Negative Selection in Social Insects

Mohammad Arshad Imrit

A Thesis submitted to the Faculty of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Science

Graduate Program in Biology

York University

Toronto, Ontario

September 2019

Abstract

Eusociality, characterized in part by cooperative brood care, and reproductive division of labor, evolved independently several times in insects. The evolution of eusociality has been hypothesized to lead to differences in the extent of both positive and negative selection. While population genomics studies of eusocial insects have so far focused on positive selection, there has been no study of the extent of negative selection in social insects, and its relationship to the evolution of caste-biased genes. To address this knowledge gap, our research estimated the extent of negative selection in honey bees, bumble bees, and paper wasps, through analysis of published population genomic datasets. Our results showed that there was a significant negative correlation between increasing social complexity and negative selection, suggesting effective population size plays a role in strength of negative selection. We identified significantly stronger negative selection in queen traits relative to worker traits in honey bees but not in bumble bees and paper wasps. Lastly, we observe stronger negative selection in drone traits relative to queen traits in honeybees, and we attribute this effect to the haplodiploidy system of honey bees.

Acknowledgements

I would like to thank Dr. Amro Zayed for the opportunity in the lab. Knowing absolutely nothing about me, Dr. Amro was willing to take me on in his lab and offered me a cool project. His constant support along the way has been incredible and I cannot thank him enough for this. He also taught me about 'quick maths!' so this was nice. Dr. Sapna Sharma has been an important source of support as well. She introduced me to Dr. Amro as she rightfully believed I would enjoy and grow in the honey bee lab at York. Dr. Sapna also knew me for merely a few years but the help she has provided during these times were remarkable. I would be nowhere near her without her. Katie Dogantzis has helped me understand my project more than anyone and she deserves a huge round of applause for this. Taking the time to help me understand the intricacies of genomic data, caste-biased genes and the theory behind eusociality, Katie made me feel at ease with the project. I would like to thank Dr. Brock Harpur as well for help with RNASeq data and saving me precious time by helping me with analyses and allowing me to collaborate with him. Dr. Clement Kent showed me the ropes with bioinformatics and understanding larger server networks and computing, big thank you to him for this. Harshil Patel and Stephen Rose helped introduce me to python and programming. Their help cannot be overstated as without them, it would have taken me months to perform some analyses. Tanushree Tiwari was a constant source of support and reassured me that I can finish the project, even when things were looking gloom. She should be a motivational speaker in fact. Laiba Khan has supported me during the first stages of the project and would always listen to me ramble on and think out loud. Cait Carew's support has been marvelous during the thesis writing part and she would frequently push me to do better in everything. Lastly but not last, the banter between Sarah MacKell and I on honey bees and bumble bees made my experience in a bee lab so much more fun. Cannot wait for more arguments with her!

Table of Contents

List of Figures

Introduction

The honey bee - a model organism

Honey bees (*Apis mellifera*) are advanced eusocial insects that live in complex societies and show cooperative and altruistic social behaviors, such as brood care and division of labour (Keller and Chapuisat, 1999). Honey bees are conceivably the most widely known social insect with the genome, physiology, and behavior of the honey bee studied intensively over the last century (Chandrasekaran *et al.*, 2011). The extensive knowledge gained from previous studies make the honey bee a model organism for understanding how social behavior evolves and how sociality influences patterns of genome evolution to enhance or limit the spread of social alleles.

Previous studies investigated how positive selection influenced the evolution of eusociality in honey bees, such as adaptive evolution of worker traits. Positive selection was also studied in other eusocial insects such as bumble bees (*Bombus impatiens*) and ants (Jancek *et al.*, 2013; Roux *et al.*, 2014; Harpur *et al.*, 2017). Moreover, a recent study compared positive selection on the genome of honey bees, bumble bees, and paper wasps (*Polistes dominulus*) (Dogantzis *et al.*, 2018). However, there are no comparable studies quantifying evidence of negative selection on social insects.

My research is the first to determine the extent of negative selection in social insects by carrying out population genomic analyses in honey bees (*Apis mellifera*), bumble bees (*Bombus impatiens*), and paper wasps (*Polistes dominulus*).

What is negative selection?

Negative selection, also known as purifying or purging selection, is defined as the removal of deleterious alleles in a population (Loewe, 2008). When a mutation occurs in the genome (such as in protein coding sequences or non-coding regions), its effect will lie on a spectrum, ranging from lethal to neutral to strongly advantageous. Most mutations are neutral, where there is no significant effect on the fitness of the individual in the population. Lethal and strongly advantageous mutations occur on the extremes of the spectrum; lethal mutations drastically reduce fitness, whereas strongly advantageous mutations enhance the fitness of individuals with these mutations.

Therefore, deleterious alleles have a wide range of effects and decrease the fitness of individuals. Negative selection decreases the frequency of such deleterious alleles by removing them from the population. This leads to reduced genetic variation in regions of the genome where deleterious alleles usually occur.

Estimating negative selection

With the availability of genotypic data, the distribution of fitness effects (dfe) parameter from population genetics can be used to understand the effect of deleterious mutations (Eyre-Walker and Keightley, 2007). Dfe gives the probability that a mutation will have a given negative effect, relative to the regions of the genome that evolve neutrally. These neutrally evolving regions are usually 4-fold degenerate/synonymous sites, where different bases in a codon (for example, the 3rd position of a codon) of a gene sequence code for the same amino acid. Dfe is estimated based on a maximum likelihood approach by comparing the allele frequencies at other sites of the genome (such as non-synonymous gene coding regions, introns, or binding sites) to the benchmark (the neutrally evolving sites) to determine if negative selection has occurred at these regions and at what strength (Zhen and Andolfatto, 2012)).

By comparing the neutrally evolving sites and the sites being investigated, dfe gives a value for the strength of selection ($N_e s$), with 'N_e' representing the effective population size and 's' the selection coefficient. Sites experiencing stronger negative selection will have larger coefficients of selection and show less variance in the genome when compared to neutrally evolving sites. Therefore, there is a shift in the allele frequency spectrum of sites impacted by negative selection, where there are more deleterious alleles having low frequencies and less deleterious alleles having high frequencies (Figure 0.1). This method of estimating dfe has been implemented into the software DFE-Alpha by Eyre-Walker and Keightley in 2007. DFE-Alpha incorporates demographic processes that allows for population size changes, thereby providing robust estimates of negative selection in the genome. This software has been used to investigate negative selection in a variety of species, such as the insects *D. melanogaster, C. pipiens, M. cinxia, M. barbarus, H. scabiosae, R. grassei* (Romiguier *et al.*, 2014; Elyashiv *et al.*, 2016), *the plants C. grandiflora and C. rubella* (Williamson *et al.*, 2014; Lafon-Placette *et al.*, 2018), *and H. sapiens* (Boyko *et al.*, 2008; Chun and Fay, 2009; Harris and Nielsen, 2016).

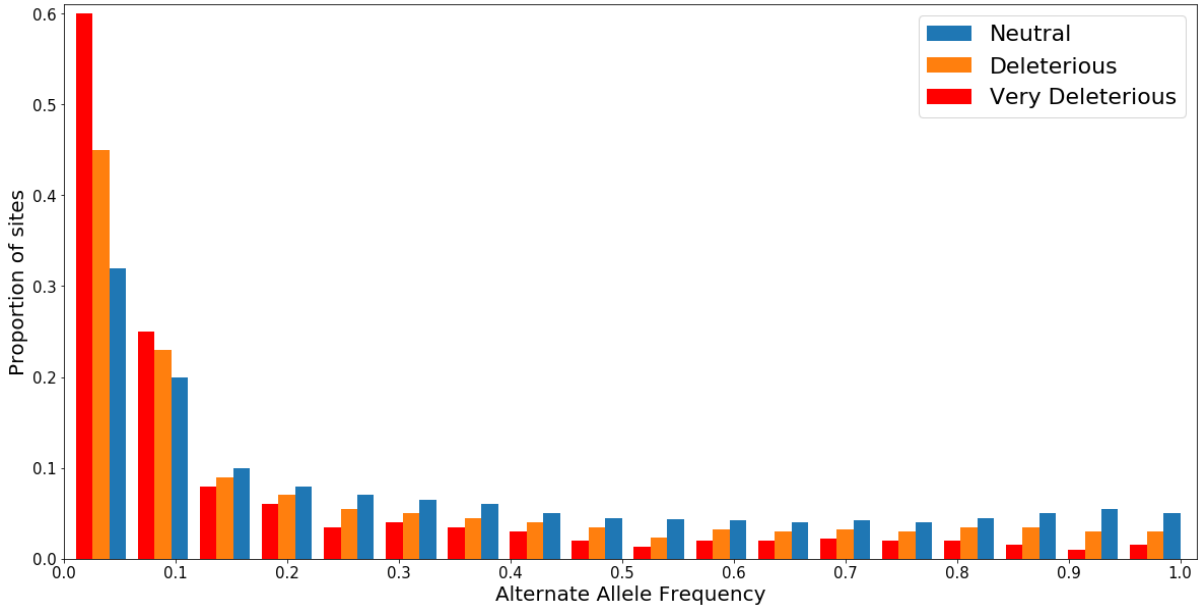

Figure 0.1: Theoretical Allele Frequency Spectrum. The y-axis represents the proportion of sites that have a given allele frequency on the x-axis. Highly deleterious sites show a leftward shift in the spectrum, where there are more sites with low alternate allele frequency (compared to the reference allele frequency) and fewer sites with high alternate frequency.

For this study, we used DFE-Alpha to estimate negative selection for *A. mellifera*, *B. impatiens*, and *P. dominula*. More specifically, we looked at negative selection in 5 regions of the genome – intronic regions, intergenic regions, 5' Untranslated Regions (UTR), 3' UTR, and protein coding regions (also known as missense sites or 0-fold sites). For each of these species, we looked at the extent of negative selection in their whole genome and compared the negative selection estimates among the species. The 3 species herein have varying levels of eusociality and we expected a correlation between levels of eusociality and negative selection (Chapter 1). Secondly, we compared the amount of negative selection in queen-biased genes against worker-biased genes in all 3 species. We expected queen-biased genes to show stronger negative selection due to queens being reproductive and affecting the fitness of the colonies directly (Chapter 2). Lastly, we compared the strength of negative selection in drone-biased genes against queen-biased genes. Both drones and queens are reproductive, with drones being haploid and queens being diploid. We expected haploid individuals to experience stronger negative selection as there are less chances for masking deleterious alleles in these individuals (Chapter 3).

Chapter 1: Negative selection and social complexity

1. Introduction

A. mellifera, *B. impatiens*, and *P. dominula* are all eusocial insects. These 3 species all show cooperative brood care, overlapping generations, and division of labor or caste differentiation - the three hallmarks of eusocial insects (Keller and Chapuisat, 1999). Cooperative brood care is when individuals of a species cooperatively take care of the brood in the nest. These young include offspring from other individuals and hence can be siblings. The second condition - overlapping generations – implies that individuals of different generations occupy the same nest during the same time (i.e. mothers and her offspring).In eusocial hymenopteran insects, there are typically 2 distinct female castes: diploid reproductive queens, or diploid sterile workers (Hölldobler and Wilson, 1990). This is a division of labor creating reproductive and nonreproductive groups. Haploid reproductive males are also present, although their role is purely reproductive. The reproductive group (queens and drones) are responsible for reproduction only while the nonreproductive workers take care of the colony, including nursing the brood. Although the 3 species studied herein show these characteristics, *A. mellifera*, *B. impatiens*, and *P. dominula* differ in their expression of eusociality.

