A Survey of Genome Size Diversity Within Scale Insects (Hemiptera: Coccoidea) and Cockroaches and Termites (Blattodea)

by

Darren James Kelly

A Thesis presented to The University of Guelph

In partial fulfilment of requirements for the degree of Master of Science in Integrative Biology

Guelph, Ontario, Canada

© Darren James Kelly, May, 2018
A SURVEY OF GENOME SIZE DIVERSITY WITHIN SCALE INSECTS (HEMIPTERA: COCCOIDEA) AND COCKROACHES AND TERMITES (BLATTODEA)

Darren James Kelly  
University of Guelph, 2018

Advisor: Dr. T. Ryan Gregory

This thesis is an investigation of the genome size diversity present within two diverse insect groups: the hemipteran superfamily Coccoidea (scale insects) and the order Blattodea (cockroaches and termites). Data have previously been lacking for these two groups, even though both present some intriguing potential patterns of genome size diversity related to unique aspects of their biology. More than 140 species from the two groups were estimated for genome size using Feulgen Image Analysis Densitometry (FIAD). Coccoidean genome sizes ranged from 0.15pg-2.1pg (n=62 species) while blattodeans ranged from 0.58pg-9.58pg (n=70 species). These data were then correlated with organismal traits related to ecology, morphology, metamorphosis, and reproduction. Genome size was found to be most strongly related to body size and geographic range size in scale insects and to the evolution of sociality in blattodeans.
Acknowledgments

To start, I am extremely thankful for this opportunity to complete my MSc under Dr. T. Ryan Gregory. I am ever grateful for being guided by an accessible, humorous, and knowledgeable advisor for this journey; I left every meeting and discussion with further understanding and confidence in my work. This MSc was also possible because of all the support and love given to me by my parents and family - it is from this that I find my drive and purpose. My MSc really started during my undergrad with the Arctic Ecology field course under Dr. Alex Smith and Dr. Sarah Adamowicz, who could not have done a better job at teaching skills and forging bonds between such a large group of undergraduate researchers. From this field course I met Dr. T. Ryan Gregory, and the tale continues from this point.

Thank you to my committee member Dr. Robert Hanner for kindly donating your time and critical eye towards this project. I am indebted to Dr. Roger Gwiazdowski who shared his passion for scale insects and phylogenetics with me, including ‘free-spirited’ sampling of scales in the local greenhouses. I am thankful for the moments of de-stressing and humour in the last two years from every Gregory lab member, members of the 2015 MSc. cohort, NJWAZ, and all my newly forged friendships; whether from nerding out, themed parties, Muay Thai sessions, or philosophical discussions on whether or not a hotdog is a sandwich.

The selflessness and willingness to help which was expressed by the scientific community of collaborators used for sampling in the current project will not be forgotten. For all the evils of academia, I am reminded of the truly good people that make up the human side of affairs. I owe thanks to: [Coccoidea] Gillian Watson, Nick Nisson, Paris Lambdin, Zvi Mendel, Lyn Cook, Yair Ben-Dov, Jean-François Germain, Disna Gunawardana, Penny Gullan, Malkie Spodek, and anyone else who contributes to ScaleNet, [Blattodea] Tim Myles, Tang Qian, Dominic Evangelista, and Alan Jeon.

In an overarchin
g way I also appreciate the environment I was raised in, Canada, for its beauty and inspiration in all things natural. I advise everyone to take time to embrace the outdoors, especially during periods of stress - from Darwin strolling along his sandwalk to the hundreds of backcountry kilometers I was able to put under my feet during my MSc; the benefits are endless.

As if to dwarf all conclusions found within the present study and their significance, keep in mind that some species of scale insects have no butt and the scientific world currently does not know why.
# Table of Contents

Chapter 1 - General Introduction: Drivers of Insect Genomic Evolution........pg. 1  
   Genome Size and the C-value Enigma..........................................................pg. 2  
   Expected Patterns between Genome Size and Organismal Traits................pg. 5  
   Difficulties in Estimating Insect Genome Sizes........................................pg. 12  
   Study Organisms.........................................................................................pg. 13  
   Scale Insects..............................................................................................pg. 13  
   Cockroaches and Termites........................................................................pg. 13  
   Research Questions and Objectives............................................................pg. 14

Chapter 2 – An Exploration of Genome Size Diversity in Scale Insects (Hemiptera: Coccoidea)..................................................................................pg. 16  
   Abstract........................................................................................................pg. 16  
   Introduction..................................................................................................pg. 16  
   Scale Insects..............................................................................................pg. 16  
   Scale Insect Traits in Relation to Genome Size........................................pg. 17  
   Methods......................................................................................................pg. 19  
   Specimen Collection....................................................................................pg. 19  
   Feulgen Staining........................................................................................pg. 20  
   Genome Size Estimation............................................................................pg. 22  
   Sources of Trait Data................................................................................pg. 24  
   Phylogenetic and Statistical Analyses.......................................................pg. 24  
   Data Exploration.........................................................................................pg. 25  
   Phylogeny...................................................................................................pg. 26  
   Phylogenetic Signal....................................................................................pg. 27  
   Phylogenetically Independent Contrasts (PICs)......................................pg. 28  
   Statistical Modelling................................................................................pg. 29  
   Results........................................................................................................pg. 29  
   Haploid Genome Size Estimates.............................................................pg. 29  
   Statistical and Phylogenetic Results.........................................................pg. 30  
   Discussion.................................................................................................pg. 31  
   Genome Size Diversity in Coccoidea.......................................................pg. 31  
   Genome Size and Metamorphosis..........................................................pg. 32  
   Genome Size Constraints in Coccoidea................................................pg. 33  
   Prospects for Future Research.................................................................pg. 35

Chapter 3 – Genome Size Diversity in Cockroaches and Termites (Blattodea): the Impacts of the Evolution of Sociality........................................pg. 37  
   Abstract......................................................................................................pg. 37  
   Introduction...............................................................................................pg. 37
List of Tables

**Table 1.** Statistical outputs for both coccoidean datasets. These datasets include contrasts of genome size and explanatory variables which are controlled for phylogeny (PICs or PGLS), while the other is the absolute relationship between genome size and variables without controlling for phylogeny (direct correlation). Asterisks (*, **, ***) denote significance levels matching the default in R. $r^2$ shading denotes a positive (green) or negative (red) significant relationship with genome size.

**Table 2.** Statistical outputs for both blattodean datasets. These datasets include contrasts of genome size and explanatory variables which are controlled for phylogeny (PICs or PGLS), while the other is the absolute relationship between genome size and variables without controlling for phylogeny (direct correlation). Asterisks (*, **, ***) denote significance levels matching the default in R. $r^2$ shading denotes a positive (green) or negative (red) significant relationship with genome size.

**Appendix Table 1.** Haploid genome size estimates for Coccoidea and Sternorrhynchan outgroups. Sample values (n) are given per family (number of species sampled in bold) and per species (number of individual specimens measured).

**Appendix Table 2.** Haploid genome size estimates for Blattodea and associated outgroups (Gryllidae and Mantodea). Sample values (n) are given per family (number of species sampled in bold) and per species (number of individual specimens measured).

List of Figures

**Figure 1.** From Gregory (2005a). Graphical distribution of C-values across insect groups with different developmental programmes. Holometabolous groups, which are above the horizontal line, appear to be genomically constrained and generally fall below the 2pg threshold. Hemimetabolous and ametabolous groups, which are below the horizontal line, comparatively show extreme variation and appear unconstrained.

**Figure 2.** (A) non-measureable smeared tissue (sheared nuclei), (B) smeared tissue that is too thick and cannot be measured, (C) a mono-layer of single nuclei, but most are slightly ruptured, (E) another mono-layer of single nuclei containing more compact nuclei than (C); these are measurable. (D) Massive polyploid bacteriome nuclei (right side) next to smaller diploid nuclei (left side). In this study, the most abundant cell type is assumed diploid.
DNA content varies across cell phases (left graph), leading to cells with similar sizes but different DNA content. The right photo is of normal cells and those that are in the process of synthesizing DNA (black dots). This reinforces the assumption that the most abundant cells will be diploid, while variation will occur based on cell phase state. Besides phases, ploidy levels also vary in gametes and the bacteriomes of scale insects.

Methods used to conduct data exploration before model fitting for the data, in this case log body size versus log genome size. These methods include residual analysis for heteroscedasticity (A and C), Q-Q plots for normality (B), and Cook’s distance for outliers (D).

Scatterplot matrix used to observe collinearity occurring between explanatory variables in Coccoidea. Upper right panels contain correlation values and their significance (with *** being highly significant), while the bottom left panels show correlation plots. Collinearity causes some of the explanatory power of one variable to already be contained within another. For example, number of plant hosts strongly correlates with scale insect geographic range size ($r^2=0.87$***). As such, the explanatory power of number of hosts is already partially contained within geographic range size as a variable; the weaker of the two variables is removed from statistical analyses.

Relationship between log genome size and log body size at the family level. There is a strong relationship for both Coccidae and Pseudococcidae ($p≤0.03$, $r^2≥0.37$) which either have soft scales or are fully exposed. Diaspididae and Eriococcidae contain highly protected species, whether under armoured scales or within plant galls, and the relationship is non-significant ($p≥0.61$). This presents an interesting idea that the relationship may be driven by factors such as predation pressure causing selection on body size, which cascades to secondarily select for genome size, and this constraint may be significantly lessened in the protected groups.

Relationship between log genome size and log reproductive frequency when analyzed at the family level. The comparative analysis across all species is significant ($p=0.004$, $r^2=0.19$), but within families there are no significant relationships observed between the two variables ($p≥0.45$). This signals that differences in these variables lie between clades and not within clades. Additionally, controlling for phylogeny removed this relationship – the pattern is explained by phylogenetic relatedness and for extant species reproductive frequency does not appear to directly impact genome size.
Figure 8

Composite coccoidean phylogeny for 63 of 65 sampled species generated from existing phylogenetic data, including three outgroup species. Bars on the right denote genome size, while bars and circles are coloured by family.

Figure 9

A positive relationship between genome size and body size is observed in Coccoidea (p=5.5e-4, r²=0.17-0.25), yet the relationship is non-existent in cockroaches (p=0.98, r²=0.02). Genome size in Blattodea may be determined by other traits, such as sociality.

Figure 10

Genome size estimates for B. dubia across different time regimes and treatment groups using tissue preserved in 70% EtOH. C-value estimates from fresh tissue align with previously published data for B. dubia, but these values then quickly decrease in value and increase in variation after one week, one month, and one year in ethanol. ‘FF’ is a freeze-flip method used on one-year preserved tissues, where tissue is frozen to a microscope slide and then sheared off, leaving behind a monolayer ‘thumbprint’ of nuclei. This is done after tissue is rehydrated with either citric acid (‘CA’), acetic acid (‘AA’), or water (‘W’), but still results in unstable estimates. Overall, preserved tissues continue to conflict with the Feulgen process and remain unreliable to measure in insects.

Figure 11

Composite blattodean phylogeny for 68 of 70 sampled species generated from existing phylogenetic data. Bars on the right denote genome size, while bars and circles within the tree are coloured by family. The evolution of sociality is shared by Cryptocerus cockroaches (subsocial, red) and termites (eusocial, brown), and a dramatic drop in genome size is observed in these groups relative to their solitary counterparts.
Chapter 1 - General Introduction:

Drivers of Insect Genomic Evolution
**Genome Size and the C-value Enigma**

Eukaryotic genomes not only provide the genetic information necessary for life, but they also influence various biological features through their physical and structural aspects. One core aspect to consider is genome size. Genome size is the amount of DNA present within a haploid chromosome set (also known as ‘C-value’ in diploid organisms; see Greilhuber *et al.* (2005) for more detailed terminology). It is typically measured in mass (picograms, pg) or nucleotide base pairs (megabase pairs, Mbp).

Genome size varies 200,000-fold across eukaryotes, and 7,000-fold in animals, yet this diversity is not correlated with organismal complexity or number of genes (Gregory, 2005a; Gregory *et al.*, 2007). The presence of large quantities of non-coding DNA explains this previously puzzling observation; most eukaryote genomes are dominated by transposable elements (TEs), along with introns, tandem repeats, and other DNA elements besides protein-coding sequences (Elliott & Gregory, 2015). Even given this understanding, many important genomic evolutionary questions remain in regards to the origin and maintenance of this genomic variation across the tree of life (Elliott & Gregory, 2015). What types of non-coding DNA are found in specific taxa? What is their abundance in different genomes? How do these elements vary across taxa, and how did this diversity originate? What is the biological significance of genome size, and which levels of organization does it impact? These questions, among others, are core components of an umbrella evolutionary problem named the ‘C-value enigma’ (Gregory, 2001).

As an investigation related to the C-value enigma, the main focus of this current work is to measure the diversity of genome sizes across species and the potential interactions between genome size and traits at different levels of organization. The latter arises by virtue of the fact that genome size correlates positively with cell size and inversely with cell divisions rates (Gregory & Hebert, 1999). This often causes organismal traits that rely on these cellular properties to have a secondary indirect relationship with genome size; traits under the categories of development, morphology, reproduction, and metabolism are known to (1) be impacted by genomic changes, creating organism-level variation which (2) selection acts on, imparting a downward
pressure on DNA fluxes (Petrov, 2001; Gregory, 2005a). Therefore, this targeted subset of the C-value enigma is often explored by measuring the genome sizes of taxa and uncovering any relationships with organismal traits within these categories. Outside of the C-value enigma, these genomic data are also directly applicable to dictating which species to target for full genomic sequencing, as the cost and difficulty of assembly are directly related to how large the genome is (Gregory, 2005b; Jeffery & Gregory, 2014).

Although many eukaryotic groups remain understudied in terms of genome size estimates, invertebrates – and especially arthropods – are particularly underrepresented in this regard, given their extraordinary diversity. Insects are the most speciose animals (with 1 million described and 6-10 million estimated species) (Novotny et al., 2002) and yet only have 1,345 recorded genome size estimates within the Animal Genome Size Database (Gregory, 2018a). This contrasts with vertebrates, whose 66,000 known species are represented by 5,191 records (Gregory, 2018a). Therefore, in the present thesis insects are targeted in order to generate new genome size data and investigate evolutionary patterns related to the C-value enigma.

Studies relating to the C-value enigma can be conducted through a discovery-based hypothesis-generating approach, where unmeasured taxa are explored for the first time, and/or a hypothesis-testing approach, where previously observed patterns and explanatory hypotheses are further tested. Within-taxon patterns can then be comparatively analyzed between-taxon across the tree of life, increasing our understanding of shared general patterns in genomic evolution. This thesis uses both approaches, one being discovery-based and the other a more focused attempt to test a proposed pattern, in a study of two very different groups of insects for which genome size data are lacking: scale insects (Hemiptera: Coccoidea; Chapter 2) and cockroaches and termites (Blattodea; Chapter 3). The scale insects represent the discovery-based approach to these questions, as the genome size diversity for this group is unmeasured thus far and any genomic patterns are unknown. Conversely, Blattodea as a group has a small set of genome size estimates available (24 species) (Gregory, 2018a); this is enough that various hypotheses have emerged which require
more rigorous sampling. This includes hypotheses regarding the evolution of sociality, which is focused on in the present project.

Foundational to exploring these groups and patterns is an understanding of the linkage between genome size and organism-level traits, which is gained through outlining how genome size itself evolves. The following flow of interactions across three categories forms the conceptual framework for the investigations in the current thesis:

**Mechanisms of genome size change at the genomic level:** There are multiple intragenomic mechanisms capable of changing genome size, with the proliferation and loss of transposable elements, ‘indels’ (insertion-deletion mutations) of various size, replication slippage, and small- and large-scale duplications being considered the major drivers of changes in DNA content (Petrov, 2001). Additionally, DNA loss is a significant factor that helps to keep mechanisms of DNA expansion in check, through mechanisms such as illegitimate recombination and indel biases towards deletions (Petrov 2002; Kapusta *et al.*, 2017). However, on average DNA fluxes from intragenomic forces (especially transposable element insertions and various modes of duplication) might be expected to tend towards gains (Gregory, 2004), creating an ‘upward’ pressure on DNA content. Not only are there many mechanisms for DNA insertion, but the deleterious effects of deletions on gene function are dependent on size as well as location, whereas even large insertions may be effectively neutral if they insert into a non-genic region (Gregory, 2004). This ‘upward’ pressure especially arises from TE behavior: TEs have their own intragenomic selection for elements with greater proliferation through acting as “selfish DNA”, while even silenced TEs remain in the genome and increase genome size (Doolittle & Sapienza, 1980; Orgel & Crick, 1980; reviewed in Rigal & Mathieu, 2011).

**Organism-level selection:** Regardless of DNA loss or gain, any DNA flux from intragenomic processes can create phenotypic variation that selection can act on if this variation promotes or conflicts with organismal survival and reproduction. Since DNA content influences nuclear and cellular volume (Bennett, 1972a; Gregory, 2001), this organism-level selection mostly acts on organismal traits that are determined by cell size and division rate. Therefore, there exists a cascade of effects from genome size to
cell size to the organism-level. Selection on this phenotypic variation can lead to a counterbalance of downward constraints which can limit DNA content expansion (Petrov, 2001). That is, if traits linked to cellular parameters are negatively burdened by excess DNA then this excess DNA will be selected against at the organism-level, and therefore limit DNA expansion. For example, in a process such as metamorphosis in insects, TE proliferation will be indirectly constrained by selection at the organism-level if it impacts negatively on this developmental process; metamorphosis needs to be completed within a time-sensitive window, which increased DNA may conflict with if it is slowing down cell division rates (Gregory, 2002). Therefore, a species’ observed genome size may often be the net result of ‘upward’ genomic drive equalizing at a threshold set by downward organism-level selection.

