bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Unexpectedly low recombination rates and presence of hotspots in termite genomes

Turid Everitt[1], Tilman Rönneburg[1], Daniel Elsner[2], Anna Olsson[1], Yuanzhen Liu[1], Tuuli Larva[1], Judith Korb[2,3], Matthew T Webster[1,4*]

1) Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

2) Evolutionary Biology and Ecology, University of Freiburg, Freiburg, Germany

3) Research Institute for the Environment and Livelihoods, Charles Darwin University, Casuarina Campus, Darwin, Australia

4) Science for Life Laboratory, Uppsala University, Uppsala, Sweden

*correspondence to matthew.webster@imbim.uu.se

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Abstract

Meiotic recombination is a fundamental evolutionary process that facilitates adaptation and the removal of deleterious genetic variation. Social Hymenoptera exhibit some of the highest recombination rates among metazoans, whereas high recombination rates have not been found among non-social species from this insect order. It is unknown whether elevated recombination rates are a ubiquitous feature of all social insects. In many metazoan taxa, recombination is mainly restricted to hotspots a few kilobases in length. However, little is known about the prevalence of recombination hotspots in insect genomes. Here we infer recombination rate and its fine-scale variation across the genomes of two social species from the insect order Blattodea: the termites *Cryptotermes secundus* and *Macrotermes bellicosus*. We used linkage-disequilibrium-based methods to infer recombination rate. We infer that recombination rates are less than 1 cM/Mb in both species, which is lower than the average metazoan rate. We also observed a highly punctate distribution of recombination in both termite genomes, indicative of the presence of recombination hotspots. We infer the presence of full-length PRDM9 genes in the genomes of both species. We also find that recombination rates in genes are correlated with inferred levels of germline DNA methylation. The finding of low recombination rates in termites indicates that eusociality is not universally connected to elevated recombination rate. We speculate that intense selection among haploid males promotes high recombination rates in social Hymenoptera. This study is one of the first to indicate the presence of PRDM9-based recombination hotspots in an insect genome.

Introduction

Meiotic recombination has two fundamental roles: It is essential for correct disjunction of chromosomes during cell division and it generates new combinations of genetic variants that form the raw material of evolution (Hartfield and Keightley 2012). However, recombination can also be deleterious, as it can generate structural mutations and break up favourable combinations of alleles. Recombination varies among species, with optimal recombination rate in a genome likely determined by the optimal trade-off of its positive and negative evolutionary and mechanistic effects (Stapley et al. 2017).

The highest recombination rates in metazoans observed so far are found in social insects (Wilfert et al. 2007; Stapley et al. 2017). However, until now recombination rate has only been studied in social insects from the order Hymenoptera, to which the majority of social insects belong. Among Hymenoptera, the highest rates are found in the genus *Apis*: the Western honey bee *Apis mellifera* (20.8 cM/Mb) and its relatives *Apis cerana* (17.4 cM/Mb) *Apis florea* (20.8 cM/Mb) and *Apis dorsata* (25.1 cM/Mb) (Beye et al. 2006; Shi et al. 2013; Liu et al. 2015; Wallberg et al. 2015; Rueppell et al. 2016; Kawakami et al. 2019). Other social Hymenoptera also have high rates including the bumblebee *Bombus terrestris* (8.9) (Liu et al. 2017; Kawakami et al. 2019), the stingless bee *Frieseomelitta varia* (9.3 – 12.5 cM/Mb) (Waiker et al. 2021), the wasp *Vespula vulgaris* (9.7 cM/Mb) (Sirviö et al. 2011a) and the ants *Pogonomyrmex rugosus* (11.1 cM/Mb) (Sirviö et al. 2011b) and *Acromymex echinatior* (6.1) (Sirviö et al. 2006). In contrast, the solitary bee *Megachile rotundata* and the solitary wasp *Nasonia* have relatively low rates (1.0 and 1.5 cM/Mb respectively) (Niehuis et al. 2010; Jones et al. 2019). For comparison, the average recombination rate of 15 insect species outside of Hymenoptera is 2.2 cM/Mb (Wilfert et al. 2007).

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

We can learn about the forces shaping the evolution of recombination rates in general by understanding why high recombination rates have evolved in social Hymenoptera. Several hypotheses have been proposed to explain this observation. One set of explanations focusses on the effects of recombination on intra-colony genetic diversity. For example, increasing the genetic diversity of a colony could promote task specialization among workers (Kent et al. 2012; Kent and Zayed 2013). Similarly, elevated genetic variation in a colony could prevent invasion by pathogens or parasites due to diversification of immune genes (Fischer and Schmid-Hempel 2005). Recombination could also reduce variance in relatedness between nestmates, thereby reducing potential kin conflict within colonies (Sherman 1979; Templeton 1979; Wilfert et al. 2007). Additionally, it has been proposed that high recombination rates could have facilitated the evolution of eusociality over a longer evolutionary timescale, because recombination between genes that take on functions in different castes permits them to evolve more independently (Kent and Zayed 2013). Several studies have addressed these hypotheses (Kent et al. 2012; Liu et al. 2015; Wallberg et al. 2015; Rueppell et al. 2016; Liu et al. 2017; Jones et al. 2019; Waiker et al. 2021; Kawakami et al. 2019) but so far none is strongly supported. It is also not clear whether eusociality *per se* selects for higher recombination rates, or whether high rates are a specific feature of eusocial Hymenoptera.

The main characteristics of eusocial insect species are the presence of a worker caste, which forgo reproduction in order to care for brood of other colony members, and a single (or very few) reproducing individuals (queens, and in termites, also kings). Eusocial insects are found mainly in Hymenoptera, among bees, wasps and ants, and also Blattodea, in which all termites (Infraorder: Isoptera) are eusocial but exhibit varying levels of social complexity. These insect orders likely diverged ~400 million years ago (Misof et al. 2014) and there are substantial differences between social insects from the two orders (Korb 2008; Korb and Thorne 2017). Hymenoptera have haplodiploid sex determination in which males develop from haploid, unfertilized eggs and females derive from diploid, fertilized eggs. In Blattodea, both males and females are diploid. Members of the worker caste in eusocial Hymenoptera are all female, whereas eusocial Blattodea workers are of both sexes. Hymenoptera colonies are headed by one or a small number of queens, whereas Blattodea colonies contain both a king and a queen.

The hypotheses proposed to explain the high recombination rates observed in eusocial Hymenoptera predict that recombination should be elevated in all social insects. However, so far it is unknown whether social insects from other orders also have high rates. As described above, there are fundamental differences in the sex determination system and colony compositions of social insects from Hymenoptera and Blattodea. The robustness of the association between eusociality and high recombination rates can therefore be addressed by assessing mean genomic recombination rates in social Blattodea (termites).

In addition to variation between species, the rate of recombination also varies along chromosomes. In mammals, crossover events are mainly restricted to regions that contain specific motifs that act as binding sites for the PRDM9 (PR/SET domain 9) protein that marks regions for double-stranded breaks, which are repaired by recombination during meiosis (Baudat et al. 2010; Myers et al. 2010; Parvanov et al. 2010). Species with an active PRDM9 include nearly all mammals and many vertebrate taxa (Baker et al. 2017). In species without PRDM9, recombination can be directed to hotspots defined by other features, such as CpG islands and promoters that have open chromatin. Such hotspots have been characterised in

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

the genomes of diverse taxa including yeast, birds, and dogs (Axelsson et al. 2012; Lam and Keeney 2015; Singhal et al. 2015; Berglund et al. 2014). In insects, the prevalence of recombination hotspots is not well studied. Fine-scale recombination maps in the fruit fly *Drosophila melanogaster* and honey bee *A. mellifera* indicate a paucity of recombination hotspots (Chan et al. 2012; Smukowski Heil et al. 2015; Wallberg et al. 2015), whereas hotspots appear to be present in the genome of the butterfly *Leptidea sinapis* (Torres et al. 2023).

Several other factors may also modulate recombination rate along chromosomes, such as germline DNA methylation. Many species have elevated recombination in CpG islands, which are usually unmethylated in the germline (Axelsson et al. 2012; Lam and Keeney 2015; Singhal et al. 2015; Berglund et al. 2014). In insect genomes, methylation is not always present, but commonly restricted to gene bodies (Arsala et al. 2022; Wang et al. 2013; Lyko et al. 2010; Ventós-Alfonso et al. 2020; Bewick et al. 2019; Harrison et al. 2018). In the honey bee *A. mellifera* and solitary bee *M. rotundata* recombination is reduced in genes, and particularly in those inferred to have the highest levels of germline methylation (Wallberg et al. 2015; Jones et al. 2019). In addition, some studies have identified correlations between caste-biases in gene expression and recombination. In honey bees, genes with worker-biased gene expression have been found to have elevated recombination rates (Kent et al. 2012) but this is likely an indirect result of differences in germline methylation between sets of genes (Wallberg et al. 2015).

Here we estimate recombination rate and its fine-scale variation across the genomes of two distantly related termite species with contrasting social complexities (Korb and Hartfelder 2008; Korb and Thorne 2017). The drywood termite *Cryptotermes secundus* (Kalotermitidae) has a low level of social complexity. Its colonies consist of a few hundred individuals and totipotent workers from which queens and kings develop (Hoffmann et al. 2012). It belongs to the wood-dwelling (one-piece nesting) termites that nest in a single piece of wood that also serves as their food source. *Macrotermes bellicosus* (Termitidae: Macrotermitinae) is a fungus-growing termite with a high level of social complexity, with colonies of several millions of individuals and sterile workers. It belongs to the foraging (multiple-pieces nesting) termites, where workers leave the nest to forage for food.

The study has three main aims. Firstly, we aim to test the robustness of the association between sociality and high recombination rate. If sociality is a universal driver of high recombination rates, we would expect both termite species to have elevated recombination rates, and that it is particularly elevated in *M. bellicosus*, which has a higher level of social complexity. Secondly, we aim to determine whether recombination hotspots exist in the two termite genomes, and infer whether active full-length PRDM9 genes are present in their genomes. Thirdly, we aim to analyse the factors that govern variation in recombination across the genome. In particular, whether genomic variation in recombination rates is associated with patterns of methylation and gene expression as has been reported in other studies.

To perform the analysis, we conducted whole-genome sequencing of 10 samples each of *C. secundus* and *M. bellicosus*, randomly sampled from different colonies. We then used patterns of SNP variation to infer fine-scale variation in recombination rate across the genomes using coalescent-based analysis of patterns of linkage disequilibrium. Our results indicate unexpectedly low rates of recombination in both termite genomes, demonstrating that sociality is not universally associated with high recombination rates as proposed from

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

analysis of social Hymenoptera. They also suggest the presence of recombination hotspots, which may be controlled by the PRDM9 protein.