A. mellifera is the most advanced eusocial insect in this species triad (Barchuk *et al.*, 2007). They have a large colony size, with a single queen, few hundreds of drones, and over 50, 000 workers (Engel, 1999). *A. mellifera* uses swarms to found new colonies, where the old mated queen will move out of the colony, along with a few thousand workers, to establish a new colony. Once settled in, the queen will start laying fertilized eggs to develop new worker brood and the workers will take care of the new offsprings (Gould and Gould, 1995). Compared to other eusocial species, *A. mellifera* workers have a highly advanced division of labor, where

specific set of individuals will perform certain tasks in the colony (Michener, 1969). For example, young worker bees are usually the ‘nurse’ bees that will take care of the larvae in the colony. These nurses will feed the larvae and also clean the hive space surrounding the larvae. As these nurses get older, they transition to foragers. Foragers have the task of finding food resources, such as nectar and pollen, for the colony. When a forager finds a resourceful location, she goes back to the hive and performs a waggle dance. The frequency and vector movement of the waggle dance notifies the other bees in the hive where that particular resource is and how far out it is (Riley *et al.*, 2005). Hence, she is able to communicate with the other bees in the hive. Other tasks performed by workers include undertaking dead bees from the colony, and guarding, where guard bees attack intruders and check if their own siblings are coming back to the colony.

Although *A. mellifera* workers can perform a wide variety of tasks, they are effectively sterile and cannot lay fertilized eggs. In a colony, only the queen is the reproductive female that will lay fertilized eggs, which will develop into workers or future queens. Hence, we see a high extent of caste differentiation in *A. mellifera*. In the absence of a queen, *A. mellifera* workers can activate their ovaries to lay haploid drone eggs, but not diploid female eggs (Cardinal and Danforth, 2011). Moreover, queens are morphologically different compared to workers, where the queens have longer abdomens. Hence, workers are vastly different from queens with extensive caste differentiation.

The paper wasp *P. dominula* is primitively eusocial, and is the least social among the three species studied herein (Dogantzis *et al.*, 2018). *P. dominula* has a small colony size of around 100 individuals (Hocherl and Tautz, 2015), much smaller than *A. mellifera*. For founding, *P. dominula* uses the foundress strategy (Eberhard, 1969). Foundresses are overwintered gynes (i.e. future queens). As winter comes along, gynes in the colony will

overwinter and hibernate while the rest of the colony dies. In the spring, these overwintered gynes, now foundresses, will build the nest and start laying eggs, with the foundress laying the most eggs emerging as the dominant queen. The other foundresses will simply help the colony. Regarding workers, the behavioral tasks are less stringent compared to *A. mellifera*. Most workers can nurse or forage, or do both at the same time (Jandt, Tibbetts and Toth, 2014). The workers also retain the ability to become fully reproductive. For example, the removal of the dominant queen, the second most dominant female from the colony will assume the dominant role and lay fertilized eggs (Theraulaz, Pratte and Gervet, 1989). Moreover, it has been shown that workers will start laying eggs when there is a decline in the colony health (Liebig, Monnin and Turillazzi, 2005). Hence, the workers are totipotent that can switch to a reproductive state and lay fertilized diploid eggs like the queen, showing a low degree of caste divergence. Morphologically, *P. dominula* queens and workers are not different (Hunt, 2007). Another example of low divergence is the presence of gynes, which are daughters destined to become queens. Due to some interactions in the colony, some workers are influenced to become gynes and will overwinter to found the next colony (Dapporto, Turillazzi and Palagi, 2006). An interesting characteristic of *P. dominula* nests is that not all the individuals of the colony are related to each other (Queller *et al.*, 2000). This means that helpers in the colony are not necessarily related to each other and are still taking care of the young in the colony. These characteristics make *P. dominula* primitively eusocial.

B. impatiens is considered intermediately eusocial, due to its life history. Compared to *A. mellifera* and *P. dominula, B. impatiens* has a intermediate colony size number between these 2 species, which is usually around 500 individuals (Michener, 1974). *B. impatiens* is similar to *P. dominula* ins many aspects, such as using gynes and foundresses for colony founding. But *B. impatiens* workers are not totipotent; they are unable to mate and lay diploid eggs, but they can lay unfertilized male eggs (Cnaani, Schmid-Hempel and Schmidt, 2002).

However, with the queen present, the workers do not attempt to lay eggs even when they had mature oocytes, showing behaviours similar to *A. mellifera*. Size of the workers plays a role in their task in the colony. Typically, smaller workers are located in the center of the nest where they care for the young while larger workers venture around the edge of the nest and act as guards or foragers (Jandt and Dornhaus, 2009). Moreover, *B. impatiens* queens and workers are morphologically distinct, similar to *A. mellifera* and unlike *P. dominula*. With these characteristics, *B. impatiens* show intermediate caste divergence, with distinct but blurred lines between queens and workers.

Hence, in the 3 species studied herein, *A. mellifera* is the most socially complex, *P. dominula* is primitively eusocial, with *B. impatiens* being intermediately social. This gradient of social complexity allowed us to investigate how social complexity relates to negative selection.

In population genetics, the effective population size (N_e), a parameter estimating the number of breeding individuals in a population, is often used in lieu of the actual population size (Halliburton, 2004). In social insects, only a few individuals are reproductive while the majority of the population do not usually reproduce. This leads to a substantial reproductive skew in eusocial species, limiting the number of reproductive individuals and decreasing the effective population size in these individuals (Crozier, 1976).

Romiguier et al. (2014) hypothesized and showed some evidence of decreasing effective population size with increasing social complexity. In their study, they looked at several insects with varying social complexity and showed that species with higher social complexity showed effective population sizes similar to vertebrates, that is low effective population size. Romiguier et al. suggested that with decreasing effective population size, the efficiency and strength of selection on such species decrease. Therefore, they expect a positive

correlation between effective population size and selection. Several studies have suggested that species with larger effective population sizes would show evidence of stronger and more effective selection, both positive and negative (Lynch and Conery, 2003; Wright and Andolfatto, 2008; Akashi, Osada and Ohta, 2012; Gossmann, Keightley and Eyre-Walker, 2012; Galtier, 2016; Harpur *et al.*, 2017).

In 2015, Kapheim et al. published their study looking at trends in genomic signatures of bees transitioning from solitary to group living. One of their main findings was that with increasing social complexity, genes in these species showed evidence of relaxed selection, akin to weaker negative selection. Kapheim et al. further showed that with increasing social complexity, there was an increased in capacity for gene regulation, mainly through transcription factors. This suggests that there is relaxed constraint, coupled with positive selection to allow better gene regulation mechanisms. Their finding that genes involved in coordinating gene regulation show evidence of rapid evolution in socially complex species further support the relaxed selection concept. Some of the rapidly evolving genes in highly eusocial species were under relaxed negative selection and Romiguier et al. (2014) hypothesized that it was due to reduced effective population size.

In this study, we suspect *A. mellifera* has the lowest effective population size and *P. dominul*a has the highest, with *B. impatiens* having an intermediate effective population size between *A. mellifera* and *P. dominula* (Dogantzis et al., 2018). The high effective population size of *P. dominula* may be attributed to dominant workers easily replacing a dead queen or the fact that 35% of a colony are genetically unrelated, meaning that multiple 'families of wasps' are living together (Queller *et al.*, 2000). In population genetics, the strength of selection is correlated with effective population size. Since negative selection is proportional to N_e (Strength of selection = N_e * s, where s is the selection coefficient determined by DFE-Alpha),

a decrease in N_e will reflect a decrease in negative selection, or a relaxation of constraint. Therefore, we hypothesized that we would see strong effects of negative selection in the genome of *P. dominulq*, and relatively weak effects in *A. mellifera* due to the difference in effective population size (Eyre-Walker, 2002). The null hypothesis was that there would be no difference in the strength of negative selection between the different species.

2. Materials and Methods

1. Data and samples

We worked with published datasets for *Apis mellifera scutellata* (49 worker samples), *Bombus impatiens* (10 drone samples), and *Polistes dominula* (10 worker samples) (Harpur et al., 2014; Harpur et al., 2017; Dogantzis et al., 2018). We obtained the allele frequency spectrums for the different species by counting sites in the variant-calling files (VCF). These VCF contain information outlining where mutations are in the genome and the details of each mutation. Invariant and variant sites were removed if: 1) the site quality was below 50; 2) read depth was not within the interquartile range (*A. mellifera*: 1300-2907, *B. impatiens*: 31-223, *P. dominula*: 69-337); or 3) the site was within 5 base pairs of indels. Moreover, sites with more than 2 alleles were removed to avoid analyzing SNPs with probable sequencing error.

2. Site annotation and filtration

We used SnpEff (Cingolani *et al.*, 2012) to annotate the sites passing all the filtering criteria and from these annotations we obtained the different site categories: synonymous sites, intergenic regions, intronic regions, missense (0-fold sites), 3' and 5' UTRs (untranslated regions). Genes with warning for incomplete transcripts, multiple stop codons, or no start codon were excluded from the analysis. Genes with loss or gain of start or stop codon were also discarded.

3. Estimation of the distribution of fitness effects

To quantify the amount of negative selection acting on the different categories of sites, we used the methods of Eyre-Walker and Keightley (2010). DFE-alpha uses the allele frequency spectrum (AFS) of a site category of interest and compares it to the AFS of neutral sites. For neutral sites, we used 4-fold degenerate sites. These are sites in the genome where any mutation causes the same protein to be coded and they are assumed to be under neutral selection in population genomic studies (Williamson et al., 2014). We bootstrapped each region 5000 times to obtain 95% confidence intervals.

3. Results

We estimated the proportion of sites under negative selection for 5 genomic regions intronic, intergenic, 3' UTR (generally found at the end of genes), 5' UTR (usually located before genes), and 0-fold. $N_e s$ bins (from DFE-Alpha) of 0-1 represents effectively neutral evolution, 1-10 represents weak negative selection, 10-100 represents moderate negative selection, and > 100 represents strong negative selection.

