2-18-2021 11:30 AM

A genetic perspective on social insect castes: A synthetic review and empirical study

Anna M. Chernyshova, *The University of Western Ontario*

Supervisor: Thompson, Graham J, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biology

ABSTRACT

The process of caste differentiation is central to understanding insect sociality because it is castes that enable division of labor. Presumably selection favors colonies that can divide labor in response to environmental demands, and for many taxa genetic factors are an important part of this equation. In my thesis, I first provide a framework for understanding genetic and epigenetic effects on caste. From mostly ant, bee and termite examples, I make clear that genotype-caste associations can evolve in different and sometimes complex ways and can involve additive or non-additive genetic effects that, in turn, may arise directly from focal individuals or indirectly via their social partners. I use this framework to launch an empirical analysis of my own. In my second chapter, I test alternative hypotheses that describe how genes evolve under direct vs. indirect selection. I predict that genes associated with reproductive castes will evolve mostly under direct selection and show patterns of nucleotide substitution that differ from those associated with non-reproductive helper castes and thus evolving under indirect selection. Using an RNA-Seq dataset for the Eastern subterranean termite, I found that caste-biased and un-biased genes evolve at similar rates, most consistent with purifying selection. I therefore did not detect an obvious pattern of molecular evolution that is diagnostic of indirect or 'kin' selection. I did discover other, more subtle patterns of nucleotide substitution that I discuss in the context of termite social biology.

SUMMARY FOR LAY AUDIENCE

Termites are eusocial insects with reproductive and non-reproductive castes, the latter of which can only evolve via indirect selection. In this thesis I first look retrospectively at the role for genes and genetic variation on the origin and maintenance of caste differences in the social insects. My synthetic review concludes that additive genetic variants must have played an essential role in the origin of castes and that gene-by-environment interactions continue to govern the conditional development of caste differences for many ant, bee, wasp and termite species. From this broader standpoint, I launch my empirical study. Here, I test if genes associated with sterile workers and soldiers in a termite have similar patterns of molecular evolution to those of reproductive castes. The Eastern subterranean termite Reticulitermes flavipes showed patterns of nucleotide substitution at most caste-associated loci consistent with purifying selection. Some soldier-associated loci did, however, differ from this general pattern and appear to be evolving close to the neutral rate. A neutral pattern is rare for highly expressed protein coding genes but is consistent with the idea that genes evolving under indirect selection are buffered from the full strength of selection and thus may approach neutrality. These two genes, encoding a titan and a feruloyl esterase-like protein, together with others identified by my analysis suggests that caste can affect the strength of selection and in a manner consistent with some prior predictions. I discuss my novel results in the context of modern sociogenomic theory and offer my own ideas on ways to further test the relationship between caste-biased gene expression and patterns of molecular evolution.

CO-AUTHORSHIP STATEMENT

Both chapters in this thesis are co-written with my supervisor Dr Graham J Thompson. I acknowledge his co-authorship on any future publications that arise. An early version of Chapter 1 is already published as cited below. My role in Chapter 1 was equal (50%) to my supervisor's: I jointly conceived, researched and wrote the review, as well as helped to edit and revise the published edition. My supervisor did take a lead role in communicating with the publication editor and it is for that reason that he is the communicating author. Chapter 2 of this thesis is not yet published but if it is, I will be the first and corresponding author. My role in Chapter 2 was to jointly conceive the project, single-handedly assemble and analyze the data, produce all of the figures and jointly write the chapter with my supervisor. Additionally, during my tenure at Western, I have co-authored several conference talks and posters with my supervisor and have co-authored other research papers with him (and others) that are not officially part of this thesis. I am therefore highly collegial and happy to contribute as first, middle or last author, as appropriate to each project. I do not anticipate any proprietary issues.

Thompson GJ, Chernyshova AM (2020) Caste differentiation: Genetic and Epigenetic Factors. pp 165-176. In: Starr, C.K. (ed.), Encyclopedia of Social Insects, Cham, Switzerland: Springer.

DEDICATION

Lyudmila, Mikhail, Evgeniy, and Sergey – my thesis is dedicated to you.

ACKNOWLEDGEMENTS

I want to recognize my MSc supervisor Dr. Graham Thompson for his excellent mentorship and continuous support throughout my study and research. I am grateful for your encouraging comments, which always have been so enlightening and propelling to my way of thinking. Thank you for this project, for broadening my perspective, and for teaching me that persistence and hard work – works. I am grateful for the many thought-provoking questions and debates that we have had and that have ultimately helped me to prepare this thesis to the best of my ability. My supervisor always encouraged me to think freely and independently, to read a lot and follow precedent but to innovate, too. I am grateful for every local and international opportunity I ventured on that allowed me to connect and exchange ideas, to think outside-the-box, and to network with people and brilliant scientists across the globe. I believe that these experiences have made a huge difference in my intellectual growth and journey to become an independent scientist.

I express my deepest appreciation to my thesis committee members Dr. Marc-André Lachance and Dr. David Smith for their encouragement, stimulating discussions, insightful comments, and challenging questions during our advisory meetings and examinations. I am especially grateful to Dr. Marc-André Lachance for the knowledge you imparted during *Evolutionary Genetics* course I had the privilege to TA, and for serving as my thesis reader. I very much appreciate your feedback. I thank Dr. Yolanda Morbey, Dr. Geoff Wild, and Dr. Jeremy McNeil as my thesis examiners and for having already raised with me many thought-provoking questions during our impromtu interactions and hallway

discussions. These interactions are not lost on me and one way or another have helped me to improve my effort and understanding.

I am endlessly grateful to Dr. Greg Gloor, Dr. Vera Tai, Dr. Art Poon, and Dr. Ryan Austin for introducing me to the art and science of coding. This project would not have moved forward without the bioinformatic skills I have learned in your courses. Your enthusiasm for the subject that back then seemed impossible to grasp, had a lasting effect – I got infected with interest and desire to solidify my knowledge in this discipline.

My thanks also extend to Dr. Amro Zayed and Dr. Brock Harpur, whose ideas were implemented in formulation of my hypotheses of Chapter 2 in this thesis. Apart from remaining an integral part of my research network, Dr. Zayed and Dr. Harpur were supervisors of my Honours thesis project at York University. I am grateful to them for their early mentorship and for connecting me with my current supervisor, Dr. Graham Thompson. My desire is to continue expanding beyond our current collaborations.

My sincere gratitude also extends to Dr. Gregor Reid and his research group (esp. Brendan Daisley and John Chmiel). It was such a great pleasure and invaluable experience to work on diverse and exciting projects, because the ideas stemming from these interactions expanded my research interests and offered new ways of thinking and of doing research. I am thus indebted to all of you – my intellectual leaders – for your innovative ideas and invaluable support in preparation of this work.

I thank Western University for the best graduate student experience and for providing me with fully equipped, multi-disciplinary platform for scientific experimentation and personal networking with the world's brightest minds. At Western, I developed outside of the lab, socially and politically, through my involvement with fellow graduate students and as elected member to both SOBGS and SOGS. I also thank my lab-mates (especially Andrew Pitek, Alex Guoth, and Kyrillos Faragalla) for all their help and discussions about this and other projects, and for being part of this great team.

To my brothers, Evgeniy and Sergey, I am thankful for the unbreakable sibling bond we share and for your unconditional love, motivation, and constant interest in my work. Most importantly, I am grateful to my parents Mikhail and Lyudmila, for all your sacrifices, investments and endless belief in me, the impossible expectations and dreams that you have for us, and for gifting me the greatest gift – my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF TERMS

ADDITIVE GENETIC EFFECTS	A quantitative inheritance mechanism that occurs when the combined effects of alleles at two or more loci are equivalent to the sum of their individual effects.
ALATE	The sexually mature developmental stage of an insect featuring wings or wing-like appendages.
ALLELE	One of the alternative forms of a gene at a chromosomal LOCUS.
ALLOPARENT	The 'other parent' or individual that performs parent-like behaviour towards other individuals that are not his or her offspring.
ALTERNATIVE SPLICING	A regulatory mechanism that generates different types of mRNAs from the same gene through variable incorporation of exons or coding regions. It results in production of related protein variants or isoforms.
ALTRUISM	In evolutionary biology, refers to reproductive ALTRUISM, which is a type of behaviour that benefits others, at a reproductive fitness cost to self, measured in terms of the expected number of offspring.
ASEXUAL QUEEN SUCCESSION (AQS) SYSTEM	A system discovered in some termite species, where queens use asexual reproduction (PARTHENOGENESIS) for generation of future queens, and sexual reproduction to produce other colony members.
AUTOMICTIC PARTHENOGENESIS	A postmeiotic process whereby a haploid cell either duplicates its chromosomes or joins with the other haploid cell, in both cases generating a diploid zygote.
AUTOSOMAL LOCUS	A LOCUS on any chromosome other than sex chromosome.
BALANCING SELECTION	Includes a number of selective pressures as a result of which alleles are maintained in population gene pool at higher frequencies than expected from genetic drift.

BLAST — Basic Local Alignment Search tool is based on a sequence comparison algorithm that is used to search the sequence databases for local alignments to a query sequence, or a sequence of interest that is being used to conduct the search.

CASTE — Subgroup of individuals within a colony that share a common, separate morphogenetic pathway. CASTES have different physical forms (i.e., soldier, queen, worker) that are usually associated with the social function performed by each individual.

CASTE DIFFERENTIATION — A biological process, whereby related individuals within a colony develop into alternative CASTES or SUB-CASTES.

CLONAL REPRODUCTION — A form of asexual reproduction that occurs through split of genetic material to generate new individuals with identical genotypes.

CLUSTAL FORMAT — A text-based file that has multiple sequences that are, or will be, aligned. The format includes information about the degree of conservation at each position of the alignment. Among the three types of symbols: "*****" indicates perfect alignment, "**:**" a site with strong similarity and "**.**" a site with low similarity. Below is an example of a fragment from one protein multiple sequence alignment for one NDE gene from this study in CLUSTAL format. The number at the end of each line is a total base count for each aligned sequence.

```
DN178023_Rf_1             VSGAVAERCNFVAYITYSAVISGFVYP 27
XP_033608955.1_Cs_1       VSGAVAERCNFVAYITYSAVISGFVYP 27
XP_023714905.1_Cs_2       VSGAVAERCNFVAYITYSAVISGFVYP 27
XP_023714897.1_Cs_3       VSGAVAERCNYIAYITYSFVISGFVYP 27
XP_021933575.1_Zn_1       VSGAVAERCNFIAYIVYSIGISGIVYP 27
                          **********::***:**.:***:***
```

COBALT — The Constraint-based Multiple Alignment Tool performs progressive multiple alignment of protein sequences and is offered by the NCBI. This tool generates robust alignments using derived constraints from protein motif database and conserved domain

	database, as well as annotated sequence similarity information from RPS-BLAST, BLASTP, and PHI-BLAST.
CODEML	A program within PAML that implements codon substitution model for DNA sequences and various models for amino acid sequences. CODEML program is used to estimate non-synonymous and synonymous substitution rates and detect selection in nucleotide and corresponding protein sequences.
CONGENERIC	Of species that belong to the same genus.
DIFFERENTIALLY EXPRESSED GENE	A gene is deemed differentially expressed if the observed change or difference in expression levels of normalized read counts is statistically significant between two experimental conditions.
d_N/d_S	Ratio that is used to measure the mode and strength of selection via comparison of synonymous substitution rates (d_S) per synonymous site with non-synonymous substitution rates (d_N) per non-synonymous site. The d_S rate is assumed to be neutral, while d_N results in a change of gene's translated amino acid composition.
DNA METHYLATION	A process by which methyl groups are added to the DNA and typically repress gene transcription. This process may therefore alter the expression or activity of the DNA molecule without changing its sequence. Out of four DNA bases, only cytosine and adenine can be methylated.
DNA METHYLTRANSFERASE	A family of enzymes, including *DNMT* and *DNA MTase*, that catalyzes the transfer of methyl groups to DNA molecules.
DOMINANCE	When one allele on a chromosome copy masks the effect of another allele for the same gene on a different chromosome copy at the same locus.
EPIGENETIC FACTORS	Factors that are 'on top of' or 'above' genetics, that modify DNA externally (i.e., DNA METHYLATION). These EPIGENETIC modifications do

not alter the DNA sequence itself but able to change gene expression by turning genes 'on' and 'off'.

EPISTASIS — The interaction between genes that affects phenotype. A gene is epistatic if its presence limits the effect of a gene at a different locus.

ERGATOID — A type of immature reproductive individuals without wing buds.

EUSOCIALITY — The most complex level of sociality organization, characterized by REPRODUCTIVE DIVISION OF LABOUR, overlapping generations, and ALLOPARENTAL brood care.

FASTA — A text-based file format used to represent either peptide or nucleotide sequences. Each sequence in FASTA format starts with a '>' symbol followed by the header containing identifying information. The header is followed by the sequence data, where single-letter codes are used to represent each nucleotide or amino acid base pair. Below is an example of one nucleotide sequence fragment in a FASTA format from this study:

```
>TRINITY_DN82301_c1_g3_i2
CCTGACAGGAGGATACTATGACGGTAAGGCTACTACAAGCACATCATGTATGTCT
```

FECUNDITY — Physiological reproductive capacity or maximum reproductive output of individual over its lifetime.

GENE FAMILY — A set of genes that arose from duplication and have similar function in a species.

GITHUB — A web-based and version-controlled collaboration system, equipped with repositories, branches and commits that facilitate data transparency. GITHUB is used for storing, tracking, and sharing any set of files or projects.

GYNES — A caste in some eusocial species that is destined to become queens. GYNES are different from workers, who are typically STERILE and do not become queens.

HAMILTON'S RULE	Asserts the conditions under which reproductive ALTRUISM evolves among relatives. The HAMILTON'S RULE is: $r \times B > C$ According to the rule, a trait is favoured by natural selection if its benefit *B* to others, when multiplied by relatedness *r*, exceeds the fitness cost *C* to self.
HEMIMETABOLOUS	Gradual type of insect development characterized by incomplete metamorphosis and lack of pupal stage.
HISTONE MODIFICATION	Post-translational covalent modification to histone proteins (i.e., DNA METHYLATION) that may influence gene expression. This occurs through modifications to the chromatin structure or via recruitment of histone modifiers.
HOLOMETABOLOUS	Type of insect development with complete metamorphosis and immature larvae stages that are different from adults. Transformation from larva to adult occurs during pupal stage.
HOMEOSTASIS	A self-regulating process that results in maintenance of a relatively stable internal state of an organism despite potential changes in the environment.
HYBRIDIZATION	The process of breeding with individuals of another species.
INDEL	Insertion-deletion mutation is one of the most common types of mutations and refers to insertion or deletion of one or more nucleotide bases in DNA/RNA sequences.
INCLUSIVE FITNESS THEORY	In evolutionary biology, also referred to as 'kin theory' or 'kin selection'. The INCLUSIVE FITNESS THEORY was proposed by W.D. Hamilton (see HAMILTON'S RULE) and models evolution of social traits or behaviours. According to the theory, individual's inclusive fitness is a sum of direct (i.e., the total number of offspring produced), and indirect fitness (i.e., the number of produced relatives) multiplied by the degree of relatedness (i.e., shared proportion of genes)

	between related individuals. One of the main tenants of INCLUSIVE FITNESS THEORY is that ALTRUISM among related organisms enables shared genes to be passed on to future generations.
INTER-SPECIFIC	Between species.
INTRA-SPECIFIC	Within species.
KIN GROUP	A group of genetic relatives that live together.
LOCUS	Genetic 'street address' or specific physical position of a gene on a chromosome.
MATRILINEAL EFFECTS	The tracing of kinship and through 'mother line' or reproductive females.
MICRO-CLIMATE	A local set of conditions or climate of a restricted area that typically differs from the surrounding environment.
NARROW-SENSE HERITABILITY	A proportion of genetic variation for a given trat that is due to ADDITIVE GENETIC EFFECTS. It is estimated as a ratio of additive genetic variance V_A to total phenotypic variance V_P. $h^2 = V_A/V_P$
NCBI	The National Center for Biotechnology Information provides access to genomic and biomedical information as part of the United States National Library of Medicine. The NCBI contains a series of databases along with relevant bioinformatic tools and services for their analysis.
NEOTENIC QUEENS	Individuals within the colony that supplement egg production and eventually replace the primary queen.
NEPOTISM	Favoritism towards relatives.
NON-ADDITIVE GENETIC EFFECTS	Describes effects that arise from interaction of different alleles at the same (i.e., DOMINANCE) or across locus (i.e., EPISTASIS).
NON-CODING RNAS	RNA molecule that does not get translated into a protein. NON-CODING RNAS may regulate gene expression at both, transcriptional and post-transcriptional levels.

NON-DIFFERENTIALLY EXPRESSED GENE	A gene is non-differentially expressed if the observed change or difference in expression levels of normalized read counts is statistically insignificant between experimental conditions.
NYMPHOID	A type of immature reproductive with wing buds.
ORF-FINDER	A bioinformatic tool offered by the NCBI that can be used manually online or piped into a script for a command line execution. This tool searches nucleotide sequences in forward and reverse directions for all six possible reading frames, specifically detecting regions flanked by START and STOP codons; then locates and extracts all potential open reading frames in a given nucleotide sequence along with corresponding amino acid translations.
ORTHOLOG	Gene that is related to another by common ancestor and encodes protein with similar functions in different species.
PAL2NAL	A program that converts paired multiple sequence alignments of nucleotide and corresponding protein sequences into codon alignments. PAL2NAL also implements PAML's CODEML program to generate d_N/d_S estimates on paired codon-by-codon alignments.
PAML	For Phylogenetic Analysis by Maximum Likelihood. PAML implements a series of programs, including CODEML to perform phylogenetic analyses of protein and nucleotide sequences using maximum likelihood.
PARTHENOGENESIS	A reproductive strategy that involves development from unfertilized ovum.
PATRILINE	The 'father line' or a descent that is exclusively established through males from a founding male ancestor.
PHEROMONE	A chemical substance that is produced and released by the animal that elicits one or more behavioural responses from other individuals of the same species.

PLEIOTROPY	When two or more phenotypic traits are influenced by a single gene, the gene is said to be pleiotropic.
PLOIDY	The number of chromosomes sets in the cells of an organism. DIPLOID organism has two sets of chromosomes and HAPLOID has one set.
POLYANDROUS	Relating to the type of mating, where one female mates with multiple males, while each male only mates once with a female.
PYTHON	An object-oriented, high-level script language that automates execution of tasks and is used to process and connect large data components.
R	A programming language and software environment for statistical data analyses, computing and graphics.
RECOMBINATION	Rearrangement of genetic material via cross-over in chromosomes.
REPRODUCTIVE DIVISION OF LABOUR	The defining feature of EUSOCIALITY, where one or a few individuals monopolize reproduction, while numerous others perform specialized and non-reproductive roles that include defense, foraging, and brood care.
ROYAL JELLY	A substance secreted and fed by honey bee workers to larvae raised as potential queens.
SEX-LINKED	A gene located on the sex chromosome (x or Y) that may therefore be inherited differently between males and females.
SMARTBLAST	NCBI's program that implements a combination of BLAST searches to generate results. This program searches query against the landmark database with BLASTP and against the non-redundant (nr) protein database with optimized version of BLAST that is specifically targeted to closely related sequences.
STERILE	Individual that is incapable of producing offspring.
SUB-CASTE	A specialized subdivision of a larger CASTE within the colony, with characterized morphological differences and functional roles.

SUB-FERTILE	Describes a form of reduced fertility, where potential for reproduction exists, but is minimal or occurs under specific conditions.
SYMPATRY	Of speciation that occurs between species that live in the same geographical area and overlap in distribution.
TAXON	Refers to any one group within the hierarchical rank of taxonomy, or biological classification of organisms from kingdom to subspecies.
THELYTOKY	A type of PARTHENOGENESIS that results in production of females only from an unfertilized ovum.
UNIPRO UGENE	Software that integrates a variety of widely used bioinformatics tools and provides visualization modules for annotation and assembly of data into multiple sequence alignments. This bioinformatic software also allows for conversion between multiple biological data formats from local and remote sources.
UNIX	An operating system with portable, multiuser and time-sharing functionality that implements a range of scripting and text-based coding. Additionally, this operating system incorporates a graphical user interface to support computing environment and file navigation.
ZYGOSITY	Characterizes DNA sequence similarity at a specific genetic locus. In DIPLOID individuals, where a copy of each allele is inherited maternally and paternally, ZYGOSITY describes whether the resulting DNA sequence is homozygous (with identical alleles) or heterozygous (with different alleles) at a locus.

CHAPTER ONE: General introduction

1.1 Goals of the thesis

The evolution of adaptively complex societies from individually selfish life histories has long been a topic of theoretical and biological interest (Ratnieks et al. 2011) – one to which studies on social insects have contributed a great deal (Williams and Williams 1957; Hamilton 1972; Lin and Michener 1972; Alexander 1974; Anderson 1984; Crozier and Pamilo 1996; Bourke and Franks 1995; Abbot et al. 2011). Central to this discussion has been the seemingly paradoxical evolution of a selfless worker caste that, through conventional understanding of how selection works, ought not to evolve at all due to seriously compromised sense of personal fitness (Darwin 1859). Thanks to a major effort by the British natural historian William D. Hamilton (Herbers 2013), we now know that fitness is not only represented through personal reproduction of descendants, but rather includes an 'indirect' fitness quotient realised through the assisted production of non-descendent kin. These direct and indirect components to a worker, queen or any individual's fitness sum up to what is now recognized as INCLUSIVE FITNESS (Levin and Grafen 2019) – the parameter that natural selection tends to maximize (Hamilton 1964).

In the following chapters, I build upon our modern understanding of the inclusive fitness theory (West and Gardner 2013; Queller 2016; Bourke 2011b) that has been fundamentally established within the field of insect sociobiology (Bourke 2011a; Alcock 2001; Marshall 2015) and contribute in novel ways by providing a study of my own. In

Chapter Two, I develop a synthetic review that describes the genetic and epigenetic effects on the evolution and development of caste systems that are found among social insects. In Chapter Three, I conduct an empirical study that implements modern bioinformatics approaches to help identify patterns of nucleotide substitution that reveal historical effects of indirect selection.

