



Research



Cite this article: Ewart KM, Ho SYW, Chowdhury A-A, Jaya FR, Kinjo Y, Bennett J, Bourguignon T, Rose HA, Lo N. 2024 Pervasive relaxed selection in termite genomes.

Proc. R. Soc. B **291**: 20232439.

<https://doi.org/10.1098/rspb.2023.2439>

Received: 30 October 2023

Accepted: 2 May 2024

Subject Category:

Evolution

Subject Areas:

evolution, genomics

Keywords:

Blattodea, effective population size, eusociality, indirect selection, phylogenomics, relaxed selection

Authors for correspondence:

Kyle M. Ewart

e-mail: kyle.ewart@sydney.edu.au

Nathan Lo

e-mail: nathan.lo@sydney.edu.au

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.7226605>.

Pervasive relaxed selection in termite genomes

Kyle M. Ewart¹, Simon Y. W. Ho¹, Al-Aabid Chowdhury¹, Frederick R. Jaya², Yukihiro Kinjo^{3,4}, Juno Bennett¹, Thomas Bourguignon³, Harley A. Rose¹ and Nathan Lo¹

¹School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales, Australia

²Ecology & Evolution, Research School of Biology, Australian National University, Acton, Australian Capital Territory, Australia

³Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

⁴Okinawa International University, Okinawa, Japan

id KME, 0000-0002-0871-3369; FRJ, 0000-0002-4019-7026; TB, 0000-0002-4035-8977; NL, 0000-0003-2176-2840

Genetic changes that enabled the evolution of eusociality have long captivated biologists. More recently, attention has focussed on the consequences of eusociality on genome evolution. Studies have reported higher molecular evolutionary rates in eusocial hymenopteran insects compared with their solitary relatives. To investigate the genomic consequences of eusociality in termites, we analysed nine genomes, including newly sequenced genomes from three non-eusocial cockroaches. Using a phylogenomic approach, we found that termite genomes have experienced lower rates of synonymous substitutions than those of cockroaches, possibly as a result of longer generation times. We identified higher rates of non-synonymous substitutions in termite genomes than in cockroach genomes, and identified pervasive relaxed selection in the former (24–31% of the genes analysed) compared with the latter (2–4%). We infer that this is due to reductions in effective population size, rather than gene-specific effects (e.g. indirect selection of caste-biased genes). We found no obvious signature of increased genetic load in termites, and postulate efficient purging of deleterious alleles at the colony level. Additionally, we identified genomic adaptations that may underpin caste differentiation, such as genes involved in post-translational modifications. Our results provide insights into the evolution of termites and the genomic consequences of eusociality more broadly.

1. Introduction

Eusocial insects represent only approximately 2% of all insect species but have an enormous ecological impact, representing over half of all insect biomass [1]. Eusocial insect colonies comprise highly specialized castes, whereby some members of the colony reproduce while the workers and soldiers focus on tasks such as foraging, defence, brood care and nest construction [2]. These considerably different behaviours arise from the same genome. The behavioural and genetic changes that enabled the evolution of eusociality, and that have facilitated the ecological dominance of these insects, have long interested biologists.

Many studies on the evolution of eusociality have focussed on the genomic features that underpin this trait. However, in recent years, the consequences of eusociality on genome evolution have also become the subject of active research [3–6]. A key parameter of interest is the molecular evolutionary rates of eusocial organisms compared with their solitary relatives [7,8]. Two main hypotheses have been put forward in relation to the consequences of eusociality on evolutionary rates. The first hypothesis posits that the presence of very few reproductive individuals within colonies should result in eusocial insects having small effective population sizes (N_e). This would be expected to increase

Table 1. Genome assembly and annotation metrics. Ab initio predictions that had no BLAST hits were not included in the numbers of annotated genes.

species	genome length (Gb)	N50	no. of contigs	complete BUSCOs (%)	fragmented BUSCOs (%)	missing BUSCOs (%)	duplicated BUSCOs (%)	no. of annotated genes
<i>G. dilatatus</i>	1.70	3.2 Mb	84 284	93.9	4.8	1.3	0.7	30 006
<i>N. hanni</i>	1.68	11.4 kb	270 019	69.0	19.5	11.5	0.7	38 314
<i>P. cribrata</i>	1.78	298 kb	204 297	85.0	10.9	4.1	1.8	39 852

the strength of genetic drift in eusocial lineages, leading to higher evolutionary rates (e.g. through weaker purifying selection) [3,5,9–12] (but see [13]). The second hypothesis states that the specific expression of genes in different castes and the reduction of multi-caste pleiotropy should result in distinct selection pressures acting on these genes for the different castes (i.e. ‘indirect selection’), leading to the reduced efficacy of selection [6,7,14]. For example, selection on worker-biased genes is largely indirect in reproductive individuals, such that selection may be weaker in worker-biased genes than in queen-biased genes. Moreover, genes under relaxed selection in eusocial ancestors might more readily undergo caste-biased expression. Under this scenario, genes that are caste-specific in eusocial species might have elevated evolutionary rates in both eusocial and non-eusocial taxa [8].

Most studies on the consequences of eusociality on genome evolution have focussed on hymenopterans, with relatively little attention given to other eusocial insects. These studies have identified elevated rates of genome-wide evolution [3], as well as elevated rates of evolution in differentially expressed caste-specific genes [7,8], supporting both of the aforementioned hypotheses. However, the relative contributions of reduced N_e and differential gene expression to these elevated rates remain unclear. Further, there has been limited use of phylogenomic approaches to investigate rates of genomic evolution in eusocial insects and their relatives, despite the availability of methods for examining differing molecular evolutionary forces across large numbers of loci. Phylogenomic methods can help to elucidate the timing of shifts in genomic rates, characterize signatures of selection associated with rate changes and identify the specific genes affected.

An important eusocial insect group with significant ecological impacts is the termites. Their evolution of eusociality from cockroach ancestors [15–19], and their diploid chromosome make-up (in contrast to the haplodiploid eusocial species in Hymenoptera), makes them a valuable comparative model system. However, past examinations of genome-wide molecular evolution in termites have typically been based on a single termite representative [3,5], or have lacked comparisons with cockroach relatives [6]. Although a number of termite genomes have been sequenced [20–23], genomes are available for only two cockroach species, the pests *Periplaneta americana* (Blattidae [24]) and *Blattella germanica* (Ectobiidae [21]). A lack of genomic resources hinders our ability to perform in-depth comparisons of termites and cockroaches to examine the consequences of eusociality on genome evolution.

Here, we facilitate comparative phylogenomic analyses of Blattodea by sequencing the genomes of three Australian species from Blaberidae, the most speciose cockroach family. *Panesthia cribrata* (subfamily Panesthiinae) nests in and feeds on rotting wood, but displays minimal social behaviour. *Geoscaphus dilatatus* and *Neogeoscaphus hanni* (subfamily Geoscaphinae) feed on dried leaf litter and form burrows in the soil up to a metre deep, in which offspring remain with their mothers for several months. The inclusion of *P. cribrata* in our analyses allowed us to control for the effects of wood nesting and wood feeding on genome evolution, in the absence of eusocial behaviour.

We use these new data in conjunction with previously generated genomes from Blattodea to achieve three main objectives: (1) increase our understanding of patterns of selection across termite and cockroach genomes; (2) to evaluate factors that have been posited to influence molecular evolutionary rates in eusocial organisms, such as a reduced N_e and indirect selection of caste-biased genes; and (3) to identify genomic adaptations that underpin the evolution of eusociality in termites. We expect that this study will shed light on the evolution of termites and eusociality more broadly.

2. Results and discussion

(a) Genome assemblies and orthologue assignment

We produced novel de novo genome assemblies for *G. dilatatus*, *N. hanni* and *P. cribrata* using a combination of high-throughput sequencing approaches (electronic supplementary material, table S1; figure S1). The genome quality for *G. dilatatus* was comparable to that of the well-studied pest cockroach species *Periplaneta americana* and *Blattella germanica* (electronic supplementary material, table S2 [21,24]), while genome completeness and continuity for *P. cribrata* and *N. hanni* were lower (table 1). Annotated protein-coding genes were BLASTed against the insect protein database, resulting in an average sequence identity of greater than 65% for each species (see electronic supplementary material, figure S2 for annotation QC results). Most annotated genes (91.7%) were assigned to orthogroups, and relatively few were found to be species-specific (electronic supplementary material, figure S3). Of the 24 455 orthogroups, 1491 were single-copy and present in all species; we used these single-copy orthologues in the subsequent evolutionary analyses.

