5-17-2017 12:00 AM

Transcriptomic correlates of caste differentiation and invasiveness in a subterranean termite

Tian Wu, *The University of Western Ontario*

Supervisor: Dr. Graham Thompson, *The University of Western Ontario*

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

Abstract

Termite colonies are characterized by a division of labour into reproductive and helper castes (soldiers and workers). Caste differentiation is associated with differences in gene expression that reflect developmental and evolutionary origins. I used RNA-seq to investigate genome-wide expression patterns of reproductive nymph, soldier, and worker castes of *Reticulitermes flavipes* from three populations. I found 93 genes differentially expressed as a function of caste, with the majority of genes being uniquely up-regulated in soldiers. My findings suggest that soldiers evolved genes that are distinct from nymphs and workers, and are signs of possible genomic novelty. I also analyzed this transcriptome as a function of population since *R. flavipes* is invasive to Canada and display supercolony-like phenotypes such as lack of kin recognition. I found 302 differentially expressed genes, with 77% of these biased toward their invasive-native status. These population gene sets indicate molecular level adaptations to the observed invasive phenotypes.

Keywords

Eastern subterranean termite, *Reticulitermes flavipes*, caste differentiation, invasiveness, RNA-seq, transcriptome, gene expression

Co-Authorship Statement

This thesis was conducted under the supervision of Dr. Graham J. Thompson. Experimental design and data analysis were done in cooperation between Dr. Thompson and myself. Any subsequent publications to this thesis will be co-authored with Dr. Thompson.

Acknowledgments

First, I like to thank my supervisor Dr. Graham Thompson for the opportunity of doing my Master's project in the Social Biology Group at UWO. Despite some trying times in the project, I am very thankful that you gave me the patience and confidence to explore and push through those challenges. Also, thank you for always striving to make my writing and presentations ever more clear. I will always keep those lessons in mind.

I like to thank my advisory committee, Dr. David Smith and Dr. Mark Daley for all your guidance and encouragements throughout the project. I also like to thank all the staff at Sharcnet for their computational and technical support. Thank you to Dr. Edward Vargo, Dr. Rebeca Rosengaus, and Vicki Simkovic for providing the termites used in this thesis.

A big thank you to all the present and former lab members of the Social Biology Group for 'putting up' with me during my Master's: Justin Croft, Alison Camiletti, Jessica Empringham, Emma Mullen, Tom Liu, Julia Sobotka, Vicki Simkovic, Anthony Gallo, Supriya Behl, Christine Scharf, and Rahul Choorakkat. Special thanks to Justin for always being there both as a labmate and as a friend. Thank you to the Mustang Bioinformatics Club for growing and learning with me. I wish everyone continued success in their own respective bioinformatic journeys. Thank you to the lunch table group for your intellectual thoughts and inputs during therapeutic freezie time.

Thank you to Emily Cornelius for always keeping me positive, smiling, and well fed. I am truly grateful. Thank you to my friends outside of school for your moral support during my Master's degree: my roommates at apt 601, Steve Trudell, and the boys in Calgary, Ottawa, Toronto, and Windsor.

Last but not least, I'd like to thank my parents and family for their endless love and support. I love you all. Without you guys, I could not have accomplished this.

Table of Contents

List of Tables

List of Figures

List of Supplemental Materials

List of Abbreviations

r: relatedness of the altruist to the receiver

b: direct fitness benefit to the receiver

c: the direct fitness cost to the altruist

GO: gene ontology

BP: biological process

CC: cellular component

MF: molecular function

DEG: differentially expressed gene

FDR: false discovery rate

ORF: open reading frame

RNA-Seq: RNA sequencing or whole transcriptome shotgun sequencing

RNAi: RNA interference

SRA: Sequence Read Archive

TPM: transcripts per million

T: Toronto, ON

B: Boston, MA

R: Raleigh, NC

Chapter 1

1 General Introduction

1.1 Evolution of eusociality

Eusociality is the highest level of social organization in animals (Wilson, 1971). It has three defining characteristics: cooperative brood care, overlapping generations, and reproductive division of labour amongst individuals living within a colony (Bourke, 2014; Crespi & Yanega, 1995). Though rare throughout animal taxa, it has independently evolved at least 20 times (Bourke, 2011a). The evolution of eusociality presents an interesting paradox to biologists (Ratnieks et al., 2011). After all, why would individuals forgo their own reproduction to improve that of others? This wasn't answered until W. D. Hamilton (1964) elegantly formulated the concept of kin selection and inclusive fitness theory. The major breakthrough of this theory is that it partitions fitness itself into direct and indirect components. Therefore, it allowed for a shift from the classic Darwinian way of thinking, in which fitness maximizes the direct reproductive success of individuals to a gene-centric way of thinking, where an individual's genes can be transmitted through descendent or non-descendent kin. It was the total of these two potential sources of individual fitness that Hamilton called 'inclusive fitness'. The key implications of this idea is that traits with no effect on direct fitness can still evolve, provided their positive effect on indirect fitness effect is sufficiently large, as explained by Hamilton's Rule:

$$r * b > c$$

Here, even genes for reproductive altruism can evolve if the direct fitness benefit (b) to a receiver, multiplied ($*$) by the degree to which the altruist is related (r) to the receiver, is greater than the direct fitness cost (c) to the altruist itself. For altruism to evolve via Hamilton's rule, the altruism must be directed towards relatives, which requires kin recognition (Penn & Frommen, 2010). To avoid the cost of indiscriminate altruism, kinship can be maintained by a closed colony system, in which non-kin are kept out (Breed, 2014).

One of the most popular examples of reproductive altruism is the workers and the queen in honeybees (*Apis mellifera*). The haplodiploid sex determination system of the honeybee and

other hymenopterans creates an unusually higher relatedness between sisters. These 'supersister' workers help to satisfy the conditions of Hamilton's rule for altruism, as the indirect fitness they gain from raising the queen's female offspring, or their sisters, is greater than the direct fitness cost of sterility for each individual (Foster et al., 2006). However, not all species that satisfy this condition evolve eusociality (Thompson & Oldroyd, 2004). There are other factors necessary for eusociality to evolve. Under the reproductive ground plan hypothesis (West-Eberhard, 1996), solitary ancestors of eusocial species possess reproductive and non-reproductive phases, where through disruptive selection, these phases eventually decoupled for reproductive division of labour into castes. Therefore, evolution of caste polyphenism into workers and queens evolved as a specialization of tasks performed by reproductive and non-reproductive individuals. These conditions are well accepted for the social hymenopterans such as ants, bees, and wasps, but its applicability is debatable for termites (Howard & Thorne, 2011).

1.2 Termite eusociality

Termites are a monophyletic group (Blattodea: Isoptera) that evolved eusociality independently once, and before the evolution of eusocial system of Hymenoptera (Grimaldi & Engel, 2005). Furthermore, there are several fundamental biological differences between termite and hymenopteran eusocial systems. All termites are diploid, as opposed to haplodiploid (Lo & Eggleton, 2011). Termite reproductive division of labour is marked by a wider degree of caste polyphenism, and there are also males and females for all castes (Thorne, 1997a). They possess both primary (king, queen) and secondary reproductives (neotenics), and soldiers as an additional non-reproductive helper caste to workers. Adding to the complexity of termite caste polyphenism, sterility for workers in some families is facultative, while obligatory in others families (Bourguignon et al., 2015). Juvenile termites progress through hemimetabolous development via a series of molts before reaching adult form. Termite hemimetaboly permits a great degree of developmental flexibility. For example, termite workers can further differentiate into soldiers or secondary reproductives depending on socio-environmental cues (Watanabe et al., 2014). Termites feed on wood, and are able to digest the tough and difficult lignocellulose materials with the assistance of symbiotic microorganisms such as bacteria, fungi, and protists (Lo & Eggleton, 2011).

Besides using wood as food, termites are also able to use wood as nest sites for their colonies (Eggleton & Tayasu, 2001).

It is clear the conditions of termite eusociality are very different than the social hymenopterans, and therefore, the origin of eusociality in termites is unique. Because of their cryptic nests and high density living, parameters for Hamilton's rule are likely met through poor dispersal and the potential for nest inheritance. These conditions favour inbreeding (i.e., high *r)* and the potential for would-be-dispersers to assume the position of the king or queen (i.e., high *b*) (Korb & Schneider, 2007). Developmentally, how termite caste polyphenism evolved is also in debate, as a recent study points out that polyphenism may have first evolved in a solitary cockroach-like ancestor before reproductive division of labour occurring between reproductive and non-reproductive individuals (Bourguignon et al., 2016). The presence of winged primary reproductives and wingless secondary reproductives in termites imply the observed extant polyphenism evolved from differential success of aerial and ground dispersal, not early reproductive skew amongst individuals. Over time, the close proximities of individuals promoted helping behaviour from kin selection, which eventually lead to a reproductive skew between these winged and wingless individuals that facilitated the evolution of eusociality.

Subsequently, with winglessness as an ancestral state for termites, it also puts in question the evolutionary status of worker and soldier castes. Soldiers have been widely accepted as an ancestral caste that evolved prior to workers (Legendre et al., 2013; Roux & Korb, 2004), but others have proposed that workers evolved prior to soldiers (Bourguignon et al., 2016; Thompson et al., 2000). Soldier ancestrality arguments are primarily based upon the fact that soldiers are ubiquitous throughout the termite lineage (secondarily lost in some species), and workers being derived because certain families possess 'true' (obligatory sterile) or 'false' (facultatively sterile) workers. However, soldiers differentiate from workers (Roisin & Korb, 2011) and are nutritionally dependent on workers via trophallaxis (Korb et al., 2012), which creates a paradox of its ancestral status. To better understand termite social evolution, it is critical to clarify the relationship and origin of the soldier and worker castes.

1.3 Breakdown of eusocial organizations

Interestingly, the rules of eusociality appear to be broken on occasions, and nowhere is it more prominent than the supercolonies of invasive termites and ants (Giraud et al., 2002; Helanterä et al., 2009; Husseneder et al., 2011; Leniaud et al., 2009). These breakdowns challenge the fundamentals of termite eusociality evolution and its maintenance moving forward. Invasive supercolonies are an extreme form of open social organization that contain multiple reproductives, consist of many individual colonies, and are often several times larger compared to the closed colonies in their native range (Evans, 2010). Unlike the closed structure of native colonies where benefits of altruism are exclusive to close reproducing relatives and aggressively protected, supercolonies freely exchange helping benefits with little to no aggression that could lead to degradation of the helper trait (Leniaud et al., 2009; Perdereau et al., 2010). As a result, mating in invasive populations occurs primarily at their natal supercolonies and new nests form by budding instead of colony founding by nuptial flights with winged adults (Thorne et al., 1999). These supercolonies can eventually become massive expansive entities that have no boundaries (Perdereau et al., 2015).

As mentioned, kin recognition is necessary for the maintenance of reproductive altruism and boundaries between neighbouring colonies (Penn & Frommen, 2010). Proper kin recognition is a behaviour that is dependent on both environmental and genetic factors, though its effects are species dependent (Adams, 1991; Florane et al., 2004; Shelton & Grace, 1997). One explanation of this observed breakdown of kin recognition in supercolonies stems from the potential genetic consequences associated with the invasion of social insects. When a species is introduced to a new environment, it suffers from a bottleneck-like effect, where any subsequent colonies are founded upon low genetic diversity (Husseneder et al., 2009; Scaduto et al., 2012). This low variation in genetic diversity can have a serious impact on the kin recognition abilities of individuals in these new colonies because they lack sufficient variation in genetic cues to effectively distinguish highly-related individuals from individuals of low relatedness (Dronnet et al., 2006). However, some argued that supercolonies are essentially all relatives of each other and kin recognition is not defective, but rather enforced (Hanna et al., 2013; Tsutsui et al., 2003). Other studies have shown there is actually sufficient genetic diversity available in invasive colonies, and diversity of recognition cues that is selected against because of the benefits of being in a supercolony (Giraud et al., 2002;

Perdereau et al., 2015). For example, invasive colonies benefit from higher worker numbers and greater competitive advantage against aggressive neighbours by having a more common recognition cue (Grosberg & Quinn, 1989; Rousset & Roze, 2007). In multi-queen colonies of the fire ant, *Solenopsis invicta*, workers heterozygous at the general protein-9 (*Gp-9*) locus selectively kill homozygous queens based on allele mismatches (Bb against BB). In this case, a particular allele - the 'little *b*' allele - encodes an odorant binding protein that is thought to directly mediate the recognition and preferential treatment of queens sharing the same alleles (Keller & Ross, 1998). For termites, heritable cues are apparent in *Microcerotermes* workers that are consistently more aggressive towards genuine non-kin, regardless of any shared environment (Adams, 1991). Where it has been examined, other species of termites may also be selectively aggressive towards less related conspecifics (Adams et al., 2007; Bulmer & Traniello, 2002; Husseneder et al., 2005). Despite promising behavioural evidence for genetic kin recognition, there isn't clear evidence to what cues serve as kinship indicators. There are several candidates to what these indicators of kinship are, such as the cuticular hydrocarbons of termite exoskeletons (van Zweden & d'Ettorre, 2010), pheromones (Kettler & Leuthold, 1994), protein secretions (Hanus et al., 2010), or a combination of these factors. Currently, there is no consensus on which of these determines kin recognition or how this process is disrupted in invasive termite populations.

1.4 Sociogenomic approach to understanding eusociality

Over recent years, there is increasing awareness of the importance of genes and their roles in driving eusocial phenotypes (Kapheim, 2016; Thompson et al., 2013). A field dubbed 'sociogenomics' has greatly facilitated this development of understanding genes underlying social traits (Robinson et al., 2005). By merging the 'why' of sociobiology with the 'how' of molecular biology, this new approach helped to overcome some of the mechanistic level challenges by quantifying and characterizing genes underlying phenotypes (Roux et al., 2014; Scharf, 2015). Specifically, the sequencing and publication of two termite genomes (Poulsen et al., 2014; Terrapon et al., 2014) revealed conservation of eusociality regulation with social hymenopterans involving important juvenile hormone regulators such as hexamerins and cytochrome P450 genes (Tarver et al., 2012; Zhou et al., 2007), and genes related to reproduction such as the egg yolk precursor protein, vitellogenin (Scharf et al., 2003). Despite these similarities, there were also many signs of alternative methods of social

organization with large differences in genome size, gene content, and rapid expansion and contraction of gene families related to chemoreception, immunity, caste differentiation, and the added genomes of their gut symbionts (Poulsen et al., 2014; Terrapon et al., 2014). More importantly, multi-caste transcriptomic level comparisons provided a more comprehensive view of co-regulatory patterns amongst genes and whole gene sets in termites. These findings encourage us to re-think the path to eusociality for termites and allow us to test new and existing hypotheses of termite social evolution with genetic level evidence. Regarding the evolutionary status of workers and soldiers, transcriptomic comparison of various castes in the dampwood termite, *Zootermopsis nevadensis*, showed a greater proportion of gene co-regulation between workers and nymphs, primary and secondary reproductives, while soldiers appeared to be uniquely regulated (Terrapon et al., 2014). Though large-scale transcriptomic studies like these are rare for termite castes (Scharf, 2015), comparisons like these for ants and bees already provide unprecedented insights into social hymenopteran evolution (Kapheim, 2016; Mikheyev & Linksvayer, 2015). Specifically, differential gene expression between queen and worker castes consistently found that the worker caste is a source of genomic novelty as supported by the up-regulation of primarily taxonomically restricted genes compared to queens up-regulating more universal and conserved genes (Feldmeyer et al., 2014; Harpur et al., 2014). The patterns have yet to be determined in termites, but more large-scale transcriptomic efforts could help clarify the evolutionary relationships of their caste evolution.

As for testing large-scale genetic differences in phenotypes such as invasive supercolonies, the application of sociogenomics could be equally as informative. Previous genetic studies of invasive population have generated foundational knowledge from a genetic diversity perspective (Evans et al., 2013). Advances in next-generation sequencing and bioinformatic analysis tools now allow non-model organisms such as invasive species to be sequenced and quantified without a prerequisite genome (Oppenheim et al., 2015). Utilizing this new transcriptomic technology, we could finally begin to test for functional genomic components to supercolony phenotypes such as lack of kin recognition, presence of multiple reproductives, and large population densities. Just as comparative transcriptomics between castes revolutionized our understanding of caste polyphenism in social insects (Robinson et al., 2005), the same application of transcriptomic technology applied to invasive and native

termite populations in a comparative context, could reveal a new level of resolution for 'invasion genes' and molecular pathways affected by invasive events (Chown et al., 2015). Furthermore, we can use these transcriptomic data to detect potential adaptive shifts, dynamics of termites' response to new selection pressure, and potential fitness trade-offs in these invasive supercolonies. Besides addressing fundamental sociobiology questions, these genetic data will also be extremely valuable from an application perspective, as they can help to develop more effective management strategies to exploit these genetic changes to prevent future invasion events.

1.5 Overall thesis objectives, hypotheses, and predictions

My overall goal was to use a comparative transcriptomics approach (RNA-seq) to measure the changes in gene expression that correlate with differences in castes and populations in the Eastern subterranean termite, *Reticulitermes flavipes*. My specific objectives were to:

I. Assemble and annotate a *de novo R. flavipes* reference transcriptome

Previous *R. flavipes* transcriptomic work (Scharf, 2015) identified some key genes and molecular pathways that are involved in the caste differentiation process and provided some of the first large-scale expression data for multi-caste comparisons. From these results, it appeared that some castes shared more similarities in gene expression than others, and is suggestive of underlying development and evolutionary patterns. However, these patterns were not definitive as there were lack of gene annotations and sequencing power. Since then two termite genomes (Terrapon et al., 2014; Poulsen et al., 2014) and several other social hymenopteran genomes (Libbrecht et al., 2013; Kapheim et al., 2015) have been published, greatly increasing the available gene reference resources. In my thesis, I used both available termite and hymenopteran gene references to annotate a reference transcriptome for the nymph, soldier, and worker castes of *R. flavipes*. Having this reference transcriptome allowed me to make better gene predictions, stronger biological inferences, and be confident in the outcomes of comparative analyses for my second and third objectives.

II. Test whether gene expression between castes is associated with development or evolutionary history

To test for caste-biased gene expression patterns, I applied an *in silico* differential gene expression analysis to the RNA-seq data that I obtained from the reproductive nymph caste, and non-reproductive soldier and worker castes. If gene expression patterns reflect termite development, then I expect there will be a reproduction bias as the gene expression profiles of soldiers and workers will be more similar compared to that of the nymphs, in the topology of [nymph, (soldier, worker)]. In terms of evolutionary history, under single origin of true worker hypothesis, then I would expect nymphs and workers to share similar gene expression profiles compared to soldiers [(nymph, worker), soldier]. Under a soldier first hypothesis, I would expect the gene expression profiles of nymphs and soldiers to be more similar compared to workers [(nymph, soldier), worker]. If gene expression is neither influenced by development or evolutionary history, then I expect the gene expression profiles to be caste-specific [nymph, soldier, worker].

III. Compare gene expression between invasive and native populations

To test for biases in gene expression associated with different termite populations, I performed a similar analysis to that outlined above for caste, but in this case I used 'population' as my predictor variable, while holding caste constant. I used RNA-seq data from one invasive population (Toronto ON) and two native populations (Boston MA, Raleigh NC) of *R. flavipes*. If invasiveness or population has an influence on gene expression, then I expect that the two native populations will share a more similar gene expression profile than either will to the single invasive population [T, (B, R)]. Because the invasive Toronto population is relatively new (Urquhart, 1953), habituated in an urban environment, and is potentially founded from a single introduction (Scaduto et al., 2012), I expect the Toronto population gene expression profile to reveal potential adaptive molecular processes.

1.6 References

Adams, E. (1991). Nest-mate recognition based on heritable odors in the termite *Microcerotermes arboreus*. *Proc Natl Acad Sci USA*, *88*, 2031–2034.

Adams, E. S., Atkinson, L., & Bulmer, M. S. (2007). Relatedness, recognition errors, and colony fusion in the termite *Nasutitermes corniger*. *Behav Ecol Sociobiol*, *61*, 1195–1201.

Bourguignon, T., Chisholm, R. A., & Evans, T. A. (2016). The termite worker phenotype evolved as a dispersal strategy for fertile wingless individuals before eusociality. *Am Nat*, *187*, 372–387.

Bourguignon, T., Lo, N., Cameron, S. L., Šobotník, J., Hayashi, Y., Shigenobu, S., et al. (2015). The evolutionary history of termites as inferred from 66 mitochondrial genomes. *Mol Biol Evol*, *32*, 406–421.

Bourke, A. F. G. (2011). *Principles of social evolution*. Oxford: Oxford University Press.

Bourke, A. F. G. (2014). Hamilton's rule and the causes of social evolution. *Philos Trans R Soc Lond B Biol Sci*, *369*, 20130362.

Breed, M. D. (2014). Kin and nestmate recognition: the influence of W. D. Hamilton on 50 years of research. *Anim Behav*, *92*, 271–279.

Bulmer, M. S., & Traniello, J. F. A. (2002). Agonistic interactions between colonies of the Formosan subterranean termite (Isoptera: Rhinotermitidae) in New Orleans, Louisiana. *J Insect Behav*, *15*, 121–126.

Chown, S. L., Hodgins, K. A., Griffin, P. C., Oakeshott, J. G., Byrne, M., & Hoffmann, A. A. (2015). Biological invasions, climate change and genomics. *Evol Appl*, *8*, 23–46.

Crespi, B. J., & Yanega, D. (1995). The definition of eusociality. *Behav Ecol*, *6*, 109–115.

Dronnet, S., Lohou, C., Christides, J. P., & Bagnères, A.-G. (2006). Cuticular hydrocarbon composition reflects genetic relationship among colonies of the introduced termite *Reticulitermes santonensis* Feytaud. *J Chem Ecol*, *32*, 1027–1042.

Eggleton, P., & Tayasu, I. (2001). Feeding groups, lifetypes and the global ecology of termites. *Ecol Res*, *16*, 941–960.

Evans, T. A. (2011). Invasive termites. In D. E. Bignell, Y. Roisin, & N. Lo (Eds.), *Biology of termites: a modern synthesis* (pp. 519–562). Dordrecht: Springer.

Evans, T. A., Forschler, B. T., & Grace, J. K. (2013). Biology of invasive termites: a worldwide review. *Annu Rev Entomol*, *58*, 455–474.

Feldmeyer, B., Elsner, D., & Foitzik, S. (2014). Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Mol Ecol*, *23*, 151–161.

Florane, C., Bland, J. M., Husseneder, C., & Raina, A. K. (2004). Diet-mediated inter-colonial aggression in the Formosan subterranean termite *Coptotermes formosanus*. *J Chem Ecol*, *30*, 2559–2574.

Foster, K. R., Wenseleers, T., & Ratnieks, F. L. W. (2006). Kin selection is the key to altruism. *Trends Eol Evol*, *21*, 57–60.

Giraud, T., Pedersen, J. S., & Keller, L. (2002). Evolution of supercolonies: The Argentine ants of southern Europe. *Proc Natl Acad Sci USA*, *99*, 6075–6079.

Grimaldi, D. A., & Engel, M. S. (2005). *Evolution of the insects*. Cambridge: Cambridge University Press.

Grosberg, R. K., & Quinn, J. F. (1989.). The evolution of selective aggression conditioned on allorecognition specificity. *Evolution*, *43*, 504-515.

Hanna, C., Cook, E. D., Thompson, A. R., Dare, L. E., Palaski, A. L., Foote, D., & Goodisman, M. A. D. (2013). Colony social structure in native and invasive populations of the social wasp *Vespula pensylvanica*. *Biol Invasions*, *16*, 283–294.

Hanus, R., Vrkoslav, V., Hrdý, I., Cvačka, J., & Šobotník, J. (2010). Beyond cuticular hydrocarbons: evidence of proteinaceous secretion specific to termite kings and queens. *Proc R Soc B-Biol Sci*, *277*, 995–1002.

Harpur, B. A., Kent, C. F., Molodtsova, D., Lebon, J., Alqarni, A. S., Owayss, A. A., & Zayed, A. (2014). Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci USA*, *111*, 2614–2619.

Helanterä, H., Strassmann, J. E., & Carrillo, J. (2009). Unicolonial ants: where do they come from, what are they and where are they going? *Trends Eol Evol*, *24*, 341–349.

Howard, K. J., & Thorne, B. L. (2011). Eusocial evolution in termites and Hymenoptera. In D. E. Bignell, Y. Roisin, & N. Lo (Eds.), *Biology of termites: a modern synthesis* (pp. 97–132). Dordrecht: Springer.

Husseneder, C., Messenger, M. T., Su, N.-Y., Grace, K. J., & Vargo, E. L. (2009). Colony social organization and population genetic structure of an introduced population of Formosan subterranean termite from New Orleans, Louisiana. *J Econ Entomol*, *107*, 1421–1434.

Husseneder, C., Simms, D. M., Delatte, J. R., Wang, C., Grace, K. J., & Vargo, E. L. (2011). Genetic diversity and colony breeding structure in native and introduced ranges of the Formosan subterranean termite, *Coptotermes formosanus*. *Biol Invasions*, *14*, 419–437.

Kapheim, K. M. (2016). Genomic sources of phenotypic novelty in the evolution of eusociality in insects. *Curr Opin Insect Sci*, *13*, 24–32.

Keller, L., & Ross, K. G. (1998). Selfish genes: a green beard in the red fire ant. *Nature*, *394*, 573–575.

Kettler, R., & Leuthold, R. H. (1994). Inter- and intraspecific alarm response in the termite *Macrotermes subhyalinus*. *Insectes Soc*, *42*, 145–156.