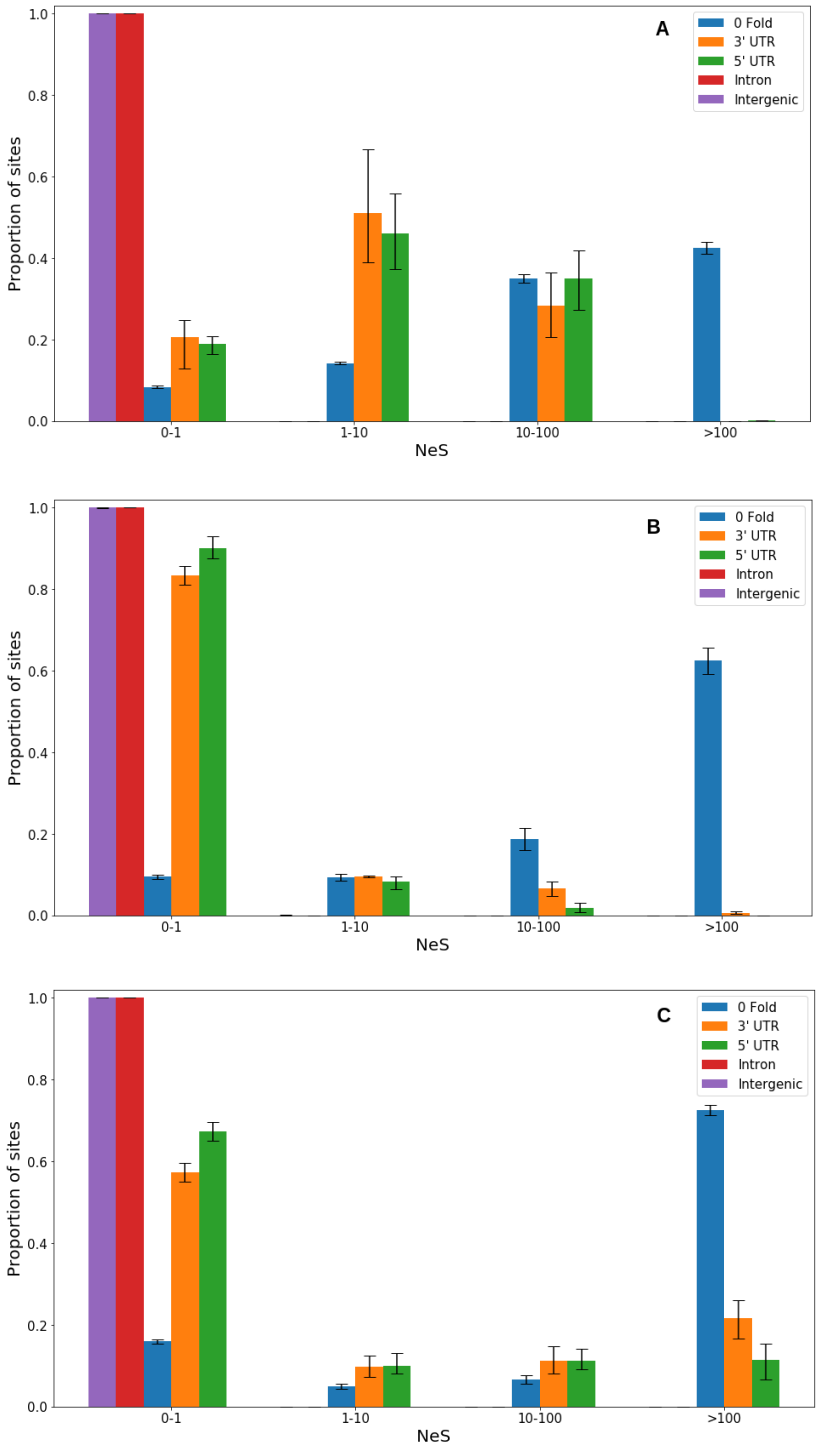

Figure 1.1: DFE-Alpha outputs for *A. mellifera scutellata* (A), *B. impatiens* (B), *P. dominula* (C). N_es bins (from DFE-Alpha) of 0-1 represents effectively neutral evolution, 1-10 represents weak negative selection, 10-100 represents moderate negative selection, and > 100 represents strong negative selection. UTR regions represent Untranslated Regions in the genome.

Figure 1.1 presents negative selection across the whole genome for each of *A. mellifera*, *B. impatiens*, and *P. dominula*, focusing on the 5 genomic regions mentioned above.

Figure 1.1.A shows the DFE-Alpha results for *A. mellifera*. Intronic and intergenic regions show no evidence of negative selection, evident by most of their sites being found in the 0-1 $N_e s$ bin. Hence, these sites are likely evolving neutrally. 5' UTR sites and 3' UTR sites show evidence of moderate negative selection (10-100 $N_e s$ bin), with 36% of 5' UTR and 31% of 3' UTR sites (Wilcoxon t-test, $p < 3e^{-15}$). Lastly, we see the most amount of negative selection in 0-fold (missense) sites, with 42% of the sites being under high negative selection ($N_e s > 100$).

For *B. impatiens* (Figure 1.1.B), intronic and intergenic regions show evidence of neutral evolution, with the majority of their sites found in the 0-1 $N_e s$ bin. Again, 5' UTR and 3' UTR sites show moderate negative selection (10-100 $N_e s$), with 9% and 3% of the sites respectively. Interestingly, we see 0.5 % of 3' UTR showing high negative selection ($N_e s > 100$). However, it should be noted that the *B. impatiens* genome is not clearly annotated as the *A. mellifera* and *P. dominula* genomes. In fact, the UTR regions were modeled from SnpEff and may not be an accurate representation of the actual UTR sites in the genome. Finally, we see a large proportion of missense sites under high negative selection (62% of sites with $N_e s > 100$).

In *P. dominula* (Figure 1.1.C), intron and intergenic regions show no sign of negative selection. 3' UTR and 5' UTR sites show evidence of strong negative selection, with 21% and 16% of their sites respectively having $N_e s > 100$. Missense sites significantly show the largest amount of negative selection here, with 72% of the sites having $N_e s > 100$.

Comparing *A. mellifera, B. impatiens* and *P. dominula,* interesting patterns stand out. In all 3 species, intronic and intergenic regions both show no signature of negative selection,

suggesting these sites are evolving neutrally. Another similarity is the evidence for moderate to strong negative selection on UTR regions in the genomes of all 3 species, for both 5' UTR sites and 3' UTR sites. We found stronger negative selection in UTR regions of *P. dominula*, compared to both *A. mellifera* and *B. impatiens* ($p < 1e\text{-}16$). Finally, in all 3 species, we see a high proportion of 0-fold (missense) sites experiencing strong negative selection, which is a high $N_es > 100$ (44% in *A. mellifera*, 62% in *B. impatiens*, and 72% in *P. dominula*; Kruskal-Wallis 1-way ANOVA, $p < 1e^{-16}$) (Figure 1.2). This is expected as 0-fold sites are in protein-coding regions of the genome and these sites code for specific amino acids. Although the proportion of 0-fold sites experiencing strong negative selection are different among the 3 species, these regions still show the most negative selection compared to other regions. Similar results were presented for these regions in *C. grandiflora* (Williamson et al., 2014).

For the 3 species, we suspect *A. mellifera* has the smallest effective population size and *P. dominula* has the largest effective population size, with *B. impatiens* being intermediate. We expected sites in *A. mellifera* to show less negative selection, with *B. impatiens* showing intermediate negative selection and *P. dominula* having strongest negative selection.

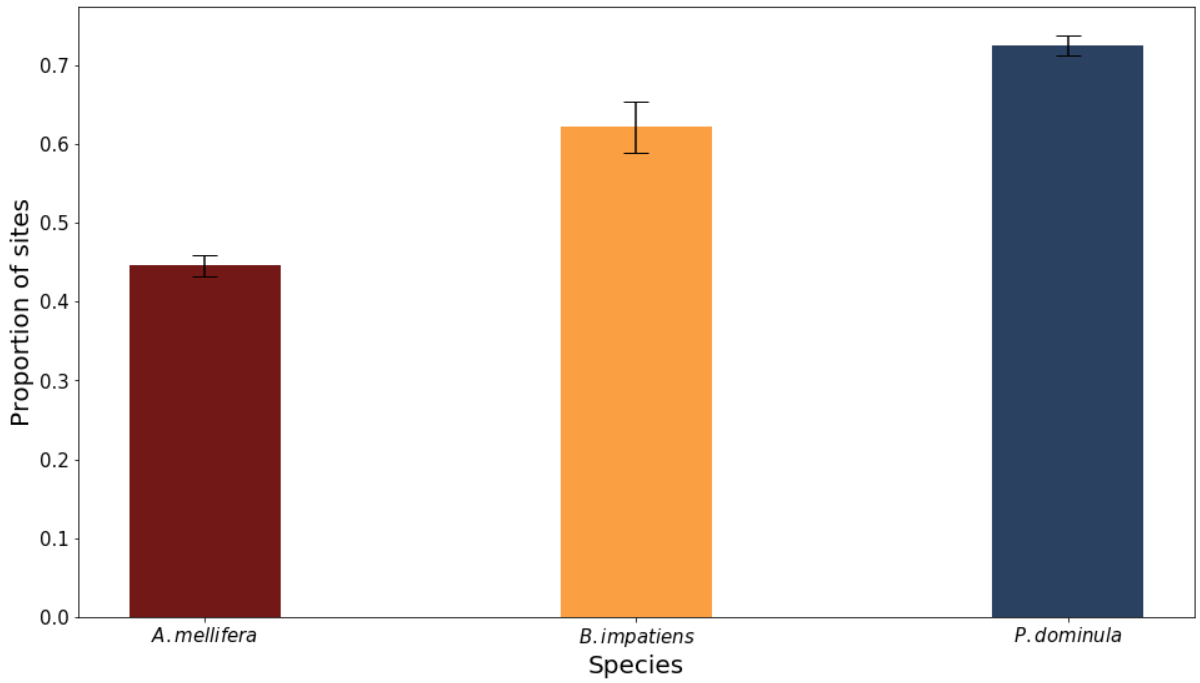

Figure 1.2: Estimates of negative selection for missense (0-fold) sites under high N_es (>100) in the genome of *A. mellifera scutellata, B. impatiens, P. dominula*. Data was obtained for 52 *Apis mellifera* workers, 10 *Bombus impatiens* drones and 10 *Polistes dominulus* workers. Sites were annotated using SnpEff and DFE-alpha was run on missense sites from these species, with 4-fold degenerate sites used as the neutral selection sites. The error bars represent the 95% bootstrapped confidence intervals. We see higher negative selection in less socially complex species ($p < 1e^{-16}$).

4. Discussion

We estimated the amount negative selection in the genome of *A. mellifera*, *B. impatiens*, and *P. dominula*.

Negative selection in the different regions of the genome

We observed no evidence of negative selection in intronic and intergenic regions, moderate to high negative selection in UTR regions and high negative selection in missense sites.

Intergenic regions are found in between genes and their functions in the genome have yet to be clearly characterized (Birney et al., 2007). These regions have widely been considered to be 'junk' DNA but recent studies, including the ENCODE project, have found some of these to be functional based on the conservation of some intergenic sequences (Djebali et al., 2012; Dunham et al., 2012). Some enhancer sequences, where proteins can bind, have been found in these intergenic regions. Protein binding to these enhancers can change gene expression during development to change the fate of a cell (Schmidt *et al.*, 2015). For example, binding to enhancers found in intergenic regions can lead to increased expression of genes found downstream of these intergenic regions, even though these genes are located thousands of base-pairs downstream. Moreover, it has been hypothesized that intergenic regions contain noncoding RNAs (Birney et al., 2007). Noncoding RNAs are sequences that are not translated into protein but rather serve as a system to control gene expression. For example, some noncoding RNAs can bind to coding RNAs to silence them and prevent them from being translated. With no translation, the amount of that protein decreases, hence a reduced expression of that gene (Zhu, 2006).

However, it's only recently that we have started identifying some of the subtle functions of intergenic regions. We see no evidence of selection on intergenic regions and these results

correlate with previous studies, such as Williamson et al. (2014). We suggest that due to the vast amount of intergenic regions in the genome, we do not have the resolution to detect negative selection on a few small intergenic regions.

Intronic regions are located between coding sequences of genes, within a gene, in contrast to intergenic regions. These introns do not code for amino acids and are not translated. Similar to intergenic regions, introns were believed to have little to no function in the genome (Penny *et al.*, 2009) until recently. (Rearick *et al.*, 2011) showed that some introns contain noncoding RNAs that can act as transfer RNAs (tRNA), ribosomal RNAs (rRNA), both important during translation, and noncoding RNAs. tRNAs catalyze the addition of amino acid to the elongating protein chain, allowing the right amino acid 'block' to be put in place (Crick, 1968). rRNAs are associated with the ribosome and help the latter function efficiently (Wolfe, 1993). The ribosome is part of the complex that reads a mRNA transcript for translation into a protein sequence and eventually a functional protein. Noncoding RNAs can also act as microRNAs and siRNAs where they regulate gene expression (Mack, 2007; Rana, 2007). These noncoding RNAs bind to functional RNAs and prevent the latter from being translated.

Introns are also involved in alternative splicing, which is a mechanism used to obtain different proteins from the same mRNA transcript. In alternative splicing, specific exons (regions coding for amino acid) are retained in specific orders, allowing different polypeptide chains to be made and giving rise to different proteins. These proteins are often related to each other and work in the same pathway in the cell (Bicknell *et al.*, 2012). Although the majority of introns are removed from the transcript during alternative splicing, an increasing number of studies have shown that some introns are retained inside the transcript (Sammeth, Foissac and Guigó, 2008; Rearick *et al.*, 2011). These retained introns downregulate gene expression by limiting the number of transcripts that will be read and successfully translated into proteins.

Given we see no evidence of negative selection on introns, we hypothesize that most of the intronic sequences are neutral, with only a few bases being important for gene regulation.

