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

# equal contribution

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

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 *Macrotermes bellicosus* and *Cryptotermes secundus*. 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, which suggests recombination hotspots in termites might be determined by PRDM9, as they are in mammals. 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.

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).

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). Hymentoptera belong to the superorder Holometabola, which is the most diverse insect superorder and contains 11 orders including Lepidoptera (butterflies, moths), Coleoptera (beetles), and Diptera (true flies), which all undergo complete metamorphosis. Blattodea is a hemimetabolous order, which undergoes partial metamorphosis. 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

107 include nearly all mammals and many vertebrate taxa (Baker et al. 2017). In species without
108 PRDM9, recombination can be directed to hotspots defined by other features, such as CpG
109 islands and promoters that have open chromatin. Such hotspots have been characterised in
110 the genomes of diverse taxa including yeast, birds, and dogs (Axelsson et al. 2012; Lam and
111 Keeney 2015; Singhal et al. 2015; Berglund et al. 2014). In insects, the prevalence of
112 recombination hotspots is not well studied. Fine-scale recombination maps in the fruit fly
113 *Drosophila melanogaster* and honey bee *A. mellifera* indicate a paucity of recombination
114 hotspots (Chan et al. 2012; Smukowski Heil et al. 2015; Wallberg et al. 2015), whereas
115 hotspots appear to be present in the genome of the butterfly *Leptidea sinapis* (Torres et al.
116 2023).

117 Several other factors may also modulate recombination rate along chromosomes, such as
118 germline DNA methylation. Many vertebrate species have elevated recombination in CpG
119 islands, which are usually unmethylated in the germline (Axelsson et al. 2012; Singhal et al.
120 2015; Berglund et al. 2014). In insect genomes, methylation is not always present, but
121 commonly restricted to gene bodies (Arsala et al. 2022; Wang et al. 2013; Lyko et al. 2010;
122 Ventós-Alfonso et al. 2020; Bewick et al. 2019; Harrison et al. 2018). In the honey bee *A.*
123 *mellifera* and solitary bee *M. rotundata* recombination is reduced in genes, and particularly
124 in those inferred to have the highest levels of germline methylation (Wallberg et al. 2015;
125 Jones et al. 2019). In addition, some studies have identified correlations between caste-
126 biases in gene expression and recombination. In honey bees, genes with worker-biased gene
127 expression have been found to have elevated recombination rates (Kent et al. 2012) but this
128 is likely an indirect result of differences in germline methylation between sets of genes
129 (Wallberg et al. 2015).

130 Here we estimate recombination rate and its fine-scale variation across the genomes of two
131 distantly related termite species with contrasting social complexities (Korb and Hartfelder
132 2008; Korb and Thorne 2017). *Macrotermes bellicosus* (Termitidae: Macrotermitinae) is a
133 fungus-growing termite with a high level of social complexity, with colonies of several
134 millions of individuals and sterile workers. It belongs to the foraging (multiple-pieces
135 nesting) termites, where workers leave the nest to forage for food. The drywood termite
136 *Cryptotermes secundus* (Kalotermitidae) has a low level of social complexity. Its colonies
137 consist of a few hundred individuals and totipotent workers from which queens and kings

138 develop (Hoffmann et al. 2012). It belongs to the wood-dwelling (one-piece nesting)
139 termites that nest in a single piece of wood that also serves as their food source.

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

150 **Results**

151 *Genetic variation in two termite species is typical for social insects*

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

157 We sequenced 10 unrelated individuals each of *M. bellicosus* and *C. secundus* to a mean
158 depth of 34x and 31x per sample, respectively (Supplementary Table S1). We mapped reads
159 to the appropriate genome assemblies and called variants in each species. In *M. bellicosus*
160 we called 5.6 million SNPs and in *C. secundus* 15.1 million SNPs (Table 1). One individual of
161 *C. secundus* (CS_8) was excluded due to a low proportion of mapped reads, poor mapping
162 quality and strand bias in mapping. Estimates of variation based on θ_W (Watterson 1975) are
163 0.14% and 0.44% in *M. bellicosus* and *C. secundus* respectively, which are broadly
164 comparable to levels of variation in the honey bee *A. mellifera* (0.3 - 0.8%) (Wallberg et al.
165 2014).

166 We used these estimates of genetic variation to estimate effective population size (N_E) for
167 each of the sampled populations (Table 1). These rely on an estimate of mutation rate, μ, in

168 each species. As there is no estimate of μ for Blattodea, we considered a range of estimates
169 from insects compiled by (Lynch et al. 2023) that includes estimates from Diptera
170 (*Drosophila*) (Keightley et al. 2009, 2014), Hymenoptera (Liu et al. 2017; Yang et al. 2015)
171 and Lepidoptera (Keightley et al. 2015). These estimates range between a minimum of 2.7 x
172 10^{-10} $bp^{-1}gen^{-1}$ in *Acyrthosiphon pisum* (Fazalova and Nevado 2020) to a maximum of 8.1 x
173 10^{-9} $bp^{-1}gen^{-1}$ in *Drosophila pseudoobscura* (Krasovec 2021). The average of multiple
174 estimates in *Drosophila melanogaster* is 4.5 x 10^{-9} $bp^{-1}gen^{-1}$, which is typical of insects (Lynch
175 et al. 2023).

176 Our estimates of N_E based on θ_W are 42,898 – 1,278,982 in *M. bellicosus*, which has high
177 social complexity, and 136,869 – 4,080,725 in *C. secundus*, which has a lower level of social
178 complexity (Table 1). Estimates of N_E assuming the typical mutation rate of 4.5 x 10^{-9} bp^{-}
179 $^{1}gen^{-1}$ are 76,876 and 245,280 respectively. These estimates are consistent with previous
180 evidence that eusocial species with large colonies and high social complexity tend to have
181 lower effective population sizes and the values are broadly comparable to levels of N_E found
182 in social and solitary Hymenoptera (Leffler et al. 2012; Romiguier et al. 2014).

183 *Low average recombination rate in termite genomes*

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

190 We estimated variation in the population recombination rate, ρ, in both datasets using
191 LDhat (Auton and McVean 2007) (Table 1). Average rates of ρ/kb are 1.7 and 4.8 in *M.*
192 *bellicosus* and *C. secundus* respectively. These estimates can be converted to cM/Mb using
193 estimates of N_E, which gives estimates of average recombination rate of 0.033 - 0.986
194 cM/Mb in *M. bellicosus* and 0.030 - 0.885 cM/Mb in *C. secundus*, which is much lower than
195 the average among insects (Stapley et al. 2017), and particularly lower than the elevated
196 recombination rates inferred in eusocial Hymenoptera. Estimates that assume a mutation
197 rate close to the average for insects are 0.550 and 0.494 cM/Mb in *M. bellicosus* and *C.*

198 *secundus* respectively. This represents the first estimation of recombination rate in social
199 insects outside of Hymenoptera and demonstrates that eusociality is not a universal driver
200 of high recombination rates.

201 *Evidence for active recombination hotspots in termite genomes*

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

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

217 We next defined hotspots as 2 kbp regions with more than 5-fold higher recombination rate
218 than the local background as defined as the surrounding 100 kbp and used STREME (Bailey
219 2021) to identify 8 to 20-mers that were enriched in hotspots compared to the rest of the
220 genome. This yielded significantly enriched motifs in both species (8 motifs for *M. bellicosus*,
221 and 12 motifs for *C. secundus* after multiple testing; Supplementary Tables S2 and S3).
222 However, none the motifs we identified are present in more than 5% of hotspots and exhibit
223 more than 2-fold difference relative to the background, suggesting they are not viable
224 candidate motifs for promoting recombination in hotspots.