Korb, J., Buschmann, M., Schafberg, S., Liebig, J., & Bagnères, A.-G. (2012). Brood care and social evolution in termites. *Proc Biol Sci*, *279*, 2662–2671.

Korb, J., & Schneider, K. (2007). Does kin structure explain the occurrence of workers in a lower termite? *Evol Ecol*, *21*, 817–828.

Legendre, F., Whiting, M. F., & Grandcolas, P. (2013). Phylogenetic analyses of termite post-embryonic sequences illuminate caste and developmental pathway evolution. *Evol Dev*, *15*, 146–157.

Leniaud, L., Pichon, A., Uva, P., & Bagnères, A.-G. (2009). Unicoloniality in *Reticulitermes urbis*: a novel feature in a potentially invasive termite species. *Bull Entomol Res*, *99*, 1–10.

Lo, N., & Eggleton, P. (2011). Termite phylogenetics and co-cladogenesis with symbionts. In D. E. Bignell, Y. Roisin, & N. Lo (Eds.), *Biology of termites: a modern synthesis s* (pp. 27–50). Dordrecht: Springer.

Mikheyev, A. S., & Linksvayer, T. (2015). Genes associated with ant social behavior show distinct transcriptional and evolutionary patterns. *eLife*, *4*, e04775.

Oppenheim, S. J., Baker, R. H., Simon, S., & DeSalle, R. (2015). We can't all be supermodels: the value of comparative transcriptomics to the study of non-model insects. *Insect Mol Biol*, *24*, 139–154.

Penn, D. J., & Frommen, J. G. (2010). Kin recognition: an overview of conceptual issues, mechanisms and evolutionary theory. In P. M. Kappeler (Ed.), *Animal behaviour: evolution and mechanisms* (pp. 55–85). Berlin Heidelberg: Springer-Verlag Berlin Heidelberg.

Perdereau, E., Bagnères, A.-G., Dupont, S., & Dedeine, F. (2010). High occurrence of colony

fusion in a European population of the American termite *Reticulitermes flavipes*. *Insectes Soc*, *57*, 393–402.

Perdereau, E., Bagnères, A.-G., Vargo, E. L., Baudouin, G., Xu, Y., Labadie, P., et al. (2015). Relationship between invasion success and colony breeding structure in a subterranean termite. *Mol Ecol*, *24*, 2125–2142.

Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S., et al. (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc Natl Acad Sci USA*, *111*, 14500–1505.

Ratnieks, F. L. W., Foster, K. R., & Wenseleers, T. (2011). Darwin's special difficulty: The evolution of "neuter insects" and current theory. *Behav Ecol Sociobiol*, *65*, 481–492.

Robinson, G. E., Grozinger, C. M., & Whitfield, C. W. (2005). Sociogenomics: social life in molecular terms. *Nat Rev Genet*, *6*, 257–270.

Roisin, Y., & Korb, J. (2011). Social organisation and the status of workers in termites. In D. E. Bignell, Y. Roisin, & N. Lo (Eds.), *Biology of termites: a modern synthesis* (pp. 133–164). Dordrecht: Springer.

Rousset, F., & Roze, D. (2007). Constraints on the origin and maintenance of genetic kin recognition. *Evolution*, *61*, 2320–2330.

Roux, E. A., & Korb, J. (2004). Evolution of eusociality and the soldier caste in termites: a validation of the intrinsic benefit hypothesis. *J Evol Biol*, *17*, 869–875.

Roux, J., Privman, E., Moretti, S., Daub, J. T., Robinson-Rechavi, M., & Keller, L. (2014). Patterns of positive selection in seven ant genomes. *Mol Biol Evol*, *31*, 1661–1685.

Scaduto, D. A., Garner, S. R., Leach, E. L., & Thompson, G. J. (2012). Genetic evidence for multiple invasions of the Eastern subterranean termite into Canada. *Environ Entomol*, *41*, 1680–1686.

Scharf, M. E. (2015). Omic research in termites: an overview and a roadmap. *Front Genet*, *6*, 76.

Scharf, M. E., Wu-Scharf, D., Pittendrigh, B. R., & Bennett, G. W. (2003). Caste- and development-associated gene expression in a lower termite. *Genome Biol*, *4*, R62.

Shelton, T., & Grace, J. (1997). Suggestion of an environmental influence on intercolony agonism of Formosan subterranean termites (Isoptera: Rhinotermitidae). *Environ Entomol*, *26*, 632–637.

Tarver, M. R., Coy, M. R., & Scharf, M. E. (2012). Cyp15F1: a novel cytochrome P450 gene

linked to juvenile hormone dependent caste differentiation in the termite *Reticulitermes flavipes*. *Arch Insect Biochem Physiol*, *80*, 92–108.

Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., et al. (2014). Molecular traces of alternative social organization in a termite genome. *Nat Commun*, *5*, 3636.

Thompson, G. J., Hurd, P. L., & Crespi, B. J. (2013). Genes underlying altruism. *Biol Lett*, *9*, 20130395.

Thompson, G. J., Kitade, O., Lo, N., & Crozier, R. H. (2000). Phylogenetic evidence for a single, ancestral origin of a "true" worker caste in termites. *J Evol Biol*, *13*, 869–881.

Thompson, G. J., & Oldroyd, B. P. (2004). Evaluating alternative hypotheses for the origin of eusociality in corbiculate bees. *Mol Phylo Evo*, *33*, 452–456.

Thorne, B. L. (1997). Evolution of eusociality in termites. *Annu Rev Ecol Evol Syst*, *28*, 27–54.

Thorne, B. L., Traniello, J. F. A., Adams, E. S., & Bulmer, M. (1999). (Isoptera Rhinotermitidae): a review of the evidence from behavioral, ecological, and genetic studies. *Ethol Ecol Evol*, *11*, 149–169.

Tsutsui, N. D., Suarez, A. V, & Grosberg, R. K. (2003). Genetic diversity, asymmetrical aggression, and recognition in a widespread invasive species. *Proc Natl Acad Sci USA*, *10*, 1078–1083.

van Zweden, J. S., & d'Ettorre, P. (2010). Nestmate recognition in social insects and the role of hydrocarbons. In G. J. Blomquist & A.-G. Bagnères (Eds.), *Insect hydrocarbons: biology biochemistry, and chemical ecology* (pp. 222–243). Cambridge: Cambridge University Press.

Watanabe, D., Gotoh, H., Miura, T., & Maekawa, K. (2014). Social interactions affecting caste development through physiological actions in termites. *Front Physiol*, *5*, 127.

West-Eberhard, M. J. (1996). Wasp societies as microcosms for the study of development and evolution. In S. Turillazzi & M. J. West-Eberhard (Eds.), *Natural history and evolution of paper wasps* (pp. 291–317). Oxford: Oxford University Press.

Wilson, E. O. (1971). *The insect societies*. Cambridge: Harvard University Press.

Zhou, X., Tarver, M. R., & Scharf, M. E. (2007). Hexamerin-based regulation of juvenile hormone-dependent gene expression underlies phenotypic plasticity in a social insect. *Development*, *134*, 601–610.

Chapter 2

2 Caste-biased gene expression in a subterranean termite reveals distinct functional specialization of the soldier caste

2.1 Introduction

Termites comprise a monophyletic clade of eusocial cockroaches (Isoptera; Inward et al., 2007; Lo et al., 2000). As such, termites live in colonies that are characterized by a pronounced reproductive division of labour. The termite king and queen castes are specialized for sexual reproduction, whereas the soldier and worker castes are specialized for non-reproductive roles to support the reproductives in the form of foraging, brood care and nest defense, among other behaviours that ultimately contribute to their indirect fitness (Shellman-Reeve, 1997; Thorne, 1997). The indirect selection of genes for reproductive altruism is predicted from inclusive fitness theory (Bourke, 2011; Hamilton, 1964; Marshall, 2015), but few sociogenomic studies of termites have attempted to relate observed gene expression patterns with developmental or evolutionary processes that give rise to each caste (e.g., Corona et al., 2015; Ishikawa et al., 2010; Miura & Scharf, 2010).

One approach towards understanding how genes affect caste differentiation in termites and other social insects is to first identify genes via their coordinated expression within castes (Robinson et al. 2005; Scharf 2015). Many termite species have a strong bifurcating caste developmental program, whereby early-instar larvae begin to differentiate along reproductive or non-reproductive lines – the former leading to sexual nymphs or imagoes (Figure 2.1). The non-reproductive line, by contrast, can differentiate into a non-reproductive worker caste, some proportion of which further differentiate into sterile soldiers (Noirot & Pasteels, 1987). For termite species that have evolved this distinct bifurcating program (in the families Termitidae, Rhinotermitidae, Hodotermitidae, Mastotermitidae; Roisin & Korb, 2011), I expect workers and soldiers to show fewer gene expression differences than either of these castes shows compared to the reproductive nymphs. After all, soldiers and workers are both reproductively altruistic castes and share a greater portion of their developmental trajectory (Korb & Hartfelder, 2008).

Alternatively, termite soldiers are unique among social insect castes (Roisin, 2000), and by some accounts, may have evolved earlier in termite social evolution than workers did (Thorne et al., 2003). The distinct soldier phenotype therefore may be under the regulation of different or otherwise specialized genes relative to those that distinguish workers from nymphs (Miura & Scharf, 2011). Early gene expression studies on *Reticulitermes* appeared to favour the former scenario, with the majority of genes in targeted screens showing biased expression between reproductive and non-reproductive castes (Scharf et al., 2005; Mitaka et al., 2016). However, this correlation between transcription and reproductive function is not always clear (Steller et al., 2010), and can vary as a function of the genes and samples screened (Table 2.1). An early 24-gene macroarray found that soldiers were distinct in their gene expression from workers and nymphs (Scharf et al., 2003). Comparable studies of caste differences in other termite groups – for example, the dampwood termites, have again found soldiers to be well-separated from other castes (Terrapon et al., 2014). Therefore, it remains uncertain whether the gene expression profile of soldiers is distinct from other castes or if they mostly align with that of workers.

In this study, I used mRNA-Sequencing (RNA-seq) technology and associated bioinformatics analyses to test the two hypotheses outlined above. Specifically, I compared transcriptome-wide gene expression profiles for reproductive (nymph) and non-reproductive (soldier and worker) castes of the Eastern subterranean termite, *R. flavipes* (Rhinotermitidae). I predicted either of two outcomes: comparison of gene-expression differences between the three castes will reveal soldier and worker expression to be more similar than either caste is to the reproductive nymphs (Korb & Hartfelder, 2008) – or reveal that worker and nymph gene expression, by default, are more similar to each other compared to a relatively novel and defensively-specialized soldier caste (Bourguignon et al., 2016). The [worker + soldier] grouping would suggest that shared development and non-reproductive functions explain the underlying expression pattern (Korb & Hartfelder, 2008; Legendre et al., 2013; Roux & Korb, 2004). By contrast, the [worker + nymph] grouping, would suggest that soldiers potentially accrued a gene regulatory pattern that is unique from both the reproductive and the prerequisite non-reproductive worker caste (Bourguignon et al., 2016; Thompson et al., 2004).

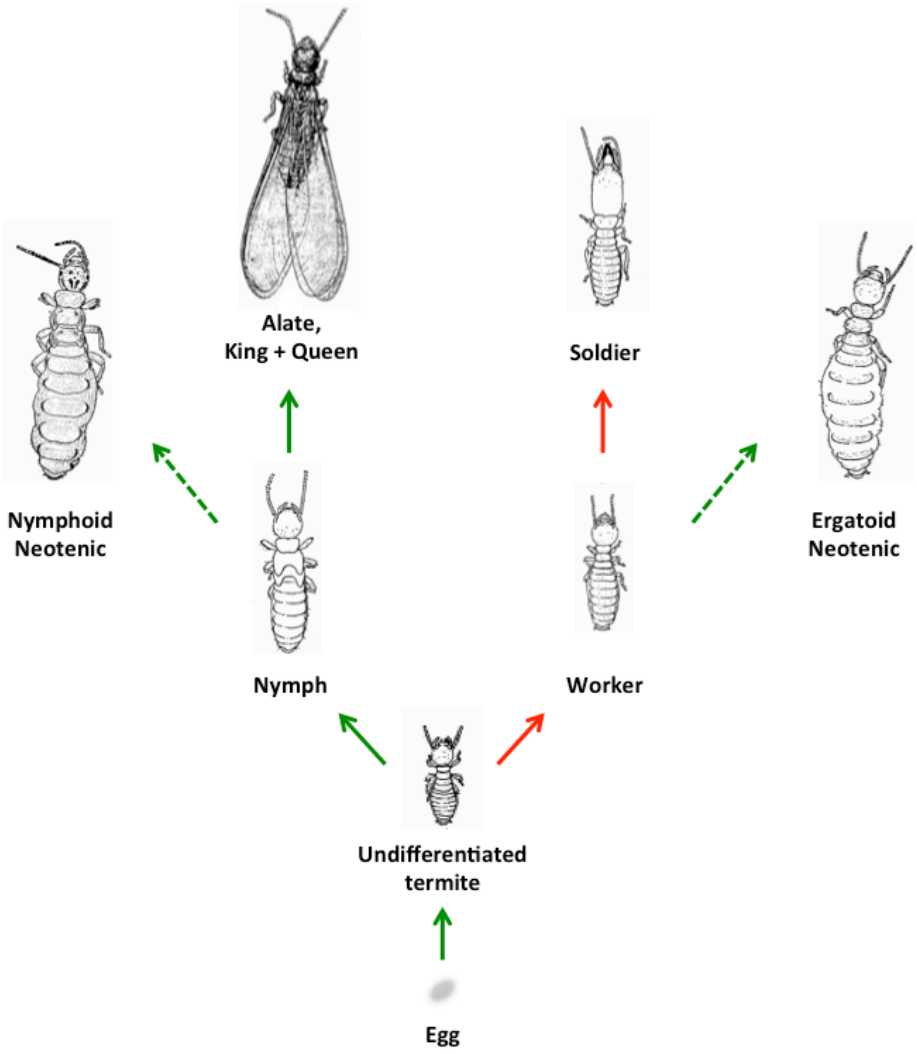


Figure 2.1 *Reticulitermes flavipes* developmental pathways. The egg matures into an undifferentiated juvenile (early-instar), which can then differentiate along reproductive (green) or essentially non-reproductive (red) lines. Nymphs can further differentiate into primary reproductives (alates, subsequently king and queen) or secondary reproductives (nymphoid neotenics). Workers, by contrast, can differentiate into a defensive soldier caste, or rarely, worker-based form of secondary reproductive (ergatoid neotenic). The lifecycle depicted here is simplified from Lainé & Wright (2003).

Table 2.1 Summary of transcriptome analysis studies that compare gene expression among termite castes. The clustering of overall gene expression patterns is shown here in nested bracket notation [N - nymph, S- soldier, W- worker] as inferred from the studies cited. Original studies may have included additional castes (e.g., alates (A), pre-soldiers, etc.), multiple developmental stages (e.g., early vs. late instar), or tissue types (e.g., mandible only, whole body, etc.). I simplify the relevant comparisons here.

Termite Family	Species	Type of screen	Overall expression pattern	Reference
Rhinotermitidae	*Reticulitermes flavipes*	24-gene macroarray	[(N, W), S]	Scharf et al., 2003
		34-gene macroarray	[N, (S, W)]	Scharf et al., 2005
		~15K EST library	[A, S, W]	Steller et al., 2010
	R. speratus	RNA-Seq of 53 chemoreception-related genes	[A, (S, W)]	Mitaka et al., 2016
Archotermopsidae	*Hodotermopsis sjostedti*	Differential display and qPCR of 12 genes	[S, W]	Koshikawa et al., 2005
	Zootermopsis nevadensis	Whole transcriptome RNA-Seq	[(N, W), S]	Terrapon et al., 2014

2.2 Methods

2.2.1 Termite sample collection

In the fall of 2014, I obtained live Eastern subterranean termite (*R. flavipes*) samples from one invasive population in Canada (Toronto, ON) and two native populations in the United States (Boston, MA; Raleigh, NC). Within each population, I sampled termites from three geographically separate colonies (> 1 km), providing a measure of biological replication (Table 2.2). I collected these termites using a live trapping method (*see* Raffoul et al. 2010 for complete details). I dug holes approximately 40 cm deep in soil near known sites of *R. flavipes* infestation and placed a roll of corrugated cardboard (10 x 10 x 12 cm) in the hole along with a piece of plywood (10 x 10 x 1 cm) as a lid. I returned to the trap sites approximately seven days later and dug out the now termite-infested cardboard and nearby soil into a clear plastic container (30 x 15 x 10 cm). The termite samples were then transported back to the lab at the University of Western Ontario (London, Canada), where I flash-froze all field-caught termite samples in liquid nitrogen and stored them at -80°C until RNA extraction. I verified the species identity of termite samples by referencing mitochondrial DNA barcodes from one individual worker per colony against two *R. flavipes*

reference sequences (ISOUS058-12, GBA8807-12) and one outgroup sequence from the congeneric *R. virginicus* (GBA8805-12) in the Barcode of Life Database (BOLD v3; Ratnasingham & Hebert, 2007). For this part of the analysis, I PCR amplified and sequenced a 658 bp fragment of the cytochrome c oxidase subunit I (*COI*) gene using primers LCO1490 and HCO2198 previously developed by Folmer et al. (1994). I then used default settings in GENEIOUS R8 alignment software (Kearse et al., 2012) to estimate the positional homology of all termite sequences. From this alignment, I used GENEIOUS R8 Tree Builder (Kearse et al., 2012) under a Tamura-Nei genetic distance model to infer a neighbour-joining tree of the termite samples. I expected sequences from the same species and population to cluster together on the tree. Finally, I used Lainé & Wright's (2003) morphological descriptions to distinguish *R. flavipes* reproductive (nymph) and non-reproductive (worker and soldier) castes, as well as males from females.

Table 2.2 Specimen localities for *R. flavipes* samples used in this study.

Population	Colony #	Locality	Date Collected	Latitude	Longitude
Boston	1	USA: Massachusetts, Newton 1	2014-09-18	42°20'10.59"N	71°12'26.01"W*
	2	USA: Massachusetts, Newton 2	2014-09-18	42°20'10.59"N	71°12'26.01"W*
	3	USA: Massachusetts, Newton 3	2014-09-18	42°20'10.59"N	71°12'26.01"W*
Raleigh	1	USA: North Carolina, Lake Wheeler	2014-09-18	35°41'40.20"N	78°41'58.10"W
	2	USA: North Carolina, Schenck Forest	2014-10-14	35°48'59.04"N	78°43'36.54"W
	3	USA: North Carolina, Yates Mills	2014-09-18	35°43'17.85"N	78°41'3.77"W
Toronto	1	CAN: Ontario, Toronto, Danforth 1	2014-09-14	43°39'52.64"N	79°21'21.20"W
	2	CAN: Ontario, Toronto, Danforth 2	2014-09-14	43°41'17.50"N	79°18'6.75"W
	3	CAN: Ontario, Toronto, Danforth 3	2014-09-14	43°41'59.75"N	79°15'54.58"W

* approximate coordinate of collection sites.

2.2.2 RNA extractions and mRNA sequencing

I extracted total RNA from six individuals per caste (a male and female pair from each of the three colonies) within each population. I chose to pool termite samples to give a better biological representation of what genes and pathways are involved in caste differentiation into nymphs, soldiers, and workers, regardless of their sexual and population-specific gene

expression differences. This overall sample size and scheme also allowed for a sufficient balance between statistical power and the detection of rare genes and transcripts (Hart et al. 2013). Therefore, I performed three pooled extractions for each caste per population for a total of nine extractions (Table 2.2). Specifically for each extraction, I homogenized whole body tissue in 500 µl of TRIzol (Life Technologies) and added 100 µl of chloroform, as per the TRIzol protocol. I then centrifuged the mixture (10,000 x *g*, 18 mins at 4°C) before transferring the RNA-containing supernatant (250 µl) with an equal volume of ethanol into an RNeasy (Qiagen) mini-column. I followed the RNeasy protocol to elute total RNA in 30 µl of RNase-free water. Finally, I treated the total RNA extract with two units of DNase (Turbo DNA-free kit, Ambion), and equilibrated each sample to ~100 ng/µl. I sent the frozen termite RNA samples to the Next-Generation Sequencing Services at Genome Québec (McGill University, Montréal, Québec; MGU-GQ) for 100 bp paired-end mRNA sequencing. As a measure of quality, the MGU-GQ facility verified that each sample had an RNA integrity number greater than '7' (Bioanalyzer 2100, Agilent Technologies), which is suitable for sequencing library preparation. The MGU-GQ facility then purified mRNA from total RNA, indexed each sample with a unique barcode (Illumina TruSeq mRNA Library Prep Kit v2), and sequenced all nine-termite mRNA libraries on two lanes of an Illumina HiSeq 2000 flow cell.

2.2.3 *De novo* transcriptome assembly

I downloaded paired-end read data files for all nine libraries (totaling ~209 Gb) onto SHARCNET's (Shared Hierarchical Academic Research Computing Network: www.sharcnet.ca) iqaluk server (1TB memory and 32-cores), and performed all downstream analyses via a UNIX command system as follows. From each data file, I trimmed individual reads of adapter sequence and removed any reads that were shorter than 36 bp or that had low base quality scores (lower than '30' in the program TRIMMOMATIC v0.32; Bolger et al., 2014). I then assessed the overall quality of the remaining pair-matched reads with FASTQC's quality control modules (v.0.11.5; Andrews, 2010). Once satisfied with all the pre-assembly criteria in FASTQC, I normalized the dataset to a maximum of 50 reads coverage per gene using TRINITY (v2.1.1; Grabherr et al., 2011). This *in silico* normalization step minimized the overall computational load of the transcriptome assembly process. In total, I assembled a

remainder of ~91 million trimmed and normalized reads under the default settings in TRINITY (Grabherr et al., 2011; Haas et al., 2013).

2.2.4 Functional annotations of genes

I assigned functional annotations to my reference transcriptome assembly using the following strategy. First, I used TRANSDECODER (v2.0.2; Haas et al., 2013) to translate all genes in my *de novo* assembled transcriptome from nucleotide to predicted amino acid sequences. To verify that the genes in my provisional assembly are of insect origin, I conducted BLASTX (for nucleotide sequences) and BLASTP (for amino acid sequences) searches against a custom-made insect-only gene database (Table 2.3). I retained only sequences that shared a minimum 70% amino acid pairwise identity with at least one other insect gene (as in Morandin et al., 2015). Otherwise, sequences were not considered further. To assign putative functions to my gene set, I again used BLASTX and BLASTP, but in this case to identify annotated homologs (e-value < $1e^{-5}$) within the Swiss-Prot database (UniProt, 2014). I further used the program HMMER (v3.1; Eddy, 2011) to infer homology of specific domains against those known from the Protein Family Database (Pfam v29.0; Finn et al., 2015). For this analysis, I used a recommended e-value of $1e^{-10}$ in the TRINOTATE pipeline (V3.0.1; Haas et al., 2013). Finally, I combined gene- and domain-level information into a single annotated gene list and assigned the most likely Gene Ontology (GO) terms using the TRINOTATE program (V3.0.1; Haas et al., 2013). This new reference transcriptome is significant because it represents a new and high quality genomic resource for three different castes of *R. flavipes*.

Table 2.3 Custom insect-only gene database that I used for BLASTX and BLASTP searches to verify the insect origin of the gene sequences in my reference transcriptome.

Order	Species	Version	Reference
Blattodea	*Zootermopsis nevadensis*	2.2	Terrapon et al., 2014
	Macrotermes natalensis	1.2	Poulsen et al., 2014
Diptera	*Drosophila melanogaster*	6.09	Attrill et al., 2016
Hymenoptera (bees)	*Apis mellifera*	3.2	Weinstock et al., 2006
	Bombus impatiens	1.0	Sadd et al., 2015
	Labioglossum albipes	5.42	Kocher et al., 2013
	Nasonia vitripennis	1.2	Werren et al., 2010
Hymenoptera (ants)	*Acromyrmex echinator*	3.8	Nygaard et al., 2011
	Atta cephalotes	1.2	Suen et al., 2011
	Camptonus floridanus	3.3	Bonasio et al., 2010
	Cardiocondyla obscurior	1.4	Schrader et al., 2014
	Harpegnathos saltator	3.3	Bonasio et al., 2010
	Linepithema humile	1.2	Smith et al., 2011a
	Pogomyrmex barbatus	1.2	Smith et al., 2011b
	Solenopsis invicta	2.2.3	Wurm et al., 2011

2.2.5 Differential gene expression and enrichment analysis

To test for caste biased expression patterns, I first used BOWTIE (v1.1.2; Langmead et al., 2009) to map trimmed paired-end reads from each of the nine libraries to the reference transcriptome. I then used the program RSEM (v1.2.25; Li & Dewey, 2011) to estimate raw read counts for each gene for each individual library. I normalized this count data for differences in gene length and library size by converting raw read counts into common units of transcripts per million (TPM). To identify genes differentially expressed between castes, I first grouped the libraries by caste regardless of population – e.g., nymph libraries from Boston, Raleigh and Toronto, and so forth. I then used DESEQ2 (1.10.1; Love et al., 2014) to test for gene expression differences between nymph, soldier, and worker castes. Put simply, DESEQ2 builds a generalized linear model of 'observed' counts for each gene based on dispersion and actual read counts for each sample group. DESEQ2 then performs a Wald test of the model gene counts against the actual read counts for each sample to determine

differential gene expression between nymphs, soldiers, and workers. In my analysis, I specified two additional criteria for individual genes to be considered differentially expressed - namely, i) they show a minimum four-fold expression difference between any two caste groups, and ii) a correction for multiple testing that maintained a false discovery rate (FDR) of less than 0.001 (as in Haas et al., 2013). To test if samples from different populations sorted by caste rather than by population, I compared how similar their transcriptome-wide expression patterns were using pair-wise Pearson's correlations. Here, I expect within-caste samples to show the highest pair-wise correlations regardless of population-of-origin.