UTR regions are untranslated regions in the genome and hence do not code for amino acids. However, they are still kept in the mRNA transcript and not spliced out like introns. These UTR regions are usually involved in the regulation of transcription of protein sequences, determining the level of gene expression and amount of protein made (Araujo *et al.*, 2012; Barrett, Fletcher and Wilton, 2012).

5' UTR regions are located upstream of genes and generally regulate the stability and translation efficiency of the mRNA of the downstream genes. Protein complexes bind to these sequences and help in recruiting of the ribosome complex, initiating translation and reading the codes to make a protein (Araujo et al., 2012). The ribosome complex is made up of RNAs and proteins, and is the main complex that reads the mRNA to translate the latter into amino acid sequences. Conversely, RNA binding proteins can bind to the 5' UTR regions and fold the sequence in such a way that it prevents the ribosome complex from being recruited, thereby preventing translation (Kozak, 2008). Although they are called untranslated, some sequences of the 5' UTR have an open reading frame (ORF) that can be used to start translation within the 5' UTR. These specific sequences have their own initiation codon that will be read by a ribosome to make proteins (Wethmar, Smink and Leutz, 2010). These proteins can then regulate translation of the main protein coding sequence of the mRNA, by either repressing or enhancing translation.

3' UTR regions are located downstream of genes and have similar functions to 5' UTR, in the sense that both help in the regulation of translation (Barrett et al., 2012). 3' UTR also help in the stability and translation efficiency of the mRNA (Pichon *et al.*, 2012). There are many intricate sequences in the 3' UTR such as microRNA response elements (MRE), AU rich

elements (ARE), and poly A tails among others. The MRE are sequences where microRNA can bind and form a complex, preventing tRNA from being recruited and halting translation, serving as a repression mechanism (Alshalalfa, 2012). At the ARE, ARE binding proteins enhance the decomposition of the mRNA or enhance the stability of the mRNA (Barrett et al., 2012). This effect depends on the types of ARE binding proteins being recruited and can hence upregulate or downregulate gene expression. The poly A tail consists of long stretches of the adenosine base, where additional proteins can bind to regulate the stability of the mRNA (Proudfoot, Furger and Dye, 2002). In general, longer 3' UTRs tend to have lower levels of mRNA expression, suggesting that with longer 3' UTRs, there are more sites for miRNAs and proteins to bind, repressing the translation of the mRNA. (Barrett et al., 2012)

5' UTRs and 3' UTRS can also interact with each other. Special RNA binding proteins bind to the 5' UTRs and 3' UTRs simultaneously and this interaction causes the mRNA to form a loop (Gilbert, 1988). With the loop, the ribosome complex cannot read the transcript, repressing gene expression. However, Kozak (2008) disputes this concep.

Therefore, mutations in UTR regions will affect the expression of multiple genes, even genes that appear to be unrelated. Deleterious alleles negatively impacting these UTR regions could eventually lead to misregulation of multiple genes, decreasing the fitness of individuals carrying these alleles. Due to these functions, we expected UTR regions to experience moderate to high negative selection. Our results are consistent with these expectations, albeit UTR sites show relaxed constraint relative to missense sites. Moreover, UTR sites showed signatures of moderate to strong negative selection in *C. grandiflora* (Williamson et al., 2014)

Missense, or 0-fold sites, are protein coding regions in the genome. These sites usually form part of a codon, a 3-base sequence, that code for an amino acid. Although there codons can be degenerate, that is multiple codons can code for the same amino acids, missense sites

are non-degenerate (Lagerkvist, 1978). Any mutation at that specific base will cause a different amino acid to be coded and potentially lead to a loss of protein function (Minde *et al.*, 2011). In fact, the majority of new mutations that occur are considered to be deleterious and lead to loss of functions (Loewe and Hill, 2010). Given mutations on these sites will adversely affect the amino acids being coded for, and hence the proteins being made, we expected negative selection to be high in these regions to reduce the frequency of deleterious alleles from potentially producing non-functional proteins. Jackson, Campos, & Zeng (2015) showed that missense sites in the fruit fly, *Drosophila melanogaster*, are under stronger negative selection compared to other regions in the genome, with around 78 % of sites under high negative selection. Williamson et al. (2014) found high levels of negative selection in protein coding regions of the plant *Capsella grandiflora*, showing that these missense regions experience strong negative selection other than insects.

Negative Selection and Eusociality

We observed strongest negative selection in the genome of *P. dominula*, followed by *B. impatiens* and *A. mellifera* showed the least amount of negative selection. Hence we see a trend that with increasing social complexity, negative selection is less efficient, supporting our hypothesis and Romiguier et al.' hypothesis (2014). As insects evolve to have more socially complex societies, reproduction becomes restricted to fewer individuals in the population. Queens are usually the only reproductive females in these 3 species but in the absence of a queen, some workers can activate their ovaries to lay eggs, becoming reproductive. In primitively eusocial species, like *B. impatiens* and *P. dominula*, even though individual colonies are smaller compared to *A. mellifera*, there are more reproductive individuals at a population level. We suspect this is due to these species not being socially advanced enough to have only a few reproductive individuals to allow colonies to reach large numbers, like in *A.*

mellifera. Since reproductive individuals are passing on their alleles to the next generation, these reproductive individuals are affecting the fitness of the population directly and will experience stronger selection (Linksvayer and Wade, 2009). With the number of reproductive individuals being smaller in highly eusocial species, selection is less efficient. This can explain why we observed weaker negative selection in *A. mellifera*, the most socially complex insect in our triad of species and strongest negative selection in *P. dominula*, the most primitively social insect in this study.

Hence, our results support the hypothesis that with decreasing social complexity, the strength and effectiveness of negative selection decreases. We also see no signs of negative selection in intronic and intergenic regions, moderate negative selection in UTR sites and strongest negative selection in missense sites. To further expand on this study, we could look at negative selection in other social complex insects, such as termites and ants, to help us confirm if strength of negative selection and levels of eusociality are indeed corelated.

Chapter 2: Negative Selection in Queen against Workers in Social Insects

1. Introduction

Social insects have a fascinating living structure. In hymenopteran social insects, the individuals are usually separated into 3 different castes: a diploid reproductive female queen, a few haploid reproductive male drones, and thousands of diploid non-reproductive female workers. The castes carry out different functions, with workers being the most diversified in tasks.

The reproductive task is typically confined to queens and males (referred to as drones in *A. mellifera* and males in *B. impatiens* and *P. dominula*). Usually, the queen is the only reproductive female in the colony and she typically releases pheromones that down regulate reproductive genes in workers, rendering the workers non-reproductive (Pirk *et al.*, 2004). Hence, almost every individual in the colony is genetically related to the queen. In the 3 social insects studied here, once the queen has mated, she stores the sperm in her spermatheca, a special organ for sperm storage. She will never mate again during her lifetime and will simply use the stored sperm to fertilize her eggs. Fertilized eggs develop into workers and unfertilized eggs develop into males (Peters *et al.*, 2017).

In *A. mellifera*, female larvae that are fed royal jelly, a highly nutritious food, by the workers will likely develop into queens (Buttstedt *et al.*, 2018). The first virgin queen to emerge will usually kill of the other queens that have yet to emerge. The surviving virgin queen will eventually go on mating nights in congregational areas, where drones gather to mate with queens (Waldbauer, 1998). The mating can occur over several nights. Once a drone has inseminated the queen, the drone's abdomen is ripped off and he will eventually die. The queen generally mates with 12 to 20 drones over the mating nights. Once the queen has mated with enough drones, she will store the sperm in her spermatheca and will gradually use the sperm to

fertilize the eggs she lays. Given that the queen mates with few drones only, genetic diversity should be alarmingly low. However, *A. mellifera* has a high genomic recombination rate (Kent *et al.*, 2012), allowing variants to be constantly introduced in the population and increasing genetic diversity. In *A. mellifera*, queens solely lay eggs and are not involved in any aspect of the colony. The workers take care of the young, build the nest, find a new nest site if needed, forage and do everything else for the colony (Ratnieks and Helanterä, 2009).

In *B. impatiens*, larvae that were fed copious amounts of food (nectar and pollen) by the workers will develop into gynes, which are females destined to become queens. A gyne will mate with a single male bee in fall, at the end of the hive life cycle. The male will mate long enough for his sperm to harden and after mating the vaginal orifice is plugged (Laverty and Harder, 1988). The mated gyne will then hibernate until spring. During spring, the gyne will emerge and find a potential nest, usually in the ground or in tree crevices, to form a colony. When a suitable nest site is found, the gyne (queen) will set up a honeypot and fill this with nectar and pollen, and will use this honey pot to feed her new brood (McAulay, Otis and Gradish, 2015). During this time, the queen has started laying eggs that will develop into offspring female workers. In late summer, the queen will lay unfertilised eggs that will develop into males and workers will start feeding larvae large amount of food to increase gyne production, hence restarting the cycle (Cnaani, Schmid-Hempel and Schmidt, 2002).

P. dominula gynes are chosen similar to *B. impatiens*, where they are fed large amounts of food. Similarly, gynes will mate during fall but here *P. dominula* gynes will mate with several males (Liebert *et al.*, 2010). In leks, *P. dominula* males will fight to prove their dominance, with the losers leaving the area. The gynes will fly to these areas and choose the dominant males, which is indicated by the variation of spots on the males' abdomens (Izzo and Tibbetts, 2012). Since *P. dominula* gynes are larger and more aggressive than the males, the

gynes are able to prevent copulation from unwanted males by biting them or stinging them (Reeve, 1991). After a number of copulations, the mated gynes (now foundresses) will hibernate with other females or alone. Similar to *B. impatiens*, the foundresses will emerge during spring, build the nest and provide food to their offspring. In some cases, multiple foundresses will emerge together and lay eggs together. The foundress that lays the most eggs usually becomes the dominant queen and the other foundresses will stop laying eggs and help the dominant foundress take care of the colony (Tibbetts, 2003). In late summer and fall, *P. dominula* is similar to *B. impatiens* where future foundresses and males are laid.

In *A. mellifera*, *B. impatiens*, and *P. dominula*, workers perform most of the tasks in the colony. These include foraging for food (foragers), guarding the colony against intruders (guards), feeding and taking care of the young in the colony (nurses), removing sick larvae and deal individuals from the colony (undertakers), among others (Jandt et al., 2014; Jandt & Dornhaus, 2009; Michener, 1969). Foragers will go outside and look for resources and come back to the colony. These foragers bring nectar, which is used as fuel and energy source, and pollen, which is high in protein and used to build tissues, to the colony. Guards check the colony for intruders and will attack them, by either biting the intruders or stinging them.

In *A. mellifera*, workers usually transition roles as they mature (Amdam and Omholt, 2003). For example, young workers become nurses and take care of larva and other young offspring in the colony. As these nurses mature, they transition to other distinct roles, such as foragers, and guards in some cases. Foragers are able to waggle dance to direct other bees to where the resources are, recruiting more workers to efficiently harvest a resource (Riley *et al.*, 2005). The transition between roles is marked by upregulation and downregulation of certain genes (Whitfield, 2003). In the absence of a queen, some *A. mellifera* workers can lay unfertilized eggs that will develop into drones.

The tasks in *B. impatiens* workers are mostly defined by the size of the individuals (Jandt et al., 2009). Smaller workers tend to be nurses and will be found in the center of the hive, where most larvae develop. Larger workers are usually foragers and guards, roaming around the edge of the hives (Jandt and Dorhaus, 2009). In this case, being at the extremities of the hive allows these guard workers to quickly detect intruders and the foragers to easily leave the hive to forage for food. *B. impatiens* workers do not typically transition to different roles during their lifetime. In the absence of a queen, *B. impatiens* workers can activate their ovaries and lay unfertilized and fertilized eggs, if they are able to mate (Cnaani et al., 2002).

