Reduced regulatory complexity associated with the evolutionary transition to sociality from cockroaches to termites despite evolutionary parallels with Hymenoptera

Alun R.C. Jones*[1], Elias Dohmen[1], Juliette Berger[2], Frédéric Legendre[2], Mark Harrison[3], Erich Bornberg-Bauer[1,4,5]

1 Institute for Evolution and Biodiversity, University of Münster, Münster, Germany

2 Institut de Systématique, Évolution, Biodiversité (ISYEB), UMR 7205, Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE-PSL, Université des Antilles, CP50, 57 rue Cuvier, F-75005 Paris, France

3 Research Centre for Health and Life Sciences, Coventry University, UK

4 Department of Protein Evolution, Max Planck Institute for Biology Tübingen, Max-Planck-Ring 5, Tübingen, Germany

5 Theoretical Sciences Visiting Program, Okinawa Institute of Science and Technology Graduate University, Onna, 904-0495, Japan

*** alunjones_101@hotmail.com**

Abstract

In social species, the phenotypic diversity within a single species has often been associated with more complex genetic regulation. Transcription factors (TFs) being key to genetic regulation, have been studied in the origins of eusociality in Hymenoptera (Bees, Ants and Wasps) but less so in Blattodea (Cockroaches and Termites). Here we show that the social transition in Blattodea, from cockroaches to termites, displays similar patterns of regulatory change to those found in Hymenoptera. Specifically, by analysing 3 cockroach and 5 termite genomes, we find more TF families with relaxed selection compared to intensified selection and lineage-specific gene family expansions in termites. We also find that genes under selection support neotenic caste determination. There are key differences in TF regulation in comparison with Hymenoptera with contractions in TF gene families and reduced diversity in enriched DNA binding motifs of TFs.

Furthermore, we show that with increased social complexity TF activating domain diversity, one of the evolutionary and structural building blocks of TFs, decreases with gene family size while DNA binding domain diversity and number increases. This study highlights similarities in social transitions between Hymenoptera and Blattodea, with evidence of large changes in transcriptional regulation followed by lineage specific adaptations. We also show that phenotypic diversity in termites does not necessitate greater transcriptional complexity.

Introduction

The evolution of eusocial organisms has remained an intriguing question in evolutionary biology as it signifies a major transition in organism complexity [45]. The defining characteristics—cooperative brood care, reproductive division of labor, and overlapping generations—have been observed in a wide variety of taxa, including in mammals, shrimp, and insects, such as bees, ants, wasps and beetles [10, 12, 32, 40, 63]. There are several other types of social organisation, such as sub sociality which involves parental care and gregarious behaviour in species which live in large groups of conspecifics [12, 29, 39]. The vast majority of eusocial research has focused on Hymenoptera [62], which harbors multiple origins of eusociality [57]. The Blattodea have a single origin of eusociality, the termites, but exhibit a full continuum of social behaviors with solitary, gregarious, sub social, primitively eusocial and complex eusocial clades allowing for comparative studies both within blattodea and between other social taxa [6, 17, 40, 66]. Other interesting features of Blattodea, which sets them apart from Hymenoptera, are that termites are hemimetabolous in contrast to the holometabolous Hymenoptera. They also possess both diploid males and females within colonies including long lived reproductives of both sexes [75] with XY sex chromosomes [37]. Blattodea, therefore are important to understand molecular mechanisms of social evolution as well as eusocial transitions because they enable comparisons of underlying molecular and genomic features and, consequently, to disentangle which of these features are more general if they are shared with other eusocial insects and which are more clade-specific.

Exploring the molecular details underlying the evolution of Blattodea has only become feasible recently, with a total of five termite genomes and three cockroach genomes having been published to date [6, 17, 26, 42, 47, 69, 77]. Remarkable insights from these species have identified mechanisms of viviparity (live bearing) and genomic changes related to social evolution. A hallmark of social insect molecular evolution is the regulation of chemical perception. Chemical perception is essential for many important tasks such as communication, identifying nest mates and their caste status or sensing food sources. Chemoreceptors in particular have been found to undergo drastic changes in the course of evolution and have been linked to the onset and evolution of social behaviour, among others. These changes include expansions in ionotropic receptors [26], in contrast to the odorant receptors found in Hymenoptera [83] although in hymenoptera there is evidence to the contrary [20]. The expansion of two different major chemoreceptor families in

Hymenoptera and Blattodea indicates that while functional changes can be conserved with major transitions the molecular mechanisms to achieve the change can vary. Another convergent pattern observed in both clades is alterations in methylation patterns which have associations with caste specific gene regulation [22, 25, 44], although this relationship is not necessarily consistent [21]. Specific regulatory pathways, such as the juvenille hormore pathway, which is known to be involved in caste determination, also convergently undergone changes in both Blattodea and Hymenoptera [26, 64].

While specific pathways have been investigated, a more general understanding of regulatory changes related to social evolution has not been conducted in Blattodea.

To gain a better understanding of gene regulatory changes linked to sociality, the present study focuses specifically on transcription factors (TFs), whose DNA binding domain and secondary activating domains control gene expression [38]. TF influence in social evolution in Hymenoptera has revealed convergent patterns of enrichment of TF binding sites in regulatory DNA regions, accompanied by significant gains in particular gene families [70].

In Hymenoptera TF-specific gene regulatory changes have been identified across eusocial origins in bees and ants. The changes included the enrichment of specific TF binding sites in promoter regions [70] in combination with significant expansions in particular gene families [11, 70]. It remains unclear whether similar convergent patterns exist in termites. To this end, Harrison et al. (2018) identified expansions in zinc fingers in termites [26] but changes in regulatory regions were not studied.

This study aims to identify changes in transcription factor regulation associated with social evolution in Blattodea. Utilizing the published genomes of five termites and three cockroach species, we investigate alterations in TF-mediated regulation. Specifically, we examine changes in TF number, binding structure, binding targets, and the interplay between all of those. Our goal is to elucidate key patterns for complex social transitions in Blattodea. By comparing these findings to analogous patterns in Hymenoptera, we aim to create a more comprehensive understanding of the genetic regulatory changes associated with social evolution. This study should reveal if socially associated TF regulatory changes are A) present in non-Hymenoptera social transitions and B) if these changes follow the same trends as those found in Hymenoptera.

Materials and Methods

Transcription factor identification

The defining features of TFs are the DNA binding domains and a secondary activating domain [38]. These protein domains can be identified from known modular DNA sequences allowing the identification of conserved protein functions and evolution [2, 73]. For this study transcription factors were identified using 70 Unique DNA binding PFAM domains [52] found in *D. melanogaster* transcription factors from FlyBase. We chose 3 cockroaches: *Blattella germanica* , *Diploptera punctata* and *Periplaneta americana* alongside 5 termite species: *Zootermopsis nevadensis* , *Cryptotermes secundus* , *Reticulitermes speratus* , *Coptotermes formosanus* and *Macrotermes natalensis* additionally two outgroup species were used; *Frankliniella occidentalis* and *Eriosoma lanigerum* [5, 17, 26, 31, 42, 43, 58, 65, 69, 77]. These species were chosen not only because at the time of the study they were all the genomic resources available but they also represented a large spectrum of sociality with solitary, gregarious, primitively and complex eusocial species being used. Pfam scans were run using the longest isoform protein sequence, and genes with any of the 70 DNA binding domains were identified as potential transcription factors. Orthogroups were identified using Orthofinder2 [15](including *D. melanogaster* 12630 TFs were identified in a total of 1621 orthogroups). Orthogroups containing genes identified as potential transcription factors were used for all analyses. Scripts for all analyses are available on https://doi.org/10.6084/m9.figshare.c.4902573.

Multiple sequence alignment

Amino acid multiple sequence alignments of TF containing orthologs were run using GUIDANCE(2.02) [68] and MAFFT(v7.397) [34] using a global pair alignment with max iterations of 1000. Unfiltered protein alignments were transformed into nucleic acid sequences using Pal2Nal [74] and aligned but only using 200 iterations and the resulting alignments were filtered according to the amino acid alignments using a script from *Fouks et al 2021* [16].