Results

Genetic variation in two termite species is typical for social insects

We utilized genome assemblies of two termite species: *C. secundus* and *M. bellicosus*. The genome assembly of *C. secundus* (Csec_1.0) (Harrison et al. 2018) is divided into 55,483 scaffolds with a scaffold N50 of 1.2 Mbp and a total length of 1.0 Gbp. The genome assembly of *M. bellicosus* (Qiu et al. 2023) is 1.3 Gbp in total length across 275 scaffolds with scaffold N50 of 22.4 Mbp.

We sequenced 10 unrelated individuals each of *C. secundus* and *M. bellicosus* to a mean depth of 34x and 31x per sample, respectively (Supplementary Table S1). We mapped reads to the appropriate genome assemblies and called variants in each species. In *C. secundus* we called 15.1 million SNPs and in *M. bellicosus* 5.6 million SNPs (Table 1). One individual of *C. secundus* (CS_8) was excluded due to a low proportion of mapped reads, poor mapping quality and strand bias in mapping. Estimates of variation based on θ_W (Watterson 1975) are 0.14% and 0.44% in *M. bellicosus* and *C. secundus* respectively, which are broadly comparable to levels of variation in the honey bee *A. mellifera* (0.3 - 0.8%) (Wallberg et al. 2014).

We used these estimates of genetic variation to estimate effective population size (N_E) for each of the sampled populations. These rely on an estimate of mutation rate, μ, in each species. As there is no estimate of μ for Blattodea, we considered a range of estimates from Diptera (*Drosophila*) (Keightley et al. 2009, 2014), Hymenoptera (bumblebee (Liu et al. 2017) and honey bee (Yang et al. 2015)) and Lepidoptera (Keightley et al. 2015) that range between 2.8 - 3.5 x 10^{-9} $bp^{-1}gen^{-1}$(Table 1).

Our estimates of N_E based on θ_W are 98,600 – 123,300 in *M. bellicosus*, which has high social complexity, and 314,900 – 393,700 in *C. secundus*, with lower level of social complexity. This is consistent with previous evidence that eusocial species with large colonies and high social complexity tend to have lower effective population sizes and the values are broadly comparable to levels of N_E found in social and solitary Hymenoptera (Leffler et al. 2012; Romiguier et al. 2014).

Low average recombination rate in termite genomes

We next examined the decay of linkage disequilibrium (LD) based on the statistic r^2. Both termite species show more extensive LD compared to *A. mellifera*, which has an extremely high recombination rate (Wallberg et al. 2015) (Figure 1). As N_E is broadly similar in all these species, these differences are expected to largely represent differences in recombination rates, and indicate that the two termite species do not exhibit the elevated genome average recombination rate that is observed in honey bees.

We estimated variation in the population recombination rate, ρ, in both datasets using LDhat (Auton and McVean 2007) (Table 1). Average rates of ρ/kb are 1.7 and 4.8 in *M. bellicosus* and *C. secundus* respectively. These estimates can be converted to cM/Mb using estimates of N_E, which gives estimates of average recombination rate of 0.34 - 0.43 cM/Mb in *M. bellicosus* and 0.31 - 0.38 cM/Mb in *C. secundus*, which is much lower than the

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

average among insects (Stapley et al. 2017), and particularly lower than the elevated recombination rates inferred in eusocial Hymenoptera. This represents the first estimation of recombination rate in social insects outside of Hymenoptera and demonstrates that eusociality is not a universal driver of high recombination rates.

Evidence for active recombination hotspots in termite genomes

Variation in recombination rate across the genome in *C. secundus* and *M. bellicosus* appears highly variable (Figure 2). We characterised the distribution of recombination events across the two termite genomes using cumulative distribution plots (Figure 3). We found that 50% of the recombination occurs in 4.8% of the genome in *C. secundus* and 3.5% of the genome in *M. bellicosus*. This indicates that the majority of recombination events are restricted to a much smaller portion of the genome than observed in *A. mellifera*, where 50% of recombination events occur in >30% of the genome (Wallberg et al. 2015).

We further investigated whether recombination hotspots are present in the two termite genomes by looking at the distribution of ρ in windows of 1 kbp, 10 kbp and 100 kbp across the genome (Figure S1). For all window sizes, a small subset of windows was observed with ρ greatly exceeding the mean in both termite species. For example, the percentage of 1 kbp windows with recombination rate at least four standard deviations above the genome wide mean value is 1.20% for *C. secundus* and 0.88% for *M. bellicosus*. By contrast, *A. mellifera* has no windows above this limit for any window size, consistent with a lack of recombination hotspots.

We next defined hotspots as 2 kbp regions with more than 5-fold higher recombination rate than the local background as defined as the surrounding 100 kbp and used STREME (Bailey 2021) to identify 8 to 20-mers that were enriched in hotspots compared to the rest of the genome. This yielded significantly enriched motifs in both species (12 motifs for *C. secundus* and 8 motifs for *M. bellicosus*, after multiple testing; Table S2 and S3). However, none the motifs we identified are present in more than 5% of hotspots and exhibit more than 2-fold difference relative to the background, suggesting they are not viable candidate motifs for promoting recombination in hotspots.

Identification and characterization of the PRDM9 gene in termites

We searched the *C. secundus* and *M. bellicosus* genomes for evidence of complete PRDM9 genes, which could potentially govern the presence of recombination hotspots in these genomes. The *C. secundus* genome contains a PRDM9 homolog (XP_023708049.2) with annotated KRAB, SSXRD, SET and zinc finger domains (Figure 4). It is plausible that this gene is responsible for initiation of recombination in hotspots as it does in many other taxa. There is no complete homolog of PRDM9 in the *M. bellicosus* annotation. We therefore used BLAST to search for sequences homologous to the conserved functional domains from *C. secundus*. We identified a putative PRDM9 orthologue on Scaffold 42, containing all domains in the correct order. However, we also identified retroviral conserved domains 5' of the zinc-finger domains in this gene in the *M. bellicosus* genome assembly.

We used a zinc-finger prediction algorithm (Persikov and Singh 2014) to predict DNA binding motifs from the zinc-finger domain present in PRDM9 in the *C. secundus* and *M. bellicosus* genome assemblies, for which an accurate annotation was available. The most probable consensus motifs were TAATAAGTAGAATCGTTTAG and TATGGAACGACAGGAACAACGACATCATCAGCCGCCTAAT for *C. secundus* and *M. bellicosus*

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

respectively. Base probability matrices for these motifs are shown in Table S4. We tested whether the presence of these motifs corresponded to elevated recombination considering 1 kbp around the motif. However, recombination rate was not significantly elevated around these predicted motifs for either species, either when searching based on the most likely motif, or by searching based on the probability matrices.

Genomic correlates of recombination rate

We analysed genomic correlates of recombination rate in 10 kbp windows. There is a significant correlation between recombination and $CpG_{O/E}$ in both species with Spearman's $\rho = 0.10$, $p < 1 \times 10^{-4}$ for *C. secundus* and Spearman's $\rho = 0.24$, $p < 1 \times 10^{-4}$ for *M. bellicosus* (Figure S2). A reduction of recombination rates in regions of low CpG content is consistent with interference of germline methylation with recombination. However, as the correlation coefficients are low, the explanatory power is weak. Correlations of similar magnitudes were observed between recombination and GC content (Spearman's ρ=0.23 for *C. secundus*, $p < 1 \times 10^{-4}$, and Spearman's ρ=0.017 for *M. bellicosus*, $p < 1 \times 10^{-4}$). This correlation has been associated with an effect of recombination in driving GC through GC biased gene conversion but the relatively low correlation coefficients indicate that this force is not strong in termites. By contrast, the correlation between GC content and recombination rate is much stronger in honey bee ($R^2 = 0.506$), which is likely caused by high recombination rates and intense GC-biased gene conversion (Wallberg et al. 2015).

There is no significant correlation between overall repeat element density and recombination rate in either of the two genomes. There is a weak positive but highly significant correlation between simple repeats and recombination in both genomes (Spearman's $\rho = 0.07$, $p < 1 \times 10^{-4}$ for *C. secundus* and Spearman's $\rho = 0.06$, $p < 1 \times 10^{-4}$ for *M. bellicosus*). SINE and LINE elements are also only weakly associated with recombination rate in both genomes (SINEs: Spearman's $\rho = 0.02$, $p < 1 \times 10^{-4}$ for *C. secundus* and Spearman's $\rho = -0.01$, $p > 0.05$ for *M. bellicosus*; LINEs: Spearman's $\rho = 0.001$, $p > 0.05$ for *C. secundus* and Spearman's $\rho = -0.01$, $p < 1 \times 10^{-4}$ for *M. bellicosus*). There is a significant negative correlation between gene density and recombination for the *M. bellicosus* genome (Spearman's $\rho = -0.20$, $p < 1 \times 10^{-4}$), but no significant correlation for *C. secundus*. We also tested whether hotspots differ from the background with respect to GC content. The GC content does not differ substantially between hotspots and the rest of the genome in either species (means of 40.1% and 40.4% for *M. bellicosus* and 40.5% in both categories for *C. secundus*).

We found a highly significant correlation between nucleotide diversity (π) and recombination in both species (Figure S2), with Spearman's $\rho = 0.32$, $p < 1 \times 10^{-4}$ for *C. secundus* and Spearman's $\rho = 0.52$, $p < 1 \times 10^{-4}$ for *M. bellicosus*. This correlation is found across a wide range of eukaryotic taxa and is generally accepted to be caused by the interaction between recombination and linked selection (Nachman 2001; Begun and Aquadro 1992).

Recombination and gene expression patterns correlate with $CpG_{O/E}$

Analysis of the distribution of $CpG_{O/E}$ in both termite species showed that it is bimodally distributed, with substantially lower values in introns and exons compared to flanking noncoding regions (Figure S3). This indicates that germline DNA methylation is biased towards gene bodies in the two termite genomes, a pattern that is also found in honey bees

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

and other insect genomes with functional DNA methylation (Harrison et al. 2018; Elango et al. 2009; Sarda et al. 2012). We found that exons have lower $CpG_{O/E}$ compared to introns, which are both lower than flanking regions (Figure 5, top row; all comparisons $p < 0.001$ after Bonferroni correction) consistent with higher levels of germline methylation in exons (Table S5). We next analysed how recombination rate varies among genomic features. We found that ρ/kbp is also significantly reduced in genes compared to flanking regions, with exons showing reduced ρ/kbp compared to introns (Figure 5, bottom row, all p-values <0.005 after Bonferroni correction). This effect was seen in both species and is more pronounced for *C. secundus* than for *M. bellicosus.* There was a significant correlation between $CpG_{O/E}$ and recombination rate among genes (Spearman's $\rho = 0.166$, $p < 1 \times 10^{-4}$ for *C. secundus* and $\rho = 0.369$, $p < 1 \times 10^{-4}$ for *M. bellicosus*).