P. dominula workers roles are defined based on interactions between individuals in the colony (O'Donnell, 1996). Most of the workers forage and nurse, with no true distinction between the roles. Dominance among individuals influence what the workers can do in the colony. The more dominant a worker is, the more likely she will be to lay eggs even if a queen is around. In fact, Liebig et al. (2005) showed that removing larvae increases the reproductivity of workers. Removing the queen also increases worker reproductivity (Strassmann *et al.*, 2004). These studies indicate that in the case where a colony seems to have an unhealthy queen (laying less eggs and hence less larvae present) or no queen at all, the more dominant workers will lay eggs and take over the role of the queen. Hence, workers are totipotent and able to perform a wide variety of roles in the colony, even reproduction, with no clear roles assigned to any worker (Jandt et al., 2014).

Here we compared caste-biased genes in female queens and female workers. Since caste-biased genes differentially expressed in these castes lead to different morphology and behavior, caste-biased genes offer an interesting way of the evolution of traits. As the queen is the primary reproductive diploid female in the population (Pamilo and Crozier, 1997), her genes and alleles are passed directly to the next generation. Hence, genes that are differentially

expressed in the queen, hereafter called queen-biased genes, will directly affect colony fitness because the queen contributes directly to the population's genotype. As workers are typically non-reproductive but do take care of the colony through foraging and nursing for example, the workers affect the overall health of the colony and indirectly contribute to colony fitness.

Linksvayer and Wade (2009) hypothesized that genes affecting social (or colony) fitness directly would show stronger negative selection compared to genes affecting colony fitness indirectly. In this study, genes affecting colony fitness directly are queen-biased genes and genes affecting colony fitness indirectly are worker-biased genes. Therefore, we hypothesized that worker-biased genes would have experienced relaxed negative selection, resulting in an accumulation of more deleterious alleles. Queen-biased genes would experience stronger negative selection. However since *B. impatiens* and *P. dominula* workers can reproduce in certain circumstances, we would expect higher negative selection in their genes compared to *A. mellifera* workers. The null hypothesis is that there will be no difference in the strength of negative selection between queen-biased genes and worker-biased genes.

2. Materials and Methods

1. Data and samples

See Chapter 1: Section 2

2. Differentially expressed genes

All caste-biased genes were obtained from RNAseq data and data for *A. mellifera* was analyzed using DESeq2 (Love, Huber and Anders, 2014). RNAseq data for *A. mellifera* were obtained from Ashby et al. (2016) and He et al. (2019). We obtained caste-biased genes data from Harpur et al. (2017) and Amsalem et al. (2015) for *B. impatiens*, and Woodard et al. (2014) and Standage et al. (2016) for *P. dominula*. For all species, queen and worker biased genes were obtained from these datasets, with the addition of drone biased genes for *A.*

mellifera. In order to obtain estimates of the distribution of fitness effects, the VCF files were divided into the different castes consisting of only regions with the respective caste-biased genes. For comparison, only missense sites (0-fold sites) under high negative selection (N_es > 100) were analyzed since these are the regions showing the largest and noticeable changes in the genome.

3. Results

We tested the hypothesis that queen-biased genes would have experienced stronger negative selection compared to worker-biased genes because queen-biased genes affect social fitness directly whereas worker-biased genes indirectly affect social fitness (Linksvayer and Wade, 2009).

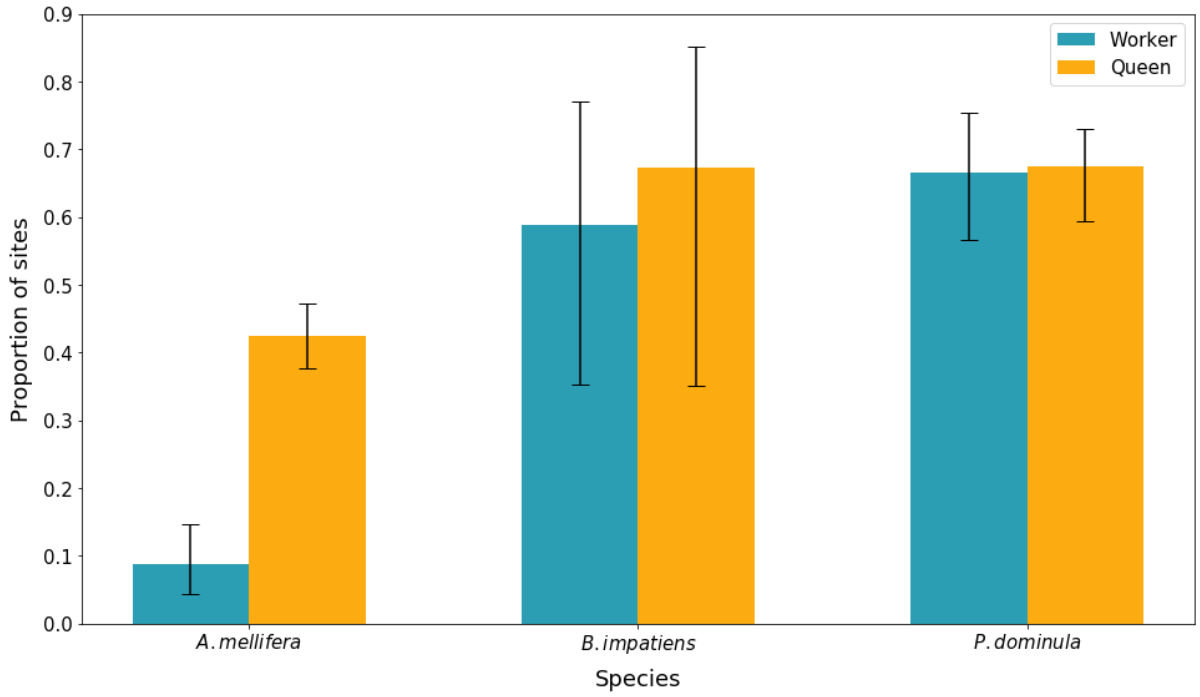

Figure 2.1: Estimates of negative selection for missense (0-fold) sites of queen-biased genes and worker-biased genes under high N_es (>100) in the genome of *A. mellifera*, *B. impatiens*, and *P. dominula*. Data was obtained for 52 *Apis mellifera* workers, 10 *Bombus impatiens* drones and 10 *Polistes dominula* workers. Sites were annotated using SnpEff and gene-biased castes were obtained from RNASeq data. DFE-alpha was run on missense sites from these caste-biased genes, with 4-fold degenerate sites used as the neutral selection sites. The error bars represent the 95% bootstrapped confidence intervals.

In *A. mellifera scutellata*, 42% of 0-fold sites from queen-biased genes are under strong negative selection compared to 8.7% for worker-biased genes (permutation test, $p < 1e^{-12}$) (Figure 2.1). For *B. impatiens*, 65 % of missense sites from queen-biased genes show evidence for strong negative selection while 57% of missense sites from worker-biased genes are under strong negative selection, with no significant difference between the caste-biased genes (permutation test, p=0.71). Lastly, we see no difference in strong negative selection between queen-biased genes (66% of sites) and worker biased genes (65% of sites) in *P. dominula* (permutation test, p=0.86). We observe an interesting pattern regarding social complexity and caste biased genes here. As social complexity increases from primitively eusocial *P. dominula* to intermediately eusocial *B. impatiens* and finally advanced eusocial *A. mellifera*, the difference in amount of negative selection between queen-biased genes and worker-biased genes increases.

To confirm that the difference in negative selection between *A. mellifera* queen-biased genes and worker-biased genes are solely due to caste-bias and not expression level, we divided the top 200 highly expressed genes and the bottom 200 lowly expressed genes for both queen-biased and worker-biased genes. There is no difference between expression levels for highly expressed genes in queens and highly expressed genes in workers (Mann-Whitney test, p=0.23). To obtain the distribution for proportion of sites under high negative selection ($N_e s > 100$), we ran DFE-Alpha on the subsets using 5000 bootstrap iterations. For highly expressed genes, there are more 0-fold sites under high negative selection for queen-biased genes compared to worker-biased genes (permutation test, $p < 1e^{-16}$). The same pattern is observed for low expression genes, where we observe stronger negative selection in queen-biased genes compared to worker-biased genes (permutation test, $p < 1e^{-16}$). These results confirm that the difference we observe in negative selection in queen-biased genes compared to worker-biased

genes is not confounded by expression levels in these castes. Hence, the differences we observe can be attributed mainly to the genes being expressed in the different castes.

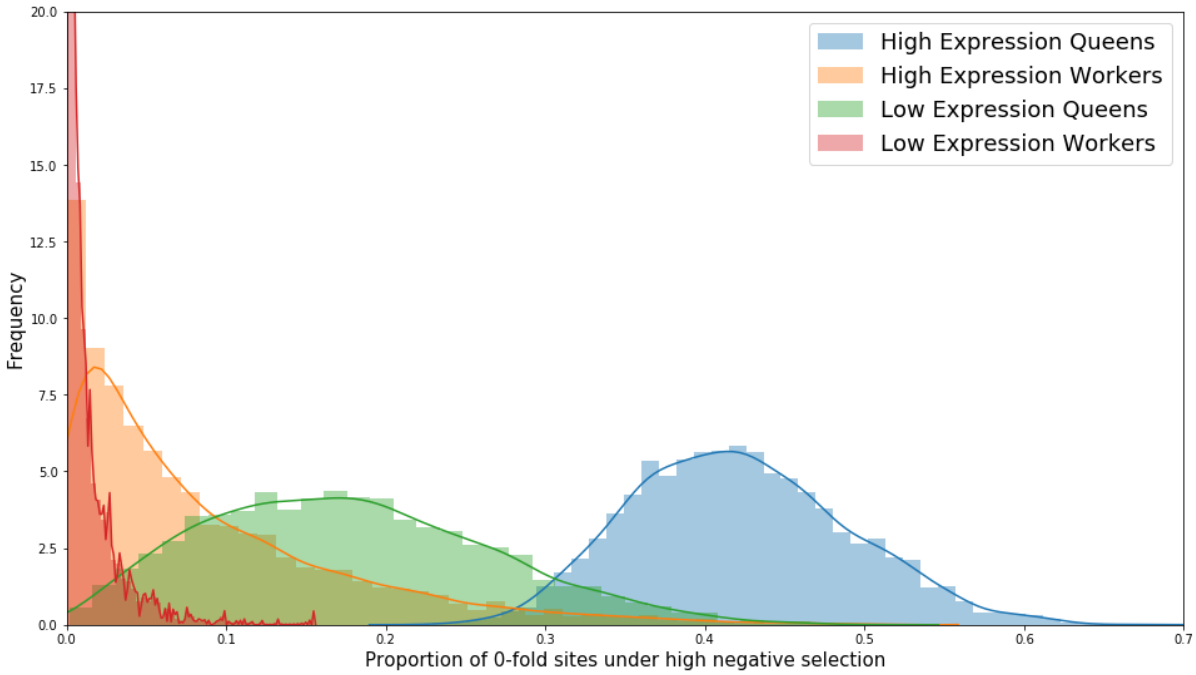

Figure 2.2: Estimates of negative selection for missense (0-fold) sites of high expression and low expression queen-biased genes and worker-biased genes under high N_es (>100) in the genome of *A. mellifera*. Data was obtained for 49 *Apis mellifera* workers. Sites were annotated using SnpEff and gene-biased castes were obtained from RNASeq data. DFE-alpha was run on missense sites from these caste-biased genes, with 4-fold degenerate sites used as the neutral selection sites. Sites were bootstrapped 5000 times to obtain the distributions. The top 200 genes highly expressed and the top 200 genes lowly expressed in both queen-biased genes and worker-biased genes were taken for this analysis. We see stronger negative selection in queen-biased genes compared to worker-biased genes, irrespective of expression level of these genes.