(b) Pervasive relaxed selection in termite genomes

By investigating patterns of non-synonymous (d_N) and synonymous (d_S) changes, we found evidence of pervasive relaxed selection on the termite terminal branches of the phylogeny. Our analyses included estimates of d_N and d_S for all single-copy

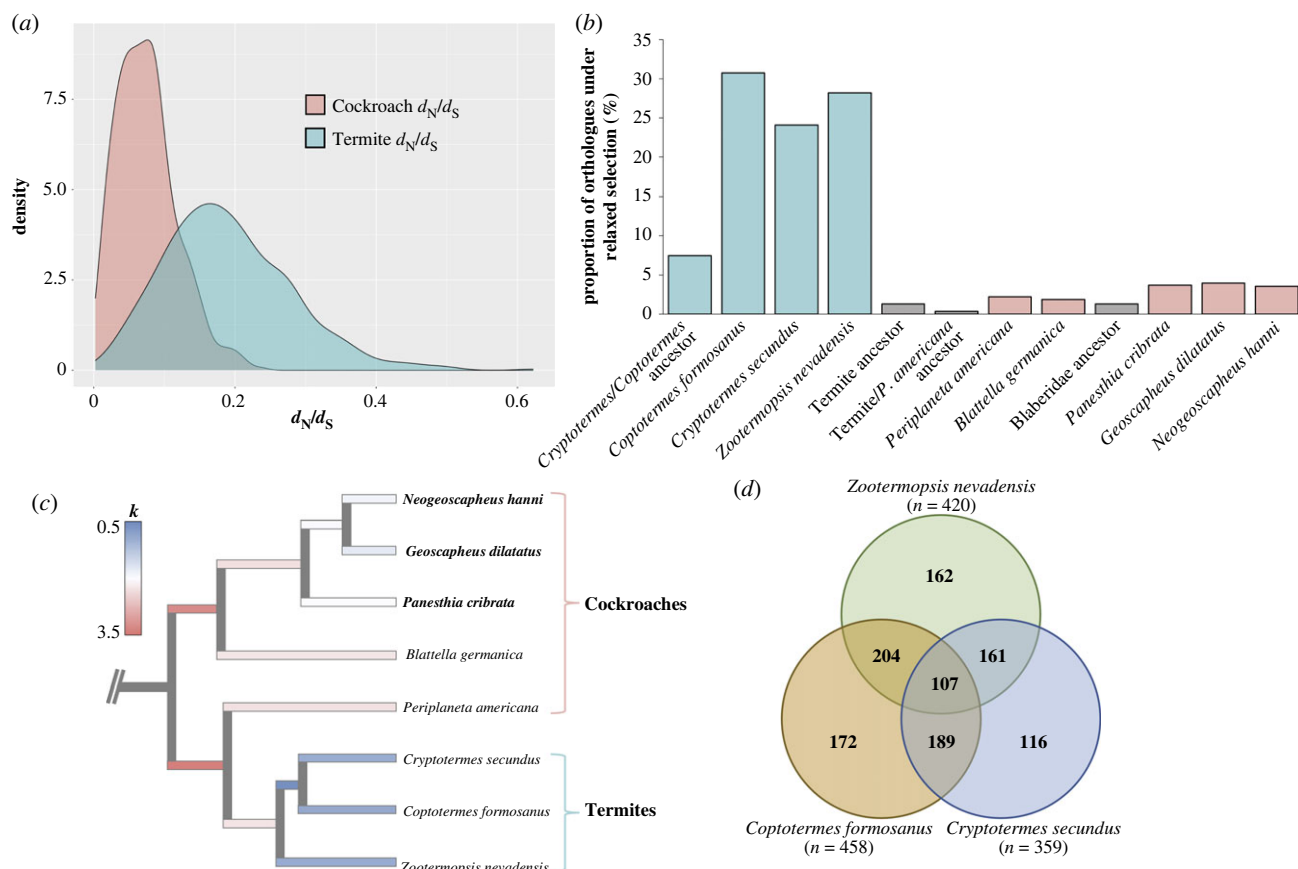


Figure 1. (a) Density plot for the three-ratio d_N/d_S model for the cockroaches and termites. (b) Proportion of genes under significant relaxed selection for each of the focal branches based on the RELAX analyses (after Holm–Bonferroni correction). (c) Phylogenetic tree of analysed termite and cockroach genomes (genomes generated in this study in bold), with branches coloured by their median relaxed selection parameter (k) determined using RELAX, whereby $k < 1$ indicates relaxed selection and $k > 1$ indicates intensified selection (the outgroup has been pruned from the tree). (d) Venn diagram displaying overlapping orthologues under relaxed selection on the termite terminal branches.

orthologues in nine examined taxa: the three newly sequenced cockroach species (*Geoscapheus dilatatus*, *Neogeoscapheus hanni* and *Panesthia cribrata*), two pest cockroach species (*Periplaneta americana* and *Blattella germanica*), three termite species (*Cryptotermes secundus*, *Coptotermes formosanus* and *Zootermopsis nevadensis*), and one outgroup species (the orthopteran *Laupala kohalensis*).

We applied the free-ratio model in CODEML [25], allowing a separate d_N and d_S on each branch, and found that median d_N rates were elevated on the termite branches (excluding the termite stem; median = 4.19×10^{-4}) compared with the cockroach branches (excluding the Blattodea stem; median = 2.85×10^{-4}). Conversely, the d_S rate was found to be lower on the termite branches (median = 2.66×10^{-3}) than on the cockroach branches (median = 3.46×10^{-3}). The median d_N/d_S for the termite branches was 0.149, while that for cockroach branches was 0.081. This median termite d_N/d_S is considerably higher than previously reported estimates for eusocial hymenopteran species based on genomic data such as bees (*Apis*, *Bombus*, and *Tetragonula* spp.; average corrected d_N/d_S = 0.109 [12]; a separate group of Apidae spp.; median d_N/d_S = 0.081 [26]), ants (comprising Formicidae spp.; average corrected d_N/d_S = 0.107 [12]) and wasps (comprising Polistinae and Vespinae spp.; average corrected d_N/d_S = 0.103 [12]), while the median cockroach d_N/d_S falls within the range of estimates for solitary hymenopteran species [12,26]. However, caution should be taken when directly comparing d_N/d_S estimates across studies based on different sets of genes with varying levels of conservation, and which have employed different models/methods to estimate d_N/d_S .

Separate analyses of d_N/d_S utilizing a three-ratio model (allowing a separate ratio for the termite clade) was preferred over a two-ratio model (allowing one ratio for the outgroup and one for the entire cockroach–termite ingroup) for 75.7% of the analysed orthologues ($p < 0.05$; 45.1% after applying a Holm–Bonferroni correction). The distribution of d_N/d_S on termite branches in this analysis was wider and had a much higher mean than that for the cockroach branches (figure 1a). In 90.8% of the analysed orthologues, the termite d_N/d_S was higher than the cockroach d_N/d_S (significantly higher for 73% when applying $p < 0.05$, and 44.5% after applying a Holm–Bonferroni correction).

To investigate further the differences in d_N , d_S and d_N/d_S between termite and cockroach branches, we tested for relaxed selection using the ‘RELAX’ method [27], implemented in HyPhy [28,29]. We detected a large proportion of the single-copy orthologues under relaxed selection on the termite terminal branches (24.09–30.74%) compared with the cockroach terminal branches (1.88–3.96%) (after a Holm–Bonferroni correction for testing multiple branches; figure 1b). When we analysed an audited subset of single-copy orthologues (i.e. gene alignments that were manually inspected and edited if required), we obtained similar proportions (electronic supplementary material, table S3). This scale of relaxed selection is comparable to inferences from a previous study on hymenopteran eusocial species [12], which identified 23.21% (after a conservative Bonferroni correction) of the analysed orthologues to be under relaxed selection (though direct comparison between these studies should be undertaken

with caution). A separate study analysed relatively small sets of genes (13–78 single-copy orthologues [30]) in termites, and found 18–27% of the genes to be under relaxed selection in four of the termite species analysed, including the three termite genera analysed in our study. However, for the other four termite species analysed, lower proportions (0–12%) of the genes were identified as being under relaxed selection (comparisons with cockroaches were not included). The difference in relaxed selection between termites and cockroaches in our analyses becomes more extreme when analysing all termite branches together (72.43% of orthologues analysed) compared with all cockroach branches (0.74%). In relaxed selection analyses, a significant result of $k < 1$ is indicative of a reduction in selection strength, while $k > 1$ indicates an intensification of selection strength. Median k values were found to be the lowest on termite branches, and well below one (figure 1c).