To assign provisional biological process (BP) terms to each caste-biased gene, I performed a GO enrichment analysis using GOSEQ (1.22.0; Young et al., 2010). To reduce redundancy amongst the enriched BP-GO terms and better highlight functional categories, I applied the web program REVIGO (Supek et al., 2011) under default parameters, and used the associated SIMREL (Schlicker et al., 2006) to gauge 'semantic similarity' among terms. Finally, I plotted the remaining non-redundant enriched BP-GO terms on a scatterplot (using REVIGO; accessed July 2016) to visualize relationships among the functional gene categories with similar but unique terms closer together and dissimilar terms further apart.

2.3 Results

2.3.1 DNA barcoding species verification

My neighbor-joining analysis of the COI barcodes suggests that all field-caught samples sort according to species and population (Figure 2.2). The *R. flavipes* in-group sequences are distinct from the single outgroup sequence of *R. virginicus*, with an average in-group distance of 0.03 versus 0.09. This pattern, together with my examination of morphological characters confirmed to my satisfaction that all field-caught termites used in this study are *R. flavipes*.

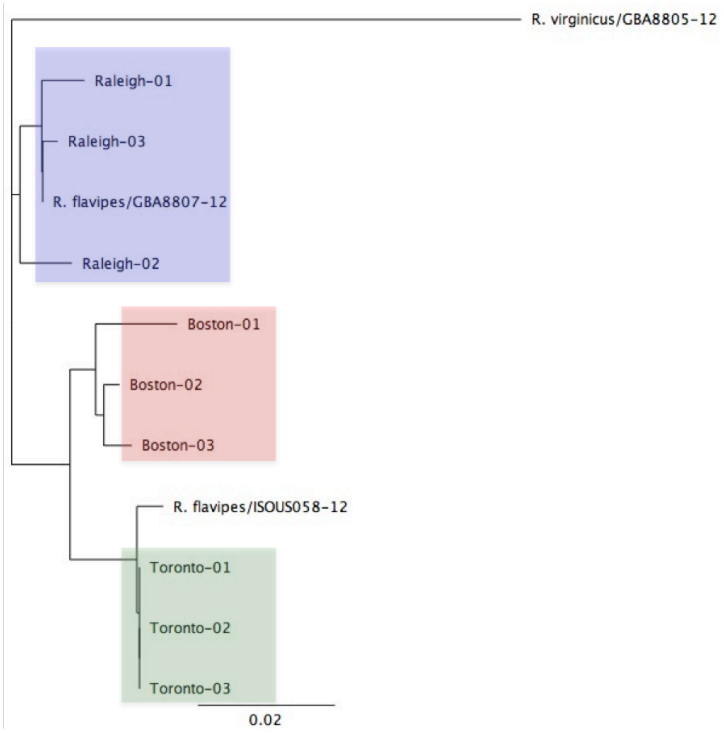


Figure 2.2 Neighbour-joining tree of *R. flavipes* worker mitochondrial COI sequences (n=9) and two *R. flavipes* Barcode of Life Database voucher COI barcodes (GBA8807-12, ISOUS058-12). The lengths of branches reflect genetic distances under a Tamura-Nei model. I also included one *R. virginicus* BOLD voucher barcode (GBA8807-12) as a congeneric outgroup. The coloured blocks simply show how sample sequences sort according to population source. Scale bar represents nucleotide substitutions per site.

2.3.2 *De novo* reference transcriptome assembly

In total, mRNA sequencing of nine *R. flavipes* libraries yielded 433,854,441 raw reads. The number of raw reads per library ranged from 39.6 to 51.3 million, with a mean of 48.2 million reads. After trimming, quality control, and normalization, I retained 91,358,892 reads for use in a *de novo* reference transcriptome assembly (Table 2.4). This assembly and its associated sequencing reads can be accessed from the NCBI Sequence Read Archive (SRA) under the accession numbers SAMN06579168-SAMN06579176. Overall, my raw assembly generated 247,174 Trinity 'genes' that represented 301,796 transcripts. The average gene length was 656 bp and ranged from 201 to 41,323 bp. To minimize the number of spurious gene predictions or contaminant genes of non-termite origin (i.e., bacterial, protist, fungal, or other), I cross-referenced the transcriptome against a custom insect-only gene database and retained only those genes with clear insect homologs (*e*-value < $1e^{-5}$, ≥70% aa identity;

BLASTX and BLASTP). Therefore, I retained a total of 13,755 genes that represented 29,641 transcripts (Table 2.4). This quality control step seemingly improved the raw assembly, as evidenced by a longer average gene length (from 656 to 2,220 bp), a higher N50 score (from 1,099 to 3,087 bp), and a higher number of large transcripts (>1 kb; n = 21,718, 73.2%).

Table 2.4 Summary statistics of sequencing and *R. flavipes de novo* reference transcriptome assembly.

Before quality processing	
Raw reads	433,854,441
After quality processing	
Clean reads	389,173,294
Normalized reads	91,358,892
Trinity assembly statistics	
Number of genes	247,174
Number of transcripts	301,796
Average gene lengths (bp)	656
Post cross reference to insect-only gene database (≥70% aa identity)	
Number of genes	13,755
Number of transcripts	29,641
GC content (%)	41.41
N50 (bp)	3,087
Median gene lengths (bp)	1,774
Average gene lengths (bp)	2,220
Minimum gene lengths (bp)	201
Maximum gene lengths (bp)	41,323

A majority of the 13,755 genes (Table 2.4) had a significant match against the Swiss-Prot database (Table 2.5). Specifically, a BLASTX query yielded 11,071 genes (80.5%) with matches, while a BLASTP query yielded 10,945 genes (79.6%) with matches. A TRANSDECODER scan for putative coding regions within the 13,755 genes also predicted a total of 51,690 ORFs. These gene predictions are more conservative than the estimate from Dedeine et al. (2015) for nymphoid neotenics of *R. flavipes* (19,375 ORFs with annotations of 64,342 total ORFs). Lastly, these genes returned a total of 15,564 unique GO term assignments, with the majority of terms (10,870; 69.8%) representing Biological Process (Table 2.5).

Table 2.5 Gene annotation summary statistics for the 13,755 genes predicted from the *R. flavipes* reference transcriptome.

Annotations	Counts (% of total)
Swiss-Prot BLASTX hits	11,071 (80.5%)
Swiss-Prot BLASTP hits	10,945 (79.6%)
BLAST hits with GO terms	10,519 (76.5%)
Unique GO terms	15,564
GO: Biological Process	10,870 (69.8%)
GO: Molecular Function	3,235 (20.7%)
GO: Cellular Component	1,419 (9.1%)
GO: Unidentified by REVIGO	40 (0.3%)

2.3.3 Gene expression patterns as a function of caste

I found only n = 93 genes (of 13,755; 0.67%) to be strictly differentially expressed as a function of caste (Table S1). These genes are conspicuously (minimum four-fold expression change) and reliably (FDR-corrected *P*-value < 0.001) associated with caste, at least via RNA-seq analysis of my study samples. This caste-biased expression is evident on a gross scale from Figure 2.3, which shows that transcriptome-wide measures of co-expression tend to cluster termite samples within their respective castes. For these genes, caste therefore appears to explain a majority of observed expression differences, with a relatively minor proportion explained by population.

A further gene-level cluster analysis reveals that these caste-informative genes are composed of three tightly co-regulated gene sets, the largest of which (n = 61 genes, Set I) is uniquely up-regulated in the soldier caste, while another set (n = 13 genes, Set III) is uniquely down-regulated in the soldier caste (Figure 2.4). The last set (n = 19 genes, Set II) is uniquely up-regulated in the reproductive (nymph) caste. There is no major gene set that is uniquely up-regulated in workers. In total, 78% (73 of 93) of caste-informative genes show a soldier-biased pattern in gene expression.

I subjected each of the three gene sets - I, II and III - identified from my differential gene expression analysis to a GO enrichment and subsequent REVIGO analysis. I specifically focused on the interpretation of GO terms assigned to Biological Process. Each of my three gene sets generated unique REVIGO scatterplots that varied in number of BP-GO terms and in the nature of their cluster representatives (Figure 2.5). Gene Set I had 129 BP-GO terms

that reduced to n = 12 clusters on the scatterplot, and two of the largest clusters are related to ‘muscle structure development’ and ‘fatty-acyl-CoA metabolism’. Meanwhile, Gene Set II had 146 terms in n = 7 clusters with ‘regulation of oocyte development’ representing nearly half of the remaining non-redundant BP-GO terms. Gene Set III was enriched for just 20 terms and n = 4 clusters, with ‘beta-glucan metabolism’ and ‘cell wall organization’ accounting of half of the remaining BP-GO terms. Furthermore, Figure 2.5 shows minimal overlap in function between cluster representative terms of the three different gene sets. Taken together, these REVIGO scatterplots and enriched GO-terms suggest that each caste-biased gene set is highly specialized for different biological functions.

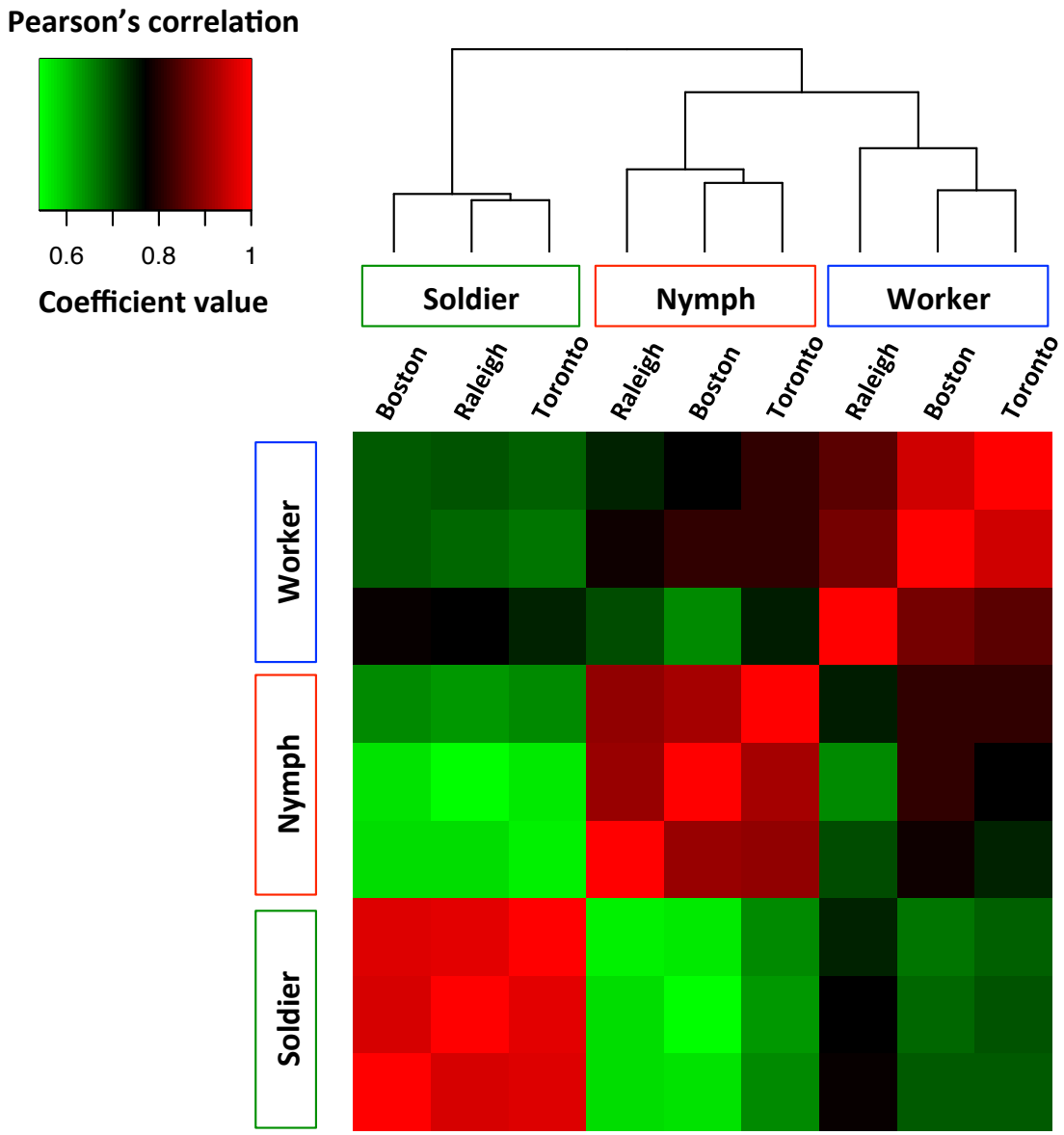


Figure 2.3 A cluster analysis of pairwise Pearson correlation coefficients arrayed in a matrix showing samples with similar (red) or dissimilar (green) gene expression profiles. Each cell represents the average coefficient of a set of n = 93 caste-informative genes. As expected, genes tend to strongly co-vary by caste (soldiers, workers, nymphs) with residual variation from this analysis explained by sample population (Boston, Toronto, Raleigh). Note from the clustering of correlation coefficients that nymphs and workers are more similar in their gene expression than either is to soldiers. Further, the diagram is symmetric across the red-cell diagonal.

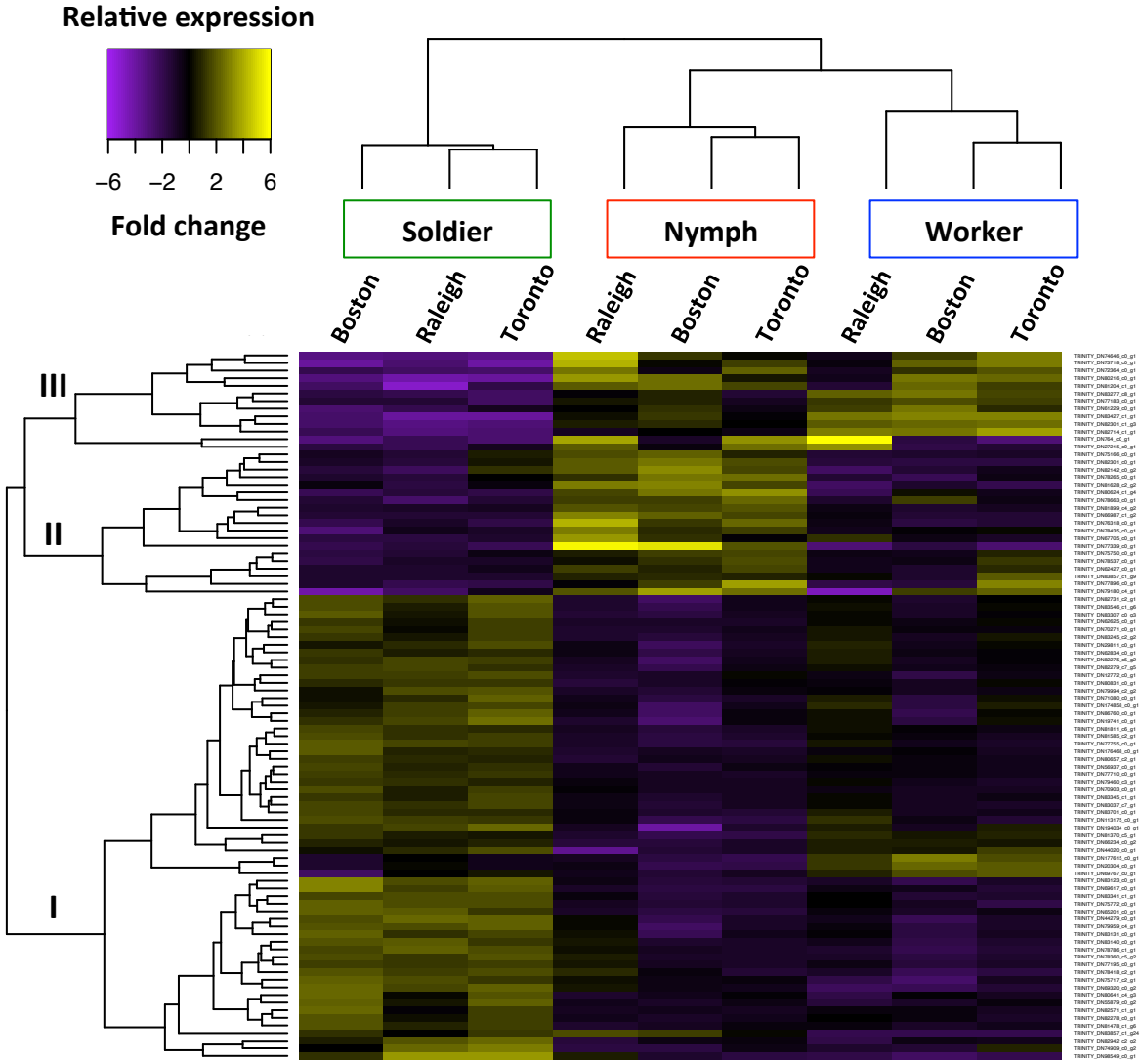


Figure 2.4 Heatmap of 93 genes differentially expressed between *R. flavipes* castes (FDR < 0.001, minimum four-fold change, *see* Methods). The hierarchical clustering of castes (top) is underpinned by the presence of three gene sets – I, II and III (left). The gene sets correspond to nymph, soldier, and nymph/worker gene functions, respectively. Gene identifiers are listed on the far right and can be read by zooming in on the graphic.

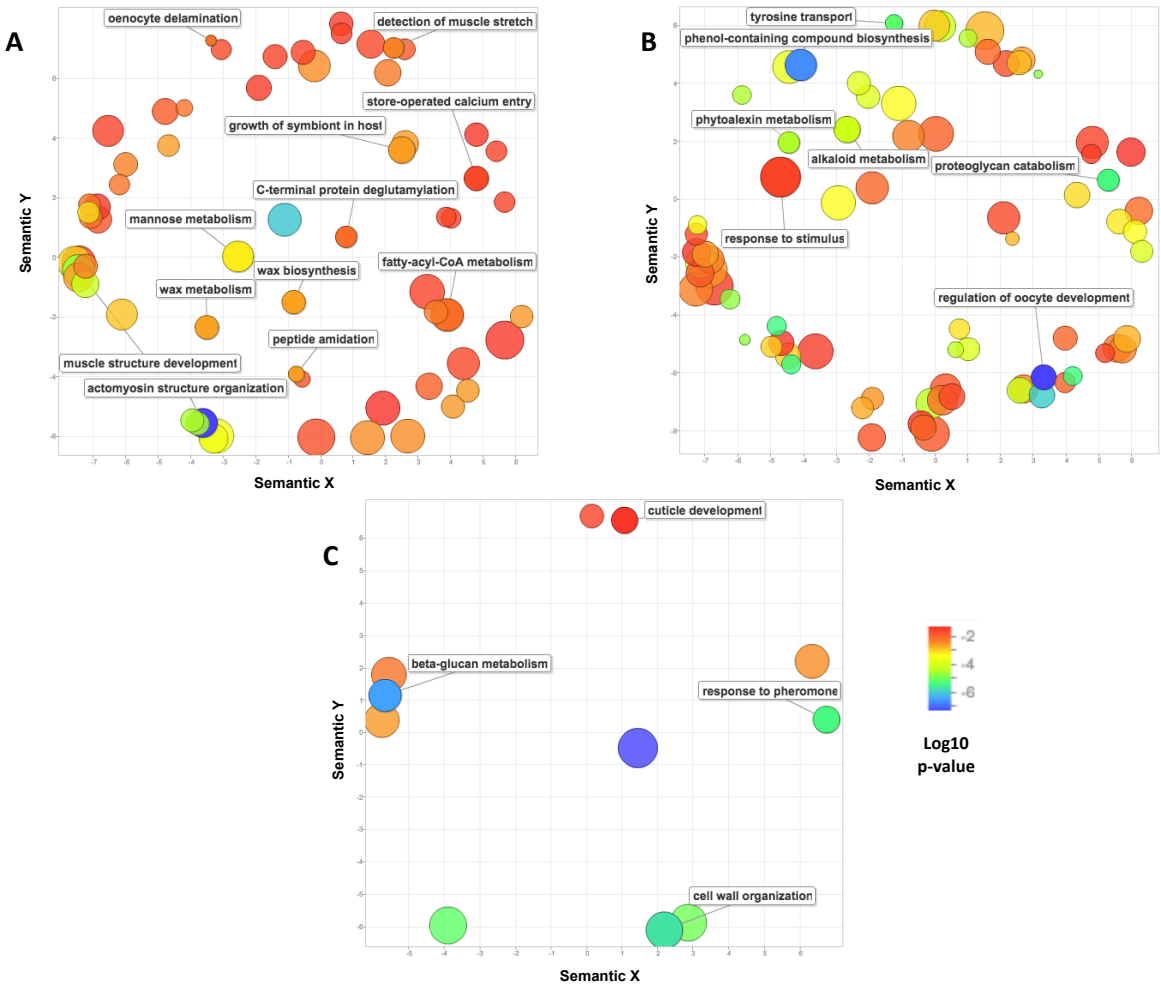


Figure 2.5 Scatterplots displaying the cluster representatives of GO terms in biological processes (BP) derived from *R. flavipes* caste-biased genes in two-dimensional space based on semantic similarity of GO terms. **A)** Gene Set I – 129 BP terms reduced to 62 non-redundant terms, which were further reduced to 12 cluster representatives (as shown), **B)** Gene Set II – 146 BP terms reduced to 76 non-redundant terms and 7 cluster representatives, and **C)** Gene Set III – 20 BP terms reduced to 11 non-redundant terms with 4 cluster representatives. The blue colored bubbles indicate the most significant $\log_{10}$ *P*-values amongst the GO terms. The size of bubbles is scaled to the corresponding term's frequency or generality in the GO annotation resource (Supek et al., 2011). No gene set has any significantly higher proportion of redundancy (49.6% vs. 52.1 % vs. 55.0%; Chi-squared statistic = 0.29, $P > 0.05$).

2.4 Discussion

In this study, I sequenced, assembled, and annotated a reference transcriptome for three castes of the Eastern subterranean termite, *R. flavipes*. This *de novo* assembly is complete with gene expression profiles from one reproductive (nymph) and two non-reproductive (soldiers, worker) castes. From this annotated assembly, I detected a total of 13,755 genes, which provides a rich bioinformatics framework for comparative analyses. In this study, I used this assembly for two purposes. First, I used it as a reference for differential gene expression analysis to identify a small but highly informative set of termite genes that differentiate soldier from worker and nymph transcriptional profiles. This analysis is significant to our understanding of termite caste differentiation because the majority of these caste-informative genes (73 of 93; 78%) are uniquely regulated in soldiers. This specialized defensive caste is therefore well differentiated from both the reproductive and the other non-reproductive caste despite a shared developmental program with workers. Second, I used the annotated assembly as a basis for inferring biological functions of the caste-informative gene sets. My gene ontology analysis revealed that nymph-biased (Set II) and soldier-biased (Sets I and III) gene sets show a high degree of functional specialization, as evidenced by the non-overlap in the three sets of inferred biological process terms. My work therefore shows that the soldier caste of *R. flavipes* has an especially distinct and specialized gene expression program. Moreover, the high-quality reference transcriptome that I have assembled for this economically important and invasive species (Rust & Su, 2012; Vargo & Husseneder, 2009) will facilitate future comparative analyses that probe for other signatures of termite caste or insect social life.

2.4.1 A *de novo* transcriptome for *Reticulitermes flavipes*

This transcriptome is assembled from whole body tissue mRNAs sampled from three different North American populations. As such, it provides a reasonable estimate of gene composition and expression for this species that compliments prior transcriptomic studies of *Reticulitermes* (Dedeine et al 2015; Hayashi et al 2015; Table 2.1) and other termites (Huang, et al., 2012; Ishikawa et al., 2010; Wu et al., 2015). Established indicators of transcriptome quality such as N50 value and average gene length generated for the present assembly (Table 2.4) are consistent with benchmark expectations for non-model organisms (Francis et al.,

2013). Further, the number of predicted genes (n = 13,755) is lower than the estimate from Dedeine et al. (2015; n = 18,323), which included data from three *Reticulitermes* spp. and have retained gene predictions derived from microorganisms such as bacteria and protists. Published gene counts from other termite species - e.g., *Zootermopsis nevadensis* (n = 17,737; Terrapon et al., 2014) and *Macrotermes natalensis* (n = 16,310; Poulsen et al., 2014) - are likewise comparable to the present study. This consistency among studies published so far suggests a fairly uniform gene numbers throughout Isoptera (Korb et al., 2015) despite large differences in overall genome size (Koshikawa, et al., 2008).

For the transcriptome as a whole, the proportion of unknown but presumably functional genes is ~20% (Table 2.5). This proportion is lower than *R. flavipes* EST-sequencing (Steller et al., 2010; Tartar et al., 2009), and is comparable to other termite RNA-seq studies (Dedeine et al., 2015; Hayashi et al., 2013; Wu et al., 2015). The set of 93 caste-informative genes presented in Figure 2.4 are well characterized by homology from other organisms in the Swiss-Prot database: only n = 29 genes (31%) did not have a homolog (Table S2). A majority (18 of 29) of 'unknown' genes were in the relatively large Set I, but the proportion of unknown genes did not vary significantly between sets (29.5% vs. 21.1% vs. 38.4%; Chi-squared statistic = 1.16, $P > 0.05$). Genes with homology assignment from Swiss-Prot were best characterized by Biological Process from GO, and thus for molecular pathways and processes that involve numerous gene products (Ashburner et al., 2000). This may simply reflect the whole-body comparisons I am making between castes, in which phenotypic differences likely involve multi-gene pathways more so than specific cellular components or molecular functions.