4. Discussion

We investigated the difference in negative selection between queen-biased genes and worker-biased genes for *A. mellifera*, *B. impatiens*, and *P. dominula*. Our results indicate higher negative selection in queen-biased genes compared to worker-biased genes for *A. mellifera*, with no difference between the two castes for both *B. impatiens* and *P. dominula*.

Compared to *A. mellifera*, *B. impatiens* and *P. dominula* are primitively eusocial (Dogantzis et al., 2018) and have similar life histories. For example, they live in small colonies and tasks are not confined to specific castes. Moreover, these 2 species show similar levels of positive selection, suggesting they are similar to each other (Dogantzis et al., 2018). In this study, I found no significant difference between queen-biased genes and worker-biased genes for both *B. impatiens* and *P. dominula* and this may be explained by the reproductive roles of *B. impatiens* and *P. dominula*. In primitively eusocial insects, the female reproductive task is not confined to queens. In fact, some workers can reproduce in *B. impatiens* and *P. dominula* in certain cases (Cnaani et al., 2002; Foster, et al., 2004; Strassmann et al., 2004). Since reproductive workers are passing on their genes to the next generation, the worker's caste-biased genes are affecting colony fitness directly. This contrasts with the indirect link to social fitness for the non-reproductive worker's caste-biased genes in *A. mellifera*. Following Linksvayer and Wade's (2009) hypothesis, reproductive worker's caste-biased genes that affect social fitness directly should experience stronger negative selection and this is indeed what we observe for *B. impatiens* and *P. dominula*.

Comparing our results to Dogantzis et al's (2008), where they studied positive selection (in the same three species studied herein), shows interesting trends. Dogantzis et al. showed that *P. dominula* and *B. impatiens* have stronger positive selection on queen biased genes, as well as similar patterns in adaptive evolution. My results complement their study, where we observe stronger negative selection in queen biased genes of *B. impatiens* and *P. dominula*, which can be again attributed to larger effective population sizes. Moreover, species that showed strong positive selection in Dogantzis' et al. study also exhibit signals of strong negative selection in this study. Both species are independently founding and species with similar life histories may be experiencing similar levels of selection, whether positive or negative.

We showed that *A. mellifera* queens experience stonger negative selection compared to workers, and there is no difference in negative selection between workers and queens of *B. impatiens* and *P. dominula*. In the future, with more RNAseq data and better gene sequences, we hope to consolidate our results and get better estimates for *B. impatiens* and *P. dominula*.

Chapter 3: Negative Selection on Drones trait versus Queens traits in Social Insects

1. Introduction

Apis mellifera are hymenopteran social insects with a haplodiploid system. Here, a major proportion of the population is diploid (queens and workers) and the rest (males) is haploid (Grimaldi and Engel, 2005). For this part of the study we are considering only *A. mellifera scutellata*.

When considering the reproductive castes in *A. mellifera*, drones are haploid males while the queen is a diploid female. Morphologically, the queen is longer than the drones but the drones are typically more stout (Koeniger, Koeniger and Phiancharoen, 2011), being larger than the average worker but smaller than queens. Drones have bigger eyes than queens or workers, where some studies suggest larger eyes in drones help them identify the queens and proceed to mating (Menzel, Wunderer and Stavenga, 1991). As mating occurs during flight, drones have musculature adapted to fast flight pursuits, possibly to help get to the queen faster and mate with her (Radloff, Hepburn and Koeniger, 2003). Drones typically do not help in taking care of the colony as their main role is to mate with queens. Drones from multiple colonies assemble in the congregational areas and wait for the queen (Zmarlicki and Morse, 1963). Once they spot a queen, they will fly to her to mate. During the process, the drones' abdomen is ripped off due to the morphology of their penis and abdomen, and the drone will die shortly after. In this respect, drones do not collect pollen or nectar. Additionally, drones do not have stingers and hence cannot defend the colony in case of attacks or intrusion. Queens are reproductive and similar to drones; they do not generally take care of the colony. In contrast to drones however, queens do have stingers.

Being a haplodiploid system, drones have only one set of chromosomes while queens have two sets of chromosomes. This haplodiploid system offers an interesting way to look at selection on deleterious allele. In diploids, a recessive allele's effect can be masked, whereas this is not the case in haploids (Kondrashov and Crow, 1991; Orr and Otto, 1994; Gerstein and Otto, 2009). We will consider a case where there is a recessive deleterious allele on one chromosome in the queen. If the other chromosome has a wild type non-deleterious allele at the same locus as the deleterious allele, this wild type allele can mask the recessive deleterious allele. For drones, however, recessive deleterious alleles cannot be masked as only one copy is present. If deleterious mutations were able to accumulate in drones, their genetic load, which is defined as the relative reduced fitness compared to the optimum fitness of the population (Klekowski, 1988), would increase and negatively impact their fitness, compared to queens. As drones are more susceptible to deleterious mutations, there would be stronger negative selection on deleterious mutations in genes that are mostly expressed in males, leading to lower alternate allele frequencies in drone-biased genes.

Our hypothesis is that drone-biased genes would show stronger negative selection compared to queen-biased genes, and this would lead to less genetic diversity in drone-biased genes. The null hypothesis is that there will be no difference in strength of negative selection when comparing drone-biased genes and queen-biased genes.

2. Materials and Methods

1. Data and samples

See Chapter 1: Section 2

2. Differentially expressed genes

We obtained RNAseq data for A. mellifera from Ashby et al. (2016) and He et al. (2017). We analyzed the RNAseq data using DESeq2 (Love et al., 2014) to get drone-biased

genes and queen-biased genes. In order to obtain estimates of the distribution of fitness effects, the VCF files were divided into drone and queen castes consisting of only regions with the respective caste-biased genes. For comparison, only missense sites (0-fold sites) under high negative selection (N_es > 100) were analyzed since these are the regions showing the largest and noticeable changes in the genome.

3. Results

We expected stronger negative selection in drone-biased genes relative to queen-biased genes due to haplodiploidy and how selection affects each ploidy (Grimaldi and Engel, 2005).

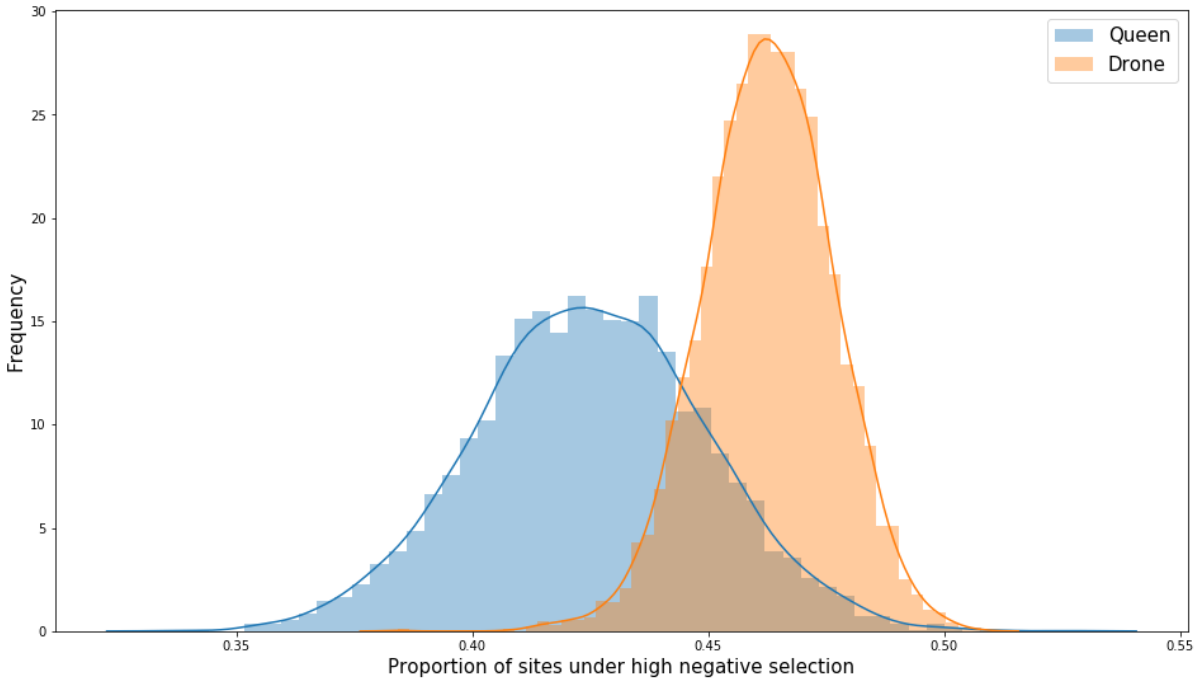

Figure 3.1: Estimates of negative selection for missense (0-fold) sites of queen-biased genes and drone-biased genes under high N_es (>100) in the genome of *A. mellifera*. Data was obtained for 49 *Apis mellifera* workers. Sites were annotated using SnpEff and gene-biased castes were obtained from RNASeq data. DFE-alpha was run on missense sites from these caste-biased genes, with 4-fold degenerate sites used as the neutral selection sites. The error bars represent the 95% bootstrapped confidence intervals. There is stronger negative selection in drone-biased genes compared to worker-biased genes ($p<1e^{-16}$).

We observed a mean of 46% of 0-fold sites in drone-biased genes are under strong negative selection (N_es > 100), compared to 42% of sites in queen-biased genes. The

differences are subtle but statistically significant (Kalmogorov-Smirnov test, $p<1e^{-16}$), supporting our hypothesis.

4. Discussion

We expected stronger negative selection in drone-biased genes compared to queen-biased genes, with our results supporting the masking hypothesis in haplodiploids systems (Orr and Otto, 1994).

In *A. mellifera*, haploid drones and diploid queens are the only reproductive individuals in the colonies (Brutscher, Baer and Niño, 2019). If the queen has a recessive deleterious allele on one chromosome, the allele's effect can be masked if the second chromosome has a dominant wild-type allele (Kondrashov and Crow, 1991). For haploid drones, with just one set of chromosomes, there is a 50% chance of having a functional allele and a 50% chance of having a recessive deleterious allele. Hence the masking effect is not possible in drones and the effects of deleterious alleles would be more prominent in drones compared to queens. Negative selection reduces the frequency of deleterious alleles in the population. As we observe stronger negative selection in drones relative to queens, deleterious alleles are purged more efficiently in these haploid drones. This phenomenon is observed in plants as well. Haploid pollens and pollen genes experience stronger negative selection relative to other diploid parts of the plants (Gerstein and Otto, 2009; Arunkumar *et al.*, 2013). Hence, there is support for the fact that haploid systems experience stronger negative selection compared to diploids.

We showed that drone-biased genes experienced stronger negative selection compared to queen-biased genes and we suggest that it is due to drones being haploid. With more caste biased gene data in the future, we hope to better clarify these results and support the haploid selection hypothesis.

Conclusion

The evolution of eusociality has been studied intensively over the last decade, with the honey bee used as a model organism. These studies focused on positive selection and adaptive selection in eusocial insects. Our study looked at negative selection in *A. mellifera*, *B. impatiens* and *P. dominula*, with interesting findings. Firstly, we see strong negative selection in protein coding regions (missense sites), moderate negative selection in regulatory regions (UTR) and no negative selection in intronic and intergenic regions, in all 3 species. Secondly, we observed stronger negative selection in queens compared to workers in *A. mellifera* but no difference in *B. impatiens* and *P. dominula*, and we attribute this difference to reproductive capabilities of the different castes. Thirdly, we see stronger negative selection in drones compared to queens, suggesting that haploids experience stronger efficient negative selection. Lastly, we observe the trend that with increasing social complexity, there is weaker negative selection, meaning negative selection is less efficient. These results confirm several hypotheses we tested and further elucidates the evolution of eusocial insects.

