Causes and consequences of eusocial genome evolution in the small carpenter bee *C. calcarata* and the honey bee *A. mellifera*

Dova B. Brenman

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOLOGY, YORK UNIVERSITY, TORONTO, ONTARIO

April 2026

Abstract

Bees provide excellent systems for understanding how social behaviour evolves and how it influences the genome, a central question within sociobiological research. Advances in genomic technologies now enable investigation of the evolutionary and molecular mechanisms underlying social complexity. I use genomic and population genetic approaches to investigate the mechanisms that contribute to the emergence of social behaviour and the genomic consequences of advanced eusociality in bees. I first review advances in bee genomics, highlighting how genome sequencing, transcriptomics, and other tools have improved our understanding of bee evolution and behaviour. Then, I suggest that combining molecular and genomic approaches provides a framework for identifying genes and molecular pathways associated with eusocial traits. Using the facultatively social *Ceratina calcarata*, I investigate genomic patterns linked to early stages of social evolution. Comparisons between social and subsocial colonies show that loci with high genetic differentiation are often located in or near regulatory regions. Comparisons with *Ceratina strenua* also identify genes under positive selection, particularly in genes involved in regulation and metabolism. Finally, I examine the genomic consequences of longevity in terms of changes in mutation and recombination rate in the advanced eusociality in *Apis mellifera*. Mutations are more likely to occur in regions with higher recombination and are less common in constrained genomic regions, suggesting that recombination and purifying selection shape patterns of mutation in the honey bee genome. Together, this work demonstrates how genomic approaches can uncover the origins of social behaviour and how eusocial life histories influence genome evolution in bees.

Acknowledgements

"בל ההתחלות קשות, אך קשה מהן היא ההתמדה"

“All beginnings are hard, but harder yet is perseverance”

-Chaim Nachman Bialik

First and foremost, I would like to acknowledge and thank Dr. Amro Zayed for his guidance and support in all thing’s bees and genomics. Dr. Zayed has demonstrated what a true leader and a life-long learner can look like. His passion for science, science communication and drive for excellence are things I admire and try to live up to. I would also like to thank my supervisory committee, Dr. Sandra Rehan and Dr. Laurence Packer and Dr. Joel Shore for their insights over the years and their access to bees. The Zayed lab has evolved over time but has consistently been a supportive environment. When I joined the lab, I did not know anything about genomics or coding but members of the lab including Tanushree, Katie, Clement, Stephen, Harshil, and Arshad were each willing to answer my basic questions and both get me started and kept me going. A special shoutout to my lab besties Tanu and Katie for keeping me motivated. I will forever be inspired by their intelligence, work ethic and grateful for their friendship. I could not have survived this without you. Or coffee.

I would be remiss if I did not dedicate a section of my acknowledgement to my husband and best friend, Josh. You have always been my biggest supporter and always believed in me even when I didn’t believe myself. Throughout my PhD journey we have welcomed two children, for whom this process is dedicated. Through this whole process, I hope I have taught my children the value of resilience—that when things are difficult, you are strong and you can pick yourself up, keep trying, and ultimately achieve something meaningful for yourself.

Table of Contents

List of Figures

List of Tables

List of Supplemental Figures

Chapter 1 Introduction

The evolution of social behaviour has long been a central question in sociobiology. With the advent of high-throughput sequencing and comparative genomic tools, understanding the genomic basis of sociality has become increasingly tractable. Comparative genomics, population genetic analyses, and genome-wide association studies (GWAS) now allow researchers to detect adaptive changes, measure natural selection, and connect genotype to phenotype at broader scales and faster rates. These tools provide powerful frameworks for testing hypotheses about the origins and preservation of social behaviour. In Chapter 2 of my dissertation, I provide a comprehensive review of how a wide range of genomic and molecular tools have transformed our understanding of bee phylogeny and behaviour. Briefly, I discuss how modern genomic techniques have greatly improved our understanding of how bees evolved, how complex social behaviours like eusociality arise, and how genes influence behaviour and caste differences. This chapter focuses on four main ideas: using genomes to trace bee evolution, studying alternative splicing and regulatory RNAs and their potential role in social and behavioural flexibility, applying new DNA-based methods to understand bee–environment interactions, and using these insights to support bee conservation under environmental change. In Chapter 3, I outline how although we have many new complete insect genomes, we still do not fully understand the genes or genetic mechanisms underlying eusocial behaviour and its evolution. Here I review the multiple approaches researchers have used to investigate the genetics and evolution of social traits, including population genomics, genome-wide association studies, and experimental gene manipulation using tools such as CRISPR/Cas9. I argue that integrating these methods offers a more powerful strategy for identifying and confirming the genes underlying eusociality than relying on any single approach in isolation.

Several theories have been proposed to explain how eusociality evolves from solitary behaviour and how it is elaborated into the complex social systems observed in insects (Figure 1-1). Two prominent frameworks are the genetic toolkit hypothesis and the novel gene hypothesis. The genetic toolkit hypothesis proposes that eusociality arises primarily through changes in the regulation of pre-existing genes or gene networks, allowing genes to be activated or repressed across developmental stages or tissues (True and Carroll 2002; Woodard *et al.* 2011; Rehan and Toth 2015). Such regulatory plasticity can generate behavioural variation among individuals and facilitate the emergence of rudimentary caste systems, a

pattern that has been observed in eusocial ants (SIMOLA *et al.* 2013; SCHRADER *et al.* 2015; GOSPOCIC *et al.* 2017), the eusocial bumblebee *Bombus terrestris* (WOODARD *et al.* 2011; WOODARD *et al.* 2014) and the halictid bee *Lasioglossum albipes* (KOCHER *et al.* 2018). In contrast, the novel gene hypothesis suggests that newly emerged, taxonomically restricted genes play an important role in the elaboration of social behaviour after sociality has already been established with supporting evidence from the advanced eusocial honey bee (JOHNSON AND TSUTSUI 2011).

In Chapter 4, I test the hypothesis that early social evolution is primarily associated with genetic differences in conserved genes consistent with the genetic toolkit hypothesis. To do this, I focus on the facultatively social bee *Ceratina calcarata* that exhibits natural variation in social behaviour, ranging from subsocial colonies (maternal care without reproductive division of labour) to social colonies with a reproductive division of labour. Specifically, in some colonies, the mother produces a small worker daughter that acts as an altruistic helper and aids the mother in performing colony tasks; following previous work (MIKÁT *et al.* 2017), I refer to these as "social" colonies. In contrast, colonies in which the mother performs all colony tasks, including provisioning and caring for her brood, without the small worker daughter are referred to as "subsocial" colonies. This variation provides a powerful framework for comparing genomic patterns between social and subsocial colonies within the same environment. Using population genomic metrics such as genetic differentiation (F_{st}), I identify genetic differences between these behavioural states to test whether divergence associated with sociality is concentrated in genes and pathways predicted by the genetic toolkit hypothesis.

In Chapter 5, I test the complementary hypothesis that positive selection on protein-coding genes, including potentially novel genes, contributes to the early stages of social evolution. While much of the previous work has focused on highly eusocial insects, where many adaptive changes are thought to be fixed, incipiently social species provide an opportunity to detect adaptive evolution as it is emerging. Here, I compare protein-coding sequences of *C. calcarata* and *C. strenua* to ask how coding sequence evolution may be linked to early social behaviour in *C. calcarata*. The small carpenter bee *C. strenua* is a close relative of *C. calcarata*, from which is diverged approximately 92,000 years ago. *C. strenua* displays extended maternal care and has been reported to act socially and produce a small worker daughter throughout its geographic range (LAWSON *et al.* 2018). However, while *C. calcarata* exhibits a well-characterized incipiently social system, including active maternal manipulation of offspring size to produce a morphologically distinct worker caste (LAWSON *et al.* 2016; LAWSON *et al.* 2017) and caste-associated divergence in brain gene expression (SHELL AND REHAN 2019), *C. strenua*, despite sharing some capacity for social behaviour (LAWSON *et al.* 2018), displays a less consistent and less elaborated social phenotype. Therefore, *C. strenua* serves as

a phylogenetically appropriate outgroup against which signatures of positive selection specific to the *C. calcarata* lineage can be identified. This comparative approach allows me to identify signatures of positive selection and to evaluate whether coding sequence evolution plays a detectable role during the earliest stages of social evolution.

Finally, in Chapter 6, I focus on the honey bee (*Apis mellifera*) to examine how maternal aging influences genome evolution with the hypothesis that *de novo* mutations will increase with maternal age likely due to increased germline cell divisions over time that have more opportunity to introduce novel mutations. I further hypothesize that these mutations will be enriched in genomic regions that have high recombination rates. Increased longevity has long been considered a consequence of eusociality, which uniquely uncouples the trade-off between reproduction and longevity in the reproductive queen, allowing her to achieve both a high reproductive output and an extended lifespan, while workers bear the costs of this trade-off (Flatt 2011; Rodrigues and Flatt 2016; Culina *et al.* 2019; Kreider *et al.* 2021). Honeybee queens are able to reproduce for several years rearing either haploid male drones or using stored sperm to generate diploid female workers (Trivers and Hare 1976; Michener 2007). This makes them an ideal model for studying the genomic consequences of aging in a eusocial context. In collaboration with colleagues from across Canada, I investigated how queen age affects the genome by sequencing and analyzing her drone offspring. Because honey bees, like other Hymenoptera, are haplodiploid, sequencing drones is akin to sequencing the unfertilized eggs of their mother queen. This provides a unique framework for isolating and detecting the effects of maternal aging on offspring genomes without the confounding influence of paternal contributions. By sequencing the genomes of drones from queens of different ages, I was able to examine how maternal aging shapes patterns of mutation and directly test the hypothesis that increased aging results in increased *de novo* mutations.

Together, my dissertation uses a genomic approach to study social behaviour across two very different forms of social organization: one that is socially flexible and one that is highly derived. By utilizing genome sequencing alongside bioinformatic and population genetic tools, my research highlights how changes at the genomic level contribute to social evolution. My work on *Ceratina calcarata* examines the genetic differences among social groups and identifies genes under positive selection in a species undergoing a transition from less complex sociality, addressing which genomic changes lead to social evolution. In contrast, my study on *Apis mellifera* examines the downstream genomic consequences of advanced eusocial evolution and how social evolution leads to genomic change. Overall, these chapters demonstrate that understanding the evolution of social behaviour requires examining many species with diverse behaviours to identify the genomic signatures left by both the transition to and elaboration of social

behaviour, and how in turn, eusociality lead to changes in genome evolution. By combining data from species representing different social strategies, this dissertation contributes to a more comprehensive understanding of how social behaviour evolves, is maintained, and continues to shape genomes in social insects.

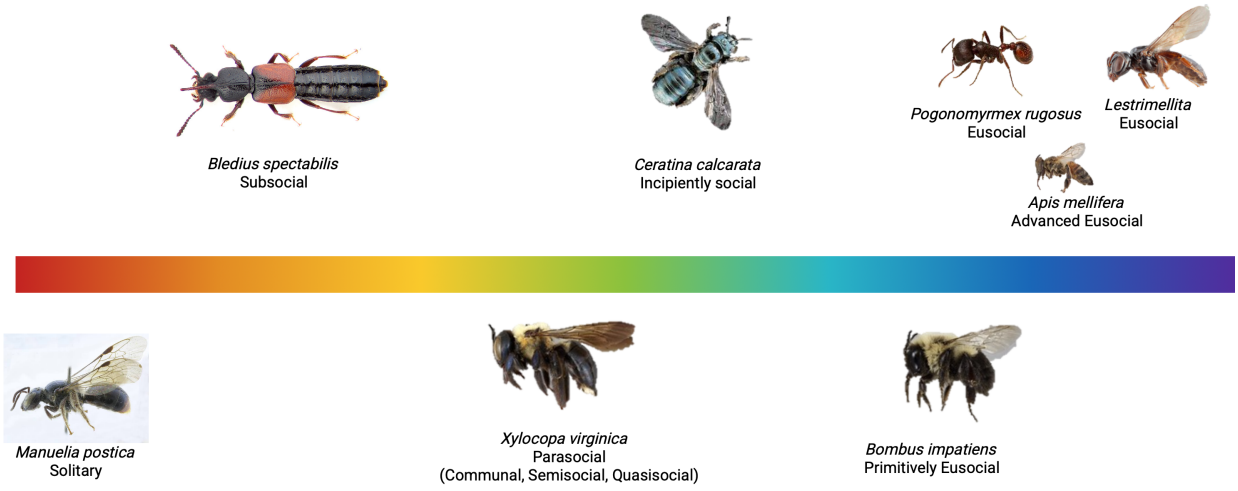


Figure 1-1 Social behaviour is a spectrum ranging from solitary behaviour to advanced eusocial behaviour.
Studying species at the terminal end of sociality can help identify genomic changes that occur as a consequence of sociality. Studying species in the middle of the spectrum, especially those with facultative behaviour, can help identify genomic changes associated with the evolution of sociality.

Chapter 2 The evolution of bees: insights from 'omic studies[1]

Dova Brenman-Suttner[1] and Amro Zayed[1]

Summary

Bees are important global pollinators that play a vital role in maintaining ecosystems and supporting global food production. They also exhibit a diversity of social organization making them ideal model organisms for studying the evolution of sociality in animals. Recent advancements in genome sequencing have enabled researchers to address longstanding questions about the evolution of social behaviour in bees, particularly in the relatively few species that exhibit complex social structures, such as *Apis*. Whole genome phylogenies have enhanced our understanding of the complex evolutionary history of bees, providing a foundation for studying the evolution of specific traits, including eusociality. Recent transcriptomic and alternative splicing studies have advanced our understanding of how gene regulation and expression patterns contribute to behavioural plasticity, caste differentiation, and the emergence of social complexity. Comparative genomics across a range of bees with varying social behaviours have aided our understanding of the genomic features associated with social evolution and has shed light on its molecular underpinnings. Genomic approaches like GWAS and population genomic comparisons, combined with advanced sequencing technologies, have revolutionized the study of bee evolution, social behaviour and environmental interactions. Pollen metabarcoding and environmental DNA (eDNA) techniques are now being used to quantify the intricate and complex interactions between bees and the plants they visit, and to identify other environmental factors, including pathogens that impact bee health. Additionally, techniques like museomics (using DNA from museum specimens) and broader genomic approaches have been instrumental in revealing how bees have been affected by anthropogenic changes. These tools offer valuable insights into population genetics, conservation biology and the impact of environmental changes on bee populations. These advancements both provide critical insights into the molecular basis of eusociality and species adaptation and offer valuable tools for addressing the urgent challenges facing bee

[1] Published in GBE (2026) https://doi.org/10.1093/gbe/evaf226.

conservation due to anthropogenic change. By leveraging these genomic approaches, researchers can inform strategies for the preservation and sustainable management of bee populations worldwide.

Significance Statement

Bees are essential to global biodiversity and food security, making them key models for studying evolution, behaviour, and adaptation. Advances in genomics—spanning ultraconserved elements (UCEs), various environmental 'omics approaches, and museomics—have revolutionized our understanding of bee phylogeny and social complexity. This review synthesizes recent findings on the molecular mechanisms driving the evolution of eusociality, a complex social system that has arisen independently multiple times. Regulatory processes, including alternative splicing, long non-coding RNAs, microRNAs, and circular RNAs, shape phenotypic plasticity and social behaviour by modulating caste differentiation, reproductive roles, and task specialization. Together, these insights highlight the dynamic genomic architecture of social evolution and provide a foundation for future research in sociogenomics and adaptive evolution.

Introduction

Bees are charismatic insects that are fundamentally important to humans on multiple levels. As pollinators, bees provide indispensable ecosystem services to a large community of flowering plants, thereby serving a critical role in maintaining plant biodiversity and that of the multitude of organisms that require flowering plants for survival and reproduction (Katumo *et al.* 2022). This of course includes humans as approximately one-third of global crop production, which represents species that are at least partially dependent on bee pollination, contributes to the human food supply (Khalifa *et al.* 2021). Some bees also produce honey which has served as a human food staple since prehistoric times (Dunne *et al.* 2021), with the first evidence of managing honey bees coming from ancient Egyptian temples in 2450 BCE (Fraser 1931; Kritsky 2017). Beyond their ecological and economic importance, bees have also served as model organisms for the study of eusocial evolution, complex behaviour, and have even been used as field monitors for ecosystem quality and environmental change (Menzel 2012; Richards 2019; Wizenberg *et al.* 2025). Bees and their intriguing behaviours have also fascinated scientists for centuries. Darwin considered the 'neuter' worker castes of bees a special difficulty for his theory of evolution by natural selection (Darwin 1809-1882), and the 1973 Nobel prize in Physiology and Medicine was partly awarded to Karl von Frisch for his work on deciphering the honey bee's infamous waggle dance (The Nobel Prize in Physiology or Medicine 1973). This fascination with bees and their biology was a large part of why the honey bee was selected as the third insect genome to be sequenced (after the fruit fly *Drosophila melanogaster*, and the malaria vector

Anopheles gambiae). Since the publication of the honey bee genome in 2006, numerous bee genomic resources have been developed. These include BeeBase (https://hymenoptera.elsiklab.missouri.edu/beebase), a comprehensive database for bee genomics that provides genome assemblies and annotations for various species, and the Hymenoptera Genome Database, which features HymenopteraMine (https://hymenoptera.elsiklab.missouri.edu), a data mining tool integrating genomic and biological information from numerous insect genomes. Additionally, the Sequence Read Archive (SRA) on the NCBI website offers extensive transcriptomic and genomic data for bees. To date, there have been 228 bee genomes sequenced (https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=34735) and 70 transcriptomes from 59 bee species (NCBI transcriptome shotgun assembly TSA; https://www.ncbi.nlm.nih.gov/Traces/wgs/).

Here we provide a review on the 'omics tools and techniques that have shaped our understanding of bee biology and evolution. Taxonomy and phylogenetics are critical for understanding both patterns of biodiversity and phenotypic evolution. Advances in genomics have been instrumental in enhancing our understanding of the bee 'tree of life'. For example, the availability of genome sequences has allowed researchers to identify ultraconserved elements that have proven very effective in generating high quality phylogenies for a large group of bees. Understanding the genetics and molecular biology underlying social behaviour has been a goal of sociogenomics since its inception with the sequencing of the honey bee genome in 2006 (The Honeybee Genome Sequencing Consortium 2006). While we are still searching of the 'genes for altruism' (Toth and Zayed 2021), research over the past decade has generated a great deal of knowledge on the emerging roles of alternative splicing and non-coding RNAs in generating behavioural plasticity in bees (Harpur *et al.* 2019; Shell *et al.* 2021; Traniello *et al.* 2023; Brenman-Suttner and Zayed 2024), which is an emerging topic that we also review here. Large genome datasets have also enabled the cross-species comparisons that shed light on the evolution of social behaviour, allowing researchers to address some long-standing sociobiological questions. Finally, metagenomics and metabarcoding have opened the doors for fully exploring the plethora of interactions between bees and plants, leading to vast new insights into bee biology and health. Additionally, advances in sequencing preserved specimens have enabled researchers to 'resurrect' bee genomes to understand threats to their population viability over time.

1. Tools for studying the evolution of bees

1.1 Genomic tools for phylogenetic analysis

With over 20,000 species belonging to 500 genera and seven families, bees are extraordinarily diverse in terms of morphology and behaviour (Almeida *et al.* 2023). Bees arose during the mid-Cretaceous period around 140 to 110 Mya (million years ago) along with flowering plants (Almeida *et al.* 2023; Ascher and Pickering 2024). The largest bee family, Apidae, includes 5,600 species spread across over 30 tribes and 170 genera, displaying a wide range of social behaviours (Michener 2007; Cardinal *et al.* 2010). The family Apidae includes several subfamilies that display a wide range of social behaviours. For example, the subfamily Apinae encompasses species with varying levels of social complexity: advanced eusocial species like honey bees (*Apis* spp.), primitively eusocial species such as bumblebees (*Bombus* spp.), and other eusocial lineages like stingless bees (*Meliponini*). It also includes orchid bees (*Euglossini*), which are largely solitary as well as Nomadinae and Xylocopinae that exhibit a range of social behaviours (Michener 1974). Reconstructing the phylogenetic relationships among bee groups helps us understand when these diverse behaviours evolved. Phylogenetic studies are also valuable for understanding whether simpler forms of social behaviour, such as communal living, are prerequisites for the evolution of more complex eusociality, and to what extent reversals to solitary living occur (Romiguier *et al.* 2022). Phylogenomic analyses, which use entire genomes, offer deeper insights into evolutionary relationships compared to traditional molecular methods that rely on limited molecular data (Ribeiro and Espindola 2024). Advances in next generation sequencing (Ballare *et al.* 2019; Johnson 2019) have allowed researchers to assemble and re-sequence the genomes of diverse bee species from across the evolutionary tree of life for phylogenetic reconstruction (Almeida *et al.* 2023; Henriquez-Piskulich *et al.* 2024). To date, 140 bee species belonging to 47 genera have been sequenced (https://www.ncbi.nlm.nih.gov/datasets/genome/), representing only 0.7% of the described bee species and 9.4% of the described genera. This limited genomic representation underscores the need for continued sequencing efforts, particularly in underrepresented lineages. Earlier phylogenetic studies, prior to the use of UCE-based methods, relied on multi-gene datasets. For example, Hedtke *et al.* (2013) combined datasets from 349 bee genera and analyzed 17,000 sites from nuclear protein coding genes, confirming the monophyly of each bee family. It also supported Melittidae as the sister group to all other bee families (Hedtke *et al.* 2013) consistent with findings from Brady et al. (2011), but differing from other studies that place Melittidae as sister only to long-tongued bees (Danforth *et al.* 2006; Michez *et al.* 2009; Debevec *et al.* 2012; Cardinal and Danforth 2013). The integration of next-generation whole genome sequencing, ultraconserved elements (UCEs) (Gueuning *et al.* 2020), and DNA barcoding (Trunz *et al.* 2016;

PRAZ *et al.* 2019) has significantly improved phylogenomic resolution, clarifying uncertain evolutionary relationships and challenging established phylogenies. In the following sections, we review major studies that have employed new techniques for generating data and their contributions to understanding the phylogeny of Apidae.

1.2 Phylogenomics and Ultra-conserved Elements (UCEs)

Early phylogenetic relationships have traditionally been inferred using morphological traits, but these can be limited by low variation between taxa, polymorphisms within species, or geographic variation that is unrelated to speciation. DNA barcoding is one method of identifying phylogenetic relationships and is often performed using the cytochrome oxidase subunit 1 (COI) gene because it differs enough between species to allow for species-level resolution (KOWALSKA *et al.* 2019). The COI gene in combination with morphology has been used previously for bee identification (PAULY *et al.* 2019; PRAZ *et al.* 2019). However, COI can be unreliable due to mitochondrial inheritance issues and biases from mitochondria-specific evolutionary forces (RUBINOFF *et al.* 2006; DESALLE *et al.* 2017). Nuclear markers like EF-1α, 28S, and ribosomal ITS regions have also been used as a DNA barcoding method (ELBOGEN 2012; ARSTINGSTALL *et al.* 2021) but these can lack phylogenetic resolution or show within-genome variation, especially in insects (RUBINOFF *et al.* 2006; DIETZ *et al.* 2022). Because of these limitations, researchers have increasingly turned to genome-scale approaches. Among these, ultraconserved elements (UCEs) have emerged as a particularly powerful and cost-effective tool, especially for non-model taxa (Figure 2-1). Despite their powerful resolution, these genomic tools are still limited by the lack of reference genomes overall, with many genera remaining unsequenced, making it difficult to build comprehensive phylogenies.

Ultraconserved elements (UCEs) are short genomic regions, usually larger than 200bp (CUMMINS *et al.* 2024), that are highly conserved among taxa. While most UCEs are found in non-coding regions, some have been shown to overlap with protein-coding regions, as seen in humans (BEJERANO *et al.* 2004; SNETKOVA *et al.* 2022). These regions have few repeats and, when captured along with adjacent DNA, can be used for phylogenetic reconstruction. Tree-building algorithms use UCE data to reconstruct evolutionary histories, accounting for the fact that DNA variability increases with distance from the UCE (ZHANG *et al.* 2019). This approach is relatively inexpensive, enabling researchers to sequence over 1000 target sequences containing exons and some intronic elements for analysis (VAN DAM *et al.* 2021). Regions of high GC content can complicate phylogenetic reconstruction because they are often affected by GC-biased gene conversion (gBGC) following recombination events (BEYE *et al.* 2006; KENT *et al.* 2012; WALLBERG *et al.* 2015; BOSSERT *et al.* 2017). Although the average genomic GC content in honey bees is relatively low (~33%) (WALLBERG *et al.*

2019), in highly eusocial species like honey bees, which exhibit very high recombination rates, this process may increase GC content in specific regions of the genome even if the overall GC content remains low. These local increases in GC can influence substitution patterns and can complicate the generation of congruent gene trees, especially when gene divergence rates vary across species, a phenomenon known as incomplete lineage sorting (Freitas *et al.* 2021). To overcome these issues, researchers have increasingly been focusing on UCEs. UCEs are enriched in AT-content, which makes them less susceptible to the effects of gBGC in GC-rich regions. As a result, UCEs provide more consistent substitution patterns across taxa and improve the accuracy and congruence of both gene and species trees (Bossert *et al.* 2017).

UCEs have been widely applied to construct phylogenies at many taxonomic levels of the bee tree of life and have led to adjustments in the phylogenetic organization and divergence times for many of the seven families within Anthophila. For example, UCEs have uncovered an increase in diversification rates among Andrenidae over the past 15 million years, particularly within the genera *Andrena* and *Perdita*, representing some of the fastest diversifying bees (Bossert *et al.* 2021). UCEs have also helped refine the phylogeny of the Halictidae, the second-largest bee family, by revealing that the Middle Eastern *Clavinomia* is closely related to the North American *Dieunomia*, leading to the inclusion of *C. clavicornis* in the tribe Dieunomiini (Bossert *et al.* 2024). Additionally, UCEs confirmed that *Homalictus* is a subgenus within *Lasioglossum*, while reinstating *Rostrohalictus* as a subgenus for *Lasioglossum longirostre* (Zhang *et al.* 2022). In Megachilidae, UCEs have helped date the origin of the mason bee subgenus *Osmia* to the late Miocene in the West Nearctic, around 5.5-4.8 Mya, before its dispersal from the East Palearctic (Branstetter *et al.* 2021). Collectively, these studies highlight the role of UCEs in establishing accurate timelines and taxonomic classifications that were previously difficult to determine with confidence, especially among non-model organisms.

Phylogenomic studies using UCEs have greatly advanced our understanding of bee evolution within the Apidae, which exhibit diverse ecological and behavioural traits (Orr *et al.* 2022). One of the most influential studies reorganized the subfamilial relationships within Apidae (Bossert *et al.* 2019). This study confirmed the monophyly of Centridini and showed that they belong with the corbiculate bees among the Apinae. It also clarified the phylogenetic positions of the oil-collecting genera *Ctenoplectra* and *Tetrapedia*, showing them as sister groups and belonging to the Xylocopinae, and suggesting shared life history traits such as large egg sizes and co-evolutionary patterns with phoretic mites (Bossert *et al.* 2019). These insights shed light on the evolution of specialized pollen-collecting structures. The study also resolved the long-standing debate over cleptoparasitism in Euglossini, supporting a single origin of the trait rather than two independent occurrences (Bossert *et al.* 2019).

UCE-based studies have made significant contributions to understanding the subfamily Eucerinae, which includes the long-horned bees – a difficult group to categorize due to their similar external morphology (Freitas *et al.* 2023). For instance, an early study using UCEs identified three main clades within Eucerinae (Freitas *et al.* 2021), while a follow-up study expanded this to seven clades and identified four new subtribes: Alloscirteticina, Gaesischiina, Thygaterina, and Melissodina (Freitas *et al.* 2023). The later study also estimated that the diversification of Eucerini began around 20 million years ago, suggesting the group likely originated in southern South America and the Andean region before dispersing northward (Freitas *et al.* 2023). Similarly, UCEs have played a key role in clarifying the origins of brood parasitism within the cleptoparasitic subfamily Nomadinae. A study using UCEs across 114 species confirmed the monophyly of the Nomadinae, revealing that all tribes within the subfamily descend from a single parasitic ancestor (Sless *et al.* 2022). This research demonstrated that parasitism initially evolved within hosts from the same family, Apidae, before some lineages shifted to parasitizing Andrenidae, among other families (Sless *et al.* 2022). Another study using 2,545 UCEs provided the first global molecular phylogeny of the genus *Nomada*, confirming its monophyly and dating its divergence from other Nomadinae to the late Cretaceous period in the Nearctic (Odanaka *et al.* 2022). Overall, UCE-based phylogenomic studies have revolutionized our understanding of bee evolution within Apidae. By providing a consistent framework for classification and resolving longstanding taxonomic challenges, these studies have uncovered the complex evolutionary history and ecological dynamics of bee subfamilies.

Understanding the phylogeny of tribes within the subfamily Apinae is important as it has implications for understanding the evolution of eusocial behaviour. Several previous studies based on morphology, behaviour, or earlier molecular approaches suggested that Euglossini diverged first within Apinae, followed by the primitively eusocial Bombini with the advanced eusocial Meliponini and Apini as sister groups where eusociality elaborated, thus suggesting a single origin of eusociality and a single origin of advances eusociality in this subfamily (Cameron 1993; Mardulyn and Cameron 1999; Cardinal and Packer 2007; Romiguier *et al.* 2016). However, a UCE-based study provided evidence for a different topology within the Apinae subfamily that has been supported by previous morphological and behavioural studies (Woodard *et al.* 2011; Hedtke *et al.* 2013), which proposes Euglossini as the basal group, followed by Meliponini and Bombini as sisters to Apini (Bossert *et al.* 2017). By rearranging the placement of these groups, we now must consider that there are dual origins of advanced eusociality within corbiculate bees.

2. Alternative splicing, regulatory RNAs and the molecular basis of eusocial evolution

Eusociality, characterized by overlapping generations, reproductive division of labour, and cooperative brood care, is a major evolutionary transition that has independently evolved multiple times in bees (Wilson 1971). Understanding the evolutionary and mechanistic perspectives of this transition remains a central focus of sociobiology, where many recent studies have focused on identifying the regulatory mechanisms underlying this behaviour. For example, epigenetic mechanisms, which are stable alterations to gene expression that do not change the DNA itself that arise during development or due to environmental changes, are known to be important for eusocial evolution (Jaenisch and Bird 2003). These mechanisms play an important role in regulating gene expression, thus facilitating behavioural adaptations essential for social evolution, as existing genes can be reversibly modified to adjust behaviour (Kribelbauer *et al.* 2020). Much of the work on epigenetics in social insects has implicated DNA methylation (Herb 2014) and acetylation (Spannhoff *et al.* 2011) as important for social behaviours, including regulating the division of labour (Lyko *et al.* 2010; Foret *et al.* 2012; Araujo and Arias 2021), and maternal care (Rehan *et al.* 2016; Arsenault *et al.* 2018). However, confirming these findings has been challenging, often yielding contradictory results (Kapheim *et al.* 2015; Libbrecht and Keller 2015; Standage *et al.* 2016). With the advancement of genome sequencing technologies, it has become easier to confidently identify the role that other genomic and transcriptomic components play in regulating behaviour, such as alternative splicing and non-coding RNA (ncRNA), which include long non-coding RNA (lncRNA), microRNA (miRNA), and circular RNA (circRNA; Figure 2-2). In the following sections, we outline the contribution of genomic studies in identifying RNA or regulatory elements that have been implicated in social phenotypes or social evolution of bees. We focus on studies that incorporate a diverse set of bees displaying varying levels of social complexity that help uncover the underlying patterns linked to social behaviour at the genomic level.

2.1 Alternative gene isoforms

Alternative splicing, a conserved, post transcriptional regulatory process, is important for generating phenotypic diversity by producing different isoforms (Kannan *et al.* 2019; Wright *et al.* 2022). This flexibility provides raw material for natural selection to act upon, contributing to the evolutionary diversification of traits, especially in taxa like bees that have evolved to have complex social structures. Alternative splicing involves several mechanisms, including exon skipping, alternative 5’ or 3’ splice sites, which alter exon length by splicing before or after the exon at the 5’ or 3’ end; intron retention, where the

intron is not spliced out; and alternative first or last exons, where the first or final exon of a transcript is either retained or spliced out (Wright *et al.* 2022). By employing these mechanisms in different combinations, a diverse array of mRNA products can emerge from a single genomic template, thereby introducing numerous avenues for phenotypic variation (Bush *et al.* 2017). Early investigations into alternative splicing in honey bees were pivotal in elucidating its role in sex determination among bees. In honey bees, the splicing cascade is triggered by the complementary sex determiner (CSD) that is a mechanism by which heterozygosity at the *csd* locus will lead to development of females, and hemizygosity or homozygosity will result in males (Otte *et al.* 2023). One of the important genes in this pathway is the feminizer gene (*fem*), which influences the sex-specific splicing of doublesex (*dsx*) and consequently affects male or female development (Gempe *et al.* 2009). One recent transcriptomic study highlighted the importance of alternative splicing in sex determination and found that embryos at various developmental stages had an increasing number of alternatively spliced genes specific to sex as development progressed (Netschitailo *et al.* 2022). Patterns of alternative splicing have been found to greatly differ between castes in *Apis mellifera*, with more isoforms produced among the worker phenotype (Zheng *et al.* 2023). These caste-specific splicing patterns likely play a role in the evolution of eusociality by allowing distinct behavioural and physiological roles to emerge from a shared genetic background. While alternative splicing provides a flexible means of generating transcriptomic diversity from a static genome, it operates alongside epigenetic mechanisms such as DNA methylation and histone modification, which also play a significant role in shaping caste-specific gene expression (Kucharski *et al.* 2008; Foret *et al.* 2012; Wojciechowski *et al.* 2018). These epigenetic modifications can influence splicing decisions as well, creating a complex regulatory network that integrates multiple layers of gene regulation to produce the distinct phenotypes observed in eusocial bees. Follow-up studies have shown that many alternatively spliced genes are often involved in processes important for social behaviour, such as communication (Liu *et al.* 2023). Similarly, in *Bombus terrestris* colonies, alternative splicing has been implicated in differences in reproductive workers and queens versus non-reproductive workers and males with around 40% of the expressed genes exhibiting multiple isoforms, some of which have been linked to genes involved in the ecdysteroid pathway, which influences various behaviours (Price *et al.* 2018). This suggests that alternative gene splicing has been co-opted independently across bee lineages to facilitate reproductive division of labour, which is essential for more advanced forms of sociality in bees. Alternative splicing also has implications for foraging behaviour, as seen in the *elavl2* gene, an ortholog of the conserved ELAV/Hu family of RNA-binding protein genes (Samson 2008; Ustaoglu *et al.* 2021). This gene can produce 40 isoforms, two of which are involved in learning and memory consolidation, which are crucial components of foraging behaviour in worker honey

bees (Ustaoglu *et al.* 2021). Overall, alternative splicing appears to be an important process for generating phenotypic plasticity, especially in social insects where different castes often arise from the same genome. This plasticity not only underpins current social organization but likely contributed to major evolutionary transitions in social complexity across bee lineages. The diverse mechanisms of alternative splicing produce a wide array of mRNA isoforms, revealing complex regulatory networks that underpin social insect biology. Future research should aim to uncover further connections between alternative splicing and social behaviours, such as identifying behaviour-specific or caste-specific isoforms or uncovering regulatory isoforms that influence key aspects of social behaviour.

2.2 Long non-coding RNA

Long non-coding RNAs (lncRNA) are a class of non-coding RNA transcripts exceeding 200 nucleotides long that lack protein-coding capacity (Kapranov *et al.* 2007). They also play diverse roles in gene regulation through their ability to alter chromatin structure which affects both local and distant transcription (Statello *et al.* 2021). They are transcribed from promoters, enhancers or intergenic regions (Fang and Fullwood 2016) and have a 5' cap (Tamtaji *et al.* 2021) and polyadenylated tail (Carninci *et al.* 2006). These transcripts can localize to both the nucleus (Derrien *et al.* 2012) and cytoplasm (Rashid *et al.* 2016), and are involved in mRNA stability, translation regulation, and microRNA (miRNA) biogenesis, or interference in a tissue-specific manner (Ingolia *et al.* 2011; Jayakodi *et al.* 2015; Montes *et al.* 2015; Paraskevopoulou and Hatzigeorgiou 2016; Noh *et al.* 2018). In bees, over 21,000 lncRNAs have been identified (Zheng *et al.* 2023), and several studies have highlighted their connection to various developmental processes and social behaviours. These findings suggest that lncRNAs may have contributed to the evolution of eusociality in bees enabling fine-tuning regulation of traits that are essential for caste differentiation and task specialization. For instance, lncRNAs such as *lncov1* and *lncov2* and others map to a QTL that has been previously linked to ovary size (Humann *et al.* 2013) and oviposition (Chen *et al.* 2017), brain function in drones (Sawata *et al.* 2002), communication through the waggle dance (Feng *et al.* 2022), and are linked to the transition from nurse to forager roles by modulating hormone levels (Amdam *et al.* 2003; Drapeau *et al.* 2006; Tadano *et al.* 2009; Liu *et al.* 2019). A specific lncRNA, *kakusei* in honey bees, has been associated with multiple behaviours and processes important for sociality including foraging behaviour (Singh *et al.* 2020) and learning and memory (Kiya *et al.* 2007). Its homolog, *Acks*, was found to be involved in temperature sensing and thus important for hot defensive bee ball activity in fighting giant hornets (Kiya *et al.* 2012).

Several studies have associated various lncRNAs with immune responses in different honey bee species and subspecies, aiding our understanding of the consequences of disease transmission within a social colony. For example, long intergenic non-coding RNAs (lincRNAs) were found to be upregulated in bees infected with deformed wing virus (DWV) or Sacbrood virus (SBV) in *Apis cerana* (Jayakodi *et al.* 2015). Additionally, differentially expressed lncRNAs (DElncRNAs) were identified in both *Apis cerana* (Guo *et al.* 2023) and *Apis mellifera ligustica* (Ye *et al.* 2022) following infection with the chalkbrood disease-causing fungal pathogen *Ascosphaera apis*, or following *Nosema ceranae* infection in *A. cerana* (Wang *et al.* 2023). The lncRNAs identified in these studies were involved in the regulation immune-related gene expression pathways, suggesting their role in the defense mechanisms of social bee colonies (Guo *et al.* 2023; Wang *et al.* 2023). Furthermore, exposure to insecticides like the neonicotinoid Dinotefuran can induce lncRNA responses, further highlighting their involvement in the response to environmental stressors through immune system activation (Huang *et al.* 2021). Overall, lncRNAs in bees serve multifaceted roles in development, behaviour, and immune function, shedding light on the intricate regulatory mechanisms within social insect colonies. Their involvement in caste-related traits, behavioural transitions, and immune responses suggests they may have played an important role in the evolutionary origins and elaboration of eusociality across bee lineages.

2.3 MicroRNA

MicroRNAs, short sequences of 18-24 nucleotides, are integral components of the gene silencing program within the RNA-induced silencing complex (RISC), enabling precise and targeted silencing of mRNA (Sovik *et al.* 2015; Bartel 2018). The advancement of genome sequencing technologies has led to the development of miRbase (www.mirbase.org), a comprehensive database encompassing microRNA profiles and annotations across 271 organisms, facilitating comparative analyses to identify both conserved and novel microRNAs (Kozomara *et al.* 2019). Early studies have highlighted the importance of microRNAs in regulating the timing and spatial expression of specific genes during development, suggesting their role in constraining the plasticity of certain genes by regulating their expression patterns (Peterson *et al.* 2009). This precise regulatory control may have been instrumental in enabling caste differentiation from a shared genome, supporting the evolution of eusociality in bees. However, the extent to which this regulation directly drives specific developmental outcomes remains an area of ongoing investigation, as many of the studies are correlational. Recent studies have revealed that, while microRNAs themselves might not be the primary drivers of eusocial genome evolution, their coevolution with target sequences significantly influences various social behaviours (Kapheim *et al.* 2020b). In a study by Vieira et al. (2021), two microRNAs,

miR-34 and *miR-317*, were found to be upregulated in queens relative to workers. These microRNAs downregulate the gene *Takeout*, which leads to a reduction in juvenile hormone levels. This hormonal shift is thought to promote neurogenesis, the process by which new neurons are formed in the brain (VIEIRA *et al.* 2021).

MicroRNAs are also suggested to be involved in caste-specific development processes. Distinct expression patterns of microRNAs have been observed in queen and worker pupae (WEAVER *et al.* 2007) and larvae (SHI *et al.* 2014; ASHBY *et al.* 2016) of *Apis mellifera*, as well as in the larvae of *Bombus terrestris* (COLLINS *et al.* 2017). Exogenous microRNAs found in bee food, with worker jelly having a higher abundance, may also play a role in caste development as seen with *ame-miR-184* that was identified as important for regulating caste morphology (GUO *et al.* 2013). Additionally, a study found similar patterns of microRNA expression among queens and workers with activated ovaries compared to virgin queens and workers with inactivated ovaries (MACEDO *et al.* 2016). When queen ovaries become active, microRNAs, such as *miR-14*, play a crucial role in regulating egg-laying as its inhibition resulted in higher egg production through targeting the mRNA of the ecdysone receptor in honey bees (CHEN AND FU 2021).

MicroRNAs exhibit nuanced regulation by partially suppressing gene expression to prevent overexpression (POSADAS AND CARTHEW 2014). Several microRNAs are specifically involved in the regulation of royal jelly-related processes. For instance, there is a positive correlation among microRNAs targeting genes upregulated in response to royal jelly ingestion, suggesting their presence to curtail the overexpression of these genes and prevent deleterious effects (GUO *et al.* 2013). Additionally, specific microRNAs have been linked to the production of royal jelly by nurse bees, while others are associated with honey processing in foragers, indicating their role in regulating glandular functions associated with different tasks (SHI *et al.* 2021). Beyond royal jelly regulation, microRNAs also play critical roles in worker polyethism within social insect societies. Changes in miRNA expression have been observed in the brains of foraging workers compared to hive workers (LIU *et al.* 2012). For example, one study found specific brain microRNAs associated with young nurses and others upregulated in older foragers, highlighting their importance in age-related behavioural changes (BEHURA AND WHITFIELD 2010). Notably, the most abundant microRNA found in brains of honey bees, *ame-miR-2796*, is known to be important for regulating forager behaviour and was found to be conserved among several social insects (GREENBERG *et al.* 2012). Additionally, microRNAs exhibit differential expression among nurses, foragers, and newly emerged bees within the hypopharyngeal glands, which undergo alterations during development, implicating these microRNAs in developmental processes (SHI *et al.* 2021). Taken together, the diverse functions and caste-specific

expression patterns of microRNAs support their role in the evolution of social complexity, providing regulatory flexibility that enables phenotypic divergence without requiring large-scale genomic changes.

2.4 Circular RNA

Circular RNAs (circRNAs) have garnered significant attention in recent years due to their unique structure and regulatory functions. Characterized by a closed-loop structure, circRNAs possess increased stability and resistance to RNAse degradation compared to linear RNAs, making them attractive candidates as stable biomarkers capable of accumulating in biological systems (Zhou *et al.* 2020). Initially considered as splicing by-products, circRNAs have emerged as important regulators of gene expression, particularly in post-transcriptional regulation, including as competing endogenous RNA (ceRNA), which are microRNAs that regulate other types of RNAs and modulates mRNA availability (Hansen *et al.* 2013; Memczak *et al.* 2013; Salzman *et al.* 2013; Ashwal-Fluss *et al.* 2014; Ala 2020). These circRNAs can act as sponges and bind to other microRNAs or RNA binding proteins in an effort to block their effects (Panda 2018). CircRNAs play diverse roles in various aspects of bee biology, including social behaviour, task specialization (Tholken *et al.* 2019), reproductive division of labour (Chen *et al.* 2019), and development (Chen *et al.* 2020). For instance, Tholken et al. (2019) identified correlations between circRNAs, DNA methylation, and task allocation in bees. They found circRNAs enriched for annotations related to memory formation and neuromuscular synaptic transmission regulation, which are crucial for tasks such as foraging and hive maintenance (Tholken *et al.* 2019). Furthermore, a specific circRNA, *ame_circ_0001780* (*circAmrad*), was correlated with the transition from nurse to forager bees, and another *circAmrsmep2* was found to increase with worker age, indicating their involvement in behavioural plasticity and response to environmental cues (Tholken *et al.* 2019). These findings suggest that circRNAs may have contributed to the evolution of eusociality by enabling flexible gene regulation necessary for task specialization and the emergence of division of labour among castes. CircRNAs have also been implicated in reproductive processes in bees. Chen et al. (2019) identified the circRNA *ame_circ_0010349* as upregulated in mating queens, suggesting its role in ovary activation and reproductive division of labour. Furthermore, Chen and Shi (2020) identified circRNAs differentially expressed among workers and queens, suggesting their involvement in the reproductive division of labour by potentially interacting with microRNAs in ecdysone and vitellogenin pathways. Such caste-specific regulatory interactions point to circRNAs as potential contributors to the evolutionary stabilization of queen and worker roles within a eusocial system.