225 *Identification and characterization of the PRDM9 gene in termites*

226 We searched the *M. bellicosus* and *C. secundus* genomes for evidence of complete PRDM9
227 genes, which could potentially govern the presence of recombination hotspots in these

228 genomes, as it has been shown to do in mammalian genomes (Myers et al. 2010; Baudat et
229 al. 2010; Parvanov et al. 2010). The *C. secundus* genome contains a PRDM9 homolog
230 (XP_023708049.2) with annotated KRAB, SSXRD, SET and zinc finger domains (Figure 4).
231 There is no complete homolog of PRDM9 in the *M. bellicosus* annotation. We therefore used
232 BLAST to search for sequences homologous to the conserved functional domains from *C.*
233 *secundus*. We identified a putative PRDM9 orthologue on Scaffold 42, containing all
234 domains in the correct order. We identified retroviral conserved domains in an intron 5' of
235 the zinc-finger domains in this gene in the *M. bellicosus* genome assembly, which are not
236 predicted to alter the transcript. We confirmed the presence of a full-length transcript as
237 well as presence and correct order of conserved domains by alignment of the
238 *Zoothermopsis nevadensis* PRDM9 protein sequence to the *M. bellicosus* genome using
239 GeneWise (Madeira et al. 2024)(Figure 4 and Supplementary Table S4).

240 We used a zinc-finger prediction algorithm (Persikov and Singh 2014) to predict DNA binding
241 motifs from the zinc-finger domain present in PRDM9 in the *M. bellicosus* and *C. secundus*
242 genome assemblies, for which an accurate annotation was available. The most probable
243 consensus motifs were TATGGAACGACAGGAACAACGACATCATCAGCCGCCTAAT and
244 TAATAAGTAGAATCGTTTAG for *M. bellicosus* and *C. secundus* respectively. Base probability
245 matrices for these motifs are shown in Supplementary Table S5. We tested whether the
246 presence of these motifs corresponded to elevated recombination considering 1 kbp around
247 the motif. However, recombination rate was not significantly elevated around these
248 predicted motifs for either species, either when searching based on the most likely motif, or
249 by searching based on the probability matrices.

250 *Genomic correlates of recombination rate*

251 We analysed genomic correlates of recombination rate in 10 kbp windows. There is a
252 significant correlation between recombination and $CpG_{O/E}$ in both species with Spearman's ρ
253 = 0.24, $p < 1 \times 10^{-4}$ for *M. bellicosus* and Spearman's $\rho = 0.10$, $p < 1 \times 10^{-4}$ for *C. secundus*
254 (Figure S2). A reduction of recombination rates in regions of low CpG content is consistent
255 with interference of germline methylation with recombination. However, as the correlation
256 coefficients are low, the explanatory power is weak. Correlations of similar magnitudes were
257 observed between recombination and GC content (Spearman's ρ =0.017 for *M. bellicosus*, p
258 $< 1 \times 10^{-4}$ and Spearman's ρ =0.23 for *C. secundus*, $p < 1 \times 10^{-4}$). This correlation has been

259 associated with an effect of recombination in driving GC through GC biased gene conversion
260 but the relatively low correlation coefficients indicate that this force is not strong in
261 termites. By contrast, the correlation between GC content and recombination rate is much
262 stronger in honey bee (R^2 = 0.506), which is likely caused by high recombination rates and
263 intense GC-biased gene conversion (Wallberg et al. 2015).

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

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

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

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

289 and other insect genomes with functional DNA methylation (Harrison et al. 2018; Elango et
290 al. 2009; Sarda et al. 2012). We found that exons have lower $CpG_{O/E}$ compared to introns,
291 which are both lower than flanking regions (Figure 5, top row; all comparisons $p < 0.001$
292 after Bonferroni correction) consistent with higher levels of germline methylation in exons
293 (Supplementary Table S6). We next analysed how recombination rate varies among genomic
294 features. We found that ρ/kbp is also significantly reduced in genes compared to flanking
295 regions, with exons showing reduced ρ/kbp compared to introns (Figure 5, bottom row, all
296 p-values <0.005 after Bonferroni correction). This effect was seen in both species and is
297 more pronounced for *C. secundus* than for *M. bellicosus.* There was a significant correlation
298 between $CpG_{O/E}$ and recombination rate among genes (Spearman's $\rho = 0.369$, $p < 1 \times 10^{-4}$ for
299 *M. bellicosus*; Spearman's $\rho = 0.166$, $p < 1 \times 10^{-4}$ for *C. secundus*).

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

316 The differences in $CpG_{O/E}$ between caste-biased gene expression categories are mirrored by
317 differences in ρ/kbp (Figure 6). Gene expression categories with elevated $CpG_{O/E}$
318 consistently show elevated ρ/kbp, and this pattern is found in both species. Studies in the
319 honey bee have found that worker-biased genes have lower CpG and higher recombination

320 rates (Kent et al. 2012; Wallberg et al. 2015). In *M. bellicosus,* we find that worker-biased
321 genes have slightly elevated $CpG_{O/E}$ and ρ/kbp. However, in *C. secundus*, these values are
322 reduced in worker-biased genes, whereas queen-biased genes have a slightly elevated
323 $CpG_{O/E}$ and ρ/kbp in this species.

324 We investigated the relative roles of $CpG_{O/E}$ and gene expression in determining ρ in genes
325 using a linear model, with $CpG_{O/E}$, ρ in flanking region and expression categories as
326 explanatory variables. In both *M. bellicosus* and *C. secundus*, we found that $CpG_{O/E}$ was the
327 strongest predictor of recombination rate variation among genes (Supplementary Tables S7
328 and S8) in line with the analyses above. We found that caste-biased expression had no
329 significant impact on the recombination rate, with the exception of king-biased genes in *M.*
330 *bellicosus*. These have higher ρ/kbp than expected based on $CpG_{O/E}$ alone. Generally, $CpG_{O/E}$
331 is a good predictor of ρ/kbp, and the variation in ρ/kbp among expression categories likely
332 mainly reflects differences in $CpG_{O/E}$, as accounting for it removes most significant
333 interactions.

334

335 **Discussion**

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

346 *Low recombination rates in termites*

347 We estimate that recombination rates in both termite species are less than 1 cM/Mb. These
348 estimates of recombination rate per meiosis depend on estimates of mutation rate. This is
349 not known in termites so we considered a range of estimates from insects. Data from a wide

range of taxa indicate that at least one crossover per chromosome tetrad is essential for correct chromosomal segregation during meiosis. The chromosome numbers *M. bellicosus* and *C. secundus* are 2n = 42 and 2n = 40 respectively (Jankásek et al. 2021). Considering the assembly lengths, this corresponds to a minumum average genomic recombination rate of 0.92 cM/Mb and 1.01 cM/Mb would be necessary for accurate meiosis to occur in the two species respectively. These values are closest to the maximum rates of 0.99 and 0.89 cM/Mb estimated when using the mutation rate estimated from *D. pseudoobscura* which is the highest estimate reported in insects (Lynch et al. 2023). The estimates based on higher mutation rates are therefore more likely to be true, indicating that termite mutation rates are likely to be among higher range for insects and that Ne in this species is relatively low (~40,000 for *M. bellicosus* and 140,000 for *C. secundus* based on Table 1).