We next investigated whether patterns of gene expression between castes and sexes (hereafter, for short 'caste-biased gene expression') were correlated with $CpG_{O/E}$ and recombination rate using gene expression data from two studies (Lin et al. 2021; Elsner et al. 2018). It has been proposed that it is evolutionarily advantageous for genes with worker-biased expression to be located in regions with elevated recombination (Kent et al. 2012). For *C. secundus*, worker-biased and queen-biased genes were identified (Lin et al. 2021). For *M. bellicosus*, we defined several categories based on gene expression difference among castes (Elsner et al. 2018). The categories we used were queen-biased, king-biased, worked-biased, male-biased, female-biased, reproduction-biased, and differentially expressed (see methods for details). Comparisons of $CpG_{O/E}$ and recombination rate for the original gene expression categories reported for *M. bellicosus* are shown in Figure S4. We find that mean $CpG_{O/E}$ in gene bodies shows significant variation between gene expression categories in both *M. bellicosus* and *C. secundus* (Figure 6). However, this variation is not consistent between species. In *M. bellicosus*, queen-biased and female-biased genes have low $CpG_{O/E}$ and king-biased and male-biased genes have high $CpG_{O/E}$. In *C. secundus*, queen-biased genes have higher than average $CpG_{O/E}$.

The differences in $CpG_{O/E}$ between caste-biased gene expression categories are mirrored by differences in ρ/kbp (Figure 6). Gene expression categories with elevated $CpG_{O/E}$ consistently show elevated ρ/kbp, and this pattern is found in both species. Studies in the honey bee have found that worker-biased genes have lower CpG and higher recombination rates (Kent et al. 2012; Wallberg et al. 2015). In *M. bellicosus,* we find that worker-biased genes have slightly elevated $CpG_{O/E}$ and ρ/kbp. However, in *C. secundus*, these values are reduced in worker-biased genes, whereas queen-biased genes have a slightly elevated $CpG_{O/E}$ and ρ/kbp in this species.

We investigated the relative roles of $CpG_{O/E}$ and gene expression in determining ρ in genes using a linear model, with $CpG_{O/E}$, ρ in flanking region and expression categories as explanatory variables. In both *C. secundus* and *M. bellicosus*, we found that $CpG_{O/E}$ was the strongest predictor of recombination rate variation among genes (Tables S6 and S7) in line with the analyses above. We found that caste-biased expression had no significant impact on the recombination rate, with the exception of king-biased genes in *M. bellicosus*. These have higher ρ/kbp than expected based on $CpG_{O/E}$ alone. Generally, $CpG_{O/E}$ is a good predictor of ρ/kbp, and the variation in ρ/kbp among expression categories likely mainly reflects differences in $CpG_{O/E}$, as accounting for it removes most significant interactions.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Discussion

The main findings we present here are 1) that two distantly related termite species with varying social complexity both have relatively low genomic average rates of meiotic recombination, 2) both genomes possess recombination hotspots and likely contain full-length copies of PRDM9 and, 3) there are reduced levels of recombination in genes and genome regions depleted in CpG sites. Our findings contrast with the extremely elevated recombination rates observed in eusocial Hymenoptera, which are likely to be generated by a feature that is specific to these taxa. The finding of recombination hotspots is one of the first in insects. This could reflect a conserved function of PRDM9 in recombination in vertebrates and invertebrates. Our results also support a role for germline methylation in suppressing recombination.

Eusociality is not a universal driver of high rates of meiotic recombination

Elevated genomic rates of recombination are characteristic of all eusocial Hymenoptera so far investigated. Several hypotheses have been advanced to explain this phenomenon, which are all based on an evolutionary advantage of recombination in eusocial species (Wilfert et al. 2007). For instance, it has been suggested that elevated recombination could increase genetic diversity in a colony. This could promote diversity in the workforce, enabling more efficient task specialisation among workers (Kent et al. 2012). Alternatively, it could render the colony less susceptible to invasion by parasites and pathogens (Fischer and Schmid-Hempel 2005). Elevated recombination might also favour rapid evolution of caste-specifically expressed genes (Kent and Zayed 2013). In addition, increased recombination could lead to reduced variance in relatedness, which could prevent kin conflict in colonies (Sherman 1979; Templeton 1979).

All these hypotheses predict that recombination rate should be elevated in all eusocial insects. Here, however, in the first report of recombination rate in social insects outside of Hymenoptera, we find overall low levels of recombination. Both termite species studied here are eusocial, and in particular *M. bellicosus* has extremely large and complex colonies characteristic of eusociality (Korb and Thorne 2017). The low recombination rate inferred in the two termite genomes prompts re-evaluation of hypotheses connecting eusociality with high recombination rates. Although both termites and social Hymenoptera share many traits, they differ in others that reflect their different ancestries.

The chromosome numbers of both termite species (*M. bellicosus*, 2n = 42 and for *C. secundus*, 2n = 40) are typical for termites, although variation exists between taxa (Jankásek et al. 2021). These values do not differ from those observed in Hymenoptera, which have a similar range (e.g. *A. mellifera*, 2n = 32) and do not vary between solitary and eusocial taxa (Cardoso et al. 2018; Cunha et al. 2021; Ross et al. 2015). This suggests that differences in chromosome number are unlikely to be relevant in explaining the differences in recombination rates among these taxa.

One proposed advantage of sex and recombination is that it is associated with sexual selection and a higher intensity of selection on males. This could be advantageous because differential male mating success can reduce mutational load in sexual populations by causing deleterious mutations to segregate at a lower equilibrium frequency on average (Siller 2001; Agrawal 2001). Recombination is essential for effective purging of deleterious alleles as it unlinks deleterious variants from nearby beneficial ones thus preventing selection interference (Hartfield and Keightley 2012). Selection has been found to be

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

stronger in males than females across animal species (Janicke et al. 2016; Winkler et al. 2021). Species with a male-biased sex ratio are expected to experience particularly intense selection on males. It is therefore possible that recombination rate could be particularly elevated in species in which males face intense competition as this increases the evolutionary advantage of recombination.

The genus *Apis* produces highly male-biased reproductive offspring and thousands of male drones compete to fertilise a small number of female queens (Winston 1991; Koeniger et al. 2011; Baer 2005). It is also in this genus that the most extreme recombination rates have been inferred (Beye et al. 2006; Shi et al. 2013; Liu et al. 2015; Wallberg et al. 2015; Rueppell et al. 2016; Kawakami et al. 2019). Male-biased sex ratios are also observed in other social bees and wasps. Both male and female biased operational sex ratios have been observed in ants, but are difficult to measure because workers invest more resources in female offspring and may manipulate sex ratios. In non-social Hymenoptera, females lay roughly equal numbers of haploid and diploid eggs (Trivers and Hare 1976). The sex ratio of reproductive males and females is also relatively even in most termite species (Roisin 2001). It is possible that the male-biased sex ratio observed in many eusocial Hymenoptera could favour elevated recombination rates in the species so far studied.

The effects of higher intensity of selection on males on recombination rate could be even more pronounced when males are haploid as in Hymentoptera. Haploid males can be considered analogous to the gametes of females. However, as many more genes are expressed in adult haploid males compared to gametes there are more targets of selection. Selection against deleterious alleles is particularly effective in haploid males because recessive deleterious alleles are exposed (Hedrick and Parker 1997).

Although its effects on overall recombination rates are unclear, haploid selection is believed be key to driving sex-specific differences in recombination rate (Immler 2019). It has been argued that differences in the intensity of selection between sexes in the haploid phase can drive differences in recombination rates between sexes among animal and plant species (heterochiasmy) (Burt et al. 1991; Trivers 1988; Lenormand 2003; Lenormand and Dutheil 2005). It is unclear how theory and observations regarding the prevalence and nature of heterochiasmy applies to haplodiploids, in which only females recombine. However, it appears that the intensity of selection on the haploid phase likely has the potential to drive selective pressure on recombination rates.

Further research is needed to determine the cause of variation in recombination rates among social and non-social insects. For example, simulations could be used to determine the effect of selection on haploid males on recombination rates, considering different sex ratios. In addition, it is necessary to estimate recombination rate in a wider range of social and non-social species to determine whether the strength of selection on males is a key factor in determining these rates.

Full length PRDM9 gene and recombination hotspots in termite genomes

Our results are also indicative of the presence of recombination hotspots in the two termite genomes. The fine-scale distribution of recombination rate in both termite genomes is similar to that in humans where 50% of recombination events occur in ~6% of the genome (Myers et al. 2005). This contrasts to the honey bee and solitary bees where fine-scale maps are available and recombination is relatively uniformly distributed in the genome (Wallberg et al. 2015; Jones et al. 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Two main mechanisms have been shown to lead to recombination hotspots. The first, which is best understood in human and mouse, is governed by the PRDM9 protein. A zinc-finger domain recognises a specific DNA sequence motif and then performs a histone modification in the vicinity, which marks the sequence for a DNA double-stranded break that is repaired by recombination during meiosis (Baudat et al. 2010; Myers et al. 2010; Parvanov et al. 2010). In species with this mechanism, the zinc finger domain is fast evolving, which results in a rapid turnover of hotspot locations (Myers et al. 2010). Full-length PRDM9 orthologues are present among a range of highly-diverged vertebrates including many fish, reptiles and mammals, but many full or partial losses of the gene have also been inferred. In a study of 225 vertebrate genomes, a minimum of six partial and three complete losses were inferred (Baker et al. 2017). However, it is unclear whether PRMD9-hotspots are common in invertebrates.

A second mechanism is found in vertebrates that lack PRDM9, in which recombination hotspots are localised to regions of open chromatin such as CpG islands and promoters. This is found in dogs and in most or all birds, in which PRDM9 has been lost, and results in hotspots with stable rather than rapidly evolving locations (Axelsson et al. 2012; Singhal et al. 2015). A similar pattern is also observed in the yeast *Saccharomyces cerevisiae* in which recombination preferentially occurs in promoter regions (Wu and Lichten 1994). In invertebrates such as *Drosophila* and *Caenorhabditis elegans*, recombination hotspots appear to be absent (Coop and Przeworski 2007; Chan et al. 2012; Smukowski Heil et al. 2015). Fine-scale maps of bees have also not revealed recombination hotspots (Wallberg et al. 2015; Jones et al. 2019).

We find evidence that recombination is directed towards hotspots in both termite species. We also identify a full-length PRDM9 gene in both species, containing all of the domains needed to initiate recombination. However, we are unable to demonstrate that PRMD9 is responsible for the presence of recombination hotspots in the two genomes. We used a motif prediction algorithm to predict the zinc-finger binding motif in the *C. secundus* genome, but we did not observe elevated recombination rate around instances of the motif in the genome. Furthermore, we did not identify any common sequence motifs that were enriched in recombination hotspots in either of the termite genomes. A plausible interpretation of these observations is that the PRDM9 protein indeed initiates recombination in hotspots in both species, but that positive selection leads to frequent shifts in the target motifs, which results in a lack of a strong association between a specific motif and LD-based hotspots (Myers et al. 2010). It is also possible that the sequences of the zinc-finger motifs are not correctly represented in the genome assemblies, which could result from problems with assembly around minisatellites. However, an unknown mechanism could also be responsible for hotspots in termites.