1.1.1 Literature review with novel synthesis

By conducting synthetic review in Chapter Two, I utilize the inclusive fitness theory to pose and answer outstanding questions on the role that genes, the environment and gene-by-environment interactions play in shaping the diversity of caste systems in nature. I highlight key reports from the empirical literature that reveal detailed insights into surprising, complex and sometimes bizarre biology of insect societies. Reflecting on the un-even interest and study effort toward certain eusocial taxa over others, this literature synthesis is inadvertently biased towards examples and insights from the eusocial Hymenoptera, particularly ants or other most-studied species of bees (i.e., the European honey bee *Apis mellifera*). I position these models against others for less-studied taxa, in particular termites, to provide the broadest view possible.

In Chapter Two, I introduce new ways-of-thinking that are conveyed to the reader through the use of conceptual figures and diagrams. For instance, where others have typically separated the developmental or proximate from evolutionary or ultimate factors that explain caste differences, I have joined them into a single conceptual framework

(**Figure 2.1**) that I believe is a unique contribution to my field. I use novel diagrams to clarify the relationship between direct and indirect selection (**Figure 2.2**) and between allelic (genic) and genotypic effects (**Figure 2.3**) – a perennial source of confusion in the field of sociobiology. This review chapter provides new approaches towards further resolution of complex and quirky breeding arrangements that evolve under a mix of sexual vs. asexual reproduction or under direct vs. indirect selection that is acting on both, diploid and haploid phases across ecologically diverse and taxonomically unrelated taxa. Marshalling this genetic, ecological and taxonomic diversity into a common framework and single review is challenging but at the same time, is precisely the point. By building upon meticulous summaries of other studies with new ideas, I am confident that this work is in itself a novel contribution to the field of sociobiology. Its inclusion as a long-form entry into the authoritative *Encyclopedia of Social Insects* (C. R. Starr, ed. Cham, Switzerland: Springer, 2021) shores-up current knowledge and accelerates the development of new studies from the framework provided.

1.1.2 An empirical study and test for selection

In my third empirical chapter, I build upon the context established in the synthetic review, to propose a novel test for scanning and identifying signatures of caste-mediated selection. Using an RNA-sequence dataset from a local species of subterranean termite (*Reticulitermes flavipes*; Rhinotermitidae), I develop a bioinformatics pipeline to compare the strength of selection and rates of molecular evolution across caste-biased and un-biased genes. My study is not yet published, but I believe that my effort is the first of its

kind from a social insect species in the order Blattodea. My approach exploits allelic diversity captured in the RNA sequences, which were derived from within our laboratory at Western (Wu et al. 2018), revealing differences in patterns of nucleotide substitution that are associated with functionally sterile worker and soldier castes.

The subterranean termite used in the Wu et al. (2018) study is native to eastern North America but has been introduced into other locales outside of this range (reviewed in Scaduto et al. 2012). In native range colonies, reproduction is monopolized by the royal pair – the queen and king, who are supported by thousands of soldiers and workers (Thorne et al. 1999; Lainé and Wright 2003). By contrast, invasive termite populations maintain a slightly different social structure, whereby colonies rely on immature ('neotenic') nymphs for reproduction (Vargo and Husseneder 2009). I reason that genes identified by Wu et al. (2018) as associated in their expression with the reproductive nymph caste evolve under direct selection, and thus will show patterns of nucleotide substitution that are different from those associated with the non-reproductive castes. I further reason that genes associated in their expression with the non-reproductive worker or soldier caste will show evidence of indirect selection, and thus evolve in a pattern opposite to that of nymphs. To determine these differences, I embrace three leading hypotheses and test their diagnostic predictions.

The three main hypotheses that are used to predict diagnostic signatures of indirect selection and against which I pose my data and analyses, are backed-up by literature and

graphically-depicted through the use of novel diagrams in Chapter Three. For example, I first paraphrase their predictions in text (**Table 3.1**) then, for the first time, graphically plot them as hypothetical results within the context of neutral theory (Figure 3.1). By consolidating these alternative hypotheses from their original, disparate sources into a common graphic, I summarize their commonalities and differences for the future audience. Moreover, my results from empirical analysis that provides the best test yet, reinforce their graphical value in explaining patterns of nucleotide substitution from available termite species. My study remains provisional as presented in this Thesis but does represent a major step forward towards its eventual completion. It is necessary to analyse the data in a manner that progresses theoretical ideas and workload from small to large, and to asses the next-steps with emerging findings. By doing so, in my work, it has become clear that my initial analysis of a set of ~90 genes from one species lacks sufficient statistical power to support or reject any one hypothesis, and that further analysis will be required. Thankfully, I do have access to an expanded data set with ~570 genes that I will incorporate in my extended analysis and future publication.

1.1.3 Towards general conclusion

In my fourth General Conclusions chapter, I summarise my main effort and results, as well as suggest a next generation of studies that build upon my work, for other researchers in the field. For example, I suggest that my approach to test for evidence of indirect selection with the neutral theory is powerful in principle but statistically lacks power, as expected, when sequence variation is minor for a relatively small RNA-sequence data set

from a sparsely sampled population of inbreed insects that is characterized by a small effective population size. In addition to using the standard d_N, d_S and d_N/d_S parameters to detect and measure variation in the direction and intensity of selection, I propose to augment my analysis with other measures for detecting selection, including analysis of allele frequency spectra as commonly implemented in the McDonald-Kreitman test (McDonald and Kreitman 1991) or its extension proposed by Akashi (Akashi 1999).

Finally, I am excited to introduce one new hypothesis of my own. In my analysis, I discovered another pattern in the strength of selection as measured by d_N/d_S that has emerged from my study. This detected pattern shows that the intensity of purifying selection ($d_N/d_S < 1$) on caste-associated genes co-varies with phylogeny, such that species with the most derived eusocial systems display evidence of most intense selection. My comparative phylogenetic analysis contains only three species; however, my observation is sufficient to propose a novel idea – that as species evolve to be ever-more social, and that their caste systems ever-more interdependent, selection likewise becomes more focussed on the genes involved. If so, elusive signatures of indirect selection might be more readily detected in derived as opposed to ancestral species on the termite (or other social insect) tree of life.

CHAPTER TWO: Genetic effects on the evolution and development of social insect castes: A synthetic review

2.1. Caste differentiation in eusocial insects

REPRODUCTIVE DIVISION OF LABOUR is a defining feature of EUSOCIALITY. Within any eusocial breeding system individuals are not generalists, but rather perform task-specialized behaviours as CASTES. The principal division of labour occurs between the royal castes that monopolize reproduction (i.e., queens, kings) and their non-reproductive helpers (i.e., workers, soldiers). In some societies this division of labor is particularly pronounced and sub-specialists within colonies are present in the form of SUB-CASTES**.** The process whereby individuals differentiate into alternative castes or sub-castes is termed CASTE DIFFERENTIATION. Caste differentiation can thus be studied for any eusocial TAXON – that is, any taxon with castes – of which there are tens of thousands of insect and other-arthropod examples (Rubenstein and Abbot 2017).

While eusocial breeding systems that feature castes have evolved more than a dozen times across the tree of life, they remain concentrated within the Arthropoda (Hölldobler and Wilson 2009) and especially in the insect orders Hymenoptera (including many species of eusocial ants, bees and wasps) and Blattodea (containing termites, all of which are eusocial cockroaches). These two insect lineages are phylogenetically remote from one another, yet both presumably evolved their caste systems in a convergent manner. As best described under the rubric of INCLUSIVE FITNESS THEORY, this likely occurred in

response to similar selection pressures and ecological opportunities that ultimately favoured complementary and contrasting behaviors within KIN GROUPS (Bourke 2011). Under certain conditions, specified by an evolutionary prophecy known as HAMILTON'S RULE, kin-mediated selection can drive the evolution of physical castes and sub-castes (Figure 2.1 A).

For any eusocial taxon, it is helpful to envision a set of developmental 'switches' that mediate caste differences. The convention of switches is sufficiently general to apply to HEMIMETABOLOUS and HOLOMETABOLOUS social insects alike and helps to point-out the single-most fundamental switch that is shared by all eusocial taxa – that which separates the more- from the less-reproductive line (Figure 2.1 B). Depending on the degree of task specialization that has evolved for a given taxon, there may be secondary or even tertiary downstream switches that mediate differentiation into sub-castes. Not all helper castes are fully STERILE, many workers and even (termite) soldiers can occasionally lay eggs.

A hymenopteran society typically consists of one or a few queens that are supported by numerous female workers. The haploid-male caste performs a strictly sexual role and is usually short-lived. By contrast, in termite societies, all individuals are diploid and can be of either sex, depending on the species. Termite colonies are formed by a sexual king and queen who reproduce many SUB-FERTILE offspring that develop into workers and soldiers. This evolutionary convergence of division of labor between the social Hymenoptera and the social Blattodea is independent still from comparable systems that are occasionally

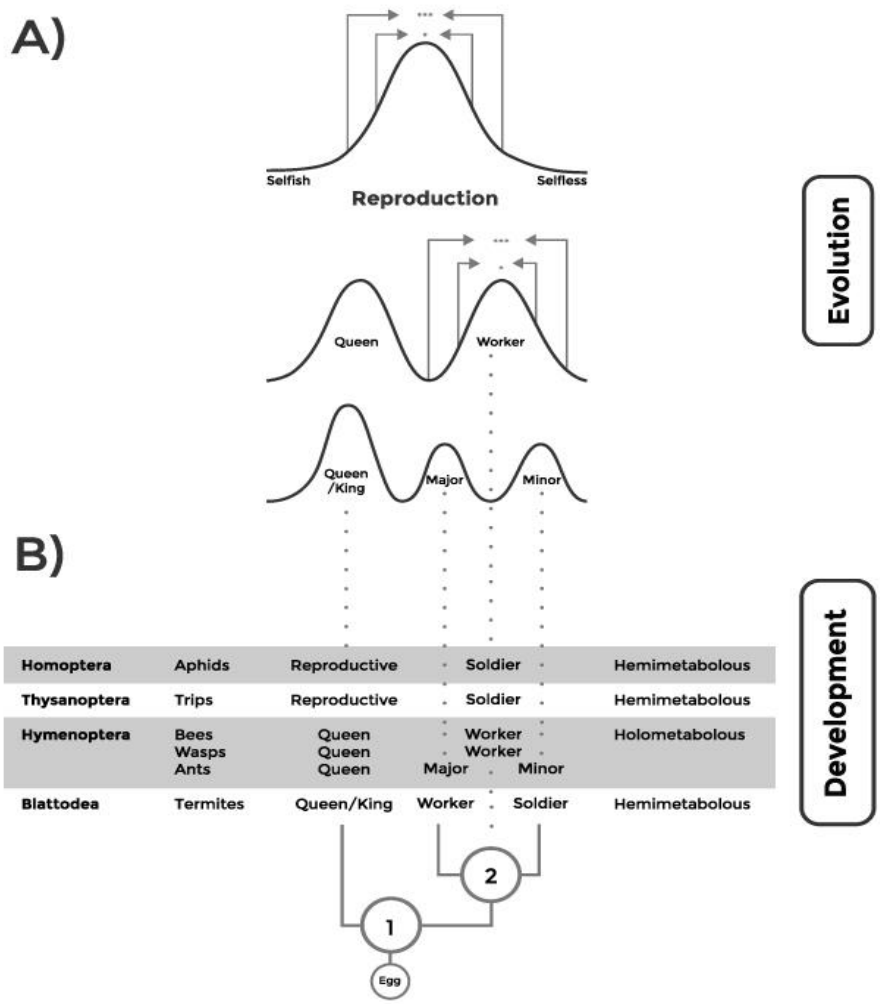

Figure 2.1. Evolution and development of insect castes and sub-castes **A)** Consider a population of individuals that vary in their reproductive disposition from selfish (parental) to selfless (alloparental). Any complementarity (*) among selfish with selfless individuals within a colony can be magnified (***) under kin-mediated selection for ever-more selfish with ever-more selfless behavioral combinations, leading to queen and worker castes, respectively. Further selection for caste complementarity can result in the emergence of sub-castes (i.e., major and minor workers). **B)** Castes and sub-castes have evolved independently in orders with gradual (HEMIMETABOLOUS) or discrete (HOLOMETABOLOUS) development. For some eusocial taxa, differentiation may involve a single switch (1) that mediates the primary division of labor between reproductive and non-reproductive castes. Caste differentiation in other eusocial taxa may involve a secondary switch (2) that mediates further division of labor among the SUB-CASTES.

described outside of these orders – for example, some species of aphids (Homoptera) and thrips (Thysanoptera) have defensive soldier castes (Costa 2006).

2.2. Genetic factors affecting caste differentiation

2.2.1. General effects of genotype on caste

There is a deep understanding of how environmental factors can actually or potentially influence caste differentiation in social insects (Nijhout 2003; Wheeler 1986). It is fundamentally established that differences in diet, temperature or social opportunity among other environmental cues can bias developmental switch-points and trigger one caste trajectory over another, even without any salient differences in individual genotype. This understanding has, however, expanded in recent years to more fully incorporate a role for individual genetic differences that might bias caste fate. For example, the honey bee *Apis mellifera* is well-known for its environmentally-mediated caste system: female larvae, when fed a ROYAL JELLY diet are predisposed to develop into queens as opposed to workers. The queen-worker switch is therefore cued to an environmental difference – namely, nutrition. Yet, even here, a role for genetic effects can be measured in at least three ways.

First, for this and other species of eusocial Hymenoptera in which sex is determined by genotype, only certain individuals – i.e., diploid females – have the genetic capacity to develop into a queen or worker at all. Individual differences in PLOIDY, and more

specifically heterozygosity at the *csd* locus (Beye et al. 2003), are therefore important genetic effects that fundamentally influence caste differentiation. Second, female larvae may be genetically variable at AUTOSOMAL LOCI in their responsiveness to the royal diet or in their ability to signal for such a diet from the adult care-giving workers (He et al. 2016; Moritz et al. 2005). Given that honey bee queens typically mate with multiple males, we can expect developing larvae to vary substantially in their genotype and, as a consequence, to compete among each other for limited food and the rare opportunity for reproductive differentiation. Finally, genetic variability among the care-giving workers can affect their propensity to provision royal jelly in certain amounts or to provision it specifically to larvae that carry the same set of paternal genes, as predicted under gene-mediated NEPOTISM. This brief consideration of genetic contrasts within honey bee colonies makes clear that the potential for genetic effects on caste are not zero, even for species like *Apis mellifera*, in which environmental differences are already known to play a strong role.

But how strong or common are genetic effects on caste in the social insects? One entry point to further explore the potential for genetic effects on caste, is to imagine caste-determining alleles that, if inherited, would predispose individual development towards one caste or another, say, an ALLELE for queen-ness or for worker-ness. A hard-wired genetic system of this type is intuitively appealing but would seem evolutionarily unstable in nature. The unconditional expression of alleles for queen-like selfishness would likely fix and, conversely, the unconditional expression of alleles for worker-like ALTRUISM would

likely be lost from the population, leaving no further variation for caste (Bourke and Franks 1995; Crozier and Pamilo 1996). Despite this reasoning and the associated expectation for low or no genetic variance at caste-biasing loci, exceptional cases, where individual differences in genotype (ΔG) are strong determinants of caste fate, apparently do exist.

Where found, genetic variance on caste may persist under a balance of within and between-colony selection. For example, within-colony selection could favor an environmentally cued system in which genotypes invariant at caste-biasing loci avoided competition with each other (for anything) but lacked any individual responsiveness to environmental demands. Between-colony selection, by contrast, might compensate for within-colony competition if variation at caste-biasing loci somehow helped the colony as a whole to fine-tune it's response to those demands – for example, by promoting more efficient divisions in colony labor (Oldroyd and Fewell 2007) or by increasing a colony's environmental performance (Mattila and Seeley 2007) or its capacity for HOMEOSTASIS (Jones et al. 2004) or its social resistance to parasites (Seeley and Tarpy 2006). Remarkably, depending on the strength of BALANCING SELECTION, caste-biasing alleles may actually be maintained at evolutionarily stable frequencies.

To the extent that any genetic effects on caste are responsive to selection, and thus display NARROW-SENSE HERITABILITY, they are likely ADDITIVE (ΔG_A) effects, whereby the combined effect of alleles at one or more loci equals the net sum of their individual

effects (Falconer and Mackay 1996). Alternatively, genetic effects that are less responsive to selection are more likely NON-ADDITIVE and arise from interactions among alleles, either at single loci as in DOMINANCE effects (ΔG_D) or among different loci as in EPISTASIS (ΔG_I). Both the non-additive ΔG_D and ΔG_I effects arise from the fortuitous combination of alleles in each generation, only to possibly break-up under RECOMBINATION in the next. For complex traits, however, heritability is a parameter of multiple loci. Heritability thus involves a multitude of genes with varying magnitude of their effects and interactions, which in combination may influence a quantitative trait or alter how a given population evolves in response to selection. For non-additive effects to respond to selection they would need to persist from one generation to the next, the probability of which varies with the likelihood that interacting alleles become reconstituted in the next generation (Falconer and Mackay 1996). This likelihood is diminished in the order Hymenoptera where recombination rates are among the highest reported for animals (Wilfert et al. 2007). Hence, non-additive effects on caste are expected to be less predictable, have a lower narrow-sense heritability and should ultimately be less responsive to selection.

How might additive or non-additive genetic effects evolve to influence caste differences in eusocial taxa? There is a growing interest in this question with the expectation that many genetic effects on caste await discovery (Anderson et al. 2008; Keller 2007; Lo et al. 2009; Schwander et al. 2010). For now, the nature of the genetic effect on caste morphology is not always clear, but specific examples I summarize below indicate that caste appears to associate with heterozygosity at one or a few loci, suggesting dominance

interaction effects, or with hybridization and genome-wide compatibility, indicating epistatic effects. Likewise, it is not yet possible to infer precisely how conditional each reported genetic effect is upon its environmental context (ΔG x ΔE). As data and evidence accumulate, it will be prudent to test these hypotheses within a quantitative genetic framework (Linksvayer and Wade 2005; Moore and Kukuk 2002), ideally one that assigns specific genetic effects to direct and indirect sources (Figure 2.2).

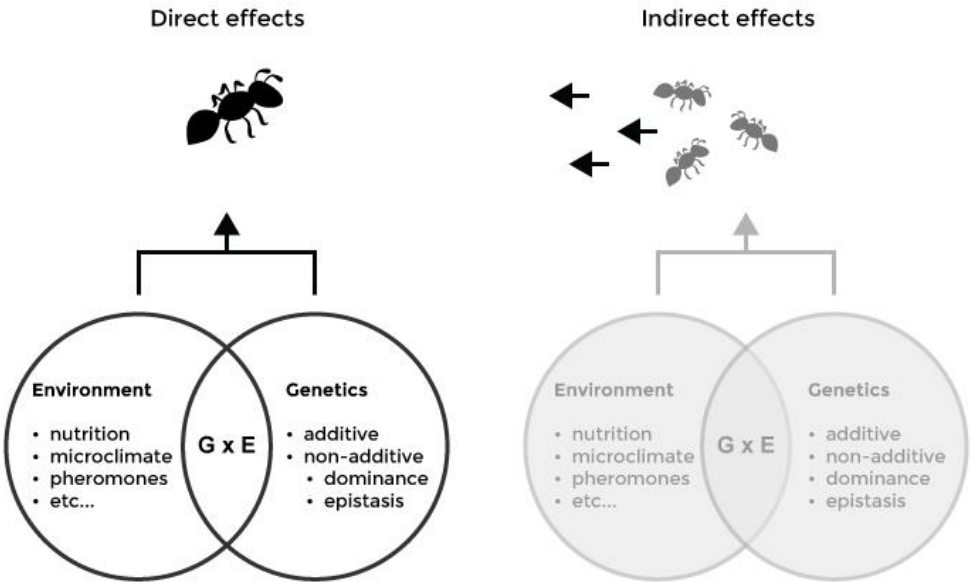

Figure 2.2. Conceptual map of genetic effects on caste differentiation. Each developmental switch depicted in Figure 2.1 can potentially be influenced by ΔE, ΔG or the interaction of environmental and genetic differences ($\Delta G \times \Delta E$). The quantitative genetics approach can be useful for distinguishing additive from non-additive genetic effects, with the latter arising through allelic interactions at single loci (DOMINANCE) or between multiple loci (EPISTASIS). Each type of effect on caste differentiation may arise directly from focal individuals or indirectly via their social partners.

2.3. Specific effects of genotype on caste differentiation and morphology

2.3.1. Tests from polyandrous species

Disentangling the effects of genetic signals from those of environmental stimuli on caste differentiation requires careful experimentation. In POLYANDROUS species, for example, a queen mates with multiple males and, consequently, multiple worker PATRILINES make up the colony structure. Polyandrous colonies can therefore be used to measure the effect of paternal genotype on caste differentiation while controlling for both environmental and MATRILINEAL EFFECTS. Hughes et al. (2003) applied this approach in the polyandrous leaf-cutting ant *Acromyrmex echinatior* and found that differences in paternal genotypes of female nestmates biased worker differentiation towards either minor or major worker sub-castes. One possible explanation for this finding is that in polyandrous species, larvae with a higher genetic sensitivity to environmental growth factors such as nutrition or PHEROMONES may be genetically predisposed to develop into major as opposed to minor workers.

In the harvester ant *Pogonomyrmex badius*, Rheindt et al. (2005) have likewise demonstrated that female nestmates from different patrilines tend to bias their development towards a particular caste phenotype. That is, workers from specific patrilines tend to differentiate into either minor or major workers, or even into reproductive GYNES. The relationship between genotype and production of caste alternatives in *P. badius*, however, is not straightforward. The direction of bias in caste

ratios is highly variable between colonies and also appears to be influenced by the food supply (Smith et al. 2008), implying that this species may have evolved a genetically-variable sensitivity to environmental cues. It is therefore apparent that the genetic effect on caste and sub-caste ratios detected in this species is dependent upon the broader social and environmental context ($\Delta G \times \Delta E$).

The sheer genetic diversity of polyandrous insect colonies, where same-aged workers from different patrilines undergo development in a shared environment, provides a powerful test for genetic effects on caste differentiation. The polyandrous army ant *Eciton burchellii* shows extreme caste polymorphism in the form of four morphologically distinct worker sub-castes. Jaffe et al. (2007) capitalized on this experimental opportunity to show that several patrilines were biased towards producing particular morphs, which indicates a significant genetic component to caste determination. By analysing the composition of worker caste-ratio variation within each patriline, the authors estimate that 15% of worker sub-caste variance can be assigned to additive effects (ΔG_A), with the remainder (~85%) being attributed to differences in rearing environment. As for *A. echinatior* and *P. badius* above, the genetic control of caste in this species is plastic, indicating it has a broad reaction norm, and is unlikely to be hardwired.