It should be noted that our estimates of recombination rate are sex-averaged rates. It is possible that recombination is absent in one of the termite sexes, which would make a lower sex-averaged recombination rate more feasible. In addition, sex-linked contigs have not been identified in the termite genome assemblies utilized here. *M. bellicosus* and *C. secundus* have X1X2/Y1Y2 and X/Y sex-determining systems respectively (Jankásek et al. 2021). The inclusion of sex chromosomes in the analysis leads to errors in determining the genome-wide recombination rate per meiosis because they have lower N_E than the rest of the genome. In addition, sex chromosomes are more prone to assembly and read mapping errors, due to reads aligning with the incorrect copy of the chromosome. However, considering that sex chromosomes only comprise a minor portion of the genome, this is unlikely to substantially alter our results.

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-

381 specifically expressed genes (Kent and Zayed 2013). In addition, increased recombination
382 could lead to reduced variance in relatedness, which could prevent kin conflict in colonies
383 (Sherman 1979; Templeton 1979).

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

392 The chromosome numbers of both termite species (*M. bellicosus*, 2n = 42; *C. secundus*, 2n =
393 40) are typical for termites, although variation exists between taxa (Jankásek et al. 2021).
394 These values do not differ from those observed in Hymenoptera, which have a similar range
395 (e.g. *A. mellifera*, 2n = 32) and do not vary between solitary and eusocial taxa (Cardoso et al.
396 2018; Cunha et al. 2021; Ross et al. 2015). This suggests that differences in chromosome
397 number are unlikely to be relevant in explaining the differences in recombination rates
398 among these taxa.

399 One proposed advantage of sex and recombination is that it is associated with sexual
400 selection and a higher intensity of selection on males. This could be advantageous because
401 differential male mating success can reduce mutational load in sexual populations by
402 causing deleterious mutations to segregate at a lower equilibrium frequency on average
403 (Siller 2001; Agrawal 2001). Recombination is essential for effective purging of deleterious
404 alleles as it unlinks deleterious variants from nearby beneficial ones thus preventing
405 selection interference (Hartfield and Keightley 2012). Selection has been found to be
406 stronger in males than females across animal species (Janicke et al. 2016; Winkler et al.
407 2021). Species with a male-biased sex ratio are expected to experience particularly intense
408 selection on males. It is therefore possible that recombination rate could be particularly
409 elevated in species in which males face intense competition as this increases the
410 evolutionary advantage of recombination.

411 The genus *Apis* produces highly male-biased reproductive offspring and thousands of male
412 drones compete to fertilise a small number of female queens (Winston 1991; Koeniger et al.
413 2011; Baer 2005). It is also in this genus that the most extreme recombination rates have
414 been inferred (Beye et al. 2006; Shi et al. 2013; Liu et al. 2015; Wallberg et al. 2015;
415 Rueppell et al. 2016; Kawakami et al. 2019). Male-biased sex ratios are also observed in
416 other social bees and wasps. Both male and female biased operational sex ratios have been
417 observed in ants, but are difficult to measure because workers invest more resources in
418 female offspring and may manipulate sex ratios. In non-social Hymenoptera, females lay
419 roughly equal numbers of haploid and diploid eggs (Trivers and Hare 1976). The sex ratio of
420 reproductive males and females is also relatively even in most termite species (Roisin 2001).
421 It is possible that the male-biased sex ratio observed in many eusocial Hymenoptera could
422 favour elevated recombination rates in the species so far studied.

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

429 Although its effects on overall recombination rates are unclear, haploid selection is believed
430 be key to driving sex-specific differences in recombination rate (Immler 2019). It has been
431 argued that differences in the intensity of selection between sexes in the haploid phase can
432 drive differences in recombination rates between sexes among animal and plant species
433 (heterochiasmy) (Burt et al. 1991; Trivers 1988; Lenormand 2003; Lenormand and Dutheil
434 2005). It is unclear how theory and observations regarding the prevalence and nature of
435 heterochiasmy applies to haplodiploids, in which only females recombine. However, it
436 appears that the intensity of selection on the haploid phase likely has the potential to drive
437 selective pressure on recombination rates.

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

442 and non-social species to determine whether the strength of selection on males is a key
443 factor in determining these rates.

444 *Full length PRDM9 gene and recombination hotspots in termite genomes*

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

451 Two main mechanisms have been shown to lead to recombination hotspots. The first, which
452 is best understood in human and mouse, is governed by the PRDM9 protein. A zinc-finger
453 domain recognises a specific DNA sequence motif and then performs a histone modification
454 in the vicinity, which marks the sequence for a DNA double-stranded break that is repaired
455 by recombination during meiosis (Baudat et al. 2010; Myers et al. 2010; Parvanov et al.
456 2010). In species with this mechanism, the zinc finger domain is fast evolving, which results
457 in a rapid turnover of hotspot locations (Myers et al. 2010). Full-length PRDM9 orthologues
458 are present among a range of highly-diverged vertebrates including many fish, reptiles and
459 mammals, but many full or partial losses of the gene have also been inferred. In a study of
460 225 vertebrate genomes, a minimum of six partial and three complete losses were inferred
461 (Baker et al. 2017). However, it is unclear whether PRMD9-hotspots are common in
462 invertebrates.

463 A second mechanism is found in vertebrates that lack PRDM9, in which recombination
464 hotspots are localised to regions of open chromatin such as CpG islands and promoters. This
465 is found in dogs and in most or all birds, in which PRDM9 has been lost, and results in
466 hotspots with stable rather than rapidly evolving locations (Axelsson et al. 2012; Singhal et
467 al. 2015). A similar pattern is also observed in the yeast *Saccharomyces cerevisiae* in which
468 recombination preferentially occurs in promoter regions (Wu and Lichten 1994). PRDM9-
469 independent hotspots can be identified in CpG islands is some vertebrate genomes even
470 when PRDM9 is present (Joseph et al. 2024). In invertebrates such as *Drosophila* and
471 *Caenorhabditis elegans*, recombination hotspots appear to be absent (Coop and Przeworski
472 2007; Chan et al. 2012; Smukowski Heil et al. 2015). In insects, fine-scale maps of social and

473 solitary bees have not revealed recombination hotspots (Wallberg et al. 2015; Jones et al.
474 2019), but there is evidence for hotspots in a butterfly genome (Torres et al. 2023).

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

490 *Evidence that germline DNA methylation suppresses recombination*

491 In insects with a functioning DNA methyltransferase machinery, the $CpG_{O/E}$ statistic shows
492 substantial variation along the genome, which mainly reflects variation in levels of germline
493 DNA methylation (Elango et al. 2009). In contrast to vertebrate genomes, methylation is
494 strongly skewed towards gene bodies in the genomes of insects (Glastad et al. 2011). The
495 overall genomic correlation between $CpG_{O/E}$ and recombination rate is significant but weak
496 in the two termite genomes. This could reflect that the majority of the genomes are not
497 methylated, so that much of the variation in $CpG_{O/E}$ on the genome scale is not strongly
498 influenced by methylation. In the two termite genomes studied here, we find lowest values
499 of $CpG_{O/E}$ in exons, whereas noncoding flanking regions are biased towards higher values,
500 indicative of lower levels of DNA methylation (Figure 5, Figure S3). All categories display a
501 predominantly bimodal distribution of $CpG_{O/E}$, likely reflective of the presence of
502 methylated and unmethylated regions.