Evidence that germline DNA methylation suppresses recombination

In insects with a functioning DNA methyltransferase machinery, the $CpG_{O/E}$ statistic shows substantial variation along the genome, which mainly reflects variation in levels of germline DNA methylation (Elango et al. 2009). In contrast to vertebrate genomes, methylation is strongly skewed towards gene bodies in the genomes of insects (Glastad et al. 2011). The overall genomic correlation between $CpG_{O/E}$ and recombination rate is significant but weak in the two termite genomes. This could reflect that the majority of the genomes are not methylated, so that much of the variation in $CpG_{O/E}$ on the genome scale is not strongly influenced by methylation. In the two termite genomes studied here, we find lowest values

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

of $CpG_{O/E}$ in exons, whereas noncoding flanking regions are biased towards higher values, indicative of lower levels of DNA methylation (Figure 5, Figure S3). All categories display a predominantly bimodal distribution of $CpG_{O/E}$, likely reflective of the presence of methylated and unmethylated regions. The differences in $CpG_{O/E}$ are also reflected by differences in recombination rate (ρ/kb), which are reduced in regions inferred to be more highly methylated.

Using linear models, we identified $CpG_{O/E}$ as the factor with the strongest influence on recombination rate variation among genes, consistent with an effect of DNA methylation on recombination rate. This factor had a stronger effect than differences in gene expression patterns across castes, a result that was also observed in the honey bee genome (Wallberg et al. 2015). This indicates that recombination rate variation is not modulated by differences in gene expression in these genomes, which might be expected if natural selection favours increased recombination rates in certain genes. For example, it has been proposed that natural selection could favour elevated recombination rates in genes with roles in worker behaviour or in immune function due to their important roles in colony function (Wilfert et al. 2007; Kent et al. 2012). However, the results presented here suggest that differences in recombination rate between genes with caste-biased patterns of gene expression are mainly due to underlying differences in methylation levels among genes.

A direct effect of germline DNA methylation in suppressing recombination could explain variation in recombination landscapes among vertebrates and invertebrates. Vertebrate genomes are usually highly methylated with the exception of CpG islands. In the absence of a functional PRDM9 gene, recombination events tend to localise to CpG islands in vertebrate genomes (for example in dogs and birds)(Singhal et al. 2015; Axelsson et al. 2012). However, in insect genomes without PRDM9, which are typically not as highly methylated, recombination events do not cluster in hotspots (for example in fruit flies and honey bees) (Chan et al. 2012; Wallberg et al. 2015). However, in the presence of a functional PRDM9 in both vertebrate or invertebrate genomes, recombination tends to cluster in hotspots determined by sequence motifs.

Conclusion

We find low rates of meiotic recombination in the genomes of two termite species from disparate parts of the phylogeny and with contrasting social complexity. This demonstrates that eusociality is not a universal driver of high recombination rates. We propose a haploid selection hypothesis that could potentially explain why recombination might be elevated in eusocial Hymenopterans, but not other social insects. This hypothesis predicts that recombination rates should depend on the intensity of selection in the haploid stage of the life cycle, but requires theoretical and empirical testing. We also provide evidence that recombination is biased towards hotspots in both genomes, and that they both contain full-length PRDM9 genes. The results are consistent with the involvement of PRDM9 in initiating recombination in hotspots in termites, and they are one of the first indications of the existence of this mechanism in invertebrates. Additional studies are required to demonstrate a role of PRDM9 in initiating recombination in hotspots in termites.

Materials and Methods

Sample collection

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Cryptotermes secundus colonies were collected from dead *Ceriops tagal* mangrove trees near Palmerston-Channel Island (Darwin Harbor, Northern Territory, Australia; 12°30'S 131°00'E). Colonies were then kept in *Pinus radiata* wood blocks in climate rooms in Germany, providing 28 °C, 70% relative humidity, and a 12-h day/night cycle (Lin et al. 2021). *Macrotermes bellicosus* samples were collected from colonies in Kakpin, next to the Comoé National Park in Côte d'Ivoire (coordinates 8°39'N 3°46'W) (Elsner et al. 2018). For both species, each of the sequenced samples came from a different colony.

Population sequencing

The termites were cut in two, approximately between head and thorax. DNA was extracted from both parts using the QIAGEN Blood and Tissue Kit following the standard protocol, including treating with 4 µL RNase and 25 µL Proteinase K. The DNA extraction product was cleaned and concentrated with ZymoResearch Genomic DNA Clean & Concentrator kit. For most samples, the RNase treatment was repeated prior to cleaning and concentrating. One DNA sample from each individual was chosen, based on the DNA amount and quality, to be sequenced. We performed Illumina short-read sequencing using NovaSeq 6000 with 2 x 150 bp paired end reads.

Read mapping and variant calling

The sequence data from the two termite species were mapped to the reference genomes for *C. secundus* (Csec_1.0) (Harrison et al. 2018) and *M. bellicosus* (Qiu et al. 2023), respectively, using Burrows-Wheeler alignment tool BWA version 0.7.17 (Li & Durbin 2009) with the BWA-MEM algorithm. Sorting and indexing of the BAM-files was done with the Samtools package version 1.17 (Li et al. 2009), followed by adding read groups and marking duplicate reads with Picard toolkit version 1.118 (http://broadinstitute.github.io/picard/). After mapping, we excluded all scaffolds < 1kb from the *C. secundus* genome, which in total correspond to 2% of the genome size, as they are potentially of lower quality and not informative for analysis of linkage disequilibrium or recombination. For *M. bellicosus* there are no scaffolds shorter than 1 kbp, but we excluded the following five scaffolds, which together correspond to 3.78% of the genome length, due to unexpectedly high heterozygosity: 93, 62, 41, 35 and 32.

One of the *C. secundus* individuals did not map well to the reference genome (71% mapped reads, unequal distribution between forward and reverse strand and an average mapping quality of 35 while all the other samples had an average mapping quality >45). This individual was excluded from all further analyses and hence the total number of termite individuals analysed in this study is 19, out of which 9 were *C. secundus*.

Variant calling was done using the tools HaplotypeCaller, GenomicsDBImport and GenotypeGVCFs from GATK version 4.3.0.0 as described in their best practices workflow (https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels-, last accessed 2023-07-11). The variants were filtered in multiple steps, first removing extended regions of low mapping quality or deviating read depth as those regions are likely to be unreliable. This was done based on statistics from the Samtools mpileup and depth tools, respectively, removing 100 kb windows with mean mapping quality below 70 or mean read depth more than two standard deviations from the genome-wide mean. In the later recombination rate and LD decay calculations, values in intervals spanning those regions were also removed. Indels were removed using VCFtools version 0.1.16 (Danecek et al. 2011) before GATK VariantFiltration was run with the

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

following limits, which are all equal to or stricter than the ones recommended (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants, last accessed 2023-07-11) and adapted to the observed distributions of the corresponding statistics: QD<2; FS>50; MQ<40; ReadPosRankSum< -4; ReadPosRankSum>4; SOR>3; ExcessHet>5 for both species as well as MQRankSum with limits -5 and 5 for *C. secundus* and limits -6 and 4 for *M. bellicosus*. This was followed by additional filtering with VCFtools to select only bi-allelic sites with a minimum quality score of 30 and maximum of 40% missing genotypes. The SNPs with mean read depth in the lowest or highest 2.5% of the depth distribution were also filtered out, which translates to a depth below 10 or above 34 for *C. secundus* and below 5 or above 40 for *M. bellicosus.* For the recombination rate calculations, the SNPs were also filtered for a minor allele count >= 2, as rare variants appearing on only one chromosome are difficult to distinguish from errors and are not informative regarding LD. The data was also phased and imputed using Beagle version 5.1 (Browning and Browning 2007; Browning et al. 2018).

Estimation of LD decay

The decay of linkage disequilibrium, r2, with increasing physical distance was estimated with the software PopLDDecay version 3.42 (Zhang et al. 2019) over a maximum distance of 10 kbp. For this analysis, sequence data from *A. mellifera scutellata* was included for comparison ((Wallberg et al. 2017); BioProject PRJNA357367). The *A. m. scutellata* sequences were mapped to the Amel_HAv3.1 reference genome (Wallberg et al. 2019) and processed and filtered in the same way as described above for the termite data (GATK VariantFiltration with the limits QD<2, FS>25, MQ<40, MQRankSum<-4, MQRankSum>4, ReadPosRankSum<-4, ReadPosRankSum>4, SOR>3, ExcessHet>5 and VCFtools to select bi-allelic sites with mean read depth between 5-13, quality score >= 30, minor allele count >=2 and <=40% missing genotypes).

Estimation of population recombination rate

The recombination rate was estimated using the complete, interval and stat functions of LDhat (https://github.com/auton1/Ldhat, downloaded 2023-04-19) (Auton and McVean 2007) with 10 million iterations, sampling every 5000th iteration, taking the first 20 samples as burn-in.

We evaluated different block penalties (1, 5 and 20), where a lower value allows the recombination rate to change more easily across the map. As there were only minor differences between the results for both species, we decided to use a block penalty of 1 as this has been used previously in similar studies (Wallberg et al. 2015; Jones et al. 2019).

The output from the LDhat stat program gives the value of the parameter ρ/kbp between each pair of consecutive SNPs. Those values were converted to equally sized windows across each scaffold, as well as the coordinates of individual genomic features, by taking a weighted average of values from overlapping intervals. The *M. bellicosus* scaffolds contain a few gaps of unknown size and the values between SNPs across such gaps were removed.

The parameter ρ estimated by LDhat is proportional to the recombination rate r and the effective population size N_E, which in turn can be estimated based on the mutation rate μ and the number of segregating sites, S. In order to convert ρ to r, the following equations were used, where n is the number of haploid sequences. The number of segregating sites

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

was calculated before the filter on minor allele count >=2 was applied, as this filter likely removes many true SNPs as well, which could affect the estimates.