With better genomic data and concurrent analyses on negative selection, positive selection, and adaptive evolution, we will be able to test the hypothesis that a relaxation of constraint is important for adaptive evolution (Hunt *et al.*, 2011). Hunt et al. suggest that relaxation of constraint allows more genetic variants into the population that may then accelerate adaptive evolution in the future. Hence, in different environments, some alleles may have positive effects instead of negative effects in the population. Moreoever, knowing the distribution effects of mutation can also help in conservation. For example, they shed light on the deleterious genetic load that is segregating within natural populations, which can help us understand how inbreeding and small effective population size can influence extinction in small populations (Frankham and Ralls, 1998).

Bibliography

Akashi, H., Osada, N. and Ohta, T. (2012) 'Weak Selection and Protein Evolution', *Genetics*, 192(1), pp. 15–31. doi: 10.1534/genetics.112.140178.

Alshalalfa, M. (2012) 'MicroRNA Response Elements-Mediated miRNA-miRNA Interactions in Prostate Cancer', *Advances in Bioinformatics*, 2012, pp. 1–10. doi: 10.1155/2012/839837.

Amdam, G. V. and Omholt, S. W. (2003) 'The hive bee to forager transition in honeybee colonies: the double repressor hypothesis', *Journal of Theoretical Biology*, 223(4), pp. 451–464. doi: 10.1016/S0022-5193(03)00121-8.

Amsalem, E. *et al.* (2015) 'Conservation and modification of genetic and physiological toolkits underpinning diapause in bumble bee queens', *Molecular Ecology*, 24(22), pp. 5596–5615. doi: 10.1111/mec.13410.

Araujo, P. R. *et al.* (2012) 'Before It Gets Started: Regulating Translation at the 5′ UTR', *Comparative and Functional Genomics*, 2012, pp. 1–8. doi: 10.1155/2012/475731.

Arunkumar, R. *et al.* (2013) 'Pollen-Specific , but Not Sperm-Specific , Genes Show Stronger Purifying Selection and Higher Rates of Positive Selection Than Sporophytic Genes in Capsella grandiflora', *Molecular Biology and Evolution*, 30(11), pp. 2475–2486. doi: 10.1093/molbev/mst149.

Ashby, R. *et al.* (2016) 'MicroRNAs in Honey Bee Caste Determination', *Scientific Reports*, 6(1), p. 18794. doi: 10.1038/srep18794.

Barchuk, A. R. *et al.* (2007) 'Molecular determinants of caste differentiation in the highly eusocial honeybee Apis mellifera', *BMC Developmental Biology*, 7(1), p. 70. doi: 10.1186/1471-213X-7-70.

Barrett, L. W., Fletcher, S. and Wilton, S. D. (2012) 'Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements', *Cellular and Molecular Life Sciences*, 69(21), pp. 3613–3634. doi: 10.1007/s00018-012-0990-9.

Bicknell, A. A. *et al.* (2012) 'Introns in UTRs: Why we should stop ignoring them', *BioEssays*, 34(12), pp. 1025–1034. doi: 10.1002/bies.201200073.

Boyko, A. R. *et al.* (2008) 'Assessing the Evolutionary Impact of Amino Acid Mutations in the Human Genome', *PLoS Genetics*. Edited by M. H. Schierup, 4(5), p. e1000083. doi: 10.1371/journal.pgen.1000083.

Brutscher, L. M., Baer, B. and Niño, E. L. (2019) 'Putative Drone Copulation Factors Regulating Honey Bee (Apis mellifera) Queen Reproduction and Health: A Review', *Insects*, 10(1), p. 8. doi: 10.3390/insects10010008.

Buttstedt, A. *et al.* (2018) 'How Honeybees Defy Gravity with Royal Jelly to Raise Queens', *Current Biology*, 28(7), pp. 1095-1100.e3. doi: 10.1016/j.cub.2018.02.022.

Cardinal, S. and Danforth, B. N. (2011) 'The Antiquity and Evolutionary History of Social Behavior in Bees', *PLoS ONE*. Edited by C. S. Moreau, 6(6), p. e21086. doi: 10.1371/journal.pone.0021086.

Chandrasekaran, S. *et al.* (2011) 'Behavior-specific changes in transcriptional modules lead to distinct and predictable neurogenomic states', *Proceedings of the National Academy of Sciences*, 108(44), pp. 18020–18025. doi: 10.1073/pnas.1114093108.

Chun, S. and Fay, J. C. (2009) 'Identification of deleterious mutations within three human genomes', *Genome Research*. Edited by M. H. Schierup, 19(9), pp. 1553–1561. doi: 10.1101/gr.092619.109.

Cingolani, P. *et al.* (2012) 'A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff', *Fly*, 6(2), pp. 80–92. doi: 10.4161/fly.19695.

Cnaani, J., Schmid-Hempel, R. and Schmidt, J. O. (2002) 'Colony development, larval development and worker reproduction in Bombus impatiens Cresson', *Insectes Sociaux*, 49(2), pp. 164–170. doi: 10.1007/s00040-002-8297-8.

Consortium, E. P. (2007) 'Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project', *Nature*, 447(7146), pp. 799–816. doi: 10.1038/nature05874.

Consortium, E. P. (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489(7414), pp. 57–74. doi: 10.1038/nature11247.

Crick, F. H. C. (1968) 'The origin of the genetic code', *Journal of Molecular Biology*, 38(3), pp. 367–379. doi: 10.1016/0022-2836(68)90392-6.

CROZIER, R. H. (1976) 'Counter-intuitive property of effective population size', *Nature*, 262(5567), p. 384. doi: 10.1038/262384a0.

Dapporto, L., Turillazzi, S. and Palagi, E. (2006) 'Dominance interactions in young adult paper wasp (Polistes dominulus) foundresses: A playlike behavior?', *Journal of Comparative Psychology*, 120(4), pp. 394–400. doi: 10.1037/0735-7036.120.4.394.

Djebali, S. *et al.* (2012) 'Landscape of transcription in human cells', *Nature*, 489(7414), pp. 101–108. doi: 10.1038/nature11233.

Dogantzis, K. A. *et al.* (2018) 'Insects with similar social complexity show convergent patterns of adaptive molecular evolution', *Scientific Reports*, 8(1), p. 10388. doi: 10.1038/s41598-018-28489-5.

Eberhard, M. J. W. (1969) 'The social biology of polistine wasps.' University of Michigan Museum of Zoology.

Elyashiv, E. *et al.* (2016) 'A Genomic Map of the Effects of Linked Selection in Drosophila', *PLOS Genetics*. Edited by N. H. Barton, 12(8), p. e1006130. doi: 10.1371/journal.pgen.1006130.

Engel, M. S. (1999) 'The Taxonomy of Recent and Fossil Honey Bees (Hymenoptera: Apidae; Apis)', *Journal of Hymenoptera research*, 8, pp. 165–196. doi: http://hdl.handle.net/1808/16476.

Eyre-Walker, A. (2002) 'Changing effective population size and the McDonald-Kreitman test.', *Genetics*, 162(4), pp. 2017–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12524367.

Eyre-Walker, A. and Keightley, P. D. (2007) ‘The distribution of fitness effects of new mutations’, *Nature Reviews Genetics*, 8(8), pp. 610–618. doi: 10.1038/nrg2146.

Foster, R. L. *et al.* (2004) ‘Reproductive physiology, dominance interactions, and division of labour among bumble bee workers’, *Physiological Entomology*, 29(4), pp. 327–334. doi: 10.1111/j.0307-6962.2004.00388.x.

Frankham, R. and Ralls, K. (1998) ‘Inbreeding leads to extinction’, *Nature*, 392(6675), pp. 441–442. doi: 10.1038/33022.

Galtier, N. (2016) ‘Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis’, *PLOS Genetics*. Edited by M. H. Schierup, 12(1), p. e1005774. doi: 10.1371/journal.pgen.1005774.

Gerstein, A. C. and Otto, S. P. (2009) ‘Ploidy and the Causes of Genomic Evolution’, *Journal of Heredity*, 100(5), pp. 571–581. doi: 10.1093/jhered/esp057.

Gilbert, S. F. (no date) ‘发育生物学 Developmental Biology’, p. 60.

Gossmann, T. I., Keightley, P. D. and Eyre-Walker, A. (2012) ‘The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes.’, *Genome biology and evolution*, 4(5), pp. 658–67. doi: 10.1093/gbe/evs027.

Gould, J. L. and Gould, C. G. (1995) *The Honey Bee*. Scientific American Library (Scientific American Library series). Available at: https://books.google.ca/books?id=TgGOHAAACAAJ.

Grimaldi, D. and Engel, M. S. (2005) *Evolution of insects*. Cambridge University Press.

Halliburton, R. (2004) *Introduction to Population Genetics*. Pearson/Prentice Hall. Available at: https://books.google.ca/books?id=loDwAAAAMAAJ.

Harpur, B. A. *et al.* (2017) ‘Queens and Workers Contribute Differently to Adaptive Evolution in Bumble Bees and Honey Bees’, *Genome Biology and Evolution*, 9(9), pp. 2395–2402. doi: 10.1093/gbe/evx182.

Harris, K. and Nielsen, R. (2016) ‘The Genetic Cost of Neanderthal Introgression’, *Genetics*, 203(2), pp. 881–891. doi: 10.1534/genetics.116.186890.

He, X. *et al.* (2019) ‘A comparison of honeybee (Apis mellifera) queen, worker and drone larvae by RNA-Seq’, *Insect Science*, 26(3), pp. 499–509. doi: 10.1111/1744-7917.12557.

Hocherl, N. and Tautz, J. (2015) ‘Nesting behavior of the paper wasp Polistes dominula in Central Europe — a flexible system for expanding into new areas’, *Ecosphere*, 6(12), p. 262.

Hölldobler, B. and Wilson, E. O. (1990) *The ants*. Harvard University Press.

Hunt, B. G. *et al.* (2011) ‘Relaxed selection is a precursor to the evolution of phenotypic plasticity’, *Proceedings of the National Academy of Sciences of the United States of America*, 108(38), pp. 15936–15941. doi: 10.1073/pnas.1104825108.

Hunt, J. H. (2007) *The Evolution of Social Wasps*. Oxford University Press. doi:

10.1093/acprof:oso/9780195307979.001.0001.

Izzo, A. S. and Tibbetts, E. A. (2012) ‘Spotting the top male: sexually selected signals in male Polistes dominulus wasps’, *Animal Behaviour*, 83(3), pp. 839–845. doi: 10.1016/j.anbehav.2012.01.005.

Jackson, B. C., Campos, J. L. and Zeng, K. (2015) ‘The effects of purifying selection on patterns of genetic differentiation between Drosophila melanogaster populations’, *Heredity*, 114(2), pp. 163–174. doi: 10.1038/hdy.2014.80.

Jancek, S. *et al.* (2013) ‘Adaptive Selection on Bracovirus Genomes Drives the Specialization of Cotesia Parasitoid Wasps’, *PLoS ONE*. Edited by B. J. Mans, 8(5), p. e64432. doi: 10.1371/journal.pone.0064432.

Jandt, J. M. and Dornhaus, A. (2009) ‘Spatial organization and division of labour in the bumblebee Bombus impatiens’, *Animal Behaviour*, 77(3), pp. 641–651. doi: 10.1016/j.anbehav.2008.11.019.

Jandt, J. M., Tibbetts, E. A. and Toth, A. L. (2014) ‘Polistes paper wasps: a model genus for the study of social dominance hierarchies’, *Insectes Sociaux*, 61(1), pp. 11–27. doi: 10.1007/s00040-013-0328-0.

Kapheim, K. M. *et al.* (2015) ‘Genomic signatures of evolutionary transitions from solitary to group living’, *Science*, 348(6239), pp. 1139–1143. doi: 10.1126/science.aaa4788.

Keller, L. and Chapuisat, M. (1999) ‘Cooperation among Selfish Individuals in Insect Societies’, *BioScience*, 49(11), pp. 899–909. doi: 10.2307/1313649.