Additionally, Chen *et al.* (2023) identified a circular RNA, *ame_circ_0002015*, that influences ovary activation in *Apis mellifera*. Experimental manipulation revealed a positive correlation between its

expression and the number of eggs laid by queens, linking circRNA activity to reproductive function, which is an essential caste-specific trait (Chen *et al.* 2023). Interestingly, *ame_circ_0002015* was also found to be regulated by *miR-14-3p*, highlighting a complex regulatory network between circRNAs and microRNAs that may fine-tune oviposition (Chen *et al.* 2023). Beyond reproduction, circRNAs also play stage-specific roles in larval development in bees. Chen et al. (2020) identified circRNAs participating in midgut development, with differential expression patterns suggesting stage-specific roles in midgut maturation. These findings provide insights into the molecular mechanisms underlying organ development and tissue specialization in bees. Recently, Zhang et al. (2023) conducted a study focusing on circRNAs within the gut of developing larvae of *Apis mellifera ligustica*. They found significant differential expression among larvae of different ages, where circRNAs that were differentially expressed were involved in growth, development, regulation, and metabolism (Zhang *et al.* 2023). Notably, they identified a new circRNAs, *Novel_circ_000838*, that interacts with the target microRNA miR-6000a-3p, providing evidence for a ceRNA networks involved in regulating microRNAs during larval development (Zhang *et al.* 2023). These findings underscore the role of circRNAs in regulating key developmental and reproductive processes. Since such traits are central to caste differentiation in eusocial insects, dynamic RNA-based regulatory networks, especially involving circRNAs and microRNAs, may have played a pivotal role in the evolution of social complexity in bees. By modulating gene expression without altering coding sequences, circRNAs provide a flexible mechanism through which developmental pathways, and thus caste-specific traits like morphology, physiology, and behaviour, can evolve and diversify.

Circular RNAs have also emerged as critical regulators of immunity in bees. For example, one study examined the effects of the neonicotinoid insecticide Dinotefuran on circRNA expression in honey bees, revealing an increase in differentially expressed circRNAs in honey bee brains over time, particularly those targeting genes involved in immune pathways (Huang *et al.* 2022a). Studies of *Apis mellifera ligustica* exposed to either *Ascosphaera apis* (Ye *et al.* 2022) or *Vairimorpha ceranae* (Chen *et al.* 2022) have highlighted the involvement of circRNAs in immune-related pathways at different ages and stages of infection. In a study on *Apis cerana* infected with *Nosema ceranae*, researchers identified novel circular RNAs implicated in immunity, along with ceRNAs that sequester microRNAs associated with various aspects of the immune pathway (Zhu *et al.* 2022). Notably, they discovered that one of the newly identified circRNA genes was found in the Hippo signaling pathway, which is known to be involved in regulating cell proliferation and growth and the regeneration of organs (Tang *et al.* 2022). This finding suggests that differentially expressed circRNAs can act as markers of damage in the midgut due to *Nosema* infection. Consequently, the host (worker) may utilize these circRNAs to regulate transcription, facilitating cellular

renewal and modulate their immune response (Zhu *et al.* 2022). As circRNA-mediated regulation can influence how bees respond to pathogens and environmental stressors, these mechanisms may play a role in shaping adaptive responses over time. Understanding the roles of circRNAs in bees not only provides insights into their complex biology but also offers opportunities for developing strategies to mitigate threats to bee health and conservation. Taken together, the diverse functions of circRNAs in development, behaviour, and immunity suggest they are not only important for individual plasticity and defense but also represent molecular innovations that may have facilitated the emergence and refinement of caste-specific traits, which are key steps in the evolution of eusocial complexity in bees.

3. Comparative genomic studies uncover features of evolution

Comparative genomic studies offer a robust framework for understanding the evolution of eusociality by examining genomic differences among species with varying social behaviour. For instance, a study using whole genome sequencing on *Lasiogossum albipes*, a species with both social and solitary populations, identified single nucleotide polymorphisms and genes expression patterns that are associated with variation in social behaviour (Kocher *et al.* 2018). They identified the gene *syntaxin 1a*, a gene involved in neurotransmission, as highly genetically differentiated between social and solitary *L. albipes* and showed behaviour-specific gene expression (Kocher *et al.* 2018). Analyses within conspecific populations exhibiting polymorphic behaviour provide inherently controlled comparisons, which reveal aspects of genome evolution related to social transitions. For example, comparative studies have identified gene regulation to be particularly important for the evolution of eusociality, as large-scale studies across multiple species have identified it as a key mechanism underling the evolution of social behaviour (Wittwer *et al.* 2017; Kocher *et al.* 2018; Rehan *et al.* 2018; Kapheim *et al.* 2020a). In a study analyzing 10 bee species with diverse origins of sociality, Kapheim et al. (2015) found that social behaviour correlates with enhanced gene regulatory capacity, evidenced by increased activity of transcription factors and a higher number of transcription factor binding sites. Similarly, studies on conserved Non-Coding Alignable Regions (NCARs) of DNA among 11 species from various origins and levels of sociality suggested that gene regulation is important for eusocial genome evolution (Rubin *et al.* 2019). Specifically, these regulatory changes are associated with the evolutionary rate shifts in non-coding regions that are linked to social behaviour and divisions of labour. Furthermore, a recent study comparing 17 different species to identify genomic features associated with the transition between solitary and social behaviour found that in lineages where sociality was lost, there were relaxed selective pressures on taxonomically restricted genes and gene regulatory elements

suggesting that the transition to social behaviour requires regulation of gene expression and selection of genes that promote enhanced regulation (JONES *et al.* 2023).

Through comparative genomics studies, researchers have gained insights into the distinct patterns of adaptive molecular evolution among various bee species, reflecting their diverse social behaviours. For instance, in a comparative analysis between three *Bombus* species and two *Apis* species, Harpur et al. (2017) revealed signs of adaptive divergence with evidence of strong positive selection found among genes associated with cognition and sensory perception in honey bees and in metabolic genes in bumble bees. Moreover, when comparing the selection coefficients of genes upregulated in queen relative to workers (queen-biased genes) and genes upregulated in workers relative to queens (worker-biased genes) in each species, the study found higher selection for queens-biased genes in bumblebees, while the opposite trend was observed in *Apis* species, suggesting distinct mechanisms of adaptive evolution (HARPUR *et al.* 2017). A comparative genomics study investigated adaptive evolution in honey bees, bumblebees, and paper wasps, hypothesizing parallel patterns of adaptive evolution among species with similar social structures (DOGANTZIS *et al.* 2018). Consistent with previous research, it found higher adaptive evolution in genes biased towards reproductive individuals in bumblebees and paper wasps compared to worker-biased genes among honey bees (DOGANTZIS *et al.* 2018). The study also observed a greater overlap of positively selected genes between primitively eusocial species that exhibit cooperative brood care and reproductive division of labour but without strong morphological differences between castes (bumblebees and paper wasps), relative to advanced eusocial honey bees, suggesting convergent adaptive molecular evolution (DOGANTZIS *et al.* 2018). Another study leveraged comparative genomics and analyzed several bumblebee species with distinct behaviour including the non-social cuckoo bumblebee (subgenus *Psithyrus*) that act as social parasites (FOUKS AND LATTORFF 2016). This study found varying patterns of selection in targeted genes known to be involved in social behaviour as well as their regulation among species exhibiting social versus parasitic behaviour, highlighting the utility of social parasites in studying mechanisms of selection within a comparative genomic framework (FOUKS AND LATTORFF 2016). Overall, comparative genomic studies illuminate the dynamic interplay between genomics and behaviour, offering a comprehensive understanding of the intricate mechanisms driving the emergence and diversification of sociality in bees.

4. Ecology

4.1 Museomics

Museum collections are a valuable source of historic DNA (Bi *et al.* 2013), whose sequence information can facilitate genomic investigations of adaptation across a wide variety of insects including bees (Cavill *et al.* 2022; Cilia *et al.* 2022), moths (Call *et al.* 2021), locusts (Goodman *et al.* 2023), beetles (Toussaint *et al.* 2021; Cohen *et al.* 2022), butterflies (Grewe *et al.* 2021), ants (Eguchi *et al.* 2020), flies (Shpak *et al.* 2023; Strunov *et al.* 2023), and others (Schmitt *et al.* 2018; Derkarabetian *et al.* 2019; Korlevic *et al.* 2021). While invaluable for conservation genomics, these data typically reflect very recent evolutionary trajectories, on the order of fewer than 100–200 generations and therefore offer unique insight into contemporary responses to anthropogenic pressures. Beyond genetic material, pollen can be extracted from both modern and preserved specimens to identify the changes in floral resource composition, assessing landscape changes, and analyzing plant-pollinator interactions over longer timeframes using tools such as pollen metabarcoding (Bell *et al.* 2023; Wizenberg *et al.* 2023a; Wizenberg *et al.* 2023b). Overall, museum samples play an important role in uncovering patterns of genetic adaptation and ecological responses to environmental change across insect taxa.

Studies of wild bee species have highlighted the power of museum collections for tracking evolutionary and ecological shifts. For example, by comparing the genomes of the small Carpenter bees *Ceratina calcarata* and *Ceratina dupla* from contemporary populations to those sampled 50 years prior, Brasil et al. (2022) identified a reduction in genetic diversity and effective population size in this relatively short time frame. These species also show population genetic evidence that proposes a potential adaptation to climate change and insecticide use (Brasil *et al.* 2022), although experimental confirmation of this has yet to be demonstrated. Similarly, a study in *Ceratina* found that with agricultural expansion, male body size decreased while female body size increased, suggesting distinct selective pressures (Kelemen and Rehan 2020; Pope *et al.* 2023). Additionally, studies have highlighted decreases in genetic diversity in at-risk species like *Bombus occidentalis*, suggesting reduced genetic resilience to environmental changes or disease resistance (Rohde *et al.* 2024). Pollen metabarcoding has expanded this type of work as seen in a study conducted in Cuxhaven, Germany, that analyzed pollen collected from contemporary bumblebees and compared them with historic specimens from 1968-1969, revealing notable changes in foraging behaviour, pollination networks and community composition (Kolter *et al.* 2023). Similarly, Simanonok et al. (2021) applied pollen metabarcoding to modern and museum samples of the endangered rusty patch bumblebee (*Bombus affinis*) spanning a century (1913-2013). Their findings indicated no significant

differences in floral resource usage over time, suggesting that regional variations, rather than floral availability, may be unrelated to the species' decline (Simanonok *et al.* 2021). In another study, Gous et al. (2021) compared pollen metabarcoding data from modern leaf-cutter bees (Megachilidae) in South Africa with museum specimens and discovered that bees with a wider foraging range exhibited a broader spectrum of floral resource use, but noted no significant temporal differences (Gous *et al.* 2021). Taken together, this demonstrates how historical DNA and pollen samples from wild bee specimens can shed light on recent demographic, adaptive or ecological dynamics.

In *Apis mellifera*, museum-based studies have offered detailed insights into population history and management. A study of California honey bees spanning 105 years revealed genetic changes over time, including the introgression of African ancestry and alleles into managed populations (Cridland *et al.* 2018). In a study of Swiss honey bee museum specimens (1879-1959), Parejo *et al.* (2020) found higher genetic diversity among modern samples compared to historical, museum samples, suggesting a lack of population bottleneck potentially due to beekeeping practices. However, this analysis did not account for the strong population structure and lineage turnover in European honey bees (e.g., from the native M lineage to C or admixed individuals). More recent studies that incorporate lineage distinctions, such as Espregueira Themudo et al. (2020) who also analyzed museum specimens and Leroy et al. (2024) who inferred demographic history from present-day linkage disequilibrium patterns, have instead reported signatures of population bottlenecks. These findings highlight the complexity of honey bee demographic history and the importance of considering population structure in such analyses.

Museomics, however, remains subject to various forms of bias that limit the reliability of its outcomes. For instance, researchers can only compare results against what has been preserved, leading to observation bias—comparisons are inherently limited to the available reference data (Shaffer *et al.* 2025). Additionally, DNA amplification methods, including the choice of Taq polymerase and the composition of primers, can introduce amplification bias (Silverman *et al.* 2021; van der Loos and Nijland 2021). Since species are typically analyzed as parts of a whole community, under-detection of one species will, by definition, inflate the apparent abundance of others within that group (McLaren *et al.* 2019). Future studies should prioritize minimizing these sources of bias by expanding reference databases through additional taxonomic sampling efforts to reduce observation bias. By improving the design of primers to either be universal or taxon-specific, researchers can overcome the limitation of poorly designed primers that may be binding to multiple species and increasing the abundance of one species artificially, thus limiting amplification bias. Another suggestion is to use a reference species to calibrate the metabarcoding data and thus be able to use absolute DNA quantities instead of relative proportions, although this method is

dependent on exact primer matches (Shaffer *et al.* 2025). Addressing these challenges through methodological improvements and careful calibration will be essential to enhance the accuracy and reliability of metabarcoding as a tool for biodiversity assessment.

4.2 Environmental 'omics

Environmental DNA (eDNA) is DNA that is collected from diverse environmental samples, such as water (Miyata *et al.* 2022), air (Klepke *et al.* 2022; Roger *et al.* 2022; Garrett *et al.* 2023), spiderwebs (Gregoric *et al.* 2022), and even honey (Bovo *et al.* 2018; Bovo *et al.* 2020; Pathiraja *et al.* 2023). eDNA has also been widely used in microbiome analysis within ecological contexts. For more information, we refer readers to several reviews (Ladin *et al.* 2021; Patin and Goodwin 2022; Cook *et al.* 2024), as we focus on how eDNA and metagenomics reveal insights into environmental genomics. eDNA can be used to estimate species richness and construct interaction networks within ecosystems, such as detecting pollinator interactions within the environment without needing direct observation. For example, eDNA analysis can be coupled with advanced sequencing techniques like next-generation sequencing (e.g., Ion Torrent) for shotgun metagenomic analysis of honey eDNA (Bovo *et al.* 2018) or used in conjunction with metabarcoding (Harper *et al.* 2022) to understand the environmental dynamics of bee habitats and their interaction networks with plants (Newton *et al.* 2023), animals (Boardman *et al.* 2023), and pathogens or viruses (Bovo *et al.* 2020; Paluoja *et al.* 2025). These approaches can help reveal how environmental pressures influence bee behaviour, pathogen resistance, and plant-pollinator co-evolution, providing indirect evidence of selective forces shaping bee genetic diversity and adaptive traits over time. For example, long-term eDNA and metabarcoding monitoring such as analyses of bee-collected pollen or environmental samples from herbarium specimens, can track shifts in plant-pollinator community composition and phenology, offering valuable insights into evolutionary responses to climate change, land use, and pathogen dynamics (Waters *et al.* 2025; Wizenberg *et al.* 2025). Additionally, eDNA has demonstrated practical utility for detecting bee-associated pests and pathogens (Ribani *et al.* 2020; Ribani *et al.* 2022) and even rare or cryptic plant species visited by bees (Batchelor *et al.* 2023), offering tools for long-term monitoring of ecological and evolutionary responses. There are several limitations that need to be considered with eDNA methods including that eDNA degrades over time due to temperature, UV light or distance which can limit its detection (Ruppert *et al.* 2019). Furthermore, eDNA is not currently able to accurately detect biomass, which is the aggregate mass of individuals of a certain species or group, based on concentrations of DNA detected (Danziger *et al.* 2022). These collected samples also lack contextual information, such as the age and life-stage of the individuals contributing to the pooled sample, which is an

inherent limitation when gathering DNA shed by the organism (Couton *et al.* 2023). Despite these limitations, eDNA remains a powerful and versatile tool that can complement other molecular techniques, and can serve as an efficient starting point for identifying the presence and diversity of organisms in an environment and guiding more targeted ecological and evolutionary studies

4.3 Pollen Metabarcoding

Pollen metabarcoding provides crucial insights into the complex and often elusive network of bee interactions (Wizenberg *et al.* 2023a; Wizenberg *et al.* 2024). Given the wide variety of plants that bees interact with, pollen metabarcoding offers a window into what plant species bees are visiting at different times of the year, enhancing our understanding of their intricate ecological relationships. Recent advances in wet lab and bioinformatic protocols have demonstrated that pollen metabarcoding can quantify the types and relative amounts of pollen collected by bees as accurately as traditional (i.e. taxonomy-based) methods, while requiring considerably less time and cost (Figure 2-3). A recent study utilizing these new methods identified that honey bees visit a much greater diversity of non-flowering plants than previous anticipated (Wizenberg *et al.* 2023b). For example, genetic material from sporophytes, bryophytes as well as different species of green algae were all detected in honey bee beebread (i.e. stored pollen provisions)(Wizenberg *et al.* 2023b). While the relevance of such visits is not fully understood, it suggests that honey bees and likely other bee species may be important dispersal vectors for non-flowering plants as well as flowering plants (Wizenberg *et al.* 2023b). Analysis of bee bread collected from a very large cohort of honey bee colonies from across Canada discovered that honey bee foraging preferences were strongly associated with dichotomous patterns of stressor exposure (Wizenberg *et al.* 2024). This pattern showed that colonies with low dietary diversity tended to be exposed to harmful agrochemicals, while colonies with high dietary diversity where more likely to be exposed to pathogens (Wizenberg *et al.* 2024). The ability of social bees to 'sample' pollen from a variety of plants over the year makes them particularly suited to monitor changes in plant phenology over time. In a small pilot study in the city of Toronto (Canada), Wizenberg *et al.* (2025) discovered that pollen sampled from honey bee colonies was able to predict the onset of flowering in a large number of plants. This finding opens the door to characterizing changes in plant phenology over time by sequencing pollen collected by bees. Pollen metabarcoding is therefore emerging as a powerful tool to elucidate landscape dynamics and plant-pollinator interactions across different temporal scales.

While these studies show a promising new direction that researchers may take to understand the behaviour of bees, plant flowering, and other ecological relationships among plants and animals, it's important to consider the biases that are still inherently present in this technology. Researchers must often

use a combination of pollen metabarcoding loci such as both ITS2 and *rbcL* to be able to capture different taxonomic groups but even those are limited by the contents of their libraries (ARSTINGSTALL *et al.* 2021). Furthermore, the reference databases are also a limiting factor in the robustness of these studies as they too are often limited to a specific geographic location or may only include certain native plants (MILLA *et al.* 2022; SAN MARTIN *et al.* 2024). Thus, important ecological features are lost among these studies due to a lack of robust libraries and databases. Future studies should therefore aim to expand these barcode libraries so studies can better capture the accurate interactions among plants and animals.

Conclusion

The use of genomic tools, methods and techniques have revolutionized our understanding of bee phylogeny, offering new insights into the evolutionary relationships and diversification of this highly varied group. Advances in next-generation sequencing, the application of ultraconserved elements (UCEs), and phylogenomic approaches have refined long-standing questions about bee evolution, enabling the reconstruction of more accurate gene and species trees. These tools have provided deeper resolution of bee lineages across families and subfamilies, revealing key evolutionary events such as the origins of cleptoparasitism, which is a form of parasitism where an insect will lay their eggs in the nest of another species and will use the hosts resources to provision for their young (CARDINAL *et al.* 2010) and the diversification of major clades like Apinae and Eucerinae. UCE-based studies have clarified taxonomic ambiguities, resolved phylogenetic positions within complex lineages, and reshaped our understanding of the timing and geographic origins of various bee groups. The expanding genomic datasets, including those derived from museum specimens and environmental DNA, are continuing to uncover both historical and contemporary patterns of bee evolution, offering powerful methods for tracking ecological shifts and genetic changes over time. These findings underscore the importance of phylogenomics on the study of bee evolution, with future research poised to explore even broader taxonomic scales and ecological contexts.

We also highlight the growing understanding of the molecular mechanisms that underpin eusocial behaviour in bees, focusing on the roles of alternative splicing, long non-coding RNA (lncRNA), microRNA (miRNA), circular RNA (circRNA), and methylomics in shaping social phenotypes. Through advances in genomics and transcriptomics, researchers have identified key regulatory pathways that influence division of labour, reproductive roles, foraging behaviour, and immune responses within bee colonies. While alternative splicing generates diversity in caste-specific behaviours, lncRNAs and microRNAs contribute to gene regulation that is critical for development, social interaction, and immune function. CircRNAs, with

their stability and regulatory potential, are emerging as important factors in behaviour and immunity. The findings reviewed underscore the complex molecular toolkit that supports social evolution in bees, but they also point to the need for further research to fully understand the interplay between these mechanisms and their roles in different social contexts. This growing body of work not only advances sociobiological theory but also offers potential insights for conservation efforts and understanding environmental impacts on bee populations.

Together, advances in phylogenomics and molecular biology have greatly expanded our understanding of bee evolution and behaviour. By combining tools that help resolve evolutionary relationships with those that uncovered the molecular basis of social traits, we are starting to build a more complete picture of how bees have diversified and adapted to different environments. This growing body of work not only sheds light on the mechanisms behind social evolution and lineage diversification, but also provides useful context for conservation, especially as bees face increasing environmental pressures. As genomic datasets continue to grow and become more accessible, future studies will be well-positioned to link broad evolutionary patterns with fine-scale molecular processes and improve our ability to understand and predict how bees may respond to changing conditions.

Figures

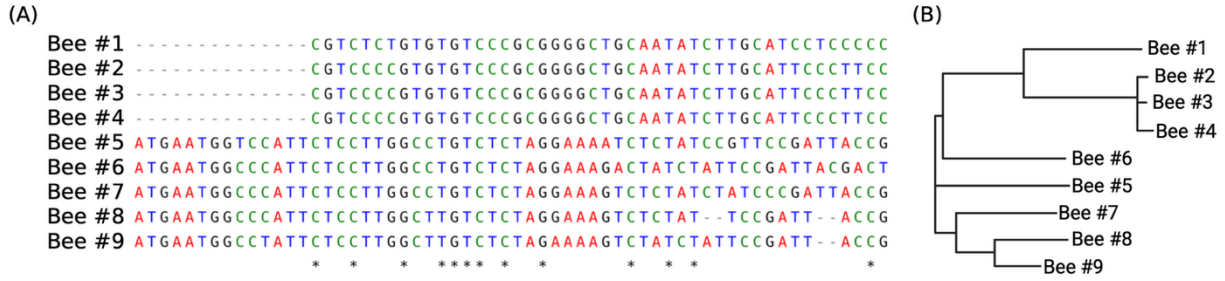


Figure 2-1. *Ultraconserved elements (UCEs) are regions of the genome that are highly conserved across diverse species. Core UCEs are first identified by comparing reference genomes of two or more species to find stretches of conserved DNA. Next, many short bait sequences are generated that span the UCEs. These baits are used to capture matching DNA from species of interest. After hybridization with bait sequences, the DNA fragments that contain the UCEs plus the flanking regions are pulled out, cleaned and sequenced. Following filtering and alignment, genetic variation within UCE sequences can be used for phylogenetic inference. (A) An example of a UCE alignment adapted from* FAIRCLOTH *et al.* (2015)*. The asterisks identify conserved bases in this region. (B) A hypothetical phylogenetic tree generated from the UCE shown in A.*

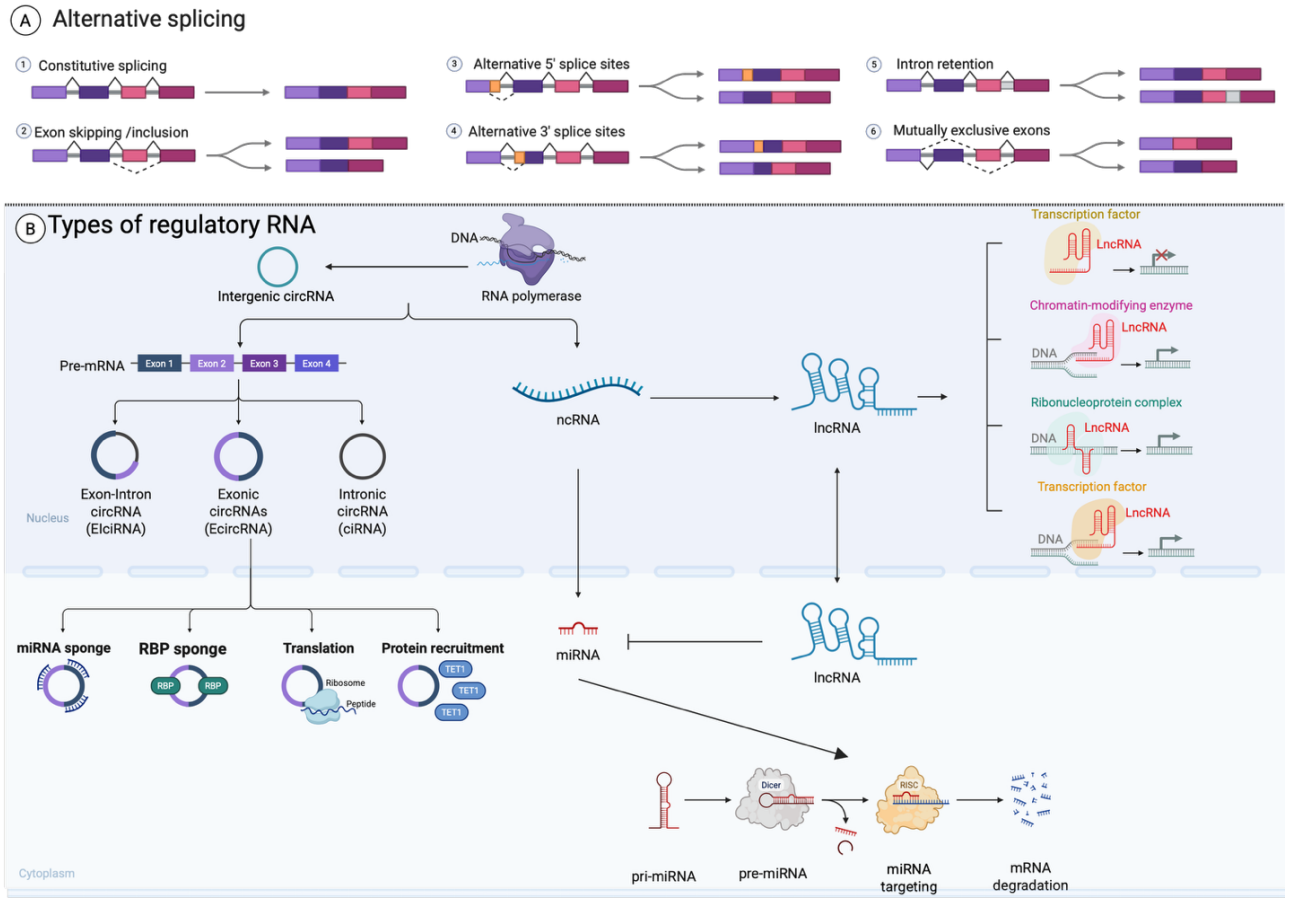


Figure 2-2 *Overview of alternative splicing mechanisms and types of regulatory RNAs that underpin eusocial behaviour in bees. (A) Alternative splicing generates transcript diversity through various mechanisms, including constitutive splicing, exon skipping/inclusion, alternative 5′ or 3′ splice sites, intron retention, and mutually exclusive exons. These events result in different mRNA isoforms from the same gene. (B) Regulatory non-coding RNAs (ncRNAs), including circular RNAs (circRNAs), microRNAs (miRNAs), and long non-coding RNAs (lncRNAs), are transcribed from pre-mRNA or intergenic regions. CircRNAs can act as miRNA sponges, RNA-binding protein (RBP) sponges, or participate in translation and protein recruitment. LncRNAs modulate gene expression through interactions with chromatin, ribonucleoprotein complexes, or transcription factors, or by inhibiting microRNAs. miRNAs are processed from non-coding RNAs and are involved regulating gene expression by binding to target mRNAs in the RISC complex that are then degraded. Each of these mechanisms has been implicated in either regulating bee behaviour or the evolution of bee behaviour.*

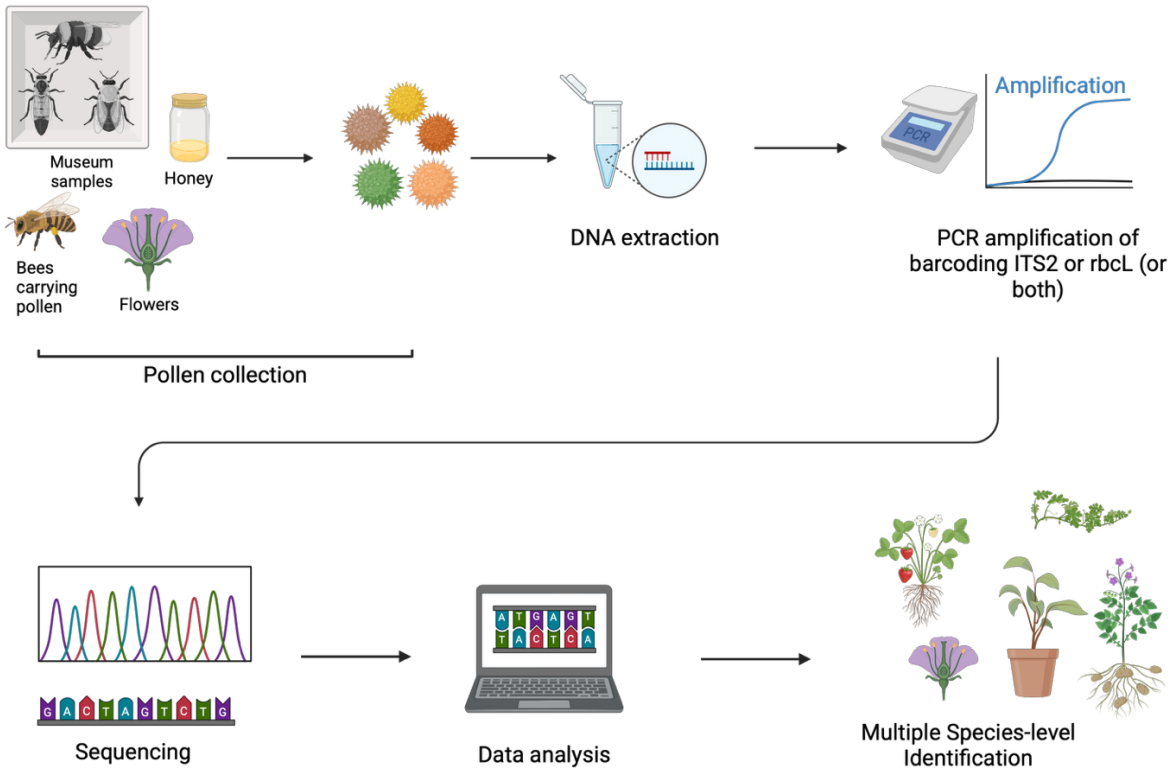


Figure 2-3 *Schematic overview of a pollen metabarcoding workflow for plant identification.*

Pollen is collected from diverse sources such as bees, flowers, museum specimens, or honey. DNA is extracted from mixed pollen samples, followed by PCR amplification targeting standard plant barcode regions (e.g., ITS2 or rbcL*, or both). Amplified products are sequenced and resulting reads are analyzed bioinformatically to identify plant taxa at the species level. This approach enables detailed reconstruction of plant-pollinator interactions and floral resource use*

Chapter 3 An integrative genomic toolkit for studying the genetic, evolutionary and molecular underpinnings of eusociality in insects[2]

Dova Brenman-Suttner[1] and Amro Zayed[1]

Summary

While genomic resources for social insects have vastly increased over the past two decades, we are still far from understanding the genetic and molecular basis of eusociality. Here, we briefly review three scientific advancements that, when integrated, can be highly synergistic for advancing our knowledge of the genetics and evolution of eusocial traits. Population genomics provides a natural way to quantify the strength of natural selection on coding and regulatory sequences, highlighting genes that have undergone adaptive evolution during the evolution or maintenance of eusociality. Genome wide association studies (GWAS) can be used to characterize the complex genetic architecture underlying eusocial traits and identify candidate causal variants. Concurrently, CRISPR/Cas9 enables the precise manipulation of gene function to both validate genotype-phenotype associations and study the molecular biology underlying interesting traits. While each approach has its own advantages and disadvantages, which we discuss herein, we argue that their combination will ultimately help us better understand the genetics and evolution of eusocial behaviour. Specifically, by triangulating across these three different approaches, researchers can directly identify and study loci that have a causal association with key phenotypes and have evidence of positive selection over the relevant timescales associated with the evolution and maintenance of eusociality in insects.

Introduction

Evolution has been punctuated by several major events that fundamentally increased the complexity and biodiversity of life (Szathmary and Maynard-Smith 1995). These evolutionary game-changers include the evolution of sex, multicellularity and eusociality. Eusociality evolved multiple times in animals

[2] Published in Current Opinion in Insect Science (2024) https://doi.org/10.1016/j.cois.2024.101231.

and involves the division of labour between related individuals that cooperate to benefit the group (Wilson and Holldobler 2005). Eusocial animals exhibit collective traits, which are emergent properties from a complex system of hundreds to thousands of individuals interacting with each other. The complex genetics and evolution of eusocial traits are not yet well understood, in part because the typical tools for studying genetics in mostly solitary organisms are not directly applicable to social animals. This lack of knowledge on the genetics of eusocial traits has hindered our ability to develop and test models for the evolution of eusociality. Most of the current theories about the evolution of eusociality are extremely difficult to test because they are either intentionally vague on the genetic architecture of these traits or assume that such traits are controlled by simple Mendelian loci. For example, perhaps the most commonly cited paper on the evolution of sociality – Hamilton's kin-selection theory – assumes a 'single locus' affecting 'helping' traits (Hamilton 1964). A more recent model of the evolution of eusociality by Nowak and colleagues (2010) assumes sociality is controlled by a single recessive allele. While there are certainly examples of variants of a single gene affecting behaviour (Osborne *et al.* 1997), most studies of complex behaviour reveal a large number of loci, often with complex forms of interaction (Anholt and Mackay 2004; Shorter *et al.* 2015). The field of sociobiology is thus stuck in a paradox: evolutionary theories are *de facto* naïve because they assume simple and likely unrealistic genetic models for social traits, while most genetic studies conducted on eusocial insects apply low-resolution techniques that preclude a meaningful understanding of the genetic architecture of eusociality. Here, we provide a brief review of three of the most promising innovations in social insect research that are relevant for understanding the ultimate and proximate causes of eusociality: 1) the use of population genomics to map patterns of adaptive evolution across the genome, 2) the use of genome-wide association studies (GWAS) to identify mutations influencing eusocial traits and 3) CRISPR/Cas9 as a tool to study gene function and test causal hypotheses about genes and phenotypes. We highlight current shortcomings of using these tools in isolation and argue that the combination of all three approaches provides a powerful system for studying the genetics of eusocial traits and how they evolved (Figure 3-1).

Linking DNA to fitness using population genomics

Population genomics broadly describes the application of genome sequencing to large population cohorts. Typically, researchers sequence multiple individuals from different populations or significant evolutionary lineages, allowing them to discover and genotype millions of single nucleotide polymorphisms (SNPs). By utilizing these dense datasets and applying an ever-growing arsenal of population genetic techniques, researchers can quickly highlight loci and genes that have signatures of adaptive evolution

(reviewed by Hasselmann *et al.* 2015; Dogantzis and Zayed 2019). These studies have been applied to several eusocial insects including bees (Harpur *et al.* 2014; Molodtsova *et al.* 2014; Wallberg *et al.* 2014; Fouks and Lattorff 2016; Dogantzis *et al.* 2018; Imrit *et al.* 2020; Jones *et al.* 2023), wasps (Miller *et al.* 2020; Xu *et al.* 2021; Li *et al.* 2024), ants (Romiguier *et al.* 2014; Roux *et al.* 2014; Engsontia *et al.* 2015; Privman *et al.* 2018; Wiernasz and Cole 2022; Barkdull and Moreau 2023) and termites (Bulmer and Crozier 2006; Harrison *et al.* 2018; Meusemann *et al.* 2020; Radford *et al.* 2022). On its own, population genomics provide invaluable insights into the evolutionary forces that have shaped genetic diversity within and between species, and has been instrumental in identify genes that have influenced fitness in both advanced (Harpur *et al.* 2014; Dogantzis *et al.* 2018; Imrit *et al.* 2020; Barkdull and Moreau 2023; Jones *et al.* 2023) and primitively eusocial (Fouks and Lattorff 2016; Harpur *et al.* 2017; Dogantzis *et al.* 2018) species. These studies have highlighted loci that have experienced natural selection over the time scale of the evolution and elaboration of eusociality, thereby providing a list of candidate genes that were potentially important for modulating eusocial traits. Notably, in some lineages, genes with signatures of positive selection are often associated with worker- and colony-level traits (Harpur *et al.* 2014; Dogantzis *et al.* 2024). There is also recent evidence of repeated use of the same genes in facilitating local adaptation to different environments (Ji *et al.* 2020; Dogantzis *et al.* 2024). However, we would like to mention two caveats with population genomic studies of positive selection. First, we note that demography can influence analysis of positive selection leading to some false positives (Beichman *et al.* 2018; Johri *et al.* 2022), although this limitation can be overcome by applying best-practices and the latest and most robust methods (Alvarez-Carretero *et al.* 2023; Kirsch-Gerweck *et al.* 2023). More importantly, while population genomics statistically links DNA sequences with fitness consequences, it does not, on its own, provide knowledge on the phenotypes and molecular mechanisms that yield to fitness differentials in the wild. This shortcoming, however, can potentially be circumvented by using the other research approaches discussed below.

Linking phenotype to candidate loci using GWAS

Another application of population genomics can be used to study the genetics underlying eusocial traits. One way to achieve this is by sequencing samples from natural or artificially selected 'populations' that differ in a specific phenotype. Population contrasts can reveal mutations that vary drastically in allele frequency between populations, and these may underlie some of the phenotypic differences between populations. This approach has been applied to study hygienic behaviour (Spotter *et al.* 2016; Harpur *et al.* 2019), the susceptibility to *Varroa* mites (Broeckx *et al.* 2019), as well as defense behaviour (Avalos *et al.* 2020) and gentleness (Guichard *et al.* 2021) in the honey bee *Apis mellifera*. Population genomic contrasts

have also identified genetic variants that vary with solitary and social phenotypes within a single species such as in the socially polymorphic sweat bee *Lasioglossum albipes* (KOCHER *et al.* 2018), the facultatively eusocial *Megalopta genalis* (KAPHEIM *et al.* 2020a) and the socially polymorphic Australian small carpenter bee *Ceratina australensis* (HARPUR AND REHAN 2021). More recently, researchers have carried out genome-wide association studies (GWAS) to link variation in phenotype with specific mutations segregating across the entire genome. Typically, GWAS involve sequencing a large cohort of individuals that have also been quantified for a trait of interest. Statistical associations between genotype and phenotype are taken as evidence that a specific mutation is causing or is linked to a causal mutation that influences that studied trait. For example, studies have found causal mutations involved in social organization in *Formica francoueri* (PIERCE *et al.* 2022) and *Formica neoclara* (MCGUIRE *et al.* 2022) and in sex ratio formation in *Formica glacialis* (LAGUNAS-ROBLES *et al.* 2021).

Population genomics and GWAS approaches have so far confirmed that eusocial traits are highly complex and multifactorial. For example, Harpur and colleagues (HARPUR *et al.* 2019) found 73 candidate genes linked to hygienic behaviour. GWAS have also revealed 322 candidate genes associated with nest defense behaviour among Africanized honey bees (HARPUR *et al.* 2020) while another study identified 1,172 SNPs associated with colony-level aggression in African honey bees. GWAS also identified 42 highly differentiated SNPs among socially polymorphic *M. genalis* (KAPHEIM *et al.* 2020a). These studies demonstrate the power of GWAS in identifying candidate loci associated with a phenotype and studying the genetic architecture of complex traits (GOLDSTEIN 2009; FADISTA *et al.* 2016; YENGO *et al.* 2022).

However, GWAS have faced criticism for their limited ability to fully explain the heritability of complex traits (TAM *et al.* 2019). Such studies typically only have power to identify common variants, potentially overlooking rarer variants (GOLDSTEIN 2009). Moreover, causal alleles may be in linkage disequilibrium (LD) with other alleles identified through GWAS, leading to false positives and uncertainties about the specific genes or mutations affecting traits if LD blocks are large (HISCHHORN *et al.* 2002; INTERNATIONAL HAPMAP 2005). Consequently, reproducing the data becomes challenging, as alleles in LD or those associated with population stratification (BERG *et al.* 2019; AW *et al.* 2024) may achieve statistical significance in one study but not in subsequent studies (HISCHHORN *et al.* 2002). GWAS also face a multiple testing burden which complicates the identification of causal variants (PE'ER *et al.* 2008). Perhaps the greatest weakness of population genomic and GWAS approaches to date is the lack of experimental validation via genetic manipulation of the candidate loci. This shortcoming however can now be addressed using the emerging CRISPR/Cas9 system, as we discuss below.

Linking DNA to phenotype and molecular function using CRISPR/Cas9

CRISPR/Cas9 has revolutionized the way in which researchers study gene function by enabling precise genetic modifications to the DNA of model and non-model organisms (Degirmenci *et al.* 2020; Liang *et al.* 2021; Ranjitha and Bhuvana 2022). CRISPR/Cas9 uses a single guide RNA (sgRNA) to direct the Cas9 endonuclease to a specific DNA sequence of interest where it induces a break, which, along with the cell's downstream repair machinery, allows for precise genetic modifications such as frameshift mutations (Degirmenci *et al.* 2020; Liang *et al.* 2021; Ranjitha and Bhuvana 2022). This approach has been very useful for studying gene function in social insects, which with a few exceptions, lack the sophisticated toolkits for genetic manipulation developed for the fruit fly, *Drosophila melanogaster*. So far, CRISPR/Cas9 has been applied to knock out genes in *Apis mellifera* (Kohno *et al.* 2016; Kohno and Kubo 2018; Hu *et al.* 2019; McAfee *et al.* 2019; Roth *et al.* 2019; Degirmenci *et al.* 2020; Sinakevitch *et al.* 2020; Chen *et al.* 2021; Liang *et al.* 2021; Nie *et al.* 2021; Wang *et al.* 2021; Wagner *et al.* 2022; Cheng *et al.* 2023; Otte *et al.* 2023), *Harpegnathos saltator* (Yan *et al.* 2017), *Lasius niger* (Konu *et al.* 2023), *Ooceraea biroi* (Ivasyk *et al.* 2023), *Solenopsis invicta* (Chiu *et al.* 2020), and *Vespula vulgaris* (Lester *et al.* 2020)(Table 1). For example, CRISPR/Cas9 has been used to knockout *Orco*, in *Ooceraea biroi* (Trible *et al.* 2017) and *Harpegnathos saltator* (Yan *et al.* 2017) to confirm the importance of this gene in olfactory communication, antennal lobe development and social behaviour. Specifically, these studies identified the effect of *Orco* mutants on the volume of the antennal lobe, the number of glomeruli, as well as the size of the remaining glomeruli (Trible *et al.* 2017; Yan *et al.* 2017). This research then served as the basis for *Orco* manipulation within honey bees (Chen *et al.* 2021). CRISPR/Cas9 has also been crucial for studying sex determination in honey bees which has enabled the induction of diploid males (Wang *et al.* 2021), knockouts leading to male development (Roth *et al.* 2019; Cheng *et al.* 2023) and the identification of a locus important for the initiation of the female sex determination pathway (Otte *et al.* 2023).

While CRISPR/Cas9 is proving to be an incredibly useful tool for studying gene function in social insects, we note that, for the most part, the list of genes manipulated using this technology thus far have been chosen without any evidence from population genomics or GWAS. For example, most genes manipulated by CRISPR/Cas9 in social insects were chosen based on a candidate gene approach, often borrowing knowledge from evolutionarily distant taxa (Table 3-1). For example, Chiu et al. (2020) targeted the gene spitz in *Solenopsis invicta* due to its function in *Drosophila* larval oenocytes, Konu and colleagues (2023) manipulated the gene *cinnabar* in *Lasius niger* due to its involvement in ommochrome biosynthesis in *D. melanogaster* (Sethurman and O'Brochta 2005) and members of the yellow gene family were knocked

out in *Apis mellifera* due to their role in melanization in *D. melanogaster* (Liang *et al.* 2021; Nie *et al.* 2021). While these studies have greatly contributed to our understanding of the molecular biology behind significant traits in social insects, the extent to which mutations in candidate genes have played a role in the evolution and maintenance of eusociality in insects remains unclear. Manipulating a gene to confirm its link to a eusocial trait does not necessarily imply that this gene has been significant in the evolution or maintenance of eusociality. Genes work within pathways (Pita-Juarez *et al.* 2018), and although manipulating a gene can confirm its importance within a pathway for a particular social trait it does not necessarily suggest that mutations in this specific gene contribute to the evolution and maintenance of eusociality in the lineage being studied. We think that data from population genomics and GWAS can be invaluable for choosing targets to manipulate using CRISPR. Such studies could directly implicate specific mutations and loci that have either played a significant adaptive role over the course of eusocial evolution (Harpur and Zayed 2013; Johnson *et al.* 2018) or exhibited associations with phenotypic variation in key eusocial traits (Rubin *et al.* 2019; Sieber *et al.* 2021). We think selecting genes to manipulate based on *a priori* knowledge from population genomics and GWAS will be key to understanding the relationship between genes and eusociality, and for testing different mechanistic hypotheses about the evolution and maintenance of eusocial behaviour.