Gene family dynamics and selection

To identify TF family expansions and contractions CAFE5 [46] was used using the Orthofinder2 results. Only orthogroups where TFs had been identified in Blattodea species were used. These results were formatted using a custom bash script. A multi-lambda model was used with two

lambda values estimated, one for the termite branches and one for the rest [46].

45 Single-copy orthologs were used to investigate signals of selection in relation to social evolution. A species tree was constructed by concatenating all single copy alignments and then running IQtree2(2.1.3) [50]. Branch-specific relaxed selection was investigated using RELAX [80], branch-specific selections using aBSREL [71] with the termites as the selected branch and sites of episodic diversifying selection using MEME [53]. Significance was determined using a pvalue of < 0.05 for MEME and aBSREL and a false discovery rate < 0.05 for RELAX. Visualisation of selection was done through Hyphy vision website.

Protein family dynamics

PFAM domain rearrangement events were computed with DomRates [13] and visualised with the DomRates visualisation script. The total number of DNA binding domains was extracted using custom python and bash scripts and then clustered together using the R pheatmap function [35]. While genes may duplicate it does not mean that the resulting duplications have the same PFAM domain arrangements. To understand the dynamics between gene duplication and how sociality might be linked to domain changes, the relation between the number of genes in an orthogroup and the number and type of PFAM domains was modeled. An Ornstein–Uhlenbeck (OU) covariance matrix was constructed using the species tree and the ape package vcv function [56]. Two different full multilevel negative binomial phylogenetic models were constructed using orthogroup and species as mixed effects to account for the multiple species level observations. Both full models contained fixed effects for each species's orthogroup. One model contained the total number of DNA binding domains and the total number of activating domains while the other contained the number of unique DNA binding domains, and the number of unique activating domains. A modifier term with sociality was added for each of these fixed effects in addition to the fixed effects. Each species was given a value for the different sociality indexs: solitary, gregarious, primitively eusocial and highly eusocial. This had: *M. natalensis* , *R. speratus* and *C. formosanus* as complex eusocial and *C. secundus* and *Z. nevadensis* as primitively eusocial *B. germanica* , *P. americana* and *D. punctata* as gregarious and the outgroups as solitary. The model was run using brms(2.20.1) [8] with 4 chains on 4 cores using weakly regularising priors and a warmup of 1500 iterations and sampling of 3000. Diagnostic plots were looked at with DHarma [27]. 2 full models were reduced and compared using leave-one-out-cross-validation (LOO) [79]. Analysis was run in R(4.3.1) using the tidyverse, dplyr, ggplot and bayesplot packages [19, 61, 81, 82].

Promoter DNA Binding

Promoter regions were extracted by taking a region of 1000bp upstream of the gene's start codon based upon the length of known promoters [36] using GTF tools and Bedtools [7, 41, 60]. These were then split into general promoters and promoters for TFs. Using the JASPAR insect-specific TF binding motif database [9], DNA binding motifs were first identified using MEME, motifs were checked for enrichment within promoter sequences using SEA and compared using TOMTOM [3, 4, 24]. Motif comparison was done in R using JASPAR packages. Known and novel motifs were compared using Pearson correlation coefficient and trees were created. Novel motifs are motifs that do not have a significant match with motifs found in the JASPAR database.

Results

TF and Orthology identification

Sequences containing known *D. melanogaster* TF DNA binding domains were collected. The cockroaches contained the following number of TFs; *B. germanica* 805, *P. americana* 1096 and *D. Punctulata* 1274 TFs. The termites contain; *C. formensus* 615, *C. secundus* 911, *M. natalensis* 567 and *R. speratus* 854. TFs were also extracted from outgroups, resulting in a total of 12630 TF sequences. Identified TFs were spread across 1621 orthogroups.

Selection

To see if TFs in Blattodea follow similar patterns of selection concerning social evolution as those found in Hymenoptera we looked at branch specific relaxed and episodic diversifying selection as well at site specific level for episodic diversifying selection. By looking at the type of selection occurring and the specific genes that it was acting in, we can identify conserved patterns.
RELAX [80] identified 7 genes undergoing relaxed selection with no genes undergoing intensified selection.
Of the 7 relaxed genes 4 had variation in DNA binding domains. The zinc finger *Klumpfuss*, which is involved in neuroblast identification [33], includes a change in the C2H2-type zinc finger domain in *M. natalensis*. *Cytochrome c oxidase assembly factor 10* (*Cox10*), which is involved in heme biosynthesis and respiration [51], gains a homeodomain in *M. natalensis*, but does not have a DNA binding domain in other species. There is a gain of C2H2 zinc fingers in the vacuolar protein sorting protein 16B (*Vps16B*) in *B. germanica* but this is absent in other species. This gene is important for maintaining vacuole stability and structure [1]. 3 Of these 7 genes; *snoo*, *tango* and *klumpfuss*, have known neuro-developmental functions. *tango* is involved in regulating *breathless* which regulates trachea development and axonogenesis [72]. *sno* binds to a different TF which changes its affinity resulting in antagonism of the product of *dpp* and facilitation of Activin signaling [78]. *vrille* (*vri*), the bzip TF, also regulates *dpp* but as an enhancer. *dpp* is important for late stage wing development [55, 67].
aBSREL identified one gene with positive selection in 2 branches with 6 other genes with selection in 1 branch. All branches identified for positive selection were terminal branches. The gene with selection in the *R. speratus a*nd *C. secundus* is *Lim3*, a zinc finger and homeodomain transcription factor that regulates neuronal subtype, particularly motor neuron identity. Single branch included:

the zinc finger *lmd*, *abstrakt*, ATPSynthaase (only gained DNA binding domain in *P. americana*), Zinc transcription factor *CG7987* (transcription repressor of RNA-polII) [76], Vps16 (only gained a DNA binding domain in *B. germanica* , Helix-loop-helix transcription factor *tango*. *lmd* is important for the specification of myoblasts [14]. *abstrakt* is a DEAD box helicase, involved in processes such as axon asymmetric cell division [30].

MEME was used to identify sites undergoing site specific episodic diversifying selection and identified 22 sites of episodic and pervasive selection [53]. 5 sites appear to undergo episodic diversifying selection. 3 of these sites are in *Vps16* and one in *CEP104*, which is important for centriole elongation. Both orthgroups only gaining DNA binding domains in cockroaches. Episodic positive selection was identified in *UPf1*, which is involved in mediating nonsense mRNA degradation. Of the 11 orthogroups in which the 17 sites were identified with purifying selection, only 7 have DNA binding domains in Blattodea. One of them only gained a DNA binding domain in *M. natalensis* and another gained a DNA binding domain in *P. americana* . Orthogroups of interest include: *tango*, *CG7897*, *CrebB*, *PSEA-binding protein*, *Inverted repeat protein* and *Lim3*. *PSEA-binding protein* is part of the small nuclear RNA activating complex which is involved in activating transcription of snRNA genes [28]. *CrebB* is a transcription factor important for controlling genes important for long-term memory formation [18]. *Inverted repeat protein* forms part of the Ku complex and is involved in double strand break repair and telomere maintenance [48].

Gene family changes

In Hymenoptera general patterns of social evolution show several lineage specific expansions, both in the TFs and in other genes [70]. Consequently, we investigated gene family changes in Blattodea TFs. Gene family expansions and contractions were only considered in orthogroups where a DNA binding domain was identified in Blattodea. The origin of termites had 25 gene family contractions and 5 gene family expansions. When looking at significant gene family changes in termites at any of the branches within the clade, there are 75 significant gene family contractions compared to 63 expansions. This shows that the majority of gene family changes occur after the social origin and contractions dominate at the phylogenetic origin of eusocial evolution.