Given that the three termite species in our study form a single monophyletic clade, we note the issue of phylogenetic non-independence in some of our analyses. However, we do not believe that this affects the key findings outlined above. When analysing individual branches, a large proportion of the genes were under relaxed selection in only one or two of the termite terminal branches (figure 1d). Hence, although reductions in N_e are likely to have occurred in the ancestral termite lineage following its emergence approximately 134 mya [19], the genomic consequences of being eusocial are likely to have manifested independently in each of the three analysed termite species since they diverged from one another approximately 97–107 mya [19].

Our analyses investigated a relatively conserved set of single-copy orthologues present among the highly divergent taxa included. The scale of d_N/d_S and relaxed selection likely differs across genes that are faster evolving, species-specific, and/or part of multigene families. We expect that the patterns and causes of relaxed selection in such genes would be similar to those that we have analysed (see below).

(c) Relaxed selection in termite genomes is better explained by reduced N_e rather than indirect selection

We evaluated two main scenarios that may explain the pervasive selection in termite genomes: (1) a reduction in N_e , and (2) indirect selection. A reduction in N_e would have a genome-wide effect, whereby the strength of genetic drift increases for every gene. Conversely, indirect selection would have gene-specific effects. Only caste-specific genes are expected to experience reduced selective constraints, as they are only expressed in a subset of the colony (e.g. worker-biased genes are predominantly expressed by workers, not reproductives). We note that a niche- or functional-specialization scenario would also confer gene-specific effects (elaborated in section ‘d’ below).

To investigate whether indirect selection underpinned signatures of relaxed selection in our data set, we identified and analysed 311 genes in our single-copy orthologue data set that had previously been identified to display either unbiased or caste-biased differential expression [6,31] (noting that caste-biased genes in *Coptotermes formosanus* were not identified in these studies). We found no significant difference between the proportion of caste-biased (33.3% for *Zootermopsis nevadensis* and 24.4% for *Cryptotermes secundus*) and unbiased (25.3% for *Z. nevadensis* and 20% for *C. secundus*) orthogroups under relaxed selection when compared with proportions in the full single-copy orthologue data set (chi-squared tests, $p = 0.29$ for *Zootermopsis nevadensis* and $p = 1$ for *Cryptotermes secundus* caste-biased genes; electronic supplementary material, table S4). Similar patterns were found for genes only expressed in one caste (electronic supplementary material, table S4). Further, if indirect selection were the main driver of these patterns of relaxed selection, we would expect to see considerably fewer genes affected. Previous RNAseq analyses showed that many caste-biased genes are species specific [6] and/or part of multi-gene families [31]. Such genes were not considered in our single-copy orthologue data set and require further investigation.

The pervasive scale of relaxed selection across the termite genomes is explained better by a reduction in N_e than by gene-specific effects resulting from indirect selection. A reduction in N_e increases the intensity of genetic drift [32], and consequently reduces the efficacy of both negative and positive selection [27,33]. Termites are expected to have a very low N_e relative to their census population size due to the complex social structure of their colonies, the small number of reproductive individuals (typically one king and one queen per colony), and their history of inbreeding within colonies [11,34,35]. Indeed, Romiguier *et al.* [3] found that the termite *Reticulitermes grassei* had a large reduction in N_e , even compared with other eusocial species, and postulated that this was due to the strict monogamy seen in this taxon.

Although a reduced N_e will theoretically not directly affect the substitution rate at neutral sites [36], a long-term low N_e has been proposed to lead to the evolution of a higher mutation rate [37]. However, we found that d_S values were lower in termites than in cockroaches. This result can be explained by the considerably longer generation time of termites (5–10 years) compared with cockroaches (typically less than 1 year). Although a longer generation time is expected to increase the mutation rate per generation (due to increased cell divisions of gametes, and greater time for DNA damage to accumulate), the per-year mutation rate is expected to decrease [38,39]. However, a reduction of N_e appears to have counteracted this generation-time effect for non-neutral sites in termites (based on observed d_N rates).

(d) Alternative explanations for relaxed selection in termite genomes

Although reduced N_e is the most likely explanation for the pervasive relaxed selection observed in our study, we are unable to rule out the possibility that indirect selection or other factors contributed to relaxed selection in at least some of the examined orthologues. Although many genes were only under significant relaxed selection on one of the three termite branches (figure 1d), there were more genes under relaxed selection that overlap in all three termite species ($n = 107$) than expected by chance. Hence, some ‘core genes’ might be under relaxed selection due to other non-random processes, such as indirect selection and/or reduced functional importance in termites. There was no evidence that caste-biased genes were undergoing relaxed selection or accelerated evolution prior to becoming caste-biased, as alluded to by previous authors [8,40,41], based on evolutionary rates of the cockroach orthologues (electronic supplementary material, figures S4, S5) and the minimal relaxed selection on the termite ancestral branch (figure 1b,c).

The intensity of selection is proportional to the product of N_e and the selection coefficient s [36]. Therefore, it is possible to observe pervasive relaxed selection without a reduction in N_e , given a change in s that affects many loci. Previous authors have hypothesized that the cooperative interactions of social species can buffer or compensate for the expression of deleterious mutations, particularly in relation to immune responses [30,42–46]. The creation of a benign environment through social cooperation (e.g. termite mounds) and/or buffering for the expression of deleterious alleles (e.g. through food allocation) will lead to a reduction in selection pressures for the genes involved [46]. For example, termite social hygiene behaviours, such as allogrooming, might result in relaxation of selection in certain innate components of their immune system [45].

Although there might have been large-scale shifts in s across termite genomes since divergence from their subsocial cockroach ancestors, we have several reasons to believe that this is not the primary driver of the patterns of relaxed selection observed. First, said patterns do not appear to be associated with specific habitats. The different termite species analysed occupy different niches: *Coptotermes formosanus* constructs nests and forages in the soil; *Zootermopsis nevadensis* lives and feeds in dead dampwood; and *Cryptotermes secundus* lives and feeds in dead drywood. Likewise, the five sampled cockroach species live in varying habitats, from urban environments to wood and soil burrows. Hence, if pervasive relaxed selection were caused by a reduced s due to differing selective pressures across different niches (and not a reduction in N_e), we might expect relaxed selection to be restricted to one or two of the termite species, or observe comparable levels of relaxed selection in some cockroach species (e.g. *Periplaneta americana*, a recently urbanized cockroach species, sister lineage to the termite species in our analysis). Second, if a reduced s across many genes were due to the social benefits of group living (e.g. allogrooming), we might expect a larger overlap of orthologues to be under relaxed selection (figure 1d; but see [30]), and a larger proportion of orthologues under relaxed selection in the termite ancestral branches. Third, given the relatively conserved data set that we analysed (i.e. single-copy orthologues across taxa approximately 350 million years divergent [47]), we expect most genes to be functionally important in termites (i.e. a relatively stable s across the analysed phylogeny). Hence, a reduction in s across greater than 24% of the relatively conserved orthologue data set is unlikely. Fourth, if some of the analysed orthologues were indeed associated with changes in niche or function, we might also expect to see signatures of positive selection in termites; however, this was not the case (figure 2a). Therefore, although we cannot rule out a pervasive reduction of s across the termite genome due to increased social cooperation or niche specialization, or a combination of reduced s and indirect selection, our results suggest that the observed patterns of relaxed selection are dominated by the genome-wide effects of a reduced N_e , rather than gene-specific effects.

(e) No obvious signature of increased genetic load in termites

A long-term small N_e and pervasive relaxed selection could result in an accumulation and fixation of deleterious alleles and a high genetic load [3,33,48,49]. Unexpectedly, however, termites do not appear to have a higher genetic load than cockroaches based on the number of derived putative deleterious substitutions (identified using PROVEAN [50]) in 581 Insecta BUSCOs (i.e. highly conserved genes). Here, we are referring only to ‘realized load’ and not ‘masked load’ [49,51], given that we are comparing amino acid sequences across an order of insects, and not considering heterozygosity. Termites had a higher number of total substitutions per amino acid than the cockroach species (significantly higher for 12/15 pairwise comparisons, and non-significantly higher for 3/15 pairwise comparisons; electronic supplementary material, figure S6; table S5). However, the number of inferred deleterious substitutions per amino acid was significantly lower or similar in termites compared with the cockroach species.