Empirical tests for genetic effects on caste for any eusocial insect, however, remain rare, and, where they have been performed, can yield negative results despite high statistical power (Keller et al. 1997; Ujvari et al. 2011). The theoretical expectation for low genetic

variance at caste-biasing loci, together with the apparent rarity of caste-biasing genetic effects in the field, make discoveries like that of *Acromyrmex* and *Pogonomyrmex* systems exceptional. Both systems demonstrate how balancing selection can maintain genetic variation at both primary and secondary switches (Figure 2.1 B). It is therefore important to consider how caste differentiation is influenced by both environmental and genetic factors, but also by the effects that arise due to their interaction. In future experimentation, it may even be possible to partition variance associated with caste into specific ΔE, ΔG and $\Delta G \times \Delta E$ components, such that their sum approaches '1' and account for nearly all phenotypic variance in caste.

2.3.2. Tests from hybrids

Dominance or epistasis may likewise affect caste differences. Schwander and Keller (2008), for example, suggest that caste morphology in *Pogonomyrmex rugosus* is influenced by genetic compatibility between the polyandrous harvester ant queens and her mates. That is, some crosses between males and females appear to produce genetic combinations in recombinant offspring that are most compatible with queen development, while other combinations lead to differentiation into workers. Cases of genetic 'compatibility' are intriguing because they imply a genetic interaction effect. Importantly, however, this interaction does not typically follow Mendelian predictions. That is, it does not result in offspring genotypes with simple predictable genotypes, as if from a Punnet square, but rather from as-yet unpredictable chimeric genotypes of inter-

specific recombination. The caste phenotypes from such hybrid matings are therefore not easily modelled. A reigning queen's ability to produce future queen vs. worker offspring is instead in some way dependent on her own genotype and the genotypes of each male she mates with. The fitness consequences to new or parental queens, as well as to the workers and males, might be complex, but one evolutionary response to increase the likelihood of favourable matri- and patri-gene combinations among a queen's progeny is for her to mate multiple times.

Other examples of idiosyncratically stable genetic polymorphisms that can bias caste development towards alternative routes are found in hybrid zones of harvester ants. In areas of the southern United States, where *Pogonomyrmex rugosis* and *Pogonomyrmex barbatus* overlap, a strong influence of genotype on caste appears to emerge from non-additive effects associated with HYBRIDIZATION. INTER-SPECIFIC crosses yield worker offspring that are all heterozygous at specific marker loci. By contrast, INTRA-SPECIFIC crosses yield mostly homozygous offspring that develop into ALATE queens (Helms Cahan et al. 2002; Julian et al. 2002; Volny and Gordon 2002). A tight association between ZYGOSITY and caste is limited to naturally occurring areas of SYMPATRY, suggesting a $\Delta G \times \Delta E$ effect. In fact, additional study on this system has begun to uncover a potentially complex hardwired genetic system, where each population consists of two cryptic and genetically-distinct lineages, yet queens are able to mate with males from both species (Schwander et al. 2007). Depending on the mating match, pure-lineage female offspring differentiate into

queens, while females derived from the inter-lineage combinations differentiate into workers (Helms Cahan and Keller 2003).

In general, inter-specific hybridizations result in the production of inviable or sterile offspring. In eusocial insects, however, these sterile worker hybrids may function as altruistic helpers. The complexity of the *Pogonomyrmex* hybrid system may help inform researchers of other cryptic cases and alternative mechanisms involved in genetic caste determination. In colonies of the southern fire ant *Solenopsis xyloni*, for example, a molecular analysis at marker loci revealed that queens, when mated with conspecific males, produced only new queens, whereas females mated with CONGENERIC males in the species *Solenopsis geminata* produced workers instead (Helms Cahan and Vinson 2003). The resulting worker offspring showed extremely high levels of heterozygosity consistent with them being F1 hybrids. These results suggest that pure-bred *S. xyloni* queens in hybrid zones may have lost their genetic potential for worker production and depend on alleles from *S. geminata* males to generate a (hybrid) worker force. The discovery of the *Pogonomyrmex* and *Solenopsis* hybrid systems suggests that inter-specific hybridization may play an important role in the evolution and maintenance of caste differences.

2.3.3. Tests from single and double locus models

Beyond the remarkability of hybrid systems, simple heterozygosity at intra-specifically variable loci can also bias caste fate. This was demonstrated in the European slave-

making ant *Harpagoxenus sublaevis*, where Winter and Buschinger (1986) showed that genotype at a single locus constrains development of female larvae into either queen or worker-like castes. Similarly, a pioneering study by Kerr (1950) showed that queens of the stingless bee genus *Melipona* differentiate from larvae that are heterozygous at as few as two unlinked loci. Larvae homozygous at these two loci are, by contrast, constrained to differentiate as workers. These findings indicate that heterozygosity at specific loci is an important component of caste fate regulation. The single or two-locus models alone, however, do not provide a full explanation for the widely-ranging caste ratios found in the *Melipona* system (Ratnieks 2001). Moreover, even heterozygote larvae differentiate into workers when deprived of food (Hartfelder et al. 2006), which implies that this genetic effect remains dependent upon environmental context ($\Delta G \times \Delta E$).

For one species of subterranean termite, *Reticulitermes speratus*, crosses between different types of immature reproductives – technically, NYMPHOIDS (with wing buds) or ERGATOIDS (without wing buds) – produce strongly differentiated caste and sex ratios, despite uniform rearing conditions (Hayashi et al. 2007). The observed worker-to-nymph ratio is plausibly explained by a genetic model featuring a single X-linked locus (*wk*) with two alleles: A and B. Offspring with genotype wk^{AB} and $wk^{A}Y$ develop into female or male workers, whereas offspring with genotypes wk^{AA} and $wk^{B}Y$ develop into female or male nymphs. Remarkably, offspring with genotype wk^{BB} are inviable. The allelic composition at specific loci in *Reticulitermes speratus* thus explains a significant proportion of within-colony variation in caste ratios. However, even in this case, the genetic effect appears to

depend upon a broader social and environmental context, again suggesting the importance of $\Delta G \times \Delta E$ interactions (Crozier and Schluns 2008).

The full extent of genetic caste determination in termites remains largely untested, but the potential for new discoveries of genetic effects on caste fate from the Blattodea is very promising (Lo et al. 2009; Schwander et al. 2010; Vargo 2019). Genetic manipulation studies in eusocial insects are already making progress and some genotypic influence of genotype on caste has been detected. One such case occurs among the non-reproductive worker and soldier castes in colonies of *Mastotermes darwiniensis*. Within single colonies collected from one site in Australia's Northern Territory, workers and soldiers had distinct genotype frequencies at specific marker loci (Goodisman and Crozier 2003). The association between caste and genotype detected here implies that certain genotypes may direct production of distinct castes in these species. Second, a number of termite species have same-sex workers or soldiers, either male or female depending on the species. The association of caste with sex suggests that some, as yet undiscovered, caste-biasing genes from termites are SEX-LINKED (Figure 2.3).

2.3.4. Tests from thelytokous parthenogens

In eusocial species of Hymenoptera and Blattodea, genetic effects can be linked to a form of asexual reproduction in which females are produced from unfertilized eggs, a reproductive system known as THELYTOKY. In the fire ant *Wasmannia auropunctata*, sterile workers are produced by normal sexual reproduction, while queens originate from CLONAL

REPRODUCTION with complete absence of genetic contribution from males (Fournier et al. 2005). The conditional use of sex for worker but not queen production may represent a balance between queens that maximize the direct transmission rate of their genes while maintaining a genetically variable worker force. In this system, patri-genes typically contribute only to production of the sterile caste and thus have zero fitness. However, patri-genes have a secondarily evolved ability to eliminate the maternal genome from the fertilized egg and, thus, effectively clone themselves, too. Such a remarkable system, where the alternate use of sexual vs. asexual reproduction is utilized to generate queen, worker or male castes, has been identified in other species of ants, including *Vollenhovia emeryi* (Ohkawara et al. 2006) and *Cataglyphis cursor* (Pearcy et al. 2004).

The subterranean termite genus *Reticulitermes* contains few species that deviate from strictly sexual reproduction. Nevertheless, in *R. speratus*, *R. virginicus* and *R. lucifugus*, queens can be occasionally replaced by means of PARTHENOGENESIS, which is a form of asexual reproduction. Queens in these three species have evolved an ASEXUAL QUEEN SUCCESSION (AQS) SYSTEM, whereby primary queens are replaced by numerous parthenogenetically produced female nymphoids (Vargo 2019). The females are produced from unfertilized eggs via thelytoky and are clonal offspring to the queen. The production of female clones is extraordinary and rendered operational through a variation on meiosis (called AUTOMICTIC PARTHENOGENESIS with terminal fusion) that yields replacement queens that are almost completely homozygous. This distinct genetic profile seems to preferentially bias the differentiation of the asexually produced nymphs as

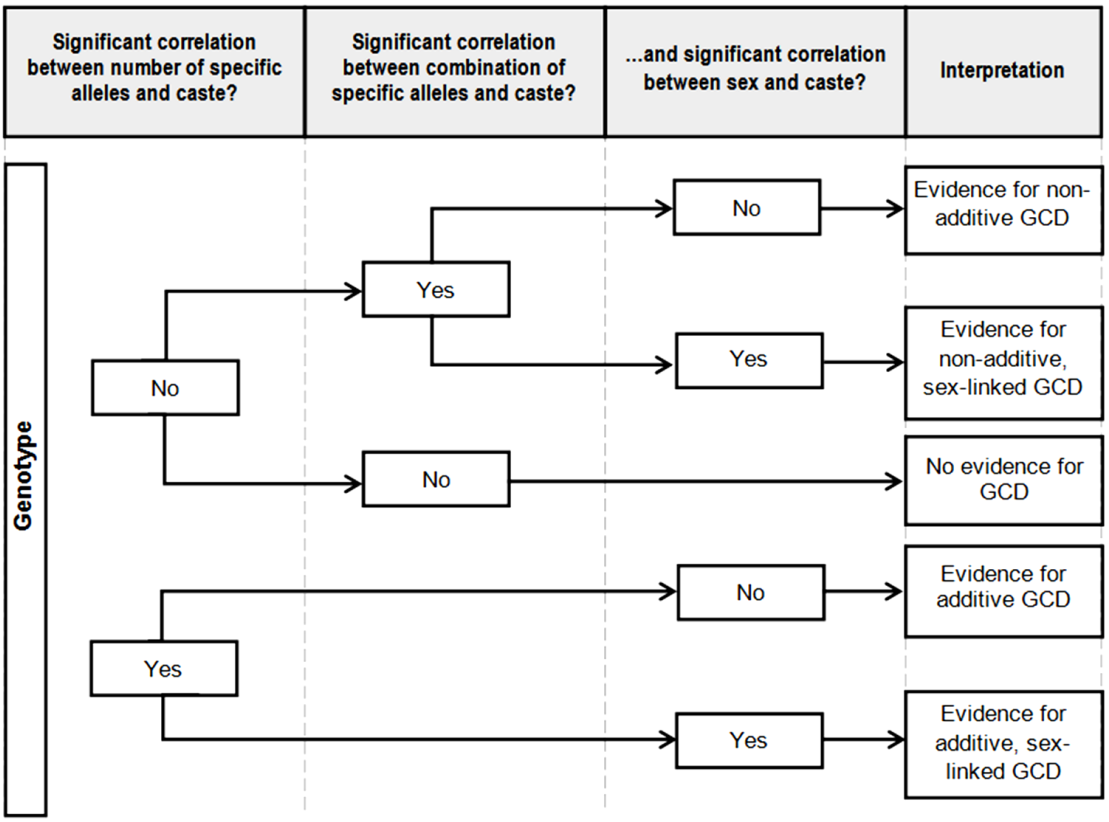

Figure 2.3. Decision tree for evidence of genetic caste determination (GCD) in termites. If variation in caste ratios is explained, in part, by variation in genotype then we expect that single alleles or whole multi-allele genotypes at marker loci will correlate with caste. Following these interpretations of genotypic data, it may be possible to distinguish additive from non-additive effects and further test if either type of effect is sex-linked.

replacements over their sexually produced counterparts. For *R. speratus*, the genetic bias was found to be specifically associated with two marker loci that correlate strongly in zygosity with future-queen status. This pattern suggests a multi-locus genetic effect on caste that is somehow linked to reproductive priority, ovary development and FECUNDITY in termites (Yamamoto and Matsuura 2012).

AQS has likewise been observed among higher termites (Termitidae). *Emiratermes neotenicus* seems to generate replacement queens using a similar variation of meiosis such that, in this case, females have nearly identical genotypes to their mother (Fougeyrollas et al. 2015). Other termite species have evolved yet different mechanisms for ploidy restoration under AQS that, in one way or another, canalize asexually-produced genotypes into NEOTENIC QUEENS (Vargo 2019). This disparate phylogenetic pattern for AQS among unrelated termite species suggests that selection can repeatedly favour breeding systems with strong genetic effects on caste, despite pronounced differences in ecology between the lower (Rhinotermitidae) and higher termite species studied so far. An order-wide pattern for genetic effects on caste in termites under sexual and asexual reproduction has yet to emerge but further cases will continue to upset the notion for strictly environmental caste determination in social cockroaches (Lo et al. 2009).

2.4. Indirect genetic effects on caste

For social insects, the physical and social environments are conflated because nestmates share and shape each other's living space. Even classic environmental factors that potentially affect caste differentiation, including differences in nutrition, pheromones and MICRO-CLIMATE, can have a socio-genetic component if these differences are explained by the genotype of queens and nestmates. Such indirect genetic effects (Cheverud 2003) can significantly complexify the quantitative genetic approach to studying caste differences (**Figure 2.2**). Indirect genetic effects – expressed, typically in eusocial colonies,

from parent or sib-social partner towards developing brood – can likewise be additive (ΔG_A) or non-additive (ΔG_D, ΔG_I) in nature, and conditional upon environmental context ($\Delta G \times \Delta E$). Indirect genetic effects arising from the action of parents to influence differentiation of their own offspring are presumably due to expression of parental care genes. Indirect genetic effects that arise from care-giving workers, by contrast, are also likely due to 'parental care' genes but in this case are not literally parental, but rather are expressed in pre- or non-reproductive ALLOPARENTS (Linksvayer and Wade 2005).

In principle, indirect effects must exist; inclusive fitness theory depends on them. However, empirical studies that distinguish direct from indirect effects in social insects remain a few (Linksvayer and Wade 2005; Moore and Kukuk 2002). In the polyandrous European honey bee, care-giving nurse workers are occasionally found to preferentially feed and rear new queens from larvae of their own patriline (Osborne and Oldroyd 1999). To the extent that workers of a patriline preferentially rear reproductives that are most related to themselves, such nepotism accelerates development of new queens and we can infer an indirect genetic effect arising from kin competition among workers. In this case, even though the cue (diet) that triggers caste fate is 'environmental', it is the natural genetic variation among nurse-age workers that likely determines which larvae receive that cue and ultimately differentiate into queens. Further evidence for indirect effects on caste from honey bees can be found in a heritable component (ΔG_A) to the production of male (drone) vs. worker comb (Page et al. 1993), a worker-controlled trait that affects the caste-fate of others.

The application of quantitative genetic thinking to fully decompose direct and indirect sources of variance on caste differences into its additive and non-additive components has not been widely attempted. One approach to isolating direct from indirect effects is to use a factorial design, such that single queens are set-up to rear related and unrelated brood within a common environment (i.e., cross-fostering), replicated across queens and environments. In this manner, it may be possible to partition variance in caste or sub-caste into direct (larvae), indirect (maternal, sib-social), and environmental genetic components. Using an indirect genetic effects approach, Linksvayer (2006) found heritable variation for queen and worker effects on developing caste size and ratio in the acorn ant *Temnothorax curvispinosus*, indicating that queen and worker influences on larval caste determination – both indirect effects – may have an additive genetic basis (ΔG_A) and can evolve. For any eusocial species in which workers actively rear developing brood, it is the worker-expressed genes that are well-suited for biasing caste differentiation.

2.5. Epigenetics of caste differentiation

I have discussed how environmental and genetic factors can trigger caste switching at developmental inflection points (**Figure 2.1**), and how these can affect focal individuals directly or indirectly via their social partners (**Figure 2.2**). EPIGENETIC FACTORS that are 'above the genome' complement this understanding of caste differences; they provide a mechanical interface between environmental experience and gene regulation, and

function to coax environmentally plastic responses from even a fixed genome (Dupont et al. 2009). The epigenomic regulation of gene expression, which can involve DNA METHYLATION, HISTONE MODIFICATION or changes to NON-CODING RNAS, among other mechanisms, is based on biochemical modifications that alter accessibility of DNA for transcription or otherwise affect gene expression. Epigenetic effects, however realized, help to record environmental experience across the genome and have been implicated as a mechanism for caste or sub-caste differentiation in a range of eusocial insects (Li-Byarlay 2016; Welch and Lister 2014; Yan et al. 2015).

DNA methylation of cytosine bases in particular has been the focus of epigenetic effects on caste differences. In mammalian genomes, the DNA METHYLTRANSFERASE-assisted addition of a methyl group to the cytosine-pyrimidine ring results in 5-methylcytosine, which is a biochemical modification typically associated with transcriptional repression. It is unclear if this epigenetic effect operates similarly in social insects, but methyl 'marks' do appear to be associated with caste-biased genes, particularly at cytosine-guanine (CpG) dinucleotides within gene bodies (introns and exons) and, in some cases, specifically at alternative splicing sites (Bonasio et al. 2012; Li-Byarlay et al. 2013; Maleszka 2016). DNA methylation may therefore represent an important epigenetic mechanism of phenotypic plasticity that co-evolved with some eusocial breeding systems (Moczek and Snell-Rood 2008), although there is as yet no strong phylogenetic support for a general evolutionary association between sociality and DNA methylation (Bewick et al. 2016).

The honey bee *Apis mellifera* has a full mammal-like complement of DNA methyltransferase genes (*DNMT1, DNMT2* and *DNMT3*) and provides a leading example of epigenetic effects on caste differentiation. Female larvae fed a diet of royal jelly are predisposed to develop into queens as opposed to workers, but a pioneering study by Kucharski et al (2008) showed that queen phenotype can be induced without nutritional signals from a larval worker's diet. Instead, by silencing a key gene involved in *de novo* DNA methylation (*DNMT3*) the authors of this study bypassed the normal nutritional cue to experimentally induce newly hatched larvae to emerge from the pupal stage with queen-like qualities, suggesting that differential methylation of cytosines is involved in the epigenetic control of caste determination. Later studies appeared to bolster this hypothesis by suggesting that nutrient-responsive pathways in the honey bee are enriched for methylated genes, and that hive-reared queen and worker larvae differed strongly in their proportion of genes methylated (Forêt et al. 2012). Caste-biased genes may also have distinct CpG profiles (Elango et al. 2009). If so, then the well-known environmental effect of diet on honey bee caste differentiation seems also to involve epigenomic modifications by DNA methyltransferases.

While the honey bee *Apis mellifera* is an emerging epigenetic model, evidence for CpG methylation appears common yet variable across the social Hymenoptera and social Blattodea (Yan et al. 2015). For example, there is little evidence for DNA methylation in the paper wasp *Polistes dominula* (Standage et al. 2016) but its congener *P. canadensis* has plenty of CpG methylated sites. Paper wasps seem also to lack the key *de novo* DNA

methylation gene *DNMT3* that is present in most other social insect genomes so far examined (Li-Byarlay 2016; Standage et al. 2016). DNA methylation is more clearly functional and associated with caste in several genera of ants. Both *Harpegnathos saltator* and *Camponotus floridanus* show evidence of differential methylation associated with caste or sub-caste differentiation (Bonasio et al. 2012). Likewise, in *Pogonomyrmex* overall methylation is lower in hybrid lines whose caste fate is genetically determined than in the respective parental lines that display environmental caste determination (Smith et al. 2012). In each case, differentially methylated genes appear to modulate environmental information when it is relevant to caste differentiation.

Epigenetic analysis of the clonal raider ant *Cerapchys biroi* revealed a robust CpG methylation system across its genome, but the methylation status of individual cytosines was not clearly associated with parental (queen-like) or alloparental (worker-like) behaviour. It is worth noting that this species has no distinct queen and worker castes, but colonies do show queen-like and worker-like phases whereby females first reproduce by parthenogenesis then shift their behaviour to care for these offspring. Libbrecht et al. (2016) showed that analysis of single samples from each phase produced only spurious correlations between methyl status and caste-like behavior and, despite careful design and replication, did not yield a consistent pattern across multiple samples. The clonal raider ant study therefore highlights the need for careful interpretation of epigenetic data as they emerge from the field of social insects.

A signature of CpG methylation is found in the termite *Coptotermes lacteus* (Lo et al. 2012) as it is for other related (Rhinotermitidae) species, including *Reticulitermes flavipes* in which differentially methylated sites are associated with caste-biased expression between reproductive alates and non-reproductive workers and soldiers (Glastad et al. 2016). The genomic sequence of *Zootermopsis nevadensis* (Termopsidae) adds further detail to these screens. First, fully ~12 % of CpG sites are methylated, compared to estimates of less than 1% of CpG sites methylated in the honey bee *Apis mellifera*. Second, a comparison of whole bodies from workers with those from reproductives revealed a large number of genes as differentially methylated, with higher levels of methylation found in the reproductive caste (Terrapon et al. 2014). These same genes were likewise associated with ALTERNATIVE SPLICING and were enriched for processes associated with development, which further implicates epigenetic effects via DNA methylation on termite caste plasticity and differentiation.

It is clear that the process of caste differentiation in any eusocial species is accompanied by massive changes in gene expression, even with no underlying genetic differences. For some species this social plasticity may well be mediated by epigenetic mechanisms that respond to environmental cues that function to coax caste-specific information out of a common genetic background. Despite the apparent adaptive suitability of epigenetic factors to the regulation of social phenotypes, the function of specific mechanisms related to methylation, histone modification, nucleosome stability or alternative splicing

are only beginning to be understood. New discoveries and analytical approaches (i.e., Morandin et al. 2019) are rapidly forthcoming.