Given the high number of contractions, we next looked into the prevalence of contractions within the termite lineages. 4 out of the 5 termite species have more contractions than expansions. Those gene families undergoing expansion are lineage-specific. The expanding orthogroups tend to

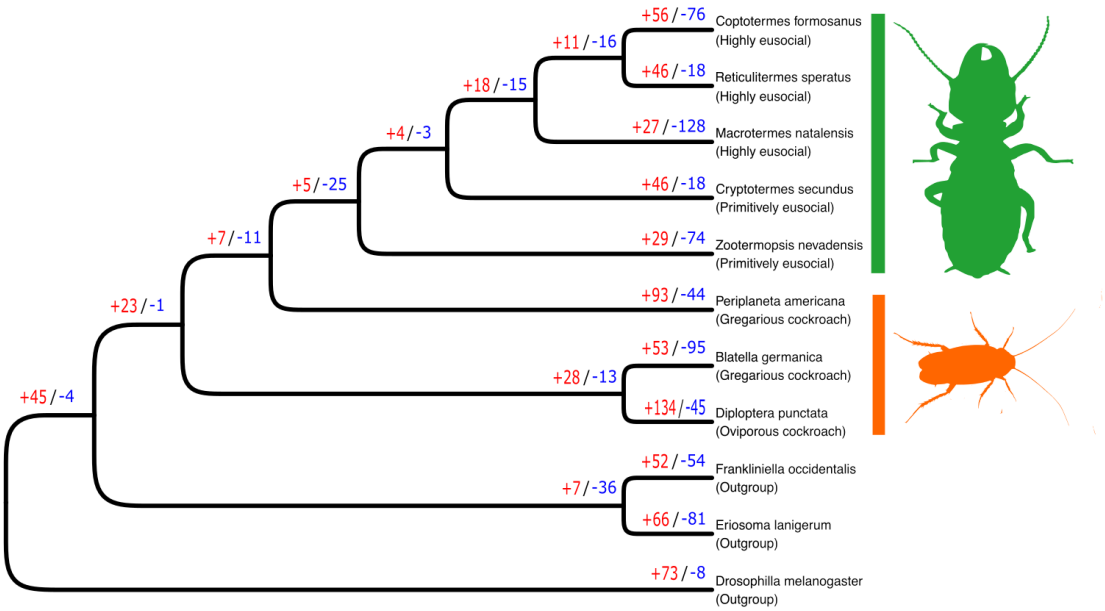

Figure 1. Cladogram of phylogenetic tree depicting all the species used, whether they are cockroaches or termites and their type of social organisation. Numbers in red are the number of TF containing orthogroups undergoing significant gene family expansions and the blue numbers are the TF containing orthogroups undergoing significant gene family contractions.

be expanding in one lineage but contracting in others.

There are 21 orthogroups found to only undergo contraction, while only 9 orthogroups undergo expansion. Genes families that we see undergoing expansions at the termite node include Neuroectoderm-expressed 2, Rfx, elbow b, noc and gemini (Ubiquitin-63E,CG11700 but doesn't have TF activity in fly,Ubiquitin-5E). To understand if there were similar processes affected by gene family contractions we conducted a GO term enrichment analysis. The GO term enrichment of contracted orthogroups did not show any particular function or process that was enriched.

PFAM dynamics

To understand large scale changes in genomic regulation it is not sufficient to just look at gene family changes but to also understand the protein function. Given TFs are defined by DNA binding domains and activating domains, we looked at changes in the domain arrangement to understand the functional changes. Looking at the overall grouping in DNA binding domain numbers based upon the heat map there are no clear increases or decreases in DNA-binding domains among transcription factors that could be observed in termites (Figure 2A).

To see if there were key changes in function brought about via changes in domain modularity we

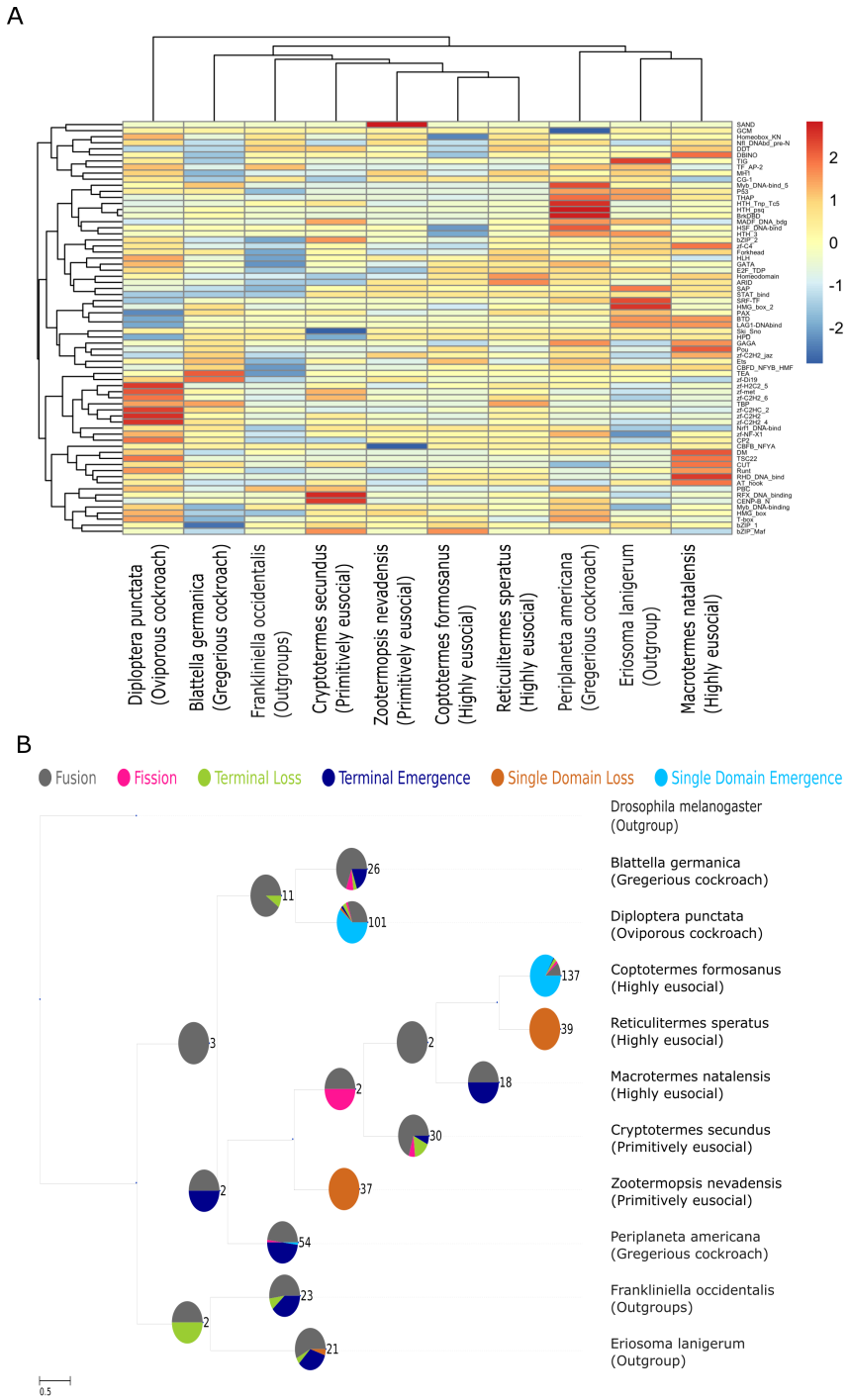

Figure 2. A) Heat map showing the prevalence of DNA binding domains. The number of domains has been standardised across the species. B) Graph showing output from Domrates.

looked at protein domain rearrangements using DomRates [13]. Domrates did not show any clear patterns in TFs but at the emergence of the termites, there were no domain changes (Figure 2).

To understand how domain changes and gene family changes relate to one another we conducted phylogenetic GLMMs. Model comparison was conducted using LOO and the best-fitting models for the total domains included a social index moderating effect (Sup table 2 and 1). When looking at total domains within orthogroups there is a strong positive relationship between the gene counts and the number of activating domains, as well as the number of dna binding domains. The moderating effect with the social index and activating domains is strongly negative (Figure 3A). So more social species tend to have proportionally less activating domains within TF gene families. With DNA binding domains there is a less pronounced positive relationship with the total number of DNA binding domains and gene family size, compared to activating domains. When looking at the social moderating effect, this is a positive moderating effect with social index. (Figure 3A). The best fitting model found from the LOO comparison using the number of unique domains included a moderating effect with the number of unique DNA binding domains and the social index (Table 1). The general trends found were a positive relationship between the number of unique DNA binding domains and gene family size and a negative relationship between unique activating domains as well as with social index. The modifying effect with social index in the best fitting unique model was a positive relationship with DNA binding domains, indicating that more social species tend to have more unique DNA binding domains proportionally than less social species (Figure 3B).