It is known that caste differentiation in termites and other social insects typically involves hundreds to thousands of genes (Corona et al., 2015), depending on how it is measured (Rau et al., 2013). In my analysis, I elected to use relatively stringent criteria (including transcriptome filtering, TPM as the standard unit of gene expression measurement, a minimum four-fold change, and FDR of less than 0.001), with the intent of focusing on the most informative gene patterns. Accordingly, the number of differentially expressed genes (DEGs) that I report between castes is relatively small (0.7%) compared to the proportion reported for other termite (~73% for seven castes; Terrapon et al., 2014) and eusocial taxa (~25-50%; Feldmeyer, et al., 2014; Harrison, et al., 2015). Applying these criteria, together

with my sampling of termites from multiple populations, should generate a strong and consistent set of marker genes that are robust to population sampling (Oppenheim et al., 2015). I understand through sampling variability and control measures used in the analysis, I may have eliminated a portion of potentially informative genes (Martin & Wang, 2011). However, my preference is for a minimum set of maximally informative genes. The soldier-biased pattern revealed by my conservative analysis is also robust under relaxed statistical criteria. If I re-test for differential expression at a two-fold level, and hold the FDR to 0.001 or to 0.05, then the number of caste informative genes increases to n = 230 and n = 570, respectively. The overall pattern of [worker + nymph] reported above does not change (Figure S1). I have made the raw assembly and sequencing reads available from the NCBI-SRA database should others wish to analyze them in different ways.

2.4.2 Biological processes associated with termite castes

The large gene Set I had many cluster representative terms (n = 12; Figure 2.5A), which suggests that soldier up-regulated genes perform a range of specialized functions. These functions include 'muscle structure development' that is linked to soldier defense via the muscle force of their enlarged mandibles (Prestwich, 1984; Scharf et al., 2003) and 'wax-' or 'fatty-acyl-CoA -metabolism' that may be involved in caste status signaling via cuticular hydrocarbon profiling (Liebig et al., 2009). I also retrieved the term ‘oenocyte delamination’, a process that has previously been implicated in soldier caste differentiation (Corona et al., 2015; Makki et al., 2014). Other terms like 'growth of symbionts' and 'peptide amidation’ for example, each suggest how soldiers may function differently to workers and nymphs. The small gene Set III has few cluster representative terms (n = 4; Figure 2.5C) but likewise reflects differences in soldier function (the genes in this set are uniquely down-regulated in soldiers, relative to workers and nymphs). These soldier-biased functions include ‘cell-wall organization’ and ‘beta-glucan metabolism’, which commonly arise in termite gene-function studies (reviewed in Lo et al., 2011). I speculate that down-regulation of genes involved in these processes for soldiers are related to their dependency on being fed by workers via trophallaxis (Korb et al., 2012). Termite workers have enzymes that can digest lignocellulose (Karl & Scharf, 2015), hyphae (Poulsen et al., 2014), and may otherwise protect against fungal pathogens (Rosengaus et al., 2014). Set III also includes the gene hexamerin, a juvenile hormone regulator, which has been singularly implicated in termite caste

differentiation (Zhou et al., 2006), and its specific down-regulation in soldiers matches results previously described in *R. flavipes* (Scharf et al., 2005). Finally, cluster representative terms in gene Set II include those related to metabolism or catabolism (3 terms), and 'regulation of oocyte development' (Figure 2.5B). This latter term is intriguing given the reproductive potential of nymphs. Other transcriptome studies of social insects have also noted this co-regulation between reproduction and metabolism (Corona et al., 2007; Hattori et al., 2013).

2.4.3 Gene expression of soldiers is unique

Reticulitermes flavipes nymphs, soldiers, and workers showed distinctive gene expression profiles that reflect caste differentiation and specialization. I found that soldiers have the most unique expression pattern that is distinct from the other two castes (Figure 2.4). As a whole, 78% of caste-informative genes (73 of 93 - i.e., Sets I and III) show soldier-specific regulation. This trend toward soldier-specific regulation is consistent with a pattern first detected by Scharf et al. (2003) in *R. flavipes*, and again by Terrapon et al. (2014) for *Z. nevadensis*. Therefore, there is precedence for an overall [(nymph, worker), soldier] pattern of transcriptional differentiation, which is consistent with the evolutionary idea that the termite soldier caste is a distinct social phenotype (Noirot & Pasteels, 1987; Roux & Korb, 2004; Scholtz et al., 2008; Yamamura, 1993). An alternative pattern: [nymph, (soldier, worker)], by contrast, in which the two non-reproductive castes are grouped is not supported. This is intriguing for two reasons. First, soldiers and workers are both non-reproductive and have most likely been indirectly selected to perform helper roles within the colony (Thorne, 1997). At some level, I expected soldiers and workers to share a common pattern of gene regulation that is related to their overlapping functional roles, regardless of any shared development. Second, future soldiers develop as workers for up to nine instars prior to their terminal differentiation via a pre-soldier stage (Lainé & Wright, 2003; Roisin & Korb, 2011). I therefore might likewise expect soldiers and workers to share a common pattern of gene regulation that is related to common development. Regardless, I instead observed that the most-informative set of caste-biased genes do not cluster soldiers with workers, but rather workers with nymphs, suggesting that soldiers are a potential source of genetic novelty in termites.

There is some uncertainty about the evolutionary status of termite soldiers. First, despite their clear developmental differentiation from workers, it is not obvious whether soldiers evolved from, or independently of a worker caste (Legendre et al., 2013), or even if workers themselves evolved once (Bourguignon et al., 2016; Thompson et al., 2004) or on multiple occasions (Roisin & Korb, 2011). Regardless of the correct macroevolutionary pattern of ancestral termite caste evolution, my analyses suggest that extant *Reticulitermes* soldiers are not just modified workers, but represent a unique caste with novel gene expression. Whether the transcriptome-wide pattern for soldier novelty observed here reflects ancient or more recent bouts of selection should be the focus of a future study. One possibility might be to test if soldier biased genes are taxonomically rare, suggesting recent gene recruitment during soldier evolution, as has been reported for helper castes in some social hymenopterans (Feldmeyer et al., 2014; Jasper et al., 2015; Johnson & Tsutsui, 2011). Alternatively, it should be possible to test soldier-biased genes for signatures of positive selection. A disproportionately high rate of non-synonymous substitution would support selection for genetic novelty during soldier evolution. This latter approach to understand the role of 'novelty' in helper caste evolution (Kapheim, 2016; Sumner, 2014) was likewise informative for recent studies of some social hymenopterans (Harpur et al., 2014; Roux et al., 2014). For termites, given that soldiers are strictly sterile and thus have no direct fitness, any evidence for positive selection would imply that past selection for novelty must have been indirect (*sensu* Hamilton 1964).

The pattern for soldier-biased gene expression is strong, but it is not uniform across all 93 caste-informative genes. For example, gene Set II (n = 19) is nymph-biased, and thus separates reproductive from non-reproductive lines, as might generally have been expected (*see* Introduction). Set II in isolation is therefore consistent with a [nymph, (soldier, worker)] pattern, and more generally is consistent with the separation of reproductive versus non-reproductive castes. This isolated pattern reinforces our earlier observation - that the (understandably) limited gene screens of early caste transcription studies can bias our view of the overall pattern. Nonetheless, this observed correlation between gene expression and division of reproductive division of labour in Set II is similar to the pattern reported for chemosensory-related genes in *R. speratus* (Mitaka et al., 2016) and for a set of 34 genes *a priori* known to be nymph-biased in a *R. flavipes* study (Scharf et al., 2005). Lastly, my

analysis did not reveal any major gene set uniquely regulated in workers. However, three genes in Set I are consistently and uniquely up-regulated in workers. These genes are: two lysosomal alpha-mannosidases and one gene of unknown function. The absence of a clear worker signal in my gene set is unexpected, but there are several possible explanations. First, unlike soldiers that are singularly specialized for defense (Noirot & Pasteels, 1987; Scholtz et al., 2008), workers are relatively generic and perform a wide range of roles (Roisin, 2000). Furthermore, workers are relatively flexible in their development, both in terms of caste end-points (Korb & Hartfelder, 2008) and instars (Lainé & Wright, 2003). Thus, while workers do not appear to show many uniquely regulated genes, they nonetheless show a unique patchwork of expression that is distinct from reproductive (nymph) and other non-reproductive (soldier) castes, at least in *R. flavipes*.

2.4.4 Conclusion

Overall, I established a reference transcriptome for nymph, soldier, and worker castes of *R. flavipes*. I found transcriptional patterns that support a close relationship between nymphs and workers, with soldiers being the most unique of the three castes examined. Specifically, the soldiers up-regulated the largest number of genes, and had the most diverse gene enrichment profile with minimal overlap with nymphs and workers. Moreover, I examined each caste-biased gene set for biological functions that contribute to the molecular caste differentiation process, and I found several well-known genes along with a small portion of unannotated genes. The findings in this study support the notions of nymph-worker first caste evolution and genetic caste differentiation. Future transcriptomic and genome-level studies should target the transitional stages of *R. flavipes* caste prior to complete differentiation, and expand sampling to include alates, kings and queens, and pre-soldiers to clarify their gene regulatory relationships and thus development and evolution. From data sets currently available for the *R. flavipes* (Dedeine et al., 2015), we have complete transcriptomes for two reproductive castes (nymph, nymphoid neotenic) and two non-reproductive castes (soldier, worker). Hence, more expansive studies seem possible, provided the meta-data can be normalized into a common analytical framework. Future studies may also examine the taxonomic novelty of caste-specific genes to test whether soldiers, or other castes, are a genetic - in addition to genomic - source of evolutionary innovation.

2.5 References

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. *Available*, http://www.bioinformatics.babraham.ac.uk/projects/.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat Genet*, *25*, 25–29.

Attrill, H., Falls, K., Goodman, J. L., Millburn, G. H., Antonazzo, G., Rey, A. J., et al. (2016). FlyBase: establishing a gene group resource for *Drosophila melanogaster*. *Nucleic Acids Res*, *44*, D1-7.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*, 2114–2120.

Bonasio, R., Zhang, G., Ye, C., Mutti, N. S., Fang, X., Qin, N., et al. (2010). Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science*, *329*, 1068–1071.

Bourguignon, T., Chisholm, R. A., & Evans, T. A. (2016). The termite worker phenotype evolved as a dispersal strategy for fertile wingless individuals before eusociality. *Am Nat*, *187*, 372–387.

Bourke, A. F. G. (2011). The validity and value of inclusive fitness theory. *Proc R Soc B-Biol Sci*, *278*, 3313–3320.

Corona, M., Libbrecht, R., & Wheeler, D. E. (2015). Molecular mechanisms of phenotypic plasticity in social insects. *Curr Opin Insect Sci*, *13*, 10–15.

Corona, M., Velarde, R. A., & Remolina, S. (2007). Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proc Natl Acad Sci USA*, *104*, 7128–7133.

Dedeine, F., Weinert, L. A., Bigot, D., Josse, T., Ballenghien, M., Cahais, V., et al. (2015). Comparative analysis of transcriptomes from secondary reproductives of three *Reticulitermes* termite species. *PLoS ONE*, *10*, e0145596.

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput Biol*, *7*, e1002195.

Feldmeyer, B., Elsner, D., & Foitzik, S. (2014). Gene expression patterns associated with caste and reproductive status in ants: worker‐specific genes are more derived than queen‐specific ones. *Mol Ecol*, *23*, 151–161.

Finn, R. D., Clements, J., Arndt, W., Miller, B. L., Wheeler, T. J., Schreiber, F., et al. (2015). HMMER web server: 2015 update. *Nucleic Acids Res*, *43*, W30-38.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*, *3*, 294–299.

Francis, W. R., Christianson, L. M., & Kiko, R. (2013). A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly. *BMC Genomics*, *14*, 167.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*, *29*, 644–652.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, *8*, 1494–1512.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *J Theor Biol*, *7*, 17–52.

Harpur, B. A., Kent, C. F., Molodtsova, D., Lebon, J., Alqarni, A. S., Owayss, A. A., & Zayed, A. (2014). Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci USA*, *111*, 2614–2619.

Harrison, M. C., Hammond, R. L., & Mallon, E. B. (2015). Reproductive workers show queenlike gene expression in an intermediately eusocial insect, the buff‐tailed bumble bee *Bombus terrestris*. *Mol Ecol*, *24*, 3043–3063.

Hattori, A., Sugime, Y., Sasa, C., Miyakawa, H., Ishikawa, Y., Miyazaki, S., et al. (2013). Soldier morphogenesis in the damp‐wood termite is regulated by the insulin signaling pathway. *J Exp Zool B Mol Dev Evol*, *320*, 295–306.

Hart, S., Therneau, T., Zhang, Y., Poland, G., & Kocher, J.-P. (2013). Calculating sample size estimates for RNA Sequencing data. *J Comput Biol*, *20*, 970–980.

Hayashi, Y., Shigenobu, S., Watanabe, D., Toga, K., Saiki, R., Shimada, K., et al. (2013). Construction and characterization of normalized cDNA libraries by 454 pyrosequencing and estimation of DNA methylation levels in three distantly related termite species. *PLoS ONE*, *8*, e76678.

Huang, Q., Sun, P., Zhou, X., & Lei, C. (2012). Characterization of head transcriptome and analysis of gene expression involved in caste differentiation and aggression in *Odontotermes formosanus* (Shiraki). *PLoS ONE*, *7*, e50383.

Ishikawa, Y., Okada, Y., Ishikawa, A., Miyakawa, H., Koshikawa, S., & Miura, T. (2010).

Gene expression changes during caste-specific neuronal development in the damp-wood termite *Hodotermopsis sjostedti*. *BMC Genomics*, *11*, 314.

Jasper, W. C., Linksvayer, T. A., Atallah, J., Friedman, D., Chiu, J. C., & Johnson, B. R. (2015). Large-scale coding sequence change underlies the evolution of postdevelopmental novelty in honey bees. *Mol Biol Evol*, *32*, 334–346.

Johnson, B. R., & Tsutsui, N. D. (2011). Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics*, *12*, 164.

Kapheim, K. M. (2016). Genomic sources of phenotypic novelty in the evolution of eusociality in insects. *Curr Opin Insect Sci*, *13*, 24–32.

Karl, Z. J., & Scharf, M. E. (2015). Effects of five diverse lignocellulosic diets on digestive enzyme biochemistry in the termite *Reticulitermes flavipes*. *Arch Insect Biochem Physiol*, *90*, 89–103.

Kearse, M., Moir, R., Wilson, A., & Stones-Havas, S. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*, 1647–1649.

Kocher, S. D., Li, C., Yang, W., Tan, H., Yi, S. V, Yang, X., et al. (2013). The draft genome of a socially polymorphic halictid bee, *Lasioglossum albipes*. *Genome Biol*, *14*, R142.

Korb, J., Buschmann, M., Schafberg, S., Liebig, J., & Bagnères, A.-G. (2012). Brood care and social evolution in termites. *Proc R Soc B-Biol Sci*, *279*, 2662–2671.

Korb, J., & Hartfelder, K. (2008). Life history and development – a framework for understanding developmental plasticity in lower termites. *Biol Rev Camb Philos Soc*, *83*, 295–313.

Korb, J., Poulsen, M., Hu, H., Li, C., Boomsma, J. J., Zhang, G., & Liebig, J. (2015). A genomic comparison of two termites with different social complexity. *Front Genet*, *6*, 9.

Koshikawa, S., Cornette, R., Hojo, M., Maekawa, K., Matsumoto, T., & Miura, T. (2005). Screening of genes expressed in developing mandibles during soldier differentiation in the termite *Hodotermopsis sjostedti*. *FEBS Lett*, *579*, 1365–1370.

Koshikawa, S., Miyazaki, S., Cornette, R., Matsumoto, T., & Miura, T. (2008). Genome size of termites (Insecta, Dictyoptera, Isoptera) and wood roaches (Insecta, Dictyoptera, Cryptocercidae). *Naturwissenschaften*, *95*, 859–867.

Lainé, L. V, & Wright, D. J. (2003). The life cycle of *Reticulitermes* spp. (Isoptera: Rhinotermitidae): what do we know? *Bull Entomol Res*, *93*, 267–378.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, *10*, R25.

Legendre, F., Whiting, M. F., & Grandcolas, P. (2013). Phylogenetic analyses of termite post‐embryonic sequences illuminate caste and developmental pathway evolution. *Evol Dev*, *15*, 146–157.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323.

Liebig, J., Eliyahu, D., & Brent, C. S. (2009). Cuticular hydrocarbon profiles indicate reproductive status in the termite *Zootermopsis nevadensis*. *Behav Ecol Sociobiol*, *63*, 1799–1807.

Lo, N., Tokuda, G., & Watanabe, D. (2011). Evolution and function of endogenous termite cellulases. In D. E. Bignell, Y. Roisin, & N. Lo (Eds.), *Biology of termites: a modern synthesis* (pp. 51–67). Dordrecht: Springer.

Lo, N., Tokuda, G., Watanabe, H., Rose, H., Slaytor, M., Maekawa, K., et al. (2000). Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Curr Biol*, *10*, 801–804.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*, 550.

Makki, R., Cinnamon, E., & Gould, A. P. (2014). The development and functions of oenocytes. *Annu Rev Entomol*, *59*, 405–425.

Marshall, J. A. R. (2015). *Social evolution and inclusive fitness theory: an introduction*. Princeton, USA: Princeton University Press.

Martin, J. A., & Wang, Z. (2011). Next-generation transcriptome assembly. *Nat Rev Genet*, *12*, 671–682.

Mitaka, Y., Kobayashi, K., Mikheyev, A., Tin, M. M. Y., Watanabe, Y., & Matsuura, K. (2016). Caste-specific and sex-specific expression of chemoreceptor genes in a termite. *PLoS ONE*, *11*, e0146125.

Miura, T., & Scharf, M. E. (2011). Molecular basis underlying caste differentiation in termites. In D. E. Bignell, Y. Roisin, & N. Lo (Eds.), *Biology of termites: a modern synthesis* (pp. 255–277). Dordrecht: Springer.

Morandin, C., Dhaygude, K., Paviala, J., Trontti, K., Wheat, C., & Helanterä, H. (2015).

Caste‐biases in gene expression are specific to developmental stage in the ant *Formica exsecta*. *J Evol Biol*, *28*, 1705–1718.

Noirot, C., & Pasteels, J. M. (1987). Ontogenetic development and evolution of the worker caste in termites. *Experientia*, *43*, 851–952.

Nygaard, S., Zhang, G., Schiøtt, M., Li, C., Wurm, Y., Hu, H., et al. (2011). The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Res*, *21*, 1339–1348.

Oppenheim, S. J., Baker, R. H., Simon, S., & DeSalle, R. (2015). We can't all be supermodels: the value of comparative transcriptomics to the study of non‐model insects. *Insect Mol Biol*, *24*, 139–154.

Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S., et al. (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc Natl Acad Sci USA*, *111*, 14500–1505.

Prestwich, G. (1984). Defense mechanisms of termites. *Annu Rev Entomol*, *29*, 201–232.

Raffoul, M., Hecnar, S. J., Prezioso, S., Hecnar, D. R., & Thompson, G. J. (2011). Trap response and genetic structure of Eastern subterranean termites (Isoptera: Rhinotermitidae) in Point Pelee National Park, Ontario, Canada. *Can Ent*, *143*, 263–271.

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol Ecol Notes*, *7*, 355–364.

Rau, A., Gallopin, M., Celeux, G., & Jaffrézic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics*, *29*, 2146–2152.

Roisin, Y. (2000). Diversity and evolution of caste patterns. In Abe, T., D. E. Bignell, & Higashi, M. (Eds), *Termites: evolution sociality, symbiosis, ecology* (pp. 95–120). Dordrecht: Kluwer Academic.

Roisin, Y., & Korb, J. (2011). Social organisation and the status of workers in termites. In D. E. Bignell, Y. Roisin, & N. Lo (Eds.), *Biology of termites: a modern synthesis* (pp. 133–164). Dordrecht: Springer.

Rosengaus, R. B., Schultheis, K. F., Yalonetskaya, A., Bulmer, M. S., DuComb, W. S., Benson, R. W., et al. (2014). Symbiont-derived β-1,3-glucanases in a social insect: mutualism beyond nutrition. *Front Microbiol*, *5*, 607.

Roux, E. A., & Korb, J. (2004). Evolution of eusociality and the soldier caste in termites: a

validation of the intrinsic benefit hypothesis. *J Evol Biol*, *17*, 869–875.

Roux, J., Privman, E., Moretti, S., Daub, J. T., Robinson-Rechavi, M., & Keller, L. (2014). Patterns of positive selection in seven ant genomes. *Mol Biol Evol*, *31*, 1661–1685.

Rust, M. K., & Su, N.-Y. (2012). Managing social insects of urban importance. *Annu Rev Entomol*, *57*, 355–375.

Sadd, B. M., Barribeau, S. M., Bloch, G., de Graaf, D. C., Dearden, P., Elsik, C. G., et al. (2015). The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol*, *16*, 76.

Scharf, M. E., Wu-Scharf, D., Pittendrigh, B. R., & Bennett, G. W. (2003). Caste- and development-associated gene expression in a lower termite. *Genome Biol*, *4*, R62.

Scharf, M. E., Wu‐Scharf, D., Zhou, X., Pittendrigh, B. R., & Bennett, G. W. (2005). Gene expression profiles among immature and adult reproductive castes of the termite *Reticulitermes flavipes*. *Insect Mol Biol*, *14*, 31–44.

Schlicker, A., Domingues, F. S., Rahnenführer, J., & Lengauer, T. (2006). A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics*, *7*, 302.

Scholtz, O. I., MacLeod, N., & Eggleton, P. (2008). Termite soldier defence strategies: a reassessment of Prestwich's classification and an examination of the evolution of defence morphology using extended eigenshape analyses of head morphology. *Zoo J Linnean Soc*, *153*, 631–650.

Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., et al. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat Commun*, *5*, 5495.

Shellman-Reeve, J. S. (1997). The spectrum of eusociality in termites. In Choe, J. C. & Crespi, B. J., (Eds.), *The evolution of social behavior in insects and arachnids* (pp. 52–93). Cambridge: Cambridge University Press.

Smith, C. D., Zimin, A., Holt, C., Abouheif, E., Benton, R., Cash, E., et al. (2011). Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proc Natl Acad Sci USA*, *108*, 5673–5678.

Smith, C. R., Smith, C. D., Robertson, H. M., Helmkampf, M., Zimin, A., Yandell, M., et al. (2011). Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc Natl Acad Sci USA*, *108*, 5667–5672.

Steller, M. M., Kambhampati, S., & Caragea, D. (2010). Comparative analysis of expressed sequence tags from three castes and two life stages of the termite *Reticulitermes flavipes*. *BMC Genomics*, *11*, 463.

Suen, G., Teiling, C., Li, L., Holt, C., Abouheif, E., Bornberg-Bauer, E., et al. (2011). The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. *PLoS Genet*, *7*, e1002007.

Sumner, S. (2014). The importance of genomic novelty in social evolution. *Mol Ecol*, *23*, 26–28.

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, *6*, e21800.

Tartar, A., Wheeler, M. M., Zhou, X., Coy, M. R., Boucias, D. G., & Scharf, M. E. (2009). Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnol Biofuels*, *2*, 25.

Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., et al. (2014). Molecular traces of alternative social organization in a termite genome. *Nat Commun*, *5*, 3636.

Thompson, G. J., Kitade, O., Lo, N., & Crozier, R. H. (2004). On the origin of termite workers: weighing up the phylogenetic evidence. *J Evol Biol*, *17*, 217–220.

Thorne, B. L. (1997). Evolution of eusociality in termites. *Annu Rev Ecol Evol Syst*, *28*, 27–54.

Thorne, B. L., Breisch, N. L., & Muscedere, M. L. (2003). Evolution of eusociality and the soldier caste in termites: influence of intraspecific competition and accelerated inheritance. *Proc Natl Acad Sci USA*, *100*, 12808–12813.

Tokuda, G., Lo, N., Watanabe, H., Arakawa, G., Matsumoto, T., & Noda, H. (2004). Major alteration of the expression site of endogenous cellulases in members of an apical termite lineage. *Mol Ecol*, *13*, 3219–3228.

UniProt. (2014). Activities at the universal protein resource (UniProt). *Nucleic Acids Res*, *42*, D191-198.

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

Weinstock, G. M., Robinson, G. E., Gibbs, R. A., Worley, K. C., Evans, J. D., Maleszka, R., et al. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature*, *443*, 931–949.

Werren, J., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., Colborne, J. K., et al. (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science*, *327*, 343–348.

Wu, W., Huang, Z., Li, Z., Zhang, S., Liu, X., & Gu, D. (2015). De novo transcriptome sequencing of *Cryptotermes domesticus* and comparative analysis of gene expression in response to different wood species. *Gene*, *575*, 655–666.

Wurm, Y., Wang, J., Riba-Grognuz, O., Corona, M., Nygaard, S., Hunt, B. G., et al. (2011). The genome of the fire ant *Solenopsis invicta*. *Proc Natl Acad Sci USA*, *108*, 5679–5684.

Yamamura, N. (1993). Different evolutionary conditions for worker and soldier castes: genetic systems explaining caste distribution among eusocial insects. *J Theor Biol*, *161*, 111–117.

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*, *11*, R14.