Therefore, despite the patterns of pervasive relaxed selection that we identified, there does not appear to be an obvious signature of increased genetic load in termites compared with cockroaches. How then have termites, an ancient and evolutionarily successful group, succeeded in diversifying and adapting to changing environments? An increased genetic load could have been alleviated through purging, which involves the removal of recessive deleterious alleles exposed by inbreeding. Purging could be particularly effective in termites due to a long-term history of inbreeding and variation in heterozygosity generated by inbreeding [52,53]. Moreover, termites could have experienced purging at the colony level. The survival of dispersing reproductive individuals that attempt to form a new colony relies on the survival of their offspring; colony survival hinges on the worker offspring gathering resources, building the nest structure, and raising offspring. Therefore, if we consider the colony as the basic unit of the population rather than individual termites (as in [34,54–56]), when primary (alate-derived) reproductives disperse and breed, purging could occur by the failure of colonies established by unfit offspring. Indeed, there is evidence that closely related primary reproductives establish colonies less frequently than non-related primary reproductives as a result of inbreeding depression [57] (but see [58–61]). This hypothesis requires further investigation.

The lack of an obvious increased genetic load in termites may also be due to the insensitivity of the methods we applied, particularly given the complexities of delineating benign and deleterious substitutions. Further investigation into the functional effects of mutations in coding regions (e.g. experimental evidence) may help to characterize the extent to which derived mutations are deleterious, and clarify any differences in genetic load in these species [51]. Further, we reiterate issues of phylogenetic non-independence in our data set, though we believe that this has a negligible effect on these genetic load inferences (detailed in the electronic supplementary material).

(f) Positive selection and gene expansions/reductions in termite and cockroach genomes

In our analyses of positive selection that allowed a separate d_N/d_S on terminal branches leading to cockroach and termite taxa, we did not detect substantial positive selection in termite genomes (detailed further in the electronic supplementary material). The number of orthologues under positive selection varied between 1.1% and 9.1% across the terminal branches (figure 2a), with the clade comprising *P. cribrata*, *G. dilatatus*, and *N. hanni* displaying the highest number. Further, no significant differences were found when comparing the proportions of genes under positive selection that had previously been demonstrated to be unbiased or differentially expressed between castes (electronic supplementary material, table S4).

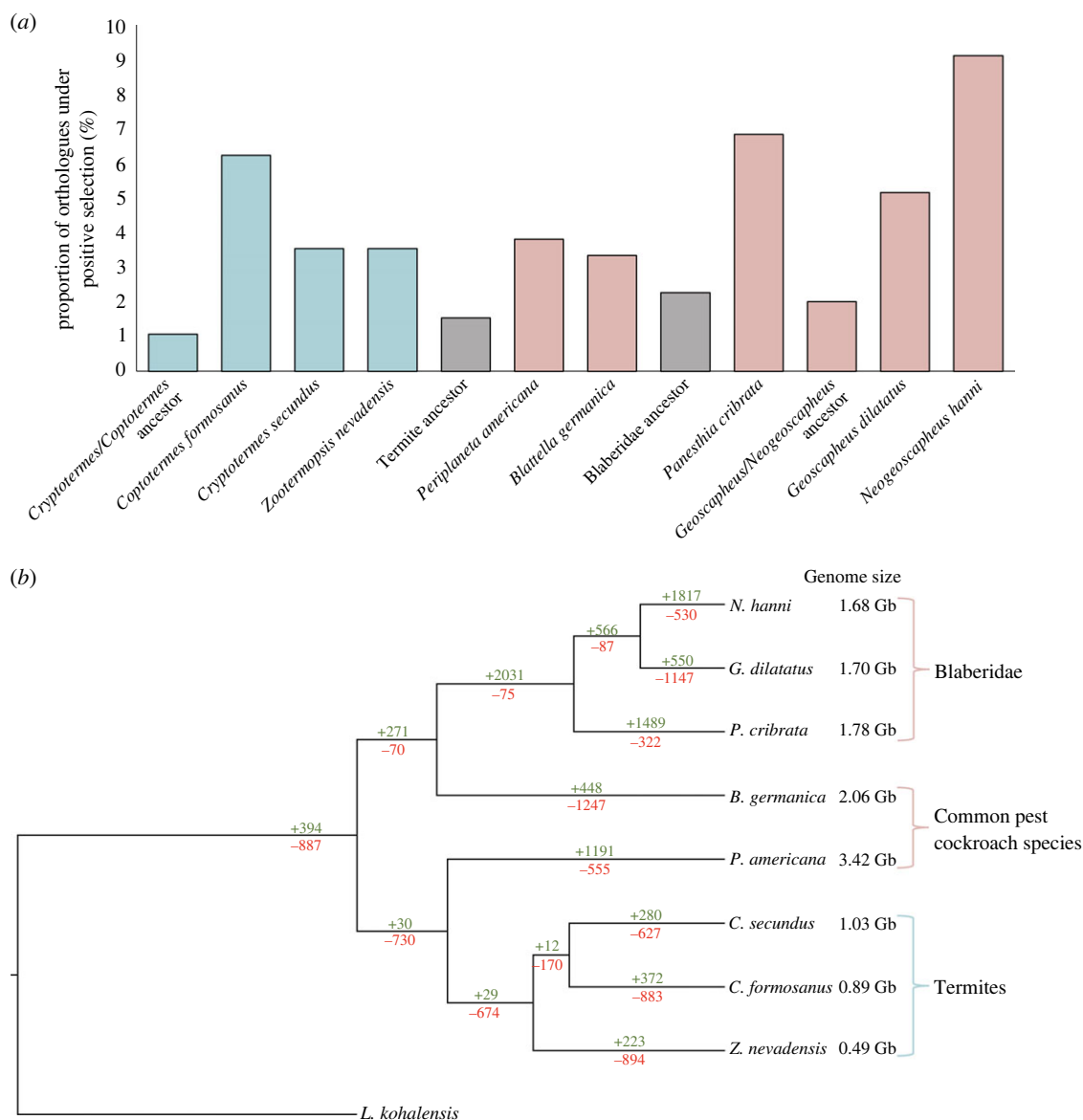


Figure 2. (a) Proportion of orthologues ($n = 1491$) on various branches undergoing positive selection, determined using aBSREL analyses (see electronic supplementary material, table S6, for names of genes undergoing positive selection in termite taxa). (b) Number of gene expansions (green) and contractions (red) on each branch of the cockroach/termite tree (see electronic supplementary material, table S7, for names of genes that had significant expansions/contractions in termite taxa). The number of gene expansions/contractions combined results from the small and large orthogroups analysed.

Previous studies on eusocial species have highlighted the importance of epigenetic mechanisms, particularly DNA methylation, for controlling caste-specific gene expression [21,56,62–67]. Of the 8538 orthogroups that we identified to be displaying gene family size changes (figure 2b; detailed further in the electronic supplementary material), numerous expanding gene families in the termite lineages were related to protein ubiquitination (electronic supplementary material, table S7; e.g. E3 ubiquitin-protein ligase, and Ubiquitin carboxyl-terminal hydrolase 45), an important post-translational mechanism, as well as other gene families involved in gene expression regulation (e.g. Histone H4 transcription factor and KAT8 regulatory NSL complex subunit 3). Protein ubiquitination results in protein degradation or modification. Histones are often ubiquitinated, which subsequently alters chromatin structure and gene function, and may therefore influence patterns of DNA methylation [68,69]. Previous studies have demonstrated that these post-translational modifications and histone structure changes influence the development of castes in the honey bee [70–72] and carpenter ant [73,74]. Evidently, various gene expression mechanisms beyond DNA methylation, such as ubiquitination, could be fundamental to the development of castes, and therefore warrant further research.

3. Concluding remarks

Incorporating additional genome resources has proven valuable in our comparative genomic analyses of termites and cockroaches. Sequencing and analysing genomes from the sister group of termites, representatives of the wood-feeding genus *Cryptocercus*, will help to elucidate the ancestral signals of the relaxed selection that we have identified, and provide insights into the evolutionary origins of caste-biased genes. Further, given that the termites represent only a single transition to eusociality, expanding the analyses to include genomes of other eusocial groups (e.g. eusocial Hymenoptera, rodents and shrimp) could provide insight into convergent patterns of eusocial evolution (e.g. [75]).

Investigating termite populations with varying N_e will help to define the relationship between the pervasive relaxed selection identified and the group's demography. Population genetic analyses will also enable categorization of realized and masked genetic load to clarify whether termites exhibit purging of deleterious alleles, and whether this occurs at the colony level. In addition, characterizing selection coefficients of various genes will help to tease apart genome-wide and gene-specific consequences of evolving eusociality, shedding light on patterns of relaxed selection.

Our findings corroborate several previous studies that suggested an association between eusociality and a reduced N_e , resulting in relaxed selection [3,5,12,38]. Although accurate estimation of N_e is particularly challenging for eusocial species [76], integrating such estimates, or relevant proxies (e.g. [5]), could help to clarify the patterns of relaxed selection that we have inferred, and the association between N_e and the evolution of eusociality. This, in turn, would help to illuminate the extent to which shifts in N_e and the emergence of eusociality have shaped genomic evolution in termites.