503 The differences in $CpG_{O/E}$ are also reflected by differences in recombination rate (ρ/kb),
504 which are reduced in regions inferred to be more highly methylated. Our inference of
505 reduced recombination rates in exons could be influenced by lower N_E in these regions due
506 to the effects of linked selection, which could result in more extensive LD (Charlesworth
507 2009). However, the observation that differences in $CpG_{O/E}$ among genes are associated
508 with recombination rate indicates that linked selection is unlikely to be a major factor
509 driving the observed differences in recombination rate.

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

522 A direct effect of germline DNA methylation in suppressing recombination could influence
523 variation in recombination landscapes among vertebrates and invertebrates. Vertebrate
524 genomes are usually highly methylated with the exception of CpG islands. In vertebrate
525 genomes that lack a functional PRDM9 gene, recombination events tend to localise to CpG
526 islands (for example in dogs and birds)(Singhal et al. 2015; Axelsson et al. 2012). The effect
527 of DNA methylation on recombination rate variation in invertebrates is less clear, but the
528 results presented here also support a role of germline methylation in suppressing
529 recombination.

530

531 **Materials and Methods**

532 *Sample collection*

533 *Macrotermes bellicosus* samples were collected from colonies in Kakpin, next to the Comoé
534 National Park in Côte d'Ivoire (coordinates 8°39'N 3°46'W) (Elsner et al. 2018). For both
535 species, each of the sequenced samples came from a different colony. *Cryptotermes*
536 *secundus* colonies were collected from dead *Ceriops tagal* mangrove trees near Palmerston-
537 Channel Island (Darwin Harbor, Northern Territory, Australia; 12°30'S 131°00'E). Colonies
538 were then kept in *Pinus radiata* wood blocks in climate rooms in Germany, providing 28 °C,
539 70% relative humidity, and a 12-h day/night cycle (Lin et al. 2021).

540 *Population sequencing*

541 The termites were cut in two, approximately between head and thorax. DNA was extracted
542 from both parts using the QIAGEN Blood and Tissue Kit following the standard protocol,
543 including treating with 4 µL RNase and 25 µL Proteinase K. The DNA extraction product was
544 cleaned and concentrated with ZymoResearch Genomic DNA Clean & Concentrator kit. For
545 most samples, the RNase treatment was repeated prior to cleaning and concentrating. One
546 DNA sample from each individual was chosen, based on the DNA amount and quality, to be
547 sequenced. Sequencing libraries were produced using the Illumina DNA PCR-free library
548 preparation kit according to the manufacturers protocol, using an input DNA quantity of 200
549 ng per sample. This protocol yields insert sizes of approximately 350 bp. We performed
550 Illumina short-read sequencing on the libraries using NovaSeq 6000 S4 flow cell with 2 x 150
551 bp paired end reads.

552 *Read mapping and variant calling*

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

563 together correspond to 3.78% of the genome length, due to unexpectedly high
564 heterozygosity: 93, 62, 41, 35 and 32.

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

570 Variant calling was done using the tools HaplotypeCaller, GenomicsDBImport and
571 GenotypeGVCFs from GATK version 4.3.0.0 as described in their best practices workflow
572 (https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-
573 discovery-SNPs-Indels-, last accessed 2023-07-11). The variants were filtered in multiple
574 steps, first removing extended regions of low mapping quality or deviating read depth as
575 those regions are likely to be unreliable. This was done based on statistics from the
576 Samtools mpileup and depth tools, respectively, removing 100 kb windows with mean
577 mapping quality below 70 or mean read depth more than two standard deviations from the
578 genome-wide mean. In the later recombination rate and LD decay calculations, values in
579 intervals spanning those regions were also removed. Indels were removed using VCFtools
580 version 0.1.16 (Danecek et al. 2011) before GATK VariantFiltration was run with the
581 following limits, which are all equal to or stricter than the ones recommended
582 (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-
583 short-variants, last accessed 2023-07-11) and adapted to the observed distributions of the
584 corresponding statistics: QD<2; FS>50; MQ<40; ReadPosRankSum< -4; ReadPosRankSum>4;
585 SOR>3; ExcessHet>5 for both species as well as MQRankSum with limits -5 and 5 for *C.*
586 *secundus* and limits -6 and 4 for *M. bellicosus*. This was followed by additional filtering with
587 VCFtools to select only bi-allelic sites with a minimum quality score of 30 and maximum of
588 40% missing genotypes. The SNPs with mean read depth in the lowest or highest 2.5% of the
589 depth distribution were also filtered out, which translates to a depth below 10 or above 34
590 for *C. secundus* and below 5 or above 40 for *M. bellicosus.* For the recombination rate
591 calculations, the SNPs were also filtered for a minor allele count >= 2, as rare variants
592 appearing on only one chromosome have a relatively high false positive rate and are not
593 informative regarding LD. The average proportion of missing genotypes per individual was

594 4.5 x 10^{-5} for *C. secundus* and 1.2 x 10^{-4} for *M. bellicosus*. After variant calling and filtering,
595 the data was phased and imputed using Beagle version 5.1 (Browning and Browning 2007;
596 Browning et al. 2018).

597 *Estimation of LD decay*

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

608 *Estimation of population recombination rate*

609 The recombination rate was estimated using the complete, interval and stat functions of
610 LDhat (https://github.com/auton1/Ldhat, downloaded 2023-04-19) (Auton and McVean
611 2007) with 10 million iterations, sampling every 5000th iteration, taking the first 20 samples
612 as burn-in.

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

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

622 The parameter ρ estimated by LDhat is proportional to the recombination rate r and the
623 effective population size N_E, which in turn can be estimated based on the mutation rate μ

624 and the number of segregating sites, S. In order to convert ρ to *r*, the following equations
625 were used, where n is the number of haploid sequences. The number of segregating sites
626 was calculated before the filter on minor allele count >=2 was applied, as this filter likely
627 removes many true SNPs as well, which could affect the estimates.

628

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

630 $$\theta_W = K \Big/ \sum_{i=1}^{n-1} \frac{1}{i}$$

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

633 $$N_E = \frac{\theta}{4\mu}$$

634 where μ is the mutation rate. We estimated N_E using a range of mutation rates estimated
635 from other insect species (Keightley et al. 2009, 2014, 2015; Yang et al. 2015; Liu et al. 2017;
636 Lynch et al. 2023). We then estimated the recombination rate using the following equation,
637 assuming a range of values of N_E.