**Comparative interspecific analyses:** The relative strengths of these upward (intragenomic) and downward (organism-level) forces vary across species, with the force of natural selection and genetic drift also varying across different population sizes (Lynch & Conery, 2003; Lynch, 2007). Therefore, the summation of these factors leads to the observed genome size diversity across life. Expanding the above example further, assuming upward drive is held constant, insects with different developmental modes will differ in the downward constraint imparted by selection at the organism-level. Holometabolous insects with complete metamorphosis will be highly constrained compared to those insects with hemimetabolic development (i.e., incomplete metamorphosis) or ametabolic development (i.e., no metamorphosis). This is due to both the complexity of their developmental regimes and the time-sensitive rate at which they must be completed (Gregory, 2002). The groups with a lessened developmental demand can be seen as instead being constrained by the next strongest threshold-setting trait.

**Expected Patterns between Genome Size and Organismal Traits**

Threshold-setting traits are most likely to be traits linked to cell size or division rates, and common trends are seen among eukaryotes in four main categories of such traits: (1) body size, (2) metabolism, (3) development, and (4) reproduction.
1) **Body Size**: this parameter changes with genome size if there is an effect on cell size but no concomitant change in cell number (Gregory, 2005a). This relationship is often observed in groups with determinate growth (i.e., growth ceases upon reaching a determined structure) and relative cell number constancy, compared to those with indeterminate growth (i.e., growth is continuous throughout ontogeny). Examples of taxa in which genome size correlates with body size include polychaete annelids, flatworms, copepod crustaceans, and aphids, among others (McLaren *et al.*, 1988; Soldi *et al.*, 1994; Finston *et al.*, 1995; Gregory *et al.*, 2000; Gregory, 2005a). It remains to be seen how widespread a genome size-body size relationship is across insects.

2) **Metabolism**: an expected relationship between genome size and metabolic rate arises from traits that rely on surface area to volume ratios of cells and the effects these have on gas and ion exchange across cellular membranes. Flight as a trait is an extreme case of metabolic demand, and in endotherms it is seen to impart a large constraint on DNA expansion (Wright *et al.*, 2014). However, the same trait is difficult to measure in insects and any linkage between flight and genome size is relatively unexplored; since insects do not have circulating respiratory cells, and they complete flight via different structural requirements than birds and bats, a directly equivalent study is difficult. As such, the category of metabolism in terms of genomics has not been the subject of much insect research (Wang *et al.*, 2014), and a rigorous investigation is beyond the scope of this current thesis.

3) **Development**: Within the category of developmental traits, there are two important parameters to understand in relation to genome size: developmental rate and developmental complexity (Gregory, 2002).

   i. **Developmental rate** refers to the amount of time it takes for an organism to undergo development. This correlates with genome size when the complexity of development (i.e., the amount of developing to be completed) is held constant, such as across similar species. This has been studied within lady beetles (Gregory *et al.*, 2003), drosophilid flies (Gregory and Johnston, 2008), and aphids (Finston *et al.*, 1995); excess DNA content in one species can be viewed as a burden towards completing a similar amount of development compared to a species with less DNA and subsequently
smaller, faster-dividing cells. Any delay or increase in cell division rates due to changes in DNA content is experienced by every single cell that replicates throughout development; the impact of genome size changes therefore is the sum of this delay or expediency across every cell and how this influences the phenotype.

ii. *Developmental complexity* also has two important components. Firstly, the amount of developing to be done within a unit of time is a parameter that can be extremely constraining for genome size. This occurs when there is a time-sensitive window for development, and a critical amount of developing needs to be completed; the main example of this phenomenon is metamorphosis, which has been studied mostly in amphibians (Werner, 1986; Gregory, 2003a) and insects (Gregory, 2005a). For insects, there appears to be a significant enough constraint that genome sizes above 2pg are rarely observed in holometabolous groups (i.e., those with complete metamorphosis), while both hemimetabolous and ametabolous groups vary considerably more in their C-values (range=0.07-16.93pg; Figure 1) (Gregory, 2018a).

Secondly, complexity is important to consider in terms of organs and situations in which the amount of space available for a tissue is limited. In this case, genome size is capable of influencing cell size, which influences cell-to-cell connections, cell numbers, and overall organ complexity. For example, in amphibians it is thought that salamanders have lost neural complexity partially due to expanding genome sizes; DNA content expansion in these salamanders has resulted in larger cell sizes, while skull size is maintained or even reduced, leading to fewer cells per unit space and resulting in simplified brain morphology (Roth et al., 1997). The opposite phenomenon is thought to drive the maintenance of complexity in some insects, specifically those that have evolved sociality – eusociality is associated with miniaturization of body sizes which forces complex neural functions to be maintained in much smaller spaces (The Honeybee Genome Sequencing Consortium, 2006; Niven & Farris, 2012). This is thought to therefore impart a secondary constraint on genome sizes, as a higher density of smaller cells is required in these smaller structures (Niven & Farris, 2012). This has been studied mostly in Hymenoptera, but the same pattern may be observed in eusocial termites (The Honeybee Genome Sequencing Consortium, 2006; Koshikawa et al.,
2008). This comprises the hypothesis-testing approach to genomic evolution in this current thesis for Blattodea, where social cockroaches and termites are predicted to have smaller genome sizes than their solitary counterparts (Chapter 3).

4) Reproduction: The linkage between reproduction and genome size is that if an organism’s intrinsic rate of increase, fecundity, or reproductive output is reliant on cell size (e.g., via egg size) or cell division rates for efficiency, then genome size changes will influence these organism-level properties. For example, a decrease in genome size was found to be related to increased reproductive fitness in sand beetles under experimental conditions, both across and within species (Arngvist et al., 2015). Under r/K selection theory, organisms may have smaller genome sizes, smaller cell sizes, and more reproductive output but with less parental investment (r-selection) compared to organisms with a larger genome, larger cells, and less offspring but with more parental investment (K-selection) (MacArthur & Wilson, 1967; Pianka, 1970). Similar to the linkage between development and C-value, an organism with a smaller genome size and smaller cells may have increased reproductive output within a set unit of time, possibly dictating how many generations per year are able to be produced. This is often the case when comparing annual and perennial plants, where annual plants have faster cell cycle timing, lower meiotic division duration, and smaller genome sizes compared to perennial plants (Bennett, 1972b). Overall, any reproductive parameter such as fecundity or frequency is expected to have a negative relationship with genome size – the rate of these processes and overall output may be slowed down in light of increased DNA content compared to smaller genomes.

This cascade of effects arises from the impacts caused by changes in genome size; the phenotype is expected to change, through cell size, when genome size shifts. However, reproductive mode (asexual versus sexual) represents an important trait since it is capable of impacting the intragenomic mechanisms which determine these initial changes in genome size. In terms of asexually reproducing organisms that lack sexual recombination, there are three main reasons why reproductive mode may be relevant for affecting genome size:
i. The Fate of TEs: TE virulence and outcrossing within a population is restricted in asexual populations, which can be thought of as contained vertical clonal lineages (barring horizontal gene transfer) (Hickey, 1982). Therefore, at the genome level asexual populations may potentially limit TE accumulation, resulting in smaller overall genome sizes. This containment of TEs is also contingent on the diversity and abundance of TEs that comprise the ancestral set that the asexual species inherits. Similar to a Muller’s ratchet scenario, a species that has a particularly virulent set of TEs that then evolves asexuality may be more likely to become extinct since it lacks reliable mechanisms for escaping this pressure over time (Lynch et al., 1993; Arkhipova & Meselson, 2005). Conversely, inheriting a more stable set of TEs while having efficient silencing and excision mechanisms may lower the overall influence and genomic pressure from TEs. An example of this is the rotifers, whose asexual species have relatively small genomes (average = 0.14pg) and limited TE abundance and diversity compared to their sexual counterparts who have much larger genomes (average = 0.72pg) (Arkhipova & Meselson, 2005; Gregory, 2018a). Overall, this contrasts with sexual populations which allow TE outcrossing between their vertical lineages, increasing their spread throughout a population (Hickey, 1982). Therefore, under TE transposition theory, sexual species should have larger observed genomes, while asexual species should have decreased DNA content in order to limit extinction risk.

ii. DNA Loss Mechanisms: A lack of recombination means that asexual species are also missing important mechanisms that can cause DNA loss, such as illegitimate recombination and unequal crossing over. These two mechanisms are major contributors towards DNA loss, so this may lead to a greater overall accumulation of DNA content over time for asexual species. Although the literature on illegitimate recombination is generally focused on plants or bacteria, it is ubiquitously found to be a large factor for DNA loss through causing large deletions (Devos et al., 2002; Kegel et al., 2006). For example, certain vertebrate groups have limited C-value ranges even in light of large TE variation and proliferation; this expansion pressure is seen to be counteracted by equal-force DNA loss, leading to C-value maintenance (Kapusta et al., 2017). This DNA loss was initially proposed as being driven by an organism’s genetic insertion-deletion bias in small (<400bp) sequences (Petrov 2002). However, these
microdeletions are not strong enough to account for the strength and size of most DNA loss that has been observed (Gregory, 2003b) – illegitimate recombination and other sources of large deletions are imperative. Overall, the loss of these mechanisms in asexual species is another reason one may expect genome sizes to differ between reproductive modes across species. In this case the expectations are reversed compared to those under TE transposition, and asexual species may have larger genomes due to a lack of large-scale deletion mechanisms compared to their sexual counterparts.

iii. The Fate of Mutations: Differences between asexual and sexual reproductive modes not only influence the fate of TEs, but also deleterious mutations. Therefore, the final difference between these modes deals with Muller’s ratchet – a lack of recombination in asexual groups makes it more difficult for deleterious mutations (regardless of their source) to be purged from the genome (Muller, 1964). This potentially increases their accumulation over time and/or the extinction risk for these species (Gabriel et al., 1993; Lynch et al., 1993). Importantly, studies often show that population size is a large driver in overcoming these issues. Modeling has shown asexual groups, given sufficient population size, to persist through TE proliferation even if transposition rates are greater than excision and silencing rates (Dolgin & Charlesworth, 2006). Furthermore, Ross et al. (2012a) posit that asexual scale insects have compensated for the negative aspects of this reproductive mode through enormous population sizes, large geographic range distributions, and increased polyphagy compared to sexual groups; the long-term advantage of sexual reproduction over asexual reproduction is much reduced at very large population sizes, and instead the high intrinsic rate of increase associated with clonal asexual populations dominates (Ross et al., 2012a). As such, asexual groups can be seen as relying on a stronger force of natural selection to purge these mutations as opposed to actual genetic mechanisms such as recombination. And indeed this has been found in invertebrate groups – a transcriptome data analysis of asexual mites found stronger purifying selection in these lineages compared to their sexual counterparts when a large enough population size was present (Brandt et al. 2017). However, even with strong purifying selection in asexual species, sexual species are able to effectively control mutation
accumulation through recombination and may have smaller genomes than asexuals. Nevertheless, both reproductive modes may have similar observed genome sizes since asexual species that have accumulated deleterious mutations and have large genomes may be rare due to heightened extinction risk. As such, an explicit prediction regarding the fate of mutations as a genomic mechanism remains unclear; that is, Muller’s ratchet is not sufficient enough to predict differences in genome size between reproductive modes. Overall, reproductive mode influences intragenomic processes that provide the source of upward drive in genome size determination. This is done through the presence, absence, and various strengths of mechanisms that change across taxa and provide the foundation for interesting comparative analyses.

The outcomes of the above processes for reproductive mode are shown to have opposing expectations for genome size. Some mechanisms lead to assumptions of increased genome size for asexual species, while this is countered by others. However, these are not taken to be equally plausible competing mechanisms or hypotheses, but rather all three of the main processes above (among others) are important for determining upward genomic pressure. Any differences found between asexual and sexual species therefore are not attributed to individual mechanisms, but rather to the summation of the effects of TEs, mutations, DNA loss, and other intragenomic processes that are not specifically analyzed herein. This applies to various cases in the present thesis, where multiple mechanisms and opposing expectations are present - in these situations it is more beneficial to raise direct research questions instead of hypotheses (e.g., “does genome size differ between sexual, asexual, and facultatively asexual species?”).

**Non-Cellular Linked Traits:** Body size, metabolism, development, and reproduction represent trait categories which have previously explored patterns with genome size in various eukaryotic groups through associations with cellular parameters. Beyond these, certain traits that are not linked to cell size or division rates may also have correlations with genome size, potentially as a side effect of collinearity with other genome-associated traits. For example, the geographic range size of scale insects is most likely not directly related to cell size or genome size. However, if scale insect
geographic distribution varies across asexual and sexual species then genome size and this distribution will be correlated through reproductive mode (assuming genomic differences in genome size between reproductive modes, as outlined above). As such, these non-cell size traits are still explored in cases where they are extremely prominent or traits unique to the targeted taxonomic group, even if a direct link to genome size is unclear. These are often very ecologically-focused traits (e.g., geographic distribution, polyphagy, trophic interactions, etc.) or intragenomic traits (e.g., chromosome number), but shared patterns in similar traits across taxa still need to be explored.

Overall, patterns between genome size and phenotypic diversity of many important taxa remain unexplored, including across both cellular-linked and non-cellular linked trait categories. Insects represent a vast majority of animal diversity, yet are under-sampled in analyses, including genomic ones. As such, our above understanding of genomic evolution within this significant group, and truly across life, is lacking.

**Difficulties in Estimating Insect Genome Sizes**

In general, insects are under-sampled in terms of genome size measurements even though they are incredibly speciose. Studies focusing on common insects (often pests) cause a large sampling bias towards these species and the majority of non-pest species are rarely sufficiently studied (Zink & Rosenheim, 2004). One difficulty that has delayed genomic analyses within the insects is that many collections of organisms are stored within preservation media. Whether ethanol, RNAlater, or another medium, preservation has historically conflicted with one of the main methodologies used to generate genome size estimates (Jeffery & Gregory, 2014), Feulgen Image Analysis Densitometry (FIAD), which is covered in Chapter 2. The breadth and diversity of available insect samples would dramatically increase if this methodological hurdle was overcome. Therefore, a new methodology was also tested for a subset of the insect samples which are from preserved collections – this is outlined and tested in Chapter 3.
**Study Organisms**

**Scale Insects**

The superfamily Coccoidea consists of approximately 8,000 scale insect species across 33 families and 1170 genera (García *et al*., 2018), all of whom are obligate phytoparasites, and are most closely related to white flies and aphids (encompassing Hemiptera: Sternorrhyncha). The unique life history of the Coccoidea involves a sedentary lifestyle of feeding on plant phloem while protected by either a secreted scale or within a plant gall (Gullan & Kosztarab, 1997; Gullan *et al*., 2005). This has led to the evolution of numerous traits that are potentially linked to genome size diversity, including polyphagy, geographic range sizes, body size, reproductive frequency, reproductive types (asexual, sexual, or facultative/both), number of predators, and chromosome number (see Chapter 2). Therefore, of the four main trait categories with linkages to cell size, morphology, reproduction, and development are analyzed herein, in addition to unique non-cell size related traits such as chromosome count and polyphagy. Because scale insects comprise the discovery-based component of the current thesis, the main focus is simply exploring the genome size diversity present within Coccoidea across their phylogeny for the first time. This can also be seen as contributing important order-level samples as well, as Hemipteran genome size records are restricted to 75 species records with a bias towards aphids and assassin bugs (range=0.18pg-7.01pg, average=1.06pg) (Gregory, 2018a).

**Cockroaches and Termites**

The order Blattodea consists of approximately 4,600 cockroach species (8 families, 460 genera) and 3,000 termite species (12 families, 300 genera). Analyzing similar organismal traits in Blattodea as those available in Coccoidea would be beneficial, but there is an unfortunate lack of trait data within the literature for this group. The few species which are regarded as pests, 200 of over 7,400 total species (Su & Scheffrahn, 2000), are well-studied but the majority of the remaining species have no consistent measurements for traits such as reproductive output, geographic range size, population sizes, or body size. For Blattodea however, the main focus is not general
correlates between organism traits and genome size, but rather a test of the hypothesis that genome size is constrained by the evolution of sociality. This has been examined previously in insects such as Hymenoptera (Ardila-Garcia et al., 2010) and termites (Koshikawa et al., 2008; Harrison et al., 2018), but it has been difficult to control for confounding factors such as phylogenetic signal or other potential correlates of small genome size. This study expands on the initial analysis by Koshikawa et al. (2008) to provide a much larger comparative dataset of not only termites but a phylogenetically broad sample of cockroaches, and controls for other factors that may constrain genome size.