θ_W was estimated for each sample using the equation below from (Watterson 1975):

$$\theta_W = K \Big/ \sum_{i=1}^{n-1} \frac{1}{i}$$

where *K* is the number of SNPs and *i* is the number of haploid genomes sampled. We used the value of θ_W to estimate N_E using:

$$N_E = \frac{\theta}{4\mu}$$

where μ is the mutation rate. We estimated N_E using a range of mutation rates estimated from other insect species (Keightley et al. 2009, 2014, 2015; Yang et al. 2015; Liu et al. 2017). We then estimated the recombination rate using the following equation, assuming a range of values of N_E.

$$r = \frac{\rho}{4N_e}$$

Cumulative plot of rho ρ/kbp

A cumulative plot of the proportion of rho versus the proportion of the physical distance along the genome was constructed based on the output from LDhat for the two termite species and previously generated LDhat output for *A. mellifera* (Wallberg et al. 2015; PRJNA236426). The ρ/kbp estimates between markers were placed in decreasing order before multiplication with the respective physical distances between the markers. Then cumulative sums were calculated for the ρ-values as well as the corresponding physical distances and divided by the total genome-wide sums.

k-mer enrichment in recombination hotspots

In order to identify k-mers associated with elevated recombination rate, we defined recombination hotspots as 2kb segments with at least five-fold elevated recombination rate compared to the local background, as defined by 50 kb flanking up and downstream of the segment, resulting in 40894 regions for *C. secundus*, and 39837 for *M. bellicosus.*

Then, we used these hotspots to search for motifs associated with elevated recombination rate by comparing against an equivalent number of randomly chosen 2kb regions from the rest of the genome. This was done using STREME 5.5.4 (Bailey 2021) using a minimum and maximum size of k-mer of 8 and 20, (--minw 8 ; --maxw 20) and otherwise default options. From all motifs with significant enrichment after adjusting for multiple testing, we then considered only motifs present in more than 5% of all hotspots and an at least two-fold enrichment compared to the background as credible candidates.

Identifying PRDM9 orthologues in C. secundus and M. bellicosus

A full-length copy of PRDM9 is present in the *C. secundus* genome annotation, but not in *M. bellicosus*. Using Orthofinder (Emms and Kelly 2019), and the genome annotation files for

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

both species, we identified a single orthologue to the *C. secundus* PRDM9 in *M. bellicosus*. We analysed this region using the NCBI conserved domain search (Wang et al. 2023). We also searched the *M. bellicosus* assembly for the PR/SET domains derived from *C. secundus* and *Zootermopsis nevadensis* using BLAST (Camacho et al. 2009) and subsequently for all other PRDM9 domains (KRAB, SSXRD, ZF-Casette) to identify hits that contain all other PRDM9 domains in the vicinity in the correct order.

Predicting binding sites from zinc-finger-motifs

Using the ZF-motifs from *C.secundus,* we obtained letter probability matrices for predicted binding motifs using a webserver (http://zf.princeton.edu/index.php ; (Persikov and Singh 2014) which were used to search for binding sites across the genome using FIMO (Grant et al. 2011) with default options. Subsequently, the mean rho for the surrounding 1 kb bin was used to test for a difference in mean rho between bins associated with a suspected binding site, and a similar number of randomly chosen bins without association. For significance testing, we repeated this draw 10.000 times to compare the associated bins against the 95% confidence interval of the resulting distribution.

Analysis of genomic correlates of recombination

We divided the genome into 10 kb windows in which average ρ and other genomic features were estimated. Windows shorter than 5 kb (half of the specified window size), which appear in short scaffolds or at the ends at longer scaffolds, were excluded from the analyses (excluding less than 1% of the sequence for each species). Correlation coefficients were estimated with Spearmans ρ, with significance estimated by permutation tests with 10,000 iterations.

In order to determine the repeat content across the *C. secundus* and *M. bellicosus* genome assemblies, we first generated custom repeat libraries for each species with RepeatModeler (Flynn et al. 2020) using default options. This was then combined with Repbase 29 (Bao et al. 2015) and DFAM 3.8 (Storer et al. 2021) and used to run RepeatMasker with default options on each genome. We estimated the proportion of sequence in each repeat class in 10 kbp windows and used Spearman's rank correlation to correlate this with estimates of recombination rate.

Per gene recombination rate and $CpG_{O/E}$ ratio

The observed/expected frequency of CpG sites ($CpG_{O/E}$) and mean ρ were calculated on a per-gene basis, as well as for exons and introns, and 50kb flanking regions 10kb upstream and downstream of the gene, akin to Wallberg et al (2015). For recombination rate, the data from the interval function in LDhat (Auton and McVean 2007) was used, utilising a weighted mean when the element spanned multiple markers. For visualisation, extreme $CpG_{O/E}$ values due to low GC content were set to a maximum of 4.

We tested for significant difference in mean between ρ and $CpG_{O/E}$ in flanking regions, exons and introns using paired t-tests corrected for multiple testing using a Bonferroni threshold.

Differential expression data

We classified genes in both species according to their caste-biased patterns of expression from two studies (Elsner et al. 2018; Lin et al. 2021). Elsner et al. (2018) analysed four castes of *M. bellicosus*, with two age classes of each: major and minor workers, queens and kings. They analysed differences in gene expression among these classes using RNAseq data. We

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

reclassified the differential expression data by grouping the direct caste-vs-caste comparisons as "queen-biased", “king-biased”, “worker-biased”, “reproduction-biased”, "male-biased" or "female-biased" if they were differentially upregulated for this caste (or group) in comparison to other castes (or groups). Since Elser *et al.* (2018) initially mapped the *M. bellicosus* reads against the *Macrotermes natalensis* genome, We used orthofinder (Emms and Kelly 2019) to identify the *M. bellicosus* orthologues corresponding to the differentially expressed genes. Only single-copy genes were considered for this analysis.

Lin et al. (2021), performed RNAseq on samples of workers and queens. Here, we used the data from untreated queens and workers. Differentially expressed *C. secundus* genes were classified as “worker-biased” or “queen-biased” if they were differentially expressed with a bias to the respective caste, or unbiased if they were not, and subsequently compared between classes.

Correlates of recombination rate variation among genes

In order to assess the relative importance of $CpG_{O/E}$ ratio and differential expression patterns on recombination rate among genes, an ordinary least squares model was built using statsmodels (Seabold and Perktold 2010). $CpG_{O/E}$, differential expression categories, flanking regions $CpG_{O/E}$ and flanking region recombination rate were used as exogenous variables.

Data access

The scripts are available on GitHub: https://github.com/troe27/termite-recombination-rate. Raw sequence from population sequencing are available at NCBI BioProject PRJNA1021607.

Competing interest statement

The authors declare no competing interests.

Acknowledgements

We acknowledge grant 2018-03896 from The Swedish Research Council to MTW and from the Deutsche Forschungsgemeinschaft (DFG) KO-1895/26-1: 417980976 to JK. We acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council. The computations were enabled by resources in project NAISS 2023/23-402 provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at UPPMAX, funded by the Swedish Research Council through grant agreement no. 2022-06725. We thank Göran Arnqvist for helpful discussions on the evolution of sex and recombination. The Parks and Wildlife Commission, Northern Territory, and the Department of the Environment, Water, Heritage and the Arts, Australia, gave permission to collect (Permit number 59044) and export (PWS2016-001559) the *C. secundus* termites. The Office Ivoirien des Parcs et Réserves (OIPR) provided sampling and export permits for the *M. bellicosus* samples. The study was conducted according to the Nagoya protocol.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

716 **References**

717 Agrawal AF. 2001. Sexual selection and the maintenance of sexual reproduction. *Nature*
718 **411**: 692–695.

719 Arsala D, Wu X, Yi SV, Lynch JA. 2022. Dnmt1a is essential for gene body methylation and
720 the regulation of the zygotic genome in a wasp. *PLOS Genetics* **18**: e1010181.

721 Auton A, McVean G. 2007. Recombination rate estimation in the presence of hotspots.
722 *Genome Research* **17**: 1219–27.

723 Axelsson E, Webster MT, Ratnakumar A, Consortium L, Ponting CP, Lindblad-Toh K. 2012.
724 Death of PRDM9 coincides with stabilization of the recombination landscape in the
725 dog genome. *Genome Res* **22**: 51–63.

726 Baer B. 2005. Sexual selection in Apis bees. *Apidologie* **36**: 187–200.

727 Bailey TL. 2021. STREME: accurate and versatile sequence motif discovery. *Bioinformatics*
728 **37**: 2834–2840.

729 Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, Przeworski M. 2017.
730 Repeated losses of PRDM9-directed recombination despite the conservation of
731 PRDM9 across vertebrates ed. B. De Massy. *eLife* **6**: e24133.

732 Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in
733 eukaryotic genomes. *Mobile DNA* **6**: 11.

734 Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. 2010.
735 PRDM9 is a major determinant of meiotic recombination hotspots in humans and
736 mice. *Science* **327**: 836–840.

737 Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with
738 recombination rates in D. melanogaster. *Nature* **356**: 519–520.

739 Berglund J, Quilez J, Arndt PF, Webster MT. 2014. Germline methylation patterns determine
740 the distribution of recombination events in the dog genome. *Genome Biol Evol* **7**:
741 522–530.

742 Bewick AJ, Sanchez Z, Mckinney EC, Moore AJ, Moore PJ, Schmitz RJ. 2019. Dnmt1 is
743 essential for egg production and embryo viability in the large milkweed bug,
744 Oncopeltus fasciatus. *Epigenetics & Chromatin* **12**: 6.

745 Beye M, Gattermeier I, Hasselmann M, Gempe T, Schioett M, Baines JF, Schlipalius D,
746 Mougel F, Emore C, Rueppell O, et al. 2006. Exceptionally high levels of
747 recombination across the honey bee genome. *Genome research* **16**: 1339–44.

748 Browning BL, Zhou Y, Browning SR. 2018. A One-Penny Imputed Genome from Next-
749 Generation Reference Panels. *The American Journal of Human Genetics* **103**: 338–
750 348.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

751 Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data
752 inference for whole-genome association studies by use of localized haplotype
753 clustering. *Am J Hum Genet* **81**: 1084–1097.

754 Burt A, Bell G, Harvey PH. 1991. Sex differences in recombination. *Journal of Evolutionary*
755 *Biology* **4**: 259–277.

756 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
757 BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

758 Cardoso DC, Santos HG, Cristiano MP. 2018. The Ant Chromosome database – ACdb: an
759 online resource for ant (Hymenoptera: Formicidae) chromosome researchers.
760 *Myrmecological News* **27**: 87–91.

761 Chan AH, Jenkins PA, Song YS. 2012. Genome-Wide Fine-Scale Recombination Rate Variation
762 in Drosophila melanogaster. *PLOS Genetics* **8**: e1003090.

763 Coop G, Przeworski M. 2007. An evolutionary view of human recombination. *Nature reviews*
764 *Genetics* **8**: 23–34.

765 Cunha MS, Cardoso DC, Cristiano MP, de Oliveira Campos LA, Lopes DM. 2021. The Bee
766 Chromosome database (Hymenoptera: Apidae). *Apidologie* **52**: 493–502.

767 Elango N, Hunt BG, Goodisman MA, Yi SV. 2009. DNA methylation is widespread and
768 associated with differential gene expression in castes of the honeybee, Apis
769 mellifera. *Proceedings of the National Academy of Sciences of the United States of*
770 *America* **106**: 11206–11.

771 Elsner D, Meusemann K, Korb J. 2018. Longevity and transposon defense, the case of
772 termite reproductives. *PNAS* **115**: 5504–5509.

