441–444 (2007).
26. Graham, A.M. *et al.* Support for the reproductive ground plan hypothesis of social evolution and major QTL for ovary traits of Africanized worker honey bees (*Apis mellifera* L.). *BMC Evol. Biol.* **11** (2011).
27. 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).
28. 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).
29. Woodard, S. H., Bloch, G. M., Band, M. R. & Robinson, G. E. Molecular heterochrony and the evolution of sociality in bumblebees (Bombus terrestris). *Proc. Biol. Sci.* **281**, 20132419 (2014).
30. Johnson, B.R. & Tsutsui, N.D. Taxonomically restricted genes are associated with the evoluton of socialiy in the honey bee. *BMC Genomics* **12** (2011).
31. Harpur, B. A. et al. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc. Natl. Acad. Sci. USA* **111**, 2614–2619 (2014).
32. Shell, W. A. & Rehan, S. M. Social modularity: conserved genes and regulatory elements underlie caste-antecedent behavioural states in an incipiently social bee. *Proc. Biol. Sci.* **286**, 20191815 (2019).
33. Arsenault, S. V., Hunt, B. G. & Rehan, S. M. The effect of maternal care on gene expression and DNA methylation in a subsocial bee. *Nat. Commun.* **9**, 3468 (2018).
34. Shell, W. A. et al. Sociality sculpts similar patterns of molecular evolution in two independently evolved lineages of eusocial bees. *Commun. Biol.* **4**, 253 (2021).
35. Chau, K. D., Shamekh, M., Huisken, J. & Rehan, S. M. The effects of maternal care on the developmental transcriptome and metatranscriptome of a wild bee. *Commun. Biol.* **6**, 904 (2023).
36. Imrit, M. A., Dogantzis, K. A., Harpur, B. A. & Zayed, A. Eusociality influences the strength of negative selection on insect genomes. *Proc. R. Soc. B Biol. Sci.* **287**, 20201512 (2020).
37. Wilfert, L., Gadau, J. & Schmid-Hempel, P. Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity (Edinb)* **98**, 189–197 (2007).
38. Sirvio, A. et al. High recombination frequency creates genotypic diversity in colonies of the leaf-cutting ant Acromyrmex echinatior. *J. Evol. Biol.* **19**, 1475–1485 (2006).
39. Kent, C. F. & Zayed, A. Evolution of recombination and genome structure in eusocial insects. *Commun. Integr. Biol.* **6**, e22919 (2013).
40. Oldroyd, B. P. et al. Adaptive, caste-specific changes to recombination rates in a thelytokous honeybee population. *Proc. Biol. Sci.* **288**, 20210729 (2021).
41. Rubin, B. E. R., Jones, B. M., Hunt, B. G. & Kocher, S. D. Rate variation in the evolution of non-coding DNA associated with social evolution in bees. *Philos. Trans. R. Soc. B* **374**, 1–13 (2019).
42. Kocher, S. *et al.* The draft genome of a socially polymorphic halictid bee, Lasioglossum albipes. *Genome Biol.* **14** (2013).
43. Kocher, S. D. et al. The genetic basis of a social polymorphism in halictid bees. *Nat. Commun.* **9**, 4338 (2018).
44. Kapheim, K. M. et al. Developmental plasticity shapes social traits and selection in a facultatively eusocial bee. *Proc. Natl. Acad. Sci.* **117**, 13615–13625 (2020).
45. Harpur, B. A. & Rehan, S. M. Connecting social polymorphism to single nucleotide polymorphism: population genomics of the small carpenter bee, *Ceratina australensis*. *Biol. J. Linnean Soc.* **132**, 945–954 (2021).
46. 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).
47. Yan, Z. et al. Evolution of a supergene that regulates a trans-species social polymorphism. *Nat. Ecol. Evol.* **4**, 240–249 (2020).
48. Yagound, B. et al. A single gene causes thelytokous parthenogenesis, the defining feature of the cape honeybee Apis mellifera capensis. *Curr. Biol.* **30**, 2248-2259 e6 (2020).
49. Harpur, B. A. et al. Integrative genomics reveals the genetics and evolution of the honey bee's social immune system. *Genome Biol. Evolut.* **11**, 937–948 (2019).
50. Spotter, A., Gupta, P., Mayer, M., Reinsch, N. & Bienefeld, K. Genome-wide association study of a varroa-specific defense behavior in honeybees (*Apis mellifera*). *J. Hered* **107**, 220–227 (2016).
51. Shell, W. A. & Rehan, S. M. Social divergence: molecular pathways underlying castes and longevity in a facultatively eusocial small carpenter bee. *Proc. Biol. Sci.* **289**, 20212663 (2022).
52. Durant, D. R., Berens, A. J., Toth, A. L. & Rehan, S. M. Transcriptional profiling of overwintering gene expression in the small carpenter bee, *Ceratina calcarata*. *Apidologie* **47**, 572–582 (2015).
53. Withee, J. R. & Rehan, S. M. Social aggression, experience, and brain gene expression in a subsocial bee. *Integr. Comp. Biol.* **57**, 640–648 (2017).