CHAPTER THREE: A molecular evolutionary analysis of caste-associated genes in the Eastern subterranean termite

3.1. Introduction

Eusociality is a rare example of a so-called 'major evolutionary transition' whereby one level of biological organization is predicated upon another, leading to notable increases in adaptive complexity (Queller 2000). In this case, eusocial breeding systems are markedly more complex than the subsocial life histories from which they presumably evolved (Bourke 2011). Despite the rarity of these transitions, however, eusociality has evolved independently more than a dozen times in insects (Crespi and Choe 1997; Rubenstein and Abbot 2017). It is most commonly found among ants, bees and wasps (Hymenoptera), as well as termites (Blattodea). Termites represent an origin of eusociality separate from that of social Hymenoptera and can be exploited to test general ideas on how insect societies evolved via kin selection (Howard and Thorne 2011).

Termites are a type of social cockroach that evolved in the mid-to-late Jurassic era (Bourguignon et al. 2014) from an ancestry shared with the extant subsocial genus *Cryptocercus* (Inward et al. 2007; Klass and Meier 2006). This relatively small clade of eusocial insects (~2500 spp.) was long known as the 'Isoptera' but is now classified at a sub-ordinal rank within the roaches (Termitoidae; Eggleton et al. 2007). Unlike other roaches, however, termites have distinct castes and extraordinary social structures (Roisin and Korb 2011), build conspicuous nests with complex architectures and some

species are agriculturalists (Nobre et al. 2011). Termites likewise have a rich gut microbiome that facilitates their pivotal role as biodecomposers of cellulose and lignin (Brune and Ohkuma 2011), especially within tropical forest ecosystems. These and other conspicuous features of termite biology are associated with their wood-feeding adaptation and subsequent transition to eusociality (Korb 2007; Wu et al. 2015).

Within the eusocial theme, termite breeding systems do vary but are essentially extended families (Vargo 2019). Parental kings and queens are specialized for reproduction while their offspring have reduced or no reproductive potential and perform other roles typically associated with colony growth, maintenance and defence (Noirot 1989; Shellman-Reeve 1997). Workers and soldiers are therefore reproductively altruistic because they help their reproductive relatives to generate large numbers of non-descendent kin (i.e., siblings, half-siblings, etc.) at the expense of their own direct fitness (Higashi et al. 2000; West et al. 2007). The evolution of altruism, for any specific taxon, is intriguing because it requires 'genes for altruism' to evolve indirectly, via selection on reproducing relatives who carry copies but do not generally express these focal genes (Charlesworth 1978; Parker 1989; Thompson et al. 2013). The notion of indirect selection at the gene level is widely discussed in the literature and extensively modelled via population genetic equations (Franks et al. 2009; Gardner et al. 2011; Hamilton 1972); yet how real genes are shaped or transformed in response to this type of selection at the molecular level remains unknown.

One proposed idea is that indirectly transmitted genes for altruism that are expressed in sterile castes but inherited through reproductive individuals are buffered from the full strength of selection. This is, the gene's proximate effect is on the worker that expresses it, but the ultimate effect is on the queen, who does not express these worker-specific genes, nor is herself exposed to the environment due to the presence of many nestmates around her. Accordingly, genes associated in their expression with reproductively altruistic castes will effectively experience relaxed molecular evolution relative to genes directly selected for selfish reproduction (Hall and Goodisman 2012; Linksvayer and Wade 2009). If this 'relaxed worker' hypothesis is validated (Table 3.1), I expect that the intensity of selection, as measured here by the ratio of non-synonymous to synonymous substitutions, or d_N/d_S (Kimura 1983), will be diminished for genes associated with the helper castes and not significantly different from a neutral value of one, in contrast to the reproductively-selfish caste genes for which this ratio will deviate from neutrality in either positive $d_N/d_S > 1$ or purifying $d_N/d_S < 1$ direction (Figure 3.1 A).

Although this hypothesis is logical, other scenarios describing caste-mediated selection are possible. Harpur et al. (2014), for example, suggests that in honey bee workers, genes associated with altruistic castes can evolve rapidly under selection, as might be expected if workers effectively buffer queens from environmental selection within a homeostatic colony. This 'adapted worker' hypothesis is essentially opposite to the nearly neutral or 'relaxed worker' idea, as depicted in Figure 3.1 B. Yet more scenarios are possible, including for genes uniquely associated with any one specific caste, altruistic or

otherwise, to evolve rapidly via drift or selection following their release from multi-caste PLEIOTROPY (Figure 3.1 C). If so, caste-associated genes may drift towards neutrality or suddenly respond more readily to directional selection than caste un-biased genes, at least for rapid bursts until a new functional equilibrium is reached (Gadagkar 1997; Hunt et al. 2011). Despite several alternative hypotheses for caste-mediated selection, none have been widely tested and, consequently, a general signature of indirect or caste-mediated selection, if there is one (Harrison et al. 2020; Helanterä and Uller 2014), remains unknown.

In my study, I propose to test for molecular signatures of kin selection using a novel approach – namely, by comparing the strength and direction of gene-level selection across caste-biased and un-biased genes from a social population genomics dataset from the insect order Blattodea. Specifically, I will exploit allelic diversity captured in a newly available RNA sequence dataset derived from within my own laboratory at Western University (Wu et al. 2018) to reveal any differences in patterns of nucleotide substitution associated with the functionally sterile worker and soldier castes. The species *R. flavipes* in the Wu et al. (2018) study is native to eastern North America, but has also been found outside of these areas (Eyer et al. 2020; Scaduto et al. 2012). In native termite colonies, reproduction is monopolized by the royal pair – the queen and king that are supported by thousands of soldiers and workers (Lainé and Wright 2003; Thorne et al. 1999). By contrast, invasive populations are characterized by the absence of a royal pair and instead maintain a slightly different social structure, whereby reproductive roles are

performed by immature ('neotenic') nymphs (Raffoul et al. 2011; Vargo and Husseneder 2009). I reason that genes identified by Wu et al. (2018) as associated in their expression with the reproductive nymph caste will evolve under direct selection, similar to the expectation for kings and queens, and therefore will show evidence of strong or weak selection, depending on which hypothesis (Figure 3.2) is supported. Conversely, I reason that genes associated in their expression with the non-reproductive worker or soldier caste will show evidence of indirect selection, and thus evolve in a pattern distinct from that of nymphs. The precise pattern will, again, depend on which hypothesis is supported.

Finally, I will compare the broadest results from my analysis of the *R. flavipes* transcriptome to what has been observed in two other termite species for which comparable data are available. Transcriptomic data from the drywood termite *Cryptotermes secundus* (Kalotermitidae; Harrison et al. 2018) and the dampwood termite *Zootermopsis nevadensis* (Archotermopsidae; Terrapon et al. 2014) provide a point of comparison against the subterranean *R. flavipes* (Rhinotermitidae) and allow me to expand my evolutionary analysis beyond a single species, to include representatives from three different termite families.

Reticulitermes spp. tend to have very large colonies (>100K individuals) that can extend to multiple sites ('multi-site nesters'; Abe 1991) within soil and have a relatively rigid caste system with sterile, or true, workers. By contrast, *Zootermopsis* and *Cryptotermes* spp. tend to have smaller colonies (100 -1000s) that consume solely the wood in which they

nest ('single-site nesters') and have a relatively flexible caste system with sub-fertile, or false, workers (Noirot and Pasteels 1987; Shellman-Reeve 1997; Thompson et al. 2000). By including this diversity of species in my extended analysis, I may be able to correlate detected patterns of nucleotide substitution with broad differences in termite social biology and ecology.

Table 3.1. Contrasting predictions from three hypotheses that describe evolved patterns of nucleotide substitution at caste-associated loci in eusocial insects. The terms 'nearly neutral' and 'genetic release' hypothesis are as used by Linksvayer and Wade (2009) and by Gadagkar (1997), respectively. The other terms are my own, as are the descriptions based on my interpretation of the listed references. More detailed explanations are available in the text.

Hypothesis	General prediction	References
Relaxed worker hypothesis		
Also known as the 'Nearly neutral' hypothesis, selection acts less effectively on loci with indirect social effects than on loci with direct effects.	Relaxed selection on genes associated with worker traits such that they more closely approach the neutral rate than genes associated with reproductive castes.	Linksvayer & Wade 2009 Linsvayer et al. 2016 Mikheyev & Linksvayer 2015 Hall & Goodisman 2012 Warner et al. 2017 Helantera & Uller 2014
Adapted worker hypothesis		
Environmental selection is strongest on loci associated with helper castes than it is on queens that are buffered from the outside environment by her workers.	Strong selection on genes associated with worker traits such that they depart more from the neutral rate than genes associated with reproductive castes.	Harpur et al 2014 Helantera & Uller 2014 Vojvodic et al 2015
Genetic release hypothesis		
Strength and possibly direction of selection on loci associated with caste can change once genes are released from the constraints of multi-caste pleiotropy.	Strong or relaxed selection on genes associated with any caste, and the precise pattern depends on which gene is 'released' from pleiotropy to take-on a caste-specific role.	Gadagkar 1997 Hunt et al. 2011

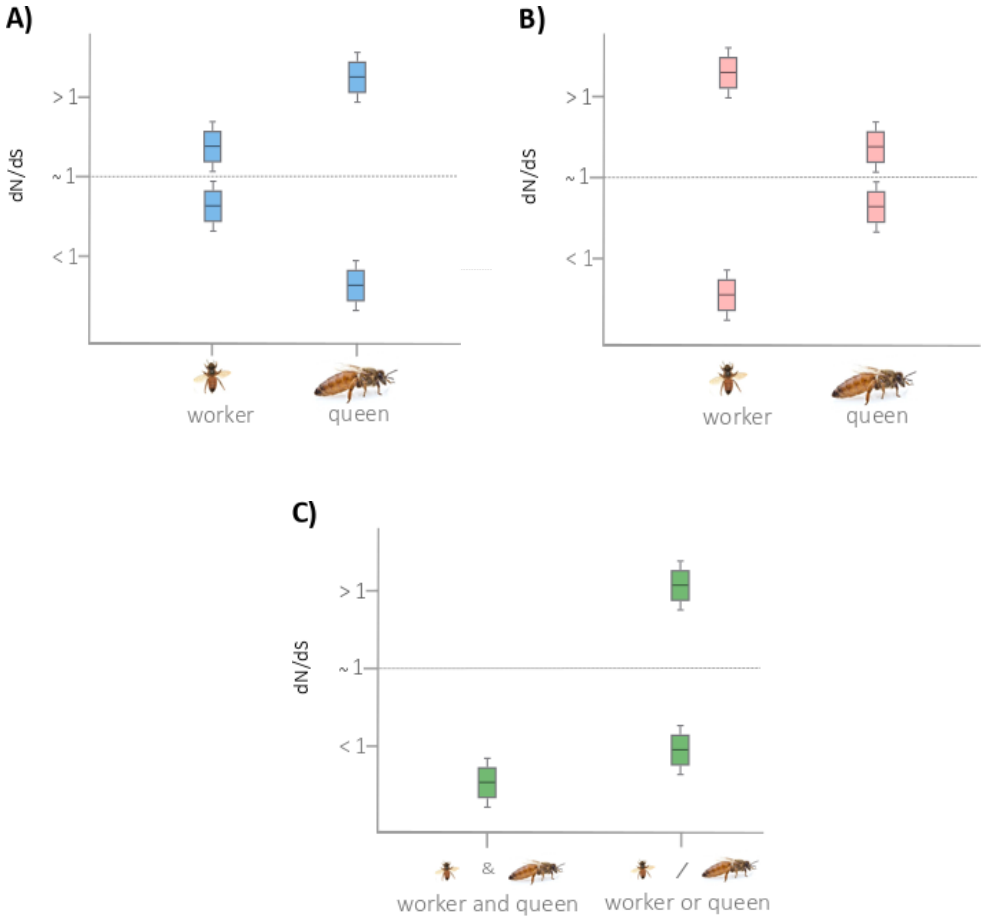

Figure 3.1. Schematic diagram displaying essential differences between three competing hypotheses that describe how caste-associated genes might evolve in social Hymenoptera. These boxplots are hypothetical and represent my interpretation of ideas stemming principally from Linksvayer and Wade (2009 - Plot **A)**, Harpur et al. (2014 - Plot **B)** and Gadagkar 1997 (Plot **C)**, and others (**Table 3.1**).

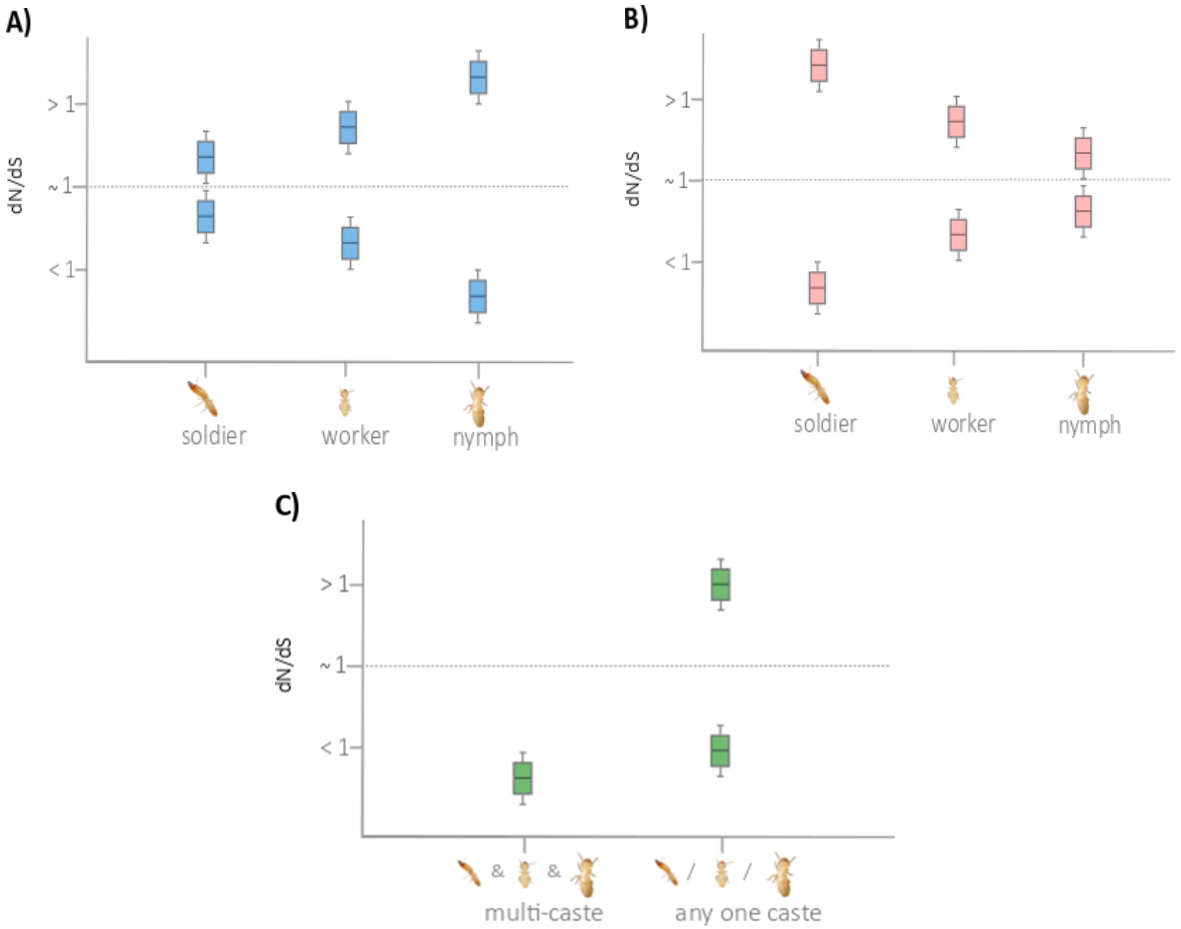

Figure 3.2. Schematic diagram showing the essential differences between three competing hypotheses that describe how caste-associated genes might evolve in the social Blattodea (termites). The box plots are hypothetical and show how the strength and direction of selection might vary as a function of direct vs. indirect selection. **A)** According to the **relaxed worker** hypothesis, genes associated with strongly altruistic castes (i.e., soldiers) ought to experience the weakest selection and thus evolve closest to the neutral rate (d_N/d_S ~1), whereas castes with increasingly selfish reproduction epitomized by the king and queen ought to experience the strongest selection and deviate from the neutral rate in either the positive (d_N/d_S >1) or purifying (d_N/d_S <1) direction, depending on the gene. **B)** The **adapted worker** hypothesis predicts an essentially opposite pattern, and **C)** the **genetic release** hypothesis predicts an altogether different pattern (see text for explanation). Finally, in all cases my null hypothesis is for no association between strength of selection and caste.

3.2. Methods

3.2.1. Code and Data Manipulations

For my analysis below, I used a combination of UNIX, PYTHON and R scripts that I developed myself or modified from others, as described in the Appendix. To start, I used the UNIX file system to set-up a directory on the Compute Canada Cedar server: www.computecanada.ca. Next, I uploaded a raw (dis-assembled) text version of the Eastern subterranean termite transcriptome and converted it to FASTA format. The resulting text file is massive, containing a total of *n = 29, 641* transcripts that correspond to *n = 13, 755* predicted genes. The transcripts were initially compiled by Wu et al. (2018) from a total of ~91M raw reads, generated via 100 bp paired-end Illumina Hi-Seq 2000 mRNA sequencing runs. This RNA-Seq dataset is organized into *n = 9* libraries that correspond to three morphological castes (nymph, soldier and worker) sampled from three separate North American populations (Boston, Raleigh, Toronto). For each population, Wu et al. (2018) pooled whole-body tissue samples from one male and one female termite per caste from each of three geographically separate colonies. The entire RNA-Seq dataset captures sequence variation from a grand total of *n=54* individuals, as described in **Table 3.2**. To enable broader use of the *R. flavipes* transcriptome, I have up-loaded all nine library files to my laboratory's public GITHUB repository: https://github.com/SocialBiologyGroupWesternU/R.flavipes_Transcriptome.

3.2.2. Differentially Expressed Genes

Wu et al. (2018) identified *n = 93* genes with strong (FDR-corrected *P*-value < 0.001, expression fold-change ≥ 4) caste-biased expression. These DIFFERENTIALLY EXPRESSED (DE) GENES were tightly co-regulated within three distinct sets (Figure 3). Gene Sets I and III were uniquely up- and down-regulated in the soldier caste, respectively, whereas gene Set II was uniquely up-regulated in nymphs. The worker caste did have a unique gene expression profile, but no distinct set of genes was uniquely dysregulated in that caste. Wu et al. (2018) worked with invasive populations and thus did not sample kings and queens. For these reasons, my tests will focus on one reproductive (nymph) and one sterile (soldier) caste.

3.2.3. Molecular Dataset Assembly

Using a custom PYTHON V.3.9.0 script (Appendix, Script 1), I first grouped all *n = 29, 641* transcripts in the FASTA file according to their gene-identifying accession numbers. I then split the gene-wise information into two FASTA files corresponding to genes differentially expressed by caste (*n = 93*) and genes NON-DIFFERENTIALLY EXPRESSED (NDE; *n = 13, 185*). Next, from the list of NDE genes, I randomly selected (Appendix, Script 2) an equivalent number (i.e., *n = 93*; corresponding to *n = 465* transcripts) of genes to serve as a comparative reference against the DE Set. I then manually filtered the same-sized DE and NDE Sets to remove sequence duplicates, individual sequences that were less than 100 nucleotides, as well as transcripts that were characterized by low (< 100 nucleotide base pairs) or no overlap with the rest of the sequences in each alignment. This filtering step removed six

genes and associated transcripts from the DE list, resulting in a final quality-controlled total of *n = 87* genes (*n = 236* transcripts). Similarly, this filtering removed two genes from the NDE list, resulting in the final quality controlled NDE dataset with *n = 91* genes (*n = 432* associated transcripts).

3.2.4. Raw Gene Alignments

Prior to performing downstream tests of selection (Yang et al. 2000), it was first necessary to align transcripts for each DE and NDE gene at the codon level. To this end, I used the ORF-FINDER (Rombel et al. 2002) program in combination with the 'get_orfs_or_cdss.py' script (Appendix, Script 3; Cock et al. 2009) to identify the longest frame (in 5' to 3' orientation) of each transcript. I then utilized a set of UNIX commands (Appendix, Script 4) to cluster all the nucleotide and corresponding amino acid transcripts into functionally related GENE FAMILIES. This step yielded two paired sets of *n= 87* and *n= 91* nucleotide and protein FASTA files, with each file containing only *R. flavipes* sequences specific to each gene. Next, I aligned all *n = 178* paired sequence sets at the nucleotide and amino acid levels using CLUSTAL OMEGA (Sievers et al. 2011) with default parameters (gap extension penalty = 0.05, gap opening penalty = 10, and using the BLOSUM amino acid substitution matrix), as implemented in UNIPRO UGENE v.35 (Okonechnikov et al. 2012). By toggling back-and-forth between the nucleotide and protein levels, I was able to manually adjust the nucleotide sequences to generate raw codon-by-codon alignments for each DE and NDE *R. flavipes* gene.

Table 3.2. Summary of termite sampling used to generate my RNA-Seq data set, along with information on the NCBI Library name and Sequence Read Archive accession numbers.