Zhou, X., Oi, F. M., & Scharf, M. E. (2006). Social exploitation of hexamerin: RNAi reveals a major caste-regulatory factor in termites. *Proc Natl Acad Sci USA*, *103*, 4499–4504.

Chapter 3

3 Invasiveness and population specific transcriptomic correlates in a subterranean termite

3.1 Introduction

Invasive insects threaten native environments if they displace or otherwise out-compete locally-adapted species (Kenis et al., 2009; Pyšek & Richardson, 2010). Inadvertently transported by humans around the world, invasive insects can be introduced into new habitats quickly (Holmes, et al., 2009; Pejchar & Mooney, 2009), and often with negative ecological and economic consequences (Lowe et al., 2000). The annual global cost of controlling invasive insects is estimated to be approximately US$70 billion (Bradshaw et al., 2016). With continued demand for international commerce and anticipated effects of climate change, the spread of invasive insect species is expected to increase (Buczkowski & Bertelsmeier, 2017; Hellmann et al., 2008). Therefore, there is strong global interest and urgency to understand how successful insect invasions occur.

Several termite species are listed among the most destructive invasive species in terrestrial ecosystems (Evans, et al., 2013, Lowe et al., 2000). Because of their division of labour into reproductive and non-reproductive helper castes, the introduction of a few sexual reproductives can quickly propagate a large number of offspring, which then work to establish resilient colonies (Grace, 1996a; Thorne & Traniello, 2003; Vargo et al., 2013). Invasive termite colonies have extremely low genetic diversities as a consequence of the 'bottleneck'-like effects from having low number of founding individuals during their introduction. Furthermore, these invasive termites have diffuse colony boundaries where multiple secondary reproductives and non-reproductives from nearby colonies can intermingle with each other, and appear to lack the typical intercolony aggression observed in native populations (Vargo & Husseneder, 2011). It is also likely this intermixing and exchange of social immunity between colonies helps them delay or overcome any negative effects associated with their introduction (Ugelvig & Cremer, 2012). As a result, these so-called 'supercolonies' can grow several times larger than their native form and are much more difficult to eradicate (Evans, 2011). For one special case, two invasive termite species

in Florida, the Asian subterranean termite (*Coptotermes formosanus*) and the Formosan subterranean termite (*C. gestroi*) were shown to produce viable hybrid colonies in experimental settings (Chouvenc, et al., 2015). It is evident there are important genetic changes and consequences associated with invasions as they could be major contributing factors to successful establishment of invasive termites.

The Eastern subterranean termite, *Reticulitermes flavipes*, is endemic to the Eastern United States (Weesner (1970), but has since invaded other parts of North (McKern et al., 2006; Urquhart, 1953) and South America (Aber & Fontes, 1993), along with multiple cities in Europe (Becker, 1970; Ghesini & Marini, 2009; Vieau, 2001). In this study, I focus on one major invasive population of *R. flavipes* in the city of Toronto, Ontario, Canada. Introduced in the 1930's, *R. flavipes* quickly established itself as a persistent soil-dwelling and wood-eating pest throughout many parts of the city (Myles, 2004; Urquhart, 1953) and beyond (Raffoul et al., 2011). Unlike their native counterparts, the invasive colonies appear to display supercolonial-like characteristics (Scaduto et al., 2012) as noted for other invasive termite populations (Perdereau et al., 2015) - specifically, their large expansive colonies intermingle with each other (Grace et al., 1989), have multiple secondary reproductives, and are resistant to pesticide control (Myles, 2004).

Previous genetic analysis suggests that the Toronto *R. flavipes* invasion was of single origin, one of a least three in the province of Ontario (Scaduto et al., 2012). This single founding event likely 'bottlenecked' allelic diversity within the population and may contribute to its open and diffuse colony structure that is characterized by the lack of intercolony aggression, as observed with other invasive social insects (Blight et al., 2012; Scaduto et al., 2012; Simkovic, 2016). In this study, I use RNA-seq analyses to explore additional transcriptomic correlates to this invasion by comparing the transcriptome of Toronto collected *R. flavipes* against that of two native populations from the United States. In doing so, I can identify genes or patterns of gene expression that are unique to *R. flavipes* from Toronto and thus potentially associated with their unique phenotypes and invasive history.

3.2 Methods

3.2.1 Termite sampling and RNA sequencing

For sequencing and *de novo* transcriptome assembly, I used the data described in Chapter 2, with the exception that I specified 'population' rather than 'caste' as the primary factor in my discriminant analysis. In brief, I used the previous three colony samples from one invasive population in Canada (Toronto) and from two native populations in the United States (Figure 3.1A). The Toronto population is distinctly urban, where colonies live in close association with human constructs (Figure 3.1B). By contrast, the native environments are characterized by semi-rural deciduous forests (Figure 3.1C). The number of populations is sampled in this study in lower compared to those in other invasive termite studies (Dedeine et al., 2015; Scharf, 2015). However, my objective for this chapter is rather qualitative and exploratory in nature, in an attempt to reveal new insights to gene expression changes associated with invasiveness. I froze the termites collected from all sites in liquid nitrogen, and extracted their total RNA following a Trizol-RNeasy protocol. I sent the RNA samples for library preparation and sequencing on two flow cell lanes of an Illumina HiSeq 2000 at the facilities of MGU-GQ. To discover gene expression patterns associated with population, I again used the TRINITY-based protocol (Haas et al., 2013), as previously outlined in Chapter 2.

3.2.2 Differential gene expression and enrichment analysis

To test for invasive Toronto-biased gene expression patterns, I again used my previous *de novo* assembly (n = 13,755 genes) and provisional annotations as a reference to compare gene expression between populations. I used BOWTIE (Langmead et al., 2009) and RSEM (Li & Dewey, 2011) to generate normalized mapped read counts for each gene for each of the nine libraries. I then used DESEQ2 (Love et al., 2014) to identify genes differentially expressed between populations. For this comparison, I created mixed-caste samples by grouping libraries by population regardless of caste – i.e., Boston nymph, soldier, and worker all grouped together as replicates for 'Boston only', and so forth. I also considered genes differentially expressed if they: i) displayed a minimum four-fold expression difference between any two population groups, and ii) maintained a false discovery rate (FDR) of less than 0.001. To see if libraries sort according to their source population regardless of caste, I correlated the average gene-wise expression of all differentially expressed genes (DEGs)

against each other in a Pearson sample correlation matrix. I also plotted a hierarchical clustered heatmap of all DEGs individually against the population replicates. For all differentially expressed gene sets, I further used GOSEQ to perform a GO enrichment analysis for biological process (Young et al., 2010). I used the 'SIMREL' feature (Pesquita et al., 2009) of REVIGO to reduce any redundancy based on semantic similarity of the GO terms, and plotted these terms with their cluster representatives on a scatterplot with REVIGO (Supek et al., 2011).

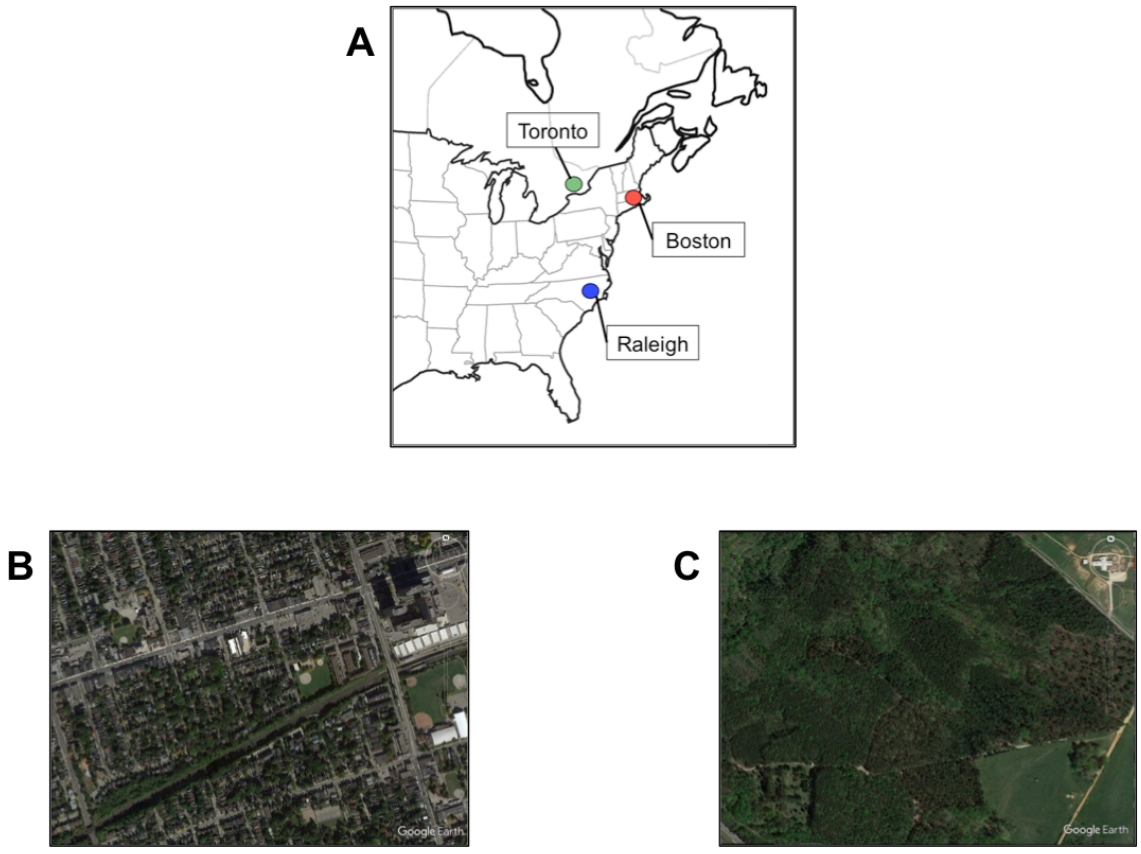


Figure 3.1 A) Map of Eastern North America showing the three termite sampling locations. **B)** Aerial photographs of a typical urban neighborhood in Toronto, and **C)** a native deciduous locale in Raleigh, where *R. flavipes* samples were collected. Images from Google Earth™.

3.3 Results

3.3.1 Gene expression patterns as a function of population

In general, I found n = 302 genes (of 13,755 genes; 2.2%) to be differentially expressed as a function of population regardless of my mixed-caste sampling (Table S3). The expression patterns of these genes support the invasive Toronto population being uniquely regulated compared to the two native populations, Boston and Raleigh (Figure 3.2). Moreover, hierarchical clustering of these population-biased genes revealed five tightly co-regulated gene sets (Figure 3.3) with the largest proportion (77.8%; 235 genes) as differences between invasive and native populations (Set II, III+V). To better facilitate the interpretation for genetic changes between invasive and native populations, I collapsed Gene Set III (141 genes) and V (18 genes) into one large native population biased gene set (n = 159 genes). From here, I refer to Gene Set II as the 'invasive gene set' and the Gene Set III + V as the 'native gene set'. A small proportion (22.2% of DEGs; 67 genes) accounted for within native population differences between the two sampling locales, Raleigh and Boston (I and IV). Overall, the entire set of DEGs is well represented by homolog annotations (~80%) in the Swiss-Prot database (Table S3). The proportions of unknown genes between the gene sets were not significant (Chi-squared statistic = 5.7366; $P > 0.05$).

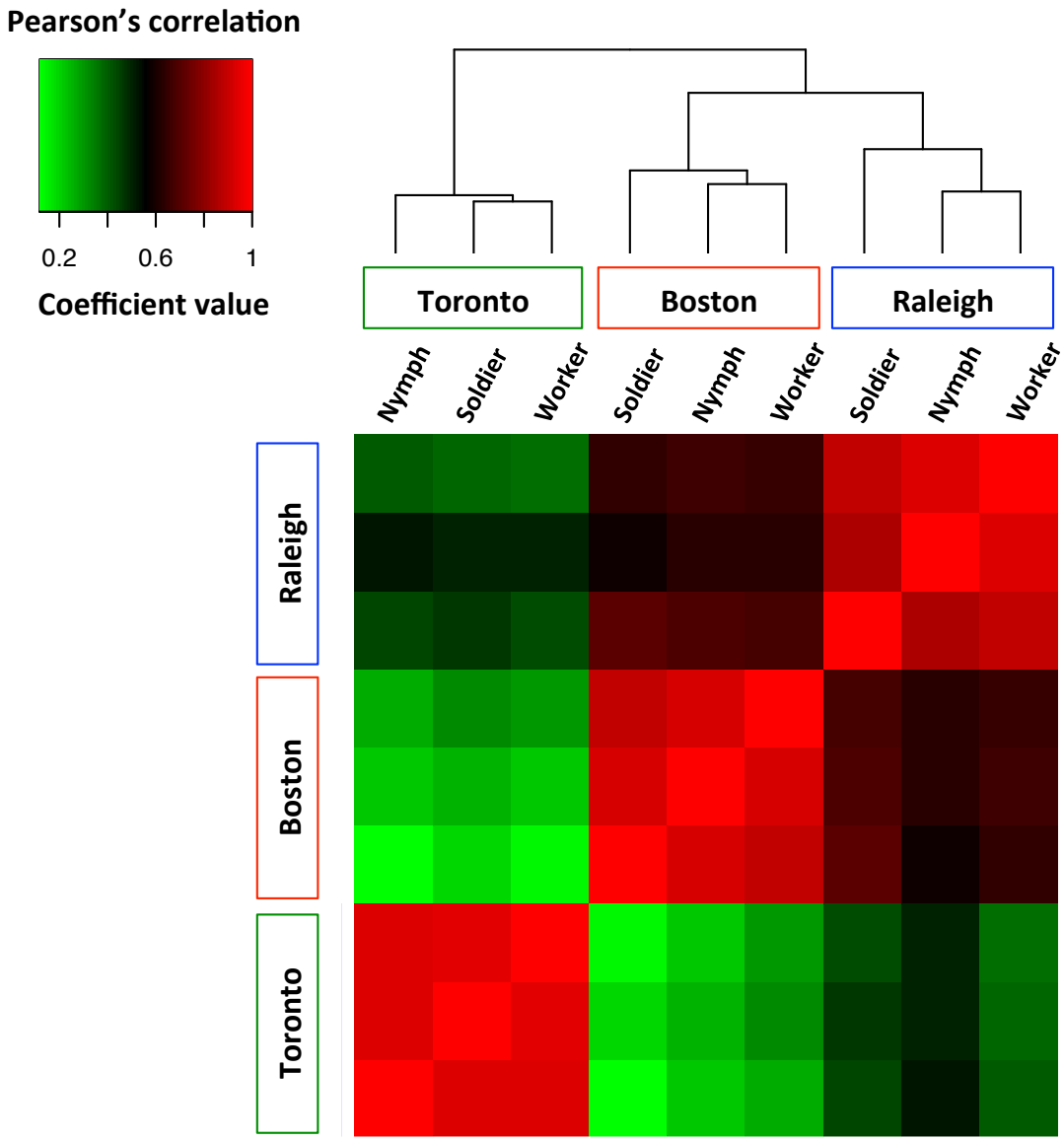


Figure 3.2 Library-by-library comparison of Pearson correlation coefficients based on average gene expression values of the 302 population-biased DEGs. Cells denote population and caste libraries that are positively (red) or negatively (green) correlated. All biological replicates sort according to population regardless of caste. Note: this pairwise diagram is symmetric along the red diagonal.

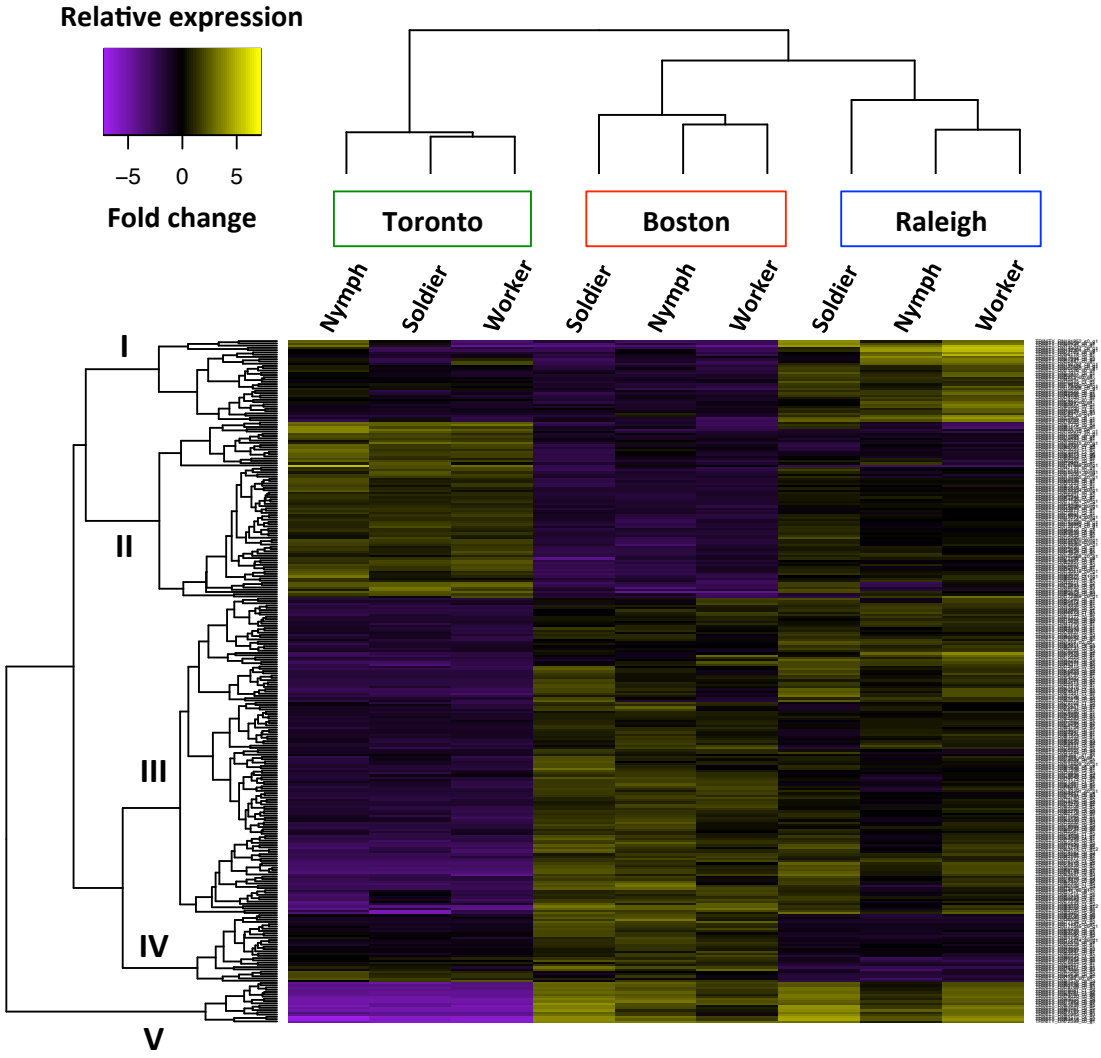


Figure 3.3 Heatmap of 302 DEGs (FDR < 0.001; minimum four-fold change, *see* Methods), clustered by gene sets (I - V) and populations of *R. flavipes*. Set I – 37 genes, Set II –76 genes, Set III – 141 genes, Set IV – 30 genes, and Set V – 18 genes. Gene identifiers are listed on the far right and can be read by zooming in on the graphic.

3.3.2 Functional analysis of invasive-native gene sets

Despite the invasive gene set having approximately half the number of up-regulated genes than the native gene set, invasive genes were enriched for more than three times the biological process (BP) terms and twice the number of cluster representatives (Figure 3.4). Specifically, three cluster representatives 'response to insecticide' (GO:0017085), 'nerve maturation' (GO:0021682), and 'glycosyl compound catabolism' (GO:1901658) accounted for more than half of the non-redundant BP-terms in the invasive gene set. In the native gene set, 'microtubule-based process' (GO:0007017) by itself accounted for more than half of the BP-terms with 'aggregation involved in sorocarp development' (GO:0031152) representing the majority of the remaining BP-terms. It is not unusual to observe an odd cluster representative like 'sorocarp (fruiting body of slime molds) development'. This GO term was generated from matching annotations in Swiss-Prot database, where the best annotation is from a phylogenetically distant taxon. It is obvious this specific cluster representative is not involved in termite biology, but rather its ontology derives from termite gene homologs. Lastly, there are no duplicates of BP-term cluster representatives between the two gene sets.

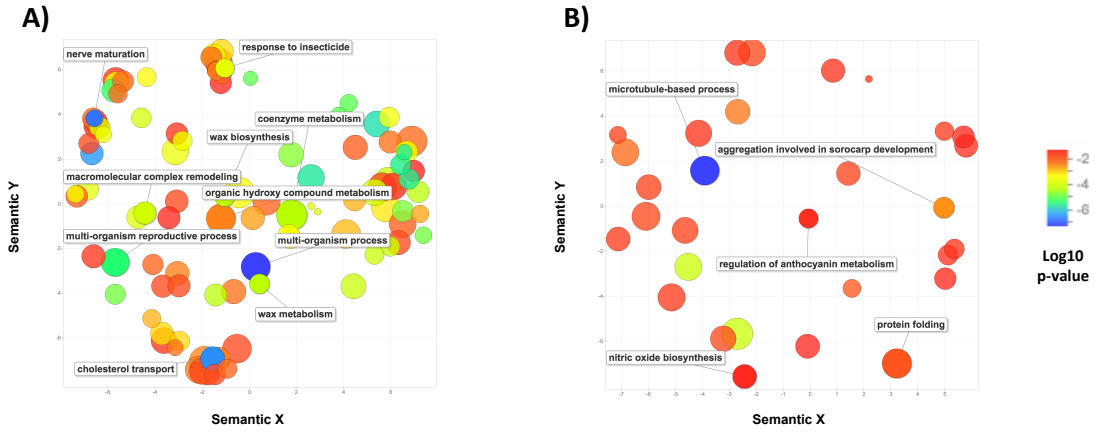


Figure 3.4 Scatterplots displaying outlined cluster representatives of GO terms in biological processes (BP) derived from *R. flavipes* genes as a function of invasive-native status in two dimensional space based on GO terms' semantic similarities. A) The invasive Gene Set II – 102 BP terms with 10 cluster representative, B) The native Gene Set III+V – 30 BP terms with 5 cluster representatives. The colour of bubbles indicates the GO term's $\log_{10}$ *P*-value. The size of bubbles represents the frequency of the GO term in the GO annotation database.

3.3.3 Functional analysis of native population differences

Between the two native population biased gene sets, the Raleigh up-regulated Gene Set I had more than twice the BP terms compared to the Boston up-regulated Gene Set IV and two additional cluster representatives (Figure 3.5). In the Raleigh up-regulated Gene Set I, majority of the BP-terms are represented by the three terms: 'glucose metabolism' (GO:0006040), 'mRNA cleavage involved in gene silencing by miRNA' (GO:0035279), and 'pupal chitin-based cuticle development' (GO:0008364). In the Boston up-regulated Gene Set IV, 'histidine biosynthesis' (GO:0000105) accounts for a large majority of BP-terms in the gene set. Again, it appears there are no duplicates of BP-term cluster representatives between the two gene sets.

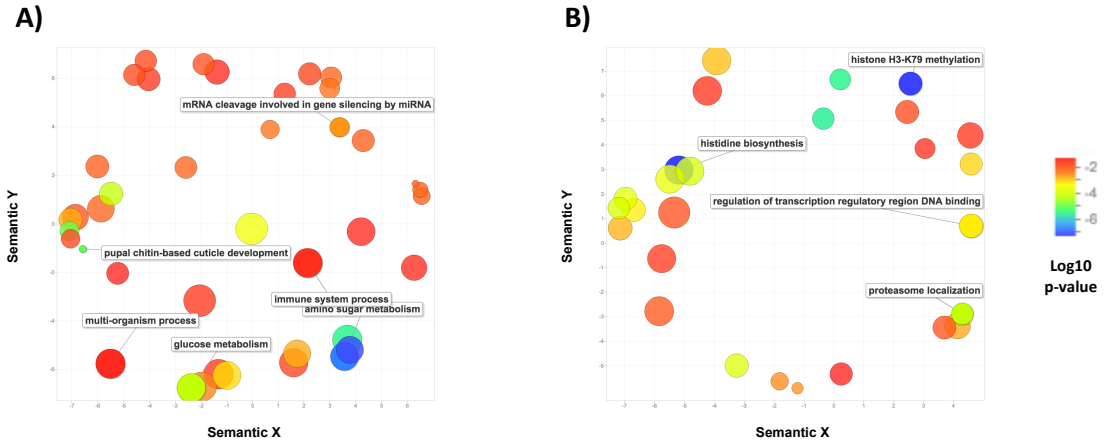


Figure 3.5 Scatterplots displaying outlined cluster representatives of GO terms in biological processes (BP) derived from *R. flavipes* native population specific genes in two dimensional space based on GO terms' semantic similarities. A) Raleigh-biased Gene Set I – 40 BP terms with 6 cluster representative, B) Boston-biased Gene Set IV – 27 BP terms with 4 cluster representatives. The colour of bubbles indicates the GO term's $\log_{10}$ *P*-value. The size of bubbles represents the frequency of the GO term in the GO annotation database.