773 Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative
774 genomics. *Genome Biology* **20**: 238.

775 Fischer O, Schmid-Hempel P. 2005. Selection by parasites may increase host recombination
776 frequency. *Biol Lett* **1**: 193–195.

777 Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020.
778 RepeatModeler2 for automated genomic discovery of transposable element families.
779 *Proceedings of the National Academy of Sciences* **117**: 9451–9457.

780 Glastad KM, Hunt BG, Yi SV, Goodisman M a. D. 2011. DNA methylation in insects: on the
781 brink of the epigenomic era. *Insect Molecular Biology* **20**: 553–565.

782 Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif.
783 *Bioinformatics* **27**: 1017–1018.

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

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

787 Hartfield M, Keightley PD. 2012. Current hypotheses for the evolution of sex and
788 recombination. *Integrative Zoology* **7**: 192–209.

789 Hedrick PW, Parker JD. 1997. Evolutionary Genetics and Genetic Variation of Haplodiploids
790 and X-Linked Genes. *Annual Review of Ecology and Systematics* **28**: 55–83.

791 Hoffmann K, Foster KR, Korb J. 2012. Nest value mediates reproductive decision making
792 within termite societies. *Behavioral Ecology* **23**: 1203–1208.

793 Immler S. 2019. Haploid Selection in "Diploid" Organisms. *Annual Review of Ecology,*
794 *Evolution, and Systematics* **50**: 219–236.

795 Janicke T, Häderer IK, Lajeunesse MJ, Anthes N. 2016. Darwinian sex roles confirmed across
796 the animal kingdom. *Science Advances* **2**: e1500983.

797 Jankásek M, Kotyková Varadínová Z, Šťáhlavský F. 2021. Blattodea Karyotype Database. *EJE*
798 **118**: 192–199.

799 Jones JC, Wallberg A, Christmas MJ, Kapheim KM, Webster MT. 2019. Extreme Differences in
800 Recombination Rate between the Genomes of a Solitary and a Social Bee. *Mol Biol*
801 *Evol* **36**: 2277–2291.

802 Kawakami T, Wallberg A, Olsson A, Wintermantel D, de Miranda JR, Allsopp M, Rundlöf M,
803 Webster MT. 2019. Substantial Heritable Variation in Recombination Rate on
804 Multiple Scales in Honeybees and Bumblebees. *Genetics* **212**: 1101–1119.

805 Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014. Estimation of the Spontaneous
806 Mutation Rate per Nucleotide Site in a Drosophila melanogaster Full-Sib Family.
807 *Genetics* **196**: 313–320.

808 Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, Davey JW,
809 Jiggins CD. 2015. Estimation of the Spontaneous Mutation Rate in Heliconius
810 melpomene. *Mol Biol Evol* **32**: 239–243.

811 Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. 2009. Analysis of the
812 genome sequences of three Drosophila melanogaster spontaneous mutation
813 accumulation lines. *Genome Res* **19**: 1195–1201.

814 Kent CF, Minaei S, Harpur BA, Zayed A. 2012. Recombination is associated with the
815 evolution of genome structure and worker behavior in honey bees. *PNAS* **109**:
816 18012–18017.

817 Kent CF, Zayed A. 2013. Evolution of recombination and genome structure in eusocial
818 insects. *Communicative & Integrative Biology* **6**: e22919.

819 Koeniger G, Koeniger N, Phiancharoen M. 2011. Comparative Reproductive Biology of
820 Honeybees. In *Honeybees of Asia* (eds. H.R. Hepburn and S.E. Radloff), pp. 159–206,
821 Springer, Berlin, Heidelberg https://doi.org/10.1007/978-3-642-16422-4_8 (Accessed
822 February 13, 2024).

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Korb J. 2008. Termites, hemimetabolous diploid white ants? *Frontiers in Zoology* **5**: 15.

Korb J, Hartfelder K. 2008. Life history and development - a framework for understanding developmental plasticity in lower termites. *Biological Reviews* **83**: 295–313.

Korb J, Thorne B. 2017. Sociality in Termites. In *Comparative Social Evolution* (eds. D.R. Rubenstein and P. Abbot), pp. 124–153, Cambridge University Press, Cambridge https://www.cambridge.org/core/books/comparative-social-evolution/sociality-in-termites/11A80CB9FA947760889676E03899E85C (Accessed August 9, 2023).

Lam I, Keeney S. 2015. Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. *Science* **350**: 932–937.

Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, Venkat A, Andolfatto P, Przeworski M. 2012. Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? *PLoS Biol* **10**: e1001388.

Lenormand T. 2003. The evolution of sex dimorphism in recombination. *Genetics* **163**: 811–22.

Lenormand T, Dutheil J. 2005. Recombination Difference between Sexes: A Role for Haploid Selection. *PLoS Biology* **3**: e63.

Lin S, Werle J, Korb J. 2021. Transcriptomic analyses of the termite, Cryptotermes secundus, reveal a gene network underlying a long lifespan and high fecundity. *Commun Biol* **4**: 1–12.

Liu H, Jia Y, Sun X, Tian D, Hurst LD, Yang S. 2017. Direct Determination of the Mutation Rate in the Bumblebee Reveals Evidence for Weak Recombination-Associated Mutation and an Approximate Rate Constancy in Insects. *Mol Biol Evol* **34**: 119–130.

Liu H, Zhang X, Huang J, Chen J-Q, Tian D, Hurst LD, Yang S. 2015. Causes and consequences of crossing-over evidenced via a high-resolution recombinational landscape of the honey bee. *Genome Biol* **16**: 15.

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.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**: 763–767.

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**: 321–4.

Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, Macfie TS, McVean G, Donnelly P. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* **327**: 876–9.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Nachman MW. 2001. Single nucleotide polymorphisms and recombination rate in humans. *Trends in genetics : TIG* **17**: 481–5.

Niehuis O, Gibson JD, Rosenberg MS, Pannebakker BA, Koevoets T, Judson AK, Desjardins CA, Kennedy K, Duggan D, Beukeboom LW, et al. 2010. Recombination and Its Impact on the Genome of the Haplodiploid Parasitoid Wasp Nasonia. *PLOS ONE* **5**: e8597.

Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 controls activation of mammalian recombination hotspots. *Science* **327**: 835–835.

Persikov AV, Singh M. 2014. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Research* **42**: 97–108.

Qiu B, Elsner D, Korb J. 2023. High-quality long-read genome assemblies reveal evolutionary patterns of transposable elements and DNA methylation in termites. 2023.10.31.564968. https://www.biorxiv.org/content/10.1101/2023.10.31.564968v1 (Accessed January 8, 2024).

Roisin Y. 2001. Caste sex ratios, sex linkage, and reproductive strategies in termites. *Insectes soc* **48**: 224–230.

Romiguier J, Lourenco J, Gayral P, Faivre N, Weinert LA, Ravel S, Ballenghien M, Cahais V, Bernard A, Loire E, et al. 2014. Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *Journal of Evolutionary Biology* **27**: 593–603.

Ross L, Blackmon H, Lorite P, Gokhman VE, Hardy NB. 2015. Recombination, chromosome number and eusociality in the Hymenoptera. *Journal of Evolutionary Biology* **28**: 105–116.

Rueppell O, Kuster R, Miller K, Fouks B, Rubio Correa S, Collazo J, Phaincharoen M, Tingek S, Koeniger N. 2016. A New Metazoan Recombination Rate Record and Consistently High Recombination Rates in the Honey Bee Genus Apis Accompanied by Frequent Inversions but Not Translocations. *Genome Biol Evol* **8**: 3653–3660.

Sarda S, Zeng J, Hunt BG, Yi SV. 2012. The Evolution of Invertebrate Gene Body Methylation. *Mol Biol Evol* **29**: 1907–1916.

Seabold S, Perktold J. 2010. Statsmodels: Econometric and Statistical Modeling with Python. In *Proceedings of the 9th Python in Science Conference* (eds. S. Van der Walt and J. Millman), pp. 92–96.

Sherman PW. 1979. Insect Chromosome Numbers and Eusociality. *The American Naturalist* **113**: 925–935.

Shi YY, Sun LX, Huang ZY, Wu XB, Zhu YQ, Zheng HJ, Zeng ZJ. 2013. A SNP Based High-Density Linkage Map of Apis cerana Reveals a High Recombination Rate Similar to Apis mellifera. *PLOS ONE* **8**: e76459.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

Siller S. 2001. Sexual selection and the maintenance of sex. *Nature* **411**: 689–692.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B, Balakrishnan CN, et al. 2015. Stable recombination hotspots in birds. *Science* **350**: 928–932.

Sirviö A, Gadau J, Rueppell O, Lamatsch D, Boomsma JJ, Pamilo P, PAGE Jr. RE. 2006. High recombination frequency creates genotypic diversity in colonies of the leaf-cutting ant Acromyrmex echinatior. *Journal of Evolutionary Biology* **19**: 1475–1485.

Sirviö A, Johnston JS, Wenseleers T, Pamilo P. 2011a. A high recombination rate in eusocial Hymenoptera: evidence from the common wasp Vespula vulgaris. *BMC Genetics* **12**: 95.

Sirviö A, Pamilo P, Johnson RA, Page Jr Robert E, Gadau J. 2011b. ORIGIN AND EVOLUTION OF THE DEPENDENT LINEAGES IN THE GENETIC CASTE DETERMINATION SYSTEM OF POGONOMYRMEX ANTS. *Evolution* **65**: 869–884.

Smukowski Heil CS, Ellison C, Dubin M, Noor MAF. 2015. Recombining without Hotspots: A Comprehensive Evolutionary Portrait of Recombination in Two Closely Related Species of Drosophila. *Genome Biology and Evolution* **7**: 2829–2842.

Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**: 20160455.

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA* **12**: 2.

Templeton AR. 1979. Chromosome Number, Quantitative Genetics and Eusociality. *The American Naturalist* **113**: 937–941.

Torres AP i, Höök L, Näsvall K, Shipilina D, Wiklund C, Vila R, Pruisscher P, Backström N. 2023. The fine-scale recombination rate variation and associations with genomic features in a butterfly. *Genome Res* **33**: 810–823.

Trivers R. 1988. Sex differences in rates of recombination and sexual selection. In *The Evolution of Sex*, pp. 270–286, Sinauer Associates, Sunderland, MA.

Trivers RL, Hare H. 1976. Haploidploidy and the Evolution of the Social Insect. *Science* **191**: 249–263.

Ventós-Alfonso A, Ylla G, Montañes J-C, Belles X. 2020. DNMT1 Promotes Genome Methylation and Early Embryo Development in Cockroaches. *iScience* **23**: 101778.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

930 Waiker P, de Abreu FCP, Luna-Lucena D, Freitas FCP, Simões ZLP, Rueppell O. 2021.
931 Recombination mapping of the Brazilian stingless bee Frieseomelitta varia confirms
932 high recombination rates in social hymenoptera. *BMC Genomics* **22**: 673.