Population	Caste	Colony	Sex of sampled individuals		Total No. samples	NCBI Library	SRA Accession
Boston	worker	A	M	F			
		B	M	F	6	Boston-worker	SAMN06579174
		C	M	F			
	soldier	A	M	F			
		B	M	F	6	Boston-soldier	SAMN06579171
		C	M	F			
	nymph	A	M	F			
		B	M	F	6	Boston-nymph	SAMN06579168
		C	M	F			
Raleigh	worker	A	M	F			
		B	M	F	6	Raleigh -worker	SAMN06579175
		C	M	F			
	soldier	A	M	F			
		B	M	F	6	Raleigh -soldier	SAMN06579172
		C	M	F			
	nymph	A	M	F			
		B	M	F	6	Raleigh -nymph	SAMN06579169
		C	M	F			
Toronto	worker	A	M	F			
		B	M	F	6	Toronto -worker	SAMN06579176
		C	M	F			
	soldier	A	M	F			
		B	M	F	6	Toronto -soldier	SAMN06579173
		C	M	F			
	nymph	A	M	F			
		B	M	F	6	Toronto -nymph	SAMN06579170
		C	M	F			
Total					54		

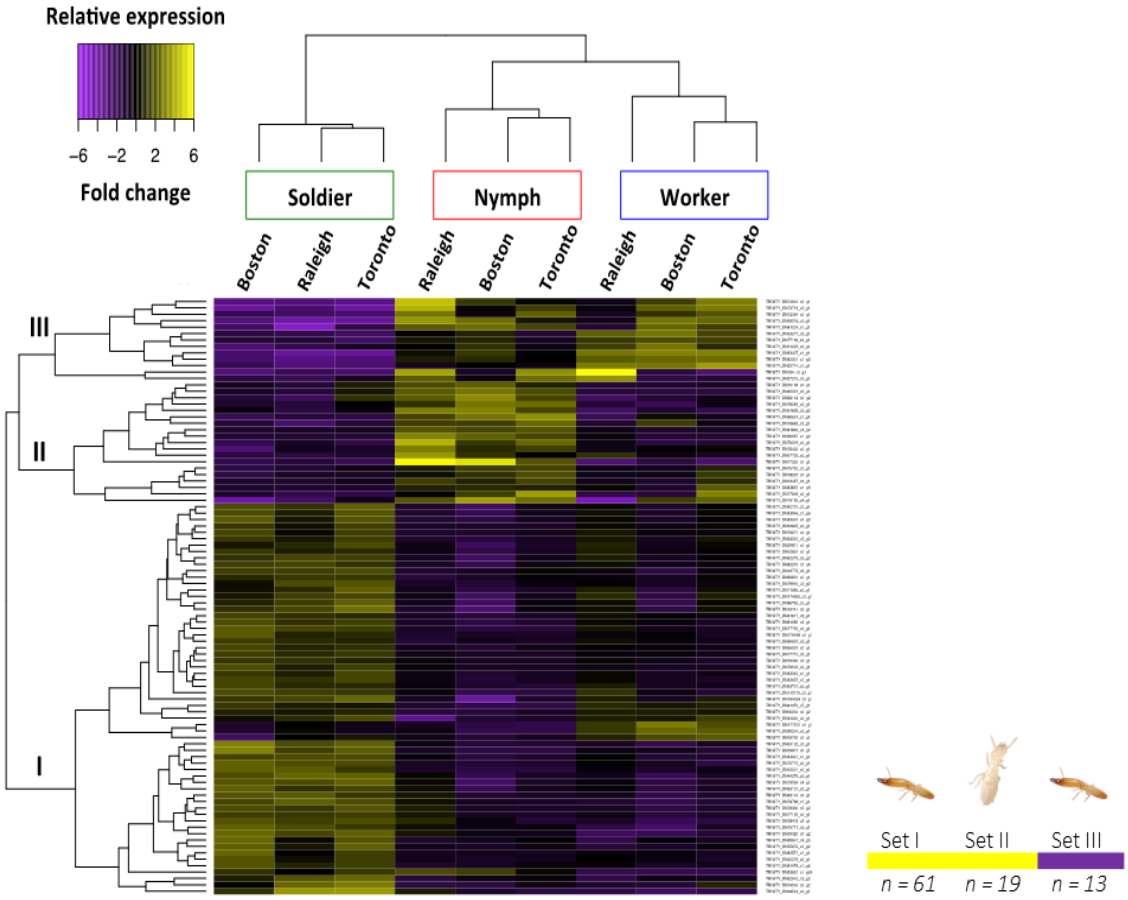

Figure 3.3. Heatmap of genes differentially expressed by caste. Adapted from Wu et al. (2018), this heatmap shows *n = 93* genes (right-most panel) that are co-expressed by caste (top panel) in three distinct sets (left-most panel). Gene Sets I and III are uniquely up- and down-regulated in soldiers, respectively. Set II is, by contrast, uniquely up-regulated in the nymph caste.

3.2.5. Adding Outgroups to Gene Family Alignments

To extend my analysis beyond one species, I amplified *R. flavipes* alignments for each gene to include homologs of two annotated and published termite genomes corresponding to *Cryptotermes secundus* (Csec_1.0; NCBI Assembly Accession no. = GCF_002891405.1) and *Zootermopsis nevadensis* (ZooNev1.0; NCBI Assembly Accession GCF_000696155.1). To achieve this for each gene, I used the longest (aa) transcript

available from *R. flavipes* to query both genomes and retrieve the top SMARTBLAST hits. Generally, I was able to match each query to a 1:1 ORTHOLOG (e-value ≤ 1e −6 and minimum 60 % sequence identity) for both outgroups simultaneously. For some genes, however, I retrieved only one or no outgroup sequences, in which case I simply selected any that were available. I therefore generated a total of three robust and complementary gene sets, with a comparable number of nucleotide and protein sequences for each of the three termite species. For reference, the phylogenetic relationship between these three species is shown in Figure 3.4.

Next, I imported each set of retrieved outgroup sequences per gene into NCBI's multiple sequence alignment program COBALT (Papadopoulos and Agarwala 2007) to produce a codon-by-codon alignment of the outgroups via detection and matching of conserved sequence motifs and domains. The resulting position-fixed alignment with optimized placement of INDELS between the outgroup sequences was exported in CLUSTAL format and up-loaded into UNIPRO UGENE, to serve as reference against the corresponding *R. flavipes* nucleotide and peptide FASTA files for each individual DE and NDE gene.

Here, I manually calibrated multi-species gene alignments, re-positioning any INDELS to fall between and not within triplet codons, as defined by the genetic code. Ultimately, I cut over-hanging (non-aligned) transcript fragments, as well as highly variable and ambiguous multi-species gene alignment regions. The resulting gene-wise alignments for DE and NDE datasets contained *R. flavipes* transcripts together with the sequences from one or both *C. secundus* and *Z. nevadensis* outgroups, where available (Figure 3.5).

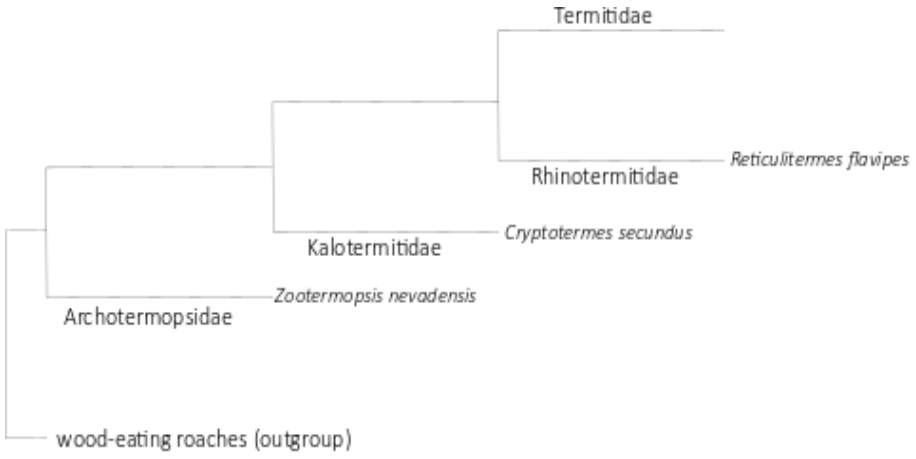

Figure 3.4. Phylogenetic relationship between the three termite species in this study, as inferred from Bourguignon et al. (2014) by taxonomic family.

3.2.6. Assessing Allelic Richness and d_N/d_S

To estimate the average d_N/d_S value for each gene, I first visually examined each codon-by-codon alignment for nucleotide and amino acid sequence diversity. Next, using UNIPRO UGENE I measured allelic richness across DE and NDE genes in all three termite species by sequentially setting any sequence variant or ALLELE present among my ingroup sequences as reference against which all intra-specific polymorphisms or inter-specific differences at both nucleotide and amino acid base levels get highlighted (**Figure 3.5**). This method allowed me to determine the total number of allelic variants present in both ingroup and outgroup sequences for each gene. Next, for each pair of alleles, where present, I split the multiple sequence file into sets of paired FASTA files corresponding to sequences that are to be compared at both intra- and inter-specific variation. For each gene family, a maximum of six paired FASTA files were possible. These are: *R. flavipes* vs. *R. flavipes, C. secundus* vs. *C. secundus, Z. nevadensis* vs. *Z. nevadensis* and *R. flavipes* vs. *C. secundus, R.*

flavipes vs. *Z. nevadensis*, and *C. secundus* vs. *Z. nevadensis*. I then imported pairs of FASTA sequences into the PAL2NAL program (Suyama et al. 2006), which implements CODEML within PAML (Yang 1997) to perform tests for selection in a pairwise manner. The d_N/d_S estimates generated from each pairwise comparison along with allele counts and gene descriptions for *R. flavipes* are listed in (**Table 3.3**). Further, they are uploaded as *caste_clusters_gene_data.xlsx* on our laboratory's GITHUB server accessible from the following link:

https://github.com/SocialBiologyGroupWesternU/R_flavipes_variation_estimates/tree/main

3.3. Results

3.3.1. Intra-specific analyses

My intra-specific alignment of *R. flavipes* transcripts identified 68 (of 87) genes that were invariant within the DE dataset and 56 (of 91) genes that were invariant within the NDE set. As expected, my intra-specific data set is not highly variable or saturated with substitutions but rather is typical of a sparsely sampled inbred population. The remaining 19 (DE) and 37 (NDE) genes were, however, variable as represented by nucleotide polymorphisms across two or more conspecific transcripts (**Figure 3.5**). The total of 19 variable genes in my DE gene set are represented by between two and seven alleles (**Table 3.3**) and are suitable for my proposed analysis of sequence variance. Gene Sets I and III are defined as up- and down-regulated in soldiers, respectively (**Figure 3.3**), and together these two sets contain the vast majority of variable genes (17 of 19). These genes include

those that encode *general odorant-binding proteins*, *hexamerin II*, *junctophilin*, three versions of *troponin* and a *twitchin*, among others. When aligned to each other, and to available homologs in GenBank, the high-quality alignment length ranges from 69 (*junctophimin*) to 3252 (*obscurin*) codons, with a majority (12 of 19) showing variation at both synonymous and non-synonymous sites (Table 3.3). These 12 genes – six genes in Set I, two genes in Set II and four genes in Set III – are therefore the focus of my proposed hypothesis testing, as first described in Figure 3.2.

My estimates of d_N/d_S range in value to suggest that 12 variable genes evolve in a manner consistent with strong purifying selection (d_N/d_S = 0.05, *tropinin I*) to slightly positive directional selection (i.e., d_N/d_S = 1.04, feruloyl esterase-like protein; d_N/d_S = 1.079, *titin*) with other values falling in between, depending on the gene and gene set (Table 3.3). Overall, there is not an obvious difference between strength or direction of selection affecting caste-associated genes. Figure 3.6 A (see *'R. flavipes'*) shows that genes associated with reproductive (Set II) and non-reproductive (Sets I and III) castes evolve at similar rates, approximating a grand average d_N/d_S value of ~ 0.46 (SD = 0.334).

The relatively few genes from the initial set of *n = 93* genes identified by Wu et al. (2018) that are suited to this analysis render my main statistical analysis lacking in power: genes in Set II only have *n=2* d_N/d_S datapoints. However, Sets I and III appear to evolve at similar rates (Wilcoxon rank sum test W = 10, df = 1; P = 0.428). Moreover, the two values associated with reproductive nymphs (Set II) are quite different from each other (0.266

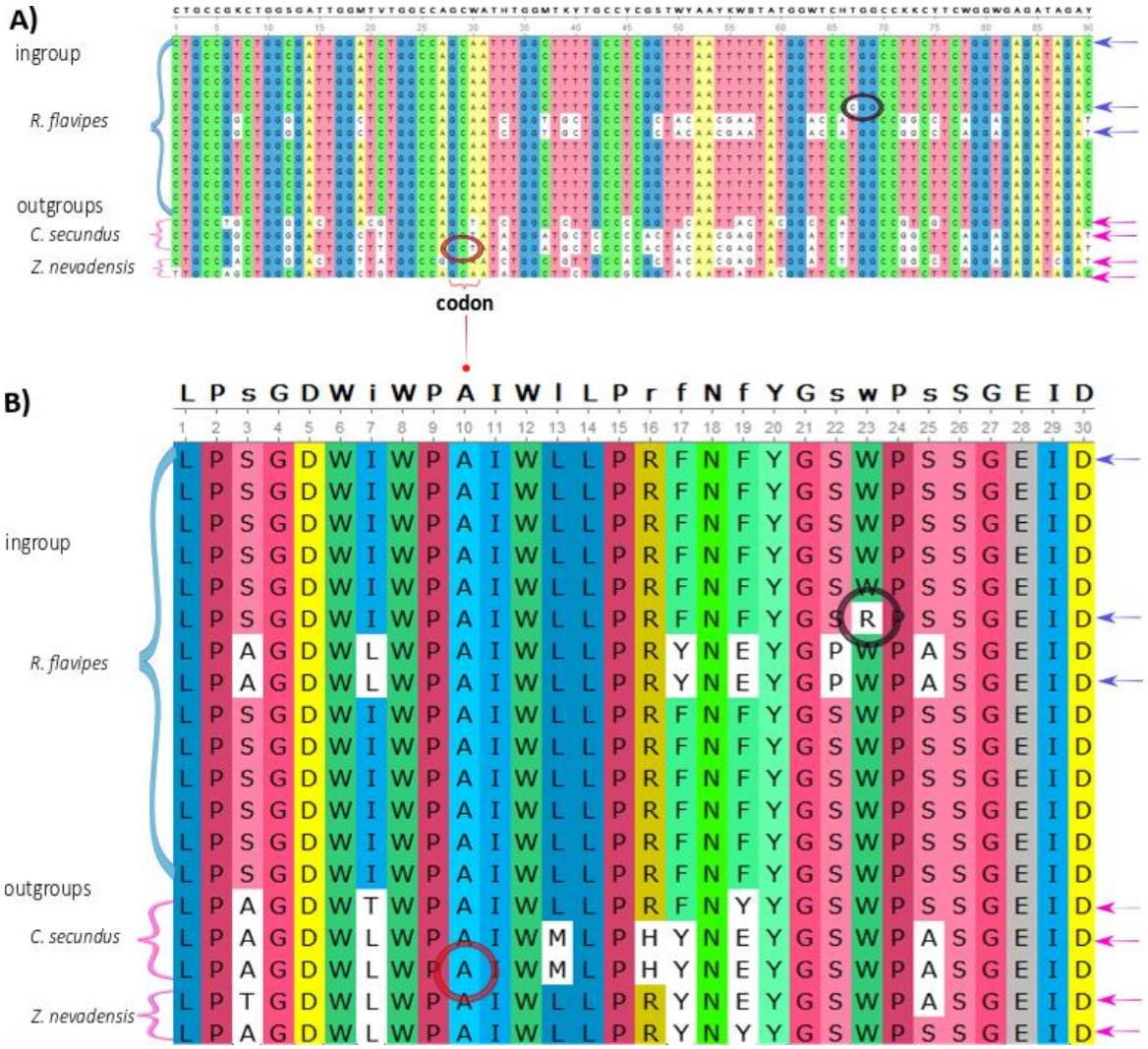

Figure 3.5. Section of a multiple sequence alignment with 90 nucleotide bases displayed in panel **A)** and corresponding translation of 30 amino acid bases shown in **B)** for a *beta-1,3-glucan binding protein (GBP)* gene (*R. flavipes* gene ID number: DN83427). The differences between base positions among all sequences are highlighted in white at both nucleotide and protein alignments. The ingroup set of *n=14* sequences correspond to *R. flavipes* with some unique sequences or 'alleles' visible (blue arrows). For this gene, a total of three from *C. secundus* and two from *Z. nevadensis* outgroup sequences were retrieved from GenBank. Here, both *C. secundus and Z. nevadensis* have two alleles (pink arrows). The black circle shows an example of how some nucleotide substitutions are non-synonymous and thus lead to a difference in amino acid sequence. The red circle displays one non-synonymous substitution, which is visible at the nucleotide sequence alignment but results in no amino acid change.

Table 3.3. Summary of molecular sequence diversity for caste-biased genes in *R. flavipes*. This table includes only those *n= 19* genes that have with at least one polymorphism at a synonymous (*S*) or non-synonymous (*N*) site. The number of variants ('alleles') and number of codons is estimated from my ingroup-only gene-wise alignments. My estimate of the synonymous (d_S) and non-synonymous (d_N) rate is derived from a single pairwise sequence comparisons between the two (or, most divergent two) alleles. For genes with non-zero d_S and d_N, I estimated the average d_N/d_S ratio for each gene. Other transcripts that make-up the remainder of the DE set are invariant for in-group transcripts and not shown. Full details for every gene as well as information on their homologs in *C. secundus* and *Z. nevadensis* is available on the Social Biology Group's GITHUB server[1].

Gene ID	Gene Annotation	Alleles	Codons	*S*	*N*	d_N	d_S	d_N/d_S
Gene Set III								
DN61229	general odorant-binding protein 56d	2	99	54	243	0	0.021	-
DN82301	beta-1,4-endoglucanase	3	441	321	996	0.083	0.243	0.343
DN83277	beta-1,3-glucan-binding protein	2	160	94	509	0.008	0.020	0.408
DN82714	beta-glucosidase	5	467	379	1020	0.065	0.132	0.492
DN83427	gram-negative bacteria binding protein	7	172	129	387	0.214	1.96	0.109
DN81204	hexamerin II	2	124	31	341	0.012	0	-
Gene Set II								
DN75166	P protein	2	567	428	1268	0.002	0.007	0.266
DN79180	hexamerin	2	481	228	1212	0.006	0.009	0.638
Gene Set I								
DN83140	GPI-anchored adhesin-like protein	2	142	153	272	0.004	0	-
DN70903	junctophilin	2	69	54	153	0.013	0	-
DN83123	hypothetical protein B7P43	2	218	161	460	0.016	0.038	0.413
DN75717	troponin C-like	2	146	88	350	0.022	0	-
DN83345	obscurin	2	3252	2154	7602	0	0.001	-
DN82278	titin isoform X1	2	818	604	1835	0.002	0.002	1.079
DN81478	ras-related growth inhibitor	2	213	130	509	0	0.008	-

DN77755	hypothetical protein B7P43	2	162	144	339	0.009	0	-
DN69320	troponin C	2	141	91	326	0.006	0.053	0.115
DN82279	troponin I	2	85	38	217	0.009	0.185	0.053
DN83701	twitchin	2	1730	1168	4022	0.004	0.007	0.552
DN69767	feruloyl esterase-like protein	2	163	146	343	0.030	0.029	1.040

[1] https://github.com/SocialBiologyGroupWesternU/R_flavipes_variation_estimates/tree/main

vs. 0.638) and are not clearly drawn from a different sampling population than are the ten genes associated with non-reproductive soldiers (Sets I and III). It is therefore not easy to statistically evaluate my data against the three alternative hypotheses outlined in the Introduction.

A comparison of evolutionary rates between all caste-associated DE genes (Sets I, II, III combined) and the caste-unassociated NDE set reveals a non-significant difference (Wilcoxon rank sum test $W = 112$, df = 1; $P = 0.226$), suggesting that caste may not affect the rate of molecular evolution and that signatures of indirect selection, if they exist, may be subtle or complex to detect. Nonetheless, it is apparent that the DE set in its entirety is evolving on average more slowly than the NDE set (**Figure 3.6 A**), indicative of slightly more intense purifying selection of DE relative to NDE genes.

My intra-specific analysis of *Cryptotermes* (**Figure 3.6 B**) and *Zootermopsis* (**Figure 3.6 C**) species yielded a consistent pattern – namely, caste-associated genes appear to be evolving under strong purifying selection relative to caste-unassociated genes (*C. secundus*, Wilcoxon rank sum test $W = 32$, df = 1, $P = 0.0013$; *Z. nevadensis*; Wilcoxon

rank sum test $W = 0$, df = 1, $P < 0.0001$). In these latter two comparisons, the genes are not strictly known to be caste-associated but inferred as such via homology to genes in the DE (and NDE) sets derived from *R. flavipes*.

Finally, Figure 3.7 shows how the variation in my sequence alignments is partitioned among gene sets. For example, the proportion of genes for which d_N/d_S estimates were unknown (red) represent genes not found in one or more termite species. Conserved genes are characterized by complete absence of variation (green), genes with a positive (non-zero) value for either d_S or d_N, but not both (aqua), and variable (purple) represent proportions of genes for which I was able to confidently able to estimate d_N/d_S. Inter-specific alignments are predictably more variable. A broad comparison of variation between DE and NDE sets shows that they are quite similar, which suggests differences in d_N/d_S ratio between these two sets (Figure 3.6) is genuinely due to different selection pressures and not simply an artifact of allelic sampling.