638 $$r = \frac{\rho}{4N_e}$$

639

640 *Cumulative plot of rho* ρ/kbp

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

648 *k-mer enrichment in recombination hotspots*

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

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

660 *Identifying PRDM9 orthologues in M. bellicosus and C. secundus*

661 A full-length copy of PRDM9 is present in the *C. secundus* genome annotation, but not in *M.*
662 *bellicosus*. Using Orthofinder (Emms and Kelly 2019), and the genome annotation files for
663 both species, we identified a single orthologue to the *C. secundus* PRDM9 in *M. bellicosus*.
664 We analysed this region using the NCBI conserved domain search (Wang et al. 2023). We
665 also searched the *M. bellicosus* assembly for the PR/SET domains derived from *C. secundus*
666 and *Zootermopsis nevadensis* using BLAST (Camacho et al. 2009) and subsequently for all
667 other PRDM9 domains (KRAB, SSXRD, ZF-Casette) to identify hits that contain all other
668 PRDM9 domains in the vicinity in the correct order. We used GeneWise (Madeira et al.
669 2024) with default options to perform gapped alignment of a the PRDM9 protein sequence
670 from *Zoothermopsis nevadensis* against the target region to predict exon-intron borders.

671 *Predicting binding sites from zinc-finger-motifs*

672 Using the ZF-motifs from *C.secundus,* we obtained letter probability matrices for predicted
673 binding motifs using a webserver (http://zf.princeton.edu/index.php ; (Persikov and Singh
674 2014) which were used to search for binding sites across the genome using FIMO (Grant et
675 al. 2011) with default options. Subsequently, the mean rho for the surrounding 1 kb bin was
676 used to test for a difference in mean rho between bins associated with a suspected binding
677 site, and a similar number of randomly chosen bins without association. For significance

678 testing, we repeated this draw 10.000 times to compare the associated bins against the 95%
679 confidence interval of the resulting distribution.

680 *Analysis of genomic correlates of recombination*

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

687 In order to determine the repeat content across the *M. bellicosus* and *C. secundus* genome
688 assemblies, we first generated custom repeat libraries for each species with RepeatModeler
689 (Flynn et al. 2020) using default options. This was then combined with Repbase 29 (Bao et
690 al. 2015) and DFAM 3.8 (Storer et al. 2021) and used to run RepeatMasker with default
691 options on each genome. We estimated the proportion of sequence in each repeat class in
692 10 kbp windows and used Spearman's rank correlation to correlate this with estimates of
693 recombination rate.

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

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

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

703 *Differential expression data*

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

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

715 Lin et al. (2021), performed RNAseq on samples of workers and queens. Here, we used the
716 data from untreated queens and workers. Differentially expressed *C. secundus* genes were
717 classified as “worker-biased” or “queen-biased” if they were differentially expressed with a
718 bias to the respective caste, or unbiased if they were not, and subsequently compared
719 between classes. The lists of genes we identified in the caste-biased categories in both
720 species are provided as Supplementary Tables S9 and S10.

721 *Correlates of recombination rate variation among genes*

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

727

728 **Data access**

729 Custom scripts are available on GitHub at https://github.com/troe27/termite-
730 recombination-rate and included as Supplemental Material. The Illumina whole-genome
731 sequencing data generated in this study have been submitted to the NCBI BioProject
732 database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number
733 PRJNA1021607.

734

735 **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.

References

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

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

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

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

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

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

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

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

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

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

776 Berglund J, Quilez J, Arndt PF, Webster MT. 2014. Germline methylation patterns determine
777 the distribution of recombination events in the dog genome. *Genome Biol Evol* **7**:
778 522–530.

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

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

785 Browning BL, Zhou Y, Browning SR. 2018. A One-Penny Imputed Genome from Next-
786 Generation Reference Panels. *The American Journal of Human Genetics* **103**: 338–
787 348.

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

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

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

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

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

800 Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and
801 patterns of molecular evolution and variation. *Nat Rev Genet* **10**: 195–205.

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

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

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

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

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

814 Fazalova V, Nevado B. 2020. Low Spontaneous Mutation Rate and Pleistocene Radiation of
815 Pea Aphids. *Mol Biol Evol* **37**: 2045–2051.

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

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

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

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

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

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

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

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

834 Immler S. 2019. Haploid Selection in “Diploid” Organisms. *Annual Review of Ecology,*
835 *Evolution, and Systematics* **50**: 219–236.

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

838 Jankásek M, Kotyková Varadínová Z, Šťáhlavský F. 2021. Blattodea Karyotype Database. *EJE*
839 **118**: 192–199.

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

843 Joseph J, Prentout D, Laverré A, Tricou T, Duret L. 2024. High prevalence of PRDM9-
844 independent recombination hotspots in placental mammals. *Proceedings of the*
845 *National Academy of Sciences* **121**: e2401973121.

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

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

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

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

858 Kent CF, Minaei S, Harpur BA, Zayed A. 2012. Recombination is associated with the
859 evolution of genome structure and worker behavior in honey bees. *PNAS* **109**:
860 18012–18017.

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

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

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

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

870 Korb J, Thorne B. 2017. Sociality in Termites. In *Comparative Social Evolution* (eds. D.R.
871 Rubenstein and P. Abbot), pp. 124–153, Cambridge University Press, Cambridge

872 https://www.cambridge.org/core/books/comparative-social-evolution/sociality-in-
873 termites/11A80CB9FA947760889676E03899E85C (Accessed August 9, 2023).

874 Krasovec M. 2021. The spontaneous mutation rate of Drosophila pseudoobscura. *G3*
875 *(Bethesda)* **11**: jkab151.

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

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

881 Lenormand T. 2003. The evolution of sex dimorphism in recombination. *Genetics* **163**: 811–
882 22.

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

885 Lin S, Werle J, Korb J. 2021. Transcriptomic analyses of the termite, Cryptotermes secundus,
886 reveal a gene network underlying a long lifespan and high fecundity. *Commun Biol* **4**:
887 1–12.

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

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

894 Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. 2010. The Honey Bee
895 Epigenomes: Differential Methylation of Brain DNA in Queens and Workers. *PLoS Biol*
896 **8**: e1000506.

897 Lynch M, Ali F, Lin T, Wang Y, Ni J, Long H. 2023. The divergence of mutation rates and
898 spectra across the Tree of Life. *EMBO reports* **24**: e57561.

899 Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, Lopez R, Butcher S.
900 2024. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024.
901 *Nucleic Acids Research* gkae241.

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

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

907 Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, Macfie TS, McVean G, Donnelly P.
908 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic
909 recombination. *Science* **327**: 876–9.

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

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

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

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

919 Qiu B, Elsner D, Korb J. 2023. High-quality long-read genome assemblies reveal evolutionary
920 patterns of transposable elements and DNA methylation in termites.
921 2023.10.31.564968.
922 https://www.biorxiv.org/content/10.1101/2023.10.31.564968v1 (Accessed January
923 8, 2024).

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

926 Romiguier J, Lourenco J, Gayral P, Faivre N, Weinert LA, Ravel S, Ballenghien M, Cahais V,
927 Bernard A, Loire E, et al. 2014. Population genomics of eusocial insects: the costs of a
928 vertebrate-like effective population size. *Journal of Evolutionary Biology* **27**: 593–
929 603.

930 Ross L, Blackmon H, Lorite P, Gokhman VE, Hardy NB. 2015. Recombination, chromosome
931 number and eusociality in the Hymenoptera. *Journal of Evolutionary Biology* **28**:
932 105–116.

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

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

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

942 Sherman PW. 1979. Insect Chromosome Numbers and Eusociality. *The American Naturalist*
943 **113**: 925–935.

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

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

948 Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B,
949 Balakrishnan CN, et al. 2015. Stable recombination hotspots in birds. *Science* **350**:
950 928–932.