DNA binding motifs

A key finding in Hymenoptera is an enrichment in DNA binding motifs in promoter regions of social species promoters [70]. To see if the trend was present in Blattodea, we investigated DNA binding motifs in promoter regions. The promoter regions of TF genes and non-TF genes were investigated separately. Cockroach TFs were found to be enriched with more motifs and by more diverse motifs compared to termites (Figure 4A). A similar pattern was found in genes other than TFs but was not as striking as that found in TFs (Sup figure 5). *B. germanica*, *P. americana* and *D. punctulata* have 272, 128 and 381 enriched motifs respectively, while the termites: *C. formensus*, *C. secundus*, *M. natalensis*, *R. speratus* and *Z. nevadensis* each have 86, 183, 153, 88 and 80 respectively.

Given the large number of 'novel' motifs we wanted to see how these motifs compare to known motifs and infer their evolutionary history. From this, we could infer if they might be more similar to known motifs and likely have similar functions or if they are more similar to each other

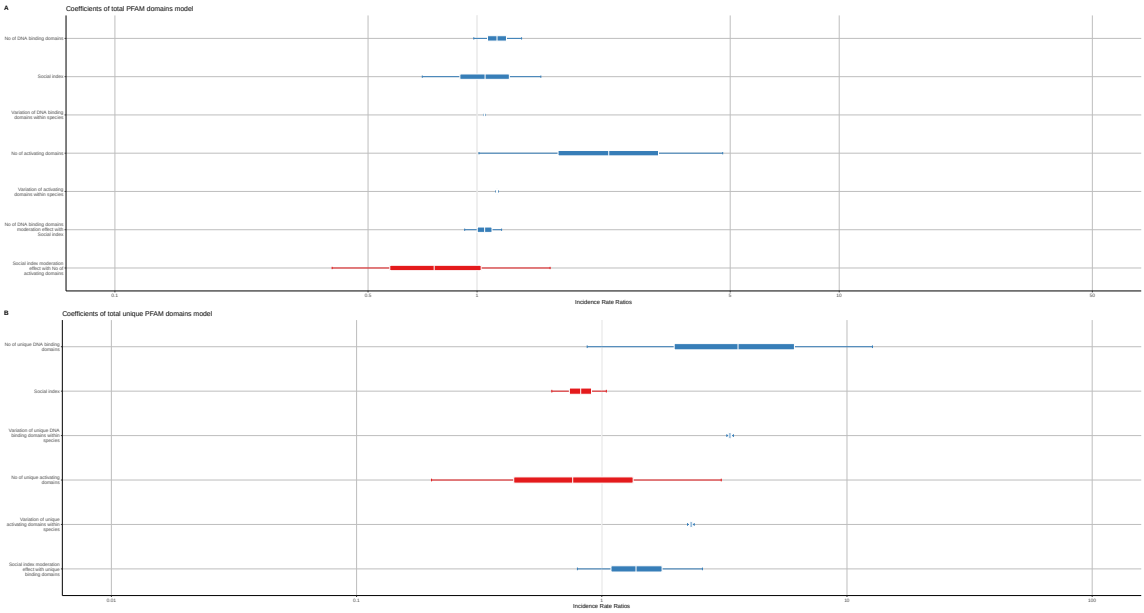

Figure 3. Figure showing the coefficients of the best fitting phylogenetic regression models, based upon LOO comparisons, predicting gene family size based upon the protein domain numbers. The x-axis A) Shows the best fitting model based on the total number of domains within the gene family. B) Shows the best fitting model based upon the number of unique domains within the Gene family.

indicating novel variation. When comparing novel motifs found in promoter regions of transcription factors, cockroaches have a large number of novel motifs that are more similar to other novel motifs (Figure 4B). Termites do not have a similar large novel motif grouping. Instead, more of the identified novel motifs are grouped with known motifs (Figure 4C).

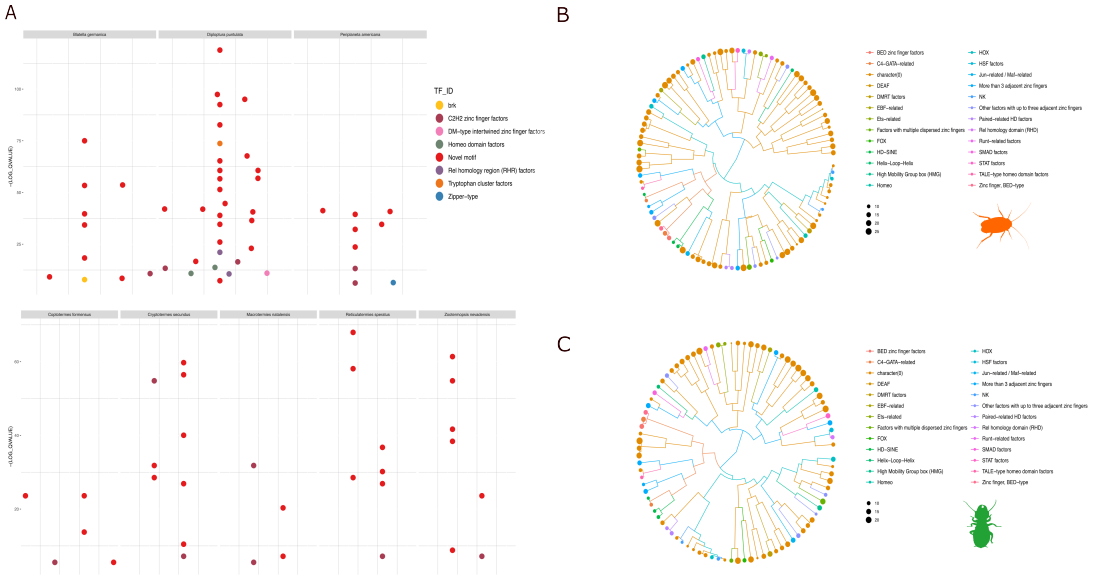

Figure 4. A) Shows the enrichment of DNA binding domain motifs in promoter regions of TFs in cockroaches as well as termites. Identified using SEA from the XStreme package. B) Phylogenetic tree showing the relationship between all novel cockroach TF motifs found in TF promoters and all the known motifs found across termites and cockroaches. Phylogenetic tree constructed using Persons correlation. The size of the tips represents the information criterion score, which is a measure of how different a motif is from a uniform distribution. C) All the termite novel motifs.

Discussion

This study identified several regulatory changes associated with the evolution of eusociality in Blattodea. Unlike Hymenoptera, which undergo an expansion of enriched TF binding motifs [70], termites exhibit a reduction in the diversity of enriched TF binding sites (Sup figure 5), particularly in TF promoters (Figure 4A). Although larger genomes in cockroaches may explain part of this pattern [6], the substantial difference in the number of enriched motif sequences in promoter regions suggests biological significance, especially considering the similarity in the number of TFs between cockroaches and termites (Figure 1). While found that expansions in motif binding regulation are found to be correlated with higher levels of social organisation within Hymenoptera [70]. Our study suggests that this is not the case for Blattodea. Increasing diversity of regulation may, therefore, not necessarily come from increasing TF binding site availability, so this would need to be considered in the context of other TF related changes, such as TF gene family changes and how the domain content within those gene families varies. Our model shows evidence that diversity is being created by changing domains, at least through the DNA binding domains (Figure 3). The models suggest that activating domains are more likely to be lost in gene

families of more social species.