**Research Questions and Objectives**

This thesis addresses several major research questions relating to genome size diversity in scale insects, cockroaches, and termites:

1. What is the genome size diversity across scale insects (Hemiptera: Coccoidea) and their relatives (Suborder: Sternorrhyncha)? (Chapter 2)

2. How does this diversity compare to the expected ≤2pg threshold found in insects with complete metamorphosis? Is there a difference in genome size between asexual (hemimetabolous) and sexual (holometabolous) scale insect species? (Chapter 2)

3. Is there a link between genome size and scale insect traits such as: geographic range size, polyphagy, body size, chromosome number, reproduction rate, and number of predators? (Chapter 2)

4. What is the genome size diversity across cockroaches and termites (Blattodea)? (Chapter 3)

5. Does citric acid rehydration of preserved blattodean tissues (in ethanol, RNAlater, etc.) increase the viability of these tissues with the FIAD methodology? (Chapter 3)

6. Do blattodean species which have evolved sociality display smaller genome sizes than their solitary counterparts? (Chapter 3)

7. For cockroaches, is there a link between genome size and body size, flight ability, or reproductive type? (Chapter 3)
Following reports on the individual studies in Chapter 2 (scale insects) and Chapter 3 (Blattodea), Chapter 4 provides a discussion of general genome size theory in relation to insect genome evolution, summarizes these research questions and their answers, provides concluding remarks, and highlights important future research directions.
Chapter 2 – An Exploration of Genome Size Diversity in
Scale Insects (Hemiptera: Coccoidea)

Abstract

The hemipteran superfamily Coccoidea (scale insects) does not yet have genome size data available, even though they present some intriguing potential patterns of genome size diversity related to unique aspects of their biology. 62 coccoidean species had their genome sizes estimated using Feulgen Image Analysis Densitometry (FIAD), resulting in genome sizes ranging from 0.15pg-2.1pg (n=62). These data were then correlated with organismal traits related to ecology, morphology, metamorphosis, and reproduction. Genome size of scale insects is found to be most strongly related to body size and geographic range size.

Introduction

Scale Insects

Members of the Hemipteran superfamily Coccoidea, the scale insects, are well known because of their role as significant agricultural pests (Grafton-Cardwell et al., 2001; Miller et al., 2005), with the cottony cushion scale (Icerya purchasi) being one of the first successful targets of ‘biological control’ in the late 1800’s (Van Driesche & Bellows, 1996). Although they appear very simple in morphology, with females being little more than an amorphous and neotenous blob, the biology of scale insects is very complex - many aspects of their ecology and evolutionary history are still unknown (Cook et al., 2002; Gwiazdowski et al., 2011). Their genetic systems include holokinetic chromosomes (Gavrilov, 2007; Ross et al., 2010) and inverse meiosis (Chandra, 1962; Ross et al., 2010), and their reproductive types range from sexual, asexual, hermaphroditic, and facultatively parthenogenetic forms (Ross et al., 2012a). Scale insects exhibit at least ten distinct sex determination mechanisms (Ross et al., 2010), including their own system ('Diaspidid') whereby haploid males are formed through paternal genome elimination (PGE) (Brown, 1965; Nur, 1980). In addition to being
enigmatic by themselves, some scale insects also harbour endosymbiotic bacteria within specialized polyploid cells, named bacteriomes, which have helped shape their evolutionary history (Ross et al., 2012b). Across nearly every level of organismal organization, there is a facet of scale insects that baffles science (Miller & Kosztarab, 1979; Gullan & Kosztarab, 1997).

Their intriguing genetic systems and importance as agricultural pests notwithstanding, scale insects remain poorly studied from a broad genomic perspective. Nothing is currently known about their genome sizes, even though this information would be useful in determining targets for whole-genome sequencing and despite the potential for interesting ecological and developmental correlates with genome size in these insects. The present study therefore represents a first, discovery-based analysis of genome size diversity in this group. These data are used to explore both the range of genome size diversity and to investigate potential correlates of DNA content variability in these unique insects.

Scale Insect Traits in Relation to Genome Size

All scale insects are obligate phytoparasites, but variability in other factors within the context of this ecological lifestyle may be of interest in terms of patterns of genome size diversity. This includes interspecific differences in body size, geographic range size, the degree of polyphagy (i.e., number of plant hosts), number of predators, reproductive rate (generations per year), and reproductive type (asexual, sexual, or both/facultative).

A number of scale insect characteristics are expected to be interrelated under the niche explosion hypothesis proposed by Normark and Johnson (2011), which posits that phytophagous insects with a pathogen-like population structure may undergo a rapid expansion of population size and host range, driven in part by the increasing effectiveness of natural selection in larger populations. If genome size is determined through an interaction between upward genomic pressure and downward organism-level constraints, then these traits can be seen as potential determinants for setting the organism-level constraints (Chapter 1). For example, strong natural selection for an
organismal trait such as body size will (1) set the threshold for cell size limits, which constrains genome size while (2) restricting TE virulence and therefore DNA content expansion.

Reproductive frequency (a measure of developmental rate) and metamorphosis type or absence (a measure of developmental complexity) are potentially linked to genome size through the reliance of these processes on cell sizes and cell division rates (Gregory, 2002). Scale insects are particularly interesting in terms of both developmental rate and complexity. As pests, they might be expected to exhibit rapid development, and there is variability among species in the number of generations per year. With regard to metamorphosis, scale insects are of special interest in that they display both holo- and hemi-metabolous development; holometabolism is a trait that is only displayed in two other non-Holometabola groups (thrips and whiteflies) (Misof et al., 2014; Vea et al., 2016). In most sexual scale insect species the male fully metamorphosizes into a winged adult (holometabolism), while all females display incomplete metamorphosis (hemimetabolism). As such, this provides an opportunity to contrast genome sizes across groups that contain both hemi- and holo-metabolism (sexual species) to ones that only have hemimetabolous females (asexual species). Comparative analyses of metamorphosis are generally done across broader taxonomic units such as orders, but scale insects represent a unique case where patterns and effects regarding metamorphosis can be observed within a single superfamily. Specifically, if the general pattern of constrained genome size among holometabolous insects (Gregory, 2005a) holds within scale insects, then we may expect species with no metamorphosing individuals (i.e., asexual, all-female species with hemimetabolous development) to have larger genomes than those with holometabolous males. At the same time, asexual reproduction may be expected to relate to genome size – although the anticipated direction of the relationship is not obvious, as explained in Chapter 1.

Overall, the scale insect traits outlined above represent either those shared among many eukaryotes that are commonly investigated (i.e., body size and reproduction), or traits that are peculiarities and/or highly prominent in scale insects (i.e., polyphagy, extensive geographic range sizes, and interspecific differences in type
of metamorphosis). The present study represents the first effort to survey genome size diversity in scale insects and to examine possible patterns relating to these parameters.

**Methods**

**Specimen Collection**

The majority of specimens used in this research were provided by a network of international collaborators that was assembled for the present study. Specifically, public contact information of coccidologists found on ScaleNet (a large information database for Coccoidea) (García *et al.*, 2018) was consulted, resulting in eight collaborators from Australia, France, Israel, and the USA. Since male scale insects are rare and short-lived, while many scale insect species are parthenogenetic, only adult female scale insects were available. Although there are sex-related genomic conflicts in the scale insects (such as paternal genome elimination in some males due to internal symbiotic bacteria), these either result in entire ploidy level changes in males only, or do not actually change the abundance of genetic material (i.e., paternal genome is heterochromatized but not removed) (Ross *et al.*, 2010; Ross *et al.*, 2012b). As such, there are assumed to be no intersexual differences in genome sizes aside from post-PGE males, which are not measured.

To facilitate specimen preparation, collaborators were provided with a sampling kit containing a microscope slide box, 25+ SuperFrost™ microscope slides (Fisher Scientific, USA), and a pamphlet with sampling and shipping information, including a hyperlink to an online spreadsheet. The spreadsheet gathered data including taxonomic identification, collector and identifier information, geographic location of the sample with coordinates, body size, host plant, and date of collection. An instructional video was also used to show the basic technique of removing an adult female scale insect from its protective scale, and then using insect pins to full-body smear the specimen across the slide. Species identifications were done by the experts providing the samples, most of whom kept voucher specimens in the event that additional taxonomic confirmation is required.
As one of the objectives was to maximize phylogenetic diversity included in the survey, phylogenetic information from Vea and Grimaldi (2016) was used to guide sampling targets (typically at the family level). R Statistical Software (R Development Core Team, Version 3.2.2) was run within an RStudio shell to highlight the missing taxa between the pool of species sampled during an initial collecting effort and the full-evidence phylogeny. With this information, further waves of sampling were conducted through the global collaboration network, with emphasis being placed on the more rare missing species and/or families. In total, three dedicated sampling waves were completed resulting in a collection of specimens that included 62 scale insect species as well as one whitefly species (Trialeurodes vaporariorum) and two aphid species (Ovatus mentharius and Sipha maydis).

**Feulgen Staining**

Genome size estimates are typically generated through Flow Cytometry, which measures the fluorescence of nuclei suspended in a buffer, or through Feulgen Image Analysis Densitometry (FIAD), which comparatively measures the density of stained nuclei against the density of a standard with a known genome size via microscopy. Due to using fresh full-body specimens, FIAD was the methodology of choice for generating coccoidean genome size estimates. It was also logistically simpler for collaborators to create tissue smears on microscope slides compared to internationally shipping tissue of pest insects for use in Flow Cytometry. Each slide was fixed, hydrolyzed in acid, and stained with Schiff reagent (a combination of basic fuschin and sodium bisulfite) which binds to the aldehydes present in biological tissues. Together, this chain of chemical processes is known as the Feulgen reaction. The specific methodology, which is described fully in Hardie et al. (2002), is as follows:

The entire Feulgen reaction takes three days. Most solutions during this process are created and stored in 1L Pyrex bottles, but when they are actually used during the Feulgen reaction to act on slides (during fixation, acid hydrolysis, or staining with the Schiff reagent), they are placed in 1L glass boats with metal racks (Fisher Scientific, USA) of up to 100 slides at one time. These solutions were always abundant enough so as to fully cover all slides present.
On the first day, the Schiff reagent was freshly prepared by mixing 11g of basic fuschin in 100mL of 1.0N HCl, which was then mixed with a solution containing 10g sodium metabisulfite (Na₂S₂O₅) in 1000mL dH₂O in a 1L glass Pyrex bottle. This 1100mL of Schiff reagent was then stored for 48-72hrs at room temperature in the dark while tightly capped. The bottle was occasionally shaken during this period in order to re-dissolve any precipitate that had settled.

On the second day, the prepared slides were placed in fixative; although the Feulgen reaction has a few different fixatives that can be used, however only MFA was used during this project. MFA is a solution of 85% methanol, 10% formalin, and 5% glacial acetic acid scaled to 1000mL. The 1L glass boat containing the slide racks and MFA were sealed with parafilm for 24 hours.

On the third day, fixing the slides was complete and they were removed from the MFA solution. They were rinsed for 10 minutes under running tepid tap water. After rinsing, slides were placed in a glass boat containing 5.0N HCl, assuring that all slides were fully submerged. This hydrolysis step lasted for 120 minutes. During this time, the previously made Schiff reagent was decolourized by adding 8.25g of activated charcoal to the 1L glass bottle, which was then shaken and magnetically stirred for two minutes. Following this, the solution was filtered through #40 Whatman filter paper into another 1L bottle. This was done using a Buchner funnel attached to a sidearm flask, which was connected to a filtration pump (KNF Neuberger Inc., USA). Once the hydrolysis step ended, the slides were dipped in a solution containing 0.1N HCl in order to prevent the carryover of the stronger 5.0N HCl. This was followed by placing the fixed and hydrolyzed slides within a glass boat containing the decolourized Schiff reagent for 120 minutes. After this time the slides were rinsed for five minutes in each of three glass boats containing a bisulfite rinse as follows: 50mL 10% sodium metabisulfite solution, 50mL 1.0N HCl, 1000mL dH₂O, for a total of 3300mL of solution evenly split across the three boats. A final rinse of running tap water for 10 minutes was followed by three changes of dH₂O for two minutes each, assuring that all excess stain and bisulfite solution was removed. Slides were air-dried and stored in slide boxes as light exposure
is known to reduce the time span in which the slides remain measurable (Dewse & Potter, 1975).

**Genome Size Estimation**

After the Feulgen reaction process, each slide was measured using image analysis densitometry, which is also fully described in Hardie et al. (2002). BioQuant Life Science software (BioQuant Image Analysis Corp., USA) was used alongside a Leica DM2500 microscope and a Retiga EXi digital camera (QImaging, Canada). Slides are measured under oil immersion conditions at 100x objective magnification. For each nucleus, Integrated Optical Density (IOD) was calculated by comparing the background light intensity to the intensity of light transmitted through the stained nuclei, where the difference between the two measurements is equivalent to how much light was absorbed by the nuclei. The IOD value for each specimen was then compared to the IOD value of a control organism with a known genome size; in this study both domestic chicken (*Gallus domesticus*) and rainbow trout (*Oncorhynchus mykiss*) blood smears were used as standards. Chicken was used as the primary standard as most samples fall within a threefold range of its haploid C-value, while rainbow trout was used as a secondary standard to confirm staining reliability and measurement accuracy. From these standards a proportionate haploid genome size is found. An example IOD value conversion is as follows:

$$\frac{\text{Mean IOD (Florida Wax Scale)}}{\text{Mean IOD (Chicken)}} \times \text{C - Value (Chicken)} = \text{C - Value (Florida Wax Scale)}$$

$$\frac{360.3}{1064.3} \times 1.25pg = 0.42pg$$

This can also be converted to mega base pairs, based on the known molecular weight of nucleotides (Doležel et al., 2003). In the above case, using a factor of 1pg=978Mbp, 0.42pg is equivalent to 411Mbp for the Florida wax scale (*Ceroplastes floridensis*).

A minimum of 20 nuclei was measured for most samples, with an average nuclei count of 70. In three species, slides were difficult to measure but 11-12 nuclei were still
able to be counted – these were included as exceptions to the threshold as they had congruent IOD values and the nuclei were in good condition. Nuclei that were undamaged (i.e., regular compaction levels and not lysed; Figure 2E) and properly stained were used for IOD measurements (which are converted to genome size estimates). In cases of possible endopolyploidy, the most consistent nuclei size and most abundant cell types were assumed to be diploid and were used to generate measurements. Although polyploidy is uncommon in insects, and previous studies have therefore assumed most somatic cells are 2C, there are important cases to consider: known polyploid tissues, haploid cells (e.g., sperm), and cells stopped during interphase of the cell cycle.

In no cases were endopolyploid cells found to be more abundant, such as the pentaploid bacteriome of scale insects, compared to general diploid somatic tissues (Figure 2D). Even within measurable, intact, stained nuclei there can be issues that arise concerning cells that have had their cellular processes stopped (due to fixation) mid-cell cycle. In this case, interphase cells in the process of synthesising extra DNA may be tetraploid (and therefore have two complete genomes before cell fission), and show IOD values that are doubled, while the nuclei look nearly identical to the diploid form (Figure 3). In the case of nuclei with IOD values appearing nearly double that of diploid cells, these are assumed to be interphase cells and are ignored completely. A final issue regarding ploidy levels in the scale insects, besides the pentaploid bacteriome, is the presence of sperm. Having used all female scale insects and full body smears, there is often a large amount of sperm found considering females hold onto sperm bundles within their spermathecae in order to regulate fertilization and control sex ratios within their populations (Ben-Dov & Hodgson, 1997; Cook, 2001). For IOD measurements, this can easily be discerned from diploid somatic tissue as the cells are not only much smaller and haploid, but also appear in the string-like form which insect sperm often takes (Figure 2A) (Jamieson, 1987). Therefore, diploid somatic cells are confirmed through their abundance, and also through an elimination process where haploid sperm, mid-division polyploid nuclei, or specialized cells such as bacteriomes are ignored.
Average IOD values, C-values, and their associated descriptive statistics were initially calculated per collaborator/sample in order to observe any significant differences between samples from either different locations or different time periods (e.g., the same species from the USA or Israel, or the same species sampled across two years). This was used to expose any confounding factors and outliers such as population differences (intraspecific variation), preservation issues, misidentification, or cryptic species complexes. In all cases there were no notable differences between samples of the same species, so all IOD values were then collapsed into a single species value and the same descriptive statistics were calculated (Appendix Table 1). These species-level average C-values were used for both species-level and family-level analyses to control for proper degrees of freedom and sampling effort.

Sources of Trait Data

The majority of trait data for scale insects was mined from ScaleNet (García et al., 2018). This includes number of predators, number of plant hosts, geographic range size (number of countries the scale insect has occurred in), reproductive type (sexual, asexual, or both/facultative), reproductive frequency (generations/year), and average adult female body size. In cases where further investigation was needed (i.e., material listed on ScaleNet was lacking), the data for reproductive frequency and type, chromosome number, and average body size was compiled from multiple sources (Brown, 1965; Hodgson, 1994; Kosztarab, 1996; Gavrilov, 2007; Ross et al., 2012a). Species in which reproductive type was unclear, but contained documentation referring to ‘males’, were listed as sexual. ‘Both/facultative’ as a reproductive type required explicit mention of the life cycle of the scale insect, in which populations commonly rotate facultatively between asexual and sexual reproduction.

Phylogenetic and Statistical Analyses

R Statistical Software (R Development Core Team, Version 3.2.2) run within an RStudio shell was used for all statistical tests at a significance level of \( \alpha=0.05 \). A Bonferroni correction was applied post-hoc to this significance threshold but was rejected on the grounds that (1) it drowned any statistical power by being hyper-
conservative due to the amount of comparisons between variables present in the study and (2) most relationships were still significant even at a lower $\alpha$ value of 0.01, while non-significant comparisons had extremely elevated p-values ($p \geq 0.5$). Therefore, altering the significance threshold did not change any statistical results. Phylogenetic analyses were done using either Mesquite v3.2 (Maddison & Maddison, 2018) or the following packages in R: ‘ape’ (Paradis et al., 2004), ‘phytools’ (Revell, 2012), ‘caper’ (Orme et al., 2013), and their various dependencies.