Conclusion

The integration of population genomics, GWAS and CRISPR/Cas9 technology presents a powerful and multifaceted approach to unraveling the genetic, molecular, and evolutionary underpinnings of eusociality in insects. Population genomics offers insights into the adaptive evolution of coding and regulatory sequences, identifying genes that have undergone positive selection during the development and maintenance of eusocial behaviour. GWAS further refines our understanding by linking specific genetic variants to eusocial traits, revealing the complex genetic architecture underlying these behaviours. CRISPR/Cas9 complements these methods by enabling precise genetic manipulation, thus validating genotype-phenotype associations, and elucidating the molecular functions of candidate genes. By leveraging the strengths of each approach and addressing their individual limitations, researchers can achieve a more comprehensive understanding of the intricate genetic networks and evolutionary processes that drive eusociality. This integrated strategy holds the promise of not only advancing our knowledge of social insect biology but also informing broader studies on the evolution of complex social behaviours.

Figures

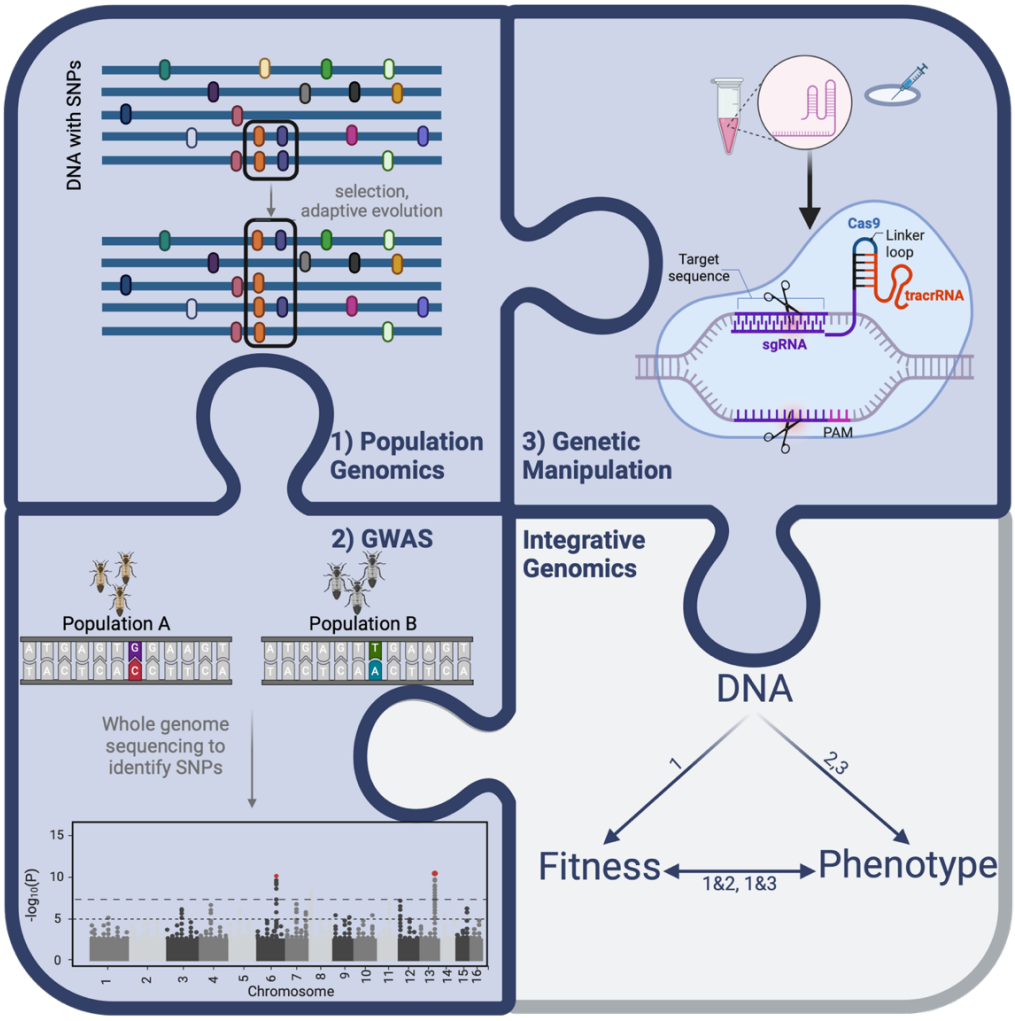


Figure 3-1 *Integrated toolkit for studying the genetics and evolution of eusocial traits*

Tables

Table 3-1. Genes selected for knockout with CRISPR/Cas9 in social insects and how they were chosen.

A comprehensive survey of CRISPR/Cas9 studies in social insects identified predominant methods for gene selection.

Species	Gene for knockout	Method of Choosing Gene*	References
Apis mellifera	*AmGR1,2,3*	candidate gene approach	(Degirmenci *et al.* 2020)
	Amyellow-y	candidate gene approach	(Liang *et al.* 2021; Nie *et al.* 2021)
	csd	candidate gene approach	(Wang *et al.* 2021; Otte *et al.* 2023)
	dsx	candidate gene approach	(Roth *et al.* 2019; Wagner *et al.* 2022)
	fem	candidate gene approach	(Cheng *et al.* 2023)
	major royal jelly protein	candidate gene approach	(Hu *et al.* 2019; Kohno and Kubo 2019)
	mGlutR1	candidate gene approach	(Sinakevitch *et al.* 2020)
	mKast	transcriptomics	(Kohno and Kubo 2018)
	Orco	candidate gene approach	(Chen *et al.* 2021)
	pax6	candidate gene approach	(Hu *et al.* 2019)
	RDL	candidate gene approach	(Sinakevitch *et al.* 2020)
Harpegnathos saltator	*Orco*	candidate gene approach	(Yan *et al.* 2017)
Lasius niger	*cinnabar*	candidate gene approach	(Konu *et al.* 2023)
Ooceraea biroi	*DNMT1*	candidate gene approach	(Ivasyk *et al.* 2023)
	Orco	candidate gene approach	(Trible *et al.* 2017)
Solenopsis invicta	*Sinv-spitz*	candidate gene approach	(Chiu *et al.* 2020)
	GP-9	candidate gene approach	(Chiu *et al.* 2020)
Vespula vulgaris	*cdc*	candidate gene approach	(Lester *et al.* 2020)
	boule	candidate gene approach	(Lester *et al.* 2020)

*Method for gene selection: Candidate gene approach- Genes are selected based on prior knowledge, often derived from studies on evolutionary distant model organisms, transcriptomic profiling- genes are chosen through analysis of gene expression patterns, phylogenetic analysis- genes are selected based on exhibiting high rates of protein evolution across social insects as a whole or within specific social lineages, population genomics- genes are identified with mutations showing distinct evolutionary patterns within a species, assessed through population genetic metrics like Fst, Tajima's D, etc., quantitative genetics- genes are pinpointed based on mutations associated with a trait of interest, identified through genome-wide association studies (GWAS) or quantitative trait loci (QTL) analyses.

Chapter 4 Exploring the genetics of social behaviour in *C. calcarata*[3]

Dova B. Brenman- Suttner[1], Sandra M. Rehan[1], and Amro Zayed[1]

Summary

Studies investigating social evolution often focus on species that are obligately eusocial, where presumably all 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 (SNPs) 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.

Introduction

Eusociality, defined by overlapping generations, cooperative brood care, and a reproductive division of labour, has evolved 11 different times within Hymenoptera, four of which occurred within bees (WILSON AND HOLLDOBLER 2005; GROOM AND REHAN 2018). 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 (HAMILTON 1964), the precise molecular and genomic changes involved in eusocial evolution are not well understood. Much of the research has focused on obligately

[3] Published in Scientific Reports (2025) https://doi.org/10.1038/s41598-025-89870-9. This article is available under the Creative Commons CC-BY-NC license and permits non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

eusocial species as a model for eusocial evolution (PAGE AND AMDAM 2007; KAPHEIM *et al.* 2015). 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 (REHAN AND TOTH 2015). 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 (WEST-EBERHARD 1987; REHAN AND RICHARDS 2010a; SHELL AND REHAN 2017). 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 (REHAN *et al.* 2018). *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 (LAWSON *et al.* 2016). 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 (REHAN *et al.* 2014; MIKÁT *et al.* 2017). 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 (WILSON 1971; MICHENER 1974; LINKSVAYER AND WADE 2005).

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 (TRUE AND CARROLL 2002; CARROLL *et al.* 2005; REHAN AND TOTH 2015). Support for the genetic toolkit hypothesis also comes from comparative studies of *Apis, Bombus, Dufourea, Euglossa, Eulaema, Eufriesea, Exoneura, Frieseomelitta, Habropoda, Lasioglossum, Melipona*, and *Megachile* (WOODARD *et al.* 2011; KAPHEIM *et al.* 2015), as well as other studies in *Apis* (GROZINGER *et al.* 2003; GROZINGER *et al.* 2007) and in fire ants of the genus *Solenopsis* (OMETTO *et al.* 2011; BERENS *et al.* 2015). Other gene regulatory hypotheses such as the Ovarian Ground-plan Hypothesis (WEST-EBERHARD AND TURILLAZZI 1996) 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 wasps (TOTH *et al.* 2007) and honey bees (AMDAM *et al.* 2004; AMDAM *et al.* 2006; GRAHAM *et al.* 2011), with evidence to the contrary found thus far in carpenter bees (REHAN *et al.* 2014). The Maternal Heterochrony Hypothesis, with support from social wasps (TOTH *et al.* 2007), bumble bees (WOODARD *et al.* 2014) and *Ceratina calcarata* (REHAN *et al.* 2014), suggests that social evolution evolves through reorganization of the timing and location of gene expression related to parental care (LINKSVAYER AND WADE 2005). 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 (JOHNSON AND TSUTSUI 2011). Evidence supporting this hypothesis is seen in the advanced eusocial honeybee, *Apis mellifera*, which possesses a large number of novel, taxonomically restricted genes (TRGs) (JOHNSON AND TSUTSUI 2011). These TRGs are predominantly expressed in workers and exhibit much higher rates of positive selection compared to non-TRGs (HARPUR *et al.* 2014). 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 (SHELL AND REHAN 2019). 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 (SHELL AND REHAN 2019). A study examining the effects of maternally cared for and orphaned offspring found differential alternatively spliced genes between the two groups (ARSENAULT *et al.* 2018; SHELL *et al.* 2021; CHAU *et al.* 2023). 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.* (2020) 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 (IMRIT *et al.* 2020). Moreover, advanced eusocial insects exhibit some of the highest recombination rates known, likely serving the purpose of increasing genotypic diversity within the worker population (SIRVIO *et al.* 2006; WILFERT *et al.* 2007; KENT AND ZAYED 2013; OLDROYD *et al.* 2021). 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 (RUBIN *et al.* 2019). 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 (RUBIN *et al.* 2019). The researcher 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 (RUBIN *et al.* 2019). 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 (SHELL *et al.* 2021). 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 (SHELL *et al.* 2021). 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 (KOCHER *et al.* 2013; KOCHER *et al.* 2018; KAPHEIM *et al.* 2020a; HARPUR AND REHAN 2021). 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)(KAPHEIM *et al.* 2020a). A similar study in the facultatively social *C. australensis* identified specific loci that were associated with social colony structure that showed signs of positive selection (HARPUR AND REHAN 2021). This suggests that alleles at genes associated with social colony 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 (KOCHER *et al.* 2013; KOCHER *et al.* 2018). 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 (KAPHEIM *et al.* 2015; KAPHEIM *et al.* 2020a). In other species, the genetic architecture of social traits often exhibit considerable variation, ranging from supergenes in ants that govern colony type formation (YAN *et al.* 2020), a single locus (and likely a single SNP) controlling parthenogenesis in the Cape honey bee, *Apis mellifera capensis* (YAGOUND *et al.* 2020), to multiple genes that may control hygienic behaviour (HARPUR *et al.* 2019) and *Varroa*-specific defense behaviour in honey bees (SPOTTER *et al.* 2016). 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 (KOCHER *et al.* 2013; KOCHER *et al.* 2018; KAPHEIM *et al.* 2020a; HARPUR AND REHAN 2021).

Previous studies on *C. calcarata* have helped us understand the transcriptomic changes associated with social behaviours among social and subsocial groups (REHAN *et al.* 2014; DURANT *et al.* 2015; WITHEE AND REHAN 2017; SHELL AND REHAN 2019; SHELL *et al.* 2021; SHELL AND REHAN 2022). However, it remains unclear whether these behavioural differences are linked to genetic variation, especially because social and subsocial colonies 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 colonies in *C. calcarata,* which are known to also exhibit transcriptional differences (ARSENAULT *et al.* 2018; SHELL AND REHAN 2019; SHELL *et al.* 2021; CHAU *et al.* 2023). 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.

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 (HARPUR *et al.* 2014; NAZARENO *et al.* 2017; EIMANIFAR *et al.* 2018; YANCAN *et al.* 2019; SHELL AND REHAN 2022) and these sample sizes have been recognized as suitable for this type of population genomic study (WILLING *et al.* 2012). 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 (SHELL *et al.* 2018), social colonies were classified as those with a female in the first brood cell, whereas subsocial colonies 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 70-100µl using established methods (CONFLITTI *et al.* 2022). Tissues were lysed using 350µl of Tissue Lysis Buffer, 20µl (20g/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, 2x150 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 364Mb. We performed quality control on all samples to check the quality, read length, GC content and other standard parameters using Qualimap 2 (OKONECHNIKOV *et al.* 2016).

Alignment and SNP Calling

We used previously published pipelines for alignment and SNP calling (HARPUR *et al.* 2019; IMRIT *et al.* 2020; HARPUR AND REHAN 2021). Briefly, adapter sequences were trimmed from raw reads using TRIMMOMATIC v0.32 (BOLGER *et al.* 2014) and then aligned to the *Ceratina calcarata* ASM165200v1 genome(REHAN *et al.* 2016) using NextGenMap v0.4.12 (SEDLAZECK *et al.* 2013). 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)(VAN DER AUWERA *et al.* 2014). We then identified single nucleotide polymorphisms (SNPs) using GATK HaplotypeCaller via an established pipeline (DOGANTZIS *et al.* 2018).

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)(VAN DER AUWERA *et al.* 2014). Following previous methods, we then filtered SNPs based on abnormally high or low read depth. The upper limit was determined by calculating 1.5x the interquartile range of the cumulative depth of each sample (IMRIT *et al.* 2020). The lower limit for depth was determined by allowing an average depth of coverage of 5 reads per

base pair per individual (5 reads x 20 samples lower pooled threshold of 100)(KAPHEIM *et al.* 2019). Low frequency SNPs using a minor allele frequency threshold of less than 0.05 were also removed from the sample set (KAPHEIM *et al.* 2019).

Population genetic analysis

We determined the population structure of social and subsocial *C. calcarata* (n = 20) using ADMIXTURE v.1.30 (ALEXANDER *et al.* 2009). 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 (HARPUR AND REHAN 2021). 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 (WICKHAM 2016).

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 (DANECEK *et al.* 2011; WAIKER *et al.* 2021). We measured genetic differentiation (F_{st}) between social (n = 10) and subsocial (n = 10) *C. calcarata* using VCFtools with the --weir-fst-pop parameter (DANECEK *et al.* 2011). 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 (NAZARENO *et al.* 2017) and empirical study (HARPUR *et al.* 2014; EIMANIFAR *et al.* 2018; SHELL AND REHAN 2022). 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 in SNPs in Group 1 and Group 2. We used python matplotlib (HUNTER 2007) 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 (CINGOLANI *et al.* 2012). 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/; Supplemental Table S4-1A). 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 (QUINLAN AND HALL 2010).

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 (WALSH *et al.* 2022). 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)(GRAMATES *et al.* 2022). 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 (Supplemental Table S4-2).

Gene ontology annotation enrichment clustering analysis was carried out using DAVID 2022 (SHERMAN *et al.* 2022) 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 (KUZNETSOV *et al.* 2022). 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 (KUZNETSOV *et al.* 2022). We downloaded flat files from the OrthoDB site, filtered them for annotations involving *Ceratina calcarata* (NCBI taxonomic ID: 156304), 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 S4-3). 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.* (2016) supplemental Table 7, fracMeth.CDS.avg > 0.01) and genes showing signs of positive selection (REHAN *et al.* (2016) 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.* (2018) supplemental materials 41467_2018_5903_MOESM4_ESM.xslx)
C. Genes associated with aggression (WITHEE AND REHAN (2017) supplemental Table S3; WW >LL)
D. Genes associated with behaviour (foraging, guarding) and social status (mothers, autumn daughters, regular daughters) (SHELL AND REHAN (2019) 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.* (2021) supplemental Table S5)

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; Supplemental Figure 4-1). 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 (Begun *et al.* 2007; Qanbari *et al.* 2012; Harpur *et al.* 2014; Feng *et al.* 2015; Harpur and Rehan 2021; Yan *et al.* 2023)- to estimate genetic differentiation between the two behavioural groups. We were able to estimate F_{st} values between social and subsocial colonies for 3,106,635 single nucleotide polymorphisms (SNPs) with an average F_{st} of 0.019 ± 0.05 SD (Figure 4-1A). 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 (Storz *et al.* 2004; Voight *et al.* 2006; Begun *et al.* 2007; Harpur *et al.* 2014; Harpur and Rehan 2021). 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 (Akey *et al.* 2002). 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 (Figure 4-1B). 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 (Christmas *et al.* 2021). 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, Supplemental Figure 4-2) than that of SNPs in windows with no outliers (mean F_{st} = 0.009 ±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 (Figure 4-2; Supplemental Table S4-1A). 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 (Figure 4-2 compared to Supplemental Figure 4-3) we see very similar results, suggesting that the patterns reflect 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 Supplemental Figure 4-3 and Supplemental Tables S4-1B,C,D, Supplemental Figure 4-4B). 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 (Supplemental Figure 4-4A, Supplemental Table S4-1E).

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 (Supplemental Table S4-4) which was also found in the highest enrichment cluster (enrichment score = 2.245; Supplemental Table S4-5), 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 (Supplemental Figure 4-6).

Functional analysis of high F_{st} genes in C. calcarata *using HymenopteraMine and* Drosophila melanogaster *orthologs*

We identified distinct functional patterns in genes with high levels of genetic differentiation between social and subsocial *Ceratina calcarata*, based on gene functions assigned via HymenopteraMine (Supplemental Table S4-2). 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;).

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 overall or those 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 (JOHNSON AND TSUTSUI 2011), 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 (Figure 4-3, Supplemental Table S4-6A). 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 (Supplemental Figure 4-7, Supplemental Table S4-6B,C). When we grouped genes based on their evolutionary age (old vs. new) we also did not find any significant enrichment (P=0.930, Supplemental Table S4-6D). 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.

Connecting genetic differentiation and candidate genes for sociality

Several previous 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 (SHELL AND REHAN 2019). 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 (SHELL *et al.* 2021). Additionally, genomic features of *C. calcarata* have been explored, including methylation patterns, alternatively spliced genes and differentially methylated genes (REHAN *et al.* 2016; ARSENAULT *et al.* 2018). 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 (Figure 4-4, Supplemental Table S4-7A). 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 ($p = 0.017$) and DEGs among guarding mothers upregulated over all others ($p = 5.33e\text{-}4$), 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 behavioural polymorphism with either subsocial or social colonies, making it an excellent model organism for studying the genetics and evolution of social behaviour (Rehan and Richards 2010b; Rehan *et al.* 2014). 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' (Rehan and Toth 2015). However, it is also possible that transitions to complex sociality do not involve step-wise evolution of social organization (Linksvayer and Johnson 2019). 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 (Storz *et al.* 2004; Voight *et al.* 2006; Begun *et al.* 2007; Qanbari *et al.* 2012; Harpur *et al.* 2014; Feng *et al.* 2015; Harpur and Rehan 2021; Yan *et al.* 2023). 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 (Akey *et al.* 2002). While this analysis on its own does not prove that any specific marker casually 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 (Rittschof and Robinson 2013; Shpigler *et al.* 2017; Groom and Rehan 2018; Jones *et al.* 2020). 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 but may be largely influenced by phenotypic plasticity. The average genetic differentiation between subsocial and social *C. calcarata* (F_{st} = 0.019 ± 0.05, Figure 4-1) is comparable to F_{st} observed between solitary and social colonies of the facultatively eusocial *Lasioglossum albipes* and *Megalopta genalis* (average F_{st} of 0.06 and 0.03, respectively)(KOCHER *et al.* 2018; KAPHEIM *et al.* 2020a). 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* (ZAYED AND WHITFIELD 2008; HARPUR *et al.* 2014; DOGANTZIS *et al.* 2021). 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 (Figure 4-2). 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 (EROKHIN *et al.* 2019), while silencers and insulators are typically located in intergenic regions (VALENZUELA AND KAMAKAKA 2006).

The Genetic Toolkit Hypothesis proposes that eusociality evolved through the repurposing of highly conserved genes and molecular pathways for various social functions (WOODARD *et al.* 2011). 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 (TRUE AND CARROLL 2002). The regulation of these patterns of expression is important for the transition to social behaviour as seen in genomic studies of ants (SIMOLA *et al.* 2013; SCHRADER *et al.* 2015; GOSPOCIC *et al.* 2017), the bumblebee *Bombus terrestris* (WOODARD *et al.* 2011; WOODARD *et al.* 2014), the halictid bee *Lasioglossum albipes* (KOCHER *et al.* 2018), in transcriptomic

studies of *C. calcarata* (Shell and Rehan 2019) and in a comparative study of *Apis mellifera*, *Solenopsis invicta*, and *Polistes metricus* (Berens *et al.* 2015). 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 (Carthew 2021). For instance, genes displaying significant genetic differentiation among behaviour groups were found to be associated with annotations related to non-coding RNA metabolism (see Supplemental Figure 4-5), highlighting the crucial role of non-coding RNA as regulators of gene expression, which is pivotal for social evolution (Simola *et al.* 2013). We also found many genes with outlier levels of F_{st} associated with behaviour related to activating, initiating or regulating transcription. Histone modifications in regulatory regions have also been shown to be important for establishing and maintaining caste-differences in honey bees (Wojciechowski *et al.* 2018) and caste-specific foraging behaviour in the carpenter ant *Camponotus floridanus* (Simola *et al.* 2016). We identified several genes associated with behaviour in *C. calcarata* that had high genetic differentiation at histone 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 (Toth and Robinson 2007).

The Novel Gene Hypothesis proposes that new genes that are uniquely found within a lineage-specific group are important for the elaboration of eusociality (Johnson and Tsutsui 2011). 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 (Johnson and Tsutsui 2011). 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 genetic differentiation associated with social or 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 because it mostly explains the elaboration of sociality once more advanced social behaviour has been established, rather than during the earlier stages of social evolution, of which *Ceratina calcarata* is a model system for studying (Rehan and Richards 2010b). However, this does support 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 numerous *Drosophila melanogaster* orthologs and our gene ontology enrichment analysis, we provide strong evidence for the collective support of 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 (LINKSVAYER AND WADE 2005). 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 (LINKSVAYER AND WADE 2005). 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 (REHAN *et al.* 2014). 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 (Figure 4-4)(SHELL AND REHAN 2019). 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 (WITTWER *et al.* 2017). Plasticity in these chemical signals is regulated at the transcriptional level to alter composition, proportion, and timing of display of these chemicals (MALKA *et al.* 2014). 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.* (2017) 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 (Gill *et al.* 2013; Wittwer *et al.* 2017). 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 (Kocher *et al.* 2013).

Antioxidant enzymes are important for bees who regularly face environmental stimuli, such as foragers who frequently encounter different flowers (Mao *et al.* 2015). This need has driven the expansion of certain gene families such as the Cyp6A family in the genome of *Apis mellifera* (Johnson *et al.* 2018). 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 (Shell and Rehan 2022). 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.* (2023) focusing on those differentially expressed in larval gonads to understand the link between polyandrous mating and ovariole number in developing reproductive, 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* (Johnson *et al.* 2018), 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. However, 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. However, 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.

Figures

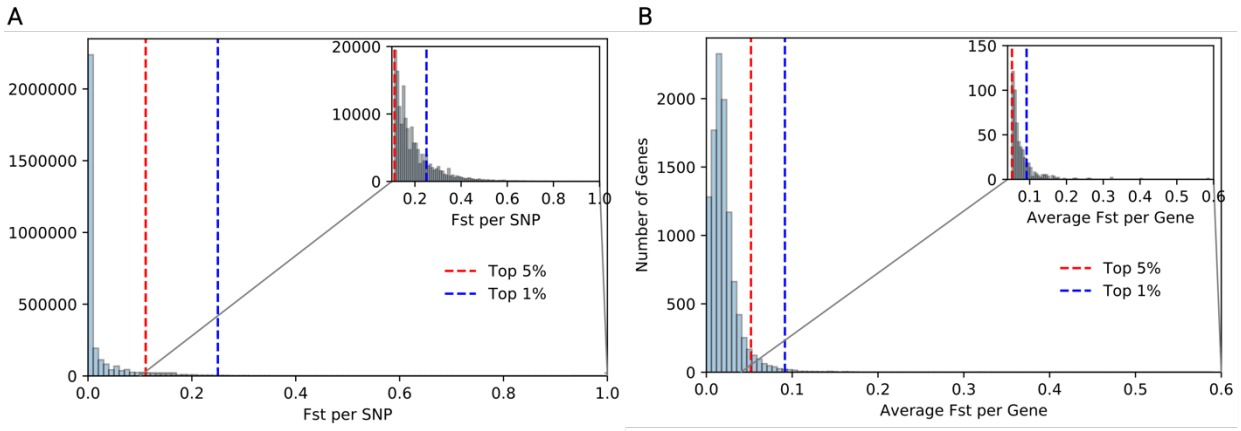


Figure 4-1. The vast majority of SNPs (A) and genes (B) show very low levels of genetic differentiation (pairwise F_{st} ~ 0) between social and subsocial *C. calcarata*

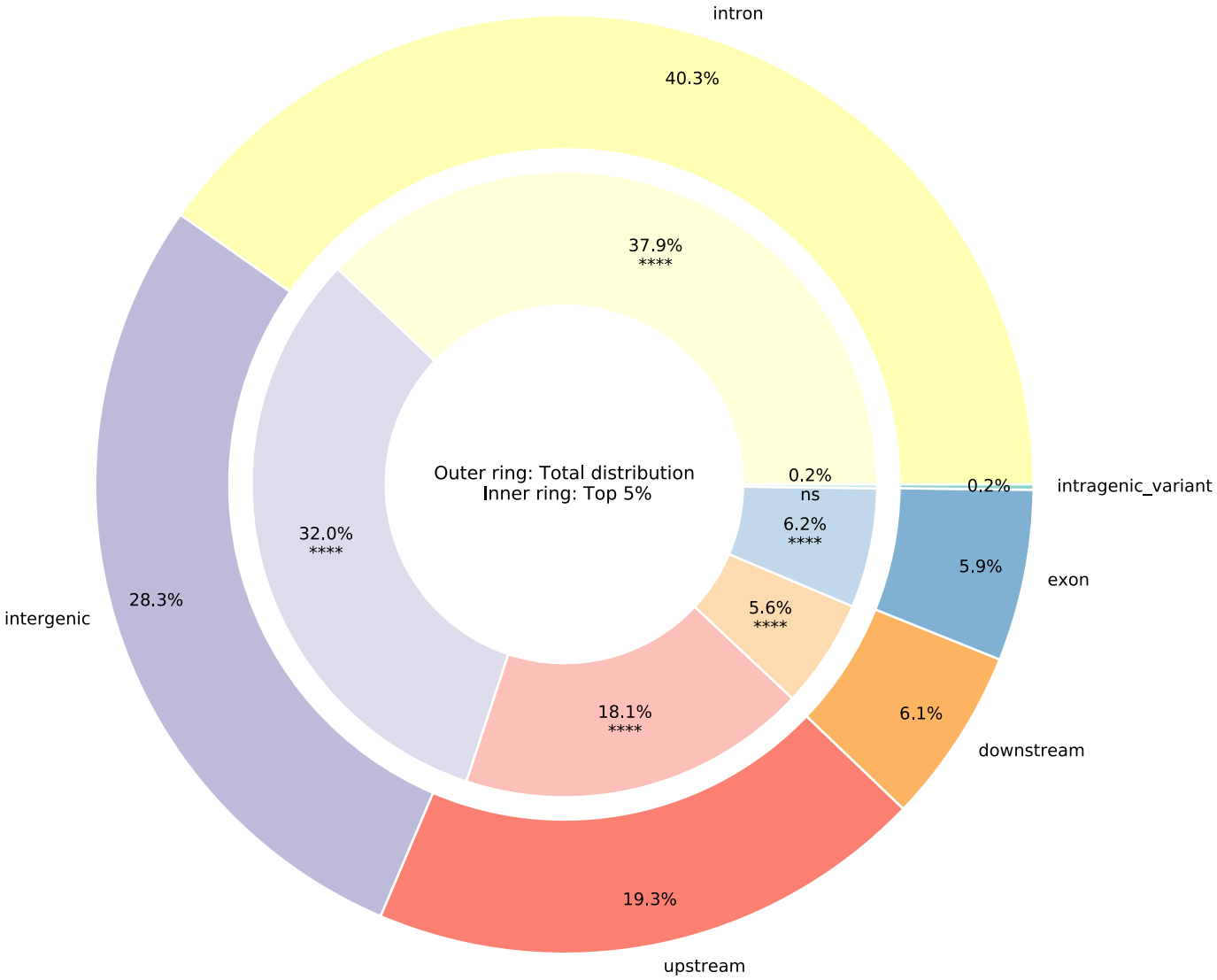


Figure 4-2. Enrichment of outlier SNPs across different mutation types in *C. calcarata*.

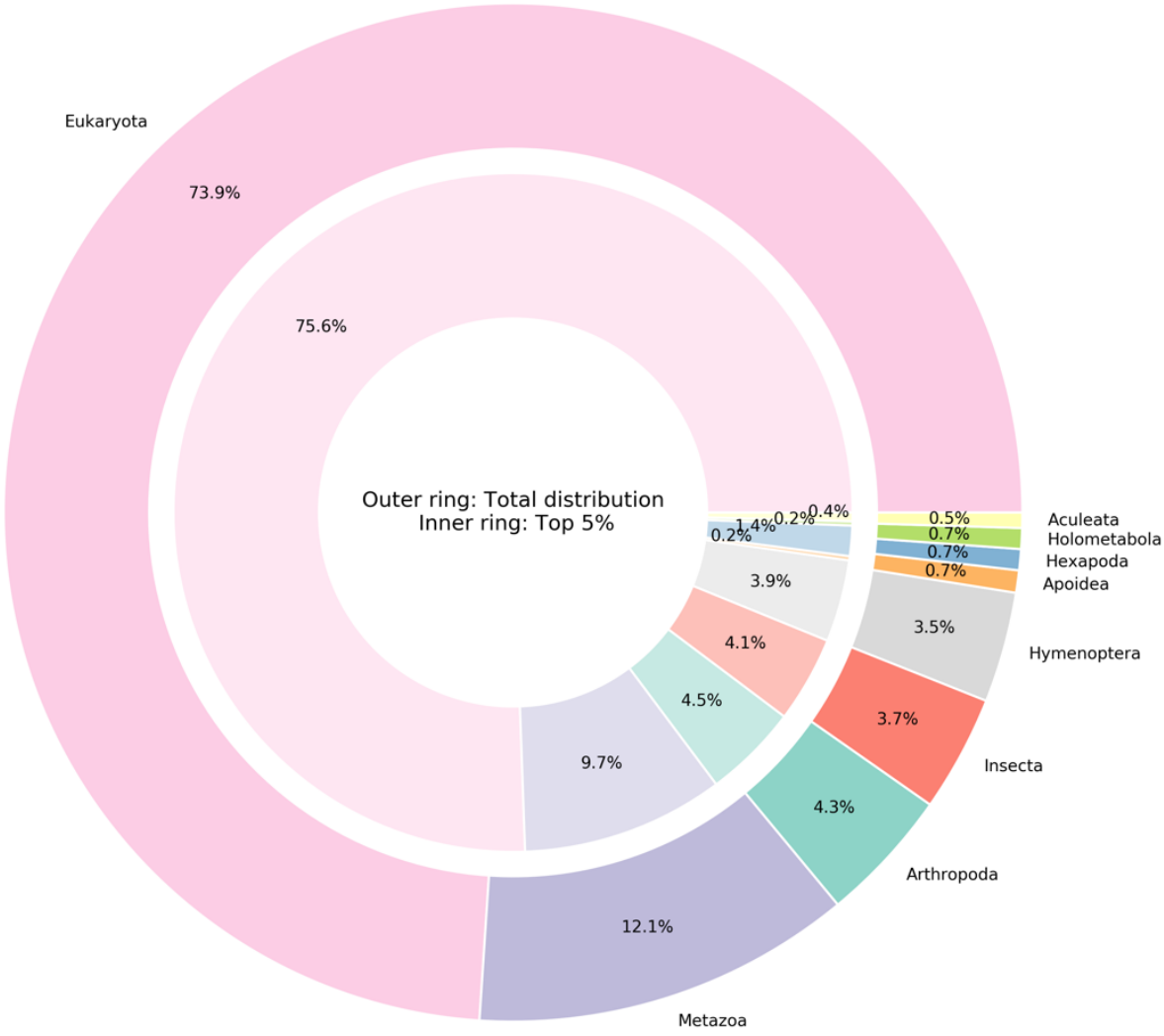


Figure 4-3. *Genes associated with social behaviour among social and subsocial groups in* C. calcarata *are predominantly evolutionarily ancient and the outlier list of genes are not enriched for any group (ns = not significant, *p ≤0.05, **p ≤0.01, ***p ≤0.001, ****p < 0.0001).*

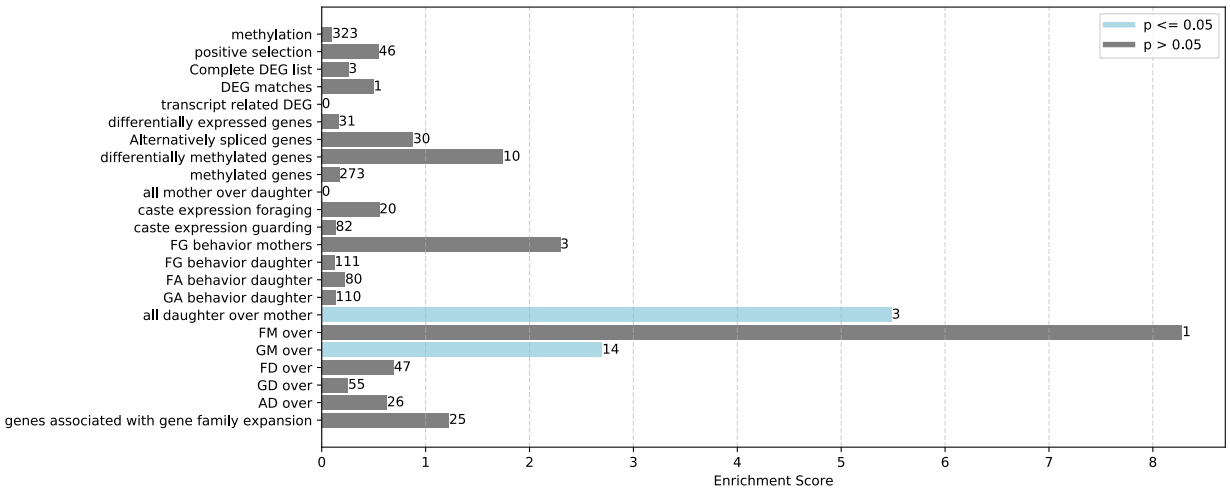
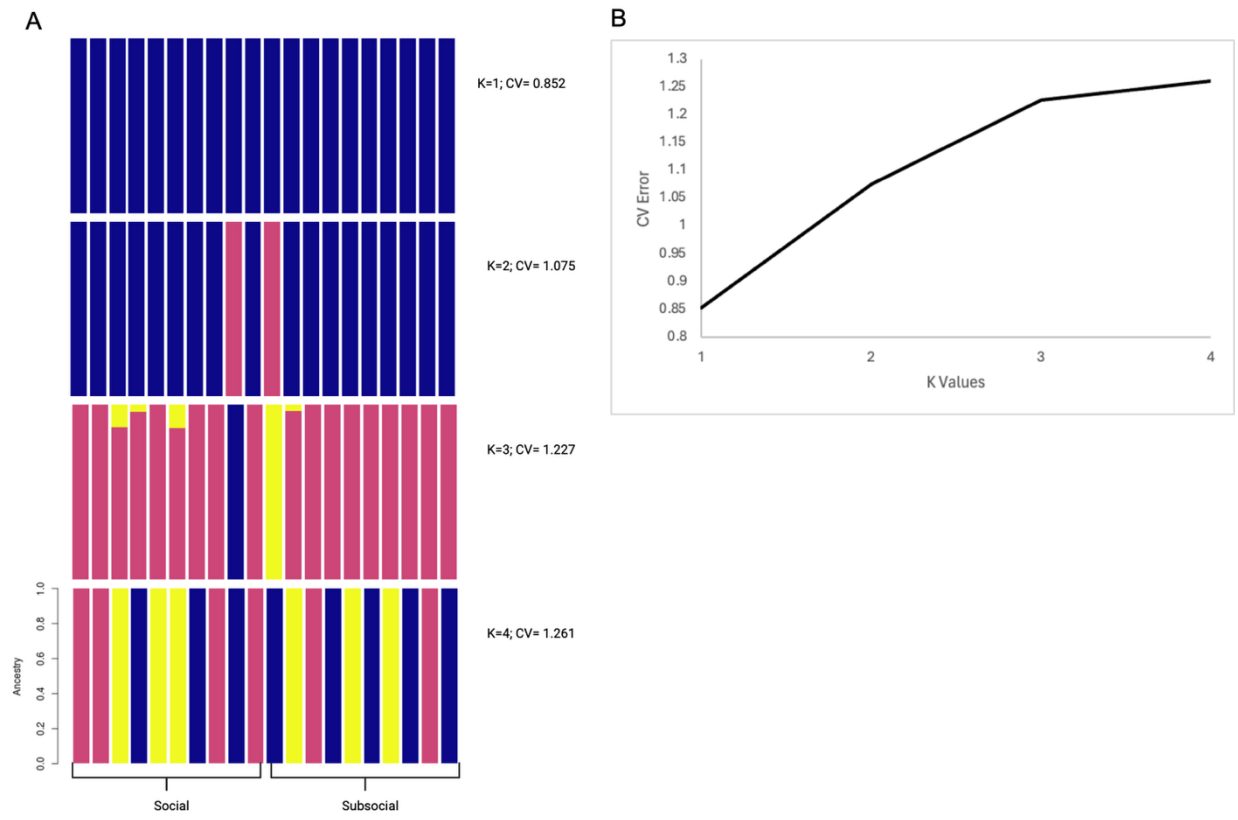
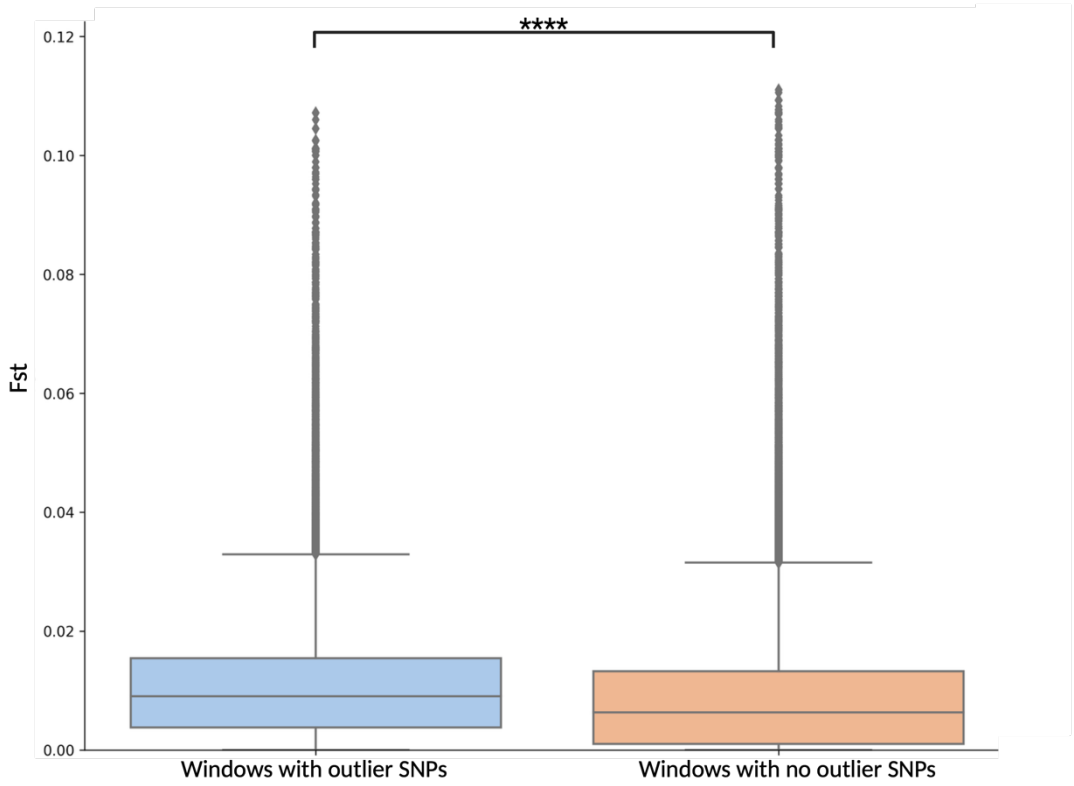


Figure 4-4. *Enrichment for genes with outlier levels of F_{st} (5%) associated with social behaviours among social and subsocial groups and previous genomic and transcriptomic studies in* C. calcarata.