The phylogeny of the novel motifs in termites reveals that the majority of these motifs are more similar to known motifs rather than to each other (Figure 4B). Despite having fewer species in this study, cockroaches exhibit more novel motifs regulating TFs, with several large clusters where many novel motifs are more similar to one another than to known motifs (Figure 4C). This suggests a loss of novel TFBS motifs in termites. Coupled with the overall reduction in TF families at the origin of termites (Figure 1), the results indicate increased regulation may not be a necessity during transitions towards eusociality. Instead, what might be more important is a general large regulatory change followed by lineage-specific changes. This trend is also observed in Hymenoptera [70], where regulation changes, such as gene family expansions, are identified alongside lineage-specific gene composition changes. Similarly, the lack of convergence in the orthogroups undergoing significant expansions in termites suggests a pattern of large changes in gene regulation followed by alterations in lineage-specific gene composition in Blattodea.

No single DNA binding domain exhibits a clear significant expansion or contraction in termites (Figure 2A). When it comes to the TFBS sites though there is a clear reduction in diversity of enriched sites in termites compared to cockroaches (Figure 4A)). This seems contradictory when looking at the two phylogenetic models where both the total number of DNA binding domains and the number of unique DNA binding domains have a positive modulating effect with social index (Figure 2A/B). When it comes to activating domains there appears to be a different relationship. The social index has a negative moderating effect with total activating domains on predicting TF-containing gene family sizes. Given there was no moderating effect with social index and the number of unique activating domains it would suggest social species are more likely to lose activating domains (Figure 3). This is supported by the DOMRates results (Figure 2B), as there are predominantly fusion events occurring within termites and cockroaches (Figure 2B). While major changes in domain arrangements have been found in the social transition to termites [49], this does not seem to be the case for TFs. This supports the finding in that a large change in regulation, in this case reductions rather than expansions, followed by lineage specific adaptations accompany social transitions.

When examining selection, similarities in TFs emerge between Hymenoptera and Blattodea, with relaxed selection being more prevalent than positive selection. Genes identified with both relaxed and positive selection exhibit similar functions, with many of them playing crucial roles in neuronal development and memory formation such as *tango*, *klumpfuss*, *Lim3* and *CrebB*. Notably,

the only positively selected orthogroup containing a DNA binding domain in all studied Blattodea species is *Lim3*. This trend persists across multiple studies, showing consistent changes in neuronal gene expression and cell number between reproductive castes, as supported by experimental evidence [23, 59].

The global patterns found in Hymenoptera [70] of relaxed selection being more prominent than intensifying selection in selection in social lineages held for Blattodea. Interestingly two genes under relaxed selection, *vri* and *sno*, are involved in regulating the *dpp* pathway which controls late stage wing development [55]. This fits in well with the neotenic workers in termites which exhibit juvenile characteristics while only alates have fully developed wings [54].

Our investigation into the regulation and composition of TFs identified that unlike expansive patterns found in Hymenoptera [70] the social transition in Blattodea is accompanied by overall reductions in TF regulation. While Hymenoptera shows an increase in transcription regulation with increased TF binding sites, Blattodea showed the opposite with fewer enriched sites, particularly in TF promoters and fewer TFs particularly losing activating domains. also identified lineage-specific changes, which is the same pattern we identify here, with social species more likely to have more unique DNA binding domains within TF containing gene families (Figure 3B). This study is the first to identify a link between TF gene family diversity and sociality in termites. Future studies must look at the relationship between changes in regulatory mechanisms instead of in isolation to understand social transitions.

Acknowledgments

ARCJ acknowledges support from DFG through grant BO 2544 / 15-1 to E.B.-B. ED was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 503348080. JB was supported by French National Reserch Agency grant no ANR-19-CE02-0023 (Project SOCIOGENOMICS). Some parts of this research were conducted while visiting the Okinawa Institute of Science and Technology (OIST) through the Theoretical Sciences Visiting Program (TSVP).

References

1. M. A. Akbar, C. Tracy, W. H. Kahr, and H. Krämer. The *full-of-bacteria* gene is required for phagosome maturation during immune defense in *Drosophila*. *Journal of Cell Biology*, 192(3):383–390, Feb. 2011.

2. G. Apic, J. Gough, and S. A. Teichmann. Domain combinations in archaeal, eubacterial and eukaryotic proteomes1. *Journal of Molecular Biology*, 310(2):311–325, July 2001.

3. T. L. Bailey and C. E. Grant. SEA: Simple Enrichment Analysis of motifs, Aug. 2021. Pages: 2021.08.23.457422 Section: New Results.

4. T. L. Bailey, J. Johnson, C. E. Grant, and W. S. Noble. The MEME Suite. *Nucleic Acids Research*, 43(W1):W39–W49, July 2015.

5. R. Biello, A. Singh, C. J. Godfrey, F. F. Fernández, S. T. Mugford, G. Powell, S. A. Hogenhout, and T. C. Mathers. A chromosome-level genome assembly of the woolly apple aphid, Eriosoma lanigerum Hausmann (Hemiptera: Aphididae). *Molecular Ecology Resources*, 21(1):316–326, Jan. 2021.

6. E. Bornberg-Bauer, M. C. Harrison, and E. Jongepier. The first cockroach genome and its significance for understanding development and the evolution of insect eusociality. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 330(5):251–253, 2018. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jez.b.22826.

7. V. Buffalo. *Bioinformatics Data Skills: Reproducible and Robust Research with Open Source Tools*. "O'Reilly Media, Inc.", July 2015. Google-Books-ID: XxERCgAAQBAJ.

8. P.-C. Bürkner. **brms** : An *R* Package for Bayesian Multilevel Models Using *Stan*. *Journal of Statistical Software*, 80(1), 2017.

9. J. A. Castro-Mondragon, R. Riudavets-Puig, I. Rauluseviciute, R. Berhanu Lemma, L. Turchi, R. Blanc-Mathieu, J. Lucas, P. Boddie, A. Khan, N. Manosalva Pérez, O. Fornes, T. Leung, A. Aguirre, F. Hammal, D. Schmelter, D. Baranasic, B. Ballester, A. Sandelin, B. Lenhard, K. Vandepoele, W. W. Wasserman, F. Parcy, and A. Mathelier. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 50(D1):D165–D173, Jan. 2022.

10. S. T. C. Chak, S. E. Harris, K. M. Hultgren, N. W. Jeffery, and D. R. Rubenstein. Eusociality in snapping shrimps is associated with larger genomes and an accumulation of transposable elements. *Proceedings of the National Academy of Sciences*, 118(24):e2025051118, June 2021. Publisher: Proceedings of the National Academy of Sciences.

11. L. M. Chau and M. A. D. Goodisman. Gene duplication and the evolution of phenotypic diversity in insect societies. *Evolution; International Journal of Organic Evolution*, 71(12):2871–2884, Dec. 2017.

12. R. H. Crozier and P. Pamilo. *Evolution of social insect colonies: sex allocation and kin selection*. Oxford series in ecology and evolution. Oxford University Press, Oxford ; New York, 1996.

13. E. Dohmen, S. Klasberg, E. Bornberg-Bauer, S. Perrey, and C. Kemena. The modular nature of protein evolution: domain rearrangement rates across eukaryotic life. *BMC Evolutionary Biology*, 20(1):30, Feb. 2020.

14. H. Duan, J. B. Skeath, and H. T. Nguyen. Drosophila Lame duck, a novel member of the Gli superfamily, acts as a key regulator of myogenesis by controlling fusion-competent myoblast development. *Development (Cambridge, England)*, 128(22):4489–4500, Nov. 2001.

15. D. M. Emms and S. Kelly. OrthoFinder: phylogenetic orthology inference for comparative genomics. preprint, Bioinformatics, Nov. 2018.

16. B. Fouks, P. Brand, H. N. Nguyen, J. Herman, F. Camara, D. Ence, D. E. Hagen, K. J. Hoff, S. Nachweide, L. Romoth, K. K. O. Walden, R. Guigo, M. Stanke, G. Narzisi, M. Yandell, H. M. Robertson, N. Koeniger, P. Chantawannakul, M. C. Schatz, K. C. Worley, G. E. Robinson, C. G. Elsik, and O. Rueppell. The genomic basis of evolutionary differentiation among honey bees. *Genome Research*, 31(7):1203–1215, July 2021. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

17. B. Fouks, M. C. Harrison, A. A. Mikhailova, E. Marchal, S. English, M. Carruthers, E. C. Jennings, E. L. Chiamaka, R. A. Frigard, M. Pippel, G. M. Attardo, J. B. Benoit,

E. Bornberg-Bauer, and S. S. Tobe. Live-bearing cockroach genome reveals convergent evolutionary mechanisms linked to viviparity in insects and beyond. *iScience*, 26(10):107832, Oct. 2023.