**Data Exploration**

First, data exploration was conducted by checking for outliers, collinearity, and normality of data (Figure 4). Outliers were determined through Cook’s distance, where data points with high leverage or a value of $D > 1$ would distort statistical outcomes for regression analyses (Figure 4D). Two outliers from Eriococcidae were removed this way, as they severely skewed many of the statistical tests due to low sampling sizes (number of measurable nuclei) and having inflated standard error for their genome size values. The underlying distribution of data and its normality was determined through Q-Q plots (Figure 4B), histograms of frequency, and Shapiro-Wilk tests. Raw values of genome size had a distribution which was skewed towards smaller $C$-values and failed a Shapiro-Wilk test, however this was not the data used in statistical analyses; all statistical tests were done using either log-transformed variables or phylogenetically independent contrasts (PICs, described below) of these log-transformed values. Both log-transformed genome size and PICs of genome size passed a Shapiro-Wilk test for normality ($p$-value=0.68 and 0.88, respectively, under a null hypothesis that the distribution is not significantly different from a normal distribution).

A scatterplot matrix (Figure 5) of each pairwise variable comparison was used to explore collinearity between explanatory variables (i.e., all non-genome size variables). For example, the number of plant hosts (polyphagy) and geographic range of scale insects may be important variables to consider when examining patterns in genome size, but these variables are highly correlated (Ross et al., 2012a). Since scale insects are obligate phytophagous parasites, their geographic range is directly related to the range of their host plants as they cannot survive otherwise (Ross et al., 2012a); an
increase in the number of potential hosts is related to a subsequent increase in geographic range size. Ignoring such collinearity leads to inflated standard errors, incorrect multifactorial models, and may skew interpretations of data (Zurr et al., 2009). This is often corrected by only including the correlated variable that explains the most observed variation in the system (X), and removing the other (Y), because the effect of Y is already partially contained within X due to underlying correlation (Zurr et al., 2009). That is, if geographic range size explains 30% of the variation in genome size within a model (e.g., r^2=0.30) while polyphagy explains 15%, only the model containing geographic range size is considered. Phylogenetic analyses were conducted after checking the dataset for collinearity, outliers, and normality.

**Phylogeny**

A composite phylogenetic tree of Coccoidea was produced in Mesquite v3.2 from previously published phylogenetic information (von Dohlen & Moran, 1995; Cook et al., 2002; Gullan & Cook, 2007; Hardy et al., 2008; Anderson et al., 2010; Cryan & Urban, 2012; Hodgson & Hardy, 2013; Vea & Grimaldi, 2016), which was also used to visualize genomic diversity across the superfamily (Figure 8). This composite phylogeny was created by stitching together lineages that matched across multiple phylogenies, resulting in an estimated tree used purely for topology between species and containing no branch lengths. In cases of topological conflict between trees, information from Vea and Grimaldi (2016) was taken to be the best hypothesized tree as theirs was formed using a strong full-evidence (fossil and molecular) approach. This contrasts with Hodgson and Hardy (2013) who had a wider breadth of species, but used morphological traits from extant and fossilized male scale insects. This justification was used in one major family-level conflict between Coccidae, Kesmesidae, and Asterolecaniidae; based on Vea and Grimaldi (2016) Coccidae diverged before Kesmesidae and Asterolecaniidae, who are sister families. Two armoured scale species, *Froggattiella penicillatta* and *Acutaspis agavis*, were omitted from the phylogeny and any phylogenetic models, as no data could be sourced for these species or genera. This composite tree was used to complete three main goals: 1) measure phylogenetic signal for all traits, 2) generate Felsenstein’s (1985) phylogenetically independent contrasts
(PICs), and 3) estimate the coccoidan ancestral genome size through parsimonious reconstruction using Mesquite v3.2.

**Phylogenetic Signal**

Pagel’s $\lambda$ (Pagel, 1999) and Blomberg’s $K$ (Blomberg et al., 2003) were used to measure phylogenetic signal. Both metrics are similar but their interpretations do have important differences; $\lambda$ is a ratio of covariance for a trait among species compared to the covariance expected under Brownian motion (‘BM’), a proxy for random genetic drift (Freckleton et al., 2002; Revell et al., 2008). As such, the expectation is that $\lambda$ will equal 1 under phylogenetic dependence. Conversely, $K$ is a value which measures the spread of covariance for a trait across groups compared to that expected under BM. Put simply, $\lambda$ measures the total amount of covariance in a trait across a phylogeny, while $K$ shows how that covariance is distributed, whether within ($K<1$) or among ($K=1$) clades. Variation within clades is evidence that trait changes are likely to be recent, as sister taxa are less similar than expected if the trait was evolving randomly. Variation being partitioned between clades is evidence that trait changes are deep in the phylogeny, as it is likely that an evolutionary change following deep divergence between major groups is then shared by most members of that group (Blomberg et al., 2003; Revell, 2012).

These signal parameters were estimated using the R package ‘phytools’, where 1,000 simulations of a randomization test are used to generate the $K$-statistic and $\lambda$ is calculated by a likelihood ratio test (Revell, 2012). These functions also provide a p-value to measure whether the phylogenetic signal is significantly different from a phylogeny with a random distribution; in this case the null hypothesis that is tested is whether $\lambda=0$, or if $K$ is different from 1 under simulation (Revell et al., 2007). These analyses were also used to observe if any traits deviated from the default assumptions of the PIC model, specifically that (1) the trait(s) are phylogenetically dependent and (2) that they are evolving by a random non-adaptive process, proxied in the model by BM (Revell et al., 2008). The latter assumption is important, as measures of phylogenetic signal can only provide evidence for or against evolution modelled by random processes. Evidence against the BM model (e.g., sister taxa are less similar than
expected due to chance) is not direct evidence towards natural selection or other adaptive, non-random modes of trait change (Blomberg *et al.*, 2003, Revell *et al.*, 2008).

Finally, employing Pagel’s λ on this data set is beneficial as it is robust to phylogenies lacking branch length data, as is present within the current study; a λ value closer to 1 signals a good fit between data and the tree used, even if the tree is built on taxonomy or has missing data (Freckleton *et al.*, 2002; Münkemüller *et al.*, 2012; Molina-Venegas & Rodriguez, 2017). One of the main factors besides tree quality is species abundance, with smaller trees producing results with less statistical power (Blomberg *et al.*, 2003). In the current study, the species representation (n=63) is well above what is generally considered to be a small tree (approximately n≤20) (Blomberg *et al.*, 2003). Overall, phylogenetic signal is used in conjunction with PICs, which generate a dataset that is independent of phylogenetic effects by only considering trait changes that have happened since the compared species have diverged.

**Phylogenetically Independent Contrasts (PICs)**

Felsenstein’s (1985) phylogenetically independent contrasts (PICs) were run in order to control for the lack of statistical independence across species (due to shared ancestry), using the composite tree stitched together in Mesquite. PICs were calculated, positivized, and forced through the origin using both the PDAP module (Midford *et al.*, 2011) in Mesquite v3.2 while also being calculated in R v3.2.2 using ‘phytools’ and its dependencies, for comparison purposes. A total of six polytomies were present within the topology, with four being species-level and two being genus-level. The PDAP module in Mesquite controlled for these by reducing the degrees of freedom by the number of polytomies, while phytools in R corrected for polytomies by assigning random branch lengths and distances in order to break up the polytomies. No difference was found between the two different correction methods in terms of impact on statistical outcomes.

**Statistical Modelling**

After generating PICs and phylogenetic signal data, relationships between explanatory variables and genome size were modelled. All continuous organism traits
(body size, number of hosts and predators, chromosome number, and reproductive frequency) were log-transformed and had their relationship with log-genome size analyzed using simple linear regression (SLR). ANOVA was used for reproductive type, the sole discrete explanatory variable for both datasets. For discrete variables, controlling for phylogeny is difficult as categorical values cannot have contrast values and do not evolve under BM as continuous variables are assumed to do (Orme et al., 2013). Therefore, the R package ‘caper’ was used to run PGLS (phylogenetic generalized least squares) models on genome size versus reproductive type. A modified script (see Appendix section 7.0) was used to allow the ‘caper’ PGLS model to run with discrete variables containing more than two levels, as the default is that the variables can only be binary; reproductive type required three categories, ‘asexual’, ‘sexual’, and ‘both/facultative’.

Finally, after all models in both datasets across all variables were run, model validation was completed. For linear models, this was done by analyzing residual plots for behaviours such as biases and homo/heteroscedasticity, confirming that a linear model was appropriate for the data used. Both statistical and phylogenetic analyses were considered completed when all data was properly explored, modelled, phylogenetically controlled, and had phylogenetic signal estimates generated.

Results

Haploid Genome Size Estimates

All measurement data are recorded in Appendix Table 1 and plotted on the coccoidean composite tree (Figure 8). Genome size estimates for Coccoidea resulted in a range from 0.15pg-2.1pg and an average of 0.65pg with a standard error of 0.05pg. This mostly falls within the previously recorded range for Hemiptera, 0.18pg-4.65pg (Gregory, 2018a). In total, 62 coccoidean species had their genome sizes estimated, 60 species had phylogenetic representation, and an additional three species acted as the phylogenetic outgroup (two aphids and one whitefly species). These measured species spanned 10 Sternorrhynchan families, with one family (Monophlebidae) representing
Archaeococcoids, seven families representing Neococcoids, and two families acting as outgroups (Aleyrodidae and Aphididae).

**Statistical and Phylogenetic Results**

The results of the regression and PIC analyses are presented in Table 1. Simple linear regression (SLR) models run on the direct correlation dataset resulted in significant relationships between genome size and body size (Figure 9), and genome size and reproductive frequency. The former was a positive relationship ($r^2=0.17$, $p=5.47\times10^{-4}$, $n=60$) while the latter was negative ($r^2=0.11$, $p=0.027$, $n=37$). Both relationships were also analyzed at the family level, to see if the patterns were explained by certain families. For reproductive frequency, all individual families either had no significant relationship between genome size and reproductive frequency ($p\geq0.45$) or lacked statistical power due to restricted sample sizes (Figure 7). Body size was mostly explained by the families Coccidae ($r^2=0.40$, $p=0.03$, $n=10$) and Pseudococcidae ($r^2=0.37$, $p=0.01$, $n=13$), with Diaspididae ($p=0.61$, $n=23$) and Eriococcidae ($p=0.70$, $n=10$) being non-significant, and all other families being under sampled (Figure 6). No relationship was found between genome size and either number of plant hosts ($p=0.47$) or number of predators ($p=0.19$), or between genome size and geographic range ($p=0.09$). Despite a nine-fold difference in chromosome numbers for the sampled scale insects (range: 4 to 36 chromosomes), this variable did not significantly correlate with genome size ($p=0.31$).

Linear models run on the PICs dataset resulted in a significant positive relationship between genome size and body size ($p=1.8\times10^{-4}$, $r^2=0.25$, $n=58$) and a negative relationship between genome size and geographic range size ($p=0.014$, $r^2=0.09$, $n=58$), while the relationship between genome size and reproductive frequency was non-significant ($p=0.2$, $r^2=0.02$, $n=35$). All other relationships were non-significant, including number of chromosomes, number of plant hosts, and number of predators ($p=0.13-0.34$). The ANOVA model for genome size versus reproductive type found no difference in mean genome size between asexual, sexual, and both/facultative scale insect species ($p=0.29$). The PGLS models run to control for phylogeny in discrete variables was also non-significant and rejected for reproductive type ($F=2.6$, $p=0.09$,
n=20). A cross-family ANOVA of average genome size values found significant differences, through a Tukey’s HSD test, between the following pairwise families: Pheonicococcidae-Margarodidae (p=0.03), Margarodidae-Asterolecaniidae (p=0.002), and Coccidae-Asterolecaniidae (p=0.01).

The ancestral reconstruction value for coccoidean genome size was 0.53pg. For phylogenetic signal, values for both Pagel’s λ and Blomberg’s K for continuous variables are given in Table 1. For λ, the null hypothesis is that λ=0. In most cases, λ was equal to 1; that is, most variables show evidence of phylogenetic dependence, or the covariance between species for a trait relies on phylogenetic structure. Otherwise, λ was not significantly different from 0 in two variables, geographic range and number of hosts (λ=0.59, p=0.06 and λ=0.44, p=0.27, respectively). For Blomberg’s K, which has a null hypothesis that K is different from 1, almost all variables accept this null hypothesis; the majority of variables had values of K significantly less than 1, providing evidence that the partitioning of variance for these traits is within clades and not among clades. That is, closely related species are less similar than expected under BM. This was true for all variables except reproductive frequency (K=0.71, p=0.001) and number of chromosomes (K=1.13, p=0.004). For these variables, the null hypothesis is rejected, and K is not significantly different from 1; species are as similar as expected under a model of random character evolution, and the partitioning of variance for these traits is across clades and not within clades.

Discussion

**Genome Size Diversity in Coccoidea**

The present study has added 62 new species from 8 families to the genome size dataset for Coccoidea, which ranges more than 14-fold (0.15pg-2.1pg). This diversity of genome sizes appears to be under genomic constraint(s) as it displays very restricted variation towards smaller values, relative to the known genome sizes for other hemipteran taxa (range=0.18pg-7.01pg) (Gregory, 2018a).
Genome Size and Metamorphosis

The vast majority of scale insect estimates are below the theoretical 2pg threshold for species that possess complete metamorphosis (Gregory, 2002, Figure 1) – only one species of 62 had a genome size of 2.1pg, while the rest fell below 2pg. However, metamorphosis does not appear to be the sole driver of this constraint, as asexual species with hemimetabolous females have equally small genome sizes as sexual species with holometabolous males. If holometabolism is genomically constraining, the removal of this trait in females may lead to the expectation of a subsequent DNA content increase, but this is not observed in the current species dataset. There are several possible explanations for why asexual scale insects still maintain smaller genome sizes. First, parthenogenesis may have evolved too recently for the removal of this constraint to be reflected through an increase in genome size. Parthenogenesis is known to have arisen independently multiple times within Coccoidea, but very little is known otherwise for the timing of evolutionary change in this trait (Nur, 1971; Ross et al., 2010). In addition, the current data shows evidence of phylogenetic signal within the covariance of genome size across the coccoidean phylogeny; closely related species are more likely to resemble each other than distantly related taxa, possibly showing that genome size tends to be maintained through inertia. Second, these smaller genome sizes may be maintained if they are tied to an organism-level trait that is under selection, especially one that is related to cell size or division rates – a plausible example of this would be selection for small body size. Thirdly, the removal of a constraint on genome size does not guarantee an increase in DNA content if mechanisms of DNA gain are weak (e.g., due to previous suppression or deletion of TEs). Finally, the ancestral genome size of Coccoidea is assumed to be small, given an ancestral reconstruction value of 0.53pg, and both Sternorrhynchan outgroups and other hemipterans showing small genome sizes (hemipteran average=1.06pg) (Gregory, 2018a). As such, the observed genome size diversity of Coccoidea is not necessarily reflective of arising only through a genomic constraint from metamorphosis; a smaller ancestral genome size may have accommodated the evolution of holometabolism in sexual species, instead of holometabolism leading to the observed smaller genome sizes. This leads to the question: if a small genome size is both the
ancestral state of Coccoidea and is still possessed by extant scale insects, what is causing this small state to be maintained?

**Genome Size Constraints in Coccoidea**

It is plausible that the maintenance of this restricted C-value range is due to selection on a non-metamorphic trait shared by most of Coccoidea. In the current dataset the body size of adult female scale insects is the most explanatory trait in terms of variation in C-value. Following this, geographic range size is the second most explanatory trait, which is linked to polyphagy through collinearity. Reproductive frequency had a correlative relationship with genome size but this is most likely an artefact of phylogeny as the relationship disappears after controlling for phylogeny; reproductive frequency as a parameter does not appear to have shaped the genome size of extant scale insects. There are two requirements for a trait to operate as a genomic constraint: a relationship between the trait and genome size, and the trait to be under organism-level selection. As such, body size, geographic range size, and polyphagy all fulfil the first criterion, which is failed by reproductive frequency and every other organismal trait considered in the current project; these latter traits are rejected as potential genomic constraints.

For the second criterion, the presence of selection, it is reasonable to expect selection to operate on body size, geographic range size, and polyphagy within scale insects. Signals of selection were not directly measured within these traits, but can be suggested at through other means. Within the phylogenetic signal data, body size is shown to rely upon phylogenetic structure and is predominantly influenced by genetic drift, but both geographic range size and polyphagy are independent of phylogeny and most likely not driven by genetic drift. This is not direct evidence that geographic range size and polyphagy are under selection, but raises them as plausible targets for further investigation, as body size is shown to more likely to be dependent on genetic drift across the phylogeny. Outside of phylogenetic signal, there are multiple reasons to expect selection to be operating on these three traits.
In the current data the relationship between genome size and body size exists across all of Coccoidea, but is absent or weaker in some families and stronger in others. Differences in life history may explain this variation between groups, in addition to showing how body size may determine differential survival and reproduction in scale insects. Members of Coccoidea vary from being completely exposed (e.g., Pseudococcidae, mealybugs), to being under lightly protective scales (e.g., Coccidae), to highly armoured scales (e.g., Diaspididae), to being enclosed within plant tissue in gall-inducing species (e.g., Eriococcidae) (Gullan & Kosztarab, 1997; Gullan et al., 2005; Ross et al., 2010). The relationship between genome size and body size is seen to be strong in the exposed groups, and non-existent in heavily protected groups. Therefore, selection may be constraining the genomes of exposed taxa, but not heavily protected groups. Alternatively, certain cases exist where a lack of relationship between body size and genome size is explained by cell number changes, resulting in a static body size (Fankhauser, 1955; Gregory et al., 2000); differences between these groups are under an assumption of cell number constancy, but this may not be the case if heavily protect groups have simply changed total cell number in response to genome size fluxes. This would allow them to reach their genetically determined structural limits regardless of cell sizes, and therefore genome size would not be constrained. If these structural limits were exceeded, it is possible that the energy being used for somatic growth is a less optimal use compared to contributions towards reproduction (Ernsting et al., 1993; Heino & Kaitala, 1999). However, this is a metric that has not been measured to date. Regardless, the conclusion is that coccidean body size appears to be a valid target for imparting a genomic constraint in this superfamily, but the behaviour of this relationship changes across taxa (mainly at the family-level).