Supplemental Figures

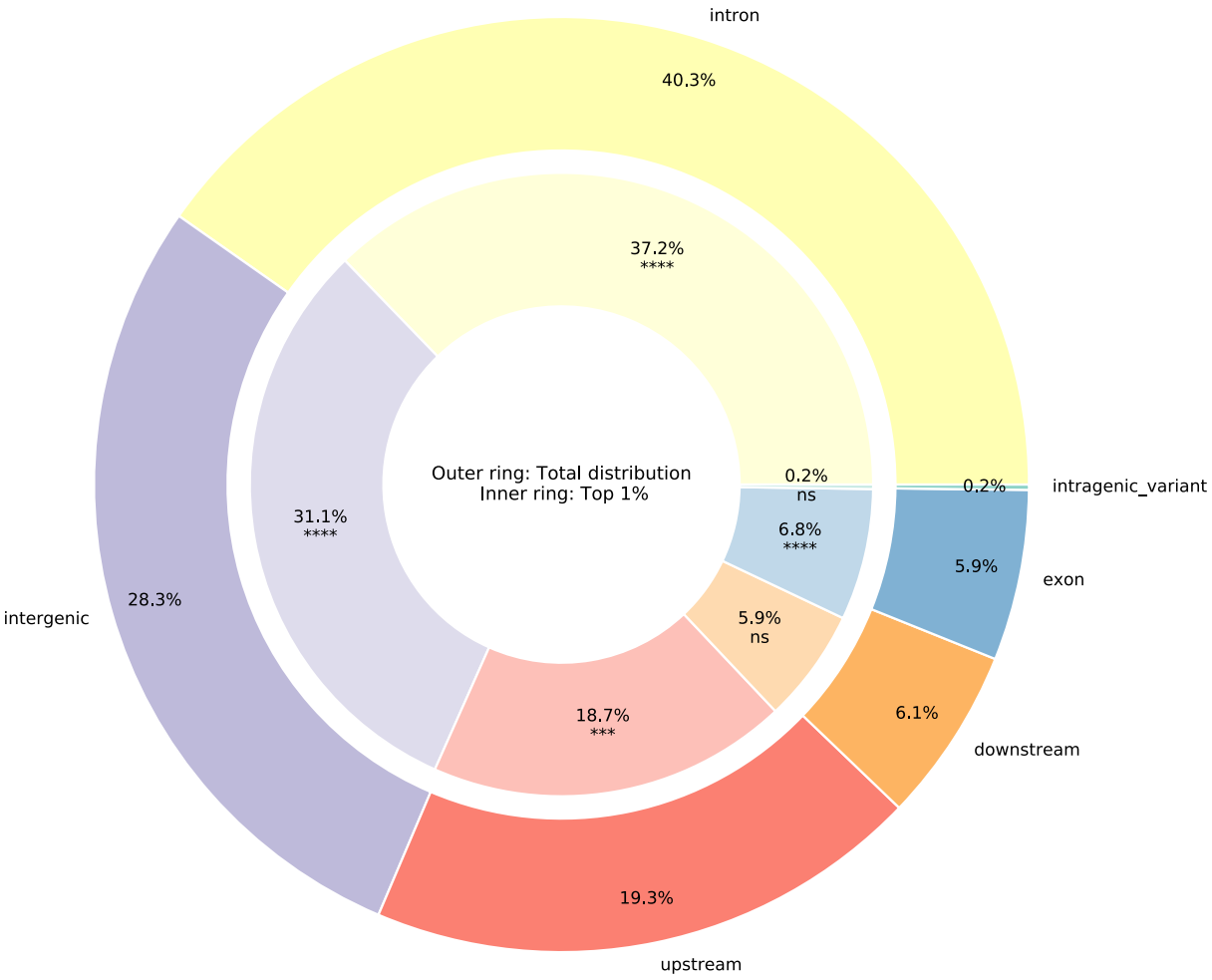


Supplemental Figure 4-1 *(A) Social (n=10, left) and subsocial (n=10, right)* Ceratina calcarata *originate from one population. (B) Cross validation error, averaged across three runs, showing support for K=1 (CV=0.852).*

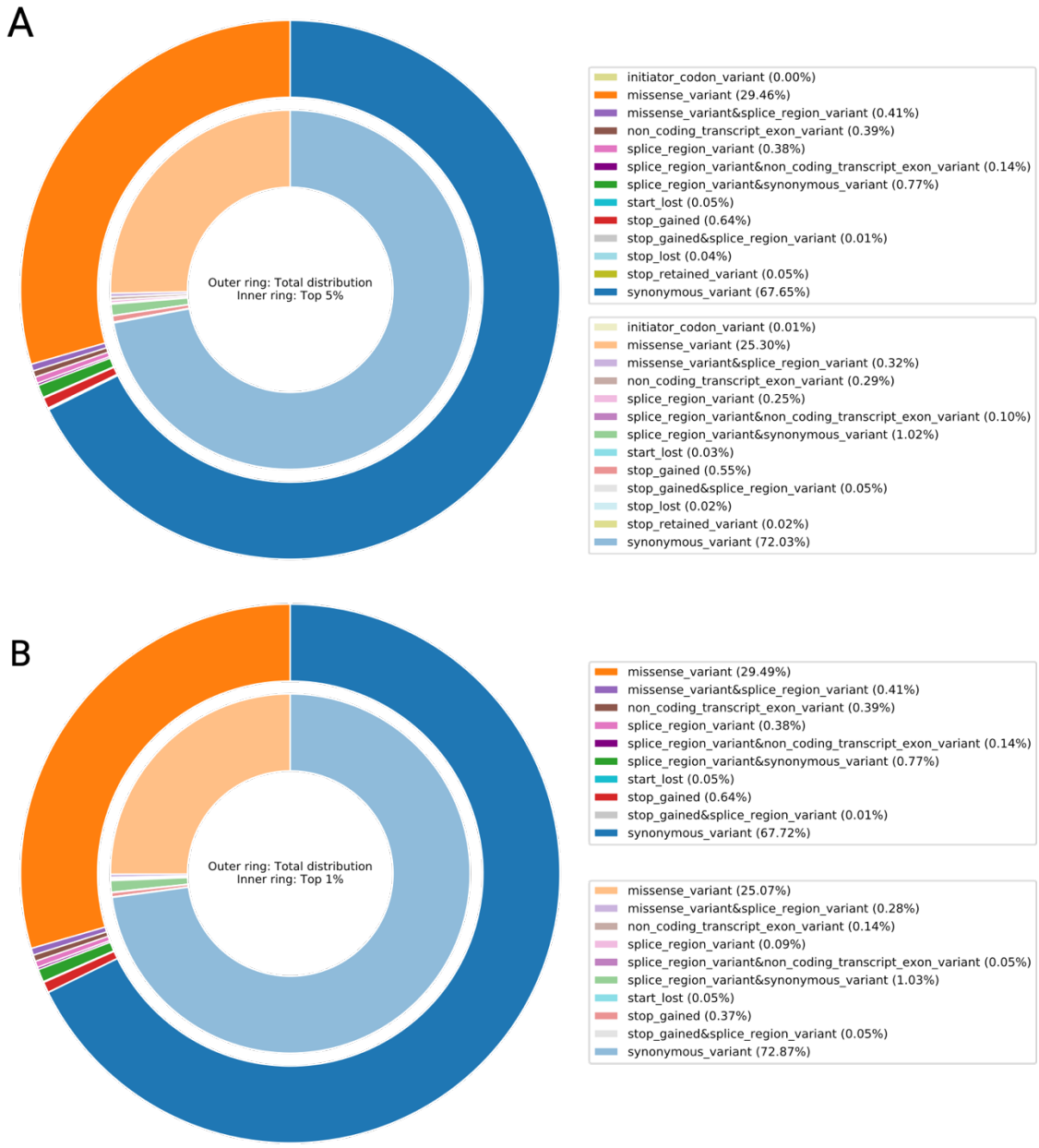


***Supplemental Figure 4-2**. Average F_{ST} of non-outlier SNPs in windows with at least one outlier SNP versus windows with no outlier SNPs in 500 base pair windows across the* Ceratina calcarata *genome.*

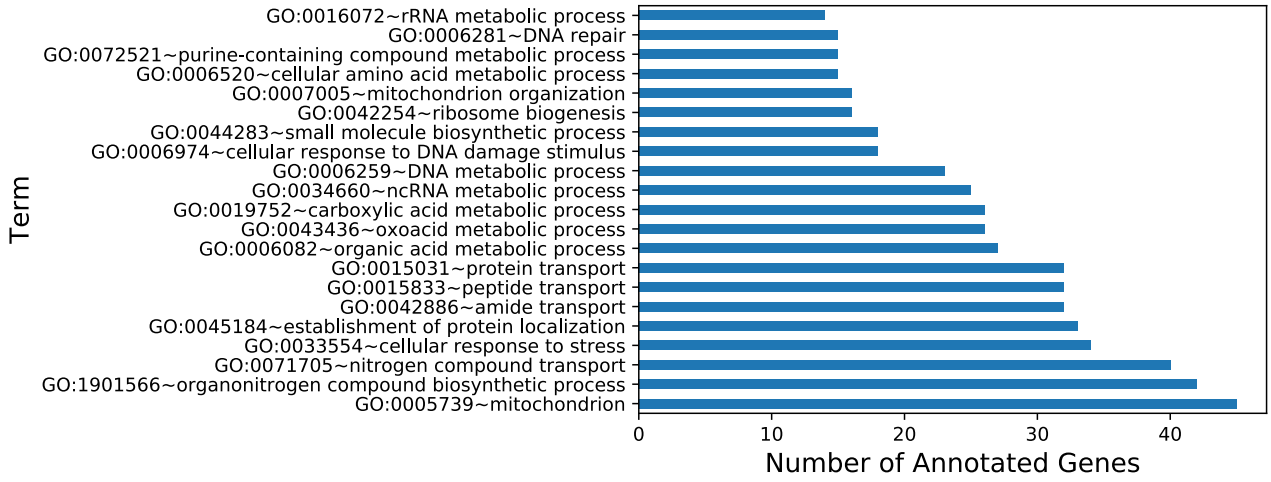
*The left boxplot (blue, Group 1) represents the distribution of average F_{st} values for non-outlier SNPs in 500bp windows containing at least one outlier SNP (threshold cut-off of SNPs with outlier levels of $F_{ST} \geq 0.11111$) with outliers removed (mean = 0.011 ± 0.010). The right boxplot (orange, Group 2) shows the distribution of average F_{st} values for windows without any outlier SNPs (mean = 0.009 ± 0.011). Significant differences were observed between Group 2 and Group 1, with t-statistics of 44.918 ($p < 0.0001$). (ns = not significant, *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p < 0.0001$).*



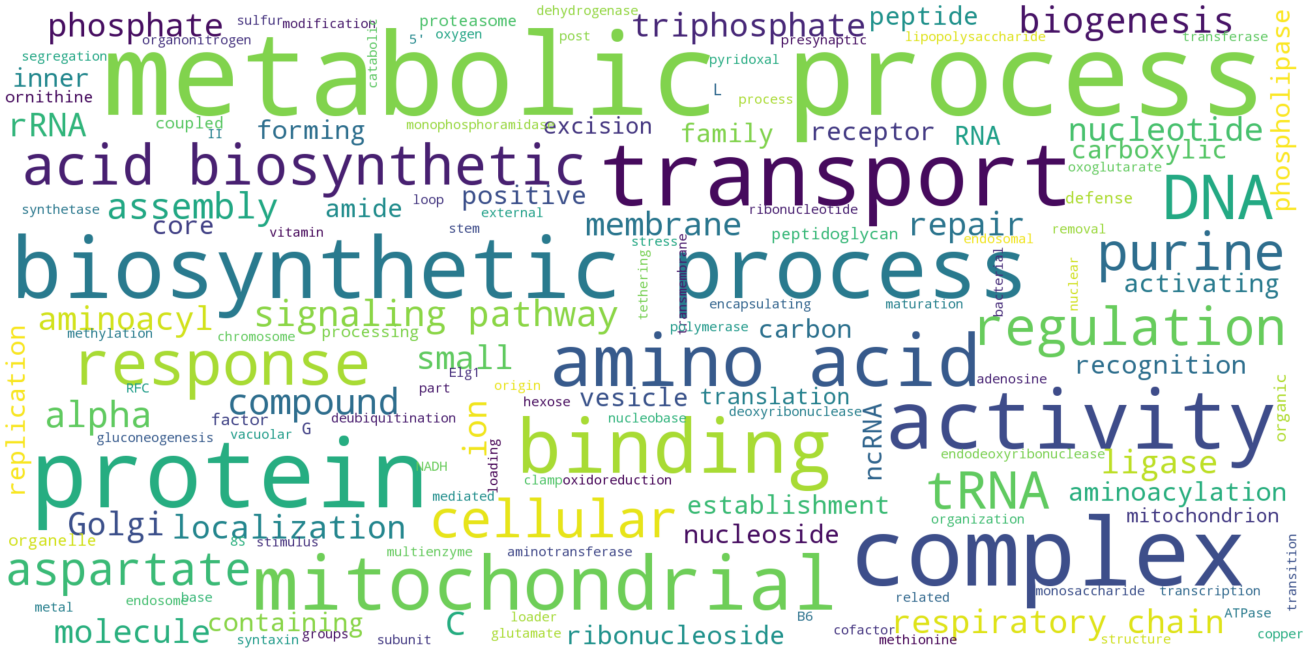
Supplemental Figure 4-3. *Enrichment of outlier SNPs (top 1%) across different mutation types in* C. calcarata.



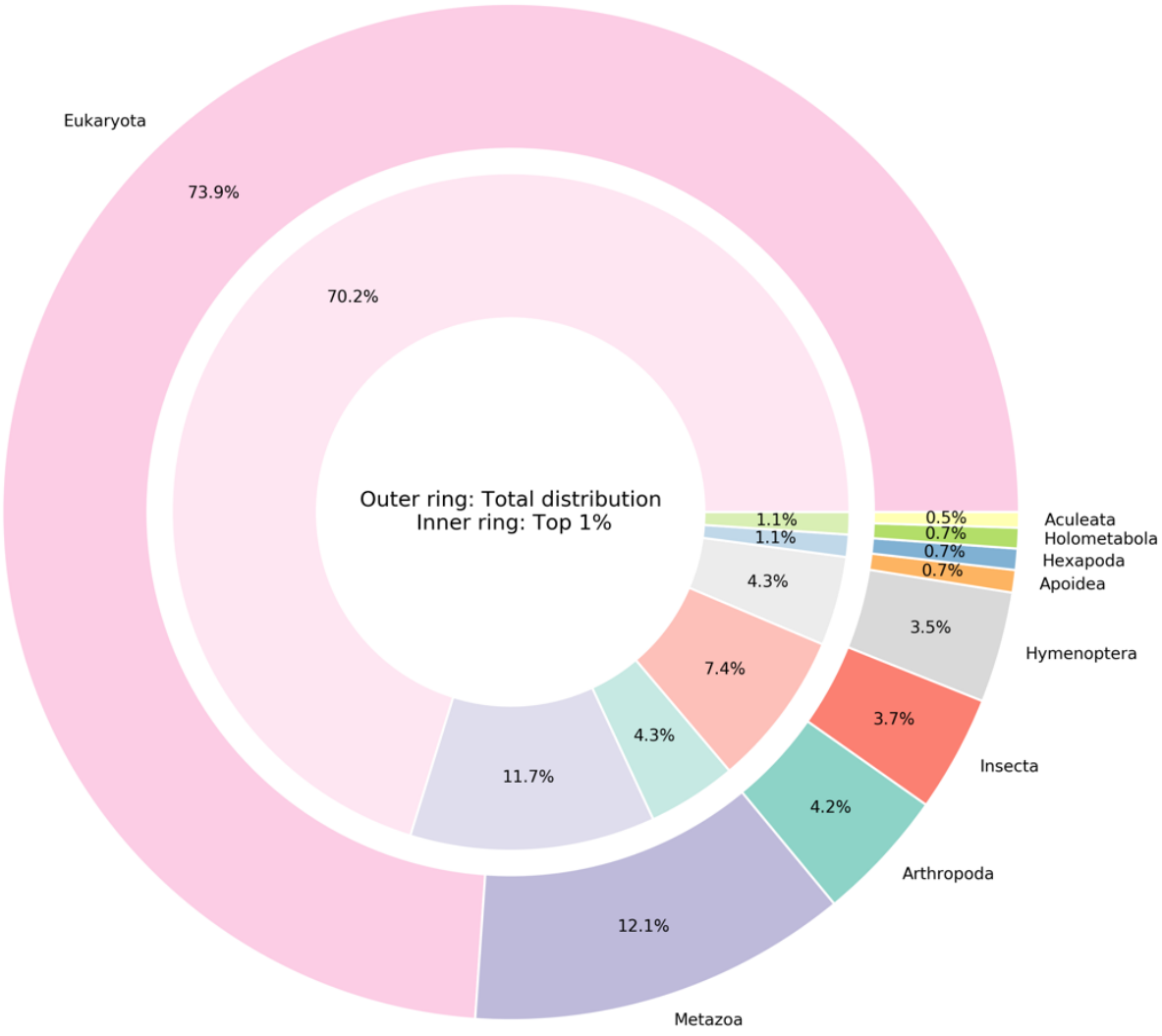
Supplemental Figure 4-4. *The type of exonic SNPs in the total distribution (outer rings) compared to the top 5% (S4A inner ring) and the top 1% (S4B inner ring).*



Supplemental Figure 4-5*. Gene Ontology enrichment analysis using DAVID GO found enrichment for genes related to metabolic processes, and biosynthesis of small molecules and amino acids (enrichment cluster score= 2.245).*



Supplemental Figure 4-6. *WordCloud generated from enrichment analysis of the top 5% of genes with high Fst associated with behaviour.*



Supplemental Figure 4-7*. Genes associated with social behaviour in* C. calcarata *are predominantly evolutionarily ancient, even within the top 1% of the distribution. (ns = not significant, *p ≤0.05, **p ≤0.01, ***p ≤0.001, ****p < 0.0001).*

Supplemental Tables

All supplemental tables can be found in the supplementary dissertation file

Chapter 5 A map of positive selection in the small carpenter bee (*Ceratina calcarata*) links gene regulation with social evolution

Dova B. Brenman- Suttner[1], Sandra M. Rehan[1], and Amro Zayed[1]

Summary

Understanding the evolution of eusocial behaviour involves elucidating the genetic mechanisms driving the transition from solitary behaviour to cooperative, altruistic behaviours. Small carpenter bees (*Ceratina* spp.) are an interesting group for studying the mechanisms underpinning the early stages of social evolution because they exhibit a wide range of social behaviours. For example, *Ceratina calcarata* and its close relative *C. strenua*, show some capacity for social behaviour but differ in their consistency and elaboration of their social organization. Comparing the genomes of these two closely related species may thus offer insights into the genetic differences that may have underpinned the evolution of sociality in *C. calcarata*. Here, we first identify genes with population genomic signatures of positive selection in *C. calcarata* to better understand the adaptive evolutionary forces acting during the early stages of eusocial evolution in this incipiently social bee. We used a Bayesian implementation of the McDonald-Kreitman test to identify 1,123 genes under strong positive selection. Genes with signatures of strong positive selection were both evolutionarily ancient and highly enriched for functions associated with gene regulation, consistent with several 'toolkit' hypotheses for the evolution of sociality. We then examined if any genes with signatures of positive selection also showed genetic differentiation between social and subsocial colonies in *C. calcarata*; we identified 20 such genes which are involved in signal transduction and chromatin remodeling. Comparisons with *Apis mellifera*, *Bombus impatiens*, and *Polistes dominula* revealed significant overlap in genes experiencing positive selection, and such genes were enriched for functions relating to metabolism and gene regulation. Taken together, our study supports the importance of regulatory genes during the evolution of primitive societies and emphasizes the conserved nature of social evolution mechanisms across species.

Introduction

Understanding the evolution of eusocial behaviour has been a central question for sociobiologists. Eusocial behaviour is characterized by overlapping generations, reproductive divisions of labour and cooperative brood care (Wilson 1971). The complex group of behaviours that characterizes eusociality has evolved independently at least eight times among insects, with five of these occurrences in bees (Hughes *et al.* 2008; Cardinal and Danforth 2011). Therefore, investigating the mechanisms underlying this convergent process is crucial for understanding the genetic basis of eusocial evolution (Woodard *et al.* 2011). The genetic toolkit hypothesis proposes that eusociality initially evolved by leveraging changes in the regulation of gene expression (True and Carroll 2002; Rehan and Toth 2015). This hypothesis proposes that reshuffling existing genes and gene networks through changes in when and where they are expressed can produce social behaviours and may represent one of the earliest pathways for the evolution of sociality (Rehan and Toth 2015). For example, changes in the regulation of reproductive genes can restrict reproductive behaviours to specific individuals, thereby creating reproductive and non-reproductive castes (Favreau *et al.* 2023). Support for this hypothesis has come from large comparative studies of bees (Woodard *et al.* 2011; Kapheim *et al.* 2015) as well as studies within honey bees (Grozinger *et al.* 2003; Grozinger *et al.* 2007; Harpur *et al.* 2019) and fire ants (Ometto *et al.* 2011; Berens *et al.* 2015). Another theory, the novel gene hypothesis (Johnson and Tsutsui 2011), suggests a different mechanism by which social behaviour can evolve; this hypothesis posits that novel (i.e. taxonomically-restricted) genes underpin derived eusocial traits. Evidence in support of the novel gene hypothesis appears in wasps (Ferreira *et al.* 2013) and honey bees (Jasper *et al.* 2015).

Studies of positive selection in species displaying social behaviours have the potential to highlight key genes and genetic processes associated with the evolution of eusociality. If eusocial behaviour is adaptive and has evolved through positive selection, then identifying genes with signatures of positive selection at specific evolutionary transitions associated with the origin or elaboration of sociality can provide insights into the genes that contributed to the evolution of eusociality (Hasselmann *et al.* 2015; Kent and Zayed 2015; Booker *et al.* 2017). We can then identify the molecular and biological functions of these genes and begin to understand the genetic and functional basis required for eusociality to evolve and increase in complexity. However, most studies investigating the evolution of eusocial behaviour have focused on insects that have been eusocial for millions of years such as honey bees (Hasselmann and Beye 2004; Zayed and Whitfield 2008; Chavez-Galarza *et al.* 2013; Harpur *et al.* 2014; Fuller *et al.* 2015; Huang *et al.* 2022b; Jones *et al.* 2023) and ants (Roux *et al.* 2014; Chandra *et al.* 2018; Romiguier *et al.* 2022). In

contrast, studying species with more flexible social lifestyles or species that transitioned to sociality or more complex forms of sociality more recently, where positive selection may still be acting on beneficial polymorphisms for social living, can provide valuable insight into the processes and mechanisms of social evolution. For example, *Lasioglossum albipes* is a socially polymorphic halictid bee that can display both solitary and eusocial colony behaviour (KOCHER *et al.* 2013). One study found accelerated evolution in genes related to metabolism and nucleotide binding (KOCHER *et al.* 2013). Similarly, a study in the facultatively social *Megalopta genalis* found that genes underlying social behaviour are enriched for signatures of positive selection and are often the same genes involved in developmental and sex-biased gene expression (KAPHEIM *et al.* 2020a). These results suggest that positive selection on social traits may act on pre-existing developmental gene networks, showing the link between adaptive changes in the genome and behavioural plasticity seen in species transitioning to more advanced social behaviours.

Ceratina calcarata is an excellent model for studying the mechanism underlying the transition to more complex social behaviours. *C. calcarata* can exhibit both a subsocial and social behavioural state. During its subsocial phase, mothers will perform all the necessary colony tasks such as caring for brood, foraging and guarding the nest (REHAN AND RICHARDS 2010b; REHAN *et al.* 2016). In some nests, mothers will create a small worker daughter called the dwarf eldest daughter (DED) through differential mass provisioning, who will then act as an altruist and aid in caring for her siblings but does not survive the winter and will therefore not have the opportunity to reproduce (LAWSON *et al.* 2016; LAWSON *et al.* 2017). Because *C. calcarata* can exhibit either of these behaviours, it is a valuable model for studying the transition to more complex forms of social behaviour, while also controlling for evolutionary history and variation in the environment. By comparing the genomes of *C. calcarata* and a closely related outgroup species, *Ceratina strenua* (divergence 92,000 years ago), which shares some capacity for social behaviour but lacks the stable, integrated social organization characteristic of *C. calcarata* (LAWSON *et al.* 2018), we can identify genes under positive selection in *C. calcarata*. These genes may represent adaptive changes underlying the increase in social complexity of *C. calcarata*.

Previous studies of positive selection in *Ceratina calcarata* carried out using dn/ds-based tests for 11,310 genes, indicated that 877 protein-coding genes exhibited signatures of positive selection, including genes related to DNA recombination and lipid transport (REHAN *et al.* 2016). This study used a PAML-based analysis which can be used to detect older instances of positive selection that happened across longer evolutionary time scales along a lineage. PAML also relies on orthology for statistical tests, which can limit its effectiveness in analyzing a broader range of protein-coding genes, particularly those that are recent and may lack orthologs in other species (MACIAS *et al.* 2020). In this study, we test for genes under positive

selection using the McDonald-Kreitman (MK) test (McDonald and Kreitman 1991), which is known to be more robust to demographic changes (Neilson 2001). By using the Bayesian MK-based method SnIPRE, we can estimate the effective population size-scaled selection coefficient ($\gamma = 2N_e s$) for the vast majority of both ancient and taxonomically-restricted genes, allowing us to quantify selective pressures acting during the social evolution of *C. calcarata* relative to its outgroup, *C. strenua*.

In this study, we identified protein-coding genes under positive selection within *Ceratina calcarata* with *Ceratina strenua* as an outgroup using a Bayesian implementation of the MK test. We then investigated the putative function of genes under strong positive selection using gene ontology. We hypothesize that genes under strong positive selection will be involved in the regulation of gene expression (Rehan and Toth 2015). We also studied the orthology of genes with signatures of positive selection to determine if adaptive evolution in *C. calcarata* is primarily driven by selection on derived ("new") genes rather than ancestral ("old") genes. We then compared genes under strong positive selection to genes that were highly divergent (high F_{st}) in subsocial versus social *C. calcarata* that we identified in a previous study (Brenman-Suttner *et al.* 2025) to connect adaptive evolution to the social behaviour observed in this species. Finally, we compare *C. calcarata* genes under positive selection with those in *Apis mellifera*, *Bombus impatiens*, and *Polistes dominula* (Dogantzis *et al.* 2018) to test the hypothesis that species experiencing similar social behaviour will share similar genes with signatures of strong positive selection.

Methods

Sample collection

Twenty diploid female *Ceratina calcarata* were collected from nests in New Hampshire, USA and flash frozen in liquid nitrogen. We used a similar sample size (20 individuals, n=40 haploid genomes) to other studies on genome wide assessments of selection (Eyre-Walker *et al.* 2002; Harpur *et al.* 2014; Molodtsova *et al.* 2014; Dogantzis *et al.* 2018; Privman *et al.* 2018). The McDonald-Kreitman test requires divergence data from a closely related species to test for selection, so we sequenced two diploid *Ceratina strenua* (2 females, n=4 haploid genomes) that were collected in Missouri, USA and stored in EtOH, and which diverged from *C. calcarata* approximately a 92,000-year years ago (Shell and Rehan 2015) and, while capable of producing worker daughters, exhibits a less consistent and less elaborated social phenotype than *C. calcarata* (Lawson *et al.* 2018), making it an appropriate outgroup for identifying selection specific to *C. calcarata*. By comparing two closely related species that differ in social behaviour, the MK test allows us to identify genes that may have experienced positive selection over the same evolutionary time frame in which sociality became more complex in *C. calcarata*.

Representative voucher specimens of *Ceratina calcarata* (n = 20) were collected via pan trap in Farmington, New Hampshire, USA, on July 20, 2018. Specimens were identified by S. Rehan in 2018 and deposited in the Rehan Collection at York University (RCYU), Toronto, Canada. Specimens used for genomic extraction and sequencing were flash-frozen in liquid nitrogen and shipped on dry ice, while associated physical vouchers are maintained as pinned specimens under accession number 181388. Each specimen bears a permanent voucher label detailing the locality (USA, NH, Farmington, GPS coordinates), collection date, and collector name as per the requirements of RCYU. Representative voucher specimens of *Ceratina strenua* (n = 2) collected in Missouri, USA, were deposited in the Rehan Collection at York University (RCYU). These specimens were preserved in ethanol. Specimens were identified by S. Rehan. Each sample container includes a permanent voucher label detailing the locality (Country, State, County, and GPS coordinates), collection date, and collector name, as per the requirements of RCYU.

Sequencing, alignment, calling and filtering SNPs

We extracted DNA from the thorax of each bee using optimized protocols (Conflitti *et al.* 2022) with the Mag-Bind® Blood & Tissue DNA HDQ 96 Kit (Omega Bio-tek Inc., USA). This kit was optimized for the KingFisher™ Flex Purification System (Thermo Fisher Scientific Inc., USA) to achieve a final elution volume of 70-100 µl. We used 350 µl of Tissue Lysis Buffer and 20 µl of Proteinase K to lyse tissue and then heated the samples overnight at 55°C. DNA quantification and quality assessment were performed using NanoDropTM 2000 Spectrophotometer (Thermo Fisher Scientific Inc., USA) and 2.0% agarose gel electrophoresis, respectively. Library preparation and sequencing were conducted at the Genome Quebec Innovation Centre (McGill University) on the Illumina HiSeq 2500 system, generating a total coverage of (27X per bee) for 364Mb *Ceratina calcarata* genome. We performed quality control on all samples to assess the read length, quality and other standard parameters using Qualimap2 (Okonechnikov *et al.* 2016). We employed previously published pipelines for alignment and SNP calling (Harpur *et al.* 2019; Imrit *et al.* 2020; Harpur and Rehan 2021). Briefly, raw reads were trimmed with TRIMMOMATIC v0.32, aligned to the *Ceratina calcarata* ASM165200v1 genome (Rehan *et al.* 2016) using NextGenMap v0.4.12 (Sedlazeck *et al.* 2013), and duplicates were removed with Picard v1.8 MarkDuplicates (http://broadinstitute.github.io/picard/). Indel realignment was performed with Genome Analysis Toolkit (GATK v3.5)(Van der Auwera *et al.* 2014). Single nucleotide polymorphisms (SNPs) were identified using GATK HaplotypeCaller through an existing pipeline (Dogantzis *et al.* 2018).

Samples were filtered based on GATK (v3.5) recommendations (MappingQuality < 20, QualitybyDepth < 2.0, FisherStrand > 60, StrandOddsRatio > 3.0, ReadPosRankSum < -8.0, MappingQualityRankSum <-12.5, MappingQualityRankSum > 12.5)(Van der Auwera *et al.* 2014). SNPs were

further filtered according to read depth, with upper and lower limits determined as 1.5x the interquartile range and an average coverage threshold of 5 reads per base pair per individual, respectively (KAPHEIM *et al.* 2019; IMRIT *et al.* 2020). Low-frequency SNPs (minor allele frequency < 0.05) were also excluded from the sample set (KAPHEIM *et al.* 2019) as well as sites with more than 30% missing data or multi-allelic sites with several alternative alleles (genotypes '0/2', '0/3', '1/2','1/3','2/3', '2/2','3/3')(MOHD-ASSAAD *et al.* 2018). Multi-allelic sites often arise in areas of the genome that are complex (such as repetitive regions)(YAO *et al.* 2021) which can be prone to sequencing or alignment errors and thus are considered to be less reliable (CAMPBELL *et al.* 2016).

Identifying SNPs with synonymous and nonsynonymous changes and the MK Test

To carry out a McDonald-Kreitman test using SnIPRE, we first identified all synonymous and nonsynonymous changes within the protein-coding sequences of the *Ceratina calcarata* reference genome. This analysis requires both the expected number of synonymous and nonsynonymous sites per gene and the observed synonymous and nonsynonymous SNPs. The expected number of synonymous and nonsynonymous sites establishes a baseline for the expected number of synonymous and nonsynonymous changes for each coding sequence, which is important for accurately detecting of selection. To estimate the expected number of synonymous and nonsynonymous sites, we downloaded all protein coding regions from NCBI *Ceratina calcarata* genome annotation file (https://ftp.ncbi.nlm.nih.gov/genomes/all/annotation_releases/156304/101/). We then parsed from the CDS FASTA file using custom bash/python scripts to divide them into codons. Briefly, coding sequences were extracted from the FASTA headers, genes lacking a start codon ('ATG') were removed, and for genes with multiple transcript isoforms we retained the longest coding sequence whose length was divisible by three. The remaining nucleotide sequences were then divided into consecutive, non-overlapping three-base triplets beginning at the first base of the coding sequence, such that each CDS was partitioned into codons for downstream site-counting analyses. We also used SnpEff (CINGOLANI *et al.* 2012) to help identify any SNP within a protein coding sequence that was missing a stop codon, contained multiple stop codons or identify where the reference sequence did not match the genome, and remove these from the dataset. For the remaining genes, we estimated the expected number of synonymous and nonsynonymous sites using the counting method (WRIGHT 1992). Briefly, each codon position was assigned a probability that a single polymorphism would result in a synonymous or non-synonymous change. These probabilities were then summed (total synonymous changes and total nonsynonymous changes) across each protein coding sequence.

Identifying genes under positive selection

We used SnpEff (Cingolani *et al.* 2012) to determine if SNPs within protein coding genes were a synonymous change (no change to the amino acid made, S) or a non-synonymous change (change to the amino acid made, R) for both the target (*Ceratina calcarata*, n=40) and outgroup (*Ceratina strenua*, n=4). Variants that were labeled as 'synonymous_variant' were considered synonymous mutations whereas variants labeled 'missense_variant', 'stop_retained_variant', 'initiator_codon_variant' and 'nonsynonymous_variant' were considered nonsynonymous mutations in accordance with their definitions within the SnpEff documentation (Cingolani *et al.* 2012). We then identified if mutations were fixed (F) meaning that all *C. calcarata* SNPs had one variant whereas all *C. strenua* had a different variant, or polymorphic (P) where either species has a mix of variants. We then combined this information and categorized each SNP as fixed or polymorphic, resulting in four categories: fixed synonymous (FS), polymorphic synonymous (PS), fixed nonsynonymous (FR), and polymorphic nonsynonymous (PR). We added this up across the protein coding gene to get the total number of FS, FR, PS, and PR SNPs per coding region. This, combined with the total predicted synonymous and nonsynonymous changes found above, formed the input table for testing positive selection.

The McDonald-Kreitman test is a powerful method to detect selection by comparing nucleotide diversity within and between species at synonymous (neutral) and non-synonymous (selected) mutations (McDonald and Kreitman 1991; Booker *et al.* 2017). To estimate selection, we used SnIPRE, a Bayesian implementation of the McDonald-Kreitman test (Eilertson *et al.* 2012). This approach utilizes genome-wide data without requiring prior knowledge of divergence parameters to estimate gamma (γ), the average population-scaled selection coefficient on non-synonymous mutations in a gene ($\gamma = 2Nes$)(Eilertson *et al.* 2012). Typically, positive selection acting on a gene leads to higher a ratio of non-synonymous to synonymous divergence (or fixed mutations) (FR/FS) relative to the ratio of non-synonymous to synonymous polymorphism (PR/PS). A gamma (γ) value of zero, or below -1 or 1, indicates neutral, or nearly-neutral evolution, respectively (Torgerson *et al.* 2009) and $\gamma > 1$ or < -1 indicate positive or negative selection, respectively.

Enrichment for genes under positive selection

We used HymenopteraMine QueryBuilder to identify orthologous *Drosophila melanogaster* GeneIDs to *C. calcarata* genes in our dataset (Walsh *et al.* 2022). We then we used Flybase ID validator (https://flybase.org/convert/id)(Gramates *et al.* 2022) to identify the *D. melanogaster* genes in the form of "FBgn" to be used for ontology analysis. All orthologs were used as "background genes" for gene ontology

enrichment analysis and orthologous genes exhibiting positive selection were used as the "target gene" group. We used DAVID GO 2022 (Sherman *et al.* 2022) to investigate biological processes (GOTERM_BP_FAT), molecular functions (GOTERM_MF_FAT), and cellular components (GOTERM_CC_FAT) associated with genes under strong positive selection ($\gamma > 1$). We then used GOrilla to visualize the relationship between enriched genes among genes with strong positive selection (Eden *et al.* 2009).

Categorizing genes based on genomic phylostratigraphy

The OrthoDB v11 database categorizes genes and proteins based on taxonomic levels (Kuznetsov *et al.* 2022). Flat files were downloaded from the OrthoDB site, filtered for *Ceratina calcarata* annotations (NCBI taxonomic ID: 156304), and merged based on common column criteria to focus on relevant orthologs for *C. calcarata*. The total number of orthologs for each gene was quantified and the lowest, most stable taxonomic classification was used for gene categorization. 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 overlapped this total list of genes, including their taxonomic classification, with our list of genes containing a selection value, focusing on those under strong positive selection ($\gamma > 1$). A G test of independence was performed to determine enrichment among genes with strong positive selection relative to the entire distribution at the levels of Apoidea and Aculeata.

Overlap genes under strong positive selection with other studies

Genes under strong positive selection were compared to SNPs exhibiting high Fst associated with behaviour (Brenman-Suttner *et al.* 2025) to find genes with both positive selection and high differentiation between social and subsocial behavioural groups. We plotted the overlap of genes with high Fst associated with behaviour and genes under positive selection using venn3 for Python 2.0 (Hunter 2007). We then converted the overlapping genes to *Drosophila melanogaster* genes and performed gene ontology analysis using DAVID GO (Sherman *et al.* 2022).

We compare genes under positive selection found here with those previously found to be under positive selection in *Apis mellifera* (advanced eusocial)(Michener 1974), *Bombus impatiens* (primitively eusocial) (Plath 1934), and *Polistes dominula* (primitively eusocial) (West-Eberhard 1969) quantified using the same MK methods used herein. We used Hymenopteramine (https://hymenopteramine.rnet.missouri.edu/hymenopteramine/begin.do) to find orthologs of *Ceratina calcarata* in other species. We then overlapped our list of genes under strong positive selection ($\gamma > 1$) with a previous study that investigated genes under positive selection in pairs of species (Dogantzis *et al.* 2018)

with our *C. calcarata* genes under positive selection. We used venn3 for Python 2.0 (HUNTER 2007) to make Venn diagrams of overlapping genes under positive selection among pairs and groups of species.

We used the Flybase ID validator (https://flybase.org/convert/id) (GRAMATES *et al.* 2022) to convert overlapping genes to *Drosophila melanogaster* IDs for gene ontology enrichment analysis using DAVID GO (SHERMAN *et al.* 2022). This allowed us to investigate the function of genes under strong positive selection in *Ceratina* that overlap with *Polistes*, *Apis*, and *Bombus*. We then visualized the results using a bar chart using matplotlib (HUNTER 2007). For overlap analyses with too few genes for ontology analysis, we identified gene descriptions using NCBI (https://www.ncbi.nlm.nih.gov). We also assessed the significance of the overlap using a Fisher's exact test.

Results

Genes under positive selection in C. calcarata

Using a Bayesian implementation of the McDonald-Kreitman test (SnIPRE), we were able to estimate selection coefficients scaled by the effective population size, gamma (γ), for 8,060 genes of *Ceratina calcarata*. We found an average γ of 0.550 ± 1.781 within the nearly-neutral range. We found 1,123 genes under strong positive selection and 3 genes with strong negative selection ($\gamma \leq -1$; Figure 5-1, Supplemental Table S5-1). For further analysis, we only consider genes under strong positive selection as the target for comparison and ontology analysis.

Gene ontology of genes under positive selection

We were able to identify *Drosophila melanogaster* orthologs for 5,673 of the *C. calcarata* protein-coding genes with a γ value, representing 70.38% of our dataset (Supplemental Table S5-2). Among these, 4,453 orthologs evolved at nearly-neutral rates, while 835 genes were under strong positive selection. We conducted gene ontology clustering analysis on the orthologs with strong positive selection and found significant annotations in the highest cluster related to the regulation of transcription, RNA metabolic processes, DNA binding, transcription factor activity, and biosynthetic processes (enrichment score = 7.177; Supplemental Table S5-3). The most common annotations included regulation of gene expression, DNA and RNA processes, biosynthetic, metabolic, and developmental processes, as well as binding activities and regulation of cellular activities (Figure 5-2, Supplemental Table S5-4).

Genes under positive selection in C. calcarata *are evolutionarily ancient*

A total of 5,872 genes were analyzed across all taxa, of which 833 showed evidence of positive selection at the level of Eukaryota. Within Metazoa, 111 out of 908 genes in *C. calcarata* were under positive

selection. Arthropoda contained 337 genes, with 52 showing positive selection, while Insecta comprised 292 genes, of which 57 were positively selected. Within Hymenoptera, 32 of 292 genes showed signatures of positive selection, representing a smaller proportion than observed in broader insect and arthropod groups. Holometabola included 47 genes, with six under positive selection, while Hexapoda contained 48 genes, of which 13 were positively selected. In contrast, the most derived groups showed very limited evidence of positive selection. Aculeata contained 37 genes, but only one was under positive selection. No positively selected genes were detected in Apoidea among the 55 genes analyzed.

Statistical tests showed that several ancient evolutionary groups exhibited significant deviations from random expectations in the distribution of positively selected genes (Eukaryota ($G = 894.444$, $p = 1.584 \times 10^{-196}$), Metazoa ($G = 116.106$, $p = 4.506 \times 10^{-27}$), Arthropoda ($G = 54.201$, $p = 1.810 \times 10^{-13}$), Hexapoda ($G = 12.812$, $p = 3.444 \times 10^{-4}$), Insecta ($G = 60.968$, $p = 5.800 \times 10^{-15}$), Holometabola ($G = 4.451$, $p = 3.489 \times 10^{-2}$). However, Hymenoptera showed a significant depletion of positively selected genes rather than enrichment ($G = 31.772$, $p = 1.734 \times 10^{-8}$). Evolutionarily younger groups, however, were either not significant such as in Aculeata ($G = 0$, $p = 1.000$), or did not have any genes under positive selection, as in the case of Apoidea (Figure 5-3; Supplemental Table S5-5). Together, these results indicate that positive selection is more commonly detected in older taxonomic groups and is rare in the most recently derived lineages.

Genes associated with social behaviour are adaptively evolving in Ceratina calcarata

A recent study identified genes with outlier levels of genetic differentiation (top 5% of the distribution) among social and subsocial *Ceratina calcarata* (Brenman-Suttner *et al.* 2025). We examined the genes present in the intersection of both gene sets (7,798 genes) and then filtered the dataset to include only genes under strong positive selection and genes with outlier levels of genetic differentiation (F_{st}) among social and subsocial behaviour groups in *C. calcarata* and we identified 20 genes (Table 5-1, Supplemental Table S5-6). This overlap was also significant based on a chi-square test (chi2 statistic = 360.41, $p = 2.29^{-80}$). Notable genes with high Fst and γ values include ubiquitin thioesterase otubain-like ($\gamma = 6.667$, $F_{st} = 0.052$), proline-rich protein 2-like ($\gamma = 4.358$, $F_{st} = 0.054$), and facilitated trehalose transporter Tret1-like ($\gamma = 3.706$, $F_{st} = 0.063$). Gene ontology enrichment analysis of these overlapping genes revealed significant enrichment for regulation of signal transduction and communication, histone acetyltransferase components, and chromatin remodeling.

Gene under positive selection in Apis, Bombus, Polistes *and* Ceratina

We identified orthologous genes under strong positive selection in *Ceratina calcarata*, *Apis mellifera*, *Bombus impatiens*, and *Polistes dominula*. We found 5,297 reciprocal blast hits with *Apis mellifera*, of which, 54 had strong positive selection in both *Ceratina* and *Apis*, which is considered a significant ($p= 1.20^{-4}$) overlap using a fisher's exact test (Supplemental Table S5-7). We found 5,675 orthologs between *Ceratina calcarata* and *Bombus impatiens*, with 51 overlapping genes that were under strong positive selection in both species, which was almost a significant overlap (p= 0.067). There were 1,815 orthologs between *Polistes dominula* and *Ceratina calcarata* but only seven genes were found to be under strong positive selection in both species (p=0.439).

We tested the hypothesis that genes under positive selection in *Ceratina calcarata* would show greater overlap with genes under positive selection in *Bombus impatiens* and *Polistes dominula* compared to *Apis mellifera*. This hypothesis stems from the idea that positive selection is more likely to act on similar sets of genes in species that share similar behavioural traits, such as sociality, rather than those that are more closely related genetically (Dogantzis *et al.* 2018). We found more genes than expected in the overlap between *Apis* and *Ceratina* ($p=1.19^{-4}$) but not in the overlaps of either *Polistes* and *Ceratina* (p=0.439) or *Bombus* and *Ceratina* (p=0.067) (Figure 5-4). We performed gene ontology enrichment analysis using the overlapping genes under strong positive selection compared to all genes that overlapped between *Ceratina calcarata* and *Apis mellifera*, *Bombus impatiens*, and *Polistes dominula* individually, to determine their function but only report the results of *Apis* and *Ceratina* as the overlap was significant. For the *Ceratina* and *Apis* overlap, we found that genes were primarily involved transmembrane transport, amino acid metabolism, cilium and binding of heme and its components (Supplemental Table S5-8A). We also looked for common genes in all other possible groupings of these 4 species and while we did find some genes that overlapped in groups of 3, these overlaps were not significant, and no genes were found to be under strong positive selection across all four species. We identified 5 genes under strong positive selection across *Apis*, *Bombus* and *Ceratina* (Supplemental Table S5-9). These genes were found to be involved in signaling such as neurotrimin-like and atrial natriuretic peptide receptor 1 and feline leukemia virus subgroup C receptor-related protein 2-like, in addition to the gene mitochondrial uncoupling protein 2-like which is important for ROS mediation (Basu Ball *et al.* 2011) and a gene mitotic apparatus protein p62-like which is important for cell division (Ye and Sloboda 1997). We also found one gene, dual oxidase, to be similarly under strong positive selection in *Apis*, *Polistes* and *Ceratina*. This gene is known to be involved in innate immunity in generating reactive oxygen species to combat pathogens and structural maintenance (Jang *et al.* 2021). We did not find any significant overlap in the number of genes under strong positive selection.

Discussion

In this study, we quantified the strength of positive selection acting on the protein coding genome of *Ceratina calcarata* since its divergence from *C. strenua*. Phylogenetic reconstructions indicate that ancestral *Ceratina* was likely social (Groom and Rehan 2020), suggesting that the divergence between these species reflects a difference in the degree to which sociality has been retained or elaborated. Identifying genes under positive selection in *C. calcarata* relative to *C. strenua* therefore provides a window into the molecular changes accompanying the increase in social complexity within an already social lineage. We found 1,123 genes under strong positive selection. These genes were enriched for annotation clusters relating to regulation of transcription and gene expression, supporting the several toolkit-type hypotheses that posit that regulatory flexibility plays an important role in the diversification of behaviour and in the early stages of eusocial evolution (True and Carroll 2002; Carroll *et al.* 2005; Rehan and Toth 2015).