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

954 Sirviö A, Johnston JS, Wenseleers T, Pamilo P. 2011a. A high recombination rate in eusocial
955 Hymenoptera: evidence from the common wasp Vespula vulgaris. *BMC Genetics* **12**:
956 95.

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

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

963 Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. 2017. Variation in
964 recombination frequency and distribution across eukaryotes: patterns and
965 processes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**:
966 20160455.

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

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

972 Torres AP i, Höök L, Näsvall K, Shipilina D, Wiklund C, Vila R, Pruisscher P, Backström N.
973 2023. The fine-scale recombination rate variation and associations with genomic
974 features in a butterfly. *Genome Res* **33**: 810–823.

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

977 Trivers RL, Hare H. 1976. Haploidploidy and the Evolution of the Social Insect. *Science* **191**:
978 249–263.

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

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

984 Wallberg A, Bunikis I, Pettersson OV, Mosbech M-B, Childers AK, Evans JD, Mikheyev AS,
985 Robertson HM, Robinson GE, Webster MT. 2019. A hybrid de novo genome assembly
986 of the honeybee, Apis mellifera, with chromosome-length scaffolds. *BMC Genomics*
987 **20**: 275.

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

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

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

998 Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, Marchler GH, Song JS,
999 Thanki N, Yamashita RA, et al. 2023. The conserved domain database in 2023.
1000 *Nucleic Acids Res* **51**: D384–D388.

1001 Wang X, Wheeler D, Avery A, Rago A, Choi J-H, Colbourne JK, Clark AG, Werren JH. 2013.
1002 Function and Evolution of DNA Methylation in Nasonia vitripennis. *PLoS Genet* **9**:
1003 e1003872.

1004 Watterson GA. 1975. On the number of segregating sites in genetical models without
1005 recombination. *Theoretical Population Biology* **7**: 256–276.

1006 Wilfert L, Gadau J, Schmid-Hempel P. 2007. Variation in genomic recombination rates
1007 among animal taxa and the case of social insects. *Heredity (Edinb)* **98**: 189–197.

1008 Winkler L, Moiron M, Morrow EH, Janicke T. 2021. Stronger net selection on males across
1009 animals eds. P.J. Wittkopp and L. Holman. *eLife* **10**: e68316.

1010 Winston ML. 1991. *The Biology of the Honey Bee*. Revised edition. Harvard University Press,
1011 Cambridge, Mass.

1012 Wu T-C, Lichten M. 1994. Meiosis-Induced Double-Strand Break Sites Determined by Yeast
1013 Chromatin Structure. *Science* **263**: 515–518.

1014 Yang S, Wang L, Huang J, Zhang X, Yuan Y, Chen J-Q, Hurst LD, Tian D. 2015. Parent–progeny
1015 sequencing indicates higher mutation rates in heterozygotes. *Nature* **523**: 463–467.

1016 Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. 2019. PopLDdecay: a fast and effective tool for
1017 linkage disequilibrium decay analysis based on variant call format files.
1018 *Bioinformatics* **35**: 1786–1788.

1019

1020 **Figure Legends:**
1021

1022 **Figure 1.** Genome-wide decay of LD measured as r^2 versus distance (bp) between SNPs for
1023 the two termite species (*M. bellicosus*, dark red; *C. secundus*, yellow) and honey bees (*A.*
1024 *mellifera scutellata*, blue; data from Wallberg et al. 2017).

1025 **Figure 2.** Variation in recombination rate estimated a ρ/kbp across representative genomic
1026 scaffolds in a) *M. bellicosus* and b) *C. secundus*. Estimates of ρ/kbp are calculated in 10 Kbp
1027 and 100 Kbp windows.

1028 **Figure 3.** Cumulative plot of proportion of recombination events versus proportion of the
1029 genome for the two termite species *M. bellicosus* and *C. secundus* compared to the honey
1030 bee *A. mellifera* (Wallberg et al. 2015). Dashed lines show the proportion of the genome
1031 where 50 % of the recombination occurs, which for *M. bellicosus* is 3.47 %, *C. secundus* is
1032 4.77 %, and *A. mellifera* is 32.05 %.

1033 **Figure 4.** PRMD9 structure and predicted zinc-finger motifs in the *M. bellicosus* and *C.*
1034 *secundus* genomes. A) Predicted structure of the PRDM9 protein sequence in *C. secundus.* B)
1035 Predicted structure of the PRDM9 gene in *M. bellicosus* identified by homology searches.
1036 Exons and conserved functional domains are marked. The mapped exons from
1037 *Zoothermopsis nevadensis* are also shown. C) Structure of PRDM9 gene in *C. secundus*.
1038 Exons and conserved functional domains are marked. D) Predicted zinc finger binding motif
1039 in *M. bellicosus,* E) Predicted zinc finger binding motif in *C. secundus*.

1040 **Figure 5.** Boxplots showing variation in levels of $CpG_{O/E}$ and recombination rate (ρ/kbp; rho)
1041 in genic and flanking regions in the *M. bellicosus* and *C. secundus* genomes. The differences
1042 between all categories are significant.

1043 **Figure 6**. Boxplots showing variation in levels of $CpG_{O/E}$ and recombination rate (ρ/kbp) in
1044 the coding region of genes in the *M. bellicosus* and *C. secundus* genomes classified according
1045 to their patterns of gene expression (Q = queen-biased, K = king-biased, W = worker-biased,
1046 F = female-biased, R = reproductive-biased).

1047

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

Species	Assembly length (Mb)	No. SNPs	θ_W/bp	ρ/kb	μ bp^{-1} gen^{-1}	N_E	r (cM/Mb)
M. bellicosus	1,139	5,581,662	0.14%	1.69	2.7×10^{-10}	1,278,982	0.033
					4.5×10^{-9}	76,876	0.550
					8.1×10^{-9}	42,898	0.986
C. secundus	992	15,037,468	0.44%	4.85	2.7×10^{-10}	4,080,725	0.030
					4.5×10^{-9}	245,280	0.494
					8.1×10^{-9}	136,869	0.885

θ_W/bp - Watterson's θ per base pair

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

μ bp^{-1} gen^{-1} - mutation rate per base pair per generation taken from literature. Estimates correspond to *Acyrthosiphon pisum* (2.7×10^{-10}), *Drosophila melanogaster* (4.5×10^{-9}) and *Drosophila pseudoobscura* (8.1×10^{-9})(Lynch et al. 2023).