3.3.2 Inter-specific analysis

Overall, my inter-specific analyses showed that genes in all datasets evolve at a neutral rate (i.e., $d_N/d_S \sim 1$). The inter-specific d_N/d_S comparison between three termite species suggests that gene Set III has a unique profile, at least in comparisons involving *R. flavipes* (*R. flavipes* vs. *C. secundus* and *R. flavipes* vs. *Z. nevadensis*). Genes associated with the soldier caste of *R. flavipes* may therefore evolve under a different selective regime or may in some way be more unique than genes in other datasets. Given that the largest

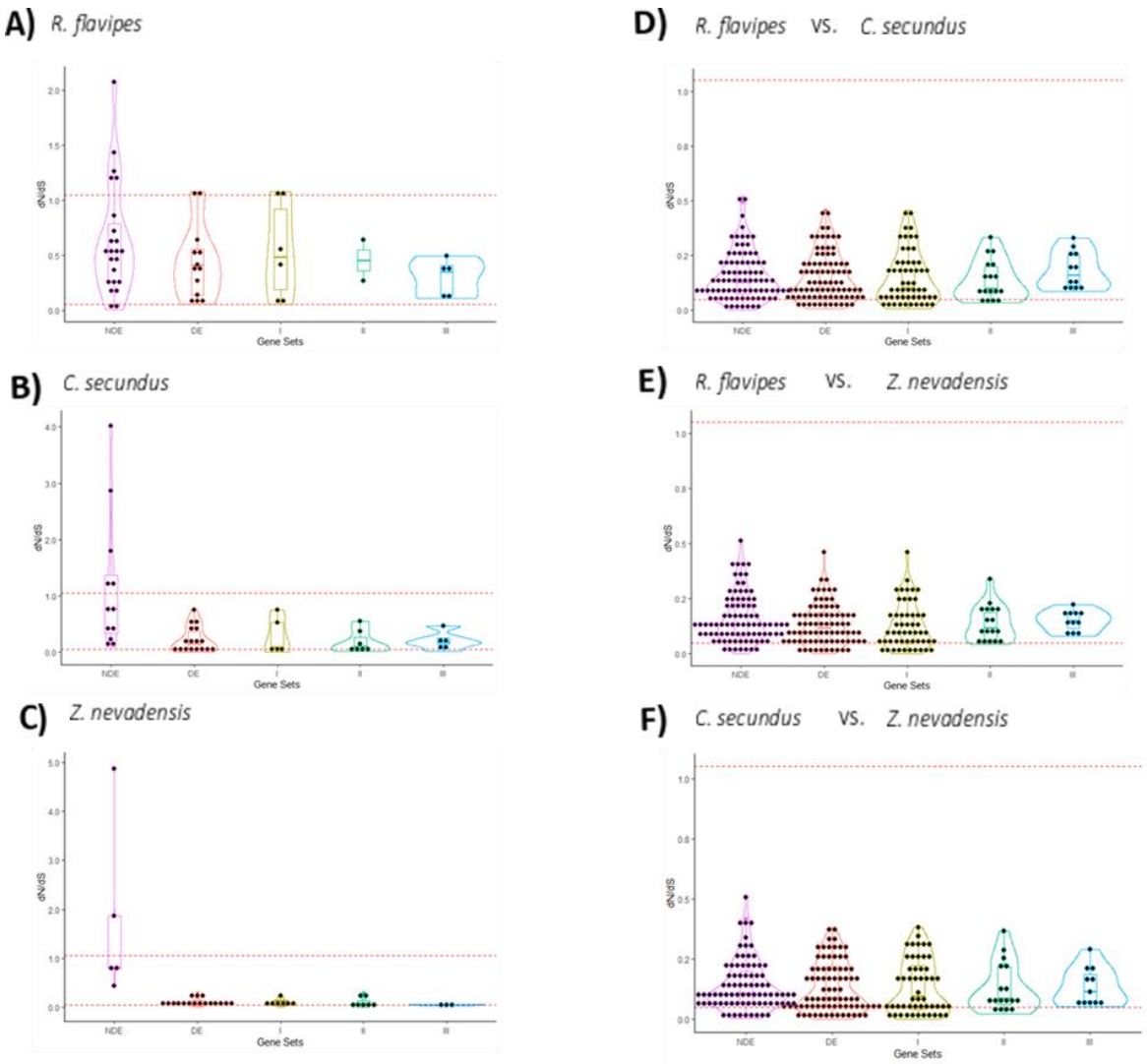

Figure 3.6. Violin plots showing average d_N/d_S ratios. Intra- **A) - C)** and inter- **D) - F)** specific values genes that are DE or NDE by caste. DE genes are associated in their up- (Set I) or down-ward (Set III) expression with the sterile soldier caste or in up-ward expression with the reproductive nymph caste (Set II). The red lines indicate cut-off values (placed at 95% confidence interval of d_N/d_S=1) for positive (>1.05) or purifying (<0.05) selection. Note: Gene Set II in Panel A has only two points so it is displayed as a box-plot).

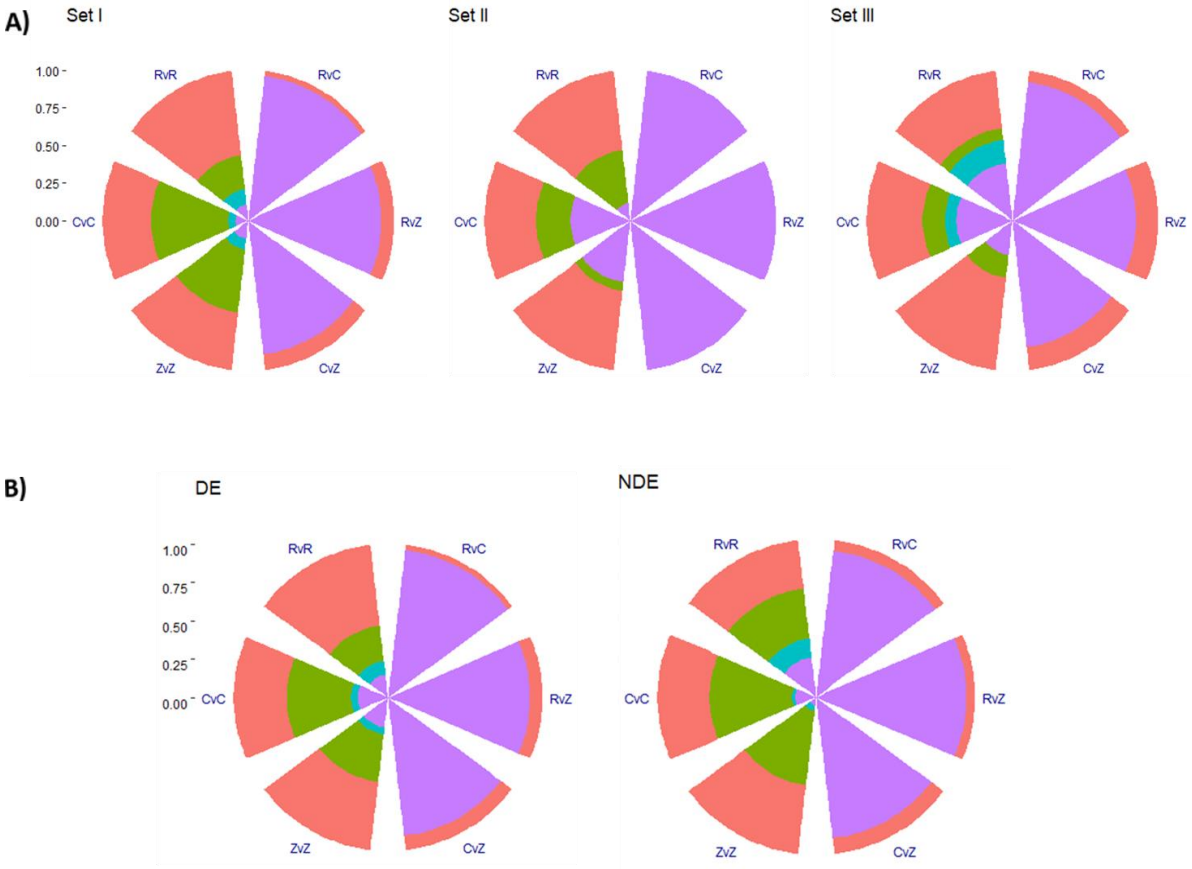

Figure 3.7. Patterns of nucleotide diversity. The circular stacked bar plots are showing the proportion of variable genes in intra-specific (R vs. R; C vs C; Z vs. Z) and inter-specific (C vs. Z; r vs Z; R vs. C) pairwise alignments for **A)** gene Sets I, II and III and **B)** DE vs. NDE sets. For intra-specific alignments the proportion of truly variable genes (purple) for which a d_N/d_S ratio can be inferred is small relative to genes invariant among ingroup transcripts. Proportion of genes for which d_N/d_S estimates were unknown (red) represent genes not found in one or more termite species. Conserved genes (green) are characterized by complete absence of variation (green), semi-variable (aqua) are genes with either d_S or d_N, but not both, and variable (purple) represent proportions of genes for which d_N/d_S estimates were made. Inter-specific alignments are predictably more variable. Proportional scale from 0.00-1.00 on the left of the plot.

differences in d_N/d_S distributions were determined for gene Set III, and that only Set I had *n = 10* genes under purifying selection, suggests that genes associated with soldier caste in *R. flavipes* are again in some way more unique than the rest of the genes.

3.4. Discussion

In this study I employed a bioinformatics approach to test alternative hypotheses describing the molecular evolution of genes purportedly under direct vs. indirect selection. Specifically, I took advantage of several genomics datasets that have recently (within five years) become available for several species of termite. Termites are eusocial insects that live in kin-based societies with a pronounced division of labour between reproductively selfish and reproductively altruistic castes. I found that for one species of subterranean termite (*Reticulitermes flavipes*) the intensity of molecular evolution was not different between genes biased in their expression towards reproductive (nymphs) or non-reproductive castes (soldiers). This relatively uniform pattern of nucleotide substitution between nymph- and soldier-associated genes suggests that direct and indirect selection affect genes similarly and apparently do not leave diagnostic signatures detectable through analysis of d_N/d_S ratios, at least not in an obvious manner with the data and analyses performed here.

I did, however, detect other patterns of nucleotide substitution that suggest caste can affect the intensity and rate of molecular evolution. First, four DE genes (of 12) from soldier-biased Gene Set I appear to experience the most intense selection in both positive

and purifying directions. These four genes are uniquely up-regulated in sterile soldiers and thus presumably evolve via indirect genetic effects that increase the fitness of others who carry but do not express these genes. This association between the level of gene expression and strength of selection is highly relevant to several sociogenomic hypotheses that are discussed below. Moreover, the rate of molecular evolution across some eukaryotic genomes (Pal et al., 2001; Krylov et al., 2003) and in mammals (Subramanian and Kumar 2004) has been reported to have an inverse relationship with the amount or intensity of gene expression. That is, the most expressed genes actually tend to show the least number of substitutions or mutations and evolve more slowly, which is consistent with the observed cumulative effects of purifying selection on caste-associated genes in this study. Second, through comparisons of genes in two other termite species *C. secundus* and *Z. nevadensis*, I discovered that in general, genes associated with caste do appear to evolve more slowly than those unassociated with caste. This implies that caste-biased expression can mediate the strength of selection and potentially, the mode of genetic transmission either from some mix of direct vs. indirect effects or by constraints imposed via multi-caste pleiotropy in caste sets. By extending my analysis to other species, I showed that overall patterns are similar with most genes evolving at the nearly-neutral rate, despite the large differences in species breeding biology, ecology and phylogenetic position.

3.4.2. Termite genes have low levels of genetic diversity

The relatively homogenous genotypes I observed in this study is not unexpected. The social biology of termites in which just one or a few pairs of sexuals monopolize reproduction within a colony of thousands or millions renders their effective population sizes (N_e) very small (Wright 1978). A low N_e should reduce standing genetic variance due to the fixation (or loss) of alleles in populations under drift. Second, termites tend to inbreed (Shellman-Reeve 1997), either at the point of colony foundation among weekly dispersing sexuals or from within-colony matings among blatant close relatives (Vargo 2019), which is typical for invasive populations of *R. flavipes* that rely on neotenic supplemental or replacement reproductives recruited from within their kin-based colonies (Eyer et al. 2020; Scaduto et al. 2012). My observed measures of d_N/d_S diversity – for example, allele counts (Table 3.3) – are therefore not surprisingly well-below a theoretical maximum. If each colony sampled is independently founded and has an effective population size of two (a male and a female reproductive) then a grand maximum of 108 alleles per locus is technically possible from my multi-population data set (54 heterozygotes). Given that termites tend to have low N_e and inbreed and thus are very unlikely to be heterozygous at more than a few loci, the actual diversity in my data set confirms this maximum.

The allelic variation summarized in Table 3.3 is derived from manually curated pairwise alignments, constructed from initially disassembled RNA-Seq reads. Figure 3.5 shows one such example corresponding to a *beta-1,3-glucan binding protein (GBP)* gene (*R. flavipes*

gene ID number: DN83427) with the highest number of alleles in my dataset. While only a portion of the gene is visible and hence only three alleles could be displayed on a page (blue arrows; Figure 3.5), the entire alignment has a total of *n = 7* alleles (Table 3.3). Another example (not shown) is from *beta-glucosidase* with *n = 5* allelic variants. Both of these genes are from Set III and are thus uniquely downregulated in soldiers. My observation for high allele counts in this one data set suggests that caste may influence rate of evolution and, specifically, that genes in the soldier-specific caste in *R. flavipes* do differ from other sets. That said, my estimates of allelic variation are conservative: First, I removed all regions of uncertainty that tend to occur at the beginning and end of each alignment. I therefore only scored variants if they were present within (or 'inside') the coding region of the gene and thus preceded by corresponding aligned amino acids. The alignments are therefore thorough but for some genes the sequence is understandably incomplete. For this reason, not all of my codon alignments represent the complete length of a gene. Nonetheless, I did have some power to measure and compare rates of nucleotide substitution between caste-biased vs. un-biased genes. In the future, I would like to extend my analysis to a larger data set that contains more genes. Under less stringent criteria (FDR-corrected p-value < 0.05, Expression fold-change ≥ 2), the number of caste-associated genes identified by Wu et al. (2018) increases to *n = 570* genes and does include a worker-specific set. In the present analysis, I chose to study the most stringent gene list to test caste patterns of nucleotide substitution, but this expanded set is available to me for future analysis.

3.4.1 Molecular signatures of kin selection: are caste-associated genes nearly neutral?

To test my predictions outlined in **Table 3.1**, I used the standard metric for estimating the strength and direction of selection, the d_N/d_S ratio. I expected this metric to vary among sets of genes associated with reproductive (nymph) or sterile (soldier) termite castes (**Figure 3.3**). So far support for either set of predictions outlined in the Introduction has been limited almost entirely to the Hymenoptera (**Table 3.1**). Termites provide a powerful model for studying molecular evolution of caste-biased and un-biased genes. My study is the first to use termites as a model and provides a starting point towards a better understanding of selective forces and functionality of genes that drive social evolution.

Under the relaxed worker hypothesis, I predicted that genes associated in their expression with reproductively altruistic castes would be partially buffered from the full strength of selection and thus more closely approach the neutral rate than genes associated with selfishly reproducing castes that presumably evolve under the full strength of direct selection. I found support for this hypothesis. The prevailing theme of my intra-specific evolutionary analysis was purifying selection, which is indeed typical of protein coding genes with conserved domains and functionality, but only a few genes were evolving under strong purifying selection, with majority evolving neutrally. I did detect few other exceptions. Two genes from the solider up-regulated sets (Gene Set I) evolved above the neutral rate (**Figure 3.6**). Given that all other genes in my DE set have d_N/d_S values well below '1', I infer them to evolve under purifying selection (despite very broad confidence intervals and my lack of statistical power to reject a neutral null).

The alternative 'adapted worker' hypothesis (Table 3.1) is generally not supported. Due to lack of statistical power, it was not possible to infer rates of reproductive caste-associated genes (Gene Set II), though the four genes did have d_N/d_S values below and above one. Only five genes in the NDE dataset were shown to evolve at adaptive rate, with the majority of genes evolving at nearly neutral or weakly-purifying rates. Importantly, my results for randomly selected non-differentially expressed genes are quite sensitive because ideally, robust analysis would involve generation of many (i.e., hundreds) pseudo-distributed data sets rather than one that may be biased or contain outliers. Despite knowing this, I still only generated one null dataset because the workload of aligning all the genes is very slow. To do hundreds of data sets would have been prohibitive. Nevertheless, if we use the NDE set as a proxy for multi-caste associated genes – that is, if we assume their non-biased expression in any one caste is because they are expressed in *all* castes – then my data are not consistent with the relaxed worker hypothesis, either. In no case did the NDE show evidence of evolving in a more constrained manner, as if under intense purifying selection. Rather, genes in the NDE set are, if anything, under more relaxed purifying selection. For two of the three species examined (*Cryptotermes* and *Zootermopsis*), this relaxation was significant (Figure 3.6). This is similar to a recent finding by Harrison et al. (2020), where using four termite species, the authors show purifying signature on caste-associated genes with low evidence for distinct signatures on genes under direct vs. indirect selection.

Problematically, the d_N/d_S values can only be estimated from variable genes, thus limiting the sample size and potential information that could be inferred from the gene sets. We thus need other methods of scanning for signatures of caste-mediated selection (Helanterä and Uller 2014). One way to search for the patterns of direct vs. indirect selection is to expand our test to include more information. For instance, including estimates of the degree of gene conservation in each dataset, or genes that either have d_N or d_S but not both. Here, even though such genes are not normally picked up by d_N/d_S, measures or computer algorithms, such genes and their proportions could be visually detected and are potentially informative. The emerging patterns provide a birds-eye view of the datasets (**Figure 3.7**), and thus offer a new way of scanning for signatures of kin selection. I believe this and other ways are necessary and should be included in the detection of selection signatures.

3.4.2. Expanded evolutionary analysis

My expanded analysis against two other termite species yielded additional insights into the molecular evolutionary patterns associated with caste. The *C. secundus* genome is 1.30 Gb in size (Harrison et al., 2018), contains a total of *n = 29, 593* transcripts that correspond to *n = 14, 313* genes, and was assembled from degutted body samples of male and female primary and worker castes. Comparatively, the genome of *Z. nevadensis* is composed of *n = 15, 103* genes and *n= 35, 617* transcripts, and is 652 Mb in size (Terrapon et al. 2014). The sampling procedure of *Z. nevadensis* was similar to *R. flavipes*

(Wu et al., 2018) and genome was assembled from samples that included whole-body tissues of male and female alate, soldier, primary and secondary reproductive castes.

The opportunistic inclusion of these species into my analysis takes advantage of two large published genomic datasets with relatively well curated gene annotations that aided my analysis. The Wu et al. (2018) study provided an essential transcriptomic background that enabled my own analysis but note that no annotated version of the *R. flavipes* genome has ever been generated or published. This lack of detailed gene-level knowledge meant that I did not have a verified reference genome against which to align my disassembled reads. It was still possible for me to generate gene-wise alignments more-or-less manually as described in the Methods section, but it was painstaking and did slow my analysis. Thankfully, my focal species is one of widespread interest due to its status as an invasive pest (Chouvenc et al. 2011; Evans et al. 2013), so my bioinformatics pipeline and analysis of genes underlying its social biology is informative. This species is phylogenetically well-suited for comparative sequence analysis against *C. secundus* and *Z. nevadensis* because all three species are roughly equally separated on the termite tree of life (Figure 3.4). The phylogenetic proximity of my outgroups is therefore important, because known functional domains in annotated *C. secundus* and *Z. nevadensis* sequences are more likely to match those in *R. flavipes* sequences and can be used to guide alignment of non-annotated transcripts.

CHAPTER FOUR: General conclusion

The gene's-eye-view of evolution, which is central to the works of John Maynard-Smith, George C Williams, Ernst Mayr and other proponents of the Modern Synthesis (Pigliucci and Muller 2010), holds that genes are central replicators in the struggle for life and the ultimate target of natural selection, regardless of the vehicles (i.e., organisms) in which they are found (Dawkins 1982). As such, the conserved, yet variable nucleotide composition of individual genes can serve as a portal into their evolutionary past to reveal 'signatures' of historical selection pressures that have shaped the evolved aspects of existent phenotypes (Nielsen 2005). In my thesis, I adopt this gene's-eye-view of evolution and use the idea of molecular signatures of selection to, first, discuss how selection acts on genes to generate diversity of caste systems in eusocial insects (Chapter Two) and, to test how patterns of nucleotide substitution vary as a function of their expression-bias toward reproductively selfish or reproductively altruistic castes (Chapter Three). In summary, my Thesis moves from theory to discovery, from general to specific, and from classic to cutting-edge, generating a novel production that will at once serve to demonstrate my skill and launch my future studies and career as a research and teaching scientist.

Throughout my Thesis, I argue that through a far-reaching review of key papers and theory, selection is most effective when additive genetic variance is high within effectively large populations characterizing eusocial insects. I indicate that there must have been additive genetic variants within social environments, that were selected for

and are driving the evolution of caste differences across ant, wasp, bee, termite and other social insect lineages. Alas the evolutionary perspective that I adopt is essential to the completeness of our understanding of castes, and by extension, to the full understanding of division of labour and origin of eusociality. The main conclusion of my chapters was not just an academic tabulation of acquired knowledge, but rather a unique synthesis and interpretation of knowledge of yet-imperfect understanding of social insect biology and evolution.

One key insight that emerged from my Thesis review was to realise that the treatment of caste differences has largely proceeded along two separate research fronts, either following early ideas on the evolution of caste differences (Oster and Wilson 1978) or following early leads on the development of caste differences (Watson et al. 1985; Brian 1983). Both avenues are interesting and important but, in my view, there had previously been very little synthesis of the two perspectives, which is surprising because they ought to be highly complementary and mutually informing. My attempt at casting (pun intended) these two processes as interdependent within a single 'evo-devo' framework is highly germane to the study of social insects (Toth and Robinson 2007; Ramsay et al. 2020) and is most visible in Figure 2.1. There I characterise caste differentiation in any social insect as a function of 'switch points' that respond to environmental and genetic cues in developmental time. Rather than separating the developmental from the evolutionary process, I show that the responsive switches have themselves evolved under selection over much longer periods of time. This evo-meets-devo realization is not novel

but is, apparently, still widely underappreciated, except by a few leading researchers in this field who are even better positioned than me to make this point. My thesis serves to highlight how the 'evo' and 'devo' components to caste variation can hardly be understood one without the other, and my conceptual framework and diagrams are a unique contribution to this understanding.

A second insight of my review was to reveal how little attention had previously been given to the notion of genetic effects on caste. It has been emphasised, perhaps to a fault, that environmental effects are important for understanding the phenotypic plasticity and developmental trajectories of insect caste systems (West-Eberhard 1987; Revely et al. 2021). My review shows that the emphasis on environmental factors is not wrong, but incomplete and stands to obscure the true diversity of caste systems. Emphasizing environmental plasticity, as if it precluded any involvement of genetic effects, likewise stands to obscure the fascinating ways in which caste and subcastes evolve in response to selection on gene variants or, in some cases, as by-products of gene interactions and hybridization events.

I acknowledge that genetic effects on caste are sometimes subtle and might initially be difficult to detect. In other cases, however, these effects are strong and obvious. Many examples that I cite are discovered serendipitously and appear to be unique to particular species or even populations. I agree with Keller (Keller 2007) who states that social insect biologists should not overlook breeding system diversity that is unexpected or goes

against accepted wisdom or that otherwise appears as an unexpected aberration. Instead, we should look deeply into these findings to characterise the conditions under which it occurs. Many of the highly cited examples of genetic effects on caste that I cite in my review chapter are associated with complex and sometimes unbelievable breeding systems that confound gender, ploidy and sex vs. asexual reproduction. Social insects therefore offer a deep well for discovery but likewise demand a high level of natural history expertise and a savvy understanding of socio-genetic theory.