3.4 Discussion

In this study, I explored the transcriptomic aspect of *R. flavipes* population associated with invasive-native status across three North American populations. The three populations showed distinct gene expression profiles, supported by the presence of five differentially regulated gene sets. My principal findings include, a large invasive-native bias of gene

expression in the form of three gene sets (II, III+V) - which represent approximately 78% of all DEGs - between invasive (Toronto) and native (Boston, Raleigh) populations. This overall expression pattern [T, (B, R)] suggests that the Toronto population is unique relative to Boston and Raleigh populations. Furthermore, the invasive Toronto population gene set is the second largest with 76 up-regulated genes, containing several potential genetic level responses related to their invasive phenotypes. The co-up-regulated Boston and Raleigh native gene set accounted for the largest proportion of total differentially expressed genes (159/302 DEGs). The co-regulation of these genes suggests there are common genetic mechanisms as a function of native status, which may alternatively be compromised in the invasive population. Lastly, my analysis revealed two small gene sets (I and IV, 37 and 30 genes each) that were native population specific to either Raleigh or Boston. Not relative to the invasive-native scenario, these smaller gene sets may otherwise be informative of different environmental conditions within the native range. Most importantly, the comparison and contrast of these transcriptional patterns revealed candidate genes and pathways that could explain the success of invasive *R. flavipes* establishment in Canada and additional population level adaptations.

3.4.1 Gene set differences between invasive-native status

Focusing on the invasive-native *R. flavipes* comparison, it appears the invasive and native gene sets are specialized for very different biological processes with minimal overlap. Within the invasive gene set, ‘nerve maturation’, ‘response to insecticide’, and many enriched enzyme metabolism related processes are genetic signatures of *R. flavipes* being exposed to neurotoxic pesticides that were used in an effort to eradicate *R. flavipes* in Toronto (Myles, 2004). In particular, this invasive gene set up-regulated genes for five enzymes in the cytochrome family with cytochrome P450 6a13 and 18 directly linked to the detoxification of synthetic insecticides (Chung et al., 2005; David et al., 2013). Indirectly, the up-regulation of these cytochrome genes could also affect the cuticular hydrocarbon (CHC) synthesis of invasive *R. flavipes*, as some cytochrome P450s genes are essential for the conversion of hydrocarbons (Reed et al., 1994). CHCs are major signaling molecules within termite colonies and a diverse CHC profile is essential to ensure kin recognition in several species of termites (Florane et al., 2004; Haverty et al., 1996; Kaib et al., 2004; Uva et al., 2004). A compromised or uniform CHC profile stemming from strong founder effect experienced by

invasive supercolonies may have lead to the lack of aggression or kin recognition. I found further support for this explanation as two possible fatty acyl-CoA reductase isoforms upstream in the CHC synthesis pathway (Chung & Carroll, 2015) were also found up-regulated in the invasive gene set. The CHC diversity explanation remains promising as a potential underlying genetic mechanism to invasiveness. Future studies should validate whether invasive colonies do indeed have lower degrees of CHC diversity compared to native colonies via techniques like gas chromatography.

As for the presence of multiple secondary neotenic reproductives in invasive supercolonies, I observed an enrichment of multi-organism reproductive process despite my mixed-sex and mixed-caste sampling. In particular, I found two homologs of protein yellow and gametogenetin-binding protein 2-like (*GGN2*) that are up-regulated in the invasive gene set. Protein yellow functions both as a pigment protein and a regulator of reproductive maturation involving major royal jelly protein in social insects (Drapeau et al, 2006). *GGN2*-like protein participates in spermatogenesis in the testis of humans (Zhao et al., 2005), though it is not known if its functions are conserved with termites. These two genes could be directly responsible for the invasive colonies' shift of reproductive strategy from king-queen reproductive monopoly to multiple secondary reproductives, or perhaps they are simply a resultant downstream effect.

In contrast to the invasive gene set, the native gene set appears to be enriched for housekeeping-related biological processes. There is an over abundance of potential isoforms or multi-copies of the same genes in comparison to the other gene sets. For example, tubulin alpha and beta chain appear several times in the gene list (Table S3). Though I am not ruling out the possibilities of mis-assemblies, the over representation of these genes in the gene set nonetheless suggest its ontological term 'microtubule-based processes' is an essential function for native *R. flavipes* populations. Besides being the structural components to microtubules, these tubulins varies greatly in their roles from mediating cell division to reproductive functions in the testis (Nielsen et al., 2010). The enrichment of 'nitric oxide biosynthesis' and 'regulation of anthocyanin metabolism' could be connected to native termites ingesting chemical compounds from a diet of common wood and plant materials in their native range (Brasseur et al., 2016; Ngugi & Brune, 2012). To further clarify the presence of 'aggregation involved in sorocarp development', this GO term derives from

genes such as heat shock proteins, which functions in acute stress response by 'protein folding' (Craig et al., 1993). Overall, it is evident from my transcriptomic comparisons that invasive supercolonies make genetic level tradeoffs to adapt to their new invasive environment compared to housekeeping functions in their native environment. However, whether this genetic tradeoff has any fitness consequences remains to be investigated.

3.4.2 Gene set differences within native populations

Lastly, my differential gene expression analysis also revealed two smaller gene sets that are biased towards either of the two native *R. flavipes* populations (Gene Set I-Raleigh, 37 genes; Gene Set IV-Boston, 30 genes). There appears to be minimal overlap in BP-GO terms between the two gene sets. However, the Gene Set I appears to have a much more diverse functional profile compared to Gene Set IV. The enrichment of 'immune system process', glucose and amino sugar metabolism in Gene Set I could all be linked to the differences in environmental conditions of the Southern native range such as parasite exposure (Ugelvig & Cremer, 2012) and diet (Shi et al., 2013). This is further supported by the over-representation of 'histidine biosynthesis', an essential amino acid, in the Boston biased Gene Set IV. These two population gene sets also have different gene regulatory mechanisms, where Gene Set I is enriched for 'mRNA cleavage involved in gene silencing by miRNA' and Gene Set IV is enriched for several processes including 'regulation of transcription regulatory region DNA binding'. All of these regulatory processes have been implicated in the caste differentiation process within termites (Scharf et al., 2003; Terrapon et al., 2014) and other social insects (Ashby et al., 2016; Li et al., 2010). The variation in the exact regulatory mechanisms could be attributed to different socioenvironmental factors from their respective locales. Nonetheless, these results suggest that termite sociality within native environments are plastic in some ways, and this plasticity may be another contributing factor to how well invasive *R. flavipes* have established in new environments such as Toronto.

3.4.3 Conclusion

To my knowledge, this is first study to investigate termite invasiveness using a comparative transcriptomics approach with multiple colonies from native and invasive populations. Results from this study identified candidate genes and biological processes that underlie the success of invasive Toronto *R. flavipes* supercolony along with transcriptomic information on

native colonies from its Northern and Southern range. Specifically, I found the invasive *R. flavipes* population to be enriched for several insecticide response and chemical metabolic processes suggesting a molecular level adaptation to the current control measures. Moreover, I found reproductive and chemical communication-related processes to be enriched in the invasive population, matching previously described supercolonial phenotypes of neotenic reproductive development and lack of kin recognition in a diffuse open colony structure. However, it is unclear whether these genetic changes in the Toronto population are the result of founder effect from being an invasive population, or it is rather these termites adapted on a molecular level to the novel conditions presented by a metropolitan environment. Future studies should focus on testing and comparing transcriptomic data from multiple invasive *R. flavipes* populations to see if they share common changes in gene regulation. Besides the invasive-native population genetic differences, I also found two other gene sets reflective of differences between populations within the native range. Overall, this study presents foundational transcriptomic knowledge to termite supercolonies, and future studies should focus on testing the identified gene candidates to develop novel control and management strategies.

3.5 References

Aber, A., & Fontes, L. R. (1993). *Reticulitermes lucifugus* (Isoptera, Rhintotermitidae) a pest of wooden structures, is introduced to the South American continent. *Sociobiology*, *21*, 335–339.

Ashby, R., Forêt, S., Searle, I., & Maleszka, R. (2016). MicroRNAs in honey bee caste determination. *Sci Rep*, *6*, 18794.

Becker, G. (1970). *Reticulitermes* in mittel und west Europa. *Z Angew Entomol*, *65*, 268–278.

Blight, O., Berville, L., Vogel, V., Hefetz, A., Renucci, M., Orgeas, J., et al. (2012). Variation in the level of aggression, chemical and genetic distance among three supercolonies of the Argentine ant in Europe. *Mol Ecol*, *21*, 4106–4121.

Bradshaw, C. J. A., Leroy, B., Bellard, C., Albert, C., Roiz, D., Barbet-Massin, M., et al. (2016). Massive yet grossly underestimated global costs of invasive insects. *Nat Commun*, *7*, 12986.

Brasseur, C., Bauwens, J., Tarayre, C., Millet, C., Mattéotti, C., Thonart, P., et al. (2016). GC×GC-TOFMS for the analysis of metabolites produced by termites (*Reticulitermes flavipes*) bred on different carbon sources. *Separations*, *3*, 19.

Buczkowski, G., & Bertelsmeier, C. (2017). Invasive termites in a changing climate: a global perspective. *Ecol Evol*, *3*, 974–985.

Chouvenc, T., Helmick, E. E., & Su, N.-Y. (2015). Hybridization of two major termite invaders as a consequence of human activity. *PLoS ONE*, *10*, e0120745.

Chung, H., & Carroll, S. B. (2015). Wax, sex and the origin of species: dual roles of insect cuticular hydrocarbons in adaptation and mating. *BioEssays*, *7*, 822–830.

Chung, H. S., Yu, T. H., Kim, B. J., Kim, S. M., Kim, J. Y., Yu, H. S., et al. (2005). Expressed sequence tags analysis of *Blattella germanica*. *Korean J Parasitol*, *43*, 149–156.

Craig, E. A., Gambill, B. D., & Nelson, R. J. (1993). Heat shock proteins: molecular chaperones of protein biogenesis. *Microbiol Rev*, *57*, 402–414.

David, J.-P., Ismail, H. M., Chandor-Proust, A., & Paine, M. J. I. (2013). Role of cytochrome

P450s in insecticide resistance: impact on the control of mosquito-borne diseases and use of insecticides on Earth. *Philos Trans R Soc Lond B Biol Sci*, *368*, 20120429.

Drapeau, M. D., Albert, S., Kucharski, R., Prusko, C., & Maleszka, R. (2006). Evolution of the Yellow / Major Royal Jelly Protein family and the emergence of social behavior in honey bees. *Genome Res*, *11*, 1385–1394.

Evans, T. A. (2011). Invasive termites. In D. E. Bignell, Y. Roisin, & N. Lo (Eds.), *Biology of termites: a modern synthesis* (pp. 519–562). Dordrecht: Springer.

Evans, T. A., Forschler, B. T., & Grace, J. K. (2013). Biology of invasive termites: a worldwide review. *Annu Rev Entomol*, *58*, 455–474.

Florane, C., Bland, J. M., Husseneder, C., & Raina, A. K. (2004). Diet-mediated inter-colonial aggression in the Formosan subterranean termite *Coptotermes formosanus*. *J Chem Ecol*, *30*, 2559–2574.

Ghesini, S., & Marini, M. (2009). Caste differentiation and growth of laboratory colonies of *Reticulitermes urbis* (Isoptera, Rhinotermitidae). *Insectes Soc*, *56*, 309–318.

Grace, J. K. (1996a). Temporal and spatial variation in caste proportions in a northern *Reticulitermes flavipes* colony (Isoptera: Rhinotermitidae). *Sociobiology*, *28*, 225–231.

Grace, J. K. (1996b). Absence of overt agonistic behavior in a northern population of *Reticulitermes flavipes*. (lsoptera: Rhinotermitidae). *Sociobiology*, *28*, 103–110.

Grace, J. K., Abdallay, A., & Farr, K. R. (1989). Eastern subterranean termite (Isoptera, Rhinotermitidae) foraging territories and populations in Toronto. *Can Entomol*, *121*, 551–556.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*, *8*, 1494–1512.

Haverty, M. I., Grace, K. J., Nelson, L. J., & Yamamoto, R. T. (1996). Intercaste, intercolony, and temporal variation in cuticular hydrocarbons of *Coptotermes formosanus* Shiraki (Isoptera: Rhinotermitidae). *J Chem Ecol*, *22*, 1813–1834.

Hellmann, J. J., Byers, J. E., Bierwagen, B. G., & Dukes, J. S. (2008). Five potential consequences of climate change for invasive species. *Conserv Biol*, *22*, 534–543.

Holmes, T. P., Aukema, J. E., Von Holle, B., Liebhold, A., & Sills, E. (2009). Economic impacts of invasive species in forests: Past, present, and future. *Ann NY Acad Sci*, *1162*, 18–38.

Kaib, M., Jmhasly, P., Wilfert, L., Durka, W., Franke, S., Francke, W., et al. (2004). Cuticular hydrocarbons and aggression in the termite *Macrotermes subhyalinus*. *J Chem Ecol*, *30*, 365–385.

Kenis, M., Auger-Rozenberg, M.-A., Roques, A., Timms, L., Péré, C., Cock, M. J. W., et al. (2009). Ecological effects of invasive alien insects. *Biol Invasions*, *11*, 21–45.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, *10*, R25.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323.

Li, J., Wu, J., Rundassa, D. B., Song, F., Zheng, A., & Fang, Y. (2010). Differential protein expression in honeybee (*Apis mellifera* L.) larvae: underlying caste differentiation. *PLoS ONE*, *5*, e13455.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550.

Lowe, S., Browne, M., Boudjelas, S., & De Poorter, M. (2000). 100 of the world's worst invasive alien species. A selection from the global invasive species database. *A Selection from the Global Perspective from the Global Invasive Species Database. ISSG, SSC, and IUCN.*, Available at www.issg.org/booklet.pdf.

McKern, J. A., Szalanski, A. L., & Austin, J. W. (2006). First record of *Reticulitermes flavipes* and *Reticulitermes hageni* in Oregon (Isoptera: Rhinotermitidae). *Fla Entomol*, *89*, 541–542.

Myles, T. G. (2004). Termite control in Canada. *Centre for Urban and Community Studies*, *23*, 1–4.

Nielsen, M. G., Gadagkar, S. R., & Gutzwiller, L. (2010). Tubulin evolution in insects: gene duplication and subfunctionalization provide specialized isoforms in a functionally constrained gene family. *BMC Evol Biol*, *10*, 113.

Ngugi, D. K., & Brune, A. (2012). Nitrate reduction, nitrous oxide formation, and anaerobic ammonia oxidation to nitrite in the gut of soil-feeding termites (*Cubitermes* and *Ophiotermes* spp.). *Environ Microbiol*, *14*, 860–871.

Pejchar, L., & Mooney, H. A. (2009). Invasive species, ecosystem services and human well-being. *Trends Eol Evol*, *24*, 497–504.

Perdereau, E., Bagnères, A.-G., Vargo, E. L., Baudouin, G., Xu, Y., Labadie, P., et al. (2015). Relationship between invasion success and colony breeding structure in a subterranean termite. *Mol Ecol*, *24*, 2125–2142.

Pesquita, C., Faria, D., Falcão, A. O., Lord, P., & Couto, F. M. (2009). Semantic similarity in biomedical ontologies. *PLoS Comput Biol*, *5*, e1000443.

Pyšek, P., & Richardson, D. M. (2010). Invasive species, environmental change and management, and health. *Annu Rev Environ Resour*, *35*, 25–55.

Raffoul, M., Hecnar, S. J., Prezioso, S., Hecnar, D. R., & Thompson, G. J. (2011). Trap response and genetic structure of Eastern subterranean termites (Isoptera: Rhinotermitidae) in Point Pelee National Park, Ontario, Canada. *Can Ent*, *143*, 263–271.

Reed, J. R., Vanderwel, D., Choi, S., Pomonis, J. G., Reitz, R. C., & Blomquist, G. J. (1994). Unusual mechanism of hydrocarbon formation in the housefly: cytochrome P450 converts aldehyde to the sex pheromone component (Z)-9-tricosene and CO2. *Proc Natl Acad Sci USA*, *91*, 10000–10004.

Scaduto, D. A., Garner, S. R., Leach, E. L., & Thompson, G. J. (2012). Genetic evidence for multiple invasions of the Eastern subterranean termite into Canada. *Environ Entomol*, *41*, 1680–1686.

Scharf, M. E., Wu-Scharf, D., Pittendrigh, B. R., & Bennett, G. W. (2003). Caste- and development-associated gene expression in a lower termite. *Genome Biol*, *4*, R62.

Shi, W., Xie, S., Chen, X., Sun, S., Zhou, X., Liu, L., et al. (2013). Comparative genomic analysis of the endosymbionts of herbivorous insects reveals eco-environmental adaptations: biotechnology applications. *PLoS Genet*, *9*, e1003131.

Simkovic, V. K. (2016). Testing for aggression and nestmate recognition in the eastern subterranean termite *(Reticulitermes flavipes)*. Electronic Thesis and Dissertation Repository. Retrieved from http://ir.lib.uwo.ca/etd/3789

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, *6*, e21800.

Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., et al. (2014). Molecular traces of alternative social organization in a termite genome. *Nat Commun*, *5*, 3636.

Thorne, B. L., & Traniello, J. F. A. (2003). Comparative social biology of basal taxa of ants and termites. *Annu Rev Entomol*, *48*, 283–306.

Ugelvig, L. V, & Cremer, S. (2012). Effects of social immunity and unicoloniality on host–parasite interactions in invasive insect societies. *Funct Ecol*, *26*, 1300–1312.

Urquhart, F. A. (1953). The introduction of the termite into Ontario. *Can Entomol*, *85*, 292–293.

Uva, P., Clément, J.-L., & Bagnères, A.-G. (2004). Colonial and geographic variations in agonistic behaviour, cuticular hydrocarbons and mtDNA of Italian populations of *Reticulitermes lucifugus* (Isoptera, Rhinotermitidae). *Insectes Soc*, *51*, 163–170.

Vargo, E. L., & Husseneder, C. (2011). Genetic structure of termite colonies and populations. In D. E. Bignell, Y. Roisin, & N. Lo (Eds.), *Biology of termites: a modern synthesis* (pp. 321–347). Dordrecht: Springer.

Vargo, E. L., Leniaud, L., Swoboda, L. E., Diamond, S. E., Weiser, M. D., Miller, D. M., & Bagnères, A.-G. (2013). Clinal variation in colony breeding structure and level of inbreeding in the subterranean termites *Reticulitermes flavipes* and *R. grassei*. *Mol Ecol*, *22*, 1447–1462.

Vieau, F. (2001). Comparison of the spatial distribution and reproductive cycle of *Reticulitermes santonensis* Feytaud and *Reticulitermes lucifugus grassei* Clément (Isoptera, Rhinotermitidae) suggests that they represent introduced and native species, respectively. *Insectes Soc*, *48*, 57–62.

Weesner, F. M. (1970). Termites of the Nearctic region. In Krishna, K. & Weesner, F. M. (Ed.), *Biology of termites, volume 2*. (pp. 477–525). New York: Academic.

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*, *11*, R14.

Zhao, Q., Zhou, Y., Cao, Z., Zhu, H., Huang, P., & Lu, B. (2005). Germ-cell specific protein gametogenetin protein 2 (*GGN2*), expression in the testis, and association with intracellular membrane. *Mol Reprod Dev*, *72*, 31–39.

Chapter 4

4 General Discussion

The Eastern subterranean termite, *Reticulitermes flavipes*, is an emerging model organism for molecular caste differentiation (Vargo & Husseneder, 2009), and a species of great economic concern around the world (Buczkowski & Bertelsmeier, 2017). The goal of this thesis was to use next-generation sequencing in a comparative transcriptomics approach to explore the genetic basis of termite caste differentiation and invasiveness phenotypes. In Chapter 2, I established a *de novo* reference transcriptome for the nymph, soldier, and worker castes of *R. flavipes* complete with gene annotations from Swiss-Prot. This new termite transcriptomic resource allowed me to identify three unique gene sets (total of 93 genes) in my comparative analyses with biased up-regulation for nymph/worker, nymphs only, and soldier only. Moreover, I found the transcriptional patterns of these three gene sets supported the relationship of soldiers being a novel caste in comparison to nymphs and worker [(nymph, worker), soldier]. In Chapter 3, I took advantage of the same reference transcriptome and identified five gene sets (302 genes) as functions of invasive-native status (Toronto vs. Boston and Raleigh), and Northern-Southern native populations. The expression patterns of these gene sets reflected a strong distinction between invasive and native populations [T, (B, R)] and further revealed many underlying genes and pathways potentially responsible for supercolony behaviours and potential environmental adaptations. In this final discussion chapter, I address some of the future outlooks, challenges, and limitations associated with my findings from these two chapters for a better understanding of termite social evolution.

4.1 Advancing the understanding of termite molecular caste differentiation

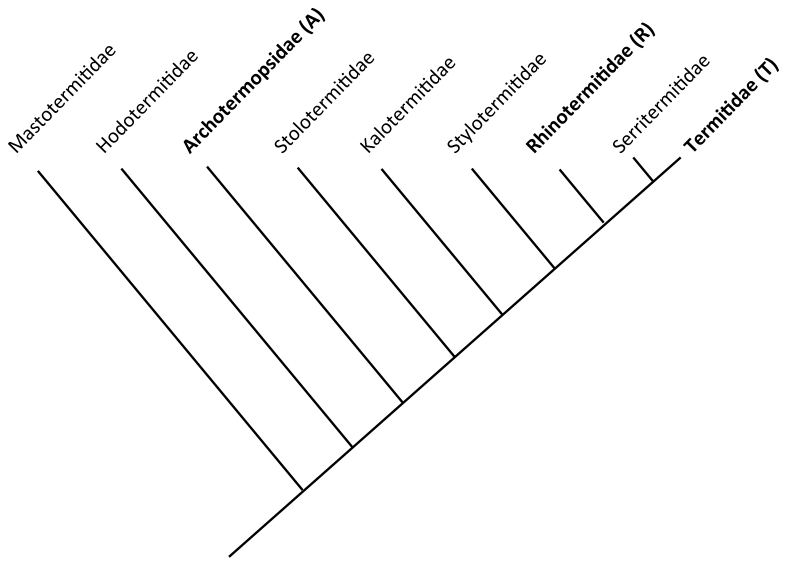
In terms of caste-biased gene expression, *R. flavipes* soldiers had the highest number of up-regulated genes and had the most unique gene expression profile with minimal overlap against nymphs and workers. This gene expression pattern matched the prediction from evolutionary history of termite castes that soldiers are relatively novel in comparison to nymphs and workers. To further validate this provisional result, future studies should focus on testing for genetic novelty associated with the identified caste-biased genes. Previous

transcriptomic and genomic studies on the social hymenopterans such as ants and bees already highlighted a trend of worker genes being more derived in origin or genetically novel compared to the queen genes (Feldmeyer et al., 2014; Harpur et al., 2014; Kapheim et al., 2016). If indeed the termite soldier caste is also more derived and more evolutionarily recent, then I expect soldier-biased genes, or a great proportion of these genes, to be more taxonomically restricted in terms of homology, and experience a more rapid rate of evolution compared to non-soldier-biased genes. For example, termite soldier-biased genes should be restricted within termites (Isoptera) versus nymph or worker genes that may match to a more phylogenetically distant or broadly inclusive taxa like all of Animalia or at least Arthropoda. It is very possible the evolution of termite soldier caste and hymenopterans workers converge on similar molecular mechanisms.

This thesis was limited to the comparisons of one reproductive (nymph) and two non-reproductive (soldier, worker) castes. To gain a more comprehensive understanding on termite molecular caste differentiation, future studies can also expand caste sampling by including other castes such as alates, primary reproductives (king, queen), and secondary neotenic reproductives (nymphoid, ergatoid). I expect the inclusion of these additional castes will show expression patterns that can further clarify the relationship between nymphs, soldiers, and workers, or reveal unexpected associations and evolutionary patterns. For example, if castes from the reproductive line share more co-regulated genes with nymphs and workers, then it would further support the notion of termite soldier evolution as a genetic novelty. One interesting aspect of this future study is the placement of nymphoid and ergatoid neotenic reproductives in terms of their gene expression profiles. Will these two castes group closer together with their pre-cursor castes (i.e. nymphs and workers), with primary reproductives, or with each other? Under the predictions outlined by winglessness theory, where secondary reproduction and polyphenism evolved prior to reproductive skew and workers (Bourguignon et al., 2016), the neotenic reproductives should group together with primary reproductives despite the large developmental difference. Previous comparative transcriptome analyses of these castes in *Z. nevadensis* (Archotermopsidae, non-bifurcate development) show support to the outlined predictions (Terrapon et al., 2014). Therefore, it would be interesting to see if *R. flavipes* (Rhinotermitidae, bifurcate development) also uphold the same regulatory pattern.

This annotated *R. flavipes* transcriptome greatly advances the field of termite sociogenomics, and social insect genomics in general. Sequencing reads from this *R. flavipes* reference transcriptome containing the nymph, soldier, and worker castes are deposited in NCBI Sequence Read Archive (NCBI-SRA) and are freely accessible to the wider community. This data should facilitate future comparative transcriptome analysis against other social or non-social insects. For example, results from two recently published *Z. nevadensis* and *M. natalensis* genomes revealed large genetic differences between these two termites of with contrasting social complexities and developmental programs (Korb et al., 2015) (Figure 4.1). Comparative insights from the *R. flavipes* transcriptome can shed light on how the transition in social complexities occurred between these species because *R. flavipes* belongs to the sister family of *M. natalensis*, and displays intermediary social traits (strictly wood diet, single and multi piece wood nesting, and facultative sterility) (Evans et al., 2013). On a broader scale for testing molecular conservation or convergent evolution of termite social traits, this transcriptomic data could also be used to compare the gene expression profiles with reproductive division of labour in other social insects like ants and bees (Korb, 2016; Oppenheim, et al., 2015). Through these multi-taxa comparisons, we should gain new perspectives to how termite social evolution occurred and whether reproductive division of labour and polyphenism are a product of conserved genes or genes with novel social functions.