4. Material and methods

(a) Sample collection and DNA extraction

Samples of *P. cribrata* were collected from North Manly, NSW, in August 2020 by N. Lo. Samples of *G. dilatatus* were collected from Gilgandra, NSW, in September 2019 by H. A. Rose. Samples of *N. hanni* were collected from Mt Molloy, QLD, in January 2020 by J. A. Walker and N. Lo. Muscle tissue was removed from the leg of one adult female of each species for DNA extraction. DNA was extracted using a phenol/chloroform extraction protocol (detailed in the electronic supplementary material).

(b) DNA sequencing approaches

Three different DNA sequencing approaches were employed to generate the cockroach genomes: linked-read sequencing, chromosomal conformation capture sequencing, and long-read sequencing (electronic supplementary material, table S1). We used two technologies for linked-read sequencing: Transposase Enzyme-Linked Long-read Sequencing (TELL-Seq™ [77]) and single-tube long fragment read (stLFR [78]). These technologies are based on co-barcoding short-read sequences that derive from the same DNA fragment, providing long-range information to enable assembly of long DNA regions.

We used Hi-C sequencing [79], a chromosomal conformation capture sequencing method based on proximity ligation. Hi-C sequencing supports scaffolding of genomic regions that are in close spatial proximity. Hi-C library preparation was performed using the Arima Hi-C Plus kit (Arima, USA) following the manufacturer's protocol for large animal tissue, using the *DpnII* and *HinfI* enzymes. Preliminary sequencing and subsequent analysis using qc3C [80] was performed to estimate the quantity of Hi-C sequencing required.

We utilized the Pacific Biosciences (PacBio) HiFi long-read sequencing technology [81], which uses circular consensus sequencing to generate long-reads that are greater than 99% accurate. The library preparation was performed using the SMRTbell Express Template Prep Kit 2.0 following the manufacturer's instructions (Pacific Biosciences). Additional information on the TELL-Seq, stLFR, Hi-C and PacBio HiFi library preparations can be found in the electronic supplementary material.

(c) Genome assembly and annotation

Different genome assembly pipelines were utilized for the different species depending on the sequencing resources available (electronic supplementary material, figure S1). The first step for all three species was a de novo assembly based on linked-read data, performed using the Supernova assembler v. 2.1.1 [82]. For the genome of *P. cribrata*, we performed a separate de novo assembly based on the PacBio HiFi sequencing data using Improved Phased Assembler (IPA) v. 1.3.1 (Pacific Biosciences, 2022), then merged this long-read assembly with the linked-read assembly using *quickmerge* [83]. For the genome of *G. dilatatus*, gaps in the Supernova de novo assembly were filled with low-coverage PacBio HiFi data using TGS-GapCloser v. 1.1.1 [84], then the assembly was scaffolded with Hi-C sequence data using SALSA2 [85,86]. All three genome assemblies were annotated using FGENESH++ v. 7.2.2 [87,88], carried out on the Nimbus cloud service (provided by the Pawsey Supercomputing Centre). We generated various statistics to assess assembly continuity using QUAST v. 4.3 [89], and assessed assembly completeness using BUSCO v. 4.0.6 (utilizing the Insecta lineage data set [90–93]). Additional information on genome assembly and annotation methods can be found in the electronic supplementary material.

(d) Orthology assignment

Proteomes of these three newly generated blaberid annotated genomes were combined with proteomes from six other species extracted from InsectBase 2.0 [94], including: American cockroach (*Periplaneta americana*), German cockroach (*Blattella germanica*), Nevada termite (*Zootermopsis nevadensis*), Formosan termite (*Coptotermes formosanus*), a drywood termite (*Cryptotermes secundus*) and a cricket (*Laupala kohalensis*) as an outgroup (electronic supplementary material, table S2). The cricket proteome was utilized as an outgroup rather than a proteome from a more closely related species (e.g. *Clitarchus hookeri*) due to its relatively high genome completeness. The longest isoform for each gene was extracted using the primary_transcript.py script supplied by OrthoFinder v. 2.4 [95].

Orthologous protein sequences among all nine species were allocated using OrthoFinder with default settings. Resultant orthogroups were annotated based on the most frequent gene annotation within the orthogroup. Single-copy orthologue sequences were identified and aligned using OrthoFinder, implementing the MAFFT sequence alignment method [96]. The corresponding nucleotide coding sequences (CDS) of these single-copy orthologue protein sequences were extracted from each of the respective genomes, and aligned using the codon-aware sequence alignment method supplied by HyPhy v. 2.5.32 [28,29]. These single-copy orthologue CDS were utilized for our subsequent analyses of selection.

(e) Gene expansions and contractions

The gene counts generated using OrthoFinder were used to investigate gene family expansions and contractions under a birth–death model using CAFE v. 5.0 [97]. To prepare the input tree for CAFE, the rooted species tree generated by OrthoFinder was converted into an ultrametric tree using the make_ultrametric.py script supplied by OrthoFinder. Gene families with ≥ 100 genes in any one lineage

were extracted using a CAFE python script (`clade_and_size_filter.py`) and analysed separately, because high variance of gene copy number can lead to non-informative parameter estimates. The orthogroup data sets were analysed with CAFE, implementing a Poisson root frequency distribution (-p), a model to account for genome assembly and annotation errors (-e), and a lambda tree (-y) specifying three separate lambda estimations for the termites, cockroaches, and outgroup (following results from CODEML analyses).

(f) Positive and relaxed selection

We estimated d_N/d_S values for the aforementioned single-copy orthologues extracted from the nine analysed genomes using CODEML, which is part of the PAML package [25]. We used a fixed tree topology, based on previous phylogenetic estimates (figure 2b [18,19,98]), and implemented various d_N/d_S partitions across the tree in separate analyses. First, we implemented the free-ratio model, which estimates a separate d_N and d_S value for each branch across the phylogeny. We ran CODEML, implementing the free-ratio model, three times on 1484 single-copy orthologue alignments (seven alignments were removed due to missingness), then extracted the estimated d_N and d_S values from the run with the highest likelihood score. To calculate the d_N and d_S rates, we scaled these values with branch lengths from chronograms derived from Evangelista *et al.* [19] and Beasley-Hall *et al.* [98].

Second, we implemented two- and three-ratio d_N/d_S models, whereby specific sets of branches are nominated as foreground branches (i.e. the branches 'of interest'), and other branches are nominated as background branches. In the two-ratio model, we set the branch to the outgroup (*Laupala kohalensis*) as the background branch, allowing a single d_N/d_S value to be estimated across Blattodea. In the three-ratio model, we allowed separate d_N/d_S values to be estimated for the termite clade (excluding the termite stem branch) and the cockroach branches (i.e. the rest of Blattodea). We ran CODEML, implementing the two- and three-ratio models, for all 1484 single-copy orthologue alignments. For each orthologue, we used likelihood-ratio tests to evaluate whether separate values for d_N/d_S for the termites and other taxa in Blattodea fit the sequence data better than a single d_N/d_S value for Blattodea.

We investigated positive selection for each gene on different branches using an adaptive branch-site random-effects model (aBSREL) [99], implemented in HyPhy. aBSREL estimates d_N/d_S values and tests whether a proportion of sites have evolved under positive selection for each user-specified test branch in the tree. First, we analysed the four termite branches (i.e. assigned them as the 'foreground' branches) to investigate signatures of positive selection related to the evolution of termites. Second, we analysed a larger subset of branches across the cockroach and termite phylogeny (i.e. 12 branches assigned as 'foreground' branches) to investigate broader patterns of positive selection in the group. Resultant p -values were corrected for multiple branches being tested using the Holm–Bonferroni method (based on the 5% level of significance). We repeated each analysis using three models accounting for multinucleotide mutations: multiple-hits 'Double', multiple hits 'Double + Triple', and the standard model. Analyses implementing the best-fitting model for each orthologue were retained.