Many genes under strong positive selection were enriched for annotations related to the regulation of gene expression and transcription. The regulation of gene expression has long been proposed as a crucial mechanism for early social evolution, allowing behavioural divergence to emerge from shared genomes (West-Eberhard and Turillazzi 1996; Linksvayer and Wade 2005). In *C. calcarata*, this type of regulation may encourage the behavioural flexibility that is required for subsocial individuals to express more social traits. These findings support theoretical and empirical studies suggesting that the evolution of eusociality is underpinned by adaptive shifts in gene regulation, which have been corroborated at the genomic level (Brenman-Suttner *et al.* 2025) and at the transcriptomic level (Rehan *et al.* 2014; Shell and Rehan 2019) in *C. calcarata* as well as in other species such as *Polistes dominula* (Dogantzis *et al.* 2018). We also identified 20 genes under strong positive selection with highly divergent SNPs between the social and subsocial behavioural groups within *C. calcarata*. These genes were enriched for annotations related to histone acetyltransferase complexes and chromatin remodeling. Chromatin remodeling is an important epigenetic mechanism that can regulate gene expression by allowing reversible modifications that help adapt behaviours to social contexts (Kribelbauer *et al.* 2020; Richard *et al.* 2021). These mechanisms have previously been shown to enable workers and queens, who share the same genome, to express different behavioural phenotypes at different times and in different locations in the body to yield caste-specific behaviours (Yan *et al.* 2015; Cardoso-Junior *et al.* 2018), as in *Camponotus floridanus*, where the chromatin corepressor CoREST epigenetically reprograms Major workers into Minor-like foragers and is therefore important for regulating the division of labour (Glastad *et al.* 2020). Several studies across social insects have also highlighted the role of epigenetic regulation in task specialization (Lyko *et al.* 2010; Foret *et al.*

2012; Araujo and Arias 2021) and maternal care (Rehan *et al.* 2016; Arsenault *et al.* 2018). Therefore, selection on genes involved in chromatin remodeling and histone modification in *C. calcarata* represents an important regulatory route to behavioural evolution in this species, supporting the genetic toolkit hypothesis that emphasizes regulatory flexibility in early social transitions (Woodard *et al.* 2011).

Genes under strong positive selection that also exhibit outlier levels of genetic differentiation between subsocial and social *Ceratina calcarata* were also enriched for signal transduction and its regulation, which have also been implicated in the molecular evolution of social behaviour. For example, a comparative genomic study analyzed 10 bee genomes and found that species with more complex sociality showed an increased abundance of transcription factor binding motifs and signatures of DNA methylation, suggesting a potential greater capacity for gene regulation (Kapheim *et al.* 2015). The same study also found that genes evolving more rapidly in advanced social species were enriched for functions related to signal transduction (Kapheim *et al.* 2015). Furthermore, another study of expressed sequence tags (ESTs) from 10 bee species with varying degrees of social behaviour revealed that signal transduction was among the fast-evolving biological processes in eusocial lineages (Woodard *et al.* 2011). This enrichment may reflect the importance of chemical communication in social evolution such as the detection of pheromones (Fischman *et al.* 2011; Van Oystaeyen *et al.* 2014), which is necessary in group living environments to prevent conflict, promote cooperation (Gonzalez-Forero and Pena 2021), recognize kin (Getz and Page 1991), and, in more advanced social species, allow reproductive individuals to suppress reproduction by subordinates (Strauss *et al.* 2008). In the socially polymorphic *Halictus rubicundus*, differences in chemical profiles among social and solitary groups also suggested the importance of changes in sensory perception and signaling for different behaviours (Steitz *et al.* 2021). The strong selection on and differentiation of signal transduction genes in *C. calcarata* that we found therefore points to both communication and behavioural plasticity as important adaptive features in the evolution of sociality.

We categorized all genes based on their most recent common ancestor using Orthology to understand the evolutionary origins of genes under strong positive selection. Most genes under strong positive selection in *C. calcarata* are evolutionarily ancient, with homologues found across broad taxonomic groups including Eukaryota, Metazoa, Arthropoda, Hexapoda, Insecta, Holometabola. In contrast, there was no significant enrichment for genes restricted to the Hymenoptera, Aculeata, and no genes under strong positive selection were found to be specific to Apoidea. Furthermore, we found that the majority of our genes for which selection could be estimated (70.38%) had a *Drosophila melanogaster* ortholog, suggesting that conserved genes play an important role in the evolution of sociality in this species. Taken together, our results show that the Orthology analysis does not support the novel gene hypothesis for the evolution of

sociality in *C. calcarata* at this time. Instead, the pattern is consistent with the genetic toolkit hypothesis, where ancient, conserved genes are repeatedly co-opted and regulated in new ways to generate social traits. This pattern is consistent with earlier transcriptomic studies in *C. calcarata*, which revealed that differentially expressed genes are also predominantly of ancient evolutionary origin (SHELL *et al.* 2021). This conclusion contrasts with work in *Apis mellifera* where many genes under positive selection are taxonomically restricted to more recent lineages such as Apoidea (HARPUR *et al.* 2014). Our findings of conserved, ancient genes being important for sociality in *C. calcarata* align with hypotheses that changes in gene regulation are crucial for early social transitions in species with less complex social behaviour.

We compared genes under strong positive selection in *Ceratina calcarata* to those identified in three other bee species including *Apis mellifera*, *Polistes dominula*, and *Bombus impatiens* (DOGANTZIS *et al.* 2018). We surprisingly found significantly more overlap than expected between genes under strong positive selection in *C. calcarata* and *A. mellifera*, despite these two species having the most divergent social phenotypes. In contrast, overlaps between *C. calcarata* and either *B. impatiens*, which is as closely related to *C. calcarata* and show primitive sociality, or *P. dominula*, which is distantly related but also exhibits similar levels of social complexity, were not significant. This pattern differs from Dogantzis et al. (2018), who found greater overlap between *Bombus* and *Polistes*. These results raise questions about the mechanisms underlying eusocial evolution even among closely related bees. Although *Ceratina* and either *Apis* or *Bombus* diverged ~105 million years ago (REHAN *et al.* 2016), they still share different sets of genes under strong positive selection, which could also reflect different environmental pressures on each species. In the overlap among *Ceratina* and *Apis*, enrichment was found for genes involved in transmembrane transport, amino acid metabolism, and heme-related binding and cilium-associated functions. We also identified only one gene (LOC108628573, dual oxidase) shared among *Apis*, *Polistes*, and *Ceratina*, a gene previously linked to innate immune function via reactive oxygen species production. This pattern fits with the idea that more social lineages often rely less on broad immune gene repertoires and more on social immunity behaviours (BARRIBEAU *et al.* 2015), which may relax selection on some individual innate immune genes. Together, these results indicate that although there are isolated convergent signals of positive selection across bee lineages, the identity and function of the genes under strong positive selection are not broadly conserved across species, even among those with similar social complexity.

Our findings reveal that adaptive evolution in *Ceratina calcarata* is primarily driven by positive selection acting on ancient, conserved genes involved in gene regulation, chromatin remodeling and signal transduction; molecular functions important for behavioural plasticity and social communication. Our findings are consistent with tool-kit hypotheses that propose that social evolution involves repurposing

regulatory and signaling networks found in solitary ancestors. We also identified genes that were both adaptively evolving and highly differentiated among behavioural groups within *C. calcarata*; these provide a list of candidate genes for the evolution of eusociality in *C. calcarata*. Together, these findings highlight conserved mechanisms of eusocial evolution and reinforce *C. calcarata* as a valuable model for studying the genetic basis of social complexity in bees.

‘

Figures

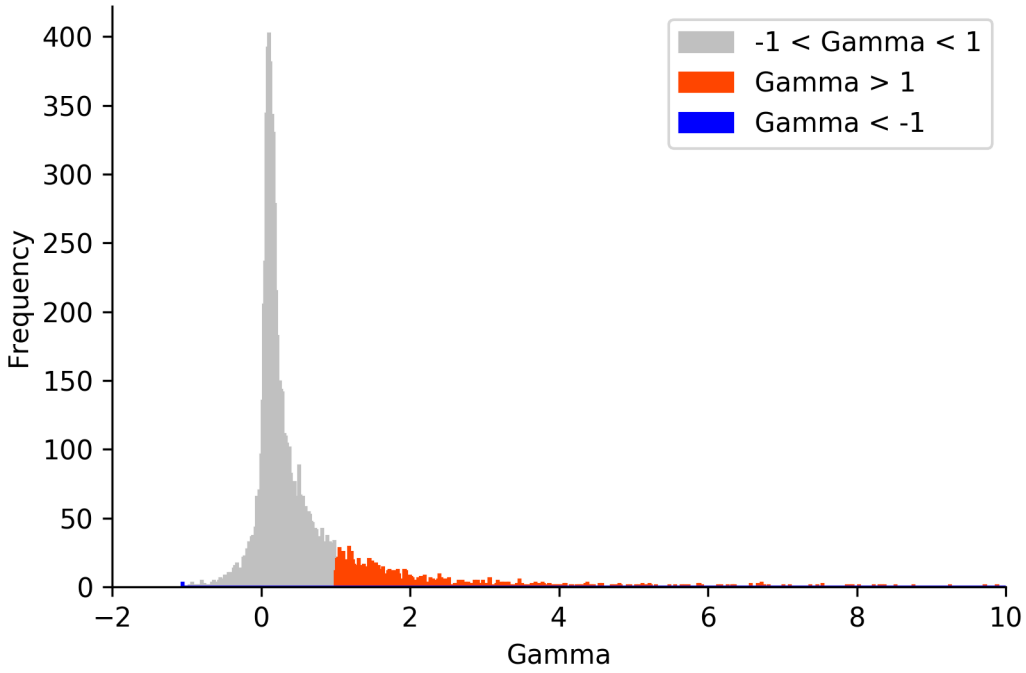


Figure 5-1. *Distribution of the selection coefficient scaled by the effective population size (γ) across the* Ceratina calcarata *genome.*

Regions highlighted in red specify loci under strong positive selection (γ>1), light grey are loci under neutral or nearly-neutral evolution (-1 <γ< 1), and those in blue are loci under strong negative selection (γ<-1).

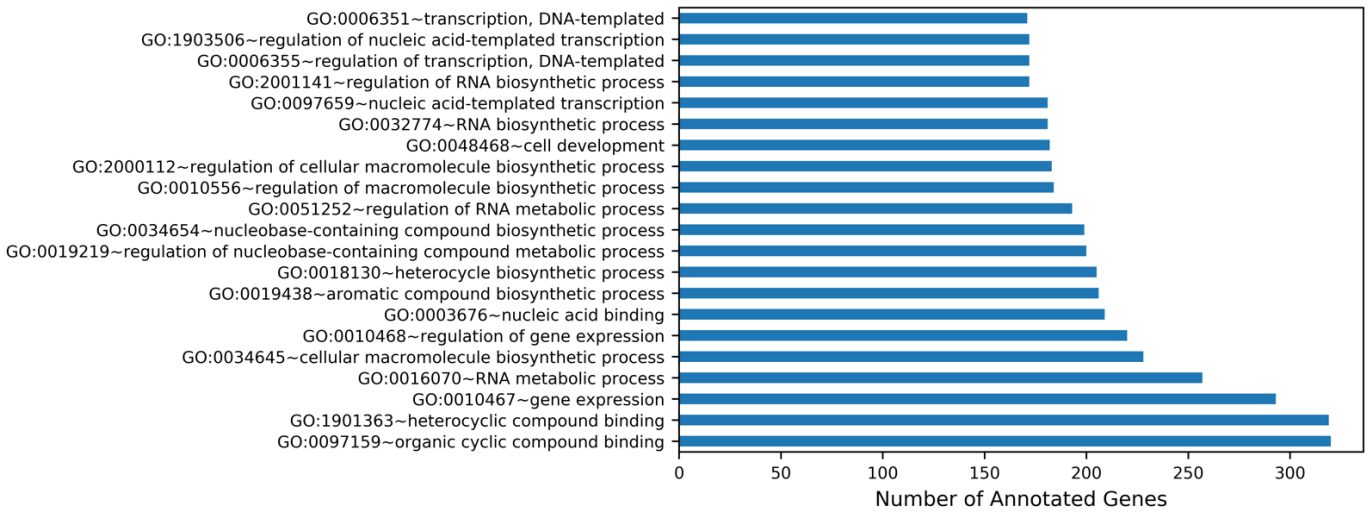


Figure 5-2. *The top 20 gene annotations significantly enriched ($p < 0.05$) among genes with outlier Fst levels between social and subsocial* Ceratina calcarata*, focusing on gene expression and transcription related to both DNA and RNA.*

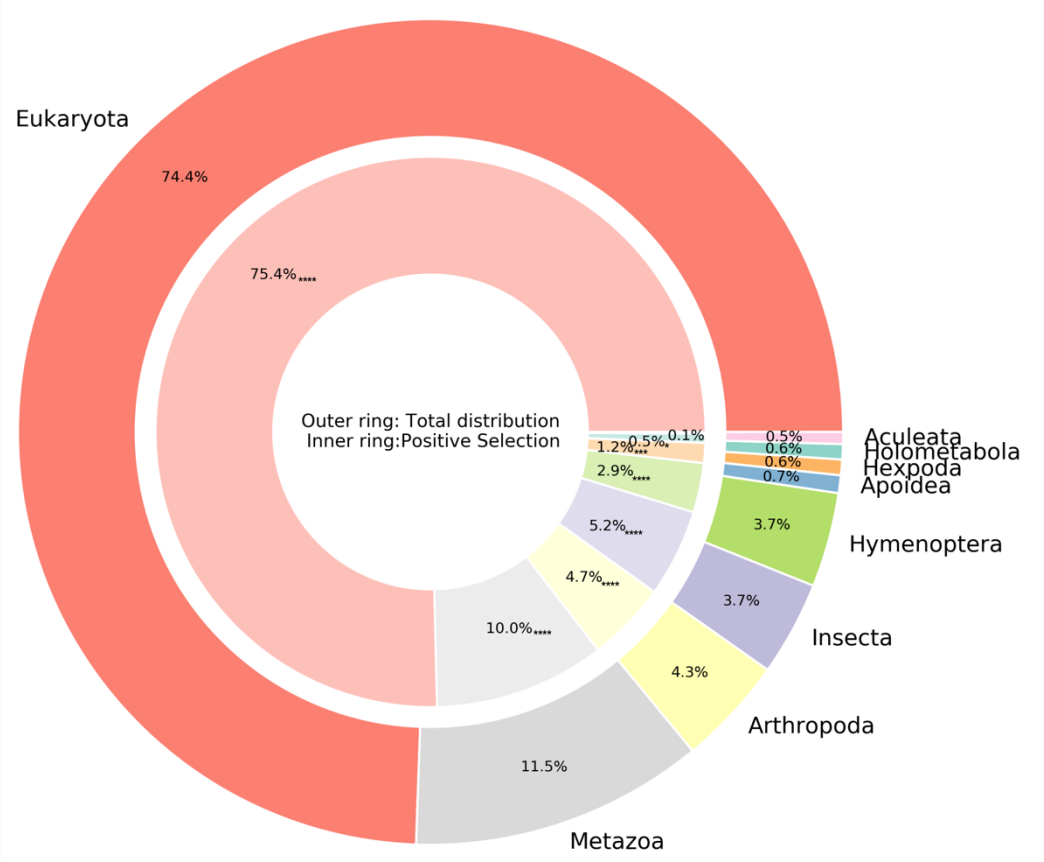


Figure 5-3. *Most genes under positive selection are evolutionarily ancient, as identified through OrthoDB.*

*The outer ring shows the total distribution of genes with a γ value, while the inner ring displays those under positive selection. Significance was assessed using a G test of independence (ns = not significant, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001).*

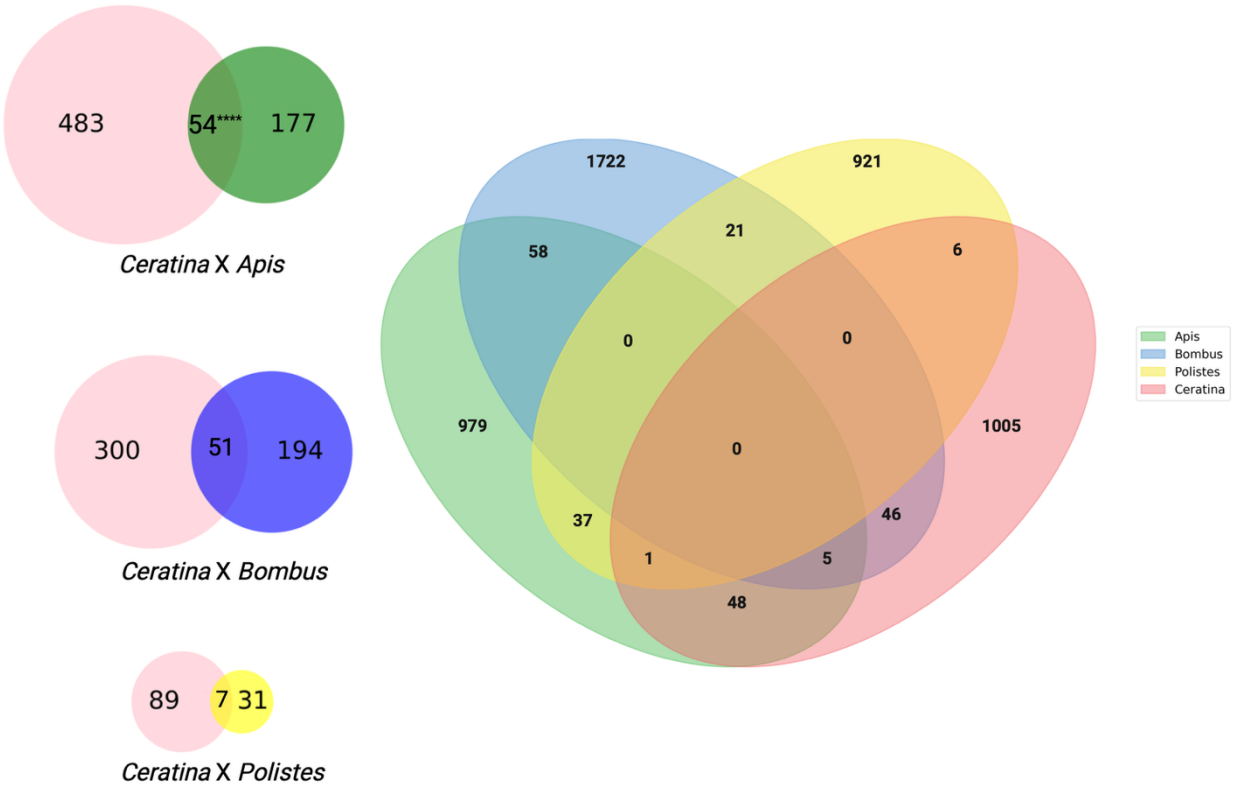


Figure 5-4. *Overlap of orthologous genes under positive selection in Apis mellifera (advanced eusocial),* Bombus impatiens *(primitively eusocial),* Polistes dominula *(primitively eusocial), and* Ceratina calcarata *(incipiently social). Significant overlap was found between genes under positive selection in Ceratina and those in Apis (p=1.19-4) but not in Polistes (p=0.439) or Bombus (p=0.067).(ns = not significant, *p ≤0.05, **p ≤0.01, ***p ≤0.001, ****p < 0.0001).*

Table 5-1. List of genes both under positive selection and with outlier levels of Fst among social and subsocial **C. calcarata.**

One biological function for the first GO term is listed here. For a complete list of all possible biological functions and GO terms, please refer to Supplemental Table S5-6.

GeneID	Gamma	Average Fst	NCBI	GO Terms	Biological Process
LOC108625050	6.667	0.052	ubiquitin thioesterase otubain-like	GO:0006508	proteolysis
LOC108625502	4.358	0.054	proline-rich protein 2-like		
LOC108622655	3.706	0.063	facilitated trehalose transporter Tret1-like	GO:0008643	carbohydrate transport
LOC108623572	1.747	0.074	U3 small nucleolar ribonucleoprotein protein IMP4-like	GO:0006364	rRNA processing
LOC108625019	1.576	0.178	erlin-1-like		
LOC108627874	1.56	0.054	facilitated trehalose transporter Tret1-2 homolog	GO:0008643	carbohydrate transport
LOC108632852	1.526	0.059	methyltransferase-like protein 13	GO:0032259	Methylation
LOC108626635	1.504	0.092	Nck associated protein 1 Hem	GO:0000902	cell morphogenesis
LOC108623377	1.439	0.06	kinesin-like protein KIF6	GO:0007017	microtubule-based process
LOC108622404	1.364	0.054	uncharacterized LOC108622404		
LOC108628187	1.323	0.056	uncharacterized LOC108628187	GO:0055085	transmembrane transport
LOC108628446	1.27	0.066	CDK-activating kinase assembly factor	GO:0000079	regulation of cyclin-dependent protein serine/threonine kinase activity
LOC108624986	1.252	0.066	receptor-binding cancer antigen expressed on SiSo cells		
LOC108625966	1.174	0.055	trichohyalin-like		
LOC108622527	1.066	0.076	F-box only protein 28	GO:0000209	protein polyubiquitination
LOC108629327	1.056	0.067	transmembrane emp24 domain-containing protein eca	GO:0006886	intracellular protein transport
LOC108629428	1.055	0.082	histone H2A-like		
LOC108623578	1.054	0.059	ruvB-like 1	GO:0000492	box C/D snoRNP assembly
LOC108623883	1.04	0.057	zinc finger protein 76-like	GO:0006357	regulation of transcription by RNA polymerase II
LOC108622777	1.013	0.053	uncharacterized LOC108622777	GO:0006629	lipid metabolic process

Supplemental Tables

All supplemental tables can be found in the supplemental dissertation file

Chapter 6 Patterns of mutation and recombination rate in aging honey bees (*Apis mellifera)*

Summary

Queen honey bees (*Apis mellifera*) have the remarkable ability to live for several years and remain reproductively active. Aging can influence genome stability but its effect on honey bee genetics and genome biology has not been fully investigated. Honey bees are haplodiploid, producing haploid males through arrhenotokous parthenogenesis. Here, we tested whether maternal age influences mutation rates and patterns across the genome by sequencing queens of different ages and their haploid sons. We found that mutation rates did not differ across age groups or in the types of mutations observed (transitions, transversions or their ratio). Novel mutations in adult drones were less frequently found in genic regions (introns and exons) relative to non-genic regions (upstream, downstream, or intergenic), which is consistent with purifying selection. GC content also did not differ between genomic windows containing at least one mutation and those without and did not change across age groups. Recombination rates inferred from linkage disequilibrium also did not change with aging, but windows that contained at least one mutation tended to occur in regions with higher recombination rates than matched control windows. Taken together, our results demonstrate that maternal age has little effect on the mutation rates or mutation types in honey bees. Instead, the distribution of mutations across the genome appears to be shaped mainly by recombination and purifying selection.

Introduction

Mutation provides the raw source of genetic material for evolution, but the rate at which mutations arise can be shaped by various biological processes such as DNA replication (Lujan *et al.* 2014; Hunter 2015; Spisak *et al.* 2024) and repair (Supek and Lehner 2015). Across insects, mutation rates appear to be relatively constant, with similar rates observed in *Drosophila melanogaster* (2.8×10^{-9})(Keightley *et al.* 2014), *Heliconius melpomene* (2.9×10^{-9})(Keightley *et al.* 2015), bumblebees (3.6×10^{-9})(Liu *et al.* 2016), and honey bees (3.4×10^{-9})(Liu *et al.* 2016) per haploid genome per generation. Despite this apparent similarity across species, mutation rates can vary across the genome and may be influenced by meiotic recombination. During meiosis, crossover between homologous chromosomes involves double-stranded breaks and DNA

repair, which have each been implicated in elevated mutation rates (HUNTER 2015). For example, in humans, crossover regions show elevated mutation rates and a strong Adenine/Thymine (A/T) bias (ARBEITHUBER *et al.* 2015). In *C. elegans*, genomic regions with higher recombination rates were enriched for indels, indicating that recombination contributes to processes of mutation beyond what is expected from replication alone (HWANG AND WANG 2023). However, evidence for a direct relationship between recombination and mutation remains mixed, as other studies have reported weak or no association between these processes (LIU *et al.* 2016).

One life-history factor that may further modulate mutation and recombination is parental age (KONG *et al.* 2012; WONG *et al.* 2016). Extended reproductive lifespans involve repeated rounds of germline cell division and prolonged exposure of germ cells to endogenous sources of DNA damage, potentially increasing the opportunity for replication errors and other mutational events to accumulate over time (KONG *et al.* 2012; GAO *et al.* 2019). Consistent with this idea, maternal age in humans is correlated with both changes in recombination rate and an increased occurrence of *de novo* mutations near crossover events (HALLDORSSON *et al.* 2019), supporting a link between aging, recombination, and mutagenesis. Despite this, direct empirical tests of aging-related changes in mutation and recombination remain scarce, particularly in systems with long-lived, highly fecund females that reproduce continuously over multiple years.

Honey bees (*Apis mellifera*) provide a powerful system in which to investigate the effects of how aging and recombination shape processes of mutation. Advanced eusocial insects, including honey bees, display a characteristic set of behaviours including a reproductive division of labour, cooperative brood care and overlapping generations (MICHENER 2007). *A. mellifera* queens mate only once early in life with multiple male drones and store sperm in a specialized organ called the spermatheca for future use (SCHLÜNS *et al.* 2005; MICHENER 2007). Queens will produce all of the eggs for the colony continuously over several years, which previous studies have reported can last up to three years (ROBERT E. PAGE JR 2001), and she can even control the quantity of her reproductive output depending on colony conditions (AAMIDOR *et al.* 2022). Whereas many insects complete their reproduction cycle over a short adult lifespan, honey bee queens can sustain continuous gamete production for several years, creating an opportunity for age-related changes in mutation and recombination to accumulate. Notably, estimates for honey bee mutation rates have been previously derived from drones generated from one year old queens (LIU *et al.* 2016), which leaves questions regarding the stability of mutation rates as queens age. No study to date has conclusively investigated aging-related genomic changes in honey bee queens, likely reflecting the logistical challenges of maintaining and studying long-lived queens over multiple years.

In addition to their extreme reproductive longevity, honey bees also exhibit some of the highest recombination rates known among animals. Recombination rates of solitary species such as *Drosophila melanogaster* (1.6 cM/Mb)(She and Jarosz 2018) , *Nasonia* spp (1.5 cM/Mb) (Wilfert *et al.* 2007) and other social Hymenoptera (6.4-14.0 cM/Mb) (Sirvio *et al.* 2006; Sirvio *et al.* 2011; Sirviö 2011; Liu *et al.* 2016) are substantially lower than those reported across several *Apis* species (17.4- 37 cM/Mb, (The Honeybee Genome Sequencing Consortium 2006; Shi *et al.* 2013; Liu *et al.* 2015; Wallberg *et al.* 2015; Rueppell *et al.* 2016). These exceptionally high recombination rates have been previously linked to the evolution of eusociality (Kent and Zayed 2013; Kawakami *et al.* 2019; Waiker *et al.* 2021). Beyond its evolutionary implications, recombination can also shape genome composition through GC-biased gene conversion (gc-BGC), in which mismatches in heteroduplex DNA are preferentially repaired in favor of Guanine or Cytosine (G/C) alleles over A/T alleles leading to increased genomic GC content over evolutionary time (Duret and Galtier 2009). Consistent with this, a comparative study across insects found that coding genes, especially near telomeres where recombination is more frequent, exhibit elevated GC content (Kyriacou *et al.* 2024). GC-bias is also especially pronounced in taxa with exceptionally high recombination rates, such as honey bees (Kent *et al.* 2012; Ross *et al.* 2015; Wallberg *et al.* 2015; Kawakami *et al.* 2019). Thus, understanding how recombination rate and GC-biased genome composition interact with mutations, and how these relationships change with aging, may be key to explaining how extreme life history traits associated with eusociality evolve.

Honey bees offer a unique opportunity to directly examine the products of female meiosis because queens can produce haploid drones that inherit their genomes entirely from their mother, providing a snapshot of the queen's meiotic output at the time eggs are laid without any paternal genetic contribution. Here, we investigate how queen aging over several years influences the genomes of their haploid sons by sequencing drones from honey bee queens of different ages across Canada. Using these data, we quantify *de novo* mutation rates, relate them to population-scale recombination rates inferred from linkage disequilibrium, and examine additional genomic features associated with recombination. We hypothesize that *de novo* mutation rates increase with maternal age and that these mutations are enriched in genomic regions with high recombination rates.

Methods

Sample collection

Apis mellifera queens of varying ages along with 4-5 of their drone sons were sourced from bee breeders or bee researchers from across Canada; these samples were flash frozen and shipped on dry ice and subsequently stored at -80°c at York University (Supplemental Table S6-1). Bee breeders or researchers

track queen age any by applying a color mark or tag on the queen's thorax during the year that the queen was produced. The universal queen breeding color scheme uses white (years 1 and 6), yellow (2 and 7), red (3 and 8), green (4 and 9) and blue (5 and 0) allowing beekeepers to quickly determine the age of their queens in years (HAGLER AND JACKSON 2001). The system, however, makes it difficult to ascertain the age of queens in months. We thus classified queen into different age bins as follows: We studied three queens aged 1 year, seven queens between the ages of 18 months and 2 years, (herein referred to as 2-year-olds), and four queens between the ages of 3 and 4 years (herein referred to as 3-year-olds).

DNA extraction and Sequencing

DNA was extracted from each bee's thorax using optimized protocols (CONFLITTI *et al.* 2022) and the Mag-Bind® Blood & Tissue DNA HDQ 96 Kit (Omega Bio-tek Inc., USA) that was augmented for the KingFisher™ Flex Purification System (Thermo Fisher Scientific Inc., USA) to a final elution volume of 70-100 µl. Tissue Lysis Buffer (350 µl) and Proteinase K (20 µl) were used to lyse tissues before heating samples at 55°C overnight. DNA quality control and quantification was assessed using a 2.0% agarose gel electrophoresis and NanoDropTM 2000 Spectrophotometer (Thermo Fisher Scientific Inc., USA), respectively. Library preparation and sequencing were performed at the Genome Quebec Innovation Centre (McGill University) on the Illumina HiSeq 2500 system, producing a total coverage of 260X (27X per bee) with 86X per queen (9 queens, 3 queens per lane) and 17X per drone (45 drones, 15 drones per lane) and a genome size of 240Mb.

We used previously published pipelines for alignment and SNP calling (HARPUR *et al.* 2019; IMRIT *et al.* 2020; HARPUR AND REHAN 2021). Briefly, raw reads were trimmed with TRIMMOMATIC v0.32, aligned to the *Apis mellifera* Amel_HAv3.1 genome (WALLBERG *et al.* 2019) using NextGenMap v0.4.12 (SEDLAZECK *et al.* 2013). Duplicate reads were removed using Picard v1.8 MarkDuplicates (http://broadinstitute.github.io/picard/). Single nucleotide polymorphisms (SNPs) were identified using GATK HaplotypeCaller through an existing pipeline (DOGANTZIS *et al.* 2018). To avoid errors of comparing haploid (drone) and diploid (queen) genomes, we called all drones as diploid (-ploidy 2) and any heterozygous drone calls were removed as errors.

Filtering and identifying unique mutations

Queens and her drone samples were compiled into one file using GATK GenotypeGVCFs. SNPs and insertion-deletions (indels) were identified using SelectVariants (-selectType SNP and-selectType INDEL). SNPs were filtered for quality using GATK's recalibration model to improve the confidence in the calls as true SNPs using GATK recommendations (MappingQuality < 20, QualitybyDepth < 2.0, FisherStrand > 60,

StrandOddsRatio > 3.0, ReadPosRankSum < -8.0, MappingQualityRankSum < -12.5, MappingQualityRankSum > 12.5)(VAN DER AUWERA *et al.* 2014). We also removed SNPs if they were within 5 base pairs of an indel or copy number variant (CNV; GATK VariantFiltration--maskExtension 5). We then separated the queen sample from drone samples and plotted the sequencing depth of all SNPs in either the queen or the drones combined to identify the threshold for which to filter out the top and bottom 5th percentile of reads that had either exceptionally low or high read depth (Supplemental Table S6-2). we then used these thresholds to filter SNPs using GATK VariantFiltration. We removed any SNP with any sort of tag using vcftools (--remove-filtered-all). We also removed any SNPs present on unplaced scaffolds (grep -v '\QIU*') or in the mitochondrial genome (grep -v 'CM009947.2') or any SNP with missing data in any sample. We also removed any SNP with an alternative alleles (genotypes '0/2', '0/3', '1/2','1/3','2/3', '2/2','3/3') (MOHD-ASSAAD *et al.* 2018). Finally, we removed any SNP within repeat regions of the genome using the RepeatMasker output file from NCBI (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/003/254/395/GCF_003254395.2_Amel_HAv3.1/GCF_003254395.2_Amel_HAv3.1_rm.out.gz).

We then classified unique SNPs (mutations from here on) as those that are present in only one drone and no other drones or queen using called jvarkit (java -jar jvarkit/dist/vcffilterjs.jar -f jvarkit/dist/select.js input.vcf > output.vcf; (LINDENBAUM AND REDON 2017)). To identify the mutation rate, we divided the total number of mutations identified in each sample divided by the number of callable loci in each sample and averaged this number across age groups (GATK CallableLoci ; (VAN DER AUWERA AND O'CONNOR 2020)). We then looked at for differences in average mutation rate across age groups using a Kruskal-Wallis test with Dunn's post hoc test to look for individual differences among age groups. We also evaluated if there was a trend of maternal aging affecting the mutation rate as a continuous variable using an ordinary least squares regression (scipy.stats, scikit_posthocs and statsmodels.formula.api for python).

Identifying the location of unique mutations across a gene

We tested whether mutations occur at different frequencies in gene bodies (introns and exons) compared to non-genic regions (upstream, downstream, and intergenic regions). This comparison was motivated by the expectation that different genomic categories experience different levels of selective constraint. Coding regions are often subject to strong purifying selection because mutation that alter amino acids are often deleterious, whereas intronic and intergenic regions generally experience weaker constraint and thus tolerate more variability (SIMOLA *et al.* 2013; MILLER AND SHEEHAN 2023). To test this hypothesis, we used SnpEff (CINGOLANI *et al.* 2012) to annotate each mutation and identify where it is located into simplified categories provided in Supplemental Table S6-3. We calculated the number of callable loci (GATK

CallableLoci ; (VAN DER AUWERA AND O'CONNOR 2020)) within each simplified genomic region to serve as the denominator for estimating mutation rates within that specific region. To avoid pseudoreplication among drones, we pooled the number of mutations and callable loci across drones within each queen-level group (sample group) and calculated queen-level mutation rates. Queens were then grouped by age category and genomic category class. For the main comparison, categories were further grouped into genic regions (exons and introns) and non-genic regions (upstream, downstream, and intergenic). We tested whether mutation rates were lower in genic regions than in non-genic regions using a one-tailed Wilcoxon signed-rank test applied to paired queen-level rates within each age group. Results were visualized using bar plots showing the mean mutation rate and standard error of the mean across queen-level groups for each age group and region category.

Transitions and transversions

To evaluate whether mutation type varied with queen age, we quantified transition and transversion mutations from unique mutations using bcftools stat (DANECEK *et al.* 2021) on VCF files for each queen-drone family. This generated mutation summary statistics per sample, from which we extracted the number of transitions, transversions, and calculated the transition/transversion (Ts/Tv) ratio. These values were compiled into a single dataset along with metadata on the number of drones per family and queen age, grouped into the three age categories (1 year, 2 years, and 3 years). We tested whether maternal age was related to the number of transitions, transversions, or the transition/transversion (Ts/Tv) ratio produced per drone. We fit three separate linear regression models, one for transitions, one for transversions, and one for the Ts/Tv ratio. The queen-level group with an undefined Ts/Tv value (because it had zero transversions) was excluded from the ratio analysis. All statistical tests were performed in Python using ordinary least-squares regression (statsmodels; (SEABOLD AND PERKTOLD 2010)).

GC content

To quantify GC content across the genome, we divided the *Apis mellifera* reference genome (Amel_HAv3.1) into non-overlapping 5 kb windows using BEDtools makewindows (QUINLAN AND HALL 2010) with the corresponding FASTA index file. GC content for each 5 kb window was then calculated using BEDtools nuc (QUINLAN AND HALL 2010), which computes the proportion of G and C nucleotides within each window based on the reference genome sequence. For each sample, the position of each mutation was extracted as chromosome–position pairs and assigned to non-overlapping 5 kb genomic windows using an interval-tree–based approach in Python. To test whether GC content and maternal age predicted the number of mutations per window, we fitted a negative binomial regression model, which is a common

method for over-dispersed count data (ZEILEIS *et al.* 2008; HILBE 2014), with mutation count as the response and GC content and age as predictors. We also fitted a logistic regression model to test whether a window contained at least one mutation, using the same predictors under the generalized linear model (GLM) framework (NELDER AND WEDDERBURN 1972). To further test whether mutation-containing windows differed in GC content from randomly sampled genomic windows, we performed a bootstrap comparison of the difference in GC content (ΔGC). Because our hypothesis predicted that mutation-containing windows would have higher GC content than randomly sampled windows, we used a one-tailed bootstrap test (EFRON AND TIBSHIRANI 1993). The null hypothesis assumed that GC content in mutation windows was equal to or lower than that in random windows, and the p-value was calculated as the proportion of bootstrap resamples producing a GC difference (ΔGC) equal to or greater than the observed value. All analyses were implemented in Python using the statsmodels package (SEABOLD AND PERKTOLD 2010).

Assessing recombination rate using LDhat

We used LDhat, a statistical genetics software package designed to infer population recombination rates (ρ) from population genetic data, particularly from phased haplotype sequences (MCVEAN AND AUTON 2011). LDhat examines patterns of linkage disequilibrium (LD), which is non-random associations between genetic variants, to estimate the population recombination parameter (ρ) that best explains the observed patterns of LD between sites in a genome. For each sample, unique SNP (mutation) positions were extracted from VCF files using bcftools query and converted to BED format. We then generated non-overlapping 5 kb windows centered on each novel mutation's position (±2,500 bp) using BEDtools slop with the *Apis mellifera* reference genome index. These novel mutation-centered windows were used to subset the full filtered VCF files. Variant records falling within each 5 kb window were extracted using a custom Python script implementing an interval-tree–based lookup, and separate VCF files were generated for each window. Because LDhat requires phased haplotypes, all window-specific VCF files were phased using Beagle prior to generating LDhat input files (BROWNING *et al.* 2021). To prepare inputs for LDhat, we generated ".sites" and ".locs" files using custom code (see Data Availability in Appendix A). The ".sites" file contains the haplotype data and segregating sites, while the ".locs" file specifies the genomic positions of those sites and is converted to megabase units to be used properly in LDhat. We used LDhat Pairwise, which estimates the population recombination rate (denoted as $\rho = 4Ner$, where Ne is the effective population size and r is the recombination rate per site per generation) between all pairs of polymorphic sites. This is achieved using a moment estimator method, which is fast and useful for generating a likelihood lookup table for further analysis. LDhat Pairwise also allows for hypothesis testing to detect the presence of recombination in the

dataset. We then used the likelihood lookup table that was generated from LDhat pairwise as one of the input files along with the sites and locs files for LDhat interval.

For each window, LDhat interval was run using a Markov Chain Monte Carlo (MCMC) approach to estimate local variation in population recombination rates (ρ) along a DNA sequence. MCMC is a computational algorithm that approximates the posterior distribution of model parameters (in this case, population recombination rates, ρ) by sampling many possible configurations. During this process, "updates" refer to the number of iterations the algorithm runs to refine its estimates. In our analyses, we ran LDhat Interval for 50,000,000 MCMC updates to allow sufficient exploration of the parameter space, and collected samples of the recombination rate every 10,000 updates, ensuring we obtained a well-sampled posterior distribution. We also specified a block penalty of 1, which controls how much the population recombination rate is allowed to change between adjacent regions where smaller values allow more frequent changes (WALLBERG *et al.* 2015). We then used LDhat stat with a burn-in of 10% of the samples applied. The mean population recombination rates (ρ) were extracted from the LDhat output and summarized across windows using custom Python scripts. In some cases, certain windows lacked sufficient information to generate the residuals file which was needed for the program that generates statistical summaries called LDhat stat and are therefore not included in the final analysis. Visualization of LD-inferred population recombination rates (ρ) across age groups was performed in Python using NumPy and Matplotlib.

To generate a null distribution for comparison with our empirical window sets, we constructed bootstrapped reference sets in 5 KB genomic windows using the *Apis mellifera* reference genome (Amel_HAv3.1) using bedtools makewindows (QUINLAN AND HALL 2010). We then randomly sampled an average number of windows for the that age group 100 times. For example, there was an average of 87 windows containing at least one unique mutation in the 1-year age group so for the control 1 year distribution (which we refer to as the null distribution) we randomly sampled from the genome 100 sets of 87 windows that did not contain any unique mutations in those windows. We continued this for 100 sets of 33 windows for the 2-year age group and 100 sets of 343 windows for the 3-year age group. We then followed the same pipeline for running preparing files for and running LDhat pairwise, LDhat interval and LDhat stat.

For the analysis, we compared the difference in the mean rho between the experimental windows (one value per queen-level colony) containing at least one mutation and the control windows that did not have any mutations in the windows (100 iterations per queen-level colony). We then compared experimental and control values by subtracting the mean control rho from the corresponding experimental

mean rho. We then computed a Δrho (difference in rho) for each paired comparison by subtracting the control mean rho from the experimental mean rho (e.g., sample A compared to 100 iterations of control A, sample B compared to 100 iterations of control B, and so on). Under the null hypothesis, the experimental and control values do not differ, so the distribution of Δrho values is expected to be centered around zero, with approximately equal numbers of positive and negative values. We therefore counted the number of negative Δrho values across all comparisons within each age group and divided this by the total number of observations to obtain a p-value. All statistical analyses were conducted in Python using pandas, NumPy, SciPy, and statsmodels, and visualizations were generated using Matplotlib (HUNTER 2007; SEABOLD AND PERKTOLD 2010; TERPILOWSKI 2019).

Results

Mutation rates with queen aging

We analyzed whole-genome data from 14 queens (via 63 drones) spanning three age groups: 1 year (n = 13 drones, 3 queens), 2 years (n = 35 drones, 7 queens), and 3 years (n = 15 drones, 4 queens), to investigate whether mutation rate in honey bee drones varies with queen age (Supplemental Table S6-4A). To avoid non-independence among drones produced by the same queen, we summarized the data at the queen level by averaging mutation rates across drones that have the same queen mother within each age, yielding 14 queen-level mean mutation rates used for downstream inference.

We then examined if queen age affects the mutation rate by evaluating the differences between mutation rate within age groups (Figure 6-1). Because the Shapiro–Wilk tests indicated a significant departure from normality in the 2-year age group ($p = 0.011$), but not in the 1-year ($p = 0.470$) or 3-year ($p = 0.238$) groups (Supplemental Table S6-4B), we used a Kruskal–Wallis test to compare mutation rates among age classes, with the null hypothesis that the distribution of mutation rates are constant across age groups. The results of this test were not significant ($H = 4.253$, $p = 0.119$; Supplemental Table S6-4C). None of the pairwise Dunn's post-hoc tests were also significant among pairs of age groups ($p > 0.05$; Supplemental Table S6-4D). We also performed an ordinary least squares regression (OLS) where queen age was treated as a continuous predictor. We found a slightly positive trend where older queens have a very modest increase in mutation rate, but the relationship was not statistically significant ($\beta = 1.747 \times 10^{-7}$, $p = 0.272$; Supplemental Table S6-4E).

Location of mutations across the gene structure with queen aging

We tested whether mutation rates differed between genic regions (introns and exons) and non-genic regions (upstream, downstream, and intergenic). Mutation rates were consistently lower in genic

regions than in non-genic regions (Figure 6-2). This difference was statistically significant in the 2-year age group (Wilcoxon signed-rank test, p = 0.009, n = 7 queen-level groups) and in the 3-year age group (p = 0.034, n = 4 queen-level groups), but was only marginally significant in the 1-year age group (p = 0.054, n = 3 queen-level groups).

Rates of transitions and transversion across queen aging

We examined whether maternal age influences the types of mutations that are generated, such as if more or less mutations are produced as transition mutations, transversion mutations or if the ratio of transition-transversions (Ts/Tv) alters with aging, meaning one type of mutation is biased over another with aging (Figure 6-3; Supplemental Table S6-5A). None of the three regression models showed a significant effect of maternal age (Supplemental Table S6-5B,C,D). For transitions, the slope was positive but non-significant (β = 29.499 ± 20.217 SE, p = 0.175, R^2 = 0.176), indicating that age did not explain meaningful variation in transition counts. Age also did not predict the number of transversions (β = 5.748 ± 3.708 SE, p = 0.152, R^2 = 0.194). The Ts/Tv ratio likewise showed no association with age (β = −0.446 ± 0.560 SE, p = 0.447, R^2 = 0.066). In all models, the 95% confidence intervals for the slopes included zero, showing no detectable increase or decrease in any mutation type with increasing queen age.

GC content across age groups

We examined the mean GC content across 5kb windows containing at least one novel mutation to identify if there is a difference in GC content within these windows across aging (Supplemental Table S6-6A). We accomplished this by testing whether GC content of genomic windows containing a novel mutation differed across age groups using a one-way ANOVA on queen-level means and found that they do not differ significantly overall (p> 0.05) or in pairwise tests (p> 0.05; Figure 6-4A; Supplemental Table S6-6B,C). We then used bootstrapping to compare the GC content of genomic windows containing at least one novel mutation with randomly selected genomic windows that did not contain a novel mutation. For each age group, GC content differences were calculated as the GC content of mutation-containing windows minus the GC content of randomly selected windows across 100 bootstrap iterations (Supplemental Table S6-6D). Under the null hypothesis that randomly selected windows have higher GC content than mutation-containing windows, this difference would be negative. We therefore calculated the proportion of iterations with negative values to obtain a p-value. We did not observe a difference in GC content between mutation-containing windows and randomly selected windows (1 year: p = 0.326, 2 years: p = 0.312, 3 years: p=0.1975; Figure 6-4B-D).