A)



B)

	Z. nevadensis (A)	*R. flavipes* (R)	*M. natalensis* (T)
Diet	Wood	Wood	Wood, Grass, Plant litter
Nesting	Single piece	Single, intermediate, and separate piece	Separate piece
Sterility	Facultative	Facultative	Obligatory

Figure 4.1 A) Simplified phylogeny of Isoptera displaying all termite families. **B)** Contrasts between diet, nesting habit, and sterility characteristics of termites from three families. Note: the topology of the phylogeny does not reflect evolutionary time.

4.2 Developing novel and target specific termite control strategies

As previously mentioned, the frequencies of biological invasions are expected to significantly increase in combination with the effects of climate change (Hellmann et al., 2008). Climate modeling of *R. flavipes* and other invasive termite species' ranges, predicted all but one species to rapidly expand their current global distribution (Buczkowski & Bertelsmeier, 2017). It is critical we understand the fundamental mechanisms of termite invasion success and their ability to adapt to new conditions outside of their native range.

My transcriptome analysis of *R. flavipes* provides fresh insight to termite invasion biology and environmental adaptations in two significant ways. First, I successfully sequenced three *R. flavipes* castes (nymph, soldier, worker) from one invasive population in Toronto and two native populations in Boston and Raleigh. The reference transcriptome assembly from the

previous caste analysis now serves as a major resource for *R. flavipes* gene annotations. To my knowledge, this is the first study to compare invasive and native termites from a transcriptomic aspect to try and understand the genetic basis of invasive phenotypes such as multiple reproductives and lack of kin recognition. Second, from my differential gene expression analysis, I found 302 genes differentially expressed as a function of populations, with 235 genes with invasive-native status biases, and 67 genes with Northern-Southern range biases. When I subjected these genes to an enrichment analysis, I found several key molecular pathways such as insecticide response, multi-organism reproduction, and fatty acid and CHC metabolism that may be linked to the invasive success of the Toronto supercolony. I also identified different diet, immune, and gene regulatory processes within native populations. I expected to see some level of genetic responses to climate conditions corresponding to Northern-Southern range, but observed none. Perhaps, my termite sampling time and the absence of environmental extremes had a large part in not being able to elicit the genetic changes in thermoregulatory responses. In the future as more termite genes become characterized, maybe some of these climate adaptive molecular processes will become more apparent in my data.

More often than not, transcriptomic studies on invasive termites focus more on termites from their native range (Scharf, 2015a). Transcriptomic sequencing studies done on termites from their invasive range are rare. There was a recent RNA-Seq study of invasive *R. flavipes* populations and two native congeneric species from France describing the transcriptomic profiles of secondary reproductives (Dedeine et al., 2015). I think it would be beneficial for future comparative studies to include various invasive populations and castes such as those in France to deduce common genetic mechanisms for termite invasion and adaptation success.

Candidate genes identified in this study are ideal targets for the development of novel pest control strategies. First, it appears the lack of conventional termiticide success in metropolitan Toronto is directly related to adaptive responses observed in the gene expression profile, such as insecticide response and the specific up-regulation of several cytochrome genes (David et al., 2013). Second, gene-level technology has been shown as a compelling concept to address some of these shortcomings of conventional toxic termiticides (Scharf, 2015b). For example, RNA interference (RNAi) is gaining strong interests from the control application perspective through its functional genomic results (Huvenne & Smagghe,

2010). RNAi has successfully knock down over 15 different genes in five species of termites, including *R. flavipes*, and can be delivered through feeding (Scharf, 2015b). Specifically, candidate genes identified from Chapter 3 can be made into dsRNA or siRNA solutions, which will be taken up and spread throughout the colony via trophallaxis. Currently, the stability of these RNA molecules needs to be improved to maximize its effectiveness in disrupting target genes, but nonetheless remains a promising new alternative (Huvenne & Smagghe, 2010). Besides RNAi, alternative control strategies such as caste or pheromone disruption can have profound effects as well. Previous laboratory studies already demonstrated juvenile hormone and soldier head extracts to have termiticidal potential in over-promoting sterile soldier development in a colony (Tarver et al., 2009). One hypothetical example is to design chemicals that simulate existing pheromones or interrupt signaling pathways for pheromone production such as acetyl-CoA and CHC catalysis identified in the invasive population. The results of these disruptions could collapse the existing termite supercolony structure through disproportionate caste composition, premature caste differentiation, or potential loss of the reproductive division of labour.

4.3 Technical challenges with mRNA-Seq

Next generation sequencing, such as RNA-Seq, has made significant advances in sociogenomic research by allowing for a comprehensive characterization and quantification of the transcriptome for various phenotypes. Specifically, we are able to quantify gene expression of transcripts with nucleotide level resolutions. This technology is a tremendous improvement compared to previous array-based technologies with limitations like probe selection biases, prerequisite sequence knowledge, and saturation induced by dynamic range limits (Wang et al., 2009). However, RNA-seq does have its own technological and experimental limitations that need to overcome (Oppenheim et al., 2015). I will address some of these problems and propose solutions below.

Reticulitermes flavipes is not a model organism and does not have its genome sequenced and annotated. Therefore, I took a *de novo* approach to assemble the transcriptome for three *R. flavipes* castes from three populations. More often than not, *de novo* assembled transcriptomes be riddled with mis-assemblies, fragmented transcripts, or difficulties with the majority of genes being unannotated (Martin & Wang, 2011). Understanding these

challenges, I was cautious in filtering gene content and using stringent parameters for differential gene expression analysis. The application of these techniques likely contributed to the small number of DEGs identified as a function of caste or invasiveness compared to other studies. Alternatively, the whole body RNA extraction method may have underestimated differences in gene expression between samples, as expression levels can vary greatly between tissues within a single individual (i.e. fat bodies vs. muscle) (Johnson et al., 2013). Regardless of these technical differences, I was still able interpret these gene sets in a biologically meaningful way. To gain a comprehensive understanding of the overall transcriptome and its regulation, future studies will need to focus on sequencing the genome, providing in depth curated annotations, and performing tissue specific RNA-seq for *R. flavipes* and other termite species. Projects like i5K (http://i5k.github.io/) and 1KITE (http://www.1kite.org/) already have plans to sequence and annotate the genome of over 20 termite species including *R. flavipes*, in the near future. Incorporating new genomic and tissue-specific transcriptomic information will clarify the portion of genes eliminated from my analyses, and discover other genes that play a role in caste differentiation and invasiveness.

Although I found promising gene candidates and molecular pathways that are potentially involved in both caste differentiation and invasiveness, further work needs to be done to confirm whether the transcript abundance reflects protein levels and *in vivo* functions in *R. flavipes*. Changes in transcript levels do not always correspond with the same changes in protein level (Vogel & Marcotte, 2012). Therefore, there are several approaches to validate these candidate genes for their gene functions and their roles. From a broad genomic level, proteomics through mass spectrophotometry of various castes and population samples can confirm whether the observed transcripts were actually translated into proteins (Wolschin & Amdam, 2007). On an *in vivo* level, functional experiments using RNAi allow us to target specific gene candidates in *R. flavipes* for knock-down experiments to observe subsequent changes in caste phenotypes (Zhou et al., 2008). For invasive phenotypes such as poor kin recognition, verifying the CHC profile with gas chromatography will confirm whether all invasive individuals share similar recognition cues (Haverty et al., 1996; Howard et al., 1977). Overall, I believe a combination of these experimental approaches will help us

reconnect genes with behavioural and physiological phenotypes for a comprehensive understanding of termite sociality.

4.4 Conclusion

This thesis was the first study to examine transcriptomic correlates for both caste differentiation and invasiveness for the Eastern subterranean termite, *R. flavipes*. Despite the species not having a genome assembly, I established a *de novo* reference transcriptome that contains gene expression information for three castes (nymph, soldier, worker) from three populations (Toronto, Boston, Raleigh). Differential gene expression and functional enrichment analysis revealed two datasets (93 and 302 genes) and several biological processes as functions of caste and invasiveness. Together with future advances in genomic and bioinformatic knowledge, studies can expand on the genes candidates identified in this thesis for a more complete understanding of termite social evolution from genes to phenotypes.

4.5 References

Bourguignon, T., Chisholm, R. A., & Evans, T. A. (2016). The termite worker phenotype evolved as a dispersal strategy for fertile wingless individuals before eusociality. *Am Nat*, *187*, 372–387.

Buczkowski, G., & Bertelsmeier, C. (2017). Invasive termites in a changing climate: a global perspective. *Ecol Evol*, *3*, 974–985.

David, J.-P., Ismail, H. M., Chandor-Proust, A., & Paine, M. J. I. (2013). Role of cytochrome P450s in insecticide resistance: impact on the control of mosquito-borne diseases and use of insecticides on Earth. *Philos Trans R Soc Lond B Biol Sci*, *368*, 20120429.

Dedeine, F., Weinert, L. A., Bigot, D., Josse, T., Ballenghien, M., Cahais, V., et al. (2015). Comparative analysis of transcriptomes from secondary reproductives of three *Reticulitermes* termite species. *PLoS ONE*, *10*, e0145596.

Evans, T. A., Forschler, B. T., & Grace, J. K. (2013). Biology of invasive termites: a worldwide review. *Annu Rev Entomol*, *58*, 455–474.

Feldmeyer, B., Elsner, D., & Foitzik, S. (2014). Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Mol Ecol*, *23*, 151–161.

Harpur, B. A., Kent, C. F., Molodtsova, D., Lebon, J., Alqarni, A. S., Owayss, A. A., & Zayed, A. (2014). Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci USA*, *111*, 2614–2619.

Haverty, M. I., Grace, K. J., Nelson, L. J., & Yamamoto, R. T. (1996). Intercaste, intercolony, and temporal variation in cuticular hydrocarbons of *Coptotermes formosanus* Shiraki (Isoptera: Rhinotermitidae). *J Chem Ecol*, *22*, 1813–1834.

Hellmann, J. J., Byers, J. E., Bierwagen, B. G., & Dukes, J. S. (2008). Five potential consequences of climate change for invasive species. *Conserv Biol*, *22*, 534–543.

Howard, R. W., McDaniel, C. A., & Blomquist, G. J. (1977). Cuticular hydrocarbons of the eastern subterranean termite, *Reticulitermes flavipes* (Kollar) (Isoptera: Rhinotermitidae). *J Chem Ecol*, *4*, 233–245.

Huvenne, H., & Smagghe, G. (2010). Mechanisms of dsRNA uptake in insects and potential of RNAi for pest control: a review. *J Insect Physiol*, *56*, 227–235.

Johnson, B. R., Atallah, J., & Plachetzki, D. C. (2013). The importance of tissue specificity for RNA-seq: highlighting the errors of composite structure extractions. *BMC*

Genomics, *14*, 586.

Kapheim, K. M. (2016). Genomic sources of phenotypic novelty in the evolution of eusociality in insects. *Curr Opin Insect Sci*, *13*, 24–32.

Korb, J. (2016). Genes underlying reproductive division of labor in termites, with comparisons to social hymenoptera. *Front Ecol Evol*, *4*, 45.

Korb, J., Poulsen, M., Hu, H., Li, C., Boomsma, J. J., Zhang, G., & Liebig, J. (2015). A genomic comparison of two termites with different social complexity. *Front Genet*, *6*, 9.

Martin, J. A., & Wang, Z. (2011). Next-generation transcriptome assembly. *Nat Rev Genet*, *12*, 671–682.

Oppenheim, S. J., Baker, R. H., Simon, S., & DeSalle, R. (2015). We can't all be supermodels: the value of comparative transcriptomics to the study of non-model insects. *Insect Mol Biol*, *24*, 139–154.

Scharf, M. E. (2015a). Omic research in termites: an overview and a roadmap. *Front Genet*, *6*, 76.

Scharf, M. E. (2015b). Termites as targets and models for biotechnology. *Annu Rev Entomol*, *60*.

Tarver, M. R., Schmelz, E. A., Rocca, J. R., & Scharf, M. E. (2009). Effects of soldier-derived terpenes on soldier caste differentiation in the termite *Reticulitermes flavipes*. *J Chem Ecol*, *35*, 256–264.

Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., et al. (2014). Molecular traces of alternative social organization in a termite genome. *Nat Commun*, *5*, 3636.

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

Vogel, C., & Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*, *13*, 227–232.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, *10*, 57–63.

Wolschin, F., & Amdam, G. V. (2007). Comparative proteomics reveal characteristics of life-history transitions in a social insect. *Proteome Sci*, *5*, 10.

Zhou, X., Wheeler, M. M., Oi, F. M., & Scharf, M. E. (2008). RNA interference in the termite *Reticulitermes flavipes* through ingestion of double-stranded RNA. *Insect Biochem Mol Biol*, *38*, 805–815.

Supplemental Materials

Table S1. Swiss-Prot annotations of *R. flavipes* caste-biased genes.

Cluster	Trinity Gene ID	Length (bp)	Gene Annotation
I-Soldier	TRINITY_DN12772_c0_g1	1080	Protein giant-lens
	TRINITY_DN29811_c0_g1	584	Ankyrin-3
	TRINITY_DN44020_c0_g1	3132	Cytochrome P450 6j1
	TRINITY_DN44279_c0_g1	1534	Adenylate kinase isoenzyme 1
	TRINITY_DN56937_c0_g1	1394	SET and MYND domain-containing protein 4
	TRINITY_DN62625_c0_g1	1904	Cytosolic carboxypeptidase 1
	TRINITY_DN62834_c0_g1	2552	Growth arrest-specific protein 2
	TRINITY_DN69320_c0_g2	775	Troponin C, isoform 1
	TRINITY_DN70903_c0_g1	2642	Junctophilin-1
	TRINITY_DN75717_c2_g1	767	Troponin C, isoform 2
	TRINITY_DN77710_c0_g1	2449	Arginine kinase
	TRINITY_DN78786_c1_g1	1218	Protein msta, isoform B
	TRINITY_DN79460_c3_g1	2594	Alpha-actinin, sarcomeric
	TRINITY_DN79959_c4_g1	1514	Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type
	TRINITY_DN79994_c2_g2	1685	Elongation of very long chain fatty acids protein 7
	TRINITY_DN80657_c2_g1	974	Muscle-specific protein 20
	TRINITY_DN80831_c0_g1	1284	Gastric triacylglycerol lipase
	TRINITY_DN81370_c5_g1	368	N-acetylmuramoyl-L-alanine amidase
	TRINITY_DN81478_c1_g6	2317	Ras-related and estrogen-regulated growth inhibitor
	TRINITY_DN81811_c6_g1	2035	Paramyosin, long form
	TRINITY_DN82275_c5_g2	3411	Muscle LIM protein Mlp84B
	TRINITY_DN82279_c7_g5	614	Troponin I
	TRINITY_DN83037_c7_g1	4768	Sarcalumenin
	TRINITY_DN83341_c1_g1	3241	Protein disulfide-isomerase A5
	TRINITY_DN83546_c1_g6	1291	Protein anoxia up-regulated
	TRINITY_DN83701_c0_g1	7430	Twitchin
	TRINITY_DN98549_c0_g1	343	Acyl-CoA Delta(11) desaturase
	TRINITY_DN174858_c0_g1	536	Tubulointerstitial nephritis antigen-like
	TRINITY_DN20304_c0_g1	665	Lysosomal alpha-mannosidase*
	TRINITY_DN177615_c0_g1	397	Lysosomal alpha-mannosidase*
	TRINITY_DN55879_c0_g2	842	Nose resistant to fluoxetine protein 6*
	TRINITY_DN80641_c4_g3	1210	Nose resistant to fluoxetine protein 6*
	TRINITY_DN69617_c0_g1	2168	Protein msta, isoform A*
	TRINITY_DN77195_c0_g1	2330	Protein msta, isoform A*
	TRINITY_DN78360_c5_g2	2481	Protein msta, isoform A*
	TRINITY_DN71080_c0_g1	1594	Muscle M-line assembly protein unc-89*
	TRINITY_DN83345_c1_g1	13232	Muscle M-line assembly protein unc-89*
	TRINITY_DN74909_c0_g2	914	Putative fatty acyl-CoA reductase CG5065*
	TRINITY_DN82942_c2_g2	1470	Putative fatty acyl-CoA reductase CG5065*
	TRINITY_DN82278_c0_g1	3124	Titin*
	TRINITY_DN82571_c1_g1	2515	Titin*
	TRINITY_DN83307_c0_g3	990	Titin*
	TRINITY_DN83857_c1_g24	1242	ATP-dependent 6-phosphofructokinase**
II-Nymph	TRINITY_DN62427_c0_g1	1590	Neither inactivation nor afterpotential protein C
	TRINITY_DN66987_c1_g2	1236	Aromatic-L-amino-acid decarboxylase
	TRINITY_DN67705_c0_g1	1953	Serine proteinase stubble
	TRINITY_DN75166_c0_g1	2763	P protein
	TRINITY_DN75750_c0_g1	2991	G2/mitotic-specific cyclin-B
	TRINITY_DN77896_c0_g1	3286	Protein aubergine
	TRINITY_DN78265_c0_g1	674	Histone H1.3
	TRINITY_DN78435_c0_g1	910	Dehydrogenase/reductase SDR family member 11
	TRINITY_DN78537_c0_g1	3360	Aurora kinase A
	TRINITY_DN78663_c0_g1	2259	A disintegrin and metalloproteinase with thrombospondin motifs 17
	TRINITY_DN79180_c4_g1	3665	Allergen Cr-PI
	TRINITY_DN80624_c1_g4	1282	Ecdysteroid UDP-glucosyltransferase
	TRINITY_DN81899_c4_g2	2610	Peroxidasin homolog
	TRINITY_DN82142_c0_g2	5971	Vitellogenin receptor
	TRINITY_DN83857_c1_g9	1238	ATP-dependent 6-phosphofructokinase**

	TRINITY_DN764_c0_g1	1109	Pro-resilin
	TRINITY_DN61229_c0_g1	868	General odorant-binding protein 56d
	TRINITY_DN77183_c0_g1	2674	PI-PLC X domain-containing protein 1
III-	TRINITY_DN81204_c1_g1	1462	Hexamerin
Nymph/Worker	TRINITY_DN82301_c1_g3	616	Endoglucanase
	TRINITY_DN82714_c1_g1	926	Myrosinase 1
	TRINITY_DN83277_c8_g1	1755	Beta-1,3-glucan-binding protein*
	TRINITY_DN83427_c1_g1	1327	Beta-1,3-glucan-binding protein*

* Duplicate in another gene set

** Duplicate within the same gene set

Table S2. *Reticulitermes flavipes* caste-biased genes without Swiss-Prot annotations.

Gene Set	Trinity Gene ID	Length (bp)
I- Soldier	TRINITY_DN19741_c0_g1	426
	TRINITY_DN65201_c0_g1	1070
	TRINITY_DN66234_c0_g2	756
	TRINITY_DN69767_c0_g1	740
	TRINITY_DN70271_c0_g1	1521
	TRINITY_DN75772_c0_g1	1713
	TRINITY_DN77755_c0_g1	5029
	TRINITY_DN78418_c2_g1	1420
	TRINITY_DN81585_c2_g1	1294
	TRINITY_DN82731_c2_g1	926
	TRINITY_DN83123_c0_g1	6787
	TRINITY_DN83131_c0_g1	1577
	TRINITY_DN83140_c0_g1	2123
	TRINITY_DN83245_c2_g2	1274
	TRINITY_DN86760_c0_g1	737
	TRINITY_DN113175_c0_g1	304
	TRINITY_DN176468_c0_g1	1101
	TRINITY_DN194034_c0_g1	317
II-Nymph	TRINITY_DN76318_c0_g1	2172
	TRINITY_DN77339_c0_g1	2542
	TRINITY_DN81628_c2_g2	353
	TRINITY_DN82301_c0_g1	1022
III-Nymph/Worker	TRINITY_DN27215_c0_g1	761
	TRINITY_DN72364_c0_g1	488
	TRINITY_DN73718_c0_g1	1851
	TRINITY_DN74646_c0_g1	2502
	TRINITY_DN80216_c0_g1	2806

Table S3. Swiss-Prot annotations of *R. flavipes* population-biased genes.

Cluster	Trinity Gene ID	Length (bp)	Gene Annotation
	TRINITY_DN563_c0_g1_i1	873	C-type lectin mannose-binding isoform
	TRINITY_DN13702_c0_g1_i1	460	Urokinase-type plasminogen activator
	TRINITY_DN17659_c0_g1_i1	758	Epoxide hydrolase 3
	TRINITY_DN45947_c0_g2_i1	1624	Putative helicase MOV-10
	TRINITY_DN52905_c0_g1_i1	926	Putative defense protein 3
	TRINITY_DN60659_c0_g1_i1	892	Cuticle protein 7
	TRINITY_DN67994_c0_g3_i1	4168	Hormone receptor 4
	TRINITY_DN70343_c1_g1_i1	1564	L-lactate dehydrogenase
	TRINITY_DN71526_c0_g1_i1	3228	Chaoptin
	TRINITY_DN74100_c1_g2_i1	3813	Transmembrane cell adhesion receptor mua-3
	TRINITY_DN78022_c0_g1_i3	1594	Mediator of RNA polymerase II transcription subunit 27
	TRINITY_DN78272_c0_g1_i2	1163	Peritrophin-1
	TRINITY_DN79063_c0_g1_i1	651	Non-specific lipid-transfer protein
	TRINITY_DN79812_c0_g1_i9	1302	Apolipoprotein D
	TRINITY_DN80560_c0_g1_i2	2241	Endocuticle structural glycoprotein SgAbd-2
	TRINITY_DN98114_c0_g1_i1	508	Ejaculatory bulb-specific protein 3
	TRINITY_DN131626_c0_g1_i1	697	Multidrug resistance protein 1B
	TRINITY_DN149047_c0_g1_i1	303	Endocuticle structural glycoprotein SgAbd-8
I-Raleigh	TRINITY_DN191852_c0_g1_i1	783	Adult-specific cuticular protein ACP-20
	TRINITY_DN198787_c0_g1_i1	636	Chymotrypsin BI
	TRINITY_DN202050_c0_g1_i1	618	Acidic mammalian chitinase
	TRINITY_DN48495_c0_g3_i1	862	Endochitinase*
	TRINITY_DN52447_c0_g1_i1	747	Hormone receptor 4*
	TRINITY_DN51590_c1_g1_i1	997	Glucose dehydrogenase [FAD, quinone]**
	TRINITY_DN65977_c0_g1_i2	803	Glucose dehydrogenase [FAD, quinone]**
	TRINITY_DN2275_c0_g1_i1	952	
	TRINITY_DN5339_c0_g1_i1	433	
	TRINITY_DN45745_c0_g1_i1	575	
	TRINITY_DN57917_c0_g1_i1	1612	
	TRINITY_DN63202_c0_g1_i3	1547	
	TRINITY_DN77165_c0_g1_i2	383	
	TRINITY_DN78347_c0_g3_i1	921	
	TRINITY_DN80222_c0_g1_i1	2698	
	TRINITY_DN82969_c0_g1_i1	1144	
	TRINITY_DN132304_c0_g1_i1	355	
	TRINITY_DN156068_c0_g1_i1	516	
	TRINITY_DN179830_c0_g1_i1	330	
	TRINITY_DN925_c0_g2_i1	757	Retrovirus-related Pol polyprotein from transposon 412
	TRINITY_DN32296_c0_g1_i1	670	Scavenger receptor class B member 1
	TRINITY_DN33652_c0_g1_i1	498	Leucine-rich repeats and immunoglobulin-like domains protein 2
	TRINITY_DN34571_c0_g1_i1	506	Tissue alpha-L-fucosidase
	TRINITY_DN34658_c0_g1_i1	345	Radical S-adenosyl methionine domain-containing protein 2
	TRINITY_DN36401_c0_g2_i1	3009	TP-binding cassette sub-family G member 1
	TRINITY_DN42672_c0_g1_i1	571	Fumarylacetoacetase
	TRINITY_DN44651_c0_g2_i1	1203	Glucose dehydrogenase
	TRINITY_DN47133_c0_g1_i1	1075	Voltage-gated potassium channel subunit beta-2
	TRINITY_DN53629_c0_g1_i1	720	ATP-binding cassette sub-family G member 4
	TRINITY_DN54435_c0_g1_i1	1685	Neuroglian
	TRINITY_DN58290_c0_g1_i2	1647	Protein N-terminal asparagine amidohydrolase
	TRINITY_DN62320_c0_g2_i1	770	GPI ethanolamine phosphate transferase 1
	TRINITY_DN62955_c0_g1_i2	632	Cytochrome b5-related protein
II-Invasive	TRINITY_DN64408_c0_g2_i1	2656	FYVE, RhoGEF and PH domain-containing protein 4
(Toronto)	TRINITY_DN65384_c0_g1_i2	1991	Putative tricarboxylate transport protein, mitochondrial
	TRINITY_DN66539_c0_g1_i1	3234	Apolipophorins
	TRINITY_DN67874_c0_g1_i2	1108	MFS-type transporter SLC18B1
	TRINITY_DN67948_c1_g1_i1	1101	Dystroglycan
	TRINITY_DN68530_c0_g1_i1	3109	Cystatin
	TRINITY_DN69394_c0_g2_i2	1616	Proton-coupled amino acid transporter 1
	TRINITY_DN70144_c0_g4_i3	504	Cytochrome P450 6B1
	TRINITY_DN72907_c0_g1_i1	1154	Ubiquitin-protein ligase E3A
	TRINITY_DN73053_c1_g1_i1	442	Retrovirus-related Pol polyprotein from transposon 17.6
	TRINITY_DN73433_c0_g2_i1	1108	Frizzled-7
	TRINITY_DN73651_c2_g1_i1	1552	Follistatin-related protein 5
	TRINITY_DN73842_c0_g2_i1	861	Haloacid dehalogenase-like hydrolase domain-containing protein 2
	TRINITY_DN75040_c3_g1_i1	1175	RNA-binding protein fusilli
	TRINITY_DN75843_c0_g1_i1	800	Eukaryotic translation initiation factor 4H
	TRINITY_DN76661_c0_g3_i1	642	Protein unc-50 homolog