N_E - effective population size

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

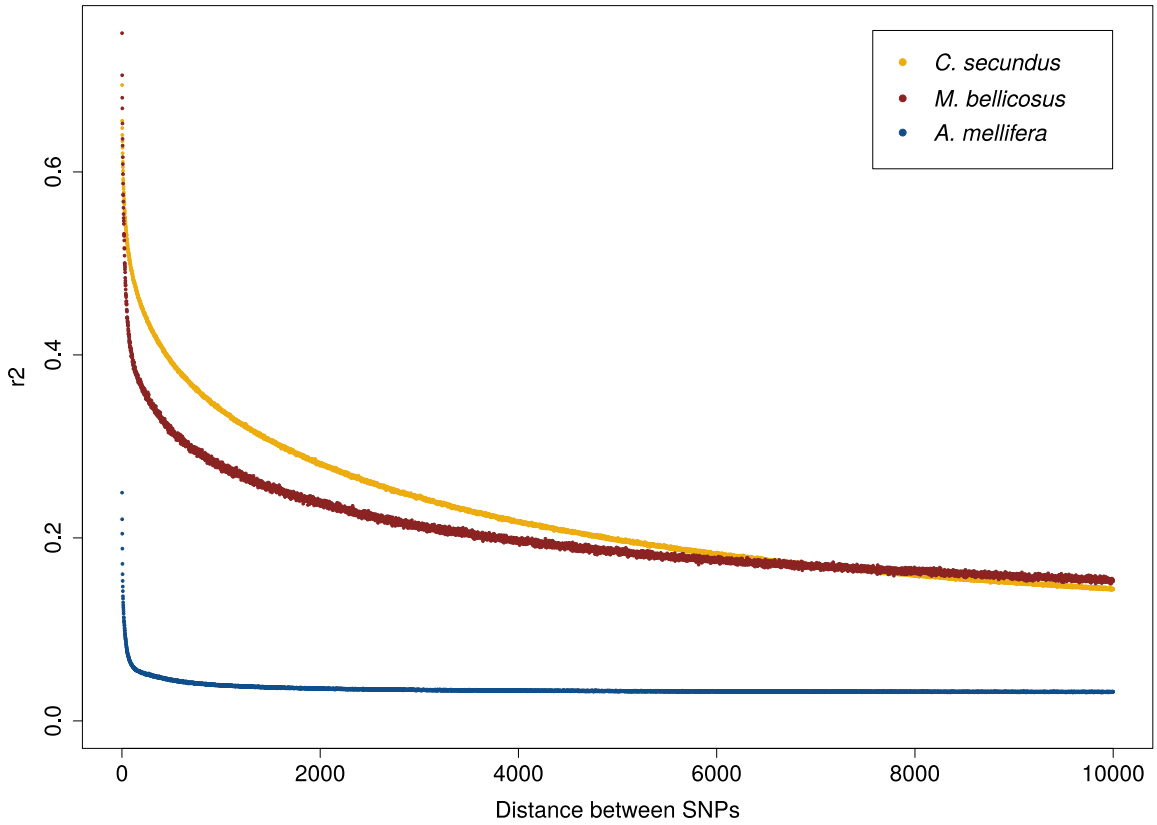
C. secundus
M. bellicosus
A. mellifera
r2
0.0
0.2
0.4
0.6
0
2000
4000
6000
8000
10000
Distance between SNPs

Figure 1

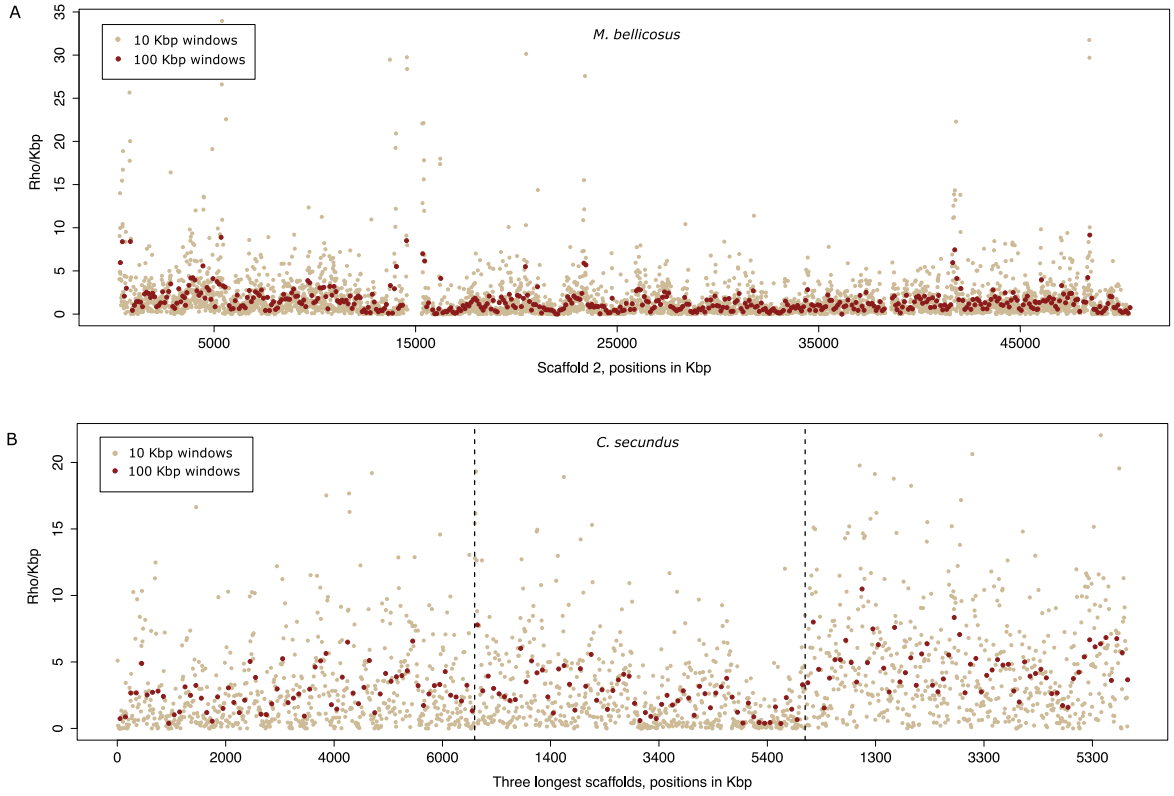
A
10 Kbp windows
100 Kbp windows
M. bellicosus
Rho/Kbp
0
5
10
15
20
25
30
35
5000
15000
25000
35000
45000
Scaffold 2, positions in Kbp
B
10 Kbp windows
100 Kbp windows
C. secundus
Rho/Kbp
0
5
10
15
20
0
2000
4000
6000
1400
3400
5400
1300
3300
5300
Three longest scaffolds, positions in Kbp