18. R. Fropf, T. C. Tubon, and J. C. Yin. Nuclear gating of a Drosophila dCREB2 activator is involved in memory formation. *Neurobiology of Learning and Memory*, 106:258–267, Nov. 2013.

19. J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman. Visualization in Bayesian Workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(2):389–402, Feb. 2019.

20. S. Gautam, S. McKenzie, J. Katzke, F. Hita Garcia, S. Yamamoto, and E. P. Economo. Evolution of odorant receptor repertoires across Hymenoptera is not linked to the evolution of eusociality. *Proceedings of the Royal Society B: Biological Sciences*, 291(2031):20241280, Sept. 2024. Publisher: Royal Society.

21. K. M. Glastad, S. V. Arsenault, K. L. Vertacnik, S. M. Geib, S. Kay, B. N. Danforth, S. M. Rehan, C. R. Linnen, S. D. Kocher, and B. G. Hunt. Variation in DNA Methylation Is Not Consistently Reflected by Sociality in Hymenoptera. *Genome Biology and Evolution*, 9(6):1687–1698, June 2017.

22. K. M. Glastad, K. Gokhale, J. Liebig, and M. A. D. Goodisman. The caste- and sex-specific DNA methylome of the termite Zootermopsis nevadensis. *Scientific Reports*, 6(1):37110, Nov. 2016.

23. J. Gospocic, K. M. Glastad, L. Sheng, E. J. Shields, S. L. Berger, and R. Bonasio. Kr-h1 maintains distinct caste-specific neurotranscriptomes in response to socially regulated hormones. *Cell*, 184(23):5807–5823.e14, Nov. 2021.

24. S. Gupta, J. A. Stamatoyannopoulos, T. L. Bailey, and W. S. Noble. Quantifying similarity between motifs. *Genome Biology*, 8(2):R24, Feb. 2007.

25. M. C. Harrison, E. Dohmen, S. George, D. Sillam-Dussès, S. Séité, and M. Vasseur-Cognet. Complex regulatory role of DNA methylation in caste- and age-specific expression of a termite. *Open Biology*, 12(7):220047, July 2022.

26. M. C. Harrison, E. Jongepier, H. M. Robertson, N. Arning, T. Bitard-Feildel, H. Chao, C. P. Childers, H. Dinh, H. Doddapaneni, S. Dugan, J. Gowin, C. Greiner, Y. Han, H. Hu, D. S. T. Hughes, A.-K. Huylmans, C. Kemena, L. P. M. Kremer, S. L. Lee, A. Lopez-Ezquerra, L. Mallet, J. M. Monroy-Kuhn, A. Moser, S. C. Murali, D. M. Muzny, S. Otani, M.-D. Piulachs, M. Poelchau, J. Qu, F. Schaub, A. Wada-Katsumata, K. C. Worley, Q. Xie, G. Ylla, M. Poulsen, R. A. Gibbs, C. Schal, S. Richards, X. Belles, J. Korb, and E. Bornberg-Bauer. Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecology & Evolution*, 2(3):557–566, Mar. 2018. Publisher: Nature Publishing Group.

27. F. Hartig. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models, Aug. 2016. Institution: Comprehensive R Archive Network Pages: 0.4.6.

28. K.-H. Hung, M. Titus, S.-C. Chiang, and W. E. Stumph. A Map of Drosophila melanogaster Small Nuclear RNA-activating Protein Complex (DmSNAPc) Domains Involved in Subunit Assembly and DNA Binding. *Journal of Biological Chemistry*, 284(34):22568–22579, Aug. 2009.

29. B. Hölldobler and E. O. Wilson. *The superorganism: the beauty, elegance, and strangeness of insect societies*. W.W. Norton, London, 2009.

30. U. Irion, M. Leptin, K. Siller, S. Fuerstenberg, Y. Cai, C. Q. Doe, W. Chia, and X. Yang. Abstrakt, a DEAD box protein, regulates Insc levels and asymmetric division of neural and mesodermal progenitors. *Current biology: CB*, 14(2):138–144, Jan. 2004.

31. S. Itakura, Y. Yoshikawa, Y. Togami, and K. Umezawa. Draft genome sequence of the termite, Coptotermes formosanus: Genetic insights into the pyruvate dehydrogenase complex of the termite. *Journal of Asia-Pacific Entomology*, 23:666–674, Aug. 2020. ADS Bibcode: 2020JAsPE..23..666I.

32. J. U. Jarvis, M. O'Riain, N. C. Bennett, and P. W. Sherman. Mammalian eusociality: a family affair. *Trends in Ecology & Evolution*, 9(2):47–51, Feb. 1994.

33. M. Kaspar, M. Schneider, W. Chia, and T. Klein. Klumpfuss is involved in the determination of sensory organ precursors in Drosophila. *Developmental Biology*, 324(2):177–191, Dec. 2008.

34. K. Katoh, G. Asimenos, and H. Toh. Multiple Alignment of DNA Sequences with MAFFT. In D. Posada, editor, *Bioinformatics for DNA Sequence Analysis*, pages 39–64. Humana Press, Totowa, NJ, 2009.

35. R. Kolde. pheatmap: Pretty Heatmaps, Dec. 2015.

36. E. Kristiansson, M. Thorsen, M. J. Tamás, and O. Nerman. Evolutionary Forces Act on Promoter Length: Identification of Enriched Cis-Regulatory Elements. *Molecular Biology and Evolution*, 26(6):1299–1307, June 2009.

37. R. C. Lacy. The Evolution of Eusociality in Termites: A Haplodiploid Analogy? *The American Naturalist*, 116(3):449–451, Sept. 1980.

38. D. S. Latchman. Transcription factors: An overview. *The International Journal of Biochemistry & Cell Biology*, 29(12):1305–1312, Dec. 1997.

39. F. Legendre and F. L. Condamine. When Darwin's Special Difficulty Promotes Diversification in Insects. *Systematic Biology*, 67(5):873–887, Sept. 2018.

40. F. Legendre and P. Grandcolas. The evolution of sociality in termites from cockroaches: A taxonomic and phylogenetic perspective. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 330(5):279–287, July 2018.

41. H.-D. Li, C.-X. Lin, and J. Zheng. GTFtools: a software package for analyzing various features of gene models. *Bioinformatics*, 38(20):4806–4808, Oct. 2022.

42. S. Li, S. Zhu, Q. Jia, D. Yuan, C. Ren, K. Li, S. Liu, Y. Cui, H. Zhao, Y. Cao, G. Fang, D. Li, X. Zhao, J. Zhang, Q. Yue, Y. Fan, X. Yu, Q. Feng, and S. Zhan. The genomic and functional landscapes of developmental plasticity in the American cockroach. *Nature Communications*, 9(1):1008, Mar. 2018.

43. S. Li, S. Zhu, Q. Jia, D. Yuan, C. Ren, K. Li, S. Liu, Y. Cui, H. Zhao, Y. Cao, G. Fang, D. Li, X. Zhao, J. Zhang, Q. Yue, Y. Fan, X. Yu, Q. Feng, and S. Zhan. The genomic and functional landscapes of developmental plasticity in the American cockroach. *Nature Communications*, 9(1):1008, Mar. 2018.

44. F. Lyko, S. Foret, R. Kucharski, S. Wolf, C. Falckenhayn, and R. Maleszka. The honey bee epigenomes: differential methylation of brain DNA in queens and workers., The Honey Bee

Epigenomes: Differential Methylation of Brain DNA in Queens and Workers. *PLoS biology, PLoS Biology*, 8, 8(11, 11):e1000506–e1000506, Nov. 2010.

45. J. Maynard Smith and E. Szathmáry. *The major transitions in evolution.* Oxford Univ. Press, Oxford, reprinted edition, 2010.