Recombination rates are stable with maternal aging

We measured linkage disequilibrium in 5kb windows containing at least one unique mutation across all age groups to identify differences in LD with aging using LDhat. When we compared the mean linkage disequilibrium as represented by rho values (rho = $\rho = 4N_e r$) in each age group, we did not observe a significant difference with aging (Figure 6-5A; Supplemental Table S6-7A,B). In pairwise comparisons of differences in LD between 5kb windows containing mutations versus windows that don't contain mutations across pairs of age groups (1 vs 2, 2 vs 3, 1 vs 3) the strongest contrast was observed between 1-year and 2-year (Welch t-test raw p = 0.034), but this did not remain significant after Holm correction (p = 0.102; Supplemental Table S6-7C). No other pairwise contrasts were significant (Holm-adjusted Welch tests: 2 vs 3 p = 0.292; 1yr vs 3yr p = 0.194). Nonparametric pairwise tests were also consistent with these results (Holm-adjusted Mann–Whitney: all p ≥ 0.205; Supplemental Table S6-7D). Overall, the data provide no significant evidence for differences in average rho across years.

To test the hypothesis that recombination is mutagenic, we compared linkage disequilibrium–inferred recombination rates ($\rho = 4N_e r$) in genomic windows containing at least one novel mutation with those in randomly selected windows that did not contain a mutation. For each sample and age group, we randomly sampled an equivalent number of 5 kb windows without mutations across 100 iterations and calculated the average recombination rate (ρ). We then averaged ρ values across mutation-containing and randomly selected windows and calculated the difference in average ρ between the two groups (mutation-containing minus random windows; Supplemental Table S6-7E), which was visualized as a histogram. As the null hypothesis is that average control rho values will be higher that experimental, resulting in negative values in the distribution (experimental minus control), we counted the number of negative values found in this distribution and divided by the total number of iterations assessed to find the P value. For 1 year old samples we found 1 negative value out of 300 iterations, resulting in a significant p value of 0.003, demonstrating that there is a significant departure from our null hypothesis and our experimental values are significantly different from our controls in that there is more LD identified in experimental windows that contain at least one unique mutation relative to control windows that do not contain unique mutations (Figure 6-5B). We found a similar pattern for 2-year samples (p <0.0001; Figure 6-5B) and 3-year samples (p = 0.001; Figure 6-5C), showing that overall, there is a link between LD and mutation-containing windows.

Discussion

We tested the hypothesis that mutation rates are influenced by maternal longevity in the honey bee genome. In many organisms, mutation rates are expected to increase with parental age due to the

cumulative effects of DNA replication errors, declines in repair efficiency, and the accumulation of cellular damage over time. For example, this phenomenon has been observed in the germline cells of *C. elegans*, which exhibit more crossover errors with aging (RAICES *et al.* 2021), and in *Drosophila melanogaster*, which shows less effective double-stranded break repair in older flies (DELABAERE *et al.* 2017). Similarly in flies, although the number of mutations introduced during continual spermatogenesis remains comparable in young and old males, older flies exhibit diminished repair capacity, leading to higher mutation loads (WITT *et al.* 2023). These age-related impairments in DNA repair and damage removal have been proposed as major contributors to reproductive aging in both somatic and germline cells (PANIER *et al.* 2024). Contrary to these expectations, we found no evidence that mutation rates differed among queens of different ages. We also did not detect any difference in the types of mutation (transition or transversion) across queen age groups. Together, these results suggest that the process of mutation itself remains largely stable across the reproductive lifespan of honey bee queens. These queens continuously produce eggs throughout their lives (CULLEN *et al.* 2023), unlike the finite number of oocytes in humans, in which gametes remain arrested until ovulation (SOLC *et al.* 2010). The need to sustain egg production over several years may therefore require particularly robust mechanisms for maintaining germline fidelity in honey bees. However, an important caveat of our study is that mutations were detected by sequencing adult drones. As a result, we can only identify mutations that successfully passed early developmental filtering. Mutations that cause lethality or severe developmental defects would not be observed because individuals carrying them would fail to develop into adults. Therefore, our findings should be interpreted as reflecting mutations that persist through development rather than the total number of mutations initially produced in the germline. Directly assessing the full mutational input would require sequencing earlier developmental stages, such as unfertilized eggs. Understanding how DNA repair and recombination processes are maintained throughout the lifespan of honey bee queens represents an important direction for future research.

We next examined whether the genomic distribution of mutations differed across functional genomic regions. Patterns of mutations are not uniform across the genome but instead reflect selective pressures acting on distinct genomic regions. As noted above, because mutations were detected by sequencing adult drones, our dataset reflects mutations that persisted through development rather than all mutations initially produced in the germline. Protein-coding regions and functionally important introns can be subject to strong purifying selection because mutations in these regions are more likely to disrupt gene function and reduce fitness (MILLER AND SHEEHAN 2023). As a result, deleterious variants are removed from the genome, leading to fewer detectable mutations in gene bodies (HALLIGAN *et al.* 2004). In contrast, many mutations in non-coding regions such as intergenic DNA are more likely to be tolerated, although they

too can still experience some constraint because they are important for regulation (CASILLAS *et al.* 2007). There are typically very low levels of genetic diversity in gene bodies relative to intergenic areas, a pattern which has been observed across many species including ants (SIMOLA *et al.* 2013), flies (CASILLAS *et al.* 2007; SINGH *et al.* 2008), humans (LINDBLAD-TOH *et al.* 2011; JABALAMELI *et al.* 2019; DUKLER *et al.* 2022) as well as primates and rodents (SUBRAMANIAN 2013). Our results are consistent with this broader pattern as mutation rates were consistently lower in genic regions than in non-genic regions across queen-level groups, suggesting that purifying selection may help shape the distribution of mutations in the honey bee genome regardless of maternal age.

Honey bees have one of the highest recombination rates of any metazoan genome (BEYE *et al.* 2006; MEZNAR *et al.* 2010). Elevated recombination rates have been implicated in eusocial genome evolution, as this phenomenon has been observed in phylogenetically distant species with similarly high recombination rates and advanced social behaviour (WILFERT *et al.* 2007; WAIKER *et al.* 2021). One hypothesis for why recombination rates may be elevated in eusocial insects relates to the evolution of more complex forms of sociality, wherein reproductive and worker caste are distinct. In these systems, effective population size may be reduced (CROZIER 1976) because relatively few individuals contribute genes to the next generation. Under these conditions, new, beneficial mutations may be less likely to reach fixation if they arise in small reproductive populations and are physically linked to deleterious alleles (MAYNARD-SMITH AND HAIGH 1974). This interference, known as the Hill–Robertson effect, can limit selection (HILL AND ROBERTSON 1966). Higher recombination rates have therefore been proposed to help alleviate this constraint by separating beneficial mutations from linked deleterious mutations when there are few reproductive individuals (HARTFIELD *et al.* 2010).

Meiotic recombination is initiated by double-strand breaks and has therefore been proposed as a direct source of new mutations. For example, in humans and other systems, crossover formation and associated double stranded break repair have been linked to higher mutation rates near recombination hotspots (ARBEITHUBER *et al.* 2015), which have also been shown to increase with maternal aging (HALLDORSSON *et al.* 2019) and paternal aging (FRANCIOLI *et al.* 2015). If recombination were also mutagenic in honey bees, we would expect windows containing *de novo* mutations to show consistently higher recombination rates than matched-control windows. Indeed, we found that windows containing at least one unique mutation were consistently associated with higher LD-inferred recombination rates than matched control windows, indicating that *de novo* mutations are non-randomly distributed with respect to the local recombination distribution. Our study thus supports the hypothesis that recombination is mutagenic by directly linking the presence of *de novo* mutations in genomic regions with high LD-inferred

recombination rates in honey bees. This link between recombination and mutation also appears to be independent of maternal age. Although mutation rates and the overall types of mutations (transition/transversions) did not change with aging in our dataset, windows with mutations were consistently found in windows with elevated LD-inferred recombination across all age groups. Together, these results suggest that the relationship between recombination and mutation reflects a stable feature of genome organization rather than a process that changes with maternal age, despite the extended reproductive lifespan of honey bee queens.

Because recombination can also influence nucleotide composition through GC-biased gene conversion (gBGC), we additionally examined whether mutation-containing windows differed in GC content from windows without mutations. Over evolutionary timescales, recombination can promote the fixation of G or C alleles, resulting in GC-rich domains that often coincide with regions of high recombination (DURET AND ARNDT 2008; KENT *et al.* 2012; PESSIA *et al.* 2012; WALLBERG *et al.* 2015). If mutations tend to occur in regions with elevated recombination, mutation-containing windows might therefore also be expected to show higher GC content. However, we found no relationship between the presence of a mutation and GC content in our dataset. This result suggests that GC composition does not provide additional support for the recombination–mutation association observed here and is consistent with the view that GC patterns largely reflect long-term recombination processes rather than short-term variation in mutation occurrence.

In this study, we exploited haplodiploidy and the long lifespan of honey bee queens to investigate how maternal aging influences patterns of mutation across the genome. Despite expectations that mutation rates might increase with maternal age due to the accumulation of replication errors or declining DNA repair fidelity (GAO *et al.* 2016), we found no evidence that overall mutation rates or mutation types differed among age groups. Instead, our results suggest that the genomic distribution of mutations is shaped mainly by the underlying structure of the genome. Mutations were consistently less frequent in genic regions than in non-genic regions across queen-level groups, consistent with purifying selection removing deleterious mutations from functional parts of the genome (SINGH *et al.* 2008; DUKLER *et al.* 2022; CHEN *et al.* 2024). We also found that windows containing mutations tended to occur in regions with higher LD-inferred recombination than matched control windows, suggesting that recombination influences where new mutations occur in the genome. This pattern was observed across all age groups and was not accompanied by differences in GC content, indicating that GC composition reflects longer-term recombination processes rather than the short-term placement of new mutations. Together, these results suggest that although honey bee queens reproduce over several years, the processes that shape where mutations occur in the genome remain largely stable throughout the reproductive lifespan of the queen. Future studies with larger

sample sizes, especially including older queens, will help clarify how genome structure influences mutation patterns in eusocial insects.

Figures

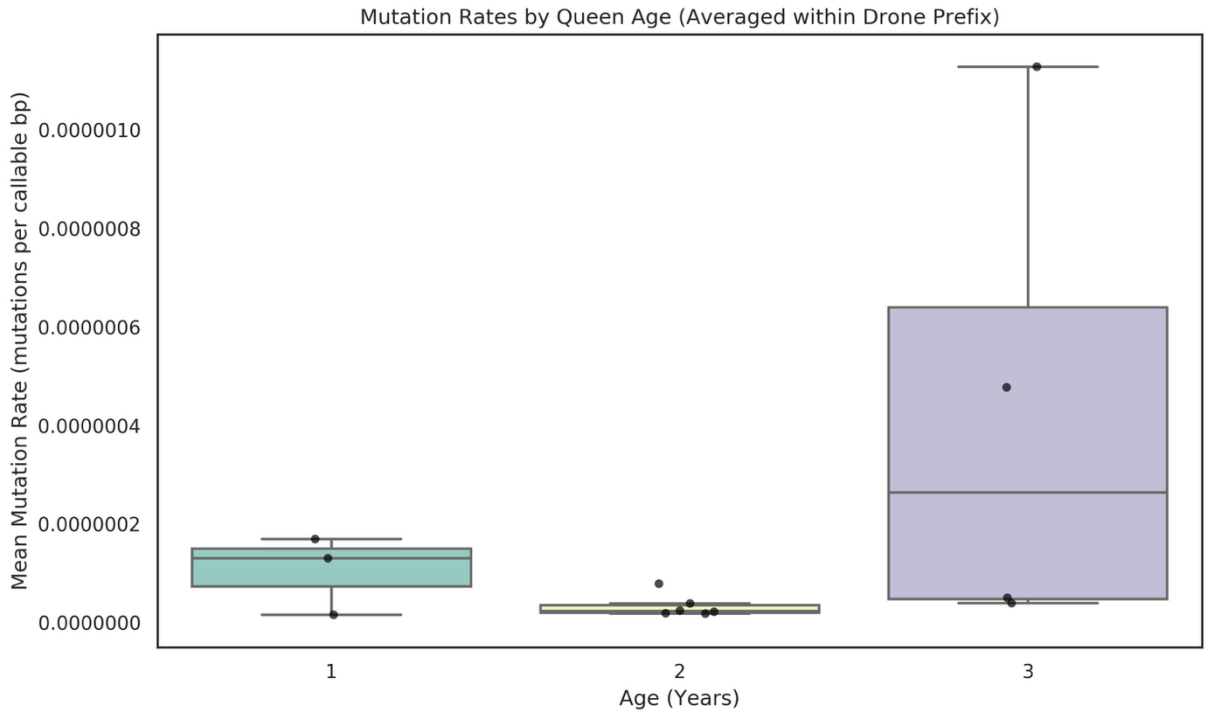


Figure 6-1. *Mutation rates do not significantly vary with queen age.*

Boxplots here show the distribution of mutation rates within each age class. Mutation rates were calculated per individual by dividing the number of observed de novo mutations by the number of callable loci in each genome and averaging across maternal age group. Overall, there is no statistical difference between age group when compared pairwise ($p > 0.05$) and there is no trend of increased mutation rate with queen aging ($p = 0.272$).

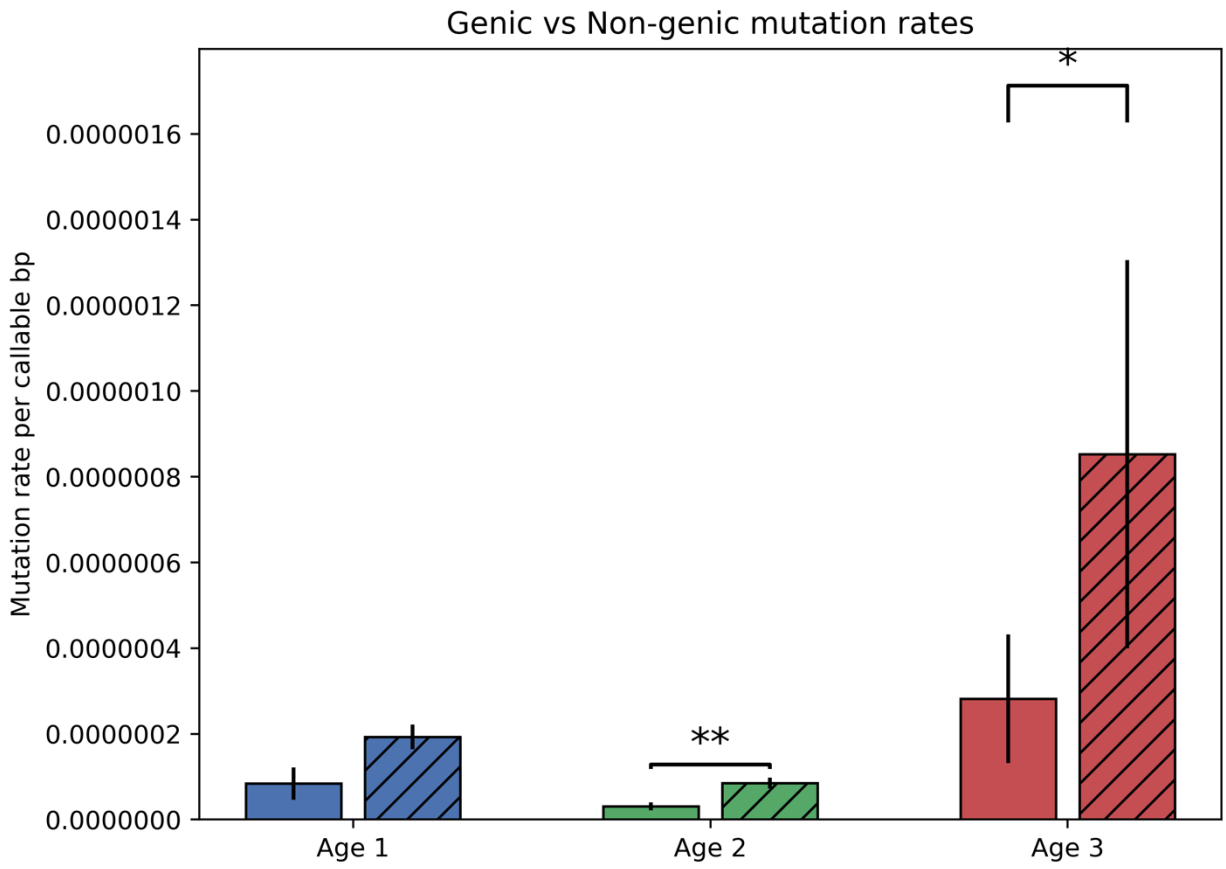


Figure 6-2. *Mutation rates in genic and non-genic regions across three maternal age groups.*

*Solid bars represent genic regions (introns and exons), and hatched bars represent non-genic regions (upstream, downstream, and intergenic combined). Mutation rates were consistently lower in genic regions than in non-genic regions. Asterisks indicate significant differences between region categories based on Wilcoxon signed-rank tests (*p < 0.05, **p < 0.01, ***p<0.001, ****p<0.0001).*

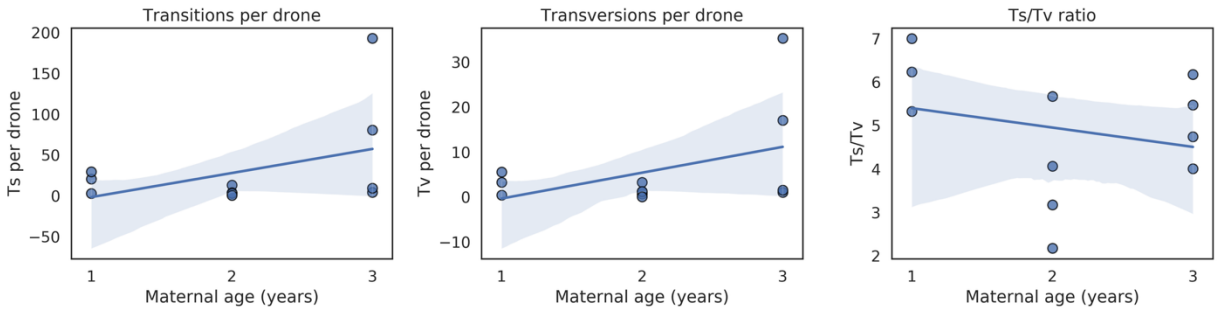


Figure 6-3. *The Relationship between maternal age and mutation types.*

The plots show the average (A) transitions, (B) transversions, and (C) the transition/transversion ratio (Ts/Tv) plotted against age. Each point represents one queen-level group, with age treated as a continuous predictor. Lines show the best-fit linear regression for each mutation type with 95% confidence intervals. No particular mutation type displayed a significant association with aging (all $p > 0.10$), indicating that the frequency of transitions, transversions, and the Ts/Tv ratio does not change with aging.

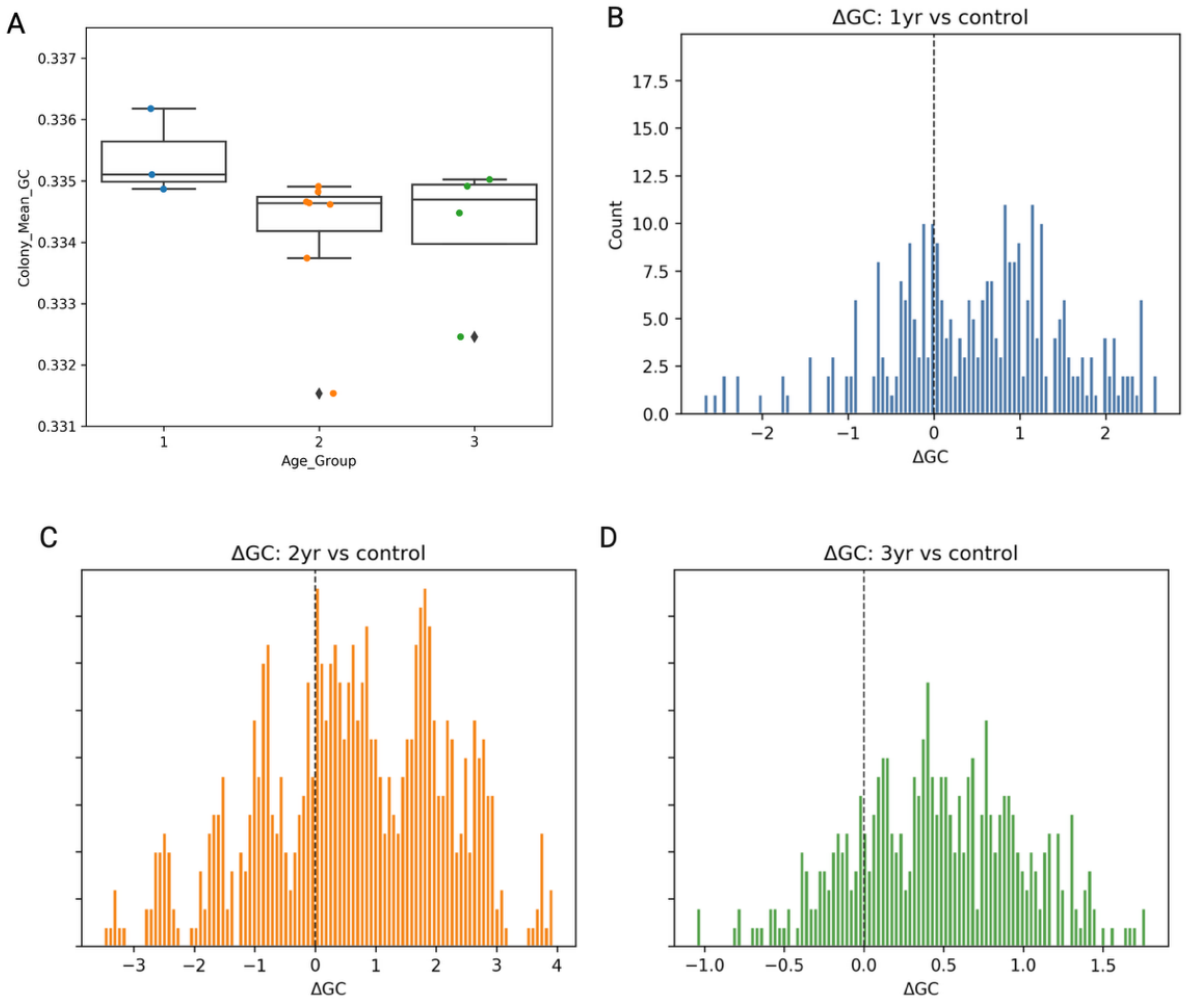


Figure 6-4. *GC content is not higher in windows that contain at least one mutation*
(A) Differences in mean GC content in windows containing at least one mutation across age groups does not significantly differ ($p > 0.05$). (B-D) Differences in GC content found in windows containing at least one unique mutation versus randomly chosen windows across age groups. The vertical dotted line shows that in each age group there was no significant difference in the GC content of windows containing a novel mutation from that of randomly chosen genomic windows ($p > 0.05$).

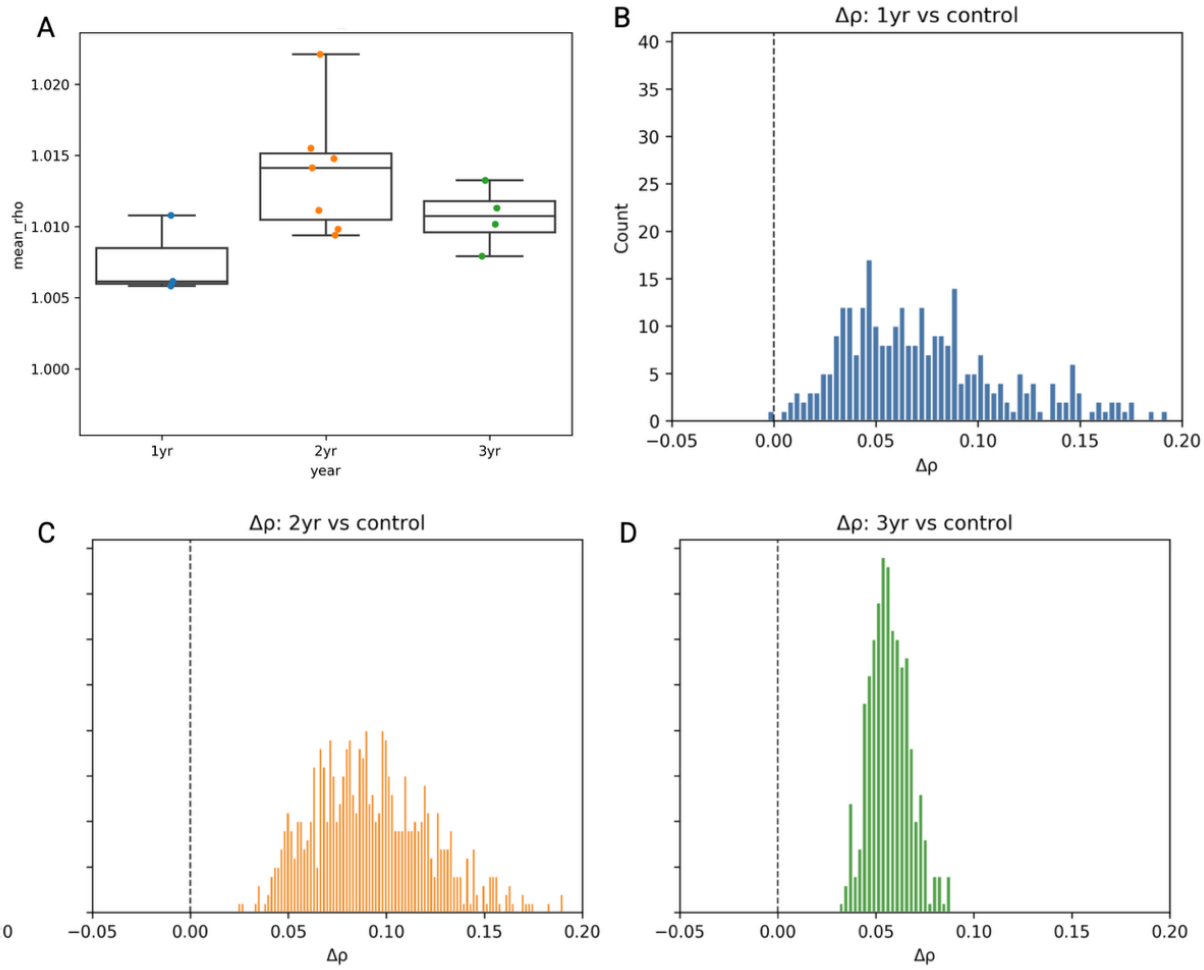


Figure 6-5 LD-inferred recombination changes in windows containing at least one unique mutation with aging. (A) There is no significant difference in the mean rho in windows containing at least one mutation across age groups ($p > 0.05$). (B-D) There are significant differences in the mean rho found among experimental and control across age groups. The dotted line at zero demonstrates the large skew in differences between experimental and control ($p < 0.05$ for each). For easier visualization, the 1- and 2-year age groups are divided into 100 bins for the histogram, but the 3-year age groups are grouped into 1000 bins to fully see the distribution.

Supplemental Tables

All supplemental tables can be found in the supplementary dissertation file

Chapter 7 Conclusions, Limitations and Future Direction

My research has advanced our understanding of how genomic approaches can inform bee evolution, the genetic basis of early stages of social evolution, and how derived features of eusociality such as long-lived queens and high recombination rates can influence processes of mutation. In Chapter 2, I highlighted how recent advances in genomic technologies have transformed the study of bee evolution and social behaviour. Phylogenomic tools such as next-generation sequencing and ultraconserved elements have enabled more accurate reconstructions of bee evolutionary relationships and clarified major evolutionary transitions across bee lineages. Furthermore, advances in genomics and transcriptomics have revealed the complex molecular mechanisms underlying eusocial behaviour that can include regulatory processes such as alternative splicing, non-coding RNAs, and DNA methylation. Together, these developments demonstrate that integrating phylogenomics with molecular biology provides a more complete understanding of both the evolutionary history of bees and the molecular mechanisms underlying their social systems.

In Chapter 3, I discuss how new genomic and experimental tools can be used together to better understand the genetics of eusocial behaviour. Combining population genomics, genome-wide association studies (GWAS), and CRISPR/Cas9 gene editing can allow researchers to connect genetic variation with social traits and directly test how specific genes influence behaviour. By pairing large-scale genomic analyses with experimental validation, this approach helps clarify the complex genetic basis of social behaviour and provides a useful framework for future studies on the evolution of sociality. Here, I argued for a more integrated approach once candidate genes or loci have been identified, an idea that serves as a useful next step to keep in mind for the findings presented in the following chapters.

In Chapters 4 and 5, I used the facultatively social small carpenter bee *Ceratina calcarata* to investigate the genomic basis of social behaviour during the early stages of social evolution. Several theories propose that social behaviour can arise either through the differential regulation of existing genes, in which genes are turned on or off at particular times or in specific tissues to generate behavioural variation, as described by the genetic toolkit hypothesis (True and Carroll 2002; Woodard *et al.* 2011; Rehan and Toth 2015), or through the evolution of novel genes that become specialized for social functions after social behaviour has already evolved (Johnson and Tsutsui 2011). Despite major advances in genomic technologies, the genetic mechanisms underlying the origin and maintenance of eusociality are yet to be fully understood. I found that overall genetic differentiation between social and subsocial behavioural groups was low, indicating that these behaviours occur within a single population rather than representing

genetically distinct lineages. However, a small number of loci showed strong differentiation between behavioural groups, and these loci were primarily located in genes involved in regulatory processes such as transcriptional control and signaling. Similarly, genes under strong positive selection in *C. calcarata* were largely enriched for regulatory functions and were evolutionarily ancient rather than taxonomically restricted. The results from both chapters support the genetic toolkit hypothesis, suggesting that early stages of social evolution are driven largely by changes in the regulation of conserved genes rather than by the evolution of novel, taxonomically restricted genes. Comparisons with other social insects further suggest that while similar functional categories of genes may be targeted by selection, the specific genes involved often differ across lineages.

In my final experimental chapter (Chapter 6), I examined genomic patterns associated with two derived eusocial traits in the honey bee, *Apis mellifera:* the extreme longevity of reproductive queens and the exceptionally high recombination rates observed in this lineage. Because queens reproduce continuously for several years and recombination can introduce mutation through double-stranded breaks of DNA and repair during meiosis, I tested the hypotheses that mutation rates increase with maternal age and that *de novo* mutations are enriched in genomic regions with high recombination. Overall, I did not find that rates of mutation change with maternal aging. However, I did find that mutations are not distributed randomly throughout the genome and are rather shaped by regions of high linkage disequilibrium and likely selection. The locations of mutations along a gene are likely influenced by processes of selection wherein regions that are functionally important, like introns and exons, possessed fewer mutations relative to less constrained regions like upstream, downstream or intergenic regions. Furthermore, I found that mutations were more likely to be found in genomic windows of high linkage disequilibrium. Together, these results provide the first evidence of a potential link between mutation and recombination in honey bees and suggest that genomic processes associated with recombination may influence patterns of mutation accumulation in long-lived eusocial queens.

While this dissertation provides insight into the genomic processes underlying eusocial evolution, several considerations should be kept in mind when interpreting these results. Across the *Ceratina* chapters, inferences about the genetic basis of social behaviour rely on population genomic patterns of differentiation and selection, which identify candidate loci but do not establish causal links between genotype and behaviour. To identify if putative loci or candidate genes are involved in these behaviours, future work will need to include functional validation studies with methods such as experimental manipulation using RNAi or gene-editing approaches such as CRISPR/Cas9 to find a mechanistic link between a locus and social behaviour.

Despite these considerations, the results of this dissertation provide a foundation for future work aimed at understanding the genomic mechanisms underlying social evolution. The candidate loci identified in the *Ceratina* analyses provide promising targets for functional studies that could experimentally test how specific genes influence social behaviour. Expanding genomic analyses to additional species across the spectrum of social complexity will also be important for determining how general the patterns identified here are and whether similar genomic mechanisms repeatedly contribute to the evolution of eusociality. In *Apis mellifera*, future work could incorporate diploid worker offspring alongside haploid drones to better understand how maternal aging interacts with paternal genomes and how mutation and recombination patterns vary among offspring with different fathers. More broadly, while this dissertation highlights regulatory evolution as an important mechanism associated with eusociality, continued improvements in genomic tools and reference genomes may reveal additional contributions from novel genes or lineage-specific adaptations

References

Aamidor, S. E., C. A. M. Cardoso-Junior, J. Harianto, C. J. Nowell, L. Cole *et al.*, 2022 Reproductive plasticity and oogenesis in the queen honey bee (*Apis mellifera*). J Insect Physiol 136**:** 104347.

Akey, J. M., G. Zhang, K. Zhang, L. Jin and M. D. Shriver, 2002 Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12**:** 1805-1814.

Ala, U., 2020 Competing Endogenous RNAs, Non-Coding RNAs and Diseases: An Intertwined Story. Cells 9.

Alexander, D. H., J. Novembre and K. Lange, 2009 Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19**:** 1655-1664.

Almeida, E. A. B., S. Bossert, B. N. Danforth, D. S. Porto, F. V. Freitas *et al.*, 2023 The evolutionary history of bees in time and space. Current Biology 33**:** 3409-3422.

Alvarez-Carretero, S., P. Kapli and Z. Yang, 2023 Beginner's Guide on the Use of PAML to Detect Positive Selection. Mol Biol Evol 40.

Amdam, G. V., A. Csondes, M. K. Fondrk and R. E. Page, Jr., 2006 Complex social behaviour derived from maternal reproductive traits. Nature 439**:** 76-78.

Amdam, G. V., K. Norberg, M. K. Fondrk and R. E. Page, Jr., 2004 Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. Proc Natl Acad Sci U S A 101**:** 11350-11355.

Amdam, G. V., K. Norberg, A. Hagen and S. W. Omholt, 2003 Social exploitation of vitellogenin. PNAS 100**:** 1799-1802.

Anholt, R. R., and T. F. Mackay, 2004 Quantitative genetic analyses of complex behaviours in *Drosophila.* Nat Rev Genet 5**:** 838-849.

Araujo, N. S., and M. C. Arias, 2021 Gene expression and epigenetics reveal species-specific mechanisms acting upon common molecular pathways in the evolution of task division in bees. Sci Rep 11**:** 3654.

Arbeithuber, B., A. J. Betancourt, T. Ebner and I. Tiemann-Boege, 2015 Crossovers are associated with mutation and biased gene conversion at recombination hotspots. Proc Natl Acad Sci U S A 112**:** 2109-2114.

Arsenault, S. V., B. G. Hunt and S. M. Rehan, 2018 The effect of maternal care on gene expression and DNA methylation in a subsocial bee. Nat Commun 9**:** 3468.

Arstingstall, K. A., S. J. DeBano, X. Li, D. E. Wooster, M. M. Rowland *et al.*, 2021 Capabilities and limitations of using DNA metabarcoding to study plant-pollinator interactions. Mol Ecol 30**:** 1-32.

Ascher, J. S., and J. Pickering, 2024 Discover life bee species guide and world checklist (Hymenoptera: Apoidea: Anthophila), pp.

Ashby, R., S. Foret, I. Searle and R. Maleszka, 2016 MicroRNAs in Honey Bee Caste Determination. Sci Rep 6**:** 18794.

Ashwal-Fluss, R., M. Meyer, N. R. Pamudurti, A. Ivanov, O. Bartok *et al.*, 2014 circRNA biogenesis competes with pre-mRNA splicing. Mol Cell 56**:** 55-66.

Avalos, A., M. Fang, H. Pan, A. Ramirez Lluch, A. E. Lipka *et al.*, 2020 Genomic regions influencing aggressive behavior in honey bees are defined by colony allele frequencies. Proc Natl Acad Sci U S A 117**:** 17135-17141.

Aw, A. J., J. McRae, E. Rahmani and Y. S. Song, 2024 Highly parameterized polygenic scores tend to overfit to population stratification via random effects. bioRxiv.

Ballare, K. M., N. S. Pope, A. R. Castilla, S. Cusser, R. P. Metz *et al.*, 2019 Utilizing field collected insects for next generation sequencing: Effects of sampling, storage, and DNA extraction methods. Ecol Evol 9**:** 13690-13705.

Barkdull, M., and C. S. Moreau, 2023 Worker Reproduction and Caste Polymorphism Impact Genome Evolution and Social Genes Across the Ants. Genome Biol Evol 15.

Barribeau, S. M., B. M. Sadd, L. du Plessis, M. J. Brown, S. D. Buechel *et al.*, 2015 A depauperate immune repertoire precedes evolution of sociality in bees. Genome Biol 16**:** 83.

Bartel, D. P., 2018 Metazoan MicroRNAs. Cell 173**:** 20-51.

Basu Ball, W., S. Kar, M. Mukherjee, A. G. Chande, R. Mukhopadhyaya *et al.*, 2011 Uncoupling protein 2 negatively regulates mitochondrial reactive oxygen species generation and induces phosphatase-mediated anti-inflammatory response in experimental visceral leishmaniasis. J Immunol 187**:** 1322-1332.

Batchelor, K. L., K. L. Bell, M. Campos and B. L. Webber, 2023 Can honey bees be used to detect rare plants? Taking an eDNA approach to find the last plants in a weed eradication program. Environmental DNA 5**:** 1516-1526.

Begun, D. J., A. K. Holloway, K. Stevens, L. W. Hillier, Y. P. Poh *et al.*, 2007 Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol 5**:** e310.

Behura, S. K., and C. W. Whitfield, 2010 Correlated expression patterns of microRNA genes with age-dependent behavioural changes in honeybee. Insect Mol Biol 19**:** 431-439.

Beichman, A. C., E. Huerta-Sanchez and K. E. Lohmueller, 2018 Using genomic data to infer historic population dynamics of nonmodel organisms. Annual Review of Ecology, Evolution, and Systematics 49**:** 433-456.

Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W. J. Kent *et al.*, 2004 Ultraconserved elements in the human genome. Science 304**:** 1321-1325.

Bell, K. L., K. J. Turo, A. Lowe, K. Nota, A. Keller *et al.*, 2023 Plants, pollinators and their interactions under global ecological change: The role of pollen DNA metabarcoding. Mol Ecol 32**:** 6345-6362.

Berens, A. J., J. H. Hunt and A. L. Toth, 2015 Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. Mol Biol Evol 32**:** 690-703.

Berg, J. J., A. Harpak, N. Sinnott-Armstrong, A. M. Joergensen, H. Mostafavi *et al.*, 2019 Reduced signal for polygenic adaptation of height in UK Biobank. Elife 8.

Beye, M., I. Gattermeier, M. Hasselmann, T. Gempe, M. Schioett *et al.*, 2006 Exceptionally high levels of recombination across the honey bee genome. Genome Res 16**:** 1339-1344.

Bi, K., T. Linderoth, D. Vanderpool, J. M. Good, R. Nielsen *et al.*, 2013 Unlocking the vault: next-generation museum population genomics. Molecular Ecology 22**:** 6018-6032.

Boardman, L., J. A. P. Marcelino, R. E. Valentin, H. Boncristiani, J. M. Standley *et al.*, 2023 Novel eDNA approaches to monitor Western honey bee (*Apis mellifera L*.) microbial and arthropod communities. Environmental DNA 00**:** 1-18.

Bolger, A. M., M. Lohse and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30**:** 2114-2120.

Booker, T. R., B. C. Jackson and P. D. Keightley, 2017 Detecting positive selection in the genome. BMC Biol 15**:** 98.

Bossert, S., E. A. Murray, E. A. B. Almeida, S. G. Brady, B. B. Blaimer *et al.*, 2019 Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. Molecular Phylogenetics and Evolution 130**:** 121-131.

Bossert, S., E. A. Murray, B. B. Blaimer and B. N. Danforth, 2017 The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data. Molecular Phylogenetics and Evolution 111**:** 149-157.

Bossert, S., A. Pauly, B. N. Danforth, M. C. Orr and E. A. Murray, 2024 Lessons from assembling UCEs: A comparison of common methods and the case of *Clavinomia* (Halictidae). Molecular Ecology Resources 00**:** 1-18.

Bossert, S., T. J. Wood, S. Patiny, D. Michez, E. A. B. Almeida *et al.*, 2021 Phylogeny, biogeography and diversification of the mining bee family Andrenidae. Systematic Entomology 47**:** 283-302.

Bovo, S., A. Ribani, V. J. Utzeri, G. Schiavo, F. Bertolini *et al.*, 2018 Shotgun metagenomics of honey DNA: Evaluation of a methodological approach to describe a multi-kingdom honey bee derived environmental DNA signature. PLoS One 13**:** 1-19.

Bovo, S., V. J. Utzeri, A. Ribani, R. Cabbri and L. Fontanesi, 2020 Shotgun sequencing of honey DNA can describe honey bee derived environmental signatures and the honey bee hologenome complexity. Scientific Reports 10**:** 1-17.

Brady, S. G., J. R. Litman and B. N. Danforth, 2011 Rooting phylogenies using gene duplications: an empirical example from the bees (Apoidea). Molecular Phylogenetics and Evolution 60**:** 295-304.

Branstetter, M. G., A. Müller, T. L. Griswold, M. C. Orr and C. D. Zhu, 2021 Ultraconserved element phylogenomics and biogeography of the agriculturally important mason bee subgenus *Osmia* (*Osmia*). Systematic Entomology 46**:** 453-472.

Brasil, S. N. R., E. P. Kelemen and S. M. Rehan, 2022 Historic DNA uncovers genetic effects of climate change and landscape alteration in two wild bee species. Conservation Genetics 24**:** 85-98.

Brenman-Suttner, D., and A. Zayed, 2024 An integrative genomic toolkit for studying the genetic, evolutionary, and molecular underpinnings of eusociality in insects. Current Opinion in Insect Science 65**:** 101231.

Brenman-Suttner, D. B., S. M. Rehan and A. Zayed, 2025 Exploring the genetics of social behaviour in *C. calcarata*. Sci Rep 15**:** 5580.

Broeckx, B. J. G., L. De Smet, T. Blacquiere, K. Maebe, M. Khalenkow *et al.*, 2019 Honey bee predisposition of resistance to ubiquitous mite infestations. Sci Rep 9**:** 7794.

Browning, B. L., X. Tian, Y. Zhou and S. R. Browning, 2021 Fast two-stage phasing of large-scale sequence data. Am J Hum Genet 108**:** 1880-1890.

Bulmer, M. S., and R. H. Crozier, 2006 Variation in positive selection in termite GNBPs and Relish. Mol Biol Evol 23**:** 317-326.

Bush, S. J., L. Chen, J. M. Tovar-Corona and A. O. Urrutia, 2017 Alternative splicing and the evolution of phenotypic novelty. Philos Trans R Soc Lond B Biol Sci 372.

Call, E., C. Mayer, V. Twort, L. Dietz, N. Wahlberg *et al.*, 2021 Museomics: Phylogenomics of the Moth Family Epicopeiidae (Lepidoptera) Using Target Enrichment. Insect Systematics and Diversity 5**:** 1-10.

Cameron, S. A., 1993 Multiple origins of advanced eusociality in bees inferred from mitochondrial DNA sequences Proceedings of the National Academy of Sciences 90**:** 8687-8691.

Campbell, I. M., T. Gambin, S. Jhangiani, M. L. Grove, N. Veeraraghavan *et al.*, 2016 Multiallelic Positions in the Human Genome: Challenges for Genetic Analyses. Hum Mutat 37**:** 231-234.

Cardinal, S., and B. N. Danforth, 2011 The antiquity and evolutionary history of social behavior in bees. PLoS One 6**:** e21086.

Cardinal, S., and B. N. Danforth, 2013 Bees diversified in the age of eudicots. Proceedings of the Royal Society B: Biological Sciences 280**:** 1-9.

Cardinal, S., and L. Packer, 2007 Phylogenetic analysis of the corbiculate Apinae based on morphology of the sting apparatus (Hymenoptera: Apidae). Cladistics 23**:** 99-118.

Cardinal, S., J. Straka and B. N. Danforth, 2010 Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. Proceedings of the National Academy of Sciences 107**:** 16207-16211.

Cardoso-Junior, C. A. M., K. R. Guidugli-Lazzarini and K. Hartfelder, 2018 DNA methylation affects the lifespan of honey bee (*Apis mellifera L.*) workers - Evidence for a regulatory module that involves vitellogenin expression but is independent of juvenile hormone function. Insect Biochem Mol Biol 92**:** 21-29.

Carninci, P., T. Kasukawa, S. Katayama, J. Gough, M. C. Frith *et al.*, 2006 The Transcriptional landscape of the mammalian genome. Science 309**:** 1559-1563.

Carroll, S. B., J. K. Grenier and S. D. Weatherbee, 2005 From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design, edited by Wiley-Blackwell.

Carthew, R. W., 2021 Gene Regulation and Cellular Metabolism: An Essential Partnership. Trends Genet 37**:** 389-400.