Figure 2

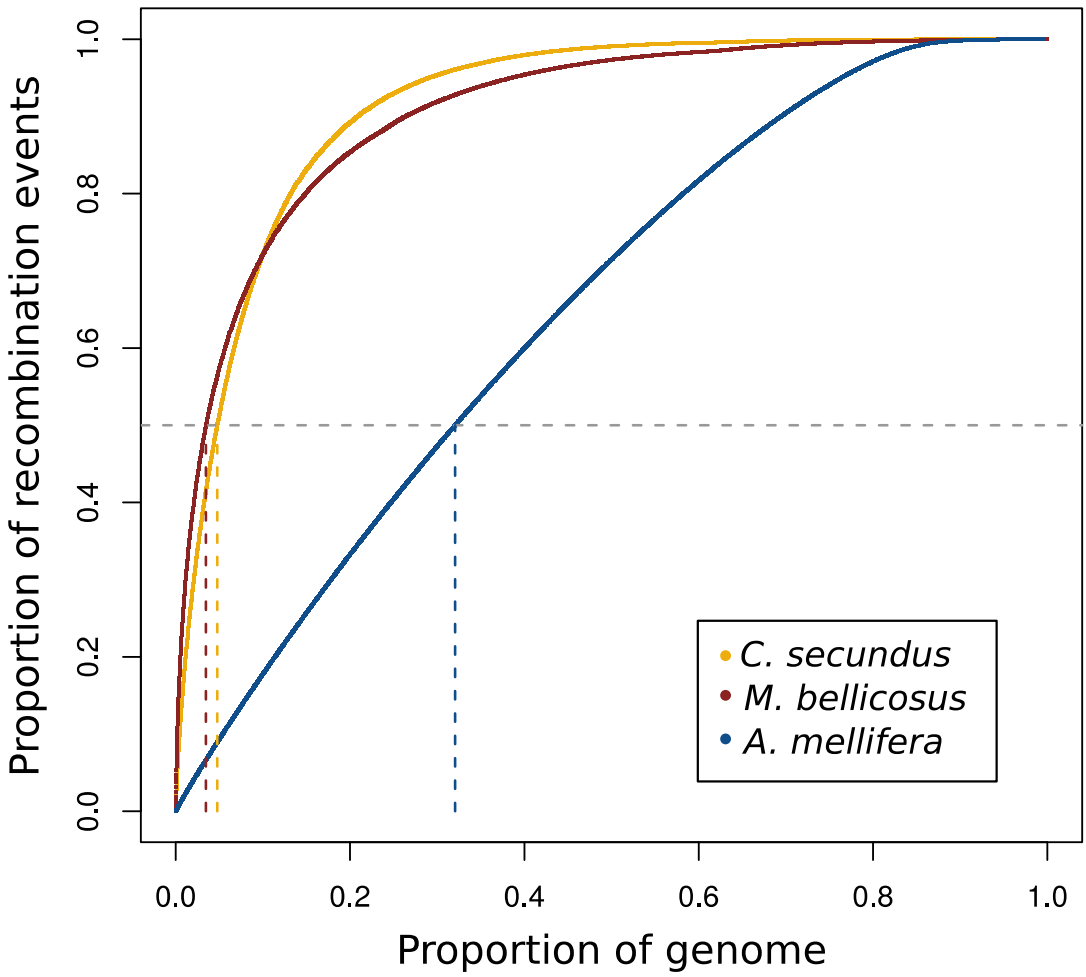
Proportion of recombination events
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of genome
0.0
0.2
0.4
0.6
0.8
1.0
C. secundus
M. bellicosus
A. mellifera

Figure 3

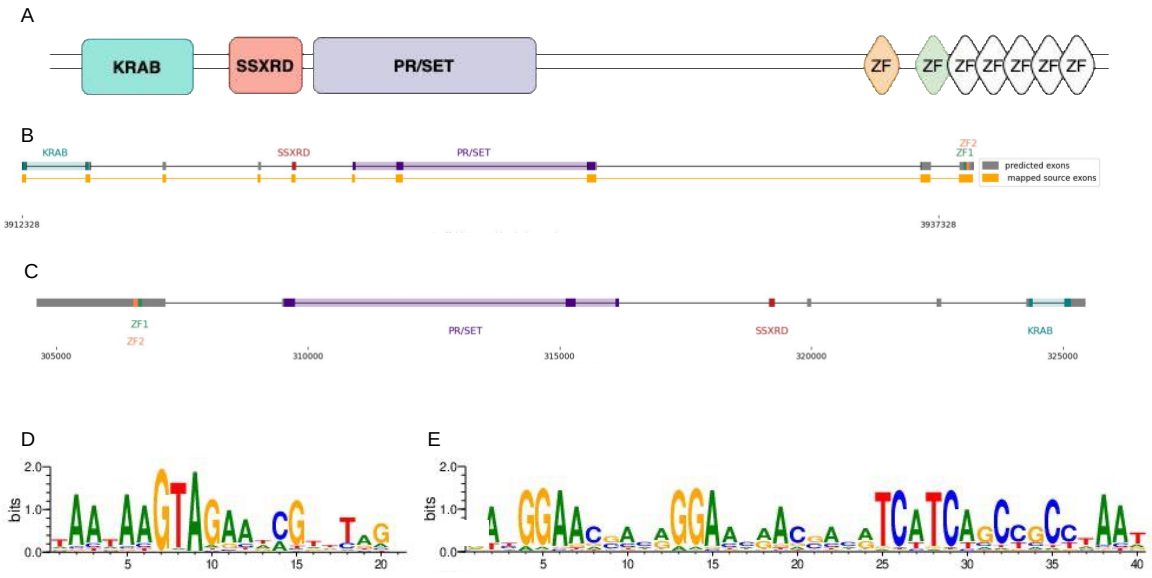
A
KRAB
SSXRD
PR/SET
ZF
ZF
ZF
ZF
ZF
ZF
ZF
B
KRAB
SSXRD
PR/SET
ZF2
ZF1
predicted exons
mapped source exons
3912328
3937328
C
ZF1
ZF2
PR/SET
SSXRD
KRAB
305000
310000
315000
320000
325000
D
bits
2.0
1.0
0.0
5
10
15
20
E
bits
2.0
1.0
0.0
5
10
15
20
25
30
35
40

Figure 4

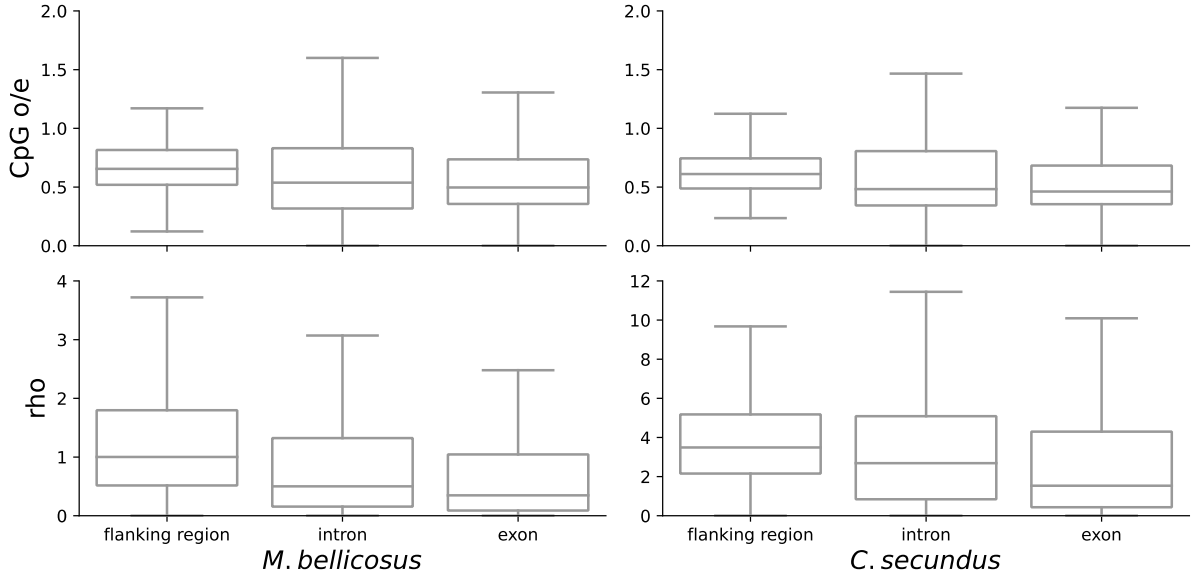
CpG o/e
rho
2.0
1.5
1.0
0.5
0.0
4
3
2
1
0
12
10
8
6
4
2
0
flanking region
intron
exon
M. bellicosus
C. secundus

Figure 5

CpG o/e
rho
Q
K
W
M
F
R
unbiased
M. bellicosus
Q
unbiased
W
C. secundus

Figure 6