Finally, my review chapter serves to bridge the gulf between environmental vs. genetics perspectives on caste through summary of the fast-advancing field of epigenetics. For example, the first draft genome of the European honey bee in 2006, demonstrated that *Apis mellifera* has a functional CpG methylation system that is linked to queen-worker differentiation. In particular the pioneering work of Ryszard Maleszka and colleagues (Australian National University, Canberra) has established that queen and workers differed heavily with respect to methylated sites, and that these sites were biased towards nutrient signaling pathways. This insight is intuitive, at least within the context of nutrient-cued caste differentiation in this and other social insect species. Other studies quickly followed from the social Hymenoptera and from termites that too, helped reveal the potential for epigenetic effects on caste. One realization from my review is that many of these studies are not highly replicated, which leads one to cast doubt on the apparent association of methyl marks with caste. I argue, instead that many examples published so far may more readily be explained by spurious inter-sample marks which, on occasion,

can be correlated with caste. To me, who had seen an explosion of epigenetic studies on social insects emerge over a decade, this was surprising, and I now realize that much work remains to be done in order to fundamentally establish the role of methylation and other environmentally acquired gene-expression markers in caste differentiation.

My implicit adoption the gene's-eye-view in Chapter Two does not typically reveal the genes involved, except in a few cases where they had been identified by specific studies. In general, my review above takes the 'phenotypic gambit' that is typical in behaviour ecology to assume genetic effects without actually knowing the genes involved (Grafen 1991). In doing so, I revealed that selection has plenty of scope to shape the highly plastic development of castes as they differentiate into reproductively-selfish vs. reproductively-altruistic forms. In Chapter Three, I move beyond this gambit to analyze specific sets of genes associated with caste differences in a species of subterranean termite. Here, I use the idea of 'molecular signatures' introduced above to not only infer evidence of past selection but to infer evidence of kin selection in the termite gene sequences. In this Chapter, I first outline the promise and potential of identifying molecular signatures of kin - i.e., indirect - selection, then offer to test three uniquely sociobiological hypotheses that describe how that signature appears.

My study is unique in that I am the first to attempt it on any one species of termite. Though the rationale is reasonably adept, I found that the power to detect any signature of kin selection was low, at least with the few variable genes that I was able to manually

align from RNA-Sequence data at hand. Going forward, I have determined new ways to increase analytical power, either by *i* - including a larger number of genes (i.e, from the expanded dataset containing ~570 genes) to generate gene alignments and *ii* - employ tests other than those involving d_N/d_S ratios in favour of population genomic tests that might be suited to my data, including tests that make use of the observed allele frequency spectra of my population, such as the McDonald-Kreitman test, the Neutrality Index measure, SNP counts, transition/transversion ratios and epigenetic modifications that have recently been shown to influence gene activity and alter protein functionality. Tests for indirect vs. direct selection in termites or other eusocial taxa are not yet widespread but I argue that such tests are informative, stand to reveal interesting and highly sought-after patterns of social evolution, and hold key to explaining the origin, evolution, and maintenance of eusociality.

REFERENCES

Abbot P, Abe J, Alcock J, Alizon S, Alpedrinha JAC, Andersson M, Andre JB, van Baalen M, Balloux F, Balshine S, Barton N, Beukeboom LW, Biernaskie JM, Bilde T, Borgia G, Breed M, Brown S, Bshary R, Buckling A, Burley NT, Burton-Chellew MN, Cant MA, Chapuisat M, Charnov EL, Clutton-Brock T, Cockburn A, Cole BJ, Colegrave N, Cosmides L, Couzin ID, Coyne JA, Creel S, Crespi B, Curry RL, Dall SRX, Day T, Dickinson JL, Dugatkin LA, El Mouden C, Emlen ST, Evans J, Ferriere R, Field J, Foitzik S, Foster K, Foster WA, Fox CW, Gadau J, Gandon S, Gardner A, Gardner MG, Getty T, Goodisman MAD, Grafen A, Grosberg R, Grozinger CM, Gouyon PH, Gwynne D, Harvey PH, Hatchwell BJ, Heinze J, Helantera H, Helms KR, Hill K, Jiricny N, Johnstone RA, Kacelnik A, Kiers ET, Kokko H, Komdeur J, Korb J, Kronauer D, Kummerli R, Lehmann L, Linksvayer TA, Lion S, Lyon B, Marshall JAR, McElreath R, Michalakis Y, Michod RE, Mock D, Monnin T, Montgomerie R, Moore AJ, Mueller UG, Noe R, Okasha S, Pamilo P, Parker GA, Pedersen JS, Pen I, Pfennig D, Queller DC, Rankin DJ, Reece SE, Reeve HK, Reuter M, Roberts G, Robson SKA, Roze D, Rousset F, Rueppell O, Sachs JL, Santorelli L, Schmid-Hempel P, Schwarz MP, Scott-Phillips T, Shellmann-Sherman J, Sherman PW, Shuker DM, Smith J, Spagna JC, Strassmann B, Suarez AV, Sundstrom L, Taborsky M, Taylor P, Thompson G, Tooby J, Tsutsui ND, Tsuji K, Turillazzi S, Ubeda F, Vargo EL, Voelkl B, Wenseleers T, West SA, West-Eberhard MJ, Westneat DF, Wiernasz DC, Wild G, Wrangham R, Young AJ, Zeh DW, Zeh JA, Zink A (2011) Inclusive fitness theory and eusociality. Nature 471:E1-E4.

Abe T (1991) Ecological factors associated with the evolution of worker and soldier castes in termites. Ann Entomol 9:101-107.

Akashi H (1999) Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. Genetics 151:221-238.

Alcock J (2001) The triumph of sociobiology. Oxford University Press, New York.

Alexander RD (1974) The evolution of social behavior. Annu Rev Ecol Syst 4:325-384.

Anderson KE, Linksvayer TA, Smith CR (2008) The causes and consequences of genetic caste determination in ants (Hymenoptera: Formicidae). Myrmecol News 11:119-132.

Anderson M (1984) The evolution of eusociality. Annu Rev Ecol Syst 15:165-191.

Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ (2016) Evolution of DNA methylation across insects. Mol Biol Evol 34:654-665.

Beye M, Hasselmann M, Fondrk MK, Page RE, Omholt SW (2003) The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. Cell 114:419-429.

Bonasio R, Li Q, Lian J, Mutti NS, Jin L, Zhao H, Zhang P, Wen P, Xiang H, Ding Y (2012) Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Curr Biol 22:1755-1764.

Bourguignon T, Lo N, Cameron SL, Šobotník J, Hayashi Y, Shigenobu S, Watanabe D, Roisin Y, Miura T, Evans TA (2014) The evolutionary history of termites as inferred from 66 mitochondrial genomes. Mol Biol Evol 32:406-421.

Bourke AFG (2011a) Principles of Social Evolution. Oxford Series in Ecology and Evolution. Oxford University Press, New York.

Bourke AFG (2011b) The validity and value of inclusive fitness theory. Proc R Soc B-Biol Sci 278:3313-3320.

Bourke AFG, Franks NR (1995) Social Evolution in Ants. Princeton.

Brian MV (1983) Comparative aspects of caste differentiation in social insects. In: Watson JAL, Okot-Kotber BM, Noirot C (eds) Caste differentiation in social insects. Pergamon, Oxford, UK, pp 385-398.

Brune A, Ohkuma M (2011) Role of the termite gut microbiota in symbiotic digestion. In: Bignell DE, Roisin Y, Lo N (eds) Biology of Termites: a Modern Synthesis. Springer Science+Business Media B.V., Dordrecht, pp 439-476.

Charlesworth B (1978) Some models of the evolution of altruistic behaviour between siblings J Theor Biol 72:297-319.

Cheverud JM (2003) Evolution in a genetically heritable social environment. Proceedings of the National Academy of Sciences 100:4357-4359.

Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422-1423.

Costa JT (2006) The Other Insect Societies. The Belknap Press of Harvard University Press, Cambridge MA.

Crespi BJ, Choe JC (1997) Explanation and evolution of social systems. In: Choe JC, Crespi BJ (eds) The Evolution of Social Behavior in Insects and Arachnids. Cambridge University Press, pp 499-524.

Crozier RH, Pamilo P (1996) Evolution of Social Insect Colonies: Sex Allocation and Kin-Selection. Oxford Series in Ecology and Evolution. Oxford University Press, Oxford.

Crozier RH, Schluns H (2008) Genetic caste determination in termites: out of the shade but not from Mars. Bioessays 30:299-302.

Darwin C (1859) The Origin of Species. Murray, London.

Dawkins R (1982) The Extended Phenotype. The Gene as the Unit of Selection. Freeman, San Francisco.

Dupont C, Armant DR, Brenner CA (2009) Epigenetics: definition, mechanisms and clinical perspective. Seminars in Reproductive Medicine 27:351-357.

Elango N, Hunt BG, Goodisman MAD, Yi SV (2009) DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. Proc Natl Acad Sci U S A 106:11206-11211.

Eyer P-A, Blumenfeld A, Johnson L, Perdereau E, Shults P, Wang S, Dedeine F, Dupont S, Bagnères A-G, Vargo E (2020) Extensive human-mediated jump dispersal within and across the native and introduced ranges of the invasive termite *Reticulitermes flavipes*. Pre-print ⟨hal-03003416⟩:DOI: 10.22541/au.160524381.14266240/v1.

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics, 4th edition. Prentice Hall, Essex, England.

Forêt S, Kucharski R, Pellegrini M, Feng SH, Jacobsen SE, Robinson GE, Maleszka R (2012) DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc Natl Acad Sci U S A 109:4968-4973.

Fougeyrollas R, Dolejšová K, Sillam-Dussès D, Roy V, Poteaux C, Hanus R, Roisin Y (2015) Asexual queen succession in the higher termite *Embiratermes neotenicus*. Proceedings of the Royal Society B: Biological Sciences 282:20150260.

Fournier D, Estoup A, Orivel J, Foucaud J, Jourdan H, Le Breton J, Keller L (2005) Clonal reproduction by males and females in the little fire ant. Nature 435:1230.

Franks NR, Dornhaus A, Marshall JA, Dechaume-Mincharmont F (2009) The dawn of a golden age in mathematical insect sociobiology. In: Gadau J, Fewell J (eds) Organization of Insect Societies: From Genome to Sociocomplexity. Harvard University Press, Cambridge, MA, pp 437-459.

Gadagkar R (1997) The evolution of caste polymorphism in social insects: genetic release followed by diversifying evolution. Journal of Genetics 76:167-179.

Gardner A, West SA, Wild G (2011) The genetical theory of kin selection. J Evol Biol 24:1020-1043.

Glastad KM, Gokhale K, Liebig J, Goodisman MA (2016) The caste-and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*. Scientific Reports 6:37110.

Goldman N (1998) Phylogenetic information and experimental design in molecular systematics. Proceedings of the Royal Society of London - Series B: Biological Sciences 265:1779-1786.

Goodisman MAD, Crozier RH (2003) Association between caste and genotype in the termite *Mastotermes darwiniensis* Froggatt (Isoptera: Mastotermitidae). Aust J Entomol 42:1-5.

Grafen A (1991) Modelling in behavioural ecology. In: Krebs JR, Davies NB (eds) Behavioural Ecology: An Evolutionary Approach, vol 3. Blackwell Scientific Publications, Oxford, pp 5-31.

Hall DW, Goodisman MAD (2012) The effects of kin selection on rates of molecular evolution in social insects. Evolution 66:2080-2093.

Hamilton WD (1964) The genetical evolution of social behaviour, I and II. J Theor Biol 7:1-52.

Hamilton WD (1972) Altruism and related phenomena, mainly in social insects. Annu Rev Ecol Syst 3:193-232.

Harrison MC, Chernyshova AM, Thompson GJ (2020) No obvious transcriptome-wide signature of indirect selection in termites J Evol Biol:https://doi.org/10.1111/jeb.13749.

Harrison MC, Jongepier E, Robertson HM, Arning N, Bitard-Feildel T, Chao H, Childers CP, Dinh H, Doddapaneni H, Dugan S (2018) Hemimetabolous genomes reveal molecular basis of termite eusociality. Nature Ecology & Evolution 2:557-566.

Hartfelder K, Makert GR, Judice CC, Pereira GAG, Santana WC, Dallacqua R, Bitondi MMG (2006) Physiological and genetic mechanisms underlying caste development, reproduction and division of labor in stingless bees. Apidologie 37:144-163.

Hayashi Y, Lo N, Miyata H, Kitade O (2007) Sex-linked genetic influence on caste determination in a termite. Science 318:985-987.

He XJ, Zhang XC, Jiang WJ, Barron AB, Zhang JH, Zeng ZJ (2016) Starving honey bee (*Apis mellifera*) larvae signal pheromonally to worker bees. Scientific Reports 6:22359.

Helanterä H, Uller T (2014) Neutral and adaptive explanations for an association between caste-biased gene expression and rate of sequence evolution. Front Genet 5:Article 297.

Helms Cahan S, Keller L (2003) Complex hybrid origin of genetic caste determination in harvester ants. Nature 424:306-309.

Helms Cahan S, Parker JD, Rissing SW, Johnson RA, Polony TS, Weiser MD, Smith DR (2002) Extreme genetic differences between queens and workers in hybridizing *Pogonomyrmex* harvester ants. Proc R Soc B-Biol Sci 269:1871-1877.

Helms Cahan S, Vinson SB (2003) Reproductive division of labor between hybrid and nonhybrid offspring in a fire ant hybrid zone. Evolution 57:1562-1570.

Higashi M, Yamamura N, Abe T (2000) Theories on the sociality of termites. In: Abe T, Bignell DE, Higashi M (eds) Termites: Evolution, Sociality, Symbiosis, Ecology. Kluwer Academic, Dordrecht, pp 169-188.

Herbers JM (2013) 50 Years on: the legacy of William Donald Hamilton. Biol Lett 9:0130792.

Hölldobler B, Wilson EO (2009) The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies. W.W. Norton, New York.

Hughes WOH, Sumner S, Van Borm S, Boomsma JJ (2003) Worker caste polymorphism has a genetic basis in *Acromyrmex* leaf-cutting ants. Proc Natl Acad Sci U S A 100:9394-9397.

Hunt BG, Ometto L, Wurm Y, Shoemaker D, Soojin VY, Keller L, Goodisman MA (2011) Relaxed selection is a precursor to the evolution of phenotypic plasticity. Proceedings of the National Academy of Sciences 108:15936-15941.

Inward D, Beccaloni G, Eggleton P (2007) Death of an oder: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches Biol Lett 3:331–335.

Jaffe R, Kronauer DJC, Kraus FB, Boomsma JJ, Moritz RFA (2007) Worker caste determination in the army ant *Eciton burchellii*. Biol Lett 3:513-516.

Jones J, Myerscough M, Graham S, Oldroyd BP (2004) Honey bee nest theromregulation: diversity promotes stability. Science 305:402-404.

Julian GE, Fewell JH, Gadau J, Johnson RA, Larrabee D (2002) Genetic determination of the queen caste in an ant hybrid zone. Proc Natl Acad Sci U S A 99:8157-8160.

Keller L (2007) Uncovering the biodiversity if genetic and reproductive systems: time for a more open approach. Am Nat 169:1-8.

Keller L, Sundström L, Chapuisat M (1997) Male reproductive success: paternity contribution to queens and workers in *Formica* ants. Behav Ecol Sociobiol 41:11-15.

Kerr WE (1950) Genetic determination of castes in the genus *Melipona*. Genetics 35:143-152.

Klass K-D, Meier R (2006) A phylogenetic analysis of Dictyoptera (Insecta) based on morphological characters. Entomologische Abhandlungen 63:3-50.

Korb J (2007) Primer: Termites. Curr Biol:R995-R999.

Krylov DM, Wolf YI, Rogozin IB, Koonin EV (2003) Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res. 13: 2229–2235.

Kucharski R, Maleszka J, Forêt S, Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. Science 319:1827-1830.

Lainé LV, Wright DJ (2003) The life cycle of *Reticulitermes* spp. (Isoptera: Rhinotermitidae): what do we know? Bull Entomol Res 93:267-278.

Levin SR, Grafen A (2019) Inclusive fitness is an indispensable approximation for understanding organismal design. Evolution.

Li-Byarlay H (2016) The function of DNA methylation marks in social insects. Frontiers in Ecology and Evolution 4:57.

Li-Byarlay H, Li Y, Stroud H, Feng SH, Newman TC, Kaneda M, Hou KK, Worley KC, Elsik CG, Wickline SA, Jacobsen SE, Ma J, Robinson GE (2013) RNA interference knockdown of DNA methyltransferase 3 affects gene alternative splicing in the honey bee. Proc Natl Acad Sci U S A 110:12750-12755.

Libbrecht R, Oxley PR, Keller L, Kronauer DJC (2016) Robust DNA methylation in the clonal raider ant brain. Curr Biol 26:391-395.

Lin N, Michener CD (1972) Evolution of eusociality in insects. The Quarterly Review of Biology 47:131-159.

Linksvayer TA, Wade MJ (2005) The evolutionary origin and elaboration of sociality in the aculeate hymenoptera: maternal effects, sib-social effects, and heterochrony. Q Rev Biol 80:317-336.

Linksvayer TA, Wade MJ (2009) Genes with social effects are expected to harbor more sequence variation within and between species. Evolution 63:1685-1696.

Lo N, Hayashi Y, Kitade O (2009) Should environmental caste determination be assumed for termites? Am Nat 173:848-853.

Lo N, Li B, Ujvari B (2012) DNA methylation in the termite *Coptotermes lacteus*. Insect Soc 59:257-261.

Maleszka R (2016) Epigenetic code and insect behavioural plasticity. Current Opinion in Insect Science 15:45-52.

Marshall JAR (2015) Social Evolution and Inclusive Fitness Theory. Princeton University Press, Princeton.

Mattila HR, Seeley TD (2007) Genetic diversity in honey bee colonies enhances productivity and fitness. Science 317:362-364.

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652-654.

Moczek AP, Snell-Rood EC (2008) The basis of bee-ing different: the role of gene silencing in plasticity. Evol Dev 10:511-513.

Moore AJ, Kukuk PF (2002) Quantitative genetic analysis of natural populations. Nat Rev Genet 3:971.

Morandin C, Brendel VP, Sundström L, Helanterä H, Mikheyev AS (2019) Changes in gene DNA methylation and expression networks accompany caste specialization and age-related physiological changes in a social insect. Mol Ecol 28:1975-1993.

Moritz RFA, Lattorff HMG, Neumann P, Kraus FB, Radloff SE, Hepburn HR (2005) Rare royal families in honeybees, *Apis mellifera*. Naturwissenschaften 92:488–491.
Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Genet 39:197-218.
Nijhout HF (2003) Development and evolution of adaptive polyphenisms. Evol Dev 5:9-18.
Nobre T, Rouland-Lefèvre C, Aanen DK (2011) Comparative biology of fungus cultivation in termites and ants. In: Bignell DE, Roisin Y, Lo N (eds) Biology of Termites: a Modern Synthesis. Springer Science+Business Media B.V., Dordrecht, pp 193-210.
Noirot C (1989) Social structure in termite societies. Ethol Ecol Evol 1:1-17.
Noirot C, Pasteels JM (1987) Ontogenetic development and evolution of the worker caste in termites. Experientia 43:851-860.
Ohkawara K, Nakayama M, Satoh A, Trindl A, Heinze J (2006) Clonal reproduction and genetic caste differences in a queen-polymorphic ant, *Vollenhovia emeryi*. Biol Lett 2:359-363.
Okonechnikov K, Golosova O, Fursov M, Team U (2012) Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28:1166-1167.
Oldroyd BP, Fewell JH (2007) Genetic diversity promotes homeostasis in insect colonies. Trends Ecol Evol 22:408-413.
Osborne KE, Oldroyd BP (1999) Possible causes of reproductive dominance during emergency queen rearing by honeybees. Anim Behav 58:267-272.
Oster GF, Wilson EO (1978) Caste and Ecology in the Social Insects. Monographs in Population Biology, no. 12. Princeton University Press, Princeton.
Page RE, Fondrk MK, Robinson GE (1993) Selectable components of sex allocation in colonies of the honeybee (*Apis mellifera* L.). Behav Ecol 4:239-245.
Pal C, Papp B, Hurst LD (2001) Highly expressed genes in yeast evolve slowly. Genetics 158: 927–931.
Papadopoulos JS, Agarwala R (2007) COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics 23:1073-1079.
Parker GA (1989) Hamilton's rule and conditionality. Ethol Ecol Evol 1:195-211.
Pearcy M, Aron S, Doums C, Keller L (2004) Conditional use of sex and parthenogenesis for worker and queen production in ants. Science 306:1780-1783.
Pigliucci M, Muller G (2010) Evolution–the extended synthesis.
Queller DC (2016) Kin selection and its discontents. Philosophy of Science 83:861-872.
Raffoul M, Hecnar SJ, Prezioso S, Hecnar DR, Thompson GJ (2011) Trap response and genetic structure of Eastern subterranean termites (Isoptera, Rhinotermitidae) in Point Pelee National Park, Ontario, Canada. Can Entomol 143:263-271.
Ramsay C, Lasko P, Abouheif E (2020) Evo-Devo lessons from the reproductive division of labor in eusocial Hymenoptera. In: Nuno de la Rosa L, Müller G (eds) Evolutionary Developmental Biology. Springer, Cham.
Ratnieks FL (2001) Heirs and spares: caste conflict and excess queen production in *Melipona* bees. Behav Ecol Sociobiol 50:467-473.
Ratnieks FLW, Foster KR, Wenseleers T (2011) Darwin's special difficulty: the evolution of "neuter insects" and current theory. Behav Ecol Sociobiol 65:481-492.
Revely L, Sumner S, Eggleton P (2021) The plasticity and developmental potential of termites. Frontiers in Ecology and Evolution 9.

Rheindt F, Strehl C, Gadau J (2005) A genetic component in the determination of worker polymorphism in the Florida harvester ant *Pogonomyrmex badius*. Insect Soc 52:163-168.