46. F. K. Mendes, D. Vanderpool, B. Fulton, and M. W. Hahn. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*, 36(22-23):5516–5518, Apr. 2021.

47. Z. Meng, S. Jiang, X. Chen, and C. Lei. The complete mitochondrial genome of fungus-growing termite, *Macrotermes natalensis* (Isoptera: Macrotermitinae). *Mitochondrial DNA Part A*, 27(3):1728–1729, May 2016.

48. C. Merigliano, A. Marzio, F. Renda, M. P. Somma, M. Gatti, and F. Vernì. A Role for the Twins Protein Phosphatase (PP2A-B55) in the Maintenance of *Drosophila* Genome Integrity. *Genetics*, 205(3):1151–1167, Mar. 2017.

49. A. A. Mikhailova, E. Dohmen, and M. C. Harrison. Major changes in domain arrangements are associated with the evolution of termites. *Journal of Evolutionary Biology*, 37(7):758–769, July 2024.

50. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, and R. Lanfear. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5):1530–1534, May 2020.

51. F. Miozzo, E. P. Valencia-Alarcón, L. Stickley, M. Majcin Dorcikova, F. Petrelli, D. Tas, N. Loncle, I. Nikonenko, P. Bou Dib, and E. Nagoshi. Maintenance of mitochondrial integrity in midbrain dopaminergic neurons governed by a conserved developmental transcription factor. *Nature Communications*, 13(1):1426, Mar. 2022.

52. J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn, and A. Bateman. Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49(D1):D412–D419, Jan. 2021.

53. B. Murrell, J. O. Wertheim, S. Moola, T. Weighill, K. Scheffler, and S. L. K. Pond. Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLOS Genetics*, 8(7):e1002764, July 2012. Publisher: Public Library of Science.

54. K. Oguchi, K. Maekawa, and T. Miura. Regulatory Mechanisms Underlying the Differentiation of Neotenic Reproductives in Termites: Partial Release From Arrested Development. *Frontiers in Ecology and Evolution*, 9, Feb. 2021. Publisher: Frontiers.

55. T. Ohde, T. Mito, and T. Niimi. A hemimetabolous wing development suggests the wing origin from lateral tergum of a wingless ancestor. *Nature Communications*, 13(1):979, Feb. 2022. Publisher: Nature Publishing Group.

56. E. Paradis and K. Schliep. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3):526–528, Feb. 2019.

57. R. S. Peters, L. Krogmann, C. Mayer, A. Donath, S. Gunkel, K. Meusemann, A. Kozlov, L. Podsiadlowski, M. Petersen, R. Lanfear, P. A. Diez, J. Heraty, K. M. Kjer, S. Klopfstein, R. Meier, C. Polidori, T. Schmitt, S. Liu, X. Zhou, T. Wappler, J. Rust, B. Misof, and O. Niehuis. Evolutionary History of the Hymenoptera. *Current Biology*, 27(7):1013–1018, Apr. 2017. Publisher: Elsevier.

58. M. Poulsen, H. Hu, C. Li, Z. Chen, L. Xu, S. Otani, S. Nygaard, T. Nobre, S. Klaubauf, P. M. Schindler, F. Hauser, H. Pan, Z. Yang, A. S. M. Sonnenberg, Z. W. de Beer, Y. Zhang, M. J. Wingfield, C. J. P. Grimmelikhuijzen, R. P. de Vries, J. Korb, D. K. Aanen, J. Wang, J. J. Boomsma, and G. Zhang. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proceedings of the National Academy of Sciences*, 111(40):14500–14505, Oct. 2014. Publisher: Proceedings of the National Academy of Sciences.

59. B. C. Pyenson and S. M. Rehan. Gene regulation supporting sociality shared across lineages and variation in complexity. *Genome*, 67(4):99–108, Apr. 2024. Publisher: NRC Research Press.

60. A. R. Quinlan and I. M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, Mar. 2010.

61. R Development Core Team. *a language and environment for statistical computing: reference index*. R Foundation for Statistical Computing, Vienna, 2010.

62. S. M. Rehan and A. L. Toth. Climbing the social ladder: the molecular evolution of sociality. *Trends in Ecology & Evolution*, 30(7):426–433, July 2015. Publisher: Elsevier.

63. L. Revely, P. Eggleton, R. Clement, C. Zhou, and T. R. Bishop. The diversity of social complexity in termites. *Proceedings of the Royal Society B: Biological Sciences*, 291(2024):20232791, June 2024.

64. G. E. Robinson and E. L. Vargo. Juvenile hormone in adult eusocial Hymenoptera: gonadotropin and behavioral pacemaker. *Archives of Insect Biochemistry and Physiology*, 35(4):559–583, 1997.

65. D. Rotenberg, A. A. Baumann, S. Ben-Mahmoud, O. Christiaens, W. Dermauw, P. Ioannidis, C. G. C. Jacobs, I. M. Vargas Jentzsch, J. E. Oliver, M. F. Poelchau, S. P. Rajarapu, D. J. Schneweis, S. Snoeck, C. N. T. Taning, D. Wei, S. M. K. Widana Gamage, D. S. T. Hughes, S. C. Murali, S. T. Bailey, N. E. Bejerman, C. J. Holmes, E. C. Jennings, A. J. Rosendale, A. Rosselot, K. Hervey, B. A. Schneweis, S. Cheng, C. Childers, F. A. Simão, R. G. Dietzgen, H. Chao, H. Dinh, H. V. Doddapaneni, S. Dugan, Y. Han, S. L. Lee, D. M. Muzny, J. Qu, K. C. Worley, J. B. Benoit, M. Friedrich, J. W. Jones, K. A. Panfilio, Y. Park, H. M. Robertson, G. Smagghe, D. E. Ullman, M. van der Zee, T. Van Leeuwen, J. A. Veenstra, R. M. Waterhouse, M. T. Weirauch, J. H. Werren, A. E. Whitfield, E. M. Zdobnov, R. A. Gibbs, and S. Richards. Genome-enabled insights into the biology of thrips as crop pests. *BMC Biology*, 18(1):142, Oct. 2020.

66. L. M. Roth and C. A. Nalepa. *Cockroaches Ecology, Behavior, and Natural History*. Johns Hopkins University Press, Baltimore, 2009. OCLC: 956771679.

67. C. Schertel, M. Albarca, C. Rockel-Bauer, N. W. Kelley, J. Bischof, K. Hens, E. Van Nimwegen, K. Basler, and B. Deplancke. A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating *Drosophila* wing development. *Genome Research*, 25(4):514–523, Apr. 2015.

68. I. Sela, H. Ashkenazy, K. Katoh, and T. Pupko. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*, 43(W1):W7–W14, July 2015.

69. S. Shigenobu, Y. Hayashi, D. Watanabe, G. Tokuda, M. Y. Hojo, K. Toga, R. Saiki, H. Yaguchi, Y. Masuoka, R. Suzuki, S. Suzuki, M. Kimura, M. Matsunami, Y. Sugime,

K. Oguchi, T. Niimi, H. Gotoh, M. K. Hojo, S. Miyazaki, A. Toyoda, T. Miura, and K. Maekawa. Genomic and transcriptomic analyses of the subterranean termite Reticulitermes speratus: Gene duplication facilitates social evolution. *Proceedings of the National Academy of Sciences*, 119(3):e2110361119, Jan. 2022. Publisher: Proceedings of the National Academy of Sciences.

70. D. F. Simola, L. Wissler, G. Donahue, R. M. Waterhouse, M. Helmkampf, J. Roux, S. Nygaard, K. M. Glastad, D. E. Hagen, L. Viljakainen, J. T. Reese, B. G. Hunt, D. Graur, E. Elhaik, E. V. Kriventseva, J. Wen, B. J. Parker, E. Cash, E. Privman, C. P. Childers, M. C. Muñoz-Torres, J. J. Boomsma, E. Bornberg-Bauer, C. R. Currie, C. G. Elsik, G. Suen, M. A. D. Goodisman, L. Keller, J. Liebig, A. Rawls, D. Reinberg, C. D. Smith, C. R. Smith, N. Tsutsui, Y. Wurm, E. M. Zdobnov, S. L. Berger, and J. Gadau. Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Research*, 23(8):1235–1247, Aug. 2013. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

71. M. D. Smith, J. O. Wertheim, S. Weaver, B. Murrell, K. Scheffler, and S. L. Kosakovsky Pond. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, 32(5):1342–1353, May 2015.