	TRINITY_DN79493_c0_g1_i1	1102	Golgi pH regulator
	TRINITY_DN80023_c0_g1_i1	1576	Gametogenetin-binding protein 2-like
	TRINITY_DN80526_c0_g1_i1	1334	Inter-alpha-trypsin inhibitor heavy chain H4
	TRINITY_DN80640_c4_g2_i2	2041	Purine nucleoside phosphorylase
	TRINITY_DN80673_c0_g3_i1	1086	Heterogeneous nuclear ribonucleoprotein A1, A2/B1 homolog
	TRINITY_DN81509_c3_g2_i1	682	Zinc finger CCCH domain-containing protein 11A
	TRINITY_DN81779_c0_g4_i1	748	Casein kinase I isoform alpha
	TRINITY_DN82072_c0_g1_i1	855	Protein Gawky
	TRINITY_DN82377_c11_g1_i1	440	Phosphoenolpyruvate carboxykinase [GTP]
	TRINITY_DN82787_c1_g3_i2	303	Probable cytochrome P450 6a13
	TRINITY_DN83158_c2_g4_i3	750	Putative cysteine proteinase CG12163
	TRINITY_DN83529_c0_g1_i1	857	Putative 1-acyl-sn-glycerol-3-phosphate acyltransferase acl-2
	TRINITY_DN84053_c1_g9_i5	563	Probable cytochrome P450 6a18
	TRINITY_DN105667_c0_g1_i1	543	Otoferlin
	TRINITY_DN110565_c0_g1_i1	486	2-amino-3-ketobutyrate coenzyme A ligase, mitochondrial
	TRINITY_DN129754_c0_g1_i1	443	Excitatory amino acid transporter 2
	TRINITY_DN130818_c0_g1_i1	527	Leucine-rich repeat-containing protein 4C
	TRINITY_DN138840_c0_g1_i1	824	Protein yellow
	TRINITY_DN139615_c0_g1_i1	526	Facilitated trehalose transporter Tret1-2 homolog
	TRINITY_DN152334_c0_g1_i1	577	Proton myo-inositol cotransporter
	TRINITY_DN152670_c0_g1_i1	846	Synaptotagmin-4
	TRINITY_DN170714_c0_g1_i1	502	Cholinesterase
	TRINITY_DN171921_c0_g1_i1	1405	Nuclear protein localization protein 4
	TRINITY_DN11684_c0_g1_i1	668	Retrovirus-related Pol polyprotein from transposon 17.6*
	TRINITY_DN39350_c0_g1_i1	1121	Apolipophorins*
	TRINITY_DN48642_c0_g3_i1	310	Tubulin alpha chain*
	TRINITY_DN64992_c0_g1_i1	1169	SH3 domain-binding protein 5 homolog*
	TRINITY_DN77138_c1_g2_i1	534	Alpha-endosulfine*
	TRINITY_DN79394_c7_g2_i1	695	Cytochrome P450 6j1*
	TRINITY_DN80139_c0_g1_i1	781	Putative fatty acyl-CoA reductase**
	TRINITY_DN81255_c2_g5_i1	601	Putative fatty acyl-CoA reductase**
	TRINITY_DN95926_c0_g1_i1	345	Glucose dehydrogenase [FAD, quinone]**
	TRINITY_DN160441_c0_g1_i1	465	Glucose dehydrogenase [FAD, quinone]**
	TRINITY_DN185680_c0_g1_i1	448	
	TRINITY_DN192088_c0_g1_i1	811	
	TRINITY_DN18898_c0_g1_i1	500	
	TRINITY_DN31253_c0_g1_i1	339	
	TRINITY_DN33873_c0_g1_i1	613	
	TRINITY_DN43995_c0_g1_i1	1601	
	TRINITY_DN64801_c0_g1_i2	515	
	TRINITY_DN82560_c0_g1_i1	924	
	TRINITY_DN84013_c2_g6_i1	426	
	TRINITY_DN100010_c0_g1_i1	314	
	TRINITY_DN105688_c0_g1_i1	625	
	TRINITY_DN107688_c0_g1_i1	497	
	TRINITY_DN148965_c0_g1_i1	483	
III-Native (Boston & Raleigh)	TRINITY_DN1012_c0_g2_i1	812	Probable G-protein coupled receptor Mth-like 1
	TRINITY_DN15625_c0_g2_i1	1236	Zinc finger protein OZF
	TRINITY_DN20472_c0_g1_i1	1990	Probable palmitoyltransferase ZDHHC14
	TRINITY_DN21918_c0_g2_i1	1109	Regulator of telomere elongation helicase 1 homolog
	TRINITY_DN22476_c0_g2_i1	935	Transforming growth factor-beta-induced protein ig-h3
	TRINITY_DN22919_c0_g1_i1	531	26S protease regulatory subunit 6A
	TRINITY_DN31249_c0_g2_i1	815	Protein tramtrack, beta isoform
	TRINITY_DN31877_c0_g2_i1	3380	Hexosaminidase D
	TRINITY_DN32884_c0_g4_i1	1826	Golgi to ER traffic protein 4 homolog
	TRINITY_DN44620_c0_g2_i1	1015	Guanine nucleotide-binding protein G(o) subunit alpha
	TRINITY_DN45311_c0_g2_i1	2399	Glycoprotein 3-alpha-L-fucosyltransferase A
	TRINITY_DN45965_c0_g1_i1	563	Cilia- and flagella-associated protein 20
	TRINITY_DN52132_c0_g1_i1	1600	Phospholipid-transporting ATPase ID
	TRINITY_DN54956_c0_g1_i1	478	Opsin-1
	TRINITY_DN56439_c0_g1_i1	364	V-type proton ATPase subunit B 1
	TRINITY_DN56737_c0_g1_i1	501	V-type proton ATPase subunit B 2
	TRINITY_DN57660_c0_g1_i1	2047	Protein prickle
	TRINITY_DN60557_c0_g2_i1	569	Endochitinase
	TRINITY_DN63111_c0_g1_i1	1511	Eukaryotic peptide chain release factor subunit 1
	TRINITY_DN64992_c0_g3_i1	1169	SH3 domain-binding protein 5 homolog
	TRINITY_DN65230_c0_g3_i1	1181	28S ribosomal protein S31, mitochondrial
	TRINITY_DN66397_c0_g1_i1	578	Peptidyl-prolyl cis-trans isomerase 6
	TRINITY_DN66673_c1_g1_i1	1440	Cytochrome b-c1 complex subunit Rieske, mitochondrial
	TRINITY_DN66853_c0_g2_i1	868	26S protease regulatory subunit 4

TRINITY_DN68257_c0_g1_i3	353	Uncharacterized protein YxjG
TRINITY_DN69134_c0_g1_i1	1200	Zinc finger protein 32
TRINITY_DN69264_c0_g2_i2	564	Putative uncharacterized oxidoreductase C513.07
TRINITY_DN71367_c1_g1_i1	321	Zinc finger protein 271
TRINITY_DN72032_c0_g3_i1	556	Histone H2A.V
TRINITY_DN72415_c0_g3_i2	311	Actin-1
TRINITY_DN73493_c0_g1_i1	675	40S ribosomal protein S5
TRINITY_DN74508_c1_g1_i1	618	Caltractin
TRINITY_DN74531_c0_g2_i1	881	Cell division cycle protein 48 homolog
TRINITY_DN75025_c0_g1_i1	3157	Nitric oxide synthase, salivary gland
TRINITY_DN75055_c0_g2_i1	369	Pathogenesis-related protein 5
TRINITY_DN75133_c0_g1_i2	405	14-3-3-like protein GF14-F
TRINITY_DN75133_c1_g2_i2	321	14-3-3 protein homolog
TRINITY_DN75212_c0_g2_i1	528	V-type proton ATPase subunit B
TRINITY_DN76143_c0_g1_i1	540	Ubiquitin-conjugating enzyme E2-16 kDa
TRINITY_DN76616_c0_g2_i1	2102	Polycomb protein Sfmbt
TRINITY_DN76837_c0_g1_i2	3092	Neurotrimin
TRINITY_DN77258_c0_g1_i2	342	Putative actin-22
TRINITY_DN77293_c0_g1_i4	357	Transient receptor potential cation channel protein painless
TRINITY_DN77969_c1_g2_i2	394	ATP-binding cassette sub-family C member 9
TRINITY_DN78094_c3_g2_i1	626	Serine protease 52
TRINITY_DN78108_c1_g6_i3	325	Histone-lysine N-methyltransferase PRDM9
TRINITY_DN78310_c4_g4_i7	512	Calmodulin
TRINITY_DN78864_c0_g2_i1	1504	GPI ethanolamine phosphate transferase 2
TRINITY_DN79069_c0_g3_i1	396	40S ribosomal protein S23
TRINITY_DN79412_c3_g5_i1	437	Ras-related protein Rab-7A
TRINITY_DN79534_c0_g2_i2	1128	Protein takeout
TRINITY_DN79845_c3_g3_i2	318	Zinc finger and SCAN domain-containing protein 12
TRINITY_DN79964_c1_g3_i2	339	60S ribosomal protein L8
TRINITY_DN80219_c1_g1_i1	359	GTP-binding protein YPTC1
TRINITY_DN80444_c0_g1_i1	1727	Adhesion G protein coupled receptor
TRINITY_DN80550_c2_g6_i1	1015	14-3-3-like protein
TRINITY_DN80554_c0_g1_i3	2629	Thymidine phosphorylase
TRINITY_DN80605_c4_g1_i1	405	Polyubiquitin-B
TRINITY_DN81518_c0_g4_i3	326	Polyubiquitin
TRINITY_DN82158_c0_g6_i1	424	40S ribosomal protein S14
TRINITY_DN82231_c3_g3_i4	425	60S ribosomal protein L10
TRINITY_DN82246_c3_g1_i6	2276	60S ribosomal protein L10a-2
TRINITY_DN82270_c3_g3_i4	3828	Cytochrome P450 6j1
TRINITY_DN82283_c0_g2_i1	1014	Endoplasmic reticulum aminopeptidase 1
TRINITY_DN82350_c1_g3_i1	336	Histone H2A.J
TRINITY_DN82586_c9_g2_i3	397	Zinc finger and SCAN domain-containing protein 22
TRINITY_DN82796_c2_g1_i1	429	78 kDa glucose-regulated protein
TRINITY_DN82954_c2_g2_i2	1380	NADH-ubiquinone oxidoreductase chain 5
TRINITY_DN83114_c0_g1_i2	841	Cytochrome b
TRINITY_DN83447_c4_g3_i3	646	Myosin regulatory light chain 2
TRINITY_DN83462_c0_g1_i5	4167	Ras guanyl-releasing protein 3
TRINITY_DN83740_c2_g2_i1	1800	Probable cytochrome P450 49a1
TRINITY_DN83754_c2_g5_i1	352	Late histone H2A.1
TRINITY_DN114829_c0_g1_i1	491	Enolase
TRINITY_DN41546_c0_g1_i1	1102	Protein scabrous*
TRINITY_DN42976_c0_g1_i1	570	Ubiquitin-conjugating enzyme E2-16 kDa*
TRINITY_DN82098_c4_g1_i1	305	Ubiquitin*
TRINITY_DN83960_c5_g4_i1	312	Actin*
TRINITY_DN75754_c0_g5_i4	309	Actin-10**
TRINITY_DN83414_c1_g12_i2	405	Actin-10**
TRINITY_DN441_c0_g1_i1	505	40S ribosomal protein S13**
TRINITY_DN78182_c0_g2_i1	561	40S ribosomal protein S13**
TRINITY_DN78182_c0_g3_i3	545	40S ribosomal protein S13**
TRINITY_DN71662_c0_g1_i1	472	40S ribosomal protein S15a*
TRINITY_DN48642_c0_g4_i1	310	Tubulin alpha chain**
TRINITY_DN80543_c0_g1_i1	325	Tubulin alpha chain**
TRINITY_DN81551_c1_g2_i4	483	Tubulin alpha chain**
TRINITY_DN83165_c2_g3_i3	411	Tubulin alpha chain**
TRINITY_DN82444_c1_g3_i4	324	Tubulin alpha-1 chain*
TRINITY_DN55283_c0_g1_i1	308	Tubulin beta chain**
TRINITY_DN77394_c0_g3_i1	522	Tubulin beta chain**
TRINITY_DN77454_c0_g1_i4	312	Tubulin beta chain**
TRINITY_DN80732_c0_g6_i2	302	Tubulin beta chain**
TRINITY_DN80732_c0_g7_i3	324	Tubulin beta chain**
TRINITY_DN74172_c1_g2_i4	309	Tubulin beta-1 chain**

	TRINITY_DN57317_c0_g2_i1	456	14-3-3-like protein GF14 iota**
	TRINITY_DN68789_c0_g1_i1	544	14-3-3-like protein GF14 iota**
	TRINITY_DN80108_c1_g3_i1	440	Histone H2A**
	TRINITY_DN82350_c1_g5_i1	305	Histone H2A**
	TRINITY_DN71370_c0_g1_i1	595	Histone H3**
	TRINITY_DN82567_c3_g1_i1	443	Histone H3**
	TRINITY_DN79666_c0_g3_i1	490	Histone H3**
	TRINITY_DN143291_c0_g1_i1	460	Histone H3.2**
	TRINITY_DN67357_c0_g2_i2	592	Histone H3.2**
	TRINITY_DN82567_c3_g5_i3	558	Histone H3.2**
	TRINITY_DN80692_c0_g1_i3	486	Histone H3.2**
	TRINITY_DN82545_c0_g2_i7	394	Histone H4**
	TRINITY_DN79898_c1_g2_i1	311	Histone H4**
	TRINITY_DN82105_c0_g7_i1	596	Heat shock protein 83**
	TRINITY_DN82105_c0_g8_i1	455	Heat shock protein 83**
	TRINITY_DN83052_c0_g1_i1	456	Elongation factor 1-alpha**
	TRINITY_DN83052_c2_g12_i1	369	Elongation factor 1-alpha**
	TRINITY_DN80273_c0_g3_i4	302	Heat shock 70 kDa protein**
	TRINITY_DN83593_c0_g1_i1	469	Heat shock 70 kDa protein**
	TRINITY_DN172989_c0_g1_i1	563	
	TRINITY_DN7166_c0_g2_i1	467	
	TRINITY_DN18469_c0_g2_i1	711	
	TRINITY_DN34667_c0_g1_i1	813	
	TRINITY_DN43995_c0_g2_i1	1601	
	TRINITY_DN45292_c0_g4_i1	1223	
	TRINITY_DN47084_c0_g2_i1	1573	
	TRINITY_DN58163_c0_g2_i1	760	
	TRINITY_DN58331_c0_g2_i1	959	
	TRINITY_DN59347_c0_g1_i1	1224	
	TRINITY_DN61377_c0_g1_i1	1561	
	TRINITY_DN61848_c0_g1_i4	797	
	TRINITY_DN64971_c0_g1_i1	1094	
	TRINITY_DN67211_c1_g3_i1	1947	
	TRINITY_DN67560_c0_g1_i1	534	
	TRINITY_DN72768_c1_g2_i1	542	
	TRINITY_DN72901_c0_g1_i1	1988	
	TRINITY_DN75584_c0_g2_i1	2605	
	TRINITY_DN77598_c0_g2_i5	455	
	TRINITY_DN78482_c2_g1_i2	385	
	TRINITY_DN78686_c0_g1_i2	2386	
	TRINITY_DN79039_c0_g1_i8	845	
	TRINITY_DN81398_c2_g1_i2	829	
	TRINITY_DN81553_c0_g1_i1	4403	
	TRINITY_DN81825_c0_g1_i4	3470	
	TRINITY_DN83108_c3_g1_i1	308	
	TRINITY_DN83687_c1_g4_i7	589	
	TRINITY_DN22636_c0_g2_i1	732	Probable palmitoyltransferase ZDHHC16
	TRINITY_DN36003_c0_g1_i1	1611	U3 small nucleolar RNA-interacting protein 2
	TRINITY_DN41546_c0_g2_i1	1102	Protein scabrous
	TRINITY_DN48317_c0_g1_i1	662	Elongation of very long chain fatty acids protein 4
	TRINITY_DN50229_c0_g1_i1	1115	Transmembrane protein 177
	TRINITY_DN63750_c0_g1_i1	614	Piezo-type mechanosensitive ion channel component 2
	TRINITY_DN67085_c0_g1_i1	1451	Sn1-specific diacylglycerol lipase beta
	TRINITY_DN68938_c0_g2_i1	2061	Coiled-coil domain-containing protein 6
	TRINITY_DN77138_c1_g5_i1	534	Alpha-endosulfine
	TRINITY_DN78825_c2_g1_i8	3558	C-1-tetrahydrofolate synthase, cytoplasmic
	TRINITY_DN80543_c1_g1_i2	314	Tubulin alpha-1/alpha-2 chain
IV-Boston	TRINITY_DN81570_c0_g3_i2	297	Peptidyl-prolyl cis-trans isomerase
	TRINITY_DN82487_c2_g1_i5	4249	Histone-lysine N-methyltransferase, H3 lysine-79 specific
	TRINITY_DN82582_c0_g2_i9	4422	Protein PTHB1
	TRINITY_DN83225_c0_g1_i1	1077	Mitochondrial intermediate peptidase
	TRINITY_DN83754_c2_g6_i8	501	Histone H2A.1
	TRINITY_DN172214_c0_g1_i1	922	Polyubiquitin 12
	TRINITY_DN74365_c2_g3_i1	377	Histone H3.2*
	TRINITY_DN171545_c0_g1_i1	636	Ubiquitin*
	TRINITY_DN82590_c0_g3_i3	436	Tubulin alpha chain**
	TRINITY_DN82590_c0_g9_i1	436	Tubulin alpha chain**
	TRINITY_DN77454_c0_g8_i2	380	Tubulin beta-1 chain*
	TRINITY_DN7588_c0_g1_i1	1512	
	TRINITY_DN11457_c0_g1_i1	635	

	TRINITY_DN25997_c0_g1_i1	1129	
	TRINITY_DN42510_c0_g1_i1	806	
	TRINITY_DN43995_c0_g3_i1	1601	
	TRINITY_DN80332_c1_g2_i1	1432	
	TRINITY_DN81138_c4_g5_i1	552	
	TRINITY_DN83535_c0_g1_i2	858	
V-Native (Boston & Raleigh)	TRINITY_DN77963_c0_g2_i2	377	Dynein light chain LC6, flagellar outer arm
	TRINITY_DN78768_c0_g3_i7	320	Heat shock protein 90
	TRINITY_DN81661_c0_g5_i1	377	Heat shock cognate 70 kDa protein 1
	TRINITY_DN82983_c0_g3_i2	760	Heat shock protein 81-1
	TRINITY_DN83414_c0_g2_i1	382	Actin-2
	TRINITY_DN81515_c4_g1_i3	401	Actin
	TRINITY_DN63138_c0_g1_i1	498	40S ribosomal protein S15a
	TRINITY_DN82495_c0_g2_i2	578	Tubulin alpha chain**
	TRINITY_DN83165_c2_g1_i2	326	Tubulin alpha chain**
	TRINITY_DN83165_c2_g6_i1	409	Tubulin alpha chain**
	TRINITY_DN77157_c0_g1_i1	459	Tubulin alpha-1 chain**
	TRINITY_DN79020_c0_g6_i1	425	Tubulin alpha-1 chain**
	TRINITY_DN78081_c1_g2_i1	492	Tubulin beta chain**
	TRINITY_DN78081_c1_g4_i1	401	Tubulin beta chain**
	TRINITY_DN77454_c0_g7_i1	431	Tubulin beta-1 chain*
	TRINITY_DN72599_c0_g1_i1	482	Elongation factor 1-alpha**
	TRINITY_DN83052_c2_g8_i2	300	Elongation factor 1-alpha**
	TRINITY_DN83638_c2_g3_i2	360	Elongation factor 1-alpha**

* Duplicate in another gene set

** Duplicate within the same gene set

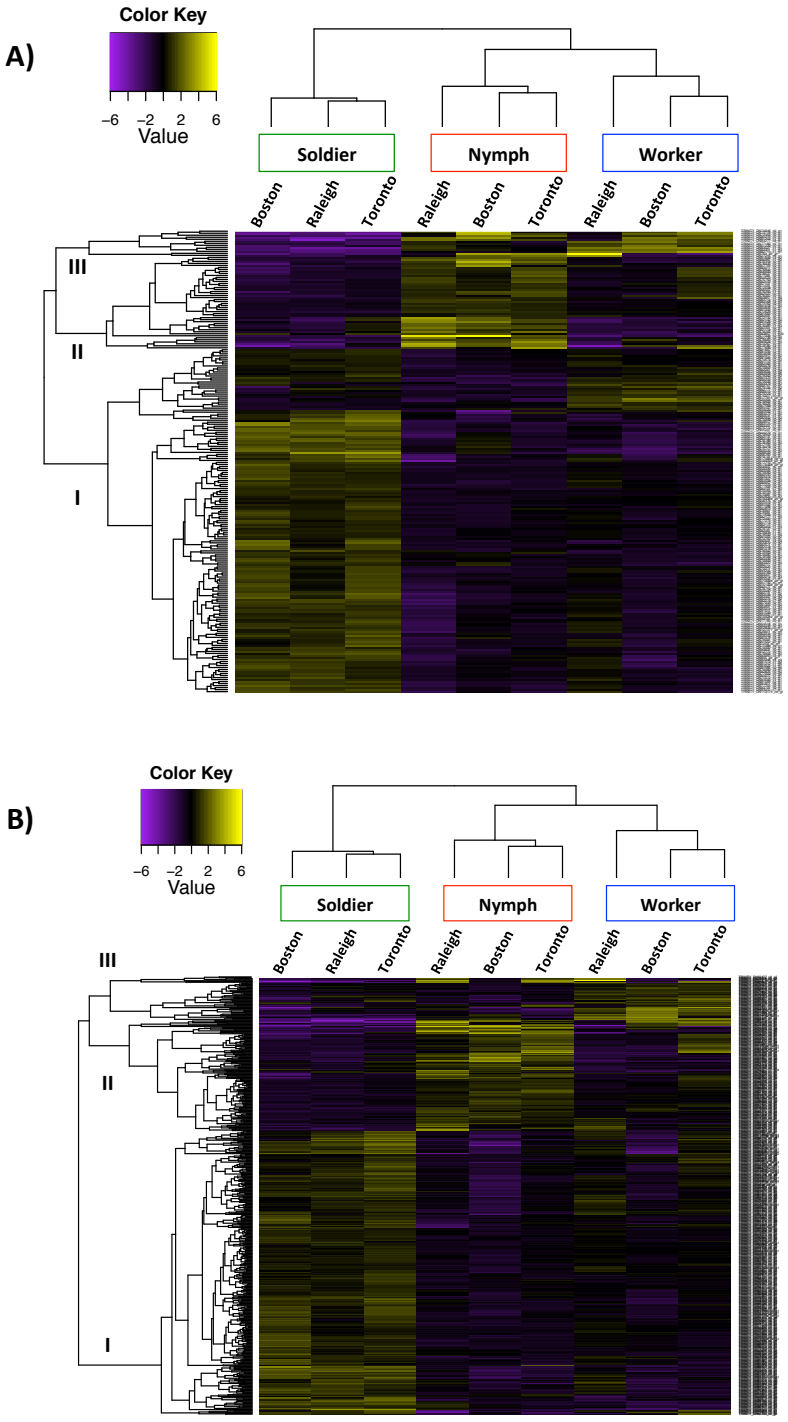


Figure S1. Alternative heatmaps of DEGs between *R. flavipes* castes supporting the relationship of soldiers being uniquely regulated relative to nymphs and workers. DEG criteria of **A)** FDR < 0.001 and 2-fold change, and **B)** FDR < 0.05 and 2-fold change.

Curriculum Vitae

Name: Tian Wu

Post-secondary Education and Degrees:

University of Western Ontario
London, Ontario, Canada
2014-2017 Master of Science in Biology

University of Ottawa
Ottawa, Ontario, Canada
2008-2012 Honours Specialization in Biology

Honours and Awards:

1st Place Graduate Student Talk
XXV International Congress of Entomology, Orlando, FL.
2016

Best Talk for Ecology and Evolution Philosophical
Department of Biology, University of Western Ontario
2016

Founder & President
Mustang Bioinformatics Club
2015-2017

Related Work Experience

Teaching Assistant
The University of Western Ontario
2014-2016

Research Technician
Agriculture and Agri-Food Canada
2012-2014

Publications:

Robideau, GP., Foottit, RG., Humble, LM., Noseworthy, MK., **Wu, T.** & Bilodeau, GJ. (2015) Real-time PCR identification of the ambrosia beetles, *Trypodendron domesticum* (L.) and *Trypodendron lineatum* (Olivier) (Coleoptera: Scolytidae). Journal of Applied Entomology, 140: 299-307.

Wu, T., Simkovic, V. & Thompson, GJ. (March and May 2015) Subterranean termites: the evolution of a pest. Pest Control Technology Canada, Pest Control Technology US, 3: 34-42. (Non-refereed)