Roisin Y, Korb J (2011) Social organisation and the status of workers in termites. In: Bignell DE, Roisin Y, Lo N (eds) Biology of Termites: a Modern Synthesis. Springer Science+Business Media B.V., Dordrecht, pp 133-164.

Rombel IT, Sykes KF, Rayner S, Johnston SA (2002) ORF-FINDER: a vector for high-throughput gene identification. Gene 282:33-41.

Rubenstein DR, Abbot P (2017) Comparative Social Evolution. In. Cambridge University Press, pp 465.

Scaduto DA, Garner SR, Leach EL, Thompson GJ (2012) Genetic evidence for multiple invasions of the Eastern subterranean termite into Canada. Environ Entomol 41:1680-1686.

Schwander T, Helms Cahan S, Keller L (2007) Characterization and distribution of *Pogonomyrmex* harvester ant lineages with genetic caste determination. Mol Ecol 16:367-387.

Schwander T, Keller L (2008) Genetic compatibility affects queen and worker caste determination. Science 322:552-552.

Schwander T, Lo N, Beekman M, Oldroyd BP, Keller L (2010) Nature versus nurture in social insect caste differentiation. Trends Ecol Evol 25:275-282.

Seeley TD, Tarpy DR (2006) Queen promiscuity lowers disease within honeybee colonies. Proceedings of the Royal Society B: Biological Sciences 274:67-72.

Shellman-Reeve JS (1997) The spectrum of eusociality in termites. In: Choe JC, Crespi BJ (eds) The Evolution of Social Behavior in Insects and Arachnids. Cambridge University Press, Cambridge, UK, pp 52-93.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539.

Smith CR, Anderson KE, Tillberg CV, Gadau J, Suarez AV (2008) Caste determination in a polymorphic social insect: Nutritional, social, and genetic factors. Am Nat 172:497-507.

Smith CR, Mutti NS, Jasper WC, Naidu A, Smith CD, Gadau J (2012) Patterns of DNA methylation in development, division of labor and hybridization in an ant with genetic caste determination. PLoS One 7:e42433.

Standage DS, Berens AJ, Glastad KM, Severin AJ, Brendel VP, Toth AL (2016) Genome, transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced DNA methylation system in a social insect. Mol Ecol 25:1769-1784.

Subramanian S, Kumar S (2004). Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics. 1;168(1):373-81.

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609-W612.

Terrapon N, Li C, Robertson HM, Ji L, Meng X, Booth W, Chen Z, Childers CP, Glastad KM, Gokhale K, Gowin J, Gronenberg W, Hermansen RA, Hu H, Hunt BG, Huylmans AK, Khalil SMS, Mitchell RD, Munoz-Torres MC, Mustard JA, Pan H, Reese JT, Scharf ME, Sun F, Vogel H, Xiao J, Yang W, Yang Z, Yang Z, Zhou J, Zhu J, Brent CS, Elsik CG, Goodisman MAD, Liberles DA, Roe RM, Vargo EL, Vilcinskas A, Wang J, Bornberg-Bauer E, Korb J, Zhang G, Liebig J (2014) Molecular traces of alternative social organization in a termite genome. Nat Commun 5:3636.

Thompson GJ, Hurd PL, Crespi BJ (2013) Genes underlying altruism. Biol Lett 9:20130395.

Thompson GJ, Kitade O, Lo N, Crozier RH (2000) Phylogenetic evidence for a single, ancestral origin of a 'true' worker caste in termites. J Evol Biol 13:869-881.

Thorne BL, Traniello JFA, Adams ES, Bulmer M (1999) Reproductive dynamics and colony structure of subterranean termites of the genus *Reticulitermes* (Isoptera Rhinotermitidae): a review of the evidence from behavioral, ecological, and genetic studies. Ethol Ecol Evol 11:149-169.

Toth AL, Robinson GE (2007) Evo-devo and the evolution of social behavior. Trends Genet 23:334-341.

Ujvari B, Li B, Evans TA, King A, Kitade O, Lo N (2011) A microsatellite-based test of the *Reticulitermes speratus* genetic caste determination model in *Coptotermes lacteus*. Insect Soc 58:365-370.

Vargo EL (2019) Diversity of termite breeding systems. Insects 10:52.

Vargo EL, Husseneder C (2009) Biology of subterranean termites: insights from molecular studies of *Reticulitermes* and *Coptotermes*. Annu Rev Entomol 54:379-403.

Volny VP, Gordon DM (2002) Genetic basis for queen-worker dimorphism in a social insect. Proc Natl Acad Sci U S A 99:6108-6111.

Watson JAL, Okot-Kotber BM, Noirot C (eds) (1985) Caste Differentiation in Social Insects, vol 3. Current Themes in Tropical Science. Pergamon Press, Oxford, Oxford.

Welch M, Lister R (2014) Epigenomics and the control of fate, form and function in social insects. Current Opinion in Insect Science 1:31-38.

West SA, Griffin AS, Gardner A (2007) Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J Evol Biol 20:415-432.

West SA, Gardner A (2013) Adaptation and inclusive fitness. Curr Biol 23:R577-R584.

West-Eberhard MJ (1987) Flexible strategy and social evolution. In: Ito Y, Brown JL, Kihhauia J (eds) Animal Societies. Theories and Facts. Japan Set. Soc. Press, Tokyo, pp 35-51.

Wheeler DE (1986) Developmental and physiological determinants of caste in social Hymenoptera: evolutionary implications. Am Nat 128:13-34.

Wilfert L, Gadau J, Schmid-Hempel P (2007) Variation in genomic recombination rates among animal taxa and the case of social insects. Heredity 98:189.

Williams GC, Williams DC (1957) Natural selection of individually harmful social adaptations among sibs with special reference to social insects. Evolution 11:32-39.

Winter U, Buschinger A (1986) Genetically mediated queen polymorphism and caste determination in the slave-making ant, *Harpagoxenus sublaevis* (Hymenoptera: Formicidae). Entomologia Generalis 11:125-137.

Wu T, Dhami GK, Thompson GJ (2018) Soldier-biased gene expression in a subterranean termite implies functional specialization of the defensive caste. Evol Dev 20:3-16.

Wu T, Simkovic V, Thompson GJ (2015) Subterranean termites: the evolution of a pest. PCT Canada 3:34-42.

Yamamoto Y, Matsuura K (2012) Genetic influence on caste determination underlying the asexual queen succession system in a termite. Behav Ecol Sociobiol 66:39-46.

Yan H, Bonasio R, Simola DF, Liebig J, Berger SL, Reinberg D (2015) DNA methylation in social insects: How epigenetics can control behavior and longevity. Annu Rev Entomol 60:435-452.

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555-556.

Yang Z (1998) On the best evolutionary rate for phylogenetic analysis. Systematic Biology 47:125-133.

Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431-449.

APPENDIX

5. 1. Scripts for Data Manipulation and Analyses

Script 1

This PYTHON script splits all *R. flavipes* transcripts into differentially expressed (DE) and non-differentially expressed (NDE) datasets using a list of required Gene IDs stored in another file. This script was applied twice, to generate two FASTA files containing DE and NDE genes.

```
import os
import sys
from Bio import SeqIO    # import required modules

unfiltered_fasta = sys.argv[1]    # input unfiltered multi-FASTA or .txt file

# input a second file which contains a list of gene IDs to be filtered out
remove_file = sys.argv[2]
filtered_fasta = sys.argv[3]    # output FASTA file with only DEs or NDEs


remove = set()
with open(remove_file) as f:
    for line in f:
        line = line.strip()
        if line != "":
            remove.add(line)


# specify the FASTA format
fasta_sequences = SeqIO.parse(open(unfiltered_fasta),'fasta')

with open(filtered_fasta, "w") as f:
    for seq in fasta_sequences:
        name = seq.id
        nuc = str(seq.seq)
        if name not in remove and len(nuc) > 0:
            SeqIO.write([seq], f, "fasta")
```

Script 2

These UNIX commands were applied to NDE dataset only, to randomly select *n=93* genes that serve as a comparative reference against an equivalent-sized DE set.

```
# create a list of all NDE headers and write to file
grep '>' NDE_Rflav_transcriptome_sequences.fasta > NDE_headers.txt

# randomly sort and select specified n=93 NDE genes from the list of headers
CONTIGS = sort -R NDE_headers.txt| head -n 93 |tr "\n" " "
```

Script 3

The 'get_orfs_or_cdss.py' PYTHON script was applied to both DE and NDE datasets to identify the longest open reading frame (in 5' to 3' orientation) of each transcript. This script has been published and is freely available from:

https://github.com/peterjc/pico_galaxy/blob/master/tools/get_orfs_or_cdss/get_orfs_or_cdss.py

```
#!/usr/bin/env python
"""Find ORFs in a nucleotide sequence file.

For more details, see the help text and argument descriptions in the
accompanying get_orfs_or_cdss.xml file which defines a Galaxy interface.

This tool is a short Python script which requires Biopython. If you use
this tool in scientific work leading to a publication, please cite the
Biopython application note:

Cock et al 2009. Biopython: freely available Python tools for computational
molecular biology and bioinformatics. Bioinformatics 25(11) 1422-3.
https://doi.org/10.1093/bioinformatics/btp163 pmid:19304878.

This script is copyright 2011-2013 by Peter Cock, The James Hutton Institute
(formerly SCRI), Dundee, UK. All rights reserved.

See accompanying text file for licence details (MIT licence).
"""

from __future__ import print_function

import re
import sys

from optparse import OptionParser
```

```
usage = r"""Use as follows:

$ python get_orfs_or_cdss.py -i genome.fa -f fasta --table 11 \
-t CDS -e open -m all -s both --on cds.nuc.fa --op cds.protein.fa \
--ob cds.bed --og cds.gff3
"""

try:
    from Bio.Seq import Seq, reverse_complement, translate
    from Bio.SeqRecord import SeqRecord
    from Bio import SeqIO
    from Bio.Data import CodonTable
except ImportError:
    sys.exit("Missing Biopython library")


parser = OptionParser(usage=usage)
parser.add_option(
    "-i",
    "--input",
    dest="input_file",
    default=None,
    help="Input fasta file",
    metavar="FILE",
)
parser.add_option(
    "-f",
    "--format",
    dest="seq_format",
    default="fasta",
    help="Sequence format (e.g. fasta, fastq, sff)",
)
parser.add_option(
    "--table", dest="table", default=1, help="NCBI Translation table",
type="int"
)
parser.add_option(
    "-t",
    "--ftype",
    dest="ftype",
    type="choice",
    choices=["CDS", "ORF"],
    default="ORF",
    help="Find ORF or CDSs",
)
parser.add_option(
    "-e",
    "--ends",
    dest="ends",
    type="choice",
    choices=["open", "closed"],
    default="closed",
    help="Open or closed. Closed ensures start/stop codons are present",
)
parser.add_option(
    "-m",
```

```
    "--mode",
    dest="mode",
    type="choice",
    choices=["all", "top", "one"],
    default="all",
    help="Output all ORFs/CDSs from sequence, all ORFs/CDSs "
    "with max length, or first with maximum length",
)
parser.add_option(
    "--min_len", dest="min_len", default=10, help="Minimum ORF/CDS length",
type="int"
)
parser.add_option(
    "-s",
    "--strand",
    dest="strand",
    type="choice",
    choices=["forward", "reverse", "both"],
    default="both",
    help="Strand to search for features on",
)
parser.add_option(
    "--on",
    dest="out_nuc_file",
    default=None,
    help="Output nucleotide sequences, or - for STDOUT",
    metavar="FILE",
)
parser.add_option(
    "--op",
    dest="out_prot_file",
    default=None,
    help="Output protein sequences, or - for STDOUT",
    metavar="FILE",
)
parser.add_option(
    "--ob",
    dest="out_bed_file",
    default=None,
    help="Output BED file, or - for STDOUT",
    metavar="FILE",
)
parser.add_option(
    "--og",
    dest="out_gff3_file",
    default=None,
    help="Output GFF3 file, or - for STDOUT",
    metavar="FILE",
)
parser.add_option(
    "-v",
    "--version",
    dest="version",
    default=False,
    action="store_true",
    help="Show version and quit",
)
```

```
options, args = parser.parse_args()

if options.version:
    print("v0.2.3")
    sys.exit(0)

if not options.input_file:
    sys.exit("Input file is required")

if not any(
    (
        options.out_nuc_file,
        options.out_prot_file,
        options.out_bed_file,
        options.out_gff3_file,
    )
):
    sys.exit("At least one output file is required")

try:
    table_obj = CodonTable.ambiguous_generic_by_id[options.table]
except KeyError:
    sys.exit("Unknown codon table %i" % options.table)

if options.seq_format.lower() == "sff":
    seq_format = "sff-trim"
elif options.seq_format.lower() == "fasta":
    seq_format = "fasta"
elif options.seq_format.lower().startswith("fastq"):
    seq_format = "fastq"
else:
    sys.exit("Unsupported file type %r" % options.seq_format)

print("Genetic code table %i" % options.table)
print("Minimum length %i aa" % options.min_len)
# print "Taking %s ORF(s) from %s strand(s)" % (mode, strand)

starts = sorted(table_obj.start_codons)
assert "NNN" not in starts
re_starts = re.compile("|".join(starts))

stops = sorted(table_obj.stop_codons)
assert "NNN" not in stops
re_stops = re.compile("|".join(stops))


def start_chop_and_trans(s, strict=True):
    """Return offset, trimmed nuc, protein."""
    if strict:
        assert s[-3:] in stops, s
    assert len(s) % 3 == 0
    for match in re_starts.finditer(s):
        # Must check the start is in frame
        start = match.start()
        if start % 3 == 0:
            n = s[start:]
```

```
            assert len(n) % 3 == 0, "%s is len %i" % (n, len(n))
            if strict:
                t = translate(n, options.table, cds=True)
            else:
                # Use when missing stop codon,
                t = "M" + translate(n[3:], options.table, to_stop=True)
            return start, n, t
    return None, None, None


def break_up_frame(s):
    """Return offset, nuc, protein."""
    start = 0
    for match in re_stops.finditer(s):
        index = match.start() + 3
        if index % 3 != 0:
            continue
        n = s[start:index]
        if options.ftype == "CDS":
            offset, n, t = start_chop_and_trans(n)
        else:
            offset = 0
            t = translate(n, options.table, to_stop=True)
        if n and len(t) >= options.min_len:
            yield start + offset, n, t
        start = index
    if options.ends == "open":
        # No stop codon, Biopython's strict CDS translate will fail
        n = s[start:]
        # Ensure we have whole codons
        # TODO - Try appending N instead?
        # TODO - Do the next four line more elegantly
        if len(n) % 3:
            n = n[:-1]
        if len(n) % 3:
            n = n[:-1]
        if options.ftype == "CDS":
            offset, n, t = start_chop_and_trans(n, strict=False)
        else:
            offset = 0
            t = translate(n, options.table, to_stop=True)
        if n and len(t) >= options.min_len:
            yield start + offset, n, t


def get_all_peptides(nuc_seq):
    """Return start, end, strand, nucleotides, protein.

    Co-ordinates are Python style zero-based.
    """
    # TODO - Refactor to use a generator function (in start order)
    # rather than making a list and sorting?
    answer = []
    full_len = len(nuc_seq)
    if options.strand != "reverse":
        for frame in range(0, 3):
            for offset, n, t in break_up_frame(nuc_seq[frame:]):
```

```
                start = frame + offset  # zero based
                answer.append((start, start + len(n), +1, n, t))
    if options.strand != "forward":
        rc = reverse_complement(nuc_seq)
        for frame in range(0, 3):
            for offset, n, t in break_up_frame(rc[frame:]):
                start = full_len - frame - offset  # zero based
                answer.append((start - len(n), start, -1, n, t))
    answer.sort()
    return answer


def get_top_peptides(nuc_seq):
    """Return all peptides of max length."""
    values = list(get_all_peptides(nuc_seq))
    if not values:
        raise StopIteration
    max_len = max(len(x[-1]) for x in values)
    for x in values:
        if len(x[-1]) == max_len:
            yield x


def get_one_peptide(nuc_seq):
    """Return first (left most) peptide with max length."""
    values = list(get_top_peptides(nuc_seq))
    if not values:
        raise StopIteration
    yield values[0]


if options.mode == "all":
    get_peptides = get_all_peptides
elif options.mode == "top":
    get_peptides = get_top_peptides
elif options.mode == "one":
    get_peptides = get_one_peptide

in_count = 0
out_count = 0
if options.out_nuc_file == "-":
    out_nuc = sys.stdout
elif options.out_nuc_file:
    out_nuc = open(options.out_nuc_file, "w")
else:
    out_nuc = None

if options.out_prot_file == "-":
    out_prot = sys.stdout
elif options.out_prot_file:
    out_prot = open(options.out_prot_file, "w")
else:
    out_prot = None

if options.out_bed_file == "-":
    out_bed = sys.stdout
elif options.out_bed_file:
```

```
    out_bed = open(options.out_bed_file, "w")
else:
    out_bed = None

if options.out_gff3_file == "-":
    out_gff3 = sys.stdout
elif options.out_gff3_file:
    out_gff3 = open(options.out_gff3_file, "w")
else:
    out_gff3 = None

if out_gff3:
    out_gff3.write("##gff-version 3\n")

for record in SeqIO.parse(options.input_file, seq_format):
    for i, (f_start, f_end, f_strand, n, t) in enumerate(
        get_peptides(str(record.seq).upper())
    ):
        out_count += 1
        if f_strand == +1:
            loc = "%i..%i" % (f_start + 1, f_end)
        else:
            loc = "complement(%i..%i)" % (f_start + 1, f_end)
        descr = "length %i aa, %i bp, from %s of %s" % (
            len(t),
            len(n),
            loc,
            record.description,
        )
        fid = record.id + "|%s%i" % (options.ftype, i + 1)
        r = SeqRecord(Seq(n), id=fid, name="", description=descr)
        t = SeqRecord(Seq(t), id=fid, name="", description=descr)
        if out_nuc:
            SeqIO.write(r, out_nuc, "fasta")
        if out_prot:
            SeqIO.write(t, out_prot, "fasta")
        nice_strand = "+" if f_strand == +1 else "-"
        if out_bed:
            out_bed.write(
                "\t".join(map(str, [record.id, f_start, f_end, fid, 0,
nice_strand]))
                + "\n"
            )
        if out_gff3:
            out_gff3.write(
                "\t".join(
                    map(
                        str,
                        [
                            record.id,
                            "getOrfsOrCds",
                            "CDS",
                            f_start + 1,
                            f_end,
                            ".",
                            nice_strand,
                            0,
```

```
                            "ID=%s%s" % (options.ftype, i + 1),
                        ],
                    )
                )
                + "\n"
            )
        in_count += 1
if out_nuc and out_nuc is not sys.stdout:
    out_nuc.close()
if out_prot and out_prot is not sys.stdout:
    out_prot.close()
if out_bed and out_bed is not sys.stdout:
    out_bed.close()

print("Found %i %ss in %i sequences" % (out_count, options.ftype, in_count))
```

Script 4

I used the following set of UNIX commands to cluster all the nucleotide and corresponding amino acid transcripts for both DE and NDE datasets into functionally related gene families using gene IDs as tags. These commands were applied a total of two times on each nucleotide and protein sequence FASTA files for DE (shown below) and NDE datasets:

```
# split all DE transcripts in a nucleotide sequences FASTA file into gene
families based on gene IDs
awk '{if(substr($0,1,1) ==
">"){split(substr($0,2,length($0)),a,/_/);filename=a[2]};print $0 > filename
}' all_DE_nucleotide.fasta

# split all DE transcripts in a corresponding protein sequences FASTA file
into gene families based on gene IDs
awk '{if(substr($0,1,1) ==
">"){split(substr($0,2,length($0)),a,/_/);filename=a[2]};print $0 > filename
}' all_DE_protein.fasta
```

CURRICULUM VITAE

Name:	Anna M Chernyshova
Post-secondary Education and Degrees:	York University, Toronto, ON, Canada iBSc in Biomedical Sciences, 2015 Western University, London, ON, Canada MSc in Biology, 2021
Honours and Awards:	Faculty of Science and Engineering Entrance Scholarship, 2008 xYU Renewable Entrance Scholarship, 2008 Biology Graduate Student Travel Award, 2018 Ruth Horner Arnold Fellowship in Biology, 2018
Related Work Experience:	Graduate Teaching Assistant Biology Department, Western University London, ON, Canada 2017-2021

Publications:

Behl S, Wu T, Chernyshova AM, Thompson GJ (2018) Caste-biased genes in a subterranean termite are taxonomically restricted: implications for novel gene recruitment during termite caste evolution. Insectes Sociaux 65 (4), 593-599.

Daisley BA, Pitek AP, Chmiel JA, Al KF, Chernyshova AM, Faragalla KM, Burton JP, Thompson GJ, Reid G (2020) Novel probiotic approach to counter Paenibacillus larvae infection in honey bees. The ISME Journal 14 (2), 476-491.

Daisley BA, Pitek AP, Chmiel JA, Gibbons S, Chernyshova AM, Al KF, Faragalla KM, Burton JP, Thompson GJ, Reid G (2020) *Lactobacillus* spp. attenuate antibiotic-induced immune and microbiota dysregulation in honey bees. Communications Biology 3 (1), 1-13.

Faragalla KM, Chernyshova AM, Gallo AJ, Thompson GJ (2018) From gene list to gene network: Recognizing functional connections that regulate behavioral traits. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution 330(6-7), 317-329.

Guoth A, Chernyshova AM, Thompson GJ (2020) Gene-regulatory context of honey bee worker sterility. Biosystems, 104235.

Harpur BA, Chernyshova A, Soltani A, Tsvetkov N, Mahjoorighasrodashti M, Xu Z, Zayed A (2014) No genetic tradeoffs between hygienic behaviour and individual innate immunity in the honey bee, *Apis mellifera*. PloS one 9 (8), e104214.

Harrison MC, Chernyshova AM, Thompson GJ (2020) No obvious transcriptome-wide signature of indirect selection in termites. Journal of Evolutionary Biology.

Thompson GJ, Chernyshova AM (2020) Caste differentiation: Genetic and Epigenetic Factors. pp 165-176. In: Starr, C.K. (ed.), Encyclopedia of Social Insects, Cham, Switzerland: Springer.