Both geographic range size and polyphagy are defining traits of Coccoidea and have shaped scale insect life history and evolution (Ross et al., 2012a), although they are more akin to general categories of traits than traits themselves (i.e., geographic range size is defined by motility, colonization success, etc). Regardless, it is unsurprising that these general traits appear to be under the influence of more evolutionary forces than predominantly genetic drift, given the absence of phylogenetic structure in these traits. This disconnect between phylogeny and both geographic range
size and polyphagy is most likely explained by evidence that most major Coccidae lineages were established before the diversification of their modern day angiosperm hosts (Vea & Grimaldi, 2016). As such, host diversity (the main determinate for polyphagy and range expansion in phytophagous insects) was determined after most coccoidean lineages originated (Vea & Grimaldi, 2016). Regardless, selection on traits associated with polyphagy and geographic range size is expected under the niche explosion hypothesis (Normark & Johnson, 2011); scale insect survival and reproduction may be reliant and explained by which plant hosts the species is capable of feeding on, across different reproductive modes, while operating across different levels of population sizes (Ross et al., 2010; Ross et al., 2012a). Therefore, traits associated with geographic range size and polyphagy may be operating as potential genomic constraints, given an expectation of selection operating on these traits and an observed connection to genome size in the present dataset.

The current dataset and genomic relationships observed within it show that coccoidean genome size diversity is most likely defined by constraints imparted by multiple traits which vary across taxa. Some groups are constrained by possessing holometabolism (e.g., sexual species with metamorphic males), while others appear to be driven by selection on body size, geographic range size, polyphagy, or a combination of traits. Many potential constraint targets remain unexplored and the current work is by no means exhaustive, but rather sets the foundation for their investigation by completing the first survey of genome size diversity in Coccoidea.

Prospects for Future Research

Two prospects for future research are highlighted within the current work, 1) various viable genome sequencing targets are now available based on genome size and 2) further pattern testing regarding metamorphosis as a genomic constraint should be completed in both whiteflies and thrips. The sequencing of a scale insect genome has not been completed to date. The Miliaris pit scale (Bambusaspis miliaris) has the smallest genome for both Coccoidea and Hemiptera at 0.15pg and is the best target if cost and difficulty of assembly are main concerns. Otherwise, key species such as notable pests (e.g., Icerya purchasi) can be evaluated based on individual research
directions. For metamorphosis, measuring the genome sizes of whiteflies and thrips is a logical next step as they are holometabolous without belonging to Holometabola through convergent evolution. This controls for any phylogenetic effects and dependence associated with belonging to Holometabola. In the current literature, only the Silverleaf whitefly (0.68pg) and four thrip species (range=0.32pg-0.49pg) have been measured to date (Gregory, 2018a), while in the present study a single outgroup whitefly species was estimated at 0.59pg (T. vaporariorum, the Greenhouse Whitefly). Therefore, all currently known whitefly and thrip genome size measurements are well below the expected 2pg threshold for species with complete metamorphosis, but an expansion of this data is needed. Through the current study, the initial survey of coccoidean genome size diversity has been conducted and the foundation for further analyses into their genome evolution has been set.
Chapter 3 – Genome Size Diversity in Cockroaches and Termites (Blattodea): the Impacts of the Evolution of Sociality

Abstract

The insect order Blattodea (cockroaches and termites) possesses a small foundation of genome size estimates (n=24), but data are otherwise lacking. These data are important as Blattodea represents a viable group for the study of the evolution of sociality and its genomic impacts. This foundation of genome sizes was expanded to 70 species through the addition of 46 newly estimated genome sizes using Feulgen Image Analysis Densitometry (FIAD), resulting in a C-value range of 0.58pg-9.58pg (n=70). Both subsocial cockroaches and eusocial termites possess significantly smaller genome sizes compared to solitary cockroaches, suggesting trait(s) associated with evolving sociality are imparting genomic constraints in these social taxa.

Introduction

Cockroaches (Blattodea) and termites (formerly classified in the order Isoptera) represent an order of insects with which many people are familiar due to their widespread geographic distributions and their potential for invading human establishments. The damage caused by termite infestations costs an estimated $40 billion per year (Rust & Su, 2012). However, the majority of cockroaches and termites are not pests; only ~30 of more than 5,000 species of cockroaches and ~183 of over 3,100 termite species are associated with causing economic damage or having pest behaviours and life histories. In general, blattodean species live in diverse environments, from Mount Everest to tropical rainforests, which has driven their species and trait diversity.

For cockroaches, this niche diversity has resulted in substantial variability in a number of characters among cockroach species, including body sizes ranging from a few millimetres in the ant cockroach to the 9.7cm long Megaloblatta longipennis, defensive sprays, camouflage, leaping abilities on par with those of grasshoppers, and
Batesian mimicry of beetles in Pseudophyllodromiinae (Vršanský, 2007; Schmied et al., 2012; Mullins, 2015). Less morphological diversity is present within termites, but this can be attributed to traits associated with living in large colonies that are often considered to operate as a eusocial superorganism. Although most termites do not inhabit human dwellings, they are nonetheless important as their colony functions affect soil composition and nutrient and carbon cycling (Lawton et al., 1996; Tayasu et al., 1997).

The Evolution of Eusociality

Eusociality is a form of sociality defined by the presence of reproductive division across castes (i.e., there are strictly non-reproductive groups) and a focus on cooperative brood care. The evolution of eusociality in insects is associated with traits such as miniaturization of body size, pheromone communication, coordinated nest creation, care of offspring, and colony defense (Andersson, 1984; Bignell et al., 2010). This reinforces how much of termite evolution and life history has been defined by this eusocial behaviour, and as such it is a main focus in the current chapter.

Eusociality is one of the primary factors contributing to the success and ecological impacts of termites. However, similar to how genomic studies of metamorphosis are biased towards vertebrates, studies of sociality are generally focused on Hymenoptera; many blattodean hypotheses remain untested (Andersson, 1984; Nowak et al., 2010). Nonetheless, recent work on the evolution of eusociality has seen an increased focus on Blattodea, including sequencing the genomes of three termite species (Zootermopsis nevadensis, Macrotermes natalensis, and Cryptotermes secundus) and one cockroach species (Blattella germanica) for the first time (Terrapon et al., 2014; Korb et al., 2015; Harrison et al., 2018). These full genome and proteome analyses have determined that changes in gene regulation (for many pathways, notably development) and chemical communication have driven the evolution of eusociality in insects (Harrison et al., 2018). All termites are eusocial, whereas their closest relatives within the Blattodea, the wood roaches (Cryptocerus spp.), display a less complex form of “subsocial” behaviour (Nalepa & Bandi, 2000). In cockroaches, subsociality as a trait
contrasts with eusociality by containing ephemeral cooperative brood care but lacking the rigid division of reproductive and non-reproductive castes. This similarity is the basis for the idea that termites arose from an ancestrally subsocial cockroach and, in strictly cladistics terms, have been considered by some authors as simply a family of eusocial cockroaches; certainly, that termites are nested phylogenetically within the cockroaches has been reinforced by comprehensive molecular phylogenetic studies (Inward et al., 2007; Ware et al., 2008; Legendre et al., 2015). Overall, this creates a testable continuum and framework within Blattodea, contrasting solitary cockroaches (ancestral state for all of Blattodea) with subsocial cockroaches and fully eusocial termites from a genomic perspective. One hypothesis to be tested is that the evolution of sociality has had a drastic impact on the genomic evolution of blattodeans, specifically on genome size.

**Eusociality and Genomics**

Although a genetic and proteomic understanding of differences between social insects is starting to become clear (Korb et al., 2015; Harrison et al., 2018), hypotheses regarding sociality and overall genome size remain relatively unexplored. Genomic differences between social levels in insects (including cockroaches and termites) have mostly been attributed to variation in diversity and abundance of transposable elements (Korb et al., 2015), even between eusocial termites with different genome sizes (Harrison et al., 2018). Therefore, sociality as a genomic constraint may explain inter-taxon differences in genome sizes (e.g., between solitary cockroaches and eusocial termites), while taxa with equal sociality may differ in genome sizes due to intra-genomic forces (e.g. variation in TEs within eusocial termites). Overall, understandings exist for how eusociality may have arisen and differences in intra-genomic forces (e.g., TE abundance and diversity). However, no hypotheses have been directly tested about what aspect(s) of sociality drive the currently observed differences in genome sizes (Koshikawa et al., 2008), or act as a downward organism-level constraint on DNA content.
It has been hypothesized, based on relationships between genome size, cell size, and neural complexity, that eusociality is associated with small genome size because it allows smaller and more numerous neurons in the brain, especially if body size is also reduced (The Honeybee Genome Sequencing Consortium, 2006; Niven & Farris, 2012). It is possible that small genome size is a prerequisite for the evolution of eusociality (and/or evolves along with eusociality), along with other possible traits such as nest-building, parental care, overlapping adult generations, and some form of chromosomal sex determination (Andersson, 1984; Nowak et al. 2010). Genome size may be connected to several of these traits in addition to neural complexity, including parental care and generation time (Roth et al., 1997; Koshikawa et al., 2008).

It is important to outline the differences between analyzing this trait in Hymenoptera versus Blattodea, and why the latter group is so important for increasing our understanding of the evolutionary effects of eusociality. Most importantly, there are several traits found among the Hymenoptera that confound the study of eusociality in the context of genome size constraints. For example, it is thought that the ancestral state in the Hymenoptera is a parasitoid lifestyle, and it has been suggested that this may also act as a constraint on genome size (Gregory, 2001; Johnston et al., 2004; Grimaldi & Engel, 2005). The Hymenoptera are also restricted in how much variability they may exhibit in genome size values as a result of their holometabolic development (Gregory, 2005a). Additionally, some major groups within the Hymenoptera (e.g., ants) are entirely eusocial, meaning that there are no extant solitary counterparts with which to draw comparisons. A study of the Blattodea is not limited in this way: the ancestral state is solitary and free-living, there is the potential for a much larger range in genome size as these insects exhibit incomplete metamorphosis, and it is possible to compare eusocial termites with subsocial and solitary cockroaches.

An initial test of the eusociality hypothesis in the Blattodea was presented by Koshikawa et al. (2008), who reported that termites do exhibit smaller genomes than the few cockroaches that they examined. However, this preliminary analysis omitted two important elements. First, the survey included only a small number of cockroaches, and therefore did not establish that termites are unique among the Blattodea in having small
genomes – that is, it is possible that small genome size is common even among non-social members of the order, which would cast doubt on the importance of eusociality as a major determinant of genome size in this group. Second, this study did not consider other potential constraints on genome size that may be more important, such as flight, reproductive mode, or body size. In addition, an effort was made to test the use of preserved specimens for genome size analysis, as this would greatly expand the breadth of available material for future studies. The present study builds upon the earlier study by Koshikawa et al. (2008) by greatly expanding the diversity of samples among roaches and includes an analysis of additional traits of interest in addition to degree of sociality.

**Methods**

**Sources of Specimens**

The majority of cockroach species collected for genome size analysis were provided by collaborators with expertise in the biology of Blattodea. Previously published literature was used for further cockroach species data, and all termite species data (Koshikawa et al., 2008; Gregory, 2018a; Harrison et al., 2018). Both sexes of cockroaches were sampled as their life histories provide no reason to prefer sampling one sex over another and there are no assumptions of intra-specific genomic variation. For termites, all samples were from diploid workers and soldiers, which are comprised of both sexes, unlike in the haplo-diploid system of the Hymenoptera (Korb et al., 2015). There is also currently no evidence of inter-caste level genome size differences (Koshikawa et al., 2008). Some of the sampling for Blattodea was done from preserved tissues; in this case collaborators shipped their samples in Eppendorf tubes containing a small amount of the preservation medium. Tissues were stored by the collaborator in either RNAlater, 70% ethanol, and/or a -80°C freezer. Slides were then prepared in the lab as opposed to by the collaborator, following the methodology below, under ‘Citric Acid Rehydration of Preserved Tissues’. For these stored tissues, preservation age ranged from <1 week to approximately 5 years. For fresh tissue, samples were smeared by collaborators on SuperFrost™ microscope slides. Fresh tissues sampled included
haemolymph, leg muscle tissue, or brain tissue for cockroaches, and full body smears for termites – a comparative analysis was then done to examine the accuracy of IOD measurements across tissue types.

Following the initial waves of sampling for both cockroaches and termites, a phylogenetic analysis was completed to identify any large missing taxonomic groups (specifically at the family level) within the initial sampling pool. Phylogenetic data was extracted in the form of Newick files from Legendre *et al.* (2015). R Statistical Software (R Development Core Team, Version 3.2.2) was run within an RStudio shell to highlight the missing taxa between the pool of currently sampled species and the full-evidence phylogeny. With this information, further waves of sampling were conducted through the global collaboration network, with emphasis being placed on the more rare missing species and/or families. In total, the final species diversity of Blattodea (n=70) was reached after two sampling waves.

**Citric Acid Rehydration of Preserved Tissues**

To test the applicability of preserved tissues with the Feulgen reaction, an initial test was completed on a set of cockroaches (*Blaptica dubia*) and termites (*Reticulitermes flavipes*) across fresh, ethanol preserved (70%), and RNAlater (Ambion Inc., USA) preserved samples for both males and females. Full-body smears of adult termites were used, while cockroach slides were prepared using leg muscle tissue or brain material. A full-body smear for cockroaches was used in cases where the specimen was too small to gather enough tissue from a specific region. Leg and head tissue was removed from each preserved sample, teased apart with dissection pins as completely as possible directly onto a microscope slide, and then rehydrated with either deionized H2O or 0.35M citric acid. The 0.35M citric acid solution was created using 7.355g of citric acid monohydrate (Fisher Scientific, USA) and 100mL of deionized H2O. Only one or two drops were used, depending on tissue quantity, until all tissue was suspended in solution. While suspended, additional teasing of tissue was done if necessary in order to further suspend tissue and to reduce the amount of large immeasurable pieces of tissue. Slides were allowed to air dry, with no further modifications needed (e.g., a cover slip).
The above methodology was also applied to a 70% EtOH-preserved sample of male and female *B. dubia* cockroaches at 1-week, 1-month, and 1-year preservation timeframes. This was done in order to observe any changes in FIAD success rates and measurement accuracy over a temporal scale on the same specimens. After this citric acid methodology was confirmed to produce measurable stained nuclei from preserved tissues with moderate success, the full amount of cockroach and termite samples received from collaboration (n=189) were put through the FIAD staining and measurement process. These collaboration samples varied in preservation time from 1-month to 5-years, allowing FIAD accuracy to be measured across a full range of 1-week to 5-years of preservation duration when the initial *B. dubia* test was included.

Finally, a second sample of 1-year preserved cockroach tissues from *B. dubia* were tested using the ‘freeze-flip’ method, across both male and female specimens. Leg tissue in 70% EtOH was dissected and teased apart fully onto a microscope slide, at which point either water, 0.35M citric acid, or 40% acetic acid (as per Jeffery & Gregory, 2014) were used as control and treatment groups, respectively. Suspended tissue then had a microscope coverslip temporarily attached to the slide via clothespins and/or binder clips. The slides were then placed within a Styrofoam box containing dry ice for 20 minutes immediately after they were created. After 20 minutes the slides were removed from the dry ice, clothespins and/or binder clips were removed, and the adhered coverslip was quickly sheared off using a razor blade. In total, six freeze-flip slides were created for *B. dubia*, across both sexes and all rehydration sources (water, citric acid and acetic acid).

**Feulgen Staining and Genome Size Estimates**

The methodology for the Feulgen reaction and the generation of genome size estimates is the same as that outlined in Chapter 2, including for those microscope slides created using preserved tissues; all slides, regardless of source tissue or preservation state, undergo the same chemical reactions to stain for DNA content and subsequent densitometry measurement process as in Hardie et al. (2002). C-values for the majority of termites and *Cryptocerus* cockroaches were taken from the dataset generated by Koshikawa *et al.* (2008), which allowed for the most rigorous analysis of
Blattodea to take place once combined with the present dataset. To confirm measurement accuracy, the few species which overlapped between the current dataset and the Animal Genome Size Database were compared and found to not significantly differ, justifying the combination of these datasets into a Blattodea-wide sample.

Finally, additional trait data for cockroach body length, flight capability ("flighted" or "flightless"), and reproductive type ("ovoviviparous" or "oviparous") were taken from cockroach breeding websites.

**Phylogenetic and Statistical Analyses**

R Statistical Software (R Development Core Team, Version 3.2.2) run within an RStudio shell was used for all statistical tests at a significance level of $\alpha=0.05$. Phylogenetic analyses were done using either Mesquite v3.2 (Maddison and Maddison, 2017) or the following packages in R: ‘ape’ (Paradis et al., 2004), ‘phytools’ (Revell, 2012), ‘caper’ (Orme et al., 2013), and their various dependencies. Outliers, collinearity, and normality of data were initially explored using the same previous methods (Figure 4); outliers were determined through Cook’s distance, collinearity was explored through pairwise variable correlations and a scatterplot matrix, while normality was determined through Q-Q plots, histograms of frequency, and Shapiro-Wilk tests.