54. Begun, D. J. et al. Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. *PLoS Biol.* **5**, e310 (2007).
55. Qanbari, S. et al. A high resolution genome-wide scan for significant selective sweeps: An application to pooled sequence data in laying chickens. *PLoS ONE* **7**, e49525 (2012).
56. Yan, D. et al. Genomic insights into genetic diversity and local adaptation of a dominant desert steppe feather grass, Stipa breviflora Griseb. *Front Plant Sci.* **14**, 1170075 (2023).
57. Feng, X. J., Jiang, G. F. & Fan, Z. Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, Ceracris kiangsu. *Sci. Rep.* **5**, 13758 (2015).
58. Storz, J. F., Payseur, B. A. & Nachman, M. W. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Mol. Biol. Evol.* **21**, 1800–1811 (2004).
59. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
60. Akey, J. M., Zhang, G., Zhang, K., Jin, L. & Shriver, M. D. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* **12**, 1805–1814 (2002).
61. Christmas, M. J. et al. Genetic barriers to historical gene flow between cryptic species of alpine bumblebees revealed by comparative population genomics. *Mol. Biol. Evol.* **38**, 3126–3143 (2021).
62. 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).
63. Rehan, S. M., Glastad, K. M., Lawson, S. P. & Hunt, B. G. The genome and methylome of a subsocial small Carpenter Bee, *Ceratina calcarata*. *Genome Biol. Evol.* **8**, 1401–1410 (2016).
64. Rehan, S. M. & Richards, M. H. Nesting biology and subsociality in Ceratina calcarata (Hymenoptera: Apidae). *Can. Entomol.* **142**, 65–74 (2010).
65. Linksvayer, T. A. & Johnson, B. R. Re-thinking the social ladder approach for elucidating the evolution and molecular basis of insect societies. *Curr. Opin. Insect. Sci.* **34**, 123–129 (2019).
66. Shpigler, H. Y. et al. Behavioral, transcriptomic and epigenetic responses to social challenge in honey bees. *Genes Brain Behav.* **16**, 579–591 (2017).
67. Jones, B.M. *et al.* Individual differences in honey bee behavior enabled by plasticity in brain gene regulatory networks. *Elife* **9** (2020).
68. Rittschof, C. C. & Robinson, G. E. Manipulation of colony environment modulates honey bee aggression and brain gene expression. *Genes Brain Behav.* **12**, 802–811 (2013).
69. Zayed, A. & Whitfield, C. W. A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee Apis mellifera. *Proc. Natl. Acad. Sci. USA* **105**, 3421–3426 (2008).
70. Dogantzis, K. A. et al. Thrice out of Asia and the adaptive radiation of the western honey bee. *Sci. Adv.* **7**, 1–10 (2021).
71. Erokhin, M. M., Mikhailova, A. V., Georgiev, P. G. & Chetverina, D. A. Effect of transcription on the white gene enhancer integrated into the intron. *Dokl. Biochem. Biophys.* **484**, 45–47 (2019).
72. Valenzuela, L. & Kamakaka, R. T. Chromatin insulators. *Annu. Rev. Genet.* **40**, 107–138 (2006).
73. Gospocic, J. et al. The neuropeptide corazonin controls social behavior and caste identity in ants. *Cell* **170**, 748-759 e12 (2017).
74. Schrader, L., Simola, D. F., Heinze, J. & Oettler, J. Sphingolipids, transcription factors, and conserved toolkit genes: Developmental plasticity in the ant Cardiocondyla obscurior. *Mol. Biol. Evol.* **32**, 1474–1486 (2015).
75. 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).
76. Carthew, R. W. Gene regulation and cellular metabolism: An essential partnership. *Trends Genet.* **37**, 389–400 (2021).
77. Wojciechowski, M. et al. Phenotypically distinct female castes in honey bees are defined by alternative chromatin states during larval development. *Genome Res.* **28**, 1532–1542 (2018).
78. Simola, D. F. et al. Epigenetic (re)programming of caste-specific behavior in the ant Camponotus floridanus. *Science* **351**, aac6633 (2016).
79. Toth, A. L. & Robinson, G. E. Evo-devo and the evolution of social behavior. *Trends Genet.* **23**, 334–341 (2007).
80. Wittwer, B. et al. Solitary bees reduce investment in communication compared with their social relatives. *Proc. Natl. Acad. Sci.* **114**, 6569–6574 (2017).
81. Malka, O., Nino, E. L., Grozinger, C. M. & Hefetz, A. Genomic analysis of the interactions between social environment and social communication systems in honey bees (Apis mellifera). *Insect Biochem. Mol. Biol.* **47**, 36–45 (2014).
82. Gill, K. P., van Wilgenburg, E., Macmillan, D. L. & Elgar, M. A. Density of antennal sensilla influences efficacy of communication in a social insect. *Am. Nat.* **182**, 834–840 (2013).
83. Mao, W., Schuler, M. A. & Berenbaum, M. R. Task-related differential expression of four cytochrome P450 genes in honeybee appendages. *Insect Mol. Biol.* **24**, 582–588 (2015).
84. Johnson, R. M., Harpur, B. A., Dogantzis, K. A., Zayed, A. & Berenbaum, M. R. Genomic footprint of evolution of eusociality in bees: Floral food use and CYPome "blooms". *Insectes Sociaux* **65**, 445–454 (2018).
85. Lago, D. C., Nora, L. C., Hasselmann, M. & Hartfelder, K. Positive selection in cytochrome P450 genes is associated with gonad phenotype and mating strategy in social bees. *Sci. Rep.* **13**, 5921 (2023).
86. Eimanifar, A., Brooks, S.A., Bustamante, T. & Ellis, J.D. Population genomics and morphometric assignment of western honey bees (*Apis mellifera* L.) in the Republic of South Africa. *BMC Genomics* **19** (2018).
87. Yancan, L., Tianle, C., Yunhan, F., Delong, L. & Guizhi, W. Population genomics and morphological features underlying the adaptive evolution of the eastern honey bee (Apis cerana). *BMC Genomics* **20**, 869 (2019).
88. Nazareno, A. G., Bemmels, J. B., Dick, C. W. & Lohmann, L. G. Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Mol. Ecol. Resources* **17**, 1136–1147 (2017).
89. Willing, E. M., Dreyer, C. & van Oosterhout, C. Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE* **7**, e42649 (2012).
90. Shell, W. A., Rehan, S. M. & Holman, L. The price of insurance: costs and benefits of worker production in a facultatively social bee. *Behav. Ecol.* **29**, 204–211 (2018).
91. Conflitti, I.M. *et al.* Bees in the six: Determinants of bumblebee habitat quality in urban landscapes. *Ecol. Evolut.* **12** (2022).
92. Okonechnikov, K., Conesa, A. & Garcia-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32**, 292–294 (2016).
93. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
94. Sedlazeck, F. J., Rescheneder, P. & von Haeseler, A. NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* **29**, 2790–2791 (2013).
95. Van der Auwera, G.A. *et al.* From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11 10 1–33 (2014).
96. Dogantzis, K. A. et al. Insects with similar social complexity show convergent patterns of adaptive molecular evolution. *Sci. Rep.* **8**, 10388 (2018).
97. Kapheim, K. M. et al. Draft Genome Assembly and Population Genetics of an Agricultural Pollinator, the Solitary Alkali Bee (Halictidae: Nomia melanderi). *G3 (Bethesda)* **9**, 625–634 (2019).

98. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
99. Wickham, H. *ggplot2: Elegant Grpahics for Data Analysis* (Springer, New York, 2016).
100. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
101. Waiker, P. et al. Recombination mapping of the Brazilian stingless bee Frieseomelitta varia confirms high recombination rates in social hymenoptera. *BMC Genomics* **22**, 673 (2021).
102. Hunter, J.D. MATPLOTLIB: A 2D graphics environment. *Sci. Program.* (2007).
103. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
104. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
105. Walsh, A. T., Triant, D. A., Le Tourneau, J. J., Shamimuzzaman, M. & Elsik, C. G. Hymenoptera Genome Database: New genomes and annotation datasets for improved go enrichment and orthologue analyses. *Nucleic Acids Res.* **50**, D1032–D1039 (2022).
106. Gramates, L. S. et al. FlyBase: a guided tour of highlighted features. *Genetics* **220**, iyac035 (2022).
107. Sherman, B.T. *et al.* DAVID: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* (2022).
108. Kuznetsov, D. *et al.* OrthoDB v11: Annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* (2022).

Acknowledgements

We would like to acknowledge Ida Conflitti for her help with organizing and helping perform bench work. We would also like to acknowledge Sandra Rehan for providing the samples and multiple genome files that helped make this project possible.

Author contributions

Research was designed in collaboration with AZ, SMR, and DBS. Whole bee samples of *C. calcarata* were provided by SMR. All sample preparation was performed by DBS with the help of Ida Conflitti. All analysis was performed by DBS. Manuscript was prepared by DBS. Guidance throughout was provided by AZ and SMR.

Funding

York University Research Chair; Natural Sciences and Engineering Research Council of Canada awarded to AZ, NSERC Discovery Grant, Supplement awarded to SMR, EWR Steacie Memorial Fellowship awarded to SMR.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-89870-9.

Correspondence and requests for materials should be addressed to A.Z.

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.