Casillas, S., A. Barbadilla and C. M. Bergman, 2007 Purifying selection maintains highly conserved noncoding sequences in *Drosophila*. Mol Biol Evol 24**:** 2222-2234.

Cavill, E. L., S. Liu, X. Zhou and M. T. P. Gilbert, 2022 To bee, or not to bee? One leg is the question. Molecular Ecology Resources 22**:** 1868-1874.

Chandra, V., I. Fetter-Pruneda, P. R. Oxley, A. Ritger, S. K. McKenzie *et al.*, 2018 Social regulation of insulin signaling and the evolution of eusociality in ants. Science 261**:** 398-402.

Chau, K. D., M. Shamekh, J. Huisken and S. M. Rehan, 2023 The effects of maternal care on the developmental transcriptome and metatranscriptome of a wild bee. Commun Biol 6**:** 904.

Chavez-Galarza, J., D. Henriques, J. S. Johnston, J. C. Azevedo, J. C. Patton *et al.*, 2013 Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. Mol Ecol 22**:** 5890-5907.

Chen, D., H. Chen, Y. Du, D. Zhou, S. Geng *et al.*, 2019 Genome-wide identification of long non-coding RNAs and their regulatory networks involved in *Apis mellifera ligustica* response to *Nosema ceranae* infection. Insects 10**:** 1-26.

Chen, D., H. Chen, Y. Du, Z. Zhu, J. Wang *et al.*, 2020 Systematic identification of circular RNAs and corresponding regulatory networks unveil their potential roles in the midguts of eastern honeybee workers. Applied Microbiology and Biotechnology 104**:** 257-276.

Chen, H., X. Fan, W. Zhang, Y. Ye, Z. Cai *et al.*, 2022 Deciphering the CircRNA-Regulated Response of Western Honey Bee (*Apis mellifera*) Workers to Microsporidian Invasion. Biology (Basel) 11**:** 1-26.

Chen, S., L. C. Francioli, J. K. Goodrich, R. L. Collins, M. Kanai *et al.*, 2024 A genomic mutational constraint map using variation in 76,156 human genomes. Nature 625**:** 92-100.

Chen, X., and J. Fu, 2021 The microRNA miR-14 regulates egg-laying by targeting EcR in honeybees (*Apis mellifera*). Insects 12**:** 1-11.

Chen, X., C. Ma, C. Chen, Q. Lu, W. Shi *et al.*, 2017 Integration of lncRNA-miRNA-mRNA reveals novel insights into oviposition regulation in honey bees. PeerJ 5**:** 1-28.

Chen, X., and W. Shi, 2020 Genome-wide characterization of coding and non-coding RNAs in the ovary of honeybee workers and queens. Apidologie 51**:** 777-792.

Chen, X., D. Wang and J. An, 2023 Circular RNA ame_circ_2015 functions as microRNA sponges in regulating egg-laying of honeybees (*Apis mellifera*). Life (Basel) 13**:** 1-10.

Chen, Z., I. M. Traniello, S. Rana, A. C. Cash-Ahmed, A. L. Sankey *et al.*, 2021 Neurodevelopmental and transcriptomic effects of CRISPR/Cas9-induced somatic *orco* mutation in honey bees. J Neurogenet 35**:** 320-332.

Cheng, F. P., X. F. Hu, L. X. Pan, Z. X. Gong, K. X. Qin *et al.*, 2023 Transcriptome changes of *Apis mellifera* female embryos with *fem* gene knockout by CRISPR/Cas9. Int J Biol Macromol 229**:** 260-267.

Chiu, Y. K., J. C. Hsu, T. Chang, Y. C. Huang and J. Wang, 2020 Mutagenesis mediated by CRISPR/Cas9 in the red imported fire ant, *Solenopsis invicta*. Insectes Sociaux 67**:** 317-326.

Christmas, M. J., J. C. Jones, A. Olsson, O. Wallerman, I. Bunikis *et al.*, 2021 Genetic Barriers to Historical Gene Flow between Cryptic Species of Alpine Bumblebees Revealed by Comparative Population Genomics. Mol Biol Evol 38**:** 3126-3143.

Cilia, G., S. Flaminio and M. Quaranta, 2022 A novel and non-invasive method for DNA extraction from dry bee specimens. Scientific Reports 12**:** 1-9.

Cingolani, P., A. Platts, L. Wang le, M. Coon, T. Nguyen *et al.*, 2012 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.

Cohen, Z. P., O. Francois and S. D. Schoville, 2022 Museum Genomics of an Agricultural Super-Pest, the Colorado Potato Beetle, *Leptinotarsa decemlineata* (Chrysomelidae), Provides Evidence of Adaptation from Standing Variation. Integrative and Comparative Biology 62**:** 1827-1837.

Collins, D. H., I. Mohorianu, M. Beckers, V. Moulton, T. Dalmay *et al.*, 2017 MicroRNAs associated with caste determination and differentiation in a primitively eusocial insect. Scientific Reports 7**:** 1-9.

Conflitti, I. M., M. Arshad Imrit, B. Morrison, S. Sharma, S. R. Colla *et al.*, 2022 Bees in the six: Determinants of bumblebee habitat quality in urban landscapes. Ecology and Evolution 12.

Cook, L. S. J., A. G. Briscoe, V. G. Fonseca, J. Boenigk, G. Woodward *et al.*, 2024 Microbial, holobiont, and Tree of Life eDNA/eRNA for enhanced ecological assessment. Trends in Microbiology**:** 1-18.

Couton, M., F. Viard and F. Altermatt, 2023 Opportunities and inherent limits of using environmental DNA for population genetics. Environmental DNA 5**:** 1048-1064.

Cridland, J. M., S. R. Ramirez, C. A. Dean, A. Sciligo and N. D. Tsutsui, 2018 Genome sequencing of museum specimens reveals rapid changes in the genetic composition of honey bees in California. Genome Biology and Evolution 10**:** 458-472.

Crozier, R. H., 1976 Counter-intuitive property of effective population size. Nature 262**:** 384.

Culina, A., D. M. Linton, R. Pradel, S. Bouwhuis and D. W. Macdonald, 2019 Live fast, don't die young: Survival-reproduction trade-offs in long-lived income breeders. J Anim Ecol 88**:** 746-756.

Cullen, G., J. B. Gilligan, J. G. Guhlin and P. K. Dearden, 2023 Germline progenitors and oocyte production in the honeybee queen ovary. Genetics 225.

Cummins, M., C. Watson, R. J. Edwards and J. S. Mattick, 2024 The Evolution of ultraconserved elements in vertebrates. Genome Biology and Evolution 41**:** 1-17.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27**:** 2156-2158.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan *et al.*, 2021 Twelve years of SAMtools and BCFtools. Gigascience 10.

Danforth, B. N., S. Sipes, J. Fang and S. G. Brady, 2006 The history of early bee diversification based on five genes plus morphology. Proceedings of the National Academy of Sciences 103**:** 15118-15123.

Danziger, A. M., Z. H. Olson and M. Frederich, 2022 Limitations of eDNA analysis for *Carcinus maenas* abundance estimations. BMC Ecol Evol 22**:** 1-11.

Darwin, C., 1809-1882 On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. John Murray, London.

Debevec, A. H., S. Cardinal and B. N. Danforth, 2012 Identifying the sister group to the bees: a molecular phylogeny of Aculeata with an emphasis on the superfamily Apoidea. Zoologica Scripta 41**:** 527-535.

Degirmenci, L., D. Geiger, F. L. Roge Ferreira, A. Keller, B. Krischke *et al.*, 2020 CRISPR/Cas 9-Mediated Mutations as a New Tool for Studying Taste in Honeybees. Chem Senses 45**:** 655-666.

Delabaere, L., H. A. Ertl, D. J. Massey, C. M. Hofley, F. Sohail *et al.*, 2017 Aging impairs double-strand break repair by homologous recombination in *Drosophila* germ cells. Aging Cell 16**:** 320-328.

Derkarabetian, S., L. R. Benavides and G. Giribet, 2019 Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. Molecular Ecology Resources 19**:** 1531-1544.

Derrien, T., R. Johnson, G. Bussotti, A. Tanzer, S. Djebali *et al.*, 2012 The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Research 22**:** 1775-1789.

Desalle, R., B. Schierwater and H. Hadrys, 2017 MtDNA: The small workhorse of evolutionary studies. Front Biosci (Landmark Ed) 22**:** 873-887.

Dietz, L., J. Eberle, C. Mayer, S. Kukowka, C. Bohacz *et al.*, 2022 Standardized nuclear markers improve and homogenize species delimitation in Metazoa. Methods in Ecology and Evolution 14**:** 543-555.

Dogantzis, K. A., B. A. Harpur, A. Rodrigues, L. Beani, A. L. Toth *et al.*, 2018 Insects with similar social complexity show convergent patterns of adaptive molecular evolution. Scientific Reports 8**:** 1-8.

Dogantzis, K. A., R. Raffiudin, R. E. Putra, I. Shaleh, I. M. Conflitti *et al.*, 2024 Post-invasion selection acts on standing genetic variation despite a severe founding bottleneck. Curr Biol 34**:** 1-8.

Dogantzis, K. A., T. Tiwari, I. M. Conflitti, A. Dey, H. M. Patch *et al.*, 2021 Thrice out of Asia and the adaptive radiation of the western honey bee. Science Advances 7**:** 1-10.

Dogantzis, K. A., and A. Zayed, 2019 Recent advances in population and quantitative genomics of honey bees. Curr Opin Insect Sci 31**:** 93-98.

Drapeau, M. D., S. Albert, R. Kucharski, C. Prusko and R. Maleszka, 2006 Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. Genome Research 16**:** 1385-1394.

Dukler, N., M. R. Mughal, R. Ramani, Y. F. Huang and A. Siepel, 2022 Extreme purifying selection against point mutations in the human genome. Nat Commun 13**:** 4312.

Dunne, J., A. Hohn, G. Franke, K. Neumann, P. Breunig *et al.*, 2021 Honey-collecting in prehistoric West Africa from 3500 years ago. Nature Communications 12**:** 1-11.

Durant, D. R., A. J. Berens, A. L. Toth and S. M. Rehan, 2015 Transcriptional profiling of overwintering gene expression in the small carpenter bee, *Ceratina calcarata*. Apidologie 47**:** 572-582.

Duret, L., and P. F. Arndt, 2008 The impact of recombination on nucleotide substitutions in the human genome. PLoS Genet 4**:** e1000071.

Duret, L., and N. Galtier, 2009 Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet 10**:** 285-311.

Eden, E., R. Navon, I. Steinfeld, D. Lipson and Z. Yakhini, 2009 GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10**:** 48.

Efron, B., and R. J. Tibshirani, 1993 *An introduction to the bootstrap*. Taylor and Francis.

Eguchi, K., E. Oguri, T. Sasaki, A. Matsuo, D. D. Nguyen *et al.*, 2020 Revisiting museum collections in the genomic era: potential of MIG-seq for retrieving phylogenetic information from aged minute dry specimens of ants (Hymenoptera: Formicidae) and other small organisms. Myrmecological News 30**:** 151-159.

Eilertson, K. E., J. G. Booth and C. D. Bustamante, 2012 SnIPRE: selection inference using a Poisson random effects model. PLoS Comput Biol 8**:** e1002806.

Eimanifar, A., S. A. Brooks, T. Bustamante and J. D. Ellis, 2018 Population genomics and morphometric assignment of western honey bees (*Apis mellifera L*.) in the Republic of South Africa. BMC Genomics 19.

Elbogen, E., 2012 Identification of Burseraceae trees from Peru: a comparison of the nuclear DNA marker ITS and the plastid DNA marker rbcL for DNA barcoding. Berkeley Scientific Journal 16**:** 1-10.

Engsontia, P., U. Sangket, H. M. Robertson and C. Satasook, 2015 Diversification of the ant odorant receptor gene family and positive selection on candidate cuticular hydrocarbon receptors. BMC Res Notes 8**:** 380.

Erokhin, M. M., A. V. Mikhailova, P. G. Georgiev and D. A. Chetverina, 2019 Effect of Transcription on the white Gene Enhancer Integrated into the Intron. Dokl Biochem Biophys 484**:** 45-47.

Espregueira Themudo, G., A. Rey-Iglesia, L. Robles Tascon, A. Bruun Jensen, R. R. da Fonseca *et al.*, 2020 Declining genetic diversity of European honeybees along the twentieth century. Sci Rep 10**:** 1-12.

Eyre-Walker, A., P. D. Keightley, N. G. C. Smith and D. Gaffney, 2002 Quantifying the Slightly Deleterious Mutation Model of Molecular Evolution. Mol. Biol. Evol. 19**:** 2142-2149.

Fadista, J., A. K. Manning, J. C. Florez and L. Groop, 2016 The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. Eur J Hum Genet 24**:** 1202-1205.

Faircloth, B. C., M. G. Branstetter, N. D. White and S. G. Brady, 2015 Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Mol Ecol Resour 15**:** 489-501.

Fang, Y., and M. J. Fullwood, 2016 Roles, functions, and mechanisms of long non-coding RNAs in cancer. Genomics Proteomics Bioinformatics 14**:** 42-54.

Favreau, E., K. S. Geist, C. D. R. Wyatt, A. L. Toth, S. Sumner *et al.*, 2023 Co-expression Gene Networks and Machine-learning Algorithms Unveil a Core Genetic Toolkit for Reproductive Division of Labour in Rudimentary Insect Societies. Genome Biol Evol 15.

Feng, W., J. Huang, Z. Zhang, H. Nie, Y. Lin *et al.*, 2022 Understanding of waggle dance in the Honey Bee (*Apis mellifera*) from the perspective of long non-coding RNA. Insects 13**:** 1-17.

Feng, X. J., G. F. Jiang and Z. Fan, 2015 Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, *Ceracris kiangsu*. Sci Rep 5**:** 13758.

Ferreira, P. G., S. Patalano, R. Chauhan, R. Ffrench-Constant, T. Gabaldón *et al.*, 2013 Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. Genome Biol 14.

Fischman, B. J., S. H. Woodard and G. E. Robinson, 2011 Molecular evolutionary analyses of insect societies. Proc Natl Acad Sci U S A 108 Suppl 2**:** 10847-10854.

Flatt, T., 2011 Survival costs of reproduction in *Drosophila*. Exp Gerontol 46**:** 369-375.

Foret, S., R. Kucharski, M. Pellegrini, S. Feng, S. E. Jacobsen *et al.*, 2012 DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proceedings of the National Academy of Sciences 109**:** 4968-4973.

Fouks, B., and H. M. G. Lattorff, 2016 Contrasting evolutionary rates between social and parasitic bumblebees for three social effect genes. Frontiers in Ecology and Evolution 4**:** 1-9.

Francioli, L. C., P. P. Polak, A. Koren, A. Menelaou, S. Chun *et al.*, 2015 Genome-wide patterns and properties of de novo mutations in humans. Nat Genet 47**:** 822-826.

Fraser, H., 1931 Beekeeping in Antiquity Univ. London Press, London.

Freitas, F. V., M. G. Branstetter, V. H. Franceschini-Santos, A. Dorchin, K. W. Wright *et al.*, 2023 UCE phylogenomics, biogeography, and classification of long-horned bees (Hymenoptera: Apidae: Eucerini), with insights on using specimens with extremely degraded DNA. Insect Systematics and Diversity 7**:** 1-21.

Freitas, F. V., M. G. Branstetter, T. Griswold and E. A. B. Almeida, 2021 Partitioned gene-tree analyses and gene-based topology testing help resolve incongruence in a phylogenomic study of host-specialist bees (Apidae: Eucerinae). Mol Biol Evol 38**:** 1090-1100.

Fuller, Z. L., E. L. Niño, H. M. Patch, O. C. Bedoya-Reina, T. Baumgarten *et al.*, 2015 Genome-wide analysis of signatures of selection in populations of African honey bees (*Apis mellifera*) using new web-based tools. BMC Genomics 16.

Gao, Z., P. Moorjani, T. A. Sasani, B. S. Pedersen, A. R. Quinlan *et al.*, 2019 Overlooked roles of DNA damage and maternal age in generating human germline mutations. Proc Natl Acad Sci U S A 116**:** 9491-9500.

Gao, Z., M. J. Wyman, G. Sella and M. Przeworski, 2016 Interpreting the Dependence of Mutation Rates on Age and Time. PLoS Biol 14**:** e1002355.

Garrett, N. R., J. Watkins, N. B. Simmons, B. Fenton, A. Maeda-Obregon *et al.*, 2023 Airborne eDNA documents a diverse and ecologically complex tropical bat and other mammal community. Environmental DNA 5**:** 350-362.

Gempe, T., M. Hasselmann, M. Schiott, G. Hause, M. Otte *et al.*, 2009 Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway. PLoS Biol 7**:** e1000222.

Getz, W. M., and R. E. Page, 1991 Chemosensory Kin-communication Systems and Kin Recognition in Honey Bees. Ethology 87**:** 298-315.

Gill, K. P., E. van Wilgenburg, D. L. Macmillan and M. A. Elgar, 2013 Density of antennal sensilla influences efficacy of communication in a social insect. Am Nat 182**:** 834-840.

Glastad, K. M., R. J. Graham, L. Ju, J. Roessler, C. M. Brady *et al.*, 2020 Epigenetic Regulator CoREST Controls Social Behavior in Ants. Mol Cell 77**:** 338-351 e336.

Goldstein, D., 2009 Common genetic variation and human trait. NEJM 360**:** 1696-1698.

Gonzalez-Forero, M., and J. Pena, 2021 Eusociality through conflict dissolution. Proc Biol Sci 288**:** 20210386.

Goodman, A., E. Tolman, R. Uche-Dike, J. Abbott, J. W. Breinholt *et al.*, 2023 Assessment of targeted enrichment locus capture across time and museums using odonate specimens. Insect Systematics and Diversity 7**:** 1-9.

Gospocic, J., E. J. Shields, K. M. Glastad, Y. Lin, C. A. Penick *et al.*, 2017 The Neuropeptide Corazonin Controls Social Behavior and Caste Identity in Ants. Cell 170**:** 748-759 e712.

Gous, A., C. D. Eardley, S. D. Johnson, D. Z. H. Swanevelder and S. Willows-Munro, 2021 Floral hosts of leaf-cutter bees (Megachilidae) in a biodiversity hotspot revealed by pollen DNA metabarcoding of historic specimens. PLoS One 16**:** 1-16.

Graham, A. M., M. D. Munday, O. Kaftanoglu, R. E. Page Jr, G. V. Amdam *et al.*, 2011 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.

Gramates, L. S., J. Agapite, H. Attrill, B. R. Calvi, M. A. Crosby *et al.*, 2022 FlyBase: a guided tour of highlighted features. Genetics 220**:** iyac035.

Greenberg, J. K., J. Xia, X. Zhou, S. R. Thatcher, X. Gu *et al.*, 2012 Behavioral plasticity in honey bees is associated with differences in brain microRNA transcriptome. Genes, Brain and Behavior 11**:** 660-670.

Gregoric, M., D. Kutnjak, K. Bacnik, C. Gostincar, A. Pecman *et al.*, 2022 Spider webs as eDNA samplers: Biodiversity assessment across the tree of life. Molecular Ecology Resources 22**:** 2534-2545.

Grewe, F., M. R. Kronforst, N. E. Pierce and C. S. Moreau, 2021 Museum genomics reveals the Xerces blue butterfly (*Glaucopsyche xerces*) was a distinct species driven to extinction. Biology Letters 17**:** 20210123.

Groom, S. V., and S. M. Rehan, 2020 Climate-mediated behavioural variability in facultatively social bees. Biological Journal of the Linnean Society 125**:** 165-170.

Groom, S. V. C., and S. M. Rehan, 2018 Climate mediated behavioral variability in faculatively social bees. Biological Journal of the Linnean Society 125**:** 165-170.

Grozinger, C. M., Y. Fan, S. E. Hoover and M. L. Winston, 2007 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.

Grozinger, C. M., N. M. Sharabash, C. W. Whitfield and G. E. Robinson, 2003 Pheromone-mediated gene expression in the honey bee brain. PNAS 100**:** 14519–14525.

Gueuning, M., J. E. Frey and C. Praz, 2020 Ultraconserved yet informative for species delimitation: Ultraconserved elements resolve long-standing systematic enigma in Central European bees. Molecular Ecology 29**:** 4203-4220.

Guichard, M., B. Dainat, S. Eynard, A. Vignal, B. Servin *et al.*, 2021 Identification of quantitative trait loci associated with calmness and gentleness in honey bees using whole-genome sequences. Anim Genet 52**:** 472-481.

Guo, R., S. Wang, S. Guo, X. Fan, H. Zang *et al.*, 2023 Regulatory roles of long non-coding RNAs relevant to antioxidant enzymes and immune responses of *Apis cerana* larvae following *Ascosphaera apis* invasion. International Journal of Molecular Sciences 24.

Guo, X., S. Su, G. Skogerboe, S. Dai, W. Li *et al.*, 2013 Recipe for a busy bee: microRNAs in honey bee caste determination. PLoS One 8**:** 1-10.

Hagler, J. R., and C. G. Jackson, 2001 Methods for marking insects: current techniques and future prospects. Annu Rev Entomol 46**:** 511-543.

Halldorsson, B. V., G. Palsson, O. A. Stefansson, H. Jonsson, M. T. Hardarson *et al.*, 2019 Characterizing mutagenic effects of recombination through a sequence-level genetic map. Science 363.

Halligan, D. L., A. Eyre-Walker, P. Andolfatto and P. D. Keightley, 2004 Patterns of evolutionary constraints in intronic and intergenic DNA of *Drosophila*. Genome Res 14**:** 273-279.

Hamilton, W. D., 1964 The Genetical Evolution of Social Behaviour. J Theor Biol 7**:** 1-16.

Hansen, T. B., T. I. Jensen, B. H. Clausen, J. B. Bramsen, B. Finsen *et al.*, 2013 Natural RNA circles function as efficient microRNA sponges. Nature 495**:** 384-388.

Harper, L. R., M. L. Niemiller, J. B. Benito, L. E. Paddock, E. Knittle *et al.*, 2022 BeeDNA: Microfluidic environmental DNA metabarcoding as a tool for connecting plant and pollinator communities. Environmental DNA 5**:** 191-211.

Harpur, B. A., A. Dey and S. P. Jennifer R. Albert, Heather M. Hines, Martin Hasselmann, Laurence Packer, Amro Zayed, 2017 Queens and workers contribute differently to adaptive evolution in bumble bees and honey bees. Genome Biology and Evolution 9**:** 2395-2402.

Harpur, B. A., M. M. Guarna, E. Huxter, H. Higo, K. M. Moon *et al.*, 2019 Integrative Genomics Reveals the Genetics and Evolution of the Honey Bee's Social Immune System. Genome Biology and Evolution 11**:** 937-948.

Harpur, B. A., S. M. Kadri, R. O. Orsi, C. W. Whitfield and A. Zayed, 2020 Defense Response in Brazilian Honey Bees (*Apis mellifera scutellata* x spp.) Is Underpinned by Complex Patterns of Admixture. Genome Biol Evol 12**:** 1367-1377.

Harpur, B. A., C. F. Kent, D. Molodtsova, J. M. Lebon, A. S. Alqarni *et al.*, 2014 Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. Proc Natl Acad Sci U S A 111**:** 2614-2619.

Harpur, B. A., and S. M. Rehan, 2021 Connecting social polymorphism to single nucleotide polymorphism: population genomics of the small carpenter bee, *Ceratina australensis*. Biological Journal of the Linnean Society 132**:** 945-954.

Harpur, B. A., and A. Zayed, 2013 Accelerated evolution of innate immunity proteins in social insects: adaptive evolution or relaxed constraint? Mol Biol Evol 30**:** 1665-1674.

Harrison, M. C., E. Jongepier, H. M. Robertson, N. Arning, T. Bitard-Feildel *et al.*, 2018 Hemimetabolous genomes reveal molecular basis of termite eusociality. Nat Ecol Evol 2**:** 557-566.

Hartfield, M., S. P. Otto and P. D. Keightley, 2010 The role of advantageous mutations in enhancing the evolution of a recombination modifier. Genetics 184**:** 1153-1164.

Hasselmann, M., and M. Beye, 2004 Signatures of selection among sex-determining alleles of the honey bee. PNAS 101.

Hasselmann, M., L. Ferretti and A. Zayed, 2015 Beyond fruit-flies: population genomic advances in non-*Drosophila* arthropods. Brief Funct Genomics 14**:** 424-431.

Hedtke, S. M., S. Patiny and B. N. Danforth, 2013 The bee tree of life: a supermatrix approach to apoid phylogeny and biogeography. BMC Evolutionary Biology 13**:** 1-13.

Henriquez-Piskulich, P., A. F. Hugall and D. Stuart-Fox, 2024 A supermatrix phylogeny of the world's bees (Hymenoptera: Anthophila). Molecular Phylogenetics and Evolution 190**:** 107963.

Herb, B. R., 2014 Epigenetics as an answer to Darwin's "special difficulty". Frontiers in Genetics 5**:** 1-5.

Hilbe, J. M., 2014 Modeling count data, pp. 126-161. Cambridge University Press.

Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genetical Research 8**:** 269-294.

Hischhorn, J., K. Lohmueller, E. Byrne and K. Hirschhorn, 2002 A comprehensive review of genetic association studies. Genetics in Medicine 4**:** 45-61.

Hu, X. F., B. Zhang, C. H. Liao and Z. J. Zeng, 2019 High-Efficiency CRISPR/Cas9-Mediated Gene Editing in Honeybee (*Apis mellifera*) Embryos. G3 (Bethesda) 9**:** 1759-1766.

Huang, M., J. Dong, H. Guo, M. Xiao and D. Wang, 2021 Identification of long noncoding RNAs reveals the effects of dinotefuran on the brain in *Apis mellifera* (Hymenopptera: Apidae). BMC Genomics 22**:** 1-13.

Huang, M., J. Dong, H. Guo, M. Xiao and D. Wang, 2022a Identification of circular RNAs and corresponding regulatory networks reveals potential roles in the brains of honey bee workers exposed to dinotefuran. Pesticide Biochemistry and Physiology 180**:** 104994.

Huang, Q., Y. Q. Zhu, B. Fouks, X. J. He, Q. S. Niu *et al.*, 2022b Signatures of Positive Selection in the Genome of *Apis mellifera carnica*: A Subspecies of European Honeybees. Life (Basel) 12.

Hughes, W. O. H., B. P. Oldroyd, M. Beekman and F. L. Ratnieks, 2008 Ancestral monogamy shows kin selection is key to the evolution of eusociality. Science 320.

Humann, F. C., G. J. Tiberio and K. Hartfelder, 2013 Sequence and expression characteristics of long noncoding RNAs in honey bee caste development--potential novel regulators for transgressive ovary size. PLoS One 8**:** 1-9.

Hunter, J. D., 2007 MATPLOTLIB: A 2D Graphics Environment. Scientific Programming

Hunter, N., 2015 Meiotic Recombination: The Essence of Heredity. Cold Spring Harb Perspect Biol 7.

Hwang, H. Y., and J. Wang, 2023 Effect of recombination on genetic diversity of *Caenorhabditis elegans*. Sci Rep 13**:** 16425.

Imrit, M. A., K. A. Dogantzis, B. A. Harpur and A. Zayed, 2020 Eusociality influences the strength of negative selection on insect genomes. Proceedings of the Royal Society B: Biological Sciences 287.

Ingolia, N. T., L. F. Lareau and J. S. Weissman, 2011 Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell 147**:** 789-802.

International HapMap, C., 2005 A haplotype map of the human genome. Nature 437**:** 1299-1320.

Ivasyk, I., L. Olivos-Cisneros, S. Valdes-Rodriguez, M. Droual, H. Jang *et al.*, 2023 DNMT1 mutant ants develop normally but have disrupted oogenesis. Nat Commun 14**:** 2201.

Jabalameli, M. R., C. Horscroft, A. Vergara-Lope, R. J. Pengelly and A. Collins, 2019 Gene-dense autosomal chromosomes show evidence for increased selection. Heredity (Edinb) 123**:** 774-783.

Jaenisch, R., and A. Bird, 2003 Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet 33 Suppl**:** 245-254.

Jang, S., P. Mergaert, T. Ohbayashi, K. Ishigami, S. Shigenobu *et al.*, 2021 Dual oxidase enables insect gut symbiosis by mediating respiratory network formation. Proc Natl Acad Sci U S A 118.

Jasper, W. C., T. A. Linksvayer, J. Atallah, D. Friedman, J. C. Chiu *et al.*, 2015 Large-scale coding sequence change underlies the evolution of postdevelopmental novelty in honey bees. Mol Biol Evol 32**:** 334-346.

Jayakodi, M., J. W. Jung, D. Park, Y. J. Ahn, S. C. Lee *et al.*, 2015 Genome-wide characterization of long intergenic non-coding RNAs (lincRNAs) provides new insight into viral diseases in honey bees *Apis cerana* and *Apis mellifera*. BMC Genomics 16**:** 1-12.

Ji, Y., X. Li, T. Ji, J. Tang, L. Qiu *et al.*, 2020 Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee. Science Advances 6**:** 1-14.

Johnson, B. R., and N. D. Tsutsui, 2011 Taxonomically restricted genes are associated with the evolution of sociality in the honey bee.

Johnson, K. P., 2019 Putting the genome in insect phylogenomics. Current Opinion in Insect Science 36**:** 111-117.

Johnson, R. M., B. A. Harpur, K. A. Dogantzis, A. Zayed and M. R. Berenbaum, 2018 Genomic footprint of evolution of eusociality in bees: floral food use and CYPome "blooms". Insectes Sociaux 65**:** 445-454.

Johri, P., C. F. Aquadro, M. Beaumont, B. Charlesworth, L. Excoffier *et al.*, 2022 Recommendations for improving statistical inference in population genomics. PLoS Biol 20**:** e3001669.

Jones, B. M., V. D. Rao, T. Gernat, T. Jagla, A. C. Cash-Ahmed *et al.*, 2020 Individual differences in honey bee behavior enabled by plasticity in brain gene regulatory networks. Elife 9.

Jones, B. M., B. E. R. Rubin, O. Dudchenko, C. J. Kingwell, I. M. Traniello *et al.*, 2023 Convergent selection on hormone signaling shaped social evolution in bees. Nature Ecology and Evolution 7**:** pages 557–569.

Kannan, K., M. Shook, Y. Li, G. E. Robinson and J. Ma, 2019 Comparative analysis of brain and fat body gene splicing patterns in the honey bee, *Apis mellifera*. G3 (Bethesda) 9**:** 1055-1063.

Kapheim, K. M., B. M. Jones, H. Pan, C. Li, B. A. Harpur *et al.*, 2020a Developmental plasticity shapes social traits and selection in a facultatively eusocial bee. Proc Natl Acad Sci U S A 117**:** 13615-13625.

Kapheim, K. M., B. M. Jones, E. Sovik, E. Stolle, R. M. Waterhouse *et al.*, 2020b Brain microRNAs among social and solitary bees. Royal Society Open Science 7**:** 1-11.

Kapheim, K. M., H. Pan, C. Li, C. Blatti, 3rd, B. A. Harpur *et al.*, 2019 Draft Genome Assembly and Population Genetics of an Agricultural Pollinator, the Solitary Alkali Bee (Halictidae: *Nomia melanderi*). G3 (Bethesda) 9**:** 625-634.

Kapheim, K. M., H. Pan, C. Li, S. L. Salzberg, Daniela Puiu *et al.*, 2015 Genomic signatures of evolutionary transitions from solitary to group living. Science 348.

Kapranov, P., J. Cheng, S. Dike, D. A. Nix, R. Duttagupta *et al.*, 2007 RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316**:** 1484-1488.

Katumo, D. M., H. Liang, A. C. Ochola, M. Lv, Q. F. Wang *et al.*, 2022 Pollinator diversity benefits natural and agricultural ecosystems, environmental health, and human welfare. Plant Divers 44**:** 429-435.

Kawakami, T., A. Wallberg, A. Olsson, D. Wintermantel, J. R. de Miranda *et al.*, 2019 Substantial Heritable Variation in Recombination Rate on Multiple Scales in Honeybees and Bumblebees. Genetics 212**:** 1101-1119.

Keightley, P. D., R. W. Ness, D. L. Halligan and P. R. Haddrill, 2014 Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. Genetics 196**:** 313-320.

Keightley, P. D., A. Pinharanda, R. W. Ness, F. Simpson, K. K. Dasmahapatra *et al.*, 2015 Estimation of the spontaneous mutation rate in *Heliconius melpomene*. Mol Biol Evol 32**:** 239-243.

Kelemen, E. P., and S. M. Rehan, 2020 Opposing pressures of climate and land-use change on a native bee. Global Change Biology 00**:** 1-10.

Kent, C. F., S. Minaei, B. A. Harpur and A. Zayed, 2012 Recombination is associated with the evolution of genome structure and worker behavior in honey bees. Proceedings of the National Academy of Sciences 109**:** 18012-18017.

Kent, C. F., and A. Zayed, 2013 Evolution of recombination and genome structure in eusocial insects. Commun Integr Biol 6**:** e22919.

Kent, C. F., and A. Zayed, 2015 Population Genomic and Phylogenomic Insights into the Evolution of Physiology and Behaviour in Social Insects, pp. 293-324 in *Genomics, Physiology and Behaviour of Social Insects*.

Khalifa, S. A. M., E. H. Elshafiey, A. A. Shetaia, A. A. A. El-Wahed, A. F. Algethami *et al.*, 2021 Overview of Bee Pollination and Its Economic Value for Crop Production. Insects 12**:** 1-23.

Kirsch-Gerweck, B., L. Bohnenkamper, M. T. Henrichs, J. N. Alanko, H. Bannai *et al.*, 2023 HaploBlocks: Efficient Detection of Positive Selection in Large Population Genomic Datasets. Mol Biol Evol 40**:** 1-12.

Kiya, T., T. Kunieda and T. Kubo, 2007 Increased neural activity of a mushroom body neuron subtype in the brains of forager honeybees. PLoS One 2**:** 1-9.

Kiya, T., A. Ugajin, T. Kunieda and T. Kubo, 2012 Identification of *kakusei,* a nuclear non-coding RNA, as an immediate early gene from the honeybee, and its application for neuroethological study. Int J Mol Sci 13**:** 15496-15509.

Klepke, M. J., E. E. Sigsgaard, M. R. Jensen, K. Olsen and P. F. Thomsen, 2022 Accumulation and diversity of airborne, eukaryotic environmental DNA. Environmental DNA 4**:** 1323-1339.

Kocher, S., C. Li, W. Yang, H. Tan, S. V. Yi *et al.*, 2013 The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. Genome Biol 14.

Kocher, S. D., R. Mallarino, B. E. R. Rubin, D. W. Yu, H. E. Hoekstra *et al.*, 2018 The genetic basis of a social polymorphism in halictid bees. Nat Commun 9**:** 4338.

Kohno, H., and T. Kubo, 2018 *mKast* is dispensable for normal development and sexual maturation of the male European honeybee. Sci Rep 8**:** 11877.

Kohno, H., and T. Kubo, 2019 Genetics in the Honey Bee: Achievements and Prospects toward the Functional Analysis of Molecular and Neural Mechanisms Underlying Social Behaviors. Insects 10.

Kohno, H., S. Suenami, H. Takeuchi, T. Sasaki and T. Kubo, 2016 Production of Knockout Mutants by CRISPR/Cas9 in the European Honeybee, *Apis mellifera* L. Zoolog Sci 33**:** 505-512.

Kolter, A., M. Husemann, L. Podsiadlowski and B. Gemeinholzer, 2023 Pollen metabarcoding of museum specimens and recently collected bumblebees (*Bombus*) indicates foraging shifts. Metabarcoding and Metagenomics 7**:** 89-119.

Kong, A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem *et al.*, 2012 Rate of de novo mutations and the importance of father's age to disease risk. Nature 488**:** 471-475.

Konu, M., J. Kulmuni and L. Viljakainen, 2023 Genetic modification of the ant *Lasius niger* using CRISPR-Cas9 technology. Insect Mol Biol 32**:** 11-25.

Korlevic, P., E. McAlister, M. Mayho, A. Makunin, P. Flicek *et al.*, 2021 A minimally morphologically destructive approach for DNA retrieval and whole-genome shotgun sequencing of pinned historic dipteran vector species. Genome Biology and Evolution 13**:** 1-13.

Kowalska, Z., F. Pniewski and A. Latała, 2019 DNA barcoding – A new device in phycologist's toolbox. Ecohydrology & Hydrobiology 19**:** 417-427.

Kozomara, A., M. Birgaoanu and S. Griffiths-Jones, 2019 miRBase: from microRNA sequences to function. Nucleic Acids Research 47**:** D155-D162.

Kreider, J. J., I. Pen and B. H. Kramer, 2021 Antagonistic pleiotropy and the evolution of extraordinary lifespans in eusocial organisms. Evolution Letters 5**:** 178-186.

Kribelbauer, J. F., X. J. Lu, R. Rohs, R. S. Mann and H. J. Bussemaker, 2020 Toward a mechanistic understanding of DNA methylation readout by transcription factors. Journal of Molecular Biology 432**:** 1801-1815.

Kritsky, G., 2017 Beekeeping from antiquity through the Middle Ages. Annual Review of Entomology 62**:** 249-264.

Kucharski, R., J. Maleszka, S. Foret and R. Maleszka, 2008 Nutritional Control of Reproductive Status in Honeybees via DNA Methylation. Science 319**:** 1827-1830.

Kuznetsov, D., F. Tegenfeldt, M. Manni, M. Seppey, M. Berkeley *et al.*, 2022 OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res.

Kyriacou, R. G., P. O. Mulhair and P. W. H. Holland, 2024 GC Content Across Insect Genomes: Phylogenetic Patterns, Causes and Consequences. J Mol Evol 92**:** 138-152.

Ladin, Z. S., B. Ferrell, J. T. Dums, R. M. Moore, D. F. Levia *et al.*, 2021 Assessing the efficacy of eDNA metabarcoding for measuring microbial biodiversity within forest ecosystems. Scientific Reports 11**:** 1-14.

Lago, D. C., L. C. Nora, M. Hasselmann and K. Hartfelder, 2023 Positive selection in cytochrome P450 genes is associated with gonad phenotype and mating strategy in social bees. Sci Rep 13**:** 5921.

Lagunas-Robles, G., J. Purcell and A. Brelsford, 2021 Linked supergenes underlie split sex ratio and social organization in an ant. Proc Natl Acad Sci U S A 118.

Lawson, S. P., K. N. Ciaccio and S. M. Rehan, 2016 Maternal manipulation of pollen provisions affects worker production in a small carpenter bee. Behavioral Ecology and Sociobiology 70**:** 1891-1900.

Lawson, S. P., S. L. Helmreich and S. M. Rehan, 2017 Effects of nutritional deprivation on development and behavior in the subsocial bee *Ceratina calcarata* (Hymenoptera: Xylocopinae). J Exp Biol 220**:** 4456-4462.

Lawson, S. P., W. A. Shell, S. S. Lombard and S. M. Rehan, 2018 Climatic variation across a latitudinal gradient affect phenology and group size, but not social complexity in small carpenter bees. Insectes Sociaux 65**:** 483-492.

Leroy, T., P. Faux, B. Basso, S. Eynard, D. Wragg *et al.*, 2024 Inferring Long-Term and Short-Term Determinants of Genetic Diversity in Honey Bees: Beekeeping Impact and Conservation Strategies. Mol Biol Evol 41**:** 1-21.

Lester, P. J., M. Bulgarella, J. W. Baty, P. K. Dearden, J. Guhlin *et al.*, 2020 The potential for a CRISPR gene drive to eradicate or suppress globally invasive social wasps. Sci Rep 10.

Li, B., Y. Duan, Z. Du, X. Wang, S. Liu *et al.*, 2024 Natural selection and genetic diversity maintenance in a parasitic wasp during continuous biological control application. Nat Commun 15**:** 1-16.

Liang, L., Z. Li, Q. Li, X. Wang, S. Su *et al.*, 2021 Expansion of CRISPR Targeting Sites Using an Integrated Gene-Editing System in *Apis mellifera*. Insects 12**:** 1-10.

Libbrecht, R., and L. Keller, 2015 The making of eusociality: insights from two bumblebee genomes. Genome Biology 16**:** 1-2.

Lindblad-Toh, K., M. Garber, O. Zuk, M. F. Lin, B. J. Parker *et al.*, 2011 A high-resolution map of human evolutionary constraint using 29 mammals. Nature 478**:** 476-482.

Lindenbaum, P., and R. Redon, 2017 bioalcidae, samjs and vcffilterjs: object-oriented formatters and filters for bioinformatics files, pp., Bioinformatics.

Linksvayer, T. A., and B. R. Johnson, 2019 Re-thinking the social ladder approach for elucidating the evolution and molecular basis of insect societies. Curr Opin Insect Sci 34**:** 123-129.

Linksvayer, T. A., and M. J. Wade, 2005 The evolutionary origin and elaboration of sociality in the aculeate hymenoptera: maternal effects, sib-social effects, and heterochrony. The Quarterly Review of Biology 80**:** 317-336.

Liu, B., Y. Xu and W. Zhang, 2023 Transcriptome analysis of *Apis mellifera* antennae reveals molecular divergence underlying the division of labour in worker bees. Insect Molecular Biology.

Liu, F., W. Peng, Z. Li, W. Li, L. Li *et al.*, 2012 Next-generation small RNA sequencing for microRNAs profiling in *Apis mellifera*: comparison between nurses and foragers. Insect Molecular Biology 21**:** 297-303.

Liu, F., T. Shi, L. Qi, X. Su, D. Wang *et al.*, 2019 lncRNA profile of *Apis mellifera* and its possible role in behavioural transition from nurses to foragers. BMC Genomics 20**:** 1-11.

Liu, H., Y. Jia, X. Sun, D. Tian, L. D. Hurst *et al.*, 2016 Direct Determination of the Mutation Rate in the Bumblebee Reveals Evidence for Weak Recombination-Associated Mutation and an Approximate Rate Constancy in Insects. Mol Biol Evol 34**:** 119-130.

Liu, H., X. Zhang, J. Huang, J. Q. Chen, D. Tian *et al.*, 2015 Causes and consequences of crossing-over evidenced via a high-resolution recombinational landscape of the honey bee. Genome Biol 16**:** 15.

Lujan, S. A., A. R. Clausen, A. B. Clark, H. K. MacAlpine, D. M. MacAlpine *et al.*, 2014 Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. Genome Res 24**:** 1751-1764.

Lyko, F., S. Foret, R. Kucharski, S. Wolf, C. Falckenhayn *et al.*, 2010 The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol 8**:** 1-12.

Macedo, L. M., F. M. Nunes, F. C. Freitas, C. V. Pires, E. D. Tanaka *et al.*, 2016 MicroRNA signatures characterizing caste-independent ovarian activity in queen and worker honeybees (*Apis mellifera* L.). Insect Mol Biol 25**:** 216-226.

Macias, L. G., E. Barrio and C. Toft, 2020 GWideCodeML: A Python Package for Testing Evolutionary Hypotheses at the Genome-Wide Level. G3 (Bethesda) 10**:** 4369-4372.

Malka, O., E. L. Nino, C. M. Grozinger and A. Hefetz, 2014 Genomic analysis of the interactions between social environment and social communication systems in honey bees (*Apis mellifera*). Insect Biochem Mol Biol 47**:** 36-45.

Mao, W., M. A. Schuler and M. R. Berenbaum, 2015 Task-related differential expression of four cytochrome P450 genes in honeybee appendages. Insect Mol Biol 24**:** 582-588.

Mardulyn, P., and S. Cameron, 1999 The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher-level phylogenetics. Molecular Phylogenetics and Evolution 12**:** 168-176.

Maynard-Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet Res Camb 23**:** 23-35.

McAfee, A., J. S. Pettis, D. R. Tarpy and L. J. Foster, 2019 Feminizer and doublesex knock-outs cause honey bees to switch sexes. PLoS Biol 17**:** 1-6.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351**:** 652-654.

McGuire, D., M. Sankovitz and J. Purcell, 2022 A novel distribution of supergene genotypes is present in the socially polymorphic ant *Formica neoclara*. BMC Ecol Evol 22**:** 47.

McLaren, M. R., A. D. Willis and B. J. Callahan, 2019 Consistent and correctable bias in metagenomic sequencing experiments. Elife 8**:** 1-31.

McVean, G., and A. Auton, 2011 LDhat 2.2: a package for the population genetic analysis of recombination.

Memczak, S., M. Jens, A. Elefsinioti, F. Torti, J. Krueger *et al.*, 2013 Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495**:** 333-338.

Menzel, R., 2012 The honeybee as a model for understanding the basis of cognition. Nat Rev Neurosci 13**:** 758-768.

Meusemann, K., J. Korb, M. Schughart and F. Staubach, 2020 No Evidence for Single-Copy Immune-Gene Specific Signals of Selection in Termites. Frontiers in Ecology and Evolution 8**:** 1-16.

Meznar, E. R., J. Gadau, N. Koeniger and O. Rueppell, 2010 Comparative linkage mapping suggests a high recombination rate in all honeybees. J Hered 101 Suppl 1**:** S118-126.