Kent, C. F. *et al.* (2012) ‘Recombination is associated with the evolution of genome structure and worker behavior in honey bees’, *Proceedings of the National Academy of Sciences*, 109(44), pp. 18012–18017. doi: 10.1073/pnas.1208094109.

Klekowski, E. (1988) ‘Genetic load and its causes in long-lived plants’, *Trees*, 2(4). doi: 10.1007/BF00202374.

Koeniger, G., Koeniger, N. and Phiancharoen, M. (2011) ‘Comparative Reproductive Biology of Honeybees’, in Hepburn, H. R. and Radloff, S. E. (eds) *Honeybees of Asia*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 159–206. doi: 10.1007/978-3-642-16422-4_8.

Kondrashov, A. S. and Crow, J. F. (1991) ‘Haploidy or diploidy: which is better?’, *Nature*, 351(6324), pp. 314–315. doi: 10.1038/351314a0.

Kozak, M. (2008) ‘Faulty old ideas about translational regulation paved the way for current confusion about how microRNAs function’, *Gene*, 423(2), pp. 108–115. doi: 10.1016/j.gene.2008.07.013.

Lafon-Placette, C. *et al.* (2018) ‘Changes in the epigenome and transcriptome of the poplar shoot apical meristem in response to water availability affect preferentially hormone pathways’, *Journal of Experimental Botany*, 69(3), pp. 537–551. doi: 10.1093/jxb/erx409.

Lagerkvist, U. (1978) ‘“Two out of three”: an alternative method for codon reading.’, *Proceedings of the National Academy of Sciences*, 75(4), pp. 1759–1762. doi: 10.1073/pnas.75.4.1759.

Laverty, T. M. and Harder, L. D. (1988) 'THE BUMBLE BEES OF EASTERN CANADA', *The Canadian Entomologist*, 120(11), pp. 965–987. doi: 10.4039/Ent120965-11.

Liebert, A. E. *et al.* (2010) 'Sexual interactions and nestmate recognition in invasive populations of Polistes dominulus wasps', *Insectes Sociaux*, 57(4), pp. 457–463. doi: 10.1007/s00040-010-0105-2.

Liebig, J., Monnin, T. and Turillazzi, S. (2005) 'Direct assessment of queen quality and lack of worker suppression in a paper wasp', *Proceedings of the Royal Society B: Biological Sciences*, 272(1570), pp. 1339–1344. doi: 10.1098/rspb.2005.3073.

Linksvayer, T. A. and Wade, M. J. (2009) 'GENES WITH SOCIAL EFFECTS ARE EXPECTED TO HARBOR MORE SEQUENCE VARIATION WITHIN AND BETWEEN SPECIES', *Evolution*, 63(7), pp. 1685–1696. doi: 10.1111/j.1558-5646.2009.00670.x.

Loewe, L. (2008) 'Negative Selection', *Nature Education*, 1(1), p. 59.

Loewe, L. and Hill, W. G. (2010) 'The population genetics of mutations: good, bad and indifferent', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1544), pp. 1153–1167. doi: 10.1098/rstb.2009.0317.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Lynch, M. and Conery, J. S. (2003) 'The Origins of Genome Complexity', *Science*, 302(5649), pp. 1401–1404. doi: 10.1126/science.1089370.

Mack, G. S. (2007) 'MicroRNA gets down to business', *Nature Biotechnology*, 25(6), pp. 631–638. doi: 10.1038/nbt0607-631.

McAulay, M. K., Otis, G. W. and Gradish, A. E. (2015) 'Honeypot visitation enables scent learning and heightens forager response in bumblebees (Bombus impatiens)', *Learning and Motivation*, 50, pp. 22–31. doi: 10.1016/j.lmot.2014.11.001.

Menzel, J. G., Wunderer, H. and Stavenga, D. G. (1991) 'Functional morphology of the divided compound eye of the honeybee drone (Apis mellifera)', *Tissue and Cell*, 23(4), pp. 525–535. doi: 10.1016/0040-8166(91)90010-Q.

Michener, C. D. (1969) 'Comparative Social Behavior of Bees', *Annual Review of Entomology*, 14(1), pp. 299–342. doi: 10.1146/annurev.en.14.010169.001503.

Michener, C. D. (1974) *The Social Behavior of the Bees: A Comparative Study*. Belknap Press of Harvard University Press (Belknap Press Series). Available at: https://books.google.ca/books?id=YjggAQAAMAAJ.

Minde, D. P. *et al.* (2011) 'Messing up disorder: how do missense mutations in the tumor suppressor protein APC lead to cancer?', *Molecular Cancer*, 10(1), p. 101. doi: 10.1186/1476-4598-10-101.

O'Donnell, S. (1996) 'Reproductive potential and division of labor in wasps: are queen and worker behavior alternative strategies?', *Ethology Ecology & Evolution*, 8(3), pp. 305–308. doi: 10.1080/08927014.1996.9522921.

Orr, A. H. and Otto, S. P. (1994) 'Does Diploidy Increase', *Genetics Society of America*, 136, pp. 1475–1480.

Pamilo, P. and Crozier, R. H. (1997) 'Population biology of social insect conservation', *Memoirs of the Museum of Victoria*, 56(2), pp. 411–419. doi: 10.24199/j.mmv.1997.56.32.

Penny, D. *et al.* (2009) 'An Overview of the Introns-First Theory', *Journal of Molecular Evolution*, 69(5), pp. 527–540. doi: 10.1007/s00239-009-9279-5.

Peters, R. S. *et al.* (2017) 'Evolutionary History of the Hymenoptera', *Current Biology*, 27(7), pp. 1013–1018. doi: 10.1016/j.cub.2017.01.027.

Pichon, X. *et al.* (2012) 'RNA Binding Protein/RNA Element Interactions and the Control of Translation', *Current Protein & Peptide Science*, 13(4), pp. 294–304. doi: 10.2174/138920312801619475.

Pirk, C. W. W. *et al.* (2004) 'Egg viability and worker policing in honey bees', *Proceedings of the National Academy of Sciences*, 101(23), pp. 8649–8651. doi: 10.1073/pnas.0402506101.

Proudfoot, N. J., Furger, A. and Dye, M. J. (2002) 'Integrating mRNA Processing with Transcription', *Cell*, 108(4), pp. 501–512. doi: 10.1016/S0092-8674(02)00617-7.

Queller, D. C. *et al.* (2000) 'Unrelated helpers in a social insect', *Nature*, 405(6788), pp. 784–787. doi: 10.1038/35015552.

Radloff, S. E., Hepburn, H. R. and Koeniger, G. (2003) 'Comparison of flight design of Asian honeybee drones', *Apidologie*, 34(4), pp. 353–358. doi: 10.1051/apido:2003031.

Rana, T. M. (2007) 'Illuminating the silence: understanding the structure and function of small RNAs', *Nature Reviews Molecular Cell Biology*, 8(1), pp. 23–36. doi: 10.1038/nrm2085.

Ratnieks, F. L. W. and Helanterä, H. (2009) 'The evolution of extreme altruism and inequality in insect societies', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533), pp. 3169–3179. doi: 10.1098/rstb.2009.0129.

Rearick, D. *et al.* (2011) 'Critical association of ncRNA with introns', *Nucleic Acids Research*, 39(6), pp. 2357–2366. doi: 10.1093/nar/gkq1080.

Reeve, H. K. (1991) *'Polistes', The Social Biology of Wasps*. Cornell University Press.

Riley, J. R. *et al.* (2005) 'The flight paths of honeybees recruited by the waggle dance', *Nature*, 435(7039), pp. 205–207. doi: 10.1038/nature03526.

Romiguier, J. *et al.* (2014) 'Population genomics of eusocial insects: the costs of a vertebrate-like effective population size', *Journal of Evolutionary Biology*, 27(3), pp. 593–603. doi: 10.1111/jeb.12331.

Roux, J. *et al.* (2014) 'Patterns of Positive Selection in Seven Ant Genomes', *Molecular Biology and Evolution*, 31(7), pp. 1661–1685. doi: 10.1093/molbev/msu141.

Sammeth, M., Foissac, S. and Guigó, R. (2008) 'A General Definition and

Nomenclature for Alternative Splicing Events', *PLoS Computational Biology*. Edited by M. R. Brent, 4(8), p. e1000147. doi: 10.1371/journal.pcbi.1000147.

Schmidt, S. *et al.* (2015) 'Expression, sorting and transport studies for the orphan carrier SLC10A4 in neuronal and non-neuronal cell lines and in Xenopus laevis oocytes.', *BMC neuroscience*, 16, p. 35. doi: 10.1186/s12868-015-0174-2.

Standage, D. S. *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(8), pp. 1769–1784. doi: 10.1111/mec.13578.

Strassmann, J. E. *et al.* (2004) 'The Cost of Queen Loss in the Social Wasp Polistes dominulus (Hymenoptera: Vespidae)', *Journal of the Kansas Entomological Society*, 77(4), pp. 343–355. doi: 10.2317/E-15.1.

Theraulaz, G., Pratte, M. and Gervet, J. (1989) 'Effects of removal of α -individuals from a Polistes dominulus Christ wasp society: Changes in behavioural patterns resulting from hierarchical changes', *Actes Coll. Insectes Soc.*, 5, pp. 169–179.

Tibbetts, E. A. (2003) 'Benefits of foundress associations in the paper wasp Polistes dominulus: increased productivity and survival, but no assurance of fitness returns', *Behavioral Ecology*, 14(4), pp. 510–514. doi: 10.1093/beheco/arg037.

Waldbauer, G. (1998) *The Birder's Bug Book*. Harvard University Press. Available at: https://books.google.ca/books?id=97XawAEACAAJ.

Wethmar, K., Smink, J. J. and Leutz, A. (2010) 'Upstream open reading frames: Molecular switches in (patho)physiology', *BioEssays*, 32(10), pp. 885–893. doi: 10.1002/bies.201000037.

Whitfield, C. W. (2003) 'Gene Expression Profiles in the Brain Predict Behavior in Individual Honey Bees', *Science*, 302(5643), pp. 296–299. doi: 10.1126/science.1086807.

Williamson, R. J. *et al.* (2014) 'Evidence for Widespread Positive and Negative Selection in Coding and Conserved Noncoding Regions of Capsella grandiflora', *PLoS Genetics*. Edited by M. W. Nachman, 10(9), p. e1004622. doi: 10.1371/journal.pgen.1004622.

Wolfe, S. L. (1993) *Molecular and Cellular Biology*. Wadsworth Publishing Company (Biology Series). Available at: https://books.google.ca/books?id=0rZqAAAAMAAJ.

Woodard, S. H. *et al.* (2014) 'Molecular heterochrony and the evolution of sociality in bumblebees (Bombus terrestris)', *Proceedings of the Royal Society B: Biological Sciences*, 281(1780), p. 20132419. doi: 10.1098/rspb.2013.2419.

Wright, S. I. and Andolfatto, P. (2008) 'The Impact of Natural Selection on the Genome: Emerging Patterns in Drosophila and Arabidopsis', *Annual Review of Ecology, Evolution, and Systematics*, 39(1), pp. 193–213. doi: 10.1146/annurev.ecolsys.39.110707.173342.

Zhen, Y. and Andolfatto, P. (2012) 'Methods to Detect Selection on Noncoding DNA', in, pp. 141–159. doi: 10.1007/978-1-61779-585-5_6.

Zhu, Y. (2006) 'Sequence analysis of RNase MRP RNA reveals its origination from

eukaryotic RNase P RNA', *RNA*, 12(5), pp. 699–706. doi: 10.1261/rna.2284906.

Zmarlicki, C. and Morse, R. A. (1963) 'Drone Congregation Areas', *Journal of Apicultural Research*, 2(1), pp. 64–66. doi: 10.1080/00218839.1963.11100059.