To test for genes with evidence of relaxed selection, we applied the RELAX method [27], implemented in HyPhy [28,29], to each alignment separately. RELAX fits three d_N/d_S classes to the phylogeny then tests for relaxed/intensified selection on a user-specified test branch. We ran RELAX for 12 different test branches in separate analyses (figure 1c). Resultant p -values were corrected for multiple branches being tested using the Holm–Bonferroni method (based on the 5% level of significance). To evaluate whether patterns of relaxed selection were due to errors in our automated sequence alignments, we repeated the RELAX analyses on a subset of audited orthologue alignments. We manually inspected and aligned sequences of 105 orthologues, including 20 randomly selected orthologues that were found to be under relaxed selection in all three termite species, 22 that were found to be under relaxed selection in two termite species, and 65 that yielded no evidence of relaxed selection. We subsequently carried out a RELAX analysis on the realigned orthologues, implementing the standard model for multinucleotide mutations. Furthermore, we ran RELAX for all termite branches together (i.e. multiple test branches were considered in a single analysis) and for all cockroach branches together (a Holm–Bonferroni correction was implemented as detailed above).

(g) Analysis of differentially expressed genes

We identified and analysed caste-biased and unbiased genes characterized by Harrison *et al.* [6] and Maekawa *et al.* [31] in our single-copy orthologue data set based on the putative annotations. We tested whether there were significantly more or fewer caste-biased orthogroups (and unbiased orthogroups) under relaxed or positive selection compared with the whole data set using chi-squared tests, implemented in the R package 'stats' [100].

To investigate the evolutionary history of these caste-biased orthogroups (and unbiased orthogroups), we assessed if they were associated with a higher or lower rate of evolution in cockroaches (see [8]). We did this by estimating total tree lengths for the cockroaches for each single-copy orthologue using IQ-TREE v. 2.2.2 [101]. Additional information on estimating tree lengths can be found in the electronic supplementary material. In addition to the tree-length analysis, we investigated the selection signatures of these caste-biased orthogroups in cockroaches, and compared them with the whole data set using chi-squared tests as above.

(h) Measuring the 'genetic load'

To investigate genetic load among the sampled taxa, we analysed amino acid substitutions in BUSCOs from the Insecta lineage data set. These highly conserved genes are expected to be under purifying selection, and most derived alleles are likely to be deleterious [102,103]. Here, we focus on 'realized load' and not 'masked load', as heterozygosity was not considered [49,51].

For this analysis, we extracted BUSCOs from the Insecta lineage data set in each of the sampled taxa using BUSCO v. 5.4.2. We then polarized the amino acid substitutions of the aligned BUSCOs relative to the corresponding BUSCOs from the cricket outgroup. We analysed the derived amino acid substitutions to infer realized genetic load based on three metrics: (1) the total number of derived substitutions per amino acid; (2) the number of putatively deleterious substitutions per amino acid; and (3) the ratio of putatively deleterious to benign substitutions. We used PROVEAN v. 1.1.5 [50] to classify substitutions as either 'benign' (PROVEAN scores > -2.5) or 'deleterious' (PROVEAN scores ≤ -2.5) based on how phylogenetically conserved the allele is among a database of homologous protein sequences [104–106]. Additional information on estimating genetic load can be found in the electronic supplementary material.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. Genomes assemblies and annotations for the three Blaberidae species (*Geoscapheus dilatatus*, *Neogeoscapheus hanni* and *Panesthia cribrata*) are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.sqv9s4n9t> [107]. The genome assemblies are also archived on

GenBank under BioProject PRJNA1065107 (accession numbers JAZBJU000000000, JAZBJV000000000 and JAZBJW000000000 for *Geoscapheus dilatatus*, *Neogeoscapheus hanni* and *Panesthia cribrata*, respectively), and all associated sequence data are archived on the SRA database (BioSample accessions SAMN39450771, SAMN39450772 and SAMN39450773 for *Geoscapheus dilatatus*, *Neogeoscapheus hanni* and *Panesthia cribrata*, respectively).

Additional data are provided in the electronic supplementary material [108].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. K.M.E.: conceptualization, data curation, formal analysis, methodology, writing—original draft, writing—review and editing; S.Y.W.H.: conceptualization, funding acquisition, methodology, writing—review and editing; A.-A.C.: formal analysis, methodology, writing—review and editing; F.R.J.: formal analysis, methodology, writing—review and editing; Y.K.: data curation, resources, writing—review and editing; J.B.: formal analysis, writing—review and editing; T.B.: data curation, resources, writing—review and editing; H.A.R.: data curation, resources, writing—review and editing; N.L.: conceptualization, funding acquisition, methodology, project administration, resources, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This work was supported by the Australian Research Council Discovery Grants FT160100463 and DP220103265.

Acknowledgements. We utilized the Artemis HPC cluster provided by the University of Sydney, the FlashLite HPC cluster provided by the University of Queensland's Research Computing Centre (RCC), and the Nimbus cloud service provided by the Pawsey Supercomputing Centre. We thank Carolyn Hogg and Parice Brandies for advice on high performance computing resources.

References

- Wilson EO. 1990 *Success and dominance in ecosystems: the case of the social insects*. Oldendorf, Germany: Ecology Institute.
- Anderson M. 1984 The evolution of eusociality. *Annu. Rev. Ecol. Evol. Syst.* **15**, 165–189. (doi:10.1146/annurev.es.15.110184.001121)
- Romiguier J *et al.* 2014 Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *J. Evol. Biol.* **27**, 593–603. (doi:10.1111/jeb.12331)
- Rubenstein DR *et al.* 2019 Coevolution of genome architecture and social behavior. *Trends Ecol. Evol.* **34**, 844–855. (doi:10.1016/j.tree.2019.04.011)
- Imrit MA, Dogantzis KA, Harpur BA, Zayed A. 2020 Eusociality influences the strength of negative selection on insect genomes. *Proc. R. Soc. B* **287**, 20201512. (doi:10.1098/rspb.2020.1512)
- Harrison MC, Chernyshova AM, Thompson GJ. 2021 No obvious transcriptome-wide signature of indirect selection in termites. *J. Evol. Biol.* **34**, 403–415. (doi:10.1111/jeb.13749)
- Hunt BG, Wyder S, Elango N, Werren JH, Zdobnov EM, Yi SV, Goodisman MA. 2010 Sociality is linked to rates of protein evolution in a highly social insect. *Mol. Biol. Evol.* **27**, 497–500. (doi:10.1093/molbev/msp225)
- Hunt BG, Ometto L, Wurm Y, Shoemaker D, Yi SV, Keller L, Goodisman MA. 2011 Relaxed selection is a precursor to the evolution of phenotypic plasticity. *Proc. Natl Acad. Sci. USA* **108**, 15 936–15 941. (doi:10.1073/pnas.1104825108)
- Helanterä H, Uller T. 2014 Neutral and adaptive explanations for an association between caste-biased gene expression and rate of sequence evolution. *Front. Genet.* **5**, 297. (doi:10.3389/fgene.2014.00297)
- Settepani V, Bechsgaard J, Bilde T. 2016 Phylogenetic analysis suggests that sociality is associated with reduced effectiveness of selection. *Nat. Ecol. Evol.* **6**, 469–477. (doi:10.1002/ece3.1886)
- Dyson CJ, Piscano OL, Durham RM, Thompson VJ, Johnson CH, Goodisman MA. 2021 Temporal analysis of effective population size and mating system in a social wasp. *J. Hered.* **112**, 626–634. (doi:10.1093/jhered/esab057)
- Weyna A, Romiguier J. 2021 Relaxation of purifying selection suggests low effective population size in eusocial Hymenoptera and solitary pollinating bees. *Peer Community Journal* **1**, e2. (doi:10.24072/pcjournal.3)
- Bromham L, Leys R. 2005 Sociality and the rate of molecular evolution. *Mol. Biol. Evol.* **22**, 1393–1402. (doi:10.1093/molbev/msi133)
- Hall DW, Goodisman MA. 2012 The effects of kin selection on rates of molecular evolution in social insects. *Evolution* **66**, 2080–2093. (doi:10.1111/j.1558-5646.2012.01602.x)
- Lo N, Tokuda G, Watanabe H, Rose H, Slaytor M, Maekawa K, Bandi C, Noda H. 2000 Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Curr. Biol.* **10**, 801–804. (doi:10.1016/S0960-9822(00)00561-3)
- Inward D, Beccaloni G, Eggleton P. 2007 Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett.* **3**, 331–335. (doi:10.1098/rsbl.2007.0102)
- Legendre F, Nel A, Svenson GJ, Robillard T, Pellens R, Grandcolas P. 2015 Phylogeny of Dictyoptera: dating the origin of cockroaches, praying mantises and termites with molecular data and controlled fossil evidence. *PLoS ONE* **10**, e0130127. (doi:10.1371/journal.pone.0130127)
- Bourguignon T *et al.* 2018 Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. *Mol. Biol. Evol.* **35**, 970–983. (doi:10.1093/molbev/msy013)
- Evangelista DA *et al.* 2019 An integrative phylogenomic approach illuminates the evolutionary history of cockroaches and termites (Blattodea). *Proc. R. Soc. B* **286**, 20182076. (doi:10.1098/rspb.2018.2076)
- Terrapon N *et al.* 2014 Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* **5**, 3636. (doi:10.1038/ncomms4636)
- Harrison MC *et al.* 2018 Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat. Ecol. Evol.* **2**, 557–566. (doi:10.1038/s41559-017-0459-1)
- Itakura S, Yoshikawa Y, Togami Y, Umezawa K. 2020 Draft genome sequence of the termite, *Coptotermes formosanus*: genetic insights into the pyruvate dehydrogenase complex of the termite. *J. Asia Pac. Entomol.* **23**, 666–674. (doi:10.1016/j.aspen.2020.05.004)
- Shigenobu S *et al.* 2022 Genomic and transcriptomic analyses of the subterranean termite *Reticulitermes speratus*: gene duplication facilitates social evolution. *Proc. Natl Acad. Sci. USA* **119**, e2110361119. (doi:10.1073/pnas.2110361119)
- Li S *et al.* 2018 The genomic and functional landscapes of developmental plasticity in the American cockroach. *Nat. Commun.* **9**, 1008. (doi:10.1038/s41467-018-03281-1)
- Yang Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
- Shell WA, Steffen MA, Pare HK, Seetharam AS, Severin AJ, Toth AL, Rehan SM. 2021 Sociality sculpts similar patterns of molecular evolution in two independently evolved lineages of eusocial bees. *Commun. Biol.* **4**, 253. (doi:10.1038/s42003-021-01770-6)
- Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015 RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832. (doi:10.1093/molbev/msu400)
- Pond SLK, Frost SD, Muse SV. 2005 HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679. (doi:10.1093/bioinformatics/bti079)