933 Wallberg A, Bunikis I, Pettersson OV, Mosbech M-B, Childers AK, Evans JD, Mikheyev AS,
934 Robertson HM, Robinson GE, Webster MT. 2019. A hybrid de novo genome assembly
935 of the honeybee, Apis mellifera, with chromosome-length scaffolds. *BMC Genomics*
936 **20**: 275.

937 Wallberg A, Glémin S, Webster MT. 2015. Extreme recombination frequencies shape
938 genome variation and evolution in the honeybee, *Apis mellifera*. *PLoS Genetics* **11**:
939 e1005189.

940 Wallberg A, Han F, Wellhagen G, Dahle B, Kawata M, Haddad N, Simões ZLP, Allsopp MH,
941 Kandemir I, De la Rúa P, et al. 2014. A worldwide survey of genome sequence
942 variation provides insight into the evolutionary history of the honeybee *Apis*
943 *mellifera*. *Nat Genet* **46**: 1081–1088.

944 Wallberg A, Schöning C, Webster MT, Hasselmann M. 2017. Two extended haplotype blocks
945 are associated with adaptation to high altitude habitats in East African honey bees.
946 *PLoS Genetics* **13**: e1006792.

947 Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, Marchler GH, Song JS,
948 Thanki N, Yamashita RA, et al. 2023. The conserved domain database in 2023.
949 *Nucleic Acids Res* **51**: D384–D388.

950 Wang X, Wheeler D, Avery A, Rago A, Choi J-H, Colbourne JK, Clark AG, Werren JH. 2013.
951 Function and Evolution of DNA Methylation in Nasonia vitripennis. *PLoS Genet* **9**:
952 e1003872.

953 Watterson GA. 1975. On the number of segregating sites in genetical models without
954 recombination. *Theoretical Population Biology* **7**: 256–276.

955 Wilfert L, Gadau J, Schmid-Hempel P. 2007. Variation in genomic recombination rates
956 among animal taxa and the case of social insects. *Heredity (Edinb)* **98**: 189–197.

957 Winkler L, Moiron M, Morrow EH, Janicke T. 2021. Stronger net selection on males across
958 animals eds. P.J. Wittkopp and L. Holman. *eLife* **10**: e68316.

959 Winston ML. 1991. *The Biology of the Honey Bee*. Revised edition. Harvard University Press,
960 Cambridge, Mass.

961 Wu T-C, Lichten M. 1994. Meiosis-Induced Double-Strand Break Sites Determined by Yeast
962 Chromatin Structure. *Science* **263**: 515–518.

963 Yang S, Wang L, Huang J, Zhang X, Yuan Y, Chen J-Q, Hurst LD, Tian D. 2015. Parent–progeny
964 sequencing indicates higher mutation rates in heterozygotes. *Nature* **523**: 463–467.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

965 Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. 2019. PopLDdecay: a fast and effective tool for
966 linkage disequilibrium decay analysis based on variant call format files.
967 *Bioinformatics* **35**: 1786–1788.

968

969

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

970 **Table 1.** Estimates of levels genetic variation, effective population size and recombination
971 rate in two termite species.

Species	Assembly length (Mb)	No. SNPs	θ_W/bp	ρ/kb	$\mu\ bp^{-1}\ gen^{-1}$	N_E	r (cM/Mb)
M. bellicosus	1,139	5,581,662	0.14%	1.69	2.8	123,292	0.34
					3.5	98,634	0.43
C. secundus	992	15,037,468	0.44%	4.85	2.8	393,662	0.31
					3.5	314,930	0.38

972 θ_W/bp - Watterson's θ per base pair

973 ρ/kb - population recombination parameter, ρ, per kilobase

974 $\mu\ bp^{-1}\ gen^{-1}$ - estimate of mutation rate per base pair per generation

975 N_E - effective population size

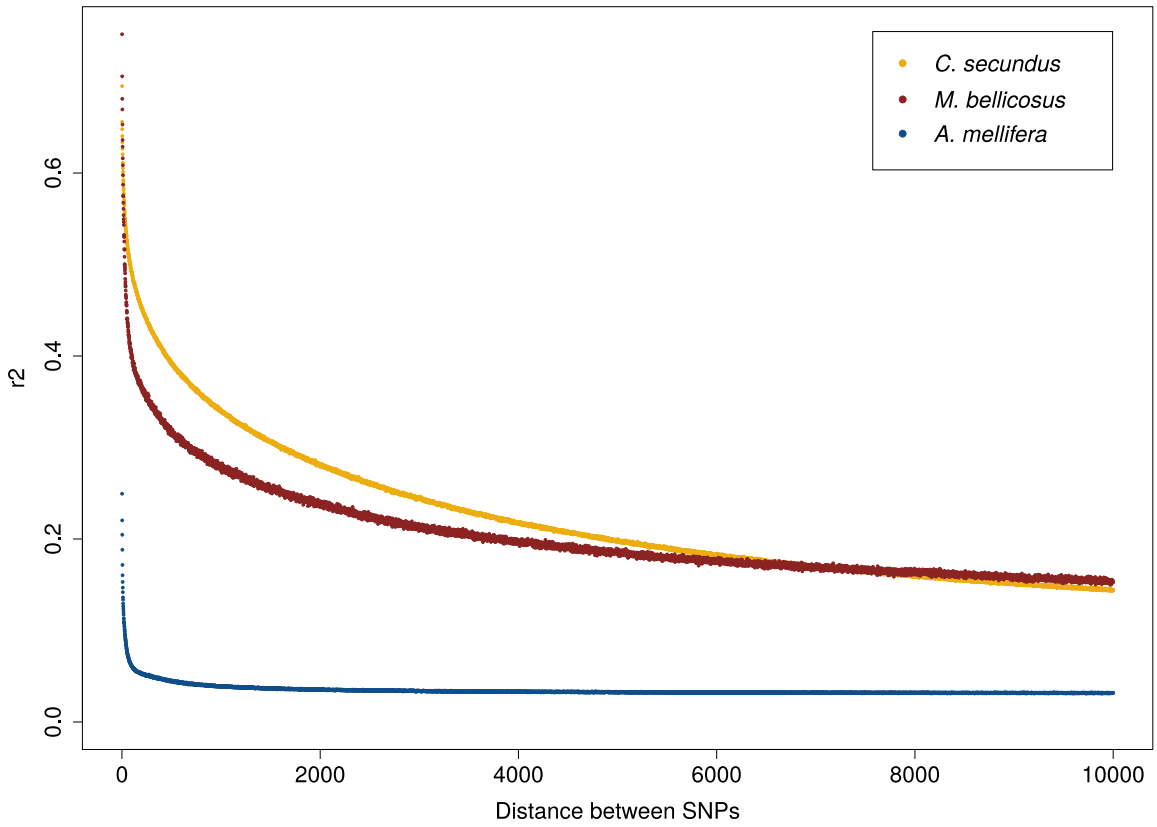
976 r (cM/Mb) - recombination rate in centimorgans per megabase

977

978

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

979



980

981

982 **Figure 1.** Genome-wide decay of LD measured as r^2 versus distance (bp) between SNPs for
983 the two termite species (*C. secundus*, yellow; *M. bellicosus*, dark red) and honey bees (*A.*
984 *mellifera scutellata*, blue; data from Wallberg et al. 2017).

985

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

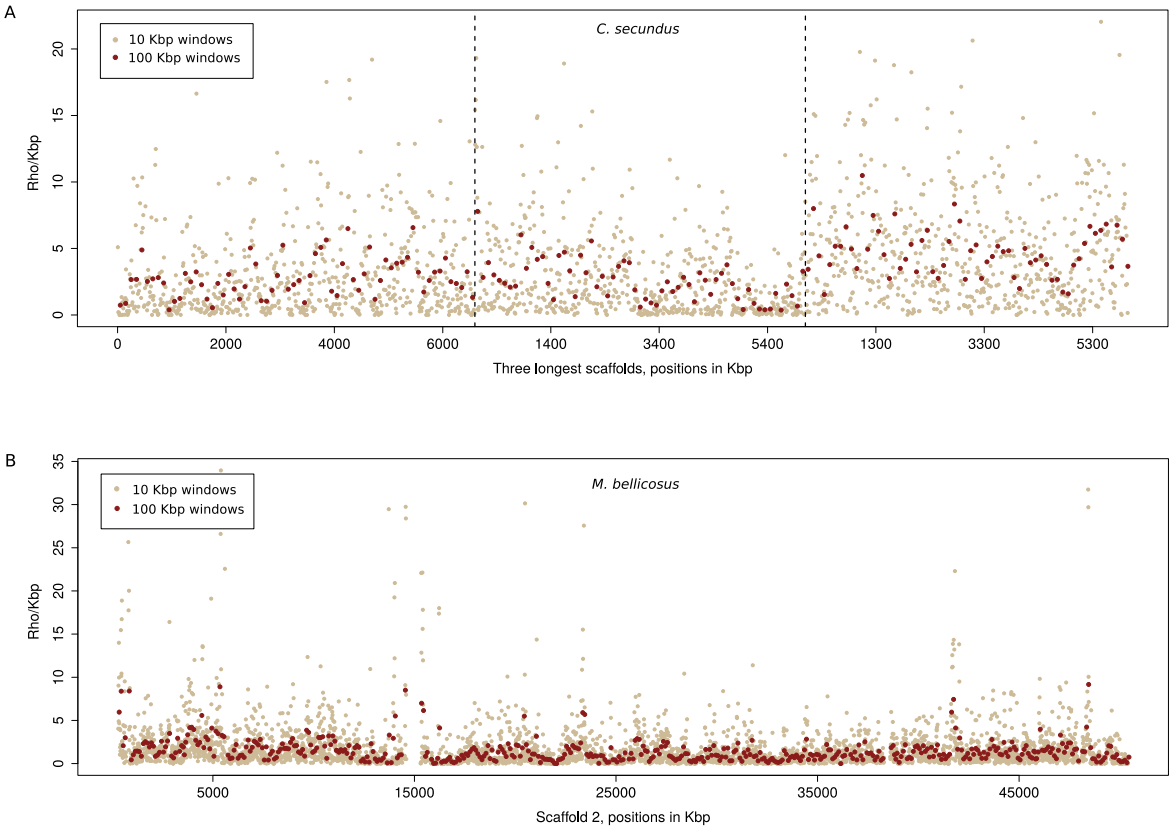


Figure 2. Variation in recombination rate estimated a ρ/kbp across representative genomic scaffolds in a) *C. secundus* and b) *M. bellicosus*. Estimates of ρ/kbp are calculated in 10 Kbp and 100 Kbp windows.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