Michener, C. D., 1974 *The social behavior of the bees*. Harvard University Press, Cambridge, MA.

Michener, C. D., 2007 *The bees of the world*.

Michez, D., S. Patiny and B. N. Danforth, 2009 Phylogeny of the bee family Melittidae (Hymenoptera: Anthophila) based on combined molecular and morphological data. Systematic Entomology 34**:** 574-597.

Mikát, M., C. Franchino and S. M. Rehan, 2017 Sociodemographic variation in foraging behavior and the adaptive significance of worker production in the facultatively social small carpenter bee, *Ceratina calcarata*. Behavioral Ecology and Sociobiology 71.

Milla, L., A. Schmidt-Lebuhn, J. Bovill and F. Encinas-Viso, 2022 Monitoring of honey bee floral resources with pollen DNA metabarcoding as a complementary tool to vegetation surveys. Ecological Solutions and Evidence 3**:** 1-14.

Miller, S. E., A. W. Legan, M. T. Henshaw, K. L. Ostevik, K. Samuk *et al.*, 2020 Evolutionary dynamics of recent selection on cognitive abilities. Proc Natl Acad Sci U S A 117**:** 3045-3052.

Miller, S. E., and M. J. Sheehan, 2023 Sex differences in deleterious genetic variants in a haplodiploid social insect. Mol Ecol 32**:** 4546-4556.

Miyata, K., Y. Inoue, Y. Amano, T. Nishioka, T. Nagaike *et al.*, 2022 Comparative environmental RNA and DNA metabarcoding analysis of river algae and arthropods for ecological surveys and water quality assessment. Scientific Reports 12**:** 1-13.

Mohd-Assaad, N., B. A. McDonald and D. Croll, 2018 Genome-Wide Detection of Genes Under Positive Selection in Worldwide Populations of the Barley Scald Pathogen. Genome Biol Evol 10**:** 1315-1332.

Molodtsova, D., B. A. Harpur, C. F. Kent, K. Seevananthan and A. Zayed, 2014 Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviors. Front Genet 5**:** 431.

Montes, M., M. M. Nielsen, G. Maglieri, A. Jacobsen, J. Hojfeldt *et al.*, 2015 The lncRNA MIR31HG regulates p16(INK4A) expression to modulate senescence. Nature Communications 6**:** 1-15.

Nazareno, A. G., J. B. Bemmels, C. W. Dick and L. G. Lohmann, 2017 Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Molecular Ecology Resources 17**:** 1136-1147.

Neilson, R., 2001 Statistical tests of selective neutrality in the age of genomics. Heredity 86**:** 641-647.

Nelder, J., and R. Wedderburn, 1972 Generalized Linear Models. Journal of the Royal Statisical Society 135**:** 370-384.

Netschitailo, O., S. Raub, O. Kaftanoglu and M. Beye, 2022 Sexual diversification of splicing regulation during embryonic development in honeybees (*Apis mellifera*), A haplodiploid system. Insect Molecular Biology 31**:** 170-176.

Newton, J. P., P. W. Bateman, M. J. Heydenrych, J. H. Kestel, K. W. Dixon *et al.*, 2023 Monitoring the birds and the bees: Environmental DNA metabarcoding of flowers detects plant–animal interactions. Environmental DNA 5**:** 488-502.

Nie, H. Y., L. Q. Liang, Q. F. Li, Z. H. Li, Y. N. Zhu *et al.*, 2021 CRISPR/Cas9 mediated knockout of Amyellow-y gene results in melanization defect of the cuticle in adult *Apis mellifera*. J Insect Physiol 132**:** 1-8.

Noh, J. H., K. M. Kim, W. G. McClusky, K. Abdelmohsen and M. Gorospe, 2018 Cytoplasmic functions of long noncoding RNAs. Wiley Interdisciplinary Reviews 9**:** 1-15.

Nowak, M. A., C. E. Tarnita and E. O. Wilson, 2010 The evolution of eusociality. Nature 466**:** 1057-1062.

Odanaka, K. A., M. G. Branstetter, K. B. Tobin and S. M. Rehan, 2022 Phylogenomics and historical biogeography of the cleptoparasitic bee genus *Nomada* (Hymenoptera: Apidae) using ultraconserved elements. Molecular Phylogenetics and Evolution 170**:** 1-9.

Okonechnikov, K., A. Conesa and F. Garcia-Alcalde, 2016 Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32**:** 292-294.

Oldroyd, B. P., B. Yagound, M. H. Allsopp, M. J. Holmes, G. Buchmann *et al.*, 2021 Adaptive, caste-specific changes to recombination rates in a thelytokous honeybee population. Proc Biol Sci 288**:** 20210729.

Ometto, L., D. Shoemaker, K. G. Ross and L. Keller, 2011 Evolution of gene expression in fire ants: the effects of developmental stage, caste, and species. Mol Biol Evol 28**:** 1381-1392.

Orr, M. C., M. G. Branstetter, J. Straka, F. Yuan, R. Leijs *et al.*, 2022 Phylogenomic interrogation revives an overlooked hypothesis for the early evolution of the bee family Apidae (Hymenoptera: Apoidea), with a focus on the subfamily Anthophorinae. Insect Systematics and Diversity 6**:** 1-15.

Osborne, K. A., A. Robichon, E. Burgess, S. Butland, R. A. Shaw *et al.*, 1997 Natural behavior polymorphism due to a cGMP-dependent protein kinase of *Drosophila*. Science 277**:** 834-836.

Otte, M., O. Netschitailo, S. Weidtkamp-Peters, C. A. M. Seidel and M. Beye, 2023 Recognition of polymorphic *Csd* proteins determines sex in the honeybee. Science Advances 9**:** 1-16.

Page, R. E., Jr., and G. V. Amdam, 2007 The making of a social insect: developmental architectures of social design. Bioessays 29**:** 334-343.

Paluoja, P., M. Vaher, H. Teder, K. Krjutskov, A. Salumets *et al.*, 2025 Honey bulk DNA metagenomic analysis to identify honey biological composition and monitor honey bee pathogens. NPJ Sci Food 9**:** 91.

Panda, A. C., 2018 Circular RNAs Act as miRNA Sponges. Adv Exp Med Biol 1087**:** 67-79.

Panier, S., S. Wang and B. Schumacher, 2024 Genome Instability and DNA Repair in Somatic and Reproductive Aging. Annu Rev Pathol 19**:** 261-290.

Paraskevopoulou, M. D., and A. G. Hatzigeorgiou, 2016 Analyzing MiRNA-LncRNA Interactions. Methods in Molecular Biology 1402**:** 271-286.

Parejo, M., D. Wragg, D. Henriques, J. D. Charriere and A. Estonba, 2020 Digging into the genomic past of Swiss honey bees by whole-genome sequencing museum specimens. Genome Biology and Evolution 12**:** 2535-2551.

Pathiraja, D., J. Cho, J. Kim and I. G. Choi, 2023 Metabarcoding of eDNA for tracking the floral and geographical origins of bee honey. Food Research International 164**:** 1-9.

Patin, N. V., and K. D. Goodwin, 2022 Capturing marine microbiomes and environmental DNA: A field sampling guide. Frontiers in Microbiology 13**:** 1-20.

Pauly, A., G. Noël, G. Sonet, D. G. Notton and J.-L. Boevé, 2019 Integrative taxonomy resuscitates two species in the *Lasioglossum villosulum* complex (Kirby, 1802) (Hymenoptera: Apoidea: Halictidae). European Journal of Taxonomy.

Pe'er, I., R. Yelensky, D. Altshuler and M. J. Daly, 2008 Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 32**:** 381-385.

Pessia, E., A. Popa, S. Mousset, C. Rezvoy, L. Duret *et al.*, 2012 Evidence for widespread GC-biased gene conversion in eukaryotes. Genome Biol Evol 4**:** 675-682.

Peterson, K. J., M. R. Dietrich and M. A. McPeek, 2009 MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. Bioessays 31**:** 736-747.

Pierce, D., P. Sun, J. Purcell and A. Brelsford, 2022 A socially polymorphic *Formica* ant species exhibits a novel distribution of social supergene genotypes. J Evol Biol 35**:** 1031-1044.

Pita-Juarez, Y., G. Altschuler, S. Kariotis, W. Wei, K. Koler *et al.*, 2018 The Pathway Coexpression Network: Revealing pathway relationships. PLoS Comput Biol 14**:** 1-28.

Plath, O. E., 1934 *Bumblebees and their ways* The Macmillan Company, New York.

Pope, N. S., A. Singh, A. K. Childers, K. M. Kapheim, J. D. Evans *et al.*, 2023 The expansion of agriculture has shaped the recent evolutionary history of a specialized squash pollinator. Proceedings of the National Academy of Sciences 120**:** 1-10.

Posadas, D. M., and R. W. Carthew, 2014 MicroRNAs and their roles in developmental canalization. Current opinion in Genetics and Development 27**:** 1-6.

Praz, C., A. Müller and D. Genoud, 2019 Hidden diversity in European bees: *Andrena amieti* sp. n., a new Alpine bee species related to Andrena bicolor (Fabricius, 1775) (Hymenoptera, Apoidea, Andrenidae). Alpine Entomology 3**:** 11-38.

Price, J., M. C. Harrison, R. L. Hammond, S. Adams, J. F. Gutierrez-Marcos *et al.*, 2018 Alternative splicing associated with phenotypic plasticity in the bumble bee *Bombus terrestris*. Molecular Ecology 27**:** 1036-1043.

Privman, E., P. Cohen, A. B. Cohanim, O. Riba-Grognuz, D. Shoemaker *et al.*, 2018 Positive selection on sociobiological traits in invasive fire ants. Mol Ecol 27**:** 3116-3130.

Qanbari, S., T. M. Strom, G. Haberer, S. Weigend, A. A. Gheyas *et al.*, 2012 A high resolution genome-wide scan for significant selective sweeps: an application to pooled sequence data in laying chickens. PLoS One 7**:** e49525.

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26**:** 841-842.

Radford, J. M., D. Chen, A. M. Chernyshova, C. Taylor, A. W. Guoth *et al.*, 2022 Differential Selection on Caste-Associated Genes in a Subterranean Termite. Insects 13**:** 1-12.

Raices, M., R. Bowman, S. Smolikove and J. L. Yanowitz, 2021 Aging Negatively Impacts DNA Repair and Bivalent Formation in the *C. elegans* Germ Line. Front Cell Dev Biol 9**:** 695333.

Ranjitha, N., and T. G. Bhuvana, 2022 Overview of CRISPR-Cas Mediated Genome Engineering. Journal of Applied Life Sciences International 25**:** 37-46.

Rashid, F., A. Shah and G. Shan, 2016 Long Non-coding RNAs in the Cytoplasm. Genomics Proteomics Bioinformatics 14**:** 73-80.

Rehan, S. M., A. J. Berens and A. L. Toth, 2014 At the brink of eusociality: transcriptomic correlates of worker behaviour in a small carpenter bee. BMC Evolutionary Biology 14**:** 1-11.

Rehan, S. M., K. M. Glastad, S. P. Lawson and B. G. Hunt, 2016 The genome and methylome of a subsocial small Carpenter Bee, *Ceratina calcarata*. Genome Biol Evol 8**:** 1401-1410.

Rehan, S. M., K. M. Glastad, M. A. Steffen, C. R. Fay, B. G. Hunt *et al.*, 2018 Conserved genes underlie phenotypic plasticity in an incipiently social bee. Genome Biology and Evolution 10**:** 2749-2758.

Rehan, S. M., and M. H. Richards, 2010a The Influence of Maternal Quality on Brood Sex Allocation in the Small Carpenter Bee, *Ceratina calcarata*. Ethology**:** no-no.

Rehan, S. M., and M. H. Richards, 2010b Nesting biology and subsociality in *Ceratina calcarata* (Hymenoptera: Apidae). The Canadian Entomologist 142**:** 65-74.

Rehan, S. M., and A. L. Toth, 2015 Climbing the social ladder: the molecular evolution of sociality. Trends Ecol Evol 30**:** 426-433.

Ribani, A., V. Taurisano, V. J. Utzeri and L. Fontanesi, 2022 Honey environmental DNA can be used to detect and monitor honey bee pests: development of methods useful to identify *Aethina tumida* and *Galleria mellonella* infestations. Veterinary Sciences 9**:** 1-11.

Ribani, A., V. J. Utzeri, V. Taurisano and L. Fontanesi, 2020 Honey as a source of environmental DNA for the detection and monitoring of honey bee pathogens and parasites. Veterinary Sciences 7**:** 1-14.

Ribeiro, T. M., and A. Espindola, 2024 Integrated phylogenomic approaches in insect systematics. Current Opinion in Insect Science 61**:** 1-6.

Richard, G., J. Jaquiery and G. Le Trionnaire, 2021 Contribution of Epigenetic Mechanisms in the Regulation of Environmentally-Induced Polyphenism in Insects. Insects 12.

Richards, M. H., 2019 Socially polymorphic bees as model organisms for studying the evolution of eusociality. Insectes Sociaux 66**:** 3-4.

Rittschof, C. C., and G. E. Robinson, 2013 Manipulation of colony environment modulates honey bee aggression and brain gene expression. Genes Brain Behav 12**:** 802-811.

Robert E. Page Jr, C. Y. S. P., 2001 Aging and development in social insects with emphasis on the honey bee, *Apis mellifera L*. Experimental Gerontology 36**:** 695-711.

Rodrigues, M. A., and T. Flatt, 2016 Endocrine uncoupling of the trade-off between reproduction and somatic maintenance in eusocial insects. Current Opinion in Insect Science 16**:** 1-8.

Roger, F., H. R. Ghanavi, N. Danielsson, N. Wahlberg, J. Löndahl *et al.*, 2022 Airborne environmental DNA metabarcoding for the monitoring of terrestrial insects—A proof of concept from the field. Environmental DNA 4**:** 790-807.

Rohde, A. T., M. G. Branstetter, K. E. Mock, J. N. Knoblett, D. S. Pilliod *et al.*, 2024 Population genetics of museum specimens indicate decreasing genetic resiliency: The case of two bumble bees of conservation concern. Biological Conservation 291**:** 1-13.

Romiguier, J., M. L. Borowiec, A. Weyna, Q. Helleu, E. Loire *et al.*, 2022 Ant phylogenomics reveals a natural selection hotspot preceding the origin of complex eusociality. Current Biology 32**:** 2942-2947 e2944.

Romiguier, J., S. A. Cameron, S. H. Woodard, B. J. Fischman, L. Keller *et al.*, 2016 Phylogenomics controlling for base compositional bias reveals a single origin of eusociality in corbiculate bees. Molecular Biology and Evolution 33**:** 670-678.

Romiguier, J., J. Lourenco, P. Gayral, N. Faivre, L. A. Weinert *et al.*, 2014 Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. J Evol Biol 27**:** 593-603.

Ross, L., H. Blackmon, P. Lorite, V. E. Gokhman and N. B. Hardy, 2015 Recombination, chromosome number and eusociality in the Hymenoptera. J Evol Biol 28**:** 105-116.

Roth, A., C. Vleurinck, O. Netschitailo, V. Bauer, M. Otte *et al.*, 2019 A genetic switch for worker nutrition-mediated traits in honeybees. PLoS Biol 17**:** 1-18.

Roux, J., E. Privman, S. Moretti, J. T. Daub, M. Robinson-Rechavi *et al.*, 2014 Patterns of positive selection in seven ant genomes. Mol Biol Evol 31**:** 1661-1685.

Rubin, B. E. R., B. M. Jones, B. G. Hunt and S. D. Kocher, 2019 Rate variation in the evolution of non-coding DNA associated with social evolution in bees. Philosophical Transactions of the Royal Society B 374**:** 1-13.

Rubinoff, D., S. Cameron and K. Will, 2006 A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. J Hered 97**:** 581-594.

Rueppell, O., R. Kuster, K. Miller, B. Fouks, S. Rubio Correa *et al.*, 2016 A New Metazoan Recombination Rate Record and Consistently High Recombination Rates in the Honey Bee Genus Apis Accompanied by Frequent Inversions but Not Translocations. Genome Biol Evol 8**:** 3653-3660.

Ruppert, K. M., R. J. Kline and M. S. Rahman, 2019 Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecology and Conservation 17**:** 1-29.

Salzman, J., R. E. Chen, M. N. Olsen, P. L. Wang and P. O. Brown, 2013 Cell-type specific features of circular RNA expression. PLoS Genetics 9**:** 1-15.

Samson, M. L., 2008 Rapid functional diversification in the structurally conserved ELAV family of neuronal RNA binding proteins. BMC Genomics 9**:** 1-11.

San Martin, G., L. Hautier, D. Mingeot and B. Dubois, 2024 How reliable is metabarcoding for pollen identification? An evaluation of different taxonomic assignment strategies by cross-validation. PeerJ 12**:** 1-26.

Sawata, M., D. Yoshino, H. Takeuchi, A. Kamikouchi, K. Ohashi *et al.*, 2002 Identification and punctate nuclear localization of a novel noncoding RNA, *Ks-1*, from the honeybee brain. RNA 8**:** 772-785.

Schlüns, H., R. F. A. Moritz, P. Neumann, P. Kryger and G. Koeniger, 2005 Multiple nuptial flights, sperm transfer and the evolution of extreme polyandry in honeybee queens. Animal Behaviour 70**:** 125-131.

Schmitt, C. J., J. A. Cook, K. R. Zamudio and S. V. Edwards, 2018 Museum specimens of terrestrial vertebrates are sensitive indicators of environmental change in the Anthropocene. Philosophical Transactions of the Royal Society B 374**:** 1-10.

Schrader, L., D. F. Simola, J. Heinze and J. Oettler, 2015 Sphingolipids, Transcription Factors, and Conserved Toolkit Genes: Developmental Plasticity in the Ant Cardiocondyla obscurior. Mol Biol Evol 32**:** 1474-1486.

Seabold, S., and J. Perktold, 2010 Statsmodels: Econometric and Statistical Modeling with Python. PROC. OF THE 9th PYTHON IN SCIENCE CONF.

Sedlazeck, F. J., P. Rescheneder and A. von Haeseler, 2013 NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics 29**:** 2790-2791.

Sethurman, N., and D. A. O'Brochta, 2005 The *Drosophila melanogaster cinnabar* Gene Is a Cell Autonomous Genetic Marker in *Aedes aegypti* (Diptera: Culicidae). Journal of Medical Entomology 42**:** 716-718.

Shaffer, M. R., E. Andruszkiewicz Allan, A. M. Van Cise, K. M. Parsons, A. O. Shelton *et al.*, 2025 Observation Bias in Metabarcoding. Mol Ecol Resour**:** 1-19.

She, R., and D. F. Jarosz, 2018 Mapping Causal Variants with Single-Nucleotide Resolution Reveals Biochemical Drivers of Phenotypic Change. Cell 172**:** 478-490 e415.

Shell, W. A., and S. M. Rehan, 2015 Recent and rapid diversification of the small carpenter bees in eastern North America. Biological Journal of the Linnean Society**:** 1-13.

Shell, W. A., and S. M. Rehan, 2017 Behavioral and genetic mechanisms of social evolution: insights from incipiently and facultatively social bees. Apidologie 49**:** 13-30.

Shell, W. A., and S. M. Rehan, 2019 Social modularity: conserved genes and regulatory elements underlie caste-antecedent behavioural states in an incipiently social bee. Proc Biol Sci 286**:** 20191815.

Shell, W. A., and S. M. Rehan, 2022 Social divergence: molecular pathways underlying castes and longevity in a facultatively eusocial small carpenter bee. Proc Biol Sci 289**:** 20212663.

Shell, W. A., S. M. Rehan and L. Holman, 2018 The price of insurance: costs and benefits of worker production in a facultatively social bee. Behavioral Ecology 29**:** 204-211.

Shell, W. A., M. A. Steffen, H. K. Pare, A. S. Seetharam, A. J. Severin *et al.*, 2021 Sociality sculpts similar patterns of molecular evolution in two independently evolved lineages of eusocial bees. Communications Biology 4**:** 253.

Sherman, B. T., M. Hao, J. Qiu, X. Jiao, M. W. Baseler *et al.*, 2022 DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res.

Shi, T., Y. Zhu, P. Liu, L. Ye, X. Jiang *et al.*, 2021 Age- and behavior-dependent differential miRNA expression in the hypopharyngeal glands of honeybees (*Apis mellifera* L.). Insects 12**:** 1-15.

Shi, Y.-Y., H.-J. Zheng, Q.-Z. Pan, Z.-L. Wang and Z.-J. Zeng, 2014 Differentially expressed microRNAs between queen and worker larvae of the honey bee (*Apis mellifera*). Apidologie 46**:** 35-45.

Shi, Y. Y., L. X. Sun, Z. Y. Huang, X. B. Wu, Y. Q. Zhu *et al.*, 2013 A SNP based high-density linkage map of Apis cerana reveals a high recombination rate similar to Apis mellifera. PLoS One 8**:** e76459.

Shorter, J., C. Couch, W. Huang, M. A. Carbone, J. Peiffer *et al.*, 2015 Genetic architecture of natural variation in *Drosophila melanogaster* aggressive behavior. Proc Natl Acad Sci U S A 112**:** E3555-3563.

Shpak, M., H. R. Ghanavi, J. D. Lange, J. E. Pool and M. C. Stensmyr, 2023 Genomes from historical *Drosophila melanogaster* specimens illuminate adaptive and demographic changes across more than 200 years of evolution. PLoS Biology 21**:** 1-31.

Shpigler, H. Y., M. C. Saul, E. E. Murdoch, A. C. Cash-Ahmed, C. H. Seward *et al.*, 2017 Behavioral, transcriptomic and epigenetic responses to social challenge in honey bees. Genes Brain Behav 16**:** 579-591.

Sieber, K. R., T. Dorman, N. Newell and H. Yan, 2021 (Epi)Genetic Mechanisms Underlying the Evolutionary Success of Eusocial Insects. Insects 12.

Silverman, J. D., R. J. Bloom, S. Jiang, H. K. Durand, E. Dallow *et al.*, 2021 Measuring and mitigating PCR bias in microbiota datasets. PLoS Comput Biol 17**:** 1-17.

Simanonok, M. P., C. R. V. Otto, R. S. Cornman, D. D. Iwanowicz, J. P. Strange *et al.*, 2021 A century of pollen foraging by the endangered rusty patched bumble bee (*Bombus affinis*): inferences from molecular sequencing of museum specimens. Biodiversity and Conservation 30**:** 123-137.

Simola, D. F., R. J. Graham, C. M. Brady, B. L. Enzmann, C. Desplan *et al.*, 2016 Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. Science 351**:** aac6633.

Simola, D. F., L. Wissler, G. Donahue, R. M. Waterhouse, M. Helmkampf *et al.*, 2013 Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. Genome Res 23**:** 1235-1247.

Sinakevitch, I., Z. Kurtzman, H. G. Choi, D. A. Ruiz Pardo, R. A. Dahan *et al.*, 2020 Anti-RDL and Anti-mGlutR1 Receptors Antibody Testing in Honeybee Brain Sections using CRISPR-Cas9. J Vis Exp**:** e59993.

Singh, A. S., M. C. Takhellambam, P. Cappelletti and M. Feligioni, 2020 Immediate early gene *kakusei* potentially plays a role in the daily foraging of honey bees. PLoS One 15**:** 1-11.

Singh, N. D., A. M. Larracuente and A. G. Clark, 2008 Contrasting the efficacy of selection on the X and autosomes in *Drosophila*. Mol Biol Evol 25**:** 454-467.

Sirvio, A., J. Gadau, O. Rueppell, D. Lamatsch, J. J. Boomsma *et al.*, 2006 High recombination frequency creates genotypic diversity in colonies of the leaf-cutting ant *Acromyrmex echinatior*. J Evol Biol 19**:** 1475-1485.

Sirvio, A., P. Pamilo, R. A. Johnson, R. E. Page, Jr. and J. Gadau, 2011 Origin and evolution of the dependent lineages in the genetic caste determination system of *Pogonomyrmex* ants. Evolution 65**:** 869-884.

Sirviö, A. J., J Spencer; Wenseleers,Tom; Pamilo,Pekka, 2011 A high recombination rate in eusocial Hymenoptera: evidence from the common wasp *Vespula vulgaris*. BMC Genetics 12**:** 1-7.

Sless, T. J. L., M. G. Branstetter, J. P. Gillung, E. A. Krichilsky, K. B. Tobin *et al.*, 2022 Phylogenetic relationships and the evolution of host preferences in the largest clade of brood parasitic bees (Apidae: Nomadinae). Molecular Phylogenetics and Evolution 166**:** 1-13.

Snetkova, V., L. A. Pennacchio, A. Visel and D. E. Dickel, 2022 Perfect and imperfect views of ultraconserved sequences. Nature Reviews Genetics 23**:** 182-194.

Solc, P., R. M. Schultz and J. Motlik, 2010 Prophase I arrest and progression to metaphase I in mouse oocytes: comparison of resumption of meiosis and recovery from G2-arrest in somatic cells. Mol Hum Reprod 16**:** 654-664.

Sovik, E., G. Bloch and Y. Ben-Shahar, 2015 Function and evolution of microRNAs in eusocial Hymenoptera. Frontiers in Genetics 6**:** 1-11.

Spannhoff, A., Y. K. Kim, N. J. Raynal, V. Gharibyan, M. B. Su *et al.*, 2011 Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. EMBO Reports 12**:** 238-243.

Spisak, N., M. de Manuel, W. Milligan, G. Sella and M. Przeworski, 2024 The clock-like accumulation of germline and somatic mutations can arise from the interplay of DNA damage and repair. PLoS Biol 22**:** e3002678.

Spotter, A., P. Gupta, M. Mayer, N. Reinsch and K. Bienefeld, 2016 Genome-Wide Association Study of a Varroa-Specific Defense Behavior in Honeybees (*Apis mellifera*). J Hered 107**:** 220-227.

Standage, D. S., A. J. Berens, K. M. Glastad, A. J. Severin, V. P. Brendel *et al.*, 2016 Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. Molecular Ecology 25**:** 1769-1784.

Statello, L., C. J. Guo, L. L. Chen and M. Huarte, 2021 Gene regulation by long non-coding RNAs and its biological functions. Nature Reviews Molecular Cell Biology 22**:** 96-118.

Steitz, I., R. J. Paxton, S. Schulz and M. Ayasse, 2021 Chemical Variation among Castes, Female Life Stages and Populations of the Facultative Eusocial Sweat Bee Halictus rubicundus (Hymenoptera: Halictidae). J Chem Ecol 47**:** 406-419.

Storz, J. F., B. A. Payseur and M. W. Nachman, 2004 Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. Mol Biol Evol 21**:** 1800-1811.

Strauss, K., H. Scharpenberg, R. M. Crewe, F. Glahn, H. Foth *et al.*, 2008 The role of the queen mandibular gland pheromone in honeybees (*Apis mellifera*): honest signal or suppressive agent? Behavioral Ecology and Sociobiology 62**:** 1523-1531.

Strunov, A., S. Kirchner, J. Schindelar, L. Kruckenhauser, E. Haring *et al.*, 2023 Historic museum samples provide evidence for a recent replacement of *Wolbachia* types in European *Drosophila melanogaster*. Molecular Biology and Evolution 40**:** 1-17.

Subramanian, S., 2013 Significance of population size on the fixation of nonsynonymous mutations in genes under varying levels of selection pressure. Genetics 193**:** 995-1002.

Supek, F., and B. Lehner, 2015 Differential DNA mismatch repair underlies mutation rate variation across the human genome. Nature 521**:** 81-84.

Szathmary, E., and J. Maynard-Smith, 1995 The major evolutionary transitions. Nature 374.

Tadano, H., Y. Yamazaki, H. Takeuchi and T. Kubo, 2009 Age- and division-of-labour-dependent differential expression of a novel non-coding RNA, Nb-1, in the brain of worker honeybees, *Apis mellifera* L. Insect Molecular Biology 18**:** 715-726.

Tam, V., N. Patel, M. Turcotte, Y. Bosse, G. Pare *et al.*, 2019 Benefits and limitations of genome-wide association studies. Nat Rev Genet 20**:** 467-484.

Tamtaji, O. R., M. Derakhshan, F. Z. Rashidi Noshabad, J. Razaviyan, R. Hadavi *et al.*, 2021 Non-coding RNAs and brain tumors: insights into their roles in apoptosis. Frontiers in Cell and Developmental Biology 9**:** 1-18.

Tang, W., M. Li, X. Yangzhong, X. Zhang, A. Zu *et al.*, 2022 Hippo signaling pathway and respiratory diseases. Cell Death Discov 8**:** 1-10.

Terpilowski, M., 2019 scikit-posthocs: Pairwise multiple comparison tests in Python. Journal of Open Source Software 4.

The Honeybee Genome Sequencing Consortium, 2006 Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443**:** 931-949.

The Nobel Prize in Physiology or Medicine, 1973, pp. Nobel Prize Outreach AB 2024, NobelPrize.org.

Tholken, C., M. Thamm, C. Erbacher and M. Lechner, 2019 Sequence and structural properties of circular RNAs in the brain of nurse and forager honeybees (*Apis mellifera*). BMC Genomics 20**:** 1-15.

Torgerson, D. G., A. R. Boyko, R. D. Hernandez, A. Indap, X. Hu *et al.*, 2009 Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. PLoS Genet 5**:** e1000592.

Toth, A. L., and G. E. Robinson, 2007 Evo-devo and the evolution of social behavior. Trends Genet 23**:** 334-341.

Toth, A. L., K. Varala, T. C. Newman, F. E. Miguez, S. K. Hutchison *et al.*, 2007 Wasp Gene Expression Supports an Evolutionary Link Between Maternal Behavior and Eusociality. Science 318**:** 441-444.

Toth, A. L., and A. Zayed, 2021 The honey bee genome-- what has it been good for? Apidologie 52**:** 45-62.

Toussaint, E. F. A., J. Gauthier, J. Bilat, C. Gillett, H. M. Gough *et al.*, 2021 HyRAD-X exome capture museomics unravels giant ground beetle evolution. Genome Biology and Evolution 13**:** 1-18.

Traniello, I. M., S. A. Bukhari, P. Dibaeinia, G. Serrano, A. Avalos *et al.*, 2023 Single-cell dissection of aggression in honeybee colonies. Nature Ecology and Evolution 7**:** 1232-1244.

Trible, W., L. Olivos-Cisneros, S. K. McKenzie, J. Saragosti, N. C. Chang *et al.*, 2017 *orco* Mutagenesis Causes Loss of Antennal Lobe Glomeruli and Impaired Social Behavior in Ants. Cell 170**:** 727-735.

Trivers, R. L., and H. Hare, 1976 Haplodiploidy and the evolution of the social insects. Science 191**:** 249-263.

True, J. R., and S. B. Carroll, 2002 Gene co-option in physiological and morphological evolution. Annu Rev Cell Dev Biol 18**:** 53-80.

Trunz, V., L. Packer, J. Vieu, N. Arrigo and C. J. Praz, 2016 Comprehensive phylogeny, biogeography and new classification of the diverse bee tribe Megachilini: can we use DNA barcodes in phylogenies of large genera? Molecular phylogenetics and evolution 103**:** 245-259.

Ustaoglu, P., J. K. Gill, N. Doubovetzky, I. U. Haussmann, T. C. Dix *et al.*, 2021 Dynamically expressed single ELAV/Hu orthologue elavl2 of bees is required for learning and memory. Communications Biology 4**:** 1-12.

Valenzuela, L., and R. T. Kamakaka, 2006 Chromatin insulators. Annu Rev Genet 40**:** 107-138.

Van Dam, M. H., J. B. Henderson, L. Esposito and M. Trautwein, 2021 Genomic Characterization and Curation of UCEs Improves Species Tree Reconstruction. Systematic Biology 70**:** 307-321.

Van der Auwera, G., and B. O'Connor, 2020 *Genomics in the cloud: Using Docker, GATK, and WDL in Terra* O'Reilly Media.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel *et al.*, 2014 From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43**:** 11 10 11-33.

van der Loos, L. M., and R. Nijland, 2021 Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. Mol Ecol 30**:** 3270-3288.

Van Oystaeyen, A., R. C. Oliveira, L. Holman, J. S. Van Zweden, C. Romero *et al.*, 2014 Conserved class of queen pheromones stops social insect workers from reproducing. Science 343.

Vieira, J., F. C. de Paula Freitas, A. Santos Cristino, D. Guariz Pinheiro, L. R. Aguiar *et al.*, 2021 Molecular underpinnings of the early brain developmental response to differential feeding in the honey bee *Apis mellifera*. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1864**:** 1-11.

Voight, B. F., S. Kudaravalli, X. Wen and J. K. Pritchard, 2006 A map of recent positive selection in the human genome. PLoS Biol 4**:** e72.

Wagner, A., J. Seiler and M. Beye, 2022 Highly efficient site-specific integration of DNA fragments into the honeybee genome using CRISPR/Cas9. G3 (Bethesda) 12**:** 1-6.

Waiker, P., F. C. P. de Abreu, D. Luna-Lucena, F. C. P. Freitas, Z. L. P. Simoes *et al.*, 2021 Recombination mapping of the Brazilian stingless bee Frieseomelitta varia confirms high recombination rates in social hymenoptera. BMC Genomics 22**:** 673.

Wallberg, A., I. Bunikis, O. V. Pettersson, M. B. Mosbech, A. K. Childers *et al.*, 2019 A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. BMC Genomics 20**:** 275.

Wallberg, A., S. Glemin and M. T. Webster, 2015 Extreme recombination frequencies shape genome variation and evolution in the honeybee, *Apis mellifera*. PLoS Genet 11**:** e1005189.

Wallberg, A., F. Han, G. Wellhagen, B. Dahle, M. Kawata *et al.*, 2014 A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. Nat Genet 46**:** 1081-1088.

Walsh, A. T., D. A. Triant, J. J. Le Tourneau, M. Shamimuzzaman and C. G. Elsik, 2022 Hymenoptera Genome Database: new genomes and annotation datasets for improved go enrichment and orthologue analyses. Nucleic Acids Res 50**:** D1032-D1039.

Wang, X., Y. Lin, L. Liang, H. Geng, M. Zhang *et al.*, 2021 Transcriptional Profiles of Diploid Mutant *Apis mellifera* Embryos after Knockout of *csd* by CRISPR/Cas9. Insects 12**:** 1-21.

Wang, Z., S. Wang, X. Fan, K. Zhang, J. Zhang *et al.*, 2023 Systematic Characterization and Regulatory Role of lncRNAs in Asian Honey Bees Responding to Microsporidian Infestation. Internaltional Journal of Molecular Sciences 24**:** 1-18.

Waters, C., C. Hurt and S. Krosnick, 2025 Looking to the past to inform the future: What eDNA from herbarium specimens can tell us about plant-animal interactions. Appl Plant Sci 13**:** 1-14.

Weaver, D. B., J. M. Anzola, J. D. Evans, J. G. Reid, J. T. Reese *et al.*, 2007 Computational and transcriptional evidence for microRNAs in the honey bee genome. Genome Biology 8**:** R97.

West-Eberhard, M., and S. Turillazzi, 1996 Natural history and evolution of paper wasps.

West-Eberhard, M. J., 1969 *The social biology of the polistine wasps*.

West-Eberhard, M. J., 1987 Flexible strategy and social evolution. Anim. Soc. Theor. Fac**:** 35-51.

Wickham, H., 2016 ggplot2: Elegant Grpahics for Data Analysis, pp., edited by Springer-Verlag, New York.
Wiernasz, D. C., and B. J. Cole, 2022 The ontogeny of selection on genetic diversity in harvester ants. Proc Biol Sci 289**:** 1-7.
Wilfert, L., J. Gadau and P. Schmid-Hempel, 2007 Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity (Edinb) 98**:** 189-197.
Willing, E. M., C. Dreyer and C. van Oosterhout, 2012 Estimates of genetic differentiation measured by F(ST) do not necessarily require large sample sizes when using many SNP markers. PLoS One 7**:** e42649.
Wilson, E. O., 1971 *The insect societies*. Harvard University Press, Cambridge, MA.
Wilson, E. O., and B. Holldobler, 2005 Eusociality: origin and consequences. Proc Natl Acad Sci U S A 102**:** 13367-13371.
Withee, J. R., and S. M. Rehan, 2017 Social Aggression, Experience, and Brain Gene Expression in a Subsocial Bee. Integr Comp Biol 57**:** 640-648.
Witt, E., C. B. Langer, N. Svetec and L. Zhao, 2023 Transcriptional and mutational signatures of the *Drosophila* ageing germline. Nat Ecol Evol 7**:** 440-449.
Wittwer, B., A. Hefetz, T. Simon, L. E. K. Murphy, M. A. Elgar *et al.*, 2017 Solitary bees reduce investment in communication compared with their social relatives. Proceedings of the National Academy of Sciences 114**:** 6569-6574.
Wizenberg, S. B., S. K. French, L. R. Newburn, M. Pepinelli, I. M. Conflitti *et al.*, 2024 Pollen foraging mediates exposure to dichotomous stressor syndromes in honey bees. PNAS Nexus 3**:** 1-13.
Wizenberg, S. B., L. R. Newburn, M. Pepinelli, I. M. Conflitti, R. T. Richardson *et al.*, 2023a Validating a multi-locus metabarcoding approach for characterizing mixed-pollen samples. Plant Methods 19**:** 120.
Wizenberg, S. B., L. R. Newburn, R. T. Richardson, M. Pepinelli, I. M. Conflitti *et al.*, 2023b Environmental metagenetics unveil novel plant-pollinator interactions. Ecology and Evolution 13**:** 1-8.
Wizenberg, S. B., M. Pepinelli, B. N. Do, M. Moubony, D. Tamashekan *et al.*, 2025 Inferring Plant Community Phenology via Bee-Collected Pollen. Environmental DNA 7**:** 1-14.
Wojciechowski, M., R. Lowe, J. Maleszka, D. Conn, R. Maleszka *et al.*, 2018 Phenotypically distinct female castes in honey bees are defined by alternative chromatin states during larval development. Genome Res 28**:** 1532-1542.
Wong, W. S., B. D. Solomon, D. L. Bodian, P. Kothiyal, G. Eley *et al.*, 2016 New observations on maternal age effect on germline de novo mutations. Nat Commun 7**:** 10486.
Woodard, S. H., G. M. Bloch, M. R. Band and G. E. Robinson, 2014 Molecular heterochrony and the evolution of sociality in bumblebees (*Bombus terrestris*). Proc Biol Sci 281**:** 20132419.
Woodard, S. H., B. J. Fischman, A. Venkat, M. E. Hudson, K. Varala *et al.*, 2011 Genes involved in convergent evolution of eusociality in bees. Proc Natl Acad Sci U S A 108**:** 7472-7477.
Wright, A. F., 1992 Codon usage in the G+C-rich *Streptomyces* genome. Gene 113**:** 55-65.
Wright, C. J., C. W. J. Smith and C. D. Jiggins, 2022 Alternative splicing as a source of phenotypic diversity. Nature Reviews Genetics 23**:** 697-710.
Xu, H., X. Ye, Y. Yang, Y. Yang, Y. H. Sun *et al.*, 2021 Comparative Genomics Sheds Light on the Convergent Evolution of Miniaturized Wasps. Mol Biol Evol 38**:** 5539-5554.
Yagound, B., K. A. Dogantzis, A. Zayed, J. Lim, P. Broekhuyse *et al.*, 2020 A Single Gene Causes Thelytokous Parthenogenesis, the Defining Feature of the Cape Honeybee *Apis mellifera capensis*. Curr Biol 30**:** 2248-2259 e2246.
Yan, D., J. Liu, Y. Fan, Z. Lian, Z. Dang *et al.*, 2023 Genomic insights into genetic diversity and local adaptation of a dominant desert steppe feather grass, *Stipa breviflora* Griseb. Front Plant Sci 14**:** 1170075.

Yan, H., R. Bonasio, D. F. Simola, J. Liebig, S. L. Berger *et al.*, 2015 DNA methylation in social insects: how epigenetics can control behavior and longevity. Annu Rev Entomol 60**:** 435-452.

Yan, H., C. Opachaloemphan, G. Mancini, H. Yang, M. Gallitto *et al.*, 2017 An Engineered *orco* Mutation Produces Aberrant Social Behavior and Defective Neural Development in Ants. Cell 170**:** 736-747.

Yan, Z., S. H. Martin, D. Gotzek, S. V. Arsenault, P. Duchen *et al.*, 2020 Evolution of a supergene that regulates a trans-species social polymorphism. Nat Ecol Evol 4**:** 240-249.

Yancan, L., C. Tianle, F. Yunhan, L. Delong and W. Guizhi, 2019 Population genomics and morphological features underlying the adaptive evolution of the eastern honey bee (*Apis cerana*). BMC Genomics 20**:** 869.

Yao, Y., K. Sun, Q. Yang, Z. Zhou, C. Shao *et al.*, 2021 Assessing Autosomal InDel Loci With Multiple Insertions or Deletions of Random DNA Sequences in Human Genome. Front Genet 12**:** 809815.

Ye, X., and R. D. Sloboda, 1997 Molecular characterization of p62, a mitotic apparatus protein required for mitotic progression. J Biol Chem 272**:** 3606-3614.

Ye, Y., X. Fan, Z. Cai, Y. Wu, W. Zhang *et al.*, 2022 Unveiling the circRNA-mediated immune responses of western honey bee larvae to *Ascosphaera apis* invasion. International Journal of Molecular Sciences 24**:** 1-17.

Yengo, L., S. Vedantam, E. Marouli, J. Sidorenko, E. Bartell *et al.*, 2022 A saturated map of common genetic variants associated with human height. Nature 610**:** 704-712.

Zayed, A., and C. W. Whitfield, 2008 A genome-wide signature of positive selection in ancient and recent invasive expansions of the honey bee Apis mellifera. Proc Natl Acad Sci U S A 105**:** 3421-3426.

Zeileis, A., C. Kleiber and S. Jackman, 2008 Regression models for count data in R. Journal of Statistical Software 27**:** 1-25.

Zhang, D., Z. Q. Niu, A. R. Luo, M. C. Orr, R. R. Ferrari *et al.*, 2022 Testing the systematic status of *Homalictus* and *Rostrohalictus* with weakened cross-vein groups within Halictini (Hymenoptera: Halictidae) using low-coverage whole-genome sequencing. Insect Science 29**:** 1819-1833.

Zhang, Y., X. Fan, H. Zang, X. Liu, P. Feng *et al.*, 2023 Novel insights into the circRNA-modulated developmental mechanism of western honey bee larval guts. Insects 14**:** 1-17.

Zhang, Y. M., J. L. Williams and A. Luck, 2019 Understanding UCEs: a comprehensive primer on using ultraconserved elements for arthropod phylogenomics. Insect Systematics and Diversity 3**:** 1-12.

Zheng, S. Y., L. X. Pan, F. P. Cheng, M. J. Jin and Z. L. Wang, 2023 A global survey of the full-length transcriptome of *Apis mellifera* by single-molecule long-read sequencing. International Journal of Molecular Sciences 24**:** 1-14.

Zhou, W. Y., Z. R. Cai, J. Liu, D. S. Wang, H. Q. Ju *et al.*, 2020 Circular RNA: metabolism, functions and interactions with proteins. Molecular Cancer 19**:** 1-19.

Zhu, Z., J. Wang, X. Fan, Q. Long, H. Chen *et al.*, 2022 CircRNA-regulated immune responses of asian honey bee workers to microsporidian infection. Frontiers in Genetics 13**:** 1-16.

Appendix A - Author contributions

Chapter 1

Author contribution

DBS wrote the introductory chapter with editing guidance from AZ.

Chapter 2

Author contribution

AZ and DBS conceived the outline and scope of the review. DBS performed all literature search and writing. AZ supervised the project and provided expert guidance on the synthesis of literature.

Chapter 3

Author contribution

AZ and DBS conceived the outline and scope of the review. DBS performed all literature search and writing. AZ supervised the project and provided expert guidance on the synthesis of literature.

Chapter 4

Author contribution

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 Sources

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

Data availability

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

Chapter 5

Author contribution

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 Sources

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

Data availability

All original FASTA files can be found uploaded to NCBI upon publication (PRJNA1028991 for *C. calcarata* raw files and PRJNA1141487 for *Ceratina strenua* files). All code for processing raw data can be found on Figshare DOI:10.6084/m9.figshare.26072632.

Chapter 6

Author contribution

Research was designed by AZ and DBS. Whole bee samples *A. mellifera* were provided by by our collaborators Rob Currie (University of Manitoba), Steve Pernal (Beaverlodge Research Farm, Alberta) and Shelley Hoover (Alberta Agriculture and Forestry Lethbridge Research Centre). All sample preparation was performed by DBS with the help of IC,MP, and KD. All analysis was performed by DBS. Manuscript was prepared by DBS. Guidance throughout was provided by AZ.

Funding Sources

York University Research Chair; Natural Sciences and Engineering Research Council of Canada Discovery grant awarded to AZ.

Data availability

All original FASTA files can be found uploaded to NCBI upon publication (PRJNA1399504). All code can be found on figshare (https://doi.org/10.6084/m9.figshare.29891894.v1).

Sincerely,

Dova Brenman

Approved by,

Amro Zayed, PhD

Professor of Biology