29. Pond SLK *et al.* 2020 HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol. Biol. Evol.* **37**, 295–299. (doi:10.1093/molbev/msz197)
30. Meusemann K, Korb J, Schughart M, Staubach F. 2020 No evidence for single-copy immune-gene specific signals of selection in termites. *Front. Ecol. Evol.* **8**, 26. (doi:10.3389/fevo.2020.00026)
31. Maekawa K, Hayashi Y, Lo N. 2022 Termite sociogenomics: evolution and regulation of caste-specific expressed genes. *Curr. Opin. Insect. Sci.* **50**, 100880. (doi:10.1016/j.cois.2022.100880)
32. Wright S. 1931 Evolution in mendelian populations. *Genetics* **16**, 97–159. (doi:10.1093/genetics/16.2.97)
33. Nei M, Tajima F. 1981 Genetic drift and estimation of effective population size. *Genetics* **98**, 625–640. (doi:10.1093/genetics/98.3.625)
34. Crozier RH. 1979 Genetics of sociality. In *Social insects* (ed. HR Hermann), pp. 223–286. New York, NY: Academic Press.
35. Glémin S. 2007 Mating systems and the efficacy of selection at the molecular level. *Genetics* **177**, 905–916. (doi:10.1534/genetics.107.073601)
36. Kimura M. 1983 *The neutral theory of molecular evolution*. Cambridge, UK: Cambridge University Press.
37. Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, Foster PL. 2016 Genetic drift, selection and the evolution of the mutation rate. *Nat. Rev. Genet.* **17**, 704–714. (doi:10.1038/nrg.2016.104)
38. Chak ST, Baeza JA, Barden P. 2021 Eusociality shapes convergent patterns of molecular evolution across mitochondrial genomes of snapping shrimps. *Mol. Biol. Evol.* **38**, 1372–1383. (doi:10.1093/molbev/msaa297)
39. Bergeron LA *et al.* 2023 Evolution of the germline mutation rate across vertebrates. *Nature* **615**, 285–291. (doi:10.1038/s41586-023-05752-y)
40. Leichty AR, Pfennig DW, Jones CD, Pfennig KS. 2012 Relaxed genetic constraint is ancestral to the evolution of phenotypic plasticity. *Integr. Comp. Biol.* **52**, 16–30. (doi:10.1093/icb/ics049)
41. Schrader L, Helantera H, Oettler J. 2017 Accelerated evolution of developmentally biased genes in the tetraphenic ant *Cardiocondyla obscurior*. *Mol. Biol. Evol.* **34**, 535–544. (doi:10.1093/molbev/msw240)
42. Traniello JF, Rosengaus RB, Savoie K. 2002 The development of immunity in a social insect: evidence for the group facilitation of disease resistance. *Proc. Natl Acad. Sci. USA* **99**, 6838–6842. (doi:10.1073/pnas.102176599)
43. Cremer S, Armitage SA, Schmid-Hempel P. 2007 Social immunity. *Curr. Biol.* **17**, R693–R702. (doi:10.1016/j.cub.2007.06.008)
44. He S *et al.* 2021 Evidence for reduced immune gene diversity and activity during the evolution of termites. *Proc. R. Soc. B* **288**, 20203168. (doi:10.1098/rspb.2020.3168)
45. Bulmer MS, Stefano AM. 2022 Termite eusociality and contrasting selective pressure on social and innate immunity. *Behav. Ecol. Sociobiol.* **76**, 4. (doi:10.1007/s00265-021-03090-5)
46. Pascoal S, Shimadzu H, Mashoodh R, Kilner RM. 2023 Parental care results in a greater mutation load, for which it is also a phenotypic antidote. *Proc. R. Soc. B* **290**, 20230115. (doi:10.1098/rspb.2023.0115)
47. Tong KJ, Duchêne S, Ho SYW, Lo N. 2015 Comment on 'Phylogenomics resolves the timing and pattern of insect evolution'. *Science* **349**, 487. (doi:10.1126/science.aaa5460)
48. Muller HJ. 1950 Our load of mutations. *Am. J. Hum. Genet.* **2**, 111–176. (doi:10.1007/BF00139458)
49. Crow JF. 1970 Genetic loads and the cost of natural selection. In *Mathematical topics in population genetics* (ed. K Kojima), pp. 128–177. Berlin, Germany: Springer.
50. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012 Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **7**, e46688. (doi:10.1371/journal.pone.0046688)
51. Bertorelle G, Raffini F, Bosse M, Bortoluzzi C, Iannucci A, Trucchi E, Morales HE, van Oosterhout C. 2022 Genetic load: genomic estimates and applications in non-model animals. *Nat. Rev. Genet.* **23**, 492–503. (doi:10.1038/s41576-022-00448-x)
52. Frankham R, Bradshaw CJ, Brook BW. 2014 Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biol. Conserv.* **170**, 56–63. (doi:10.1016/j.biocon.2013.12.036)
53. Frankham R, Ballou JD, Ralls K, Eldridge MDB, Dudash MR, Fenster CB, Lacy RC, Sunnucks P. 2017 *Genetic management of fragmented animal and plant populations*. Oxford, UK: Oxford University Press.
54. Thorne BL, Traniello JFA, Adams ES, Bulmer M. 1999 Reproductive dynamics and colony structure of subterranean termites of the genus *Reticulitermes* (Isoptera Rhinotermitidae): a review of the evidence from behavioral, ecological, and genetic studies. *Ethol. Ecol. Evol.* **11**, 149–169. (doi:10.1080/08927014.1999.9522833)
55. Lepage M, Darlington JPEC. 2000 Population dynamics of termites. In *Termites: evolution, sociality, symbioses, ecology* (eds T Abe, D Bignell, M Higashi), pp. 333–361. Dordrecht, The Netherlands: Kluwer Academic Press.
56. Smith CR, Toth AL, Suarez AV, Robinson GE. 2008 Genetic and genomic analyses of the division of labour in insect societies. *Nat. Rev. Genet.* **9**, 735–748. (doi:10.1038/nrg2429)
57. DeHeer CJ, Vargo EL. 2006 An indirect test of inbreeding depression in the termites *Reticulitermes flavipes* and *Reticulitermes virginicus*. *Behav. Ecol. Sociobiol.* **59**, 753–761. (doi:10.1007/s00265-005-0105-9)
58. Rosengaus RB, Traniello JF. 1993 Disease risk as a cost of outbreeding in the termite *Zootermopsis angusticollis*. *Proc. Natl Acad. Sci. USA* **90**, 6641–6645. (doi:10.1073/pnas.90.14.6641)
59. Fei HX, Henderson G. 2003 Comparative study of incipient colony development in the Formosan subterranean termite, *Coptotermes formosanus* Shiraki (Isoptera, Rhinotermitidae). *Insectes Soc.* **50**, 226–233. (doi:10.1007/s00040-003-0666-4)
60. Calleri DV, Rosengaus RB, Traniello JF. 2005 Disease and colony foundation in the dampwood termite *Zootermopsis angusticollis*: the survival advantage of nestmate pairs. *Naturwissenschaften* **92**, 300–304. (doi:10.1007/s00114-005-0630-4)
61. Eyer PA, Vargo EL. 2022 Short and long-term costs of inbreeding in the lifelong-partnership in a termite. *Commun. Biol.* **5**, 389. (doi:10.1038/s42003-022-03317-9)
62. Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. 2010 The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* **8**, e1000506. (doi:10.1371/journal.pbio.1000506)
63. Bonasio R *et al.* 2012 Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* **22**, 1755–1764. (doi:10.1016/j.cub.2012.07.042)
64. Kapheim KM *et al.* 2015 Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 1139–1143. (doi:10.1126/science.aaa4788)
65. Glastad KM, Gokhale K, Liebig J, Goodisman MA. 2016 The caste-and sex-specific DNA methylome of the termite *Zootermopsis nevadensis*. *Sci. Rep.* **6**, 37110. (doi:10.1038/srep37110)
66. Oldroyd BP, Yagound B. 2021 The role of epigenetics, particularly DNA methylation, in the evolution of caste in insect societies. *Philos. Trans. R. Soc. B* **376**, 20200115. (doi:10.1098/rstb.2020.0115)
67. Sieber KR, Dorman T, Newell N, Yan H. 2021 (Epi)genetic mechanisms underlying the evolutionary success of eusocial insects. *Insects* **12**, 498. (doi:10.3390/insects12060498)
68. Cedar H, Bergman Y. 2009 Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10**, 295–304. (doi:10.1038/nrg2540)