A composite phylogeny of Blattodea was produced from previously published phylogenetic information (Djernaes et al., 2012; Legendre et al., 2015; Wang et al., 2017) (Figure 9) in order to visualize genomic changes evolutionarily and to control for phylogenetic effects within statistical models. This composite phylogeny was created by stitching together lineages that matched across multiple phylogenies, resulting in an estimated tree used purely for topology between species and containing no branch lengths. Two cockroach species, *Cariblatta minima* and *Hemiblabera tenebricosa*, were omitted from all phylogenetic models as no phylogenetic data could be sourced for these species or their genera. Using this composite phylogeny, Felsenstein’s (1985) phylogenetically independent contrasts (PICs) were run in order to control for the lack of statistical independence across species. PICs were calculated, positivized, and forced through the origin using both the PDAP module (Midford et al., 2011) in Mesquite v3.2
while also being calculated in R v3.2.2 using ‘phytools’ and its dependencies, for comparison purposes. A total of ten polytomies were present within the topology, with seven being species-level and three being genus-level. The PDAP module in Mesquite controlled for these by reducing the degrees of freedom by the number of polytomies, while phytools in R corrected for polytomies by assigning random branch lengths and distances in order to break up the polytomies. No difference was found between the two different correction methods in terms of impact on statistical outcomes.

Cockroach body size data were log-transformed prior to comparison with genome size via simple linear regression (SLR) (Figure 9). Phylogenetic contrasts of body size and genome size were then generated in R, and a separate phylogenetic model was run on this phylogenetically-controlled dataset using SLR. ANOVA was used for comparing genome size against discrete variables in the study (flight capability and reproductive type). To control for the effects of phylogeny in these variables, a PGLS model was used in R, under the package ‘caper’, with an edited script allowing for ANOVA models to be run on variables with greater than two state levels (see appendices 7.0).

Both Pagel’s λ (Pagel, 1999) and Blomberg’s K (Blomberg et al., 2003) were used to measure phylogenetic signal in both body size and genome size against the phylogenetic data. This was done using the R package ‘phytools’, where 1,000 simulations of a randomization test are used to generate the K-statistic and λ is calculated by a likelihood ratio test (Revell, 2012). These functions also provide a p-value to measure whether the phylogenetic signal is significantly different from a phylogeny with a random distribution (Jeffery et al., 2016); in this case the null hypothesis that is tested is whether λ=0, or K is different from 1 under simulation (Revell et al., 2007).

Finally, after correcting both continuous and discrete variables for statistical independence and running their applicable models, model validation was completed. For linear models, this was done by analyzing residual plots for behaviours such as biases and homo/heteroscedasticity, confirming that a linear model was appropriate for the data used.
Results

Timeline of Genome Size Estimates from Preserved Tissues

Genome size estimates derived from samples preserved with RNAlater were found to be extremely underestimated (approximately 20% lower than the known genome size for *B. dubia*) and also unreliable to measure. That is, nuclei in good condition were rare and the majority were lysed or overall in an abnormal state, leading to decreased measurements available to generate estimates from. As such, only estimates from 70% EtOH were used to create the timeline of FIAD measurement accuracy for preserved tissues. There were also no significant differences in genome sizes across sexes or tissue types (haemolymph, brain tissue, or leg tissue) within the same species (ANOVA, p>0.05), so all *B. dubia* C-value averages were collapsed into one value per time category (fresh, 1-week, 1-month, or 1-year preserved). Average C-value from fresh tissue was 4.65pg, which is comparable to the known range of 4.5-4.7pg for *B. dubia* in the Animal Genome Size Database (Gregory, 2018a). From this average, estimated C-value was seen to decrease to 4.40pg (1-week), 3.58pg (1-month), and 3.65pg (1-year) (Figure 10). 1-year EtOH samples that were freeze-flipped using acetic acid had an average genome size of 4.15pg, freeze-flipping with citric acid resulted in an average of 4.05pg, while freezing-flipping with water (a control measure) resulted in an average of 3.51pg. Finally, FIAD was found to fail for the test termite (*R. flavipes*) sample when using preserved tissues, so no termite genome size data were generated using preserved samples.

Haploid Genome Size Estimates

46 new blattodean species had their genome sizes estimated. Including published data, blattodean estimates from a total of 70 species now range in size from 0.58pg-9.58pg with a mean of 3.2pg and a standard error of 0.05pg (Appendix Table 2). These are shown mapped on a phylogeny in Figure 11.
Statistical and Phylogenetic Results

The results of the regression and PIC analyses are presented in Table 2. No significant relationship was found between log genome size and log body size across cockroaches \((p=0.97, r^2=0.02, n=44)\) including after controlling for phylogeny using PICs \((p=0.59, r^2=0.02, n=42)\). Three ANOVA models were run on genome size versus flight capability, reproductive type, and sociality type, and were followed by a phylogenetically-controlled PGLS model for each variable. Both ANOVA and PGLS models were non-significant for flight capability \((\text{ANOVA, } p=0.427, \text{ PGLS, } p=0.13)\) and reproductive type \((\text{ANOVA, } p=0.732, \text{ PGLS, } p=0.15)\). For sociality type, both the ANOVA and PGLS models were highly significant \((\text{ANOVA, } p=9.9\times10^{-4}, \text{ PGLS, } p=9.0\times10^{-4})\). A post-hoc Tukey’s HSD revealed that solitary cockroaches have significantly larger genome sizes than both subsocial cockroaches and eusocial termites \((p\leq9.8\times10^{-4})\).

For phylogenetic signal, values for both Pagel’s \(\lambda\) and Blomberg’s \(K\) for continuous variables are given in Table 2. For \(\lambda\), the null hypothesis is that \(\lambda=0\). For both body size and genome size \(\lambda\) was significantly different from 0 \((p\leq1.56\times10^{-6}, \lambda=1)\); that is, these variables show evidence of phylogenetic dependence, or that the covariance between species for these traits relies on phylogenetic structure. For Blomberg’s \(K\), which has a null hypothesis that \(K\) is different from 1, genome size accepts this null hypothesis while body size rejects it. That is, genome size had a \(K\) value significantly less than 1 \((p=0.064)\), providing evidence that the partitioning of covariance for this trait is within clades and not among clades (i.e. closely related species are less similar than expected under Brownian motion). For body size, the null hypothesis is rejected and \(K\) is not significantly different from 1 \((p=0.002)\); species are as similar as expected under a model of Brownian motion evolution.

Discussion

Genome Size Diversity in Blattodea

The present study has added 46 new species from 4 families to the genome size dataset for Blattodea, which now ranges more than 16-fold \((0.58\text{pg}-9.58\text{pg})\) (Gregory,
These data were solely from fresh samples, as preserved tissues continue to conflict with the FIAD methodology (Figure 10). Regardless, this genome size range greatly exceeds the theoretical 2pg threshold for genome size expected for groups with complete metamorphosis (Gregory, 2002); as a hemimetabolic group (Bignell et al., 2010), Blattodea is suitable for comparisons that could otherwise be confounded by metamorphic constraints. Traditional study groups of sociality (e.g., Hymenoptera) are not free of this confounding factor, in addition to factors such as parasitism (a trait also associated with small genome sizes), small ancestral genome sizes, and a lack of proper solitary comparative groups in some taxa (e.g., ants) (Gregory, 2001; Johnston et al., 2004; Grimaldi & Engel, 2005). For Blattodea, genome sizes greater than 2pg belong only to families of cockroaches, with all termites having genome size estimates below 2pg (0.58pg-1.90pg, average=1.25pg; Figure 11). This implies that some additional constraints are acting in termites that are not related to metamorphosis. That no other organismal traits analyzed in the current study (flight ability or reproductive type) significantly relate to genome size suggests that they are not major constraints either. This reinforces the idea that the marked difference in genome sizes between most cockroaches and termites is due to a constraint related to the evolution of eusociality in termites, and subsociality in one family of cockroaches (Cryptoceridae spp.) (Koshikawa et al., 2008; Appendix Table 2). Furthermore, testing for phylogenetic signal in genome size across the phylogeny shows that covariance in this trait relies on phylogenetic structure (i.e., λ=1); covariance in genome size is similar within termites and within cockroaches but differs between these groups. Rather than being typical of most of Blattodea, the small genomes of both subsocial cockroaches and eusocial termites appear to have resulted from a secondary decrease in DNA content from a larger ancestral value (known to be multiple Gbs in size; Koshikawa et al., 2008).

**Genome Size and Sociality**

Sociality per se is not a constraining factor, but rather represents a suite of traits, several of which may act to constrain genome size. Socially-related traits include brood care, overlapping generations, reproductive division of labour, changes in development,
miniaturization of body size, and neural complexity (Koshikawa et al., 2008). Some of these features, such as brood care, overlapping generations, and reproductive division of labour, are unlikely to relate causally to genome size as they are not clearly associated with cell size or cell division rate. Developmental differences between cockroaches and termites are also unlikely to explain the genomic differences between these groups. In termites, workers have slower developmental rates and reduced developmental complexity compared to cockroaches, with this caste essentially being neotenic (i.e., workers are immature and may remain in this state for a prolonged period); neoteny is generally associated with larger, not smaller, genomes (Gregory, 2002). As such, termites display constrained genomes in light of possessing both hemimetabolism and reduced developmental rate and complexity, traits which have expectations of larger and not smaller genomes. Therefore, of the core traits associated with sociality, miniaturization of body size and neural complexity remain as plausible candidates for generating a constraint on genome size in termites, as both can be affected by cell size and/or cell division rate.

In the current dataset body size was unrelated to genome size across cockroaches, which may suggest that this has not been a strong determinant of genome size variability among Blattodea as a whole. However, body size data were only available for solitary cockroaches and it remains to be seen if a relationship exists within both subsocial cockroaches and the termites. The role of body size as a driver of genomic evolution within these social taxa is contingent on investigating this relationship, which should be elucidated as it has been in multiple other invertebrate taxa (McLaren et al., 1988; Soldi et al., 1994; Finston et al., 1995; Gregory et al., 2000; Gregory, 2005a).

Neural complexity is a second viable option for acting as a constraint on genome size. This arises through organismal selection for smaller, more densely packed neural cells in the brain, leading to a constraint on DNA content in order to maintain levels of neural complexity (in terms of behaviour, communication, and overall neural function) (The Honeybee Genome Sequencing Consortium, 2006; Niven & Farris, 2012). Importantly, this is not mutually exclusive from body size acting as a constraint, and
indeed selection for smaller cells be intensified when body size is reduced, both as a mechanism of body size reduction and because smaller bodies can accommodate fewer neurons than larger ones without a change in cell size. In other words, the strength of selection for small genome size may be especially severe if the evolution of sociality involves both an increase in neural complexity and a reduction in body size. The relationship between neural complexity and genome size has been demonstrated in aves (Andrews & Gregory, 2009; Gregory, 2018b), is thought to occur in Hymenoptera (The Honeybee Genome Sequencing Consortium, 2006), and the opposite phenomena (an expansion of DNA content leading to reduced neural complexity) has likely occurred in salamanders (Roth et al., 1997). For future research, measurement of neural complexity in insects needs to be completed as a direct test of the hypothesis that the genomes of social taxa are constrained due to maintenance of neural function. This could include measures of brain cell density per unit of tissue, as has been done in birds (Gregory, 2018b), or in terms of the degree of neural tissue differentiation, as reported for amphibians (Roth et al., 1997). This would include contrasting termite and subsocial cockroach neural density to that of solitary cockroaches, to observe if overall density of neural cells has increased and/or neural cell size has been reduced in these social groups. Overall, of all the organismal traits within the concept of sociality, both miniaturization of body size and maintenance of neural complexity are seen as the two most likely sources of genomic constraint given previously observed patterns between these traits and genome size in multiple taxa.

**Mechanisms of Genome Size Change**

A final question in the current project is the underlying mechanism(s) responsible for how subsocial cockroaches and eusocial termites evolved smaller genomes than solitary cockroaches. However, not only are small genomes observed in social groups, but there is also a sizeable rift in genome sizes across Blattodea as solitary cockroaches also display large, unconstrained C-values relative to the ancestral state (Koshikawa et al., 2008). This is shown through the 16-fold difference in genome sizes across Blattodea and the presence of genome sizes upwards of 9.58pg in the solitary Big Black Beetle Mimic cockroach (*Ergaula pilosa*). No intragenomic mechanisms that
cause DNA fluxes are expected to vary between cockroaches and termites in terms of presence or absence of entire mechanisms (e.g., as opposed to a lack of recombination when comparing asexual and sexual groups). Rather, the genome size diversity of Blattodea may be reflective of shared genetic mechanisms operating differentially and independently since the divergence of their lineages.

Although cockroaches and termites share an ancestral genome size and initial set of transposable elements, all lineages have since undergone independent genomic evolution following divergence. It is possible for termites and subsocial cockroaches to have expunged more TEs and resisted increasing DNA content from upward mutation pressure, while solitary cockroaches may have experienced relatively unconstrained DNA content expansions. The role of TEs and repetitive DNA in creating these genomic differences between cockroaches and termites is supported through analyses of the sequenced blattodean genomes available to date (Korb et al., 2015; Harrison et al., 2018). A comparison of two termite genomes (0.54pg and 1.34pg) showed this difference to be nearly entirely explained by non-coding sequences (Korb et al., 2015). Reinforcing this termite analysis, a comparison across Blattodea (i.e., the inclusion of a cockroach genome sequence) found the same results - genome size differences between groups are often explained by repetitive DNA and non-coding sequences (Harrison et al., 2018). For example, the 0.54pg genome size of the termite Z. nevadensis contains 28% repetitive content, compared to the termites M. natalensis (1.34pg, 46% content) and C. secundus (1.33pg, 55% content), and the solitary cockroach B. germanica (2.0pg, 55% content) (Harrison et al., 2018); although restricted by the number of available full sequences, the current data show a strong positive association between repetitive sequences, TE abundance, and DNA content.

In tandem with differences in TE diversity and abundance, it is also possible that intense downward constraints on DNA content in social groups have led to neutral and nearly neutral deletion mutations to be maintained by selection. This would compare to solitary cockroaches, whose genomes may expand if the upward genomic drive from TEs is unconstrained. The overall summation of intense constraint(s) in social groups
with differences in responses to TE activity across Blattodea have most likely lead to the observed genome size diversity in this group.

**Prospects for Future Research**

A main direction for future research is the sequencing of a subsocial cockroach genome, in to the measurement of social blattodean body sizes and insect neural complexity noted above. The currently sequenced blattodean genomes represent three eusocial termites and one solitary cockroach (Terrapon *et al*., 2014; Korb *et al*., 2015; Harrison *et al*., 2018). Therefore, the sequencing of a subsocial cockroach genome would be of interest in order to complete a full genomic sequence for each blattodean social class (solitary, subsocial, and eusocial), effectively expanding the analyses of these genomes from a binary approach (solitary versus eusocial) to a more robust and wider spectrum. Since the cost and difficulty of sequence assembly are key factors when deciding on a species target (Gregory, 2005b; Jeffery & Gregory, 2014), *Cryptocercus kyebangensis* is the best target as it has the smallest non-termite genome size in the current sample (1.16pg) and is subsocial.

In conclusion, the current work has expanded our knowledge of insect genomic evolution through a more phylogenetically expansive view of genome size diversity in Blattodea. Multiple future research directions have been highlighted, and overall this insect order is seen to be a great candidate for testing hypotheses regarding insect genomics and the evolution of sociality.
Chapter 4 - General Discussion and Concluding Remarks

The purpose of the current chapter is to revisit the research questions initially outlined in Chapter 1 and to discuss insect-wide genomic evolution through the two specific insect groups analyzed in the present study.

Review of Major Research Questions

As noted in Chapter 1, the one million known insect species have 1,345 genome size measurements within the Animal Genome Size Database, while the 66,000 known vertebrate species have 5,191 (Gregory, 2018a). Therefore, the first goal of the present work was to increase the number of available insect genome size estimates, and did so in two important insect groups – Coccoidea (scale insects), Blattodea (cockroaches and termites), and their relatives.

Research question 1: What is the genome size diversity across scale insects (Hemiptera: Coccoidea) and their relatives (Suborder: Sternorrhyncha)?

Scale insect genome size was found to range from 0.15pg-2.1pg with an average of 0.65pg. The other sternorrhynchan species used as phylogenetic outgroups showed similar diversity, ranging from 0.43pg-0.59pg with an average of 0.49pg. Previously, 75 Hemipteran species had their genome sizes estimated, and these measurements were biased towards aphids (Aphididae) and assassin bugs (Reduviidae) (Gregory, 2018a). The current work expands these measurements from 75 to 140 species by adding 65 new sternorrhynchan species. This does not solve the bias in hemipteran genome size measurements (indeed, it simply adds scale insects to the bias), but rather highlights the need to continue sampling the extraordinary diversity of insects available.

Research question 2: How does this diversity compare to the expected ≤2pg threshold found in insects with complete metamorphosis? Is there a difference in genome size between asexual (hemimetabolous) and sexual (holometabolous) scale insect species?

This diversity matches the expectation that taxa with complete metamorphosis will possess constrained genome sizes. Essentially all C-values within Coccoidea fall
under the theoretical 2pg threshold, with one species having an estimated C-value of 2.1pg. However, no difference in genome sizes was observed between asexual (hemimetabolous) and sexual (holometabolous) scale insects. That is, the removal of this constraint in hemimetabolous females does not appear to have subsequently resulted in an increased genome size for these species. Explanations for this include phylogenetic inertia, the presence of an equally strong alternative constraint, or that the coccoidean ancestral genome size state is small to begin with, among others.