72. M. Sonnenfeld, M. Ward, G. Nystrom, J. Mosher, S. Stahl, and S. Crews. The *Drosophila tango* gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development. *Development*, 124(22):4571–4582, Nov. 1997.

73. E. L. Sonnhammer, S. R. Eddy, and R. Durbin. Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins*, 28(3):405–420, July 1997.

74. M. Suyama, D. Torrents, and P. Bork. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34(suppl_2):W609–W612, July 2006.

75. E. Tasaki, M. Takata, and K. Matsuura. Why and how do termite kings and queens live so long? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1823):20190740, Apr. 2021.

76. I. Tegeder, K. Thiel, S. Erkek, P. D. Johann, J. Berlandi, V. Thatikonda, M. C. Frühwald, M. Kool, A. Jeibmann, and M. Hasselblatt. Functional relevance of genes predicted to be affected by epigenetic alterations in atypical teratoid/rhabdoid tumors. *Journal of Neuro-Oncology*, 141(1):43–55, Jan. 2019.

77. N. Terrapon, C. Li, H. M. Robertson, L. Ji, X. Meng, W. Booth, Z. Chen, C. P. Childers, K. M. Glastad, K. Gokhale, J. Gowin, W. Gronenberg, R. A. Hermansen, H. Hu, B. G. Hunt, A. K. Huylmans, S. M. S. Khalil, R. D. Mitchell, M. C. Munoz-Torres, J. A. Mustard, H. Pan, J. T. Reese, M. E. Scharf, F. Sun, H. Vogel, J. Xiao, W. Yang, Z. Yang, Z. Yang, J. Zhou, J. Zhu, C. S. Brent, C. G. Elsik, M. A. D. Goodisman, D. A. Liberles, R. M. Roe, E. L. Vargo, A. Vilcinskas, J. Wang, E. Bornberg-Bauer, J. Korb, G. Zhang, and J. Liebig. Molecular traces of alternative social organization in a termite genome. *Nature Communications*, 5(1):3636, May 2014. Publisher: Nature Publishing Group.

78. L. Tsuda, R. Nagaraj, S. Zipursky, and U. Banerjee. An EGFR/Ebi/Sno Pathway Promotes Delta Expression by Inactivating Su(H)/SMRTER Repression during Inductive Notch Signaling. *Cell*, 110(5):625–637, Sept. 2002.

79. A. Vehtari, A. Gelman, and J. Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432, Sept. 2017.

80. J. O. Wertheim, B. Murrell, M. D. Smith, S. L. Kosakovsky Pond, and K. Scheffler. RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. *Molecular Biology and Evolution*, 32(3):820–832, Mar. 2015.

81. H. Wickham. The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software*, 40(1), 2011.

82. H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Use R! Springer International Publishing : Imprint: Springer, Cham, 2nd ed. 2016 edition, 2016.

83. X. Zhou, A. Rokas, S. L. Berger, J. Liebig, A. Ray, and L. J. Zwiebel. Chemoreceptor Evolution in Hymenoptera and Its Implications for the Evolution of Eusociality. *Genome Biology and Evolution*, 7(8):2407–2416, Aug. 2015.

Supplementary materials

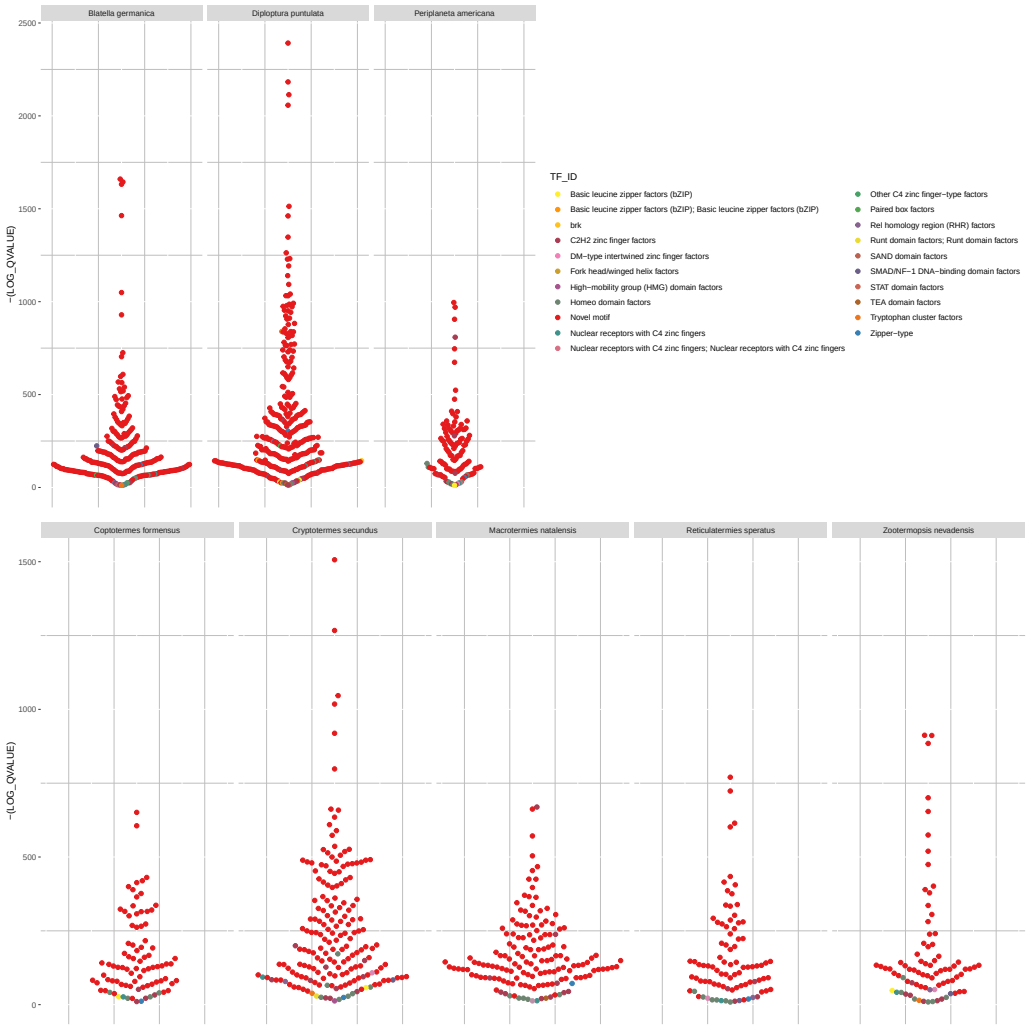

Figure 5. Figure showing the number of enriched motifs identified using SEA. The number of known TF families are for *B. germanica* : C2H2 zinc finger factors High-mobility group 8, High mobility group (HMG) domain factors 1, Homeo domain factors 4, Nuclear receptors with C4 zinc fingers 2,

Model	Expected log predictive density	Standard error of component-wise differences
Unique domains model and social index modifier with DNA binding domains	0	0
Unique domains model and social index modifier with activating domains	−4.2	5.0
Unique domains model and social index modifier with both domain types	−6.4	4.7
Unique domains model and no modifier	−8.4	5.8

Table 1. Table showing the LOO comparison between phylogenetic models using the total number of unique DNA binding domains and total number of unique activating domains as predictors.

Model	Expected log predictive density	standard error of component-wise differences
Total no domains and modifier with social index	0	0
Total no activating domains responding with social index	−1.7	6.2
Total no binding domains domains responding with social index	−6.5	7.4
Total no domains and no modifier	−5.4	7.4
Total no binding domains domains responding with social index	−10.8	8.4

Table 2. Table showing the LOO comparison between phylogenetic models using the total number of activating and DNA binding domains as a predictor.