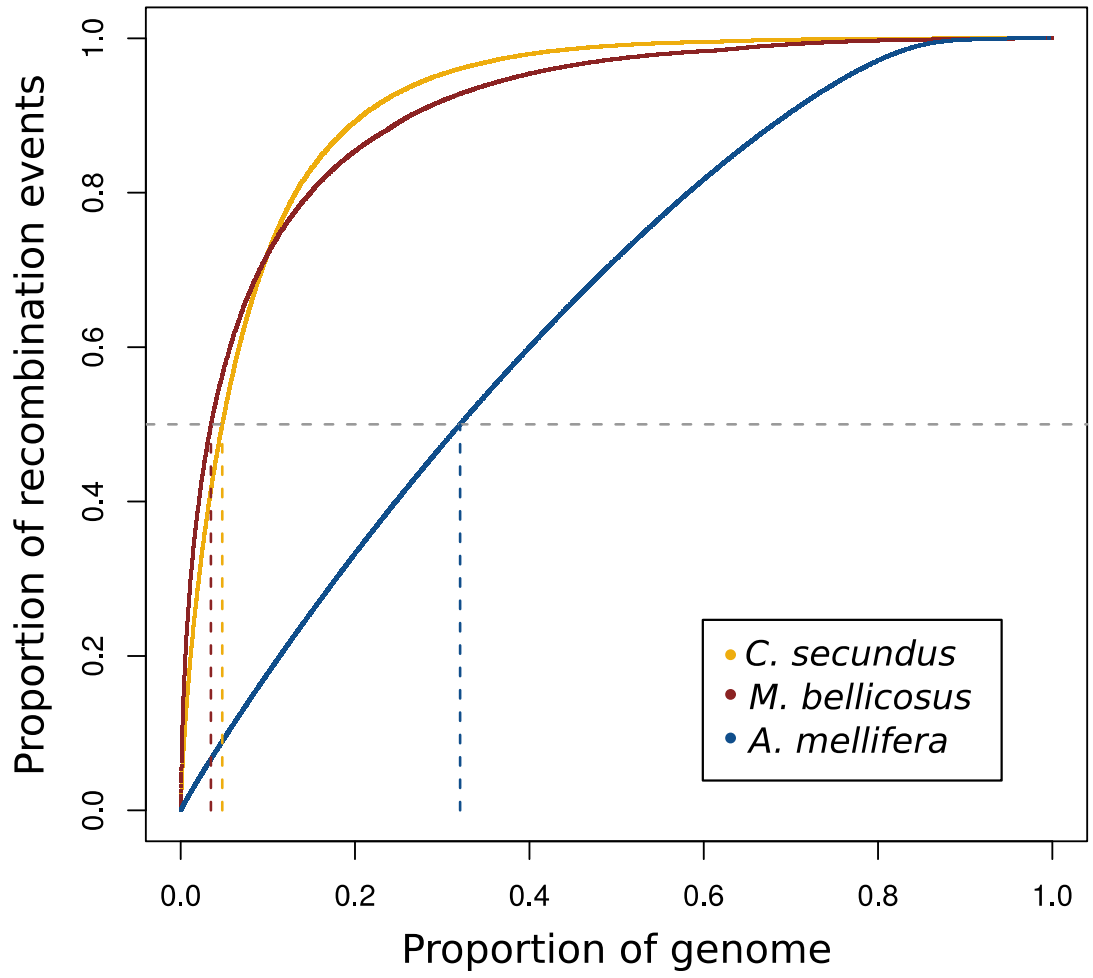
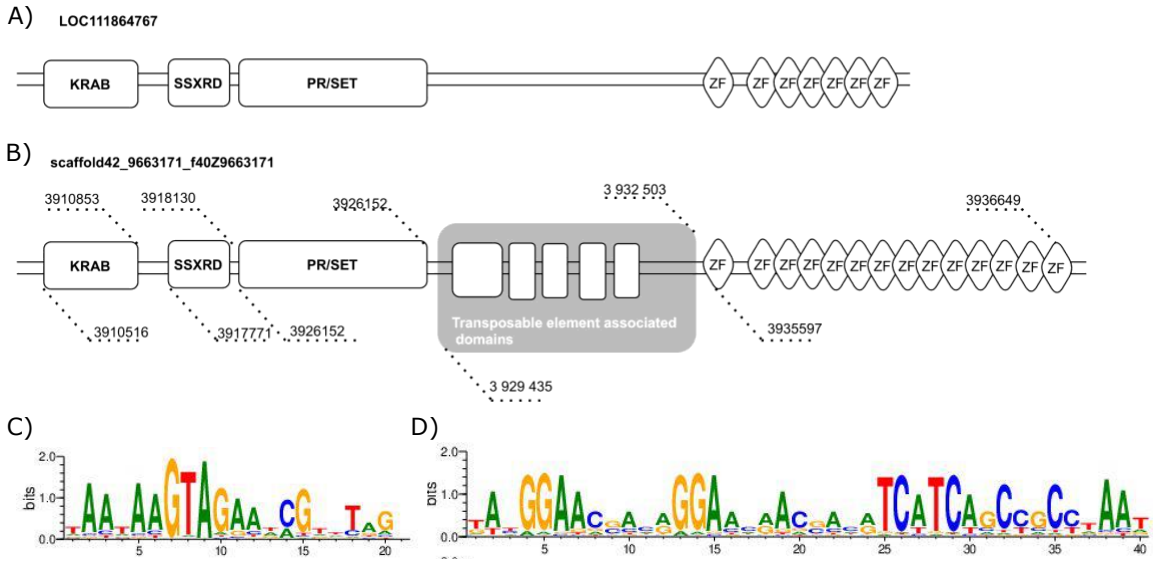


Figure 3. Cumulative plot of proportion of recombination events versus proportion of the genome for the two termite species *C. secundus* and *M. bellicosus* compared to the honey bee *A. mellifera* (Wallberg et al. 2015). Dashed lines show the proportion of the genome where 50 % of the recombination occurs, which for *C. secundus* is 4.77 %, *M. bellicosus* is 3.47 % and *A. mellifera* is 32.05 %.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

1004



1005

1006

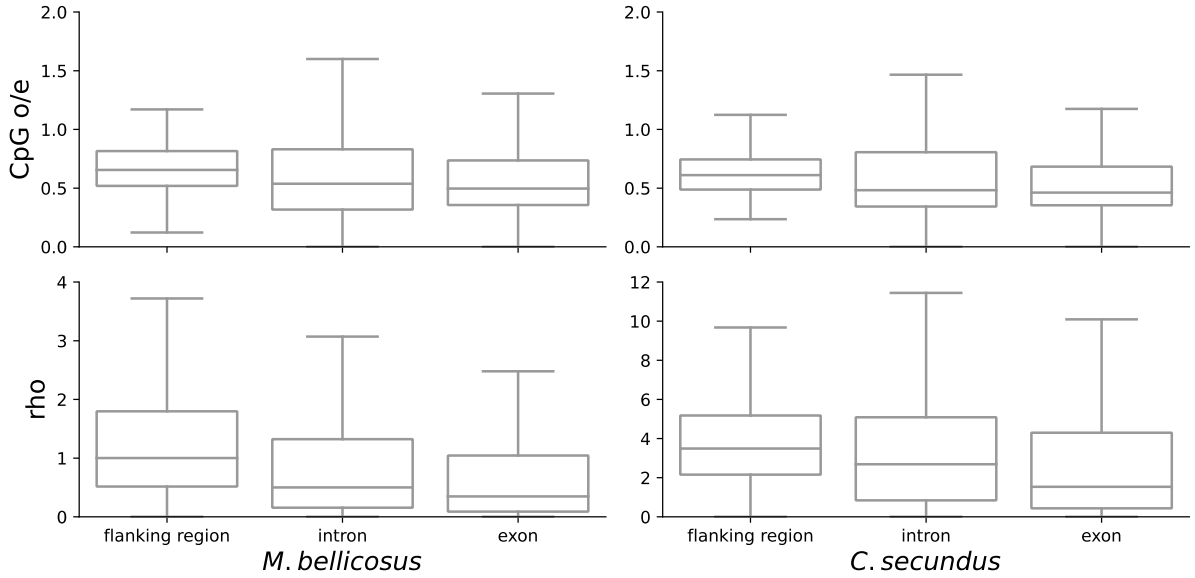
1007 **Figure 4.** PRMD9 structure and predicted zinc-finger motifs. A) Structure of PRDM9 in *C.*
1008 *secundus* genome annotation, B) predicted structure of PRDM9 in *M. bellicosus* identified by
1009 homology searches. C) predicted zinc finger binding motif in *C. secundus*, D) predicted zinc
1010 finger binding motif in *M. bellicosus*.

1011

1012

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

1013



1014

1015

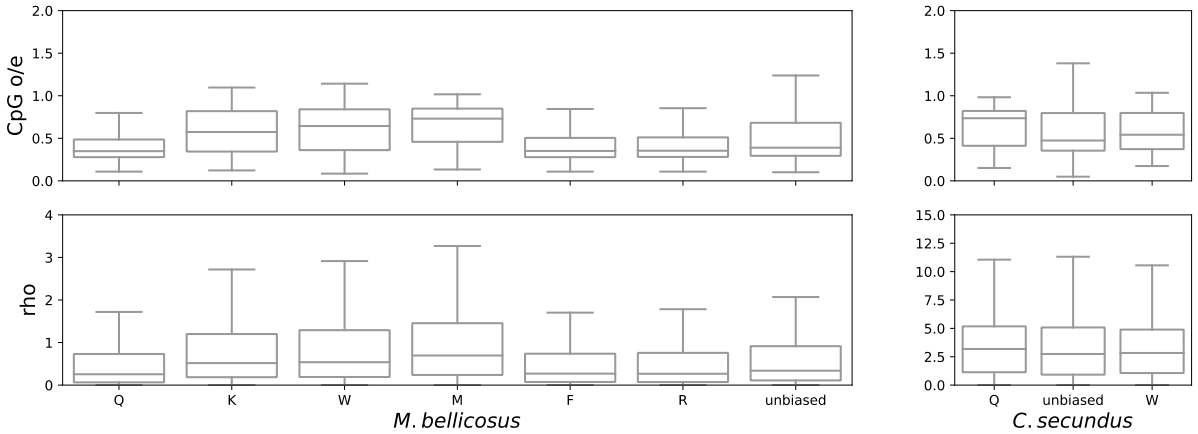
1016 **Figure 5.** Boxplots showing variation in levels of $CpG_{O/E}$ and recombination rate (ρ/kbp; rho)
1017 in genic and flanking regions in the *M. bellicosus* and *C. secundus* genomes. The differences
1018 between all categories are significant.

1019

1020

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.22.586269; this version posted March 25, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

1021



1022

1023

1024 **Figure 6**. Boxplots showing variation in levels of $CpG_{O/E}$ and recombination rate (ρ/kbp) in
1025 the coding region of genes in the *M. bellicosus* and *C. secundus* genomes classified according
1026 to their patterns of gene expression (Q = queen-biased, K = king-biased, W = worker-biased,
1027 F = female-biased, R = reproductive-biased).

1028