69. Vaughan RM, Kupai A, Rothbart SB. 2021 Chromatin regulation through ubiquitin and ubiquitin-like histone modifications. *Trends Biochem. Sci.* **46**, 258–269. (doi:10.1016/j.tibs.2020.11.005)
70. Spannhoff A *et al.* 2011 Histone deacetylase inhibitor activity in royal jelly might facilitate caste switching in bees. *EMBO Rep.* **12**, 238–243. (doi:10.1038/embor.2011.9)
71. Dickman MJ, Kucharski R, Maleszka R, Hurd PJ. 2013 Extensive histone post-translational modification in honey bees. *Insect Biochem. Mol. Biol.* **43**, 125–137. (doi:10.1016/j.ibmb.2012.11.003)
72. Wojciechowski M, Lowe R, Maleszka J, Conn D, Maleszka R, Hurd PJ. 2018 Phenotypically distinct female castes in honey bees are defined by alternative chromatin states during larval development. *Genome Res.* **28**, 1532–1542. (doi:10.1101/gr.236497.118)
73. Simola DF, Ye C, Mutti NS, Dolezal K, Bonasio R, Liebig J, Reinberg D, Berger SL. 2013 A chromatin link to caste identity in the carpenter ant *Camponotus floridanus*. *Genome Res.* **23**, 486–496. (doi:10.1101/gr.148361.112)
74. Simola DF *et al.* 2016 Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. *Science* **351**, aac6633. (doi:10.1126/science.aac6633)
75. Warner MR, Qiu L, Holmes MJ, Mikheyev AS, Linksvayer TA. 2019 Convergent eusocial evolution is based on a shared reproductive groundplan plus lineage-specific plastic genes. *Nat. Commun.* **10**, 2651. (doi:10.1038/s41467-019-10546-w)
76. Wang J, Santiago E, Caballero A. 2016 Prediction and estimation of effective population size. *Heredity* **117**, 193–206. (doi:10.1038/hdy.2016.43)
77. Chen Z *et al.* 2020 Ultralow-input single-tube linked-read library method enables short-read second-generation sequencing systems to routinely generate highly accurate and economical long-range sequencing information. *Genome Res.* **30**, 898–909. (doi:10.1101/gr.260380.119)
78. Wang O *et al.* 2019 Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res.* **29**, 798–808. (doi:10.1101/gr.245126.118)
79. Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012 Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276. (doi:10.1016/j.ymeth.2012.05.001)
80. DeMaere MZ, Darling AE. 2021 qc3C: reference-free quality control for Hi-C sequencing data. *PLoS Comput. Biol.* **17**, e1008839. (doi:10.1371/journal.pcbi.1008839)
81. Wenger AM *et al.* 2019 Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **37**, 1155–1162. (doi:10.1038/s41587-019-0217-9)
82. Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. 2017 Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767. (doi:10.1101/gr.214874.116)
83. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. 2016 Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res.* **44**, e147. (doi:10.1093/nar/gkw654)
84. Xu M *et al.* 2020 TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience* **9**, g1aa094. (doi:10.1093/gigascience/g1aa094)
85. Ghurye J, Pop M, Koren S, Bickhart D, Chin CS. 2017 Scaffolding of long read assemblies using long range contact information. *BMC Genom.* **18**, 527. (doi:10.1186/s12864-017-3879-z)
86. Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. 2019 Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput. Biol.* **15**, e1007273. (doi:10.1371/journal.pcbi.1007273)
87. Salamov AA, Solovyev VV. 2000 Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516–522. (doi:10.1101/gr.10.4.516)
88. Solovyev V, Kosarev P, Seledsov I, Vorobyev D. 2006 Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **7**, S10. (doi:10.1186/gb-2006-7-s1-s10)
89. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013 QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075. (doi:10.1093/bioinformatics/btt086)
90. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212. (doi:10.1093/bioinformatics/btv351)
91. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Kliuchnikov G, Kriventseva EV, Zdobnov EM. 2018 BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548. (doi:10.1093/molbev/msx319)
92. Waterhouse RM, Seppey M, Simão FA, Zdobnov EM. 2019 Using BUSCO to assess insect genomic resources. In *Insect genomics: methods in molecular biology*, vol. 1858 (eds S Brown, M Pfrender). New York, NY: Humana Press.
93. Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021 BUSCO: assessing genomic data quality and beyond. *Curr. Protoc.* **1**, e323. (doi:10.1002/cpz1.323)
94. Mei Y *et al.* 2022 InsectBase 2.0: a comprehensive gene resource for insects. *Nucleic Acids Res.* **50**, D1040–D1045. (doi:10.1093/nar/gkab1090)
95. Emms DM, Kelly S. 2019 OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238. (doi:10.1186/s13059-019-1832-y)
96. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
97. Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2020 CAFÉ 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518. (doi:10.1093/bioinformatics/btaa1022)
98. Beasley-Hall PG, Rose HA, Walker J, Kinjo Y, Bourguignon T, Lo N. 2021 Digging deep: a revised phylogeny of Australian burrowing cockroaches (Blaberidae: Panesthiinae, Geoscapheinae) confirms extensive nonmonophyly and provides insights into biogeography and evolution of burrowing. *Syst. Entomol.* **46**, 767–783. (doi:10.1111/syen.12487)
99. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015 Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353. (doi:10.1093/molbev/msv022)
100. R Core Team. 2022 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. See <https://www.R-project.org/>.
101. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020 IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534. (doi:10.1093/molbev/msaa015)
102. Eyre-Walker A, Keightley PD. 2007 The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618. (doi:10.1038/nrg2146)
103. Boyko AR *et al.* 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* **4**, e1000083. (doi:10.1371/journal.pgen.1000083)
104. Perrier C, Ferchaud AL, Sirois P, Thibault I, Bernatchez L. 2017 Do genetic drift and accumulation of deleterious mutations preclude adaptation? Empirical investigation using RAD seq in a northern lacustrine fish. *Mol. Ecol.* **26**, 6317–6335. (doi:10.1111/mec.14361)
105. Ochoa A, Gibbs HL. 2021 Genomic signatures of inbreeding and mutation load in a threatened rattlesnake. *Mol. Ecol.* **30**, 5454–5469. (doi:10.1111/mec.16147)
106. Kleinman-Ruiz D *et al.* 2022 Purging of deleterious burden in the endangered Iberian lynx. *Proc. Natl Acad. Sci. USA* **119**, e2110614119. (doi:10.1073/pnas.2110614119)
107. Ewart KM, Ho SYW, Chowdhury A-A, Jaya FR, Kinjo Y, Bennett J, Bourguignon T, Rose HA, Lo N. 2024 Data from: Pervasive relaxed selection in termite genomes. Dryad Digital Repository. (doi:10.5061/dryad.sqv9s4n9t)
108. Ewart KM, Ho SYW, Chowdhury A-A, Jaya FR, Kinjo Y, Bennett J, Bourguignon T, Rose HA, Lo N. 2024 Pervasive relaxed selection in termite genomes. Figshare. (doi:10.6084/m9.figshare.c.7226605)