**Research question 3: Is there a link between genome size and scale insect traits such as: geographic range size, polyphagy, body size, chromosome number, reproduction rate, and number of predators?**

Relationships were found between genome size and body size, geographic range size and (solely through collinearity with geographic range size) polyphagy. Of these, body size was the most explanatory in terms of genome size variation, followed by geographic range size and polyphagy.

**Research question 4: What is the genome size diversity across cockroaches and termites (Blattodea)?**

The genome size dataset for Blattodean has been expanded from 28 to 74 species by adding new estimates for 46 cockroach species. The previous range of these estimates was 0.58pg-5.15pg and has been updated to 0.58pg-9.58pg with an average of 3.2pg; the upper range of this range has almost doubled in size, with eight cockroach species being above the previous highest value of 5.15pg. Within Blattodea, termites had a range of 0.58pg-1.90pg with an average of 1.25pg while cockroaches had a range of 1.16pg-9.58pg with an average of 3.7pg (Figure 11).

**Research question 5: Does citric acid rehydration of preserved blattodean tissues (in ethanol, RINAlater, etc.) increase the viability of these tissues with the FIAD methodology?**

The rehydration methodology used in the present study does not appear to reliably increase the compatibility of preserved insect tissues with the FIAD process.
Very temporary storage of tissues (≤1 week) in 70% EtOH or freeze-flipping tissues with acetic acid are the two options which resulted in C-value estimates closest to the known fresh value for the cockroach *B. dubia*. However, 1) these findings are only reliable for *B. dubia* and other insect taxa and tissues need to be individually tested, while 2) these estimates also had unreliable levels of variation regardless of how similar their mean C-values were found to be compared to the fresh value for this species.

**Research question 6: Do blattodean species which have evolved sociality display smaller genome sizes than their solitary counterparts?**

In the current dataset it does appear that some aspect of evolving sociality is associated with possessing smaller genome sizes; eusocial termites and subsocial cockroaches predominantly show much smaller genome sizes than solitary cockroaches (Figure 11). The exact suite of trait(s) that are under selection and imparting a genomic constraint on these social taxa remains to be elucidated, with both miniaturization of body size and neural complexity being plausible targets.

**Research question 7: For cockroaches, is there a link between genome size and body size, flight ability, or reproductive type?**

No relationships between genome size and body size, flight ability, or reproductive type were observed in the current dataset. However, uncovering a relationship between genome size and termite body size remains to be completed, as only solitary cockroach body sizes were able to be used to form the model which analyzed the relationship between genome size and blattodean body size.

**Similarities and Differences between Coccoidea and Blattodea**

Relationships between organismal traits and genome size are of interest in genomic evolution as selection on these organismal traits may constrain genome size fluxes if they are linked to cellular parameters (namely cell size and division rates) (Gregory & Hebert, 1999). This helps reach a stronger explanatory understanding of genome size diversity as opposed to descriptive or predictive approaches (Lynch, 2007). An aspect of these genomic patterns is to observe if they are maintained or vary
across different taxa, illuminating shared effects arising from similar phenomena operating across the tree of life. For genome size, these patterns often deal with traits that fall into four categories of characters which rely upon cellular parameters – body size, reproduction, development, and metabolism (Gregory, 2005a). Contrasting Coccoidea and Blattodea can therefore show how these patterns behave on an insect-wide level.

**Body Size:** A relationship between genome size and body size was observed within Coccoidea but not Blattodea. Additionally, the relationship was stronger in some coccoidean families while it was much weaker or non-existent in others. Variation in these results, including within closely related taxa (i.e., within Coccoidea), can be explained by two important factors (Gregory et al., 2000) – the amount of genome size variation to explain, and taxa-specific responses to changes in DNA content. For any trait to be explanatory (not only body size) there needs to be a sizeable amount of variation within genome size to be explained. In Coccoidea, a range of 0.15pg-2.1pg is relatively constrained compared to the range of Blattodea which is much larger at 0.58pg-9.58pg; and yet, no relationship was found within Blattodea even given more variation in genome size for body size to explain. Therefore, a lack of relationship in some coccoidean families is attributed to both low sample sizes at the family level and a low amount of genome size variation to explain. For Blattodea, the lack of relationship between genome size and body size may be explained by either differences in cell number or responses in cell size to DNA content changes. Since the main linkage between genome size and body size is through both cell size and cell number, changes in these parameters may cause the relationship to not exist. That is, cell size responds to DNA fluxes differently across taxa (Cavalier-Smith, 1982) and Blattodea may display a low response in cell size to genomic changes. Secondly, a change in cell size due to DNA content fluctuation may be counteracted by a subsequent change in cell number in the opposing direction, causing body size to remain static and no relationship between genome size and body size to be observed (Fankhauser, 1955; Gregory et al., 2000). These parameters (cell number and size) should be directly tested in groups with genome size data in order to continue investigating the relationship between insect genomics and morphology.
**Reproduction and Metabolism:** In both Coccoidea and Blattodea no reproductive or metabolic parameters had significant relationships with genome size, removing them as potential constraints or explanatory factors for the observed genome size diversity in the current dataset. However, future research can target these categories for insects in general through either a phylogenetic breadth increase (i.e., inclusion of different taxa) or by targeting more diverse parameters. For reproduction these parameters may include fecundity, egg size, or intrinsic rate of increase, among others. Studies of metabolism are often defined by extreme examples of metabolic demand, such as flight in endothermic animals (Wright *et al*., 2014). Although studies on insect flight and metabolism are well documented (Nutting, 1969; Chino *et al*., 1992), studies approaching this category of traits from a genomic perspective are rare for insects, with only the locust having been studied (Wang *et al*., 2014). Therefore, it is of interest to pursue a genome-metabolism linkage in insects through the use of insect-relevant traits that mirror wing loading and muscle abundance used in vertebrate studies (Wright *et al*., 2014).

**Development:** Metamorphosis is seen to potentially operate as a genomic constraint within Coccoidea (0.15pg-2.1pg; complete metamorphosis) but most likely not within Blattodea (0.58pg-9.58pg; incomplete metamorphosis). As such, both taxa analyzed in the current work contribute to our understanding of insect metamorphosis and genome size, whether directly as a group that contains holometabolism and constrained genome sizes (Coccoidea) or by providing a contrast group that has an expectation of increased diversity in genome size (Blattodea). Both of these expectations have been met in the current data, providing further evidence towards the idea that selection on metamorphosis and developmental parameters may drive insect genome size evolution (Figure 1).

As mentioned, neural complexity is a viable target for imparting a genomic constraint in social blattodean taxa, explaining the disparity in genome sizes between social and non-social groups. This trait deals with developmental complexity, where organ complexity is under selection to maintain function while under space (i.e., brain case) limitations, especially in light of miniaturization of body size. This represents a
future research direction that can increase our understanding of the evolution of sociality from pattern observing to reaching a more explanatory framework.

**Non-Cellular Linked Traits:** This final category contains traits which have no direct expectation of relating to cell size or division rates, and therefore genome size. These traits tend to be more ecological or intragenomic (without impacting DNA abundance), yet in some cases there are observed correlations between genome size and these traits even if the underlying connection is unclear. In the present work, traits such as geographic range size and polyphagy in Coccoidea are contained in this category. Since the connection between these traits and genome size is unclear, this leads to questions about which aspect(s) of organismal biology are actually under selection and causing the observed variation in genome sizes across taxa. Furthermore, traits in this category are often taxa-specific and are traits that are prominent or defining of the targeted group, so it is difficult to draw general conclusions for ‘non-cellular linked traits’ across all insects. As more taxa are investigated and genome sizes are estimated, it is possible that shared non-cellular correlations across eukaryotes will surface and become better understood. For now, the relationships are explored but not fully explained.

**Directions for Future Research**

**Coccoidea:** A logical next step in the study of insect metamorphosis and genomics is the generation of genome size estimates for both whiteflies and thrips. Together with Coccoidea, these three groups represent the known insect groups who have convergently evolved complete metamorphosis outside of Holometabola (Misof et al., 2014; Vea et al., 2016). As such, observing the same genomic constraint (typically ≤2pg) in these groups will provide strong evidence that the observed small genome sizes are due to selection on development and metamorphosis and not due to phylogenetic consequences (e.g., inertia) of belonging to Holometabola. In terms of Coccoidea genomics specifically, the phylogenetic diversity present in the current study is sufficient for setting an initial survey of diversity in this group; however, future measurements could take place in the missing families (8 of 33 coccid families are
measured in the current study; García *et al.*, 2018), while Archaeococcidae is a major scale insect lineage that is currently only represented by two species.

**Blattodea:** A further understanding of the genomic impacts of the evolution of sociality can be gained from two approaches. First, other eusocial groups can be sampled to help further confirm the pattern of genome size restrictions in social clades across a more broad phylogenetic representation of invertebrates, expanding outside of **Insecta.** Within **Insecta** the remaining known eusocial clades include aphids (Stern & Foster, 1996) and thrips (Crespi, 1992), while within invertebrates eusociality has been observed in some species of parasitic shrimp (Duffy, 1996). This phylogenetic breadth would help increase the potential observation of this genomic pattern outside of **Hymenoptera** and **Blattodea,** where insect sociality research has predominately taken place. Secondly, within **Blattodea** explicit hypothesis testing through the measurement of termite body sizes and neural complexity are logical next steps, expanding this pattern observation to a more explanatory understanding. Neural complexity can be measured similarly as has been done in aves (Andrews & Gregory, 2009; Gregory, 2018b), where brain size and density of neurons per unit of tissue can be measured through flow cytometry or densitometry methodologies. This would need to be completed in solitary cockroaches, subsocial cockroaches, and eusocial termites to form an adequate spectrum to compare and contrast across. Following this, the measurement of termite body sizes would assist in confirming whether or not this parameter is a plausible driver of genome size reduction in termites, as the relationship is otherwise lacking across all of **Blattodea** (i.e., solitary cockroaches). This can also be combined with neural complexity data as these two traits may be operating in tandem if miniaturization of body sizes amplifies selection for smaller cells (including neurons). Phylogenetically, the genome size diversity of **Blattodea** can be further elucidated by estimating C-values in two distinct families: the cave cockroaches (**Nocticolidae**) and a further exploration of subsocial **Cryptoceridae** genome sizes. **Nocticolidae** is the last unmeasured cockroach family in terms of genome size estimates, while **Cryptoceridae** (which is an important component of the sociality spectrum) has two species estimates to date.
Insects: Of the main trait categories with known genomic correlates (body size, development, reproduction, and metabolism), metabolism remains weakly investigated in insects. Similar to measurements of neural complexity, this category can be approached through similar directions as research done in birds (Wright et al., 2014). However, this would require insect parallels of flight parameters used in avian research (e.g., wing loading, pectoral muscle abundance, etc.), and measurements of insect metabolic rate through oxygen intake, resource acquisition, and CO$_2$ exhalation. Phylogenetically, the vast majority of insects remain under sampled in terms of genome size estimates, and our knowledge of insect genome sizes would benefit from a large sweeping survey of more insect diversity and representation as opposed to individually targeted groups.

Concluding Remarks

The present thesis has completed both a discovery-based and hypothesis-testing approach to insect genomic evolution. In both cases our understanding of insect genome sizes has been increased, whether for the first time to date (62 coccoidean species) or through expanding previous foundations (70 blattodean species total). Multiple genomic-phenotypic correlations have been illuminated between genome size and organismal traits, while established patterns in insect metamorphosis and sociality have been further confirmed. From this data, optimum genome sequencing targets have been identified for both Coccoidea and Blattodea through abundance of DNA, while future research should pursue a deeper understanding of underlying mechanisms responsible for the observed relationships.
References


http://CRAN.R-project.org/package=caper


Vea, IM, Grimaldi, DA (2016) Putting scales into evolutionary time: the divergence of major
scale insect lineages (Hemiptera) predates the radiation of modern angiosperm hosts. Nat
Sci Rep, 6, 23487.

juvenile hormone variations in scale insect extreme sexual dimorphism. PLoS One, 11,
doi: e0149459.


Vršanský, P (2007) Jumping cockroaches (Blattaria, Skokidae fam. n.) from the late Jurassic of

Wang, W, et al. (2014) The locust genome provides insight into swarm formation and long-

support for interfamilial relationships and major clades. Sci Rep, 7, 3903.

Ware, JL, Litman, J, Klass, K-D, Spearman, LA (2008) Relationships among the major lineages
of Dictyoptera: the effects of outgroup selection on dictyopteran tree topology. Syst
Entomol, 33, 429-450.

Werner, EE (1986) Amphibian metamorphosis: growth rate, predation risk, and the optimal size

Wright, NA, Gregory, TR, Witt, CC (2014) Metabolic ‘engines’ of flight drive genome size

monitoring western tarnished plant bugs. Entomol Exp Appl, 113, 117-123.

Tables and Figures

Figure 1 – From Gregory (2005a). Graphical distribution of C-values across insect groups with different developmental programmes. Holometabolous groups, which are above the horizontal line, appear to be genomically constrained and generally fall below the 2pg threshold. Hemimetabolous and ammetabolous groups, which are below the horizontal line, comparatively show extreme variation and appear unconstrained.

Figure 2 – (A) non-measureable smeared tissue (sheared nuclei), (B) smeared tissue that is too thick and cannot be measured, (C) a mono-layer of single nuclei, but most are slightly ruptured, (E) another mono-layer of single nuclei containing more compact nuclei than (C); these are measurable. (D) Massive polyploid bacteriome nuclei (right side) next to smaller diploid nuclei (left side). In this study, the most abundant cell type is assumed diploid.
**Figure 3** - DNA content varies across cell phases (left graph), leading to cells with similar sizes but different DNA content. The right photo is of normal cells and those that are in the process of synthesizing DNA (black dots). This reinforces the assumption that the most abundant cells will be diploid, while variation will occur based on cell phase state. Besides phases, ploidy levels also vary in gametes and the bacteriomes of scale insects (from: [http://www.utoledo.edu/corelabs/amic/vickers.html](http://www.utoledo.edu/corelabs/amic/vickers.html)).

**Figure 4** – Methods used to conduct data exploration before model fitting for the data, in this case log body size versus log genome size. These methods include residual analysis for heteroscedasticity (A and C), Q-Q plots for normality (B), and Cook’s distance for outliers (D).
Figure 5 – Scatterplot matrix used to observe collinearity occurring between explanatory variables in Coccoidea. Upper right panels contain correlation values and their significance (with *** being highly significant), while the bottom left panels show correlation plots. Collinearity causes some of the explanatory power of one variable to already be contained within another. For example, number of plant hosts strongly correlates with scale insect geographic range size \((r^2=0.87^{***})\). As such, the explanatory power of number of hosts is already partially contained within geographic range size as a variable; the weaker of the two variables is removed from statistical analyses.
**Table 1.** Statistical outputs for both coccidean datasets. These datasets include contrasts of genome size and explanatory variables which are controlled for phylogeny (PICs or PGLS), while the other is the absolute relationship between genome size and variables without controlling for phylogeny (direct correlation). Asterisks (*, **, ***) denote significance levels matching the default in R. $r^2$ shading denotes a positive (green) or negative (red) significant relationship with genome size.

<table>
<thead>
<tr>
<th>Log-Variable Comparison (Genome size versus...)</th>
<th>Direct Correlation (SLR or ANOVA)</th>
<th>Phylogenetic Model (PICs or PGLS)</th>
<th>Pagel’s $\lambda$</th>
<th>Blomberg’s $K$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>p</td>
<td>$r^2$</td>
<td>df</td>
</tr>
<tr>
<td>Genome size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of chromosomes</td>
<td>40</td>
<td>0.311</td>
<td>0.001</td>
<td>38</td>
</tr>
<tr>
<td>Geographic range</td>
<td>60</td>
<td>0.09</td>
<td>0.03</td>
<td>58</td>
</tr>
<tr>
<td>Body size</td>
<td>60</td>
<td>5.5e-4***</td>
<td>0.17</td>
<td>58</td>
</tr>
<tr>
<td># of hosts (genera)</td>
<td>59</td>
<td>0.47</td>
<td>0.008</td>
<td>57</td>
</tr>
<tr>
<td># of foes (genera)</td>
<td>45</td>
<td>0.188</td>
<td>0.017</td>
<td>43</td>
</tr>
<tr>
<td>Reproductive frequency</td>
<td>37</td>
<td>0.027*</td>
<td>0.11</td>
<td>35</td>
</tr>
<tr>
<td>Reproductive type</td>
<td>22</td>
<td>0.292</td>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>
Figure 6 – Relationship between log genome size and log body size at the family level. There is a strong relationship for both Coccidae and Pseudococcidae ($p \leq 0.03$, $r^2 \geq 0.37$) which either have soft scales or are fully exposed. Diaspididae and Eriococcidae contain highly protected species, whether under armoured scales or within plant galls, and the relationship is non-significant ($p \geq 0.61$). This presents an interesting idea that the relationship may be driven by factors such as predation pressure causing selection on body size, which cascades to secondarily select for genome size, and this constraint may be significantly lessened in the protected groups.

Figure 7 – Relationship between log genome size and log reproductive frequency when analyzed at the family level. The comparative analysis across all species is significant ($p=0.004$, $r^2=0.19$), but within families there are no significant relationships observed between the two variables ($p \geq 0.45$). This signals that differences in these variables lie between clades and not within clades. Additionally, controlling for phylogeny removed this relationship – the pattern is explained by phylogenetic relatedness and for extant species reproductive frequency does not appear to directly impact genome size.
Figure 8 – Composite coccoidean phylogeny for 63 of 65 sampled species generated from existing phylogenetic data, including three outgroup species. Bars on the right denote genome size, while bars and circles are coloured by family.
Figure 9 – A positive relationship between genome size and body size is observed in Coccoidea ($p=5.5\times10^{-4}$, $r^2=0.17-0.25$), yet the relationship is non-existent in cockroaches ($p=0.98$, $r^2=0.02$). Genome size in Blattodea may be determined by other traits, such as sociality.
Figure 10 – Genome size estimates for *B. dubia* across different time regimes and treatment groups using tissue preserved in 70% EtOH. C-value estimates from fresh tissue align with previously published data for *B. dubia*, but these values then quickly decrease in value and increase in variation after one week, one month, and one year in ethanol. ‘FF’ is a freeze-flip method used on one-year preserved tissues, where tissue is frozen to a microscope slide and then sheared off, leaving behind a monolayer ‘thumbprint’ of nuclei. This is done after tissue is rehydrated with either citric acid (‘CA’), acetic acid (‘AA’), or water (‘W’), but still results in unstable estimates. Overall, preserved tissues continue to conflict with the Feulgen process and remain unreliable to measure in insects.
Table 2. Statistical outputs for both blattodean datasets. These datasets include contrasts of genome size and explanatory variables which are controlled for phylogeny (PICs or PGLS), while the other is the absolute relationship between genome size and variables without controlling for phylogeny (direct correlation). Asterisks (*, **, *** ) denote significance levels matching the default in R. $r^2$ shading denotes a positive (green) or negative (red) significant relationship with genome size.

<table>
<thead>
<tr>
<th>Log-Variable Comparison (Genome size versus...)</th>
<th>Direct Correlation (SLR or ANOVA)</th>
<th>Phylogenetic Model (PICs or PGLS)</th>
<th>Pagel’s λ</th>
<th>Blomberg’s K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>df</td>
<td>p</td>
<td>$r^2$</td>
<td>df</td>
</tr>
<tr>
<td>Genome size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body size</td>
<td>44</td>
<td>0.97</td>
<td>-0.02</td>
<td>42</td>
</tr>
<tr>
<td>Flight Ability</td>
<td>44</td>
<td>0.427</td>
<td></td>
<td>44</td>
</tr>
<tr>
<td>Reproductive type</td>
<td>44</td>
<td>0.732</td>
<td></td>
<td>44</td>
</tr>
<tr>
<td>Sociality Type</td>
<td>72</td>
<td>&gt;9.9e-4***</td>
<td></td>
<td>72</td>
</tr>
</tbody>
</table>
Figure 11 - Composite blattodean phylogeny for 68 of 70 sampled species generated from existing phylogenetic data. Bars on the right denoted genome size, while bars and circles within the tree are coloured by family. The evolution of sociality is shared by Cryptocerus cockroaches (subsocial, red) and termites (eusocial, brown), and a dramatic drop in genome size is observed in these groups relative to their solitary counterparts.
### Table 1 – Haploid genome size estimates for Coccoidea and Sternorrhynchan outgroups. Sample values (n) are given per family (number of species sampled in bold) and per species (number of individual specimens measured).

<table>
<thead>
<tr>
<th>Family</th>
<th>Species name</th>
<th>1C GS (pg)</th>
<th>Standard Error (±pg)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aleyrodidae</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td><em>Trialeurodes vaporariorum</em></td>
<td>0.59</td>
<td>0.17</td>
<td>3</td>
</tr>
<tr>
<td>Aphididae</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td><em>Ovatus mentharius</em></td>
<td>0.434</td>
<td>0.017</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Sipha maydis</em></td>
<td>0.463</td>
<td>0.005</td>
<td>5</td>
</tr>
<tr>
<td>Asterolecaniidae</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td><em>Bambusaspis bambusae</em></td>
<td>0.203</td>
<td>0.006</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Bambusaspis miliaris</em></td>
<td>0.150</td>
<td>0.001</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td><em>Planchonia stentae</em></td>
<td>0.364</td>
<td>0.003</td>
<td>5</td>
</tr>
<tr>
<td>Coccidae</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td><em>Ceroplastes floridensis</em></td>
<td>0.434</td>
<td>0.006</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Ceroplastes japonicus</em></td>
<td>0.818</td>
<td>0.017</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Ceroplastes sinensis</em></td>
<td>1.173</td>
<td>0.016</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Coccus hesperidium</em></td>
<td>0.673</td>
<td>0.016</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Coccus viridis</em></td>
<td>0.336</td>
<td>0.002</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Kenima galilt</em></td>
<td>1.575</td>
<td>0.012</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td><em>Parthenolecanium corni</em></td>
<td>1.229</td>
<td>0.038</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Protopulvinaria pyriformis</em></td>
<td>0.440</td>
<td>0.007</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td><em>Pulvinaria lloccifera</em></td>
<td>1.245</td>
<td>0.033</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Stozia ephedrae</em></td>
<td>0.595</td>
<td>0.010</td>
<td>5</td>
</tr>
<tr>
<td>Diaspididae</td>
<td></td>
<td></td>
<td></td>
<td>23</td>
</tr>
<tr>
<td></td>
<td><em>Acutaspis agavis</em></td>
<td>0.262</td>
<td>0.017</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Aonidiella auranti</em></td>
<td>0.647</td>
<td>0.003</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td><em>Aspidiotus nerii</em></td>
<td>0.391</td>
<td>0.007</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td><em>Chionaspis pinifoliae</em></td>
<td>0.534</td>
<td>0.011</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td><em>Chrysomphalus aonidum</em></td>
<td>0.630</td>
<td>0.004</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Diaspidiotus juglansregiae</em></td>
<td>0.802</td>
<td>0.007</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Diaspis boisduvalii</em></td>
<td>0.362</td>
<td>0.003</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td><em>Diaspis echinocacti</em></td>
<td>0.451</td>
<td>0.005</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td><em>Epidiaspis leperii</em></td>
<td>0.701</td>
<td>0.032</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Fiorinia externa</em></td>
<td>0.436</td>
<td>0.006</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td><em>Froggattiella penicillata</em></td>
<td>0.488</td>
<td>0.017</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td><em>Furcaspis bifomis</em></td>
<td>0.408</td>
<td>0.006</td>
<td>5</td>
</tr>
<tr>
<td>Species</td>
<td>Density</td>
<td>Error</td>
<td>ID</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>---------</td>
<td>-------</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>Hemiberlesia lataniae</td>
<td>0.664</td>
<td>0.013</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Hemiberlesia palmae</td>
<td>0.769</td>
<td>0.003</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Hemiberlesia rapax</td>
<td>0.633</td>
<td>0.003</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Lepidosaphes beckii</td>
<td>0.359</td>
<td>0.003</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Melanaspis obscura</td>
<td>0.877</td>
<td>0.02</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Parlatoria pergandii</td>
<td>0.678</td>
<td>0.008</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Parlatoria pittospori</td>
<td>0.597</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pinnaspis aspidistrae</td>
<td>0.183</td>
<td>0.002</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Pseudaulacaspis cockerelli</td>
<td>0.356</td>
<td>0.009</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Targonia vitis</td>
<td>1.231</td>
<td>0.030</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Unaspis euonymi</td>
<td>0.514</td>
<td>0.011</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Apiomorpha maliformis</td>
<td>1.170</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apiomorpha pileata</td>
<td>1.052</td>
<td>0.012</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Apiomorpha pomaphora</td>
<td>2.125</td>
<td>0.022</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Apiomorpha subconica</td>
<td>0.926</td>
<td>0.014</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Ascelis praemollis</td>
<td>0.288</td>
<td>0.003</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Ascelis schraderi</td>
<td>0.247</td>
<td>0.002</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Eriococcus coriaceus</td>
<td>0.300</td>
<td>0.002</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Eriococcus sp. 1</td>
<td>0.532</td>
<td>0.003</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Kermes echinatus</td>
<td>0.686</td>
<td>0.014</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Nidularia balachowskii</td>
<td>0.682</td>
<td>0.010</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Icerya purchasi</td>
<td>1.131</td>
<td>0.016</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Icerya seychellarum</td>
<td>2.062</td>
<td>0.028</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Phoenicococcus marlatti</td>
<td>0.198</td>
<td>0.002</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Coccidohystrix lubersaci</td>
<td>0.890</td>
<td>0.025</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Dysmicoccus brevipes</td>
<td>0.412</td>
<td>0.011</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Paracoccus marginatus</td>
<td>0.651</td>
<td>0.033</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Paracoccus sp.</td>
<td>0.566</td>
<td>0.006</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Phenacoccus gossypii</td>
<td>0.496</td>
<td>0.006</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Phenacoccus peruvianus</td>
<td>0.197</td>
<td>0.004</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Phenacoccus solani</td>
<td>0.363</td>
<td>0.008</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Phenacoccus solenopsis</td>
<td>0.508</td>
<td>0.014</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Planococcus citri</td>
<td>0.561</td>
<td>0.016</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Planococcus ficus</td>
<td>0.580</td>
<td>0.009</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
Table 2 – Haploid genome size estimates for Blattodea and associated outgroups (Gryllidae and Mantodea). Sample values (n) are given per family (number of species sampled in bold) and per species (number of individual specimens measured).

<table>
<thead>
<tr>
<th>Family</th>
<th>Species name</th>
<th>1C GS (pg)</th>
<th>Standard Error (±pg)</th>
<th>n</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blaberidae</td>
<td><em>Blaberus colloseus</em></td>
<td>4.08</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Blaberus craniifer</em></td>
<td>3.86</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Blaberus discoidalis</em></td>
<td>4.82</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Blaberus fuscus</em></td>
<td>3.36</td>
<td></td>
<td></td>
<td>Bier &amp; Muller (1969)</td>
</tr>
<tr>
<td></td>
<td><em>Blaptica dubia</em></td>
<td>4.77</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td>*Byrsotria sp. “Cuba”</td>
<td>2.64</td>
<td>0.05</td>
<td>2</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Elliptorhina javanica</em></td>
<td>2.98</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Epilampra maya</em></td>
<td>4.21</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td>*Eublaberus sp. “Ivory”</td>
<td>6.02</td>
<td>0.1</td>
<td>2</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Gromphadorhina portentosa</em></td>
<td>3.32</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Gyna caffrorum</em></td>
<td>1.94</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Gyna centurio</em></td>
<td>2.01</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Gyna lurida</em></td>
<td>2.29</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Hemiblabera tenebricosa</em></td>
<td>3.45</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Henschoutedenia flexivitta</em></td>
<td>6.79</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Nauphoeta cinerea</em></td>
<td>5.15</td>
<td></td>
<td></td>
<td>Koshikawa <em>et al.</em>, 2008</td>
</tr>
<tr>
<td></td>
<td><em>Panchlora nivea</em></td>
<td>1.60</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Paraplecta sp. 1</em></td>
<td>2.09</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Paraplecta sp. 2</em></td>
<td>2.27</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Pycnoscelus femapterus</em></td>
<td>5.57</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Pycnoscelus nigra</em></td>
<td>7.49</td>
<td>0.29</td>
<td>2</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td>*Pycnoscelus sp. “Thailand”</td>
<td>7.97</td>
<td>0.4</td>
<td>2</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Pycnoscelus striatus</em></td>
<td>2.84</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Rhabdoblatta formosana</em></td>
<td>2.15</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td></td>
<td><em>Schultesia lampyridiformis</em></td>
<td>4.11</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td>Blattidae</td>
<td><em>Blatta orientalis</em></td>
<td>3.03</td>
<td></td>
<td></td>
<td>Gregory, 2018</td>
</tr>
<tr>
<td></td>
<td><em>Eurycotis lixa</em></td>
<td>4.06</td>
<td></td>
<td>1</td>
<td>Present study</td>
</tr>
<tr>
<td>Family</td>
<td>Species</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neostylopyga</td>
<td>propinqua</td>
<td>2.20 0.19 2 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Periplaneta</td>
<td>americana</td>
<td>2.72 Gregory, 2018</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Periplaneta</td>
<td>australasiae</td>
<td>5.20 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Periplaneta</td>
<td>fuliginosa</td>
<td>4.53 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corydiidae</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arenivaga</td>
<td>bolliana</td>
<td>2.07 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compsodes</td>
<td>schwarzi</td>
<td>3.39 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ergaula</td>
<td>pilosa</td>
<td>9.58 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cryptoceridae</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cryptocerus</td>
<td>kyebangensis</td>
<td>1.16 Koshikawa et al., 2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cryptocerus</td>
<td>punctulatus</td>
<td>1.32 Koshikawa et al., 2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ectobiidae</td>
<td></td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balta</td>
<td>vilis</td>
<td>2.97 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blattella</td>
<td>asahinai</td>
<td>2.17 0.14 2 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blattella</td>
<td>germanica</td>
<td>2.00 Bier &amp; Muller (1969)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blattella</td>
<td>nipponica</td>
<td>2.24 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blattella</td>
<td>vaga</td>
<td>3.44 0.11 2 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cariblatta</td>
<td>minima</td>
<td>4.70 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischnoptera</td>
<td>bilunata</td>
<td>3.60 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischnoptera</td>
<td>deropteltiformis</td>
<td>2.42 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischnoptera</td>
<td>rufa</td>
<td>2.80 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parcoblatta</td>
<td>americana</td>
<td>5.35 0.31 2 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parcoblatta</td>
<td>bolliana</td>
<td>4.00 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parcoblatta</td>
<td>caudelli</td>
<td>4.77 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parcoblatta</td>
<td>desertae</td>
<td>5.04 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parcoblatta</td>
<td>divisa</td>
<td>3.05 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parcoblatta</td>
<td>fulvescens</td>
<td>3.82 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parcoblatta</td>
<td>lata</td>
<td>3.57 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parcoblatta</td>
<td>uhlerianna</td>
<td>3.34 0.08 2 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parcoblatta</td>
<td>virginica</td>
<td>4.24 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parcoblatta</td>
<td>zebra</td>
<td>3.09 1 Present study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gryllidae</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acheta</td>
<td>domesticus</td>
<td>2.38 Koshikawa et al., 2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hodotermitidae</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hodotermites</td>
<td>mossambicus</td>
<td>1.00 Koshikawa et al., 2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>Family</td>
<td>Species</td>
<td>Distance (km)</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------</td>
<td>----------------------------------</td>
<td>---------------</td>
<td>----------------------------</td>
<td></td>
</tr>
<tr>
<td>Hymenopodidae</td>
<td></td>
<td>Acromantis japonica</td>
<td>4.53</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td>Kalotermitidae</td>
<td></td>
<td>Glyptotermes fuscus</td>
<td>1.56</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Glyptotermes nakajimai</td>
<td>0.89</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Neotermes koshunensis</td>
<td>1.87</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cryptotermes secundus</td>
<td>1.33</td>
<td>Harrsion et al., 2018</td>
<td></td>
</tr>
<tr>
<td>Mantidae</td>
<td></td>
<td>Statilia maculata</td>
<td>3.05</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tenodera aridifolia</td>
<td>2.92</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td>Mastotermitidae</td>
<td></td>
<td>Mastotermes darwiniensis</td>
<td>1.30</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td>Rhinotermitidae</td>
<td></td>
<td>Coptotermes formosanus</td>
<td>0.86</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reticulitermes speratus</td>
<td>1.07</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reticulitermes speratus kyushuensis</td>
<td>1.07</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reticulitermes speratus</td>
<td>1.01</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reticulitermes yaeyamanus</td>
<td>1.00</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reticulitermes flavipes</td>
<td>1.07</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td>Termitidae</td>
<td></td>
<td>Nasutitermes takasagoensis</td>
<td>1.68</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Odontotermes formosanus</td>
<td>1.48</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pericapritermes nitobei</td>
<td>1.90</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Macrotermes natalensis</td>
<td>1.34</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td>Termopsidae</td>
<td></td>
<td>Hodotermopsis sjostedti</td>
<td>1.27</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zootermopsis nevadensis</td>
<td>0.58</td>
<td>Koshikawa et al., 2008</td>
<td></td>
</tr>
</tbody>
</table>
Modified R script for running ANOVA/PGLS models using discrete variables with 2+ levels in the R package ‘caper’

```r
anova.pgl.s.fixed <- function (object) {
  data <- object$data
  tlabels <- attr(terms(object$formula), "term.labels")
  k <- object$k
  n <- object$n
  NR <- length(tlabels) + 1
  rss <- resdf <- rep(NA, NR)
  rss[1] <- object$NSSQ
  resdf[1] <- n - 1
  lm <- object$param["lambda"]
  dl <- object$param["delta"]
  kp <- object$param["kappa"]
  for (i in 1:length(tlabels)) {
    fmla <- as.formula(paste(object$namey, " ~ ",
                         paste(tlabels[1:i], collapse = "+")))
    plm <- pgls(fmla, data, lambda = lm, delta = dl, kappa = kp)
    rss[i + 1] <- plm$RSSQ
    resdf[i + 1] <- (n - 1) - plm$k + 1
  }
  ss <- c(abs(diff(rss)), object$RSSQ)
  df <- c(abs(diff(resdf)), n - k)
  ms <- ss/df
  fval <- ms/ms[NR]
  P <- pf(fval, df, df[NR], lower.tail = FALSE)
  table <- data.frame(df, ss, ms, f = fval, P)
  table[length(P), 4:5] <- NA
  dimnames(table) <- list(c(tlabels, "Residuals"), c("Df", "Sum Sq", "Mean Sq", "F value", "Pr(>F)"))
  structure(table, heading = c("Analysis of Variance Table",
                              sprintf("Sequential SS for pgls: lambda = %0.2f, delta = %0.2f, kappa = %0.2f\n", lm, dl, kp),
                              paste("Response:",
                              deparse(formula(object)[[2L]]))), class = c("anova", "data.frame"))
}
```