Received: 12 June 2024 | Accepted: 2 September 2024

DOI: 10.1111/imb.12961

RESEARCH ARTICLE

Insect Molecular Biology

Royal Entomological Society

Alternative double strand break repair pathways shape the evolution of high recombination in the honey bee, *Apis mellifera*

Bertrand Fouks[1,2,3] | **Katelyn J. Miller**[1,4] | **Caitlin Ross**[5,6] | **Corbin Jones**[7] | **Olav Rueppell**[1,8]

[1]Department of Biology, University of North Carolina at Greensboro, Greensboro, North Carolina, USA

[2]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

[3]CIRAD, UMR AGAP Institut, Montpellier, France

[4]Smithers PDS, Gaithersburg, Maryland, USA

[5]Department of Computer Sciences, University of North Carolina at Greensboro, Greensboro, North Carolina, USA

[6]Kitware, Minneapolis, Minnesota, USA

[7]Department of Biology, University of North Carolina at Chapel Hill & Carolina Center for Genome Sciences, Chapel Hill, North Carolina, USA

[8]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

Correspondence
Olav Rueppell, Department of Biology, University of North Carolina at Greensboro, Greensboro, NC 27412, USA.
Email: olav@ualberta.ca

Funding information
US Department of Defence, Grant/Award Numbers: RGPIN-2022-03629, W911NF1520045, W911NF2210195; National Institute of General Medical Sciences; Natural Sciences and Engineering Research Council of Canada; US National Institutes of Health, Grant/Award Number: R15GM102753

Associate Editor: Emily Remnant

Abstract

Social insects, particularly honey bees, have exceptionally high genomic frequencies of genetic recombination. This phenomenon and underlying mechanisms are poorly understood. To characterise the patterns of crossovers and gene conversion in the honey bee genome, a recombination map of 187 honey bee brothers was generated by whole-genome resequencing. Recombination events were heterogeneously distributed without many true hotspots. The tract lengths between phase shifts were bimodally distributed, indicating distinct crossover and gene conversion events. While crossovers predominantly occurred in G/C-rich regions and seemed to cause G/C enrichment, the gene conversions were found predominantly in A/T-rich regions. The nucleotide composition of sequences involved in gene conversions that were associated with or distant from crossovers corresponded to the differences between crossovers and gene conversions. These combined results suggest two types of DNA double-strand break repair during honey bee meiosis: non-canonical homologous recombination, leading to gene conversion and A/T enrichment of the genome, and the canonical homologous recombination based on completed double Holliday Junctions, which can result in gene conversion or crossover and is associated with G/C bias. This G/C bias may be selected for to balance the A/T-rich base composition of eusocial hymenopteran genomes. The lack of evidence for a preference of the canonical homologous recombination for double-strand break repair suggests that the high genomic recombination rate of honey bees is mainly the consequence of a high rate of double-strand breaks, which could in turn result from the life history of honey bees and their A/T-rich genome.

KEYWORDS
crossover, gene conversion, genetic mapping, genome evolution, recombination rate

INTRODUCTION

Sexual reproduction is widespread in eukaryotes despite its fitness costs. The main explanation for the ubiquity of sex is its benefits of increasing genetic variation via independent chromosome segregation and recombination (Maynard-Smith, 1978; Otto & Lenormand, 2002). Thus, sexual reproduction facilitates long-term genome evolution, specifically preventing the accumulation of deleterious mutations (Muller's Ratchet) and combining novel, potentially beneficial mutations at independent loci (Agrawal, 2006; Gandon & Otto, 2007; Hill & Robertson, 1966; Roze, 2012). Crossover formation is required in most species to achieve successful meiosis (John, 2005), and a lack or reduction of crossovers can lead to severe disease (Baker et al., 1976; Shi et al., 2001). However, elevated rates of crossover formation have also been associated with human disease (Louis & Borts, 2003; Purandare & Patel, 1997) and increase the risk of disrupting adaptive allele combinations (Agrawal, 2006). Thus, it is thought that antagonistic selection pressures should select for an optimal frequency of crossovers (Dumont & Payseur, 2008; Louis & Borts, 2003). Nevertheless, genome-wide recombination rate fluctuates greatly among species, individuals and sexes (Stapley et al., 2017), suggesting that an increase or decrease of genome-wide recombination rates can result from adaptive evolution. This argument is supported by genes involved in the recombination pathway demonstrating signatures of positive selection in mammals (reviewed in (Baudat et al., 2013)), *Drosophila* (Hunter et al., 2016) and honey bees (Fouks et al., 2021) and by signatures of adaptive evolution underpinning recombination rate differences between *Drosophila* populations (Samuk et al., 2020). Despite variation in genome-wide recombination rate among species, recombination events are often restricted to one per chromosome or per chromosome arm (Dumont & Payseur, 2008).

Social insects in the order Hymenoptera are an important exception, exhibiting higher levels of recombination than solitary species (DeLory et al., 2024; Sirviö et al., 2006; Waiker et al., 2021; Wilfert et al., 2007). The recombination rate of the Western Honey Bee (*Apis mellifera*) is particularly high, with most estimates close to 20 cM/Mbp (Beye et al., 2006; Hunt & Page, 1995; Ross et al., 2015; Solignac et al., 2007). Even before the first empirical evidence, it was suggested that recombination would benefit the evolution of eusociality (Sherman, 1979), and multiple other explanations have been put forward to explain the evolution of the exceptionally high level of recombination in the social Hymenoptera. The main arguments can be summarised into two independent hypotheses: the "genotypic diversity hypothesis" and the "social innovation hypothesis". The first hypothesis postulates that high recombination rates have evolved and are maintained as a means to increase within-colony genotypic diversity (DeLory et al., 2024; Sirviö et al., 2006), improving division of labour (Mattila & Seeley, 2007) and disease resistance (Seeley & Tarpy, 2007). However, the theoretical appeal of these arguments is diminished by the finding that within-colony genotypic variance is not quantitatively increased by elevated recombination rates in the parental generation (Rueppell et al., 2012). The second hypothesis postulates that high recombination rates of social insects have evolved to facilitate evolutionary innovations, such as the evolution of the worker caste. Although this argument can also be related to division of labour and disease resistance, it is distinct from the first hypothesis because it relies on a multi-generational argument and thus relates conceptually to the general explanation for the evolution of sex and recombination in terms of long-term benefits (Otto & Lenormand, 2002). According to this "social innovation hypothesis", recombination rates in social insects are elevated because social insects have longer generation times, smaller population sizes and stronger selection for evolutionary innovation and decoupling of genes than solitary species (Kent & Zayed, 2013). The associations between local recombination rates and worker biased gene expression support this model (Kent et al., 2012; Liu et al., 2015). However, the causality of these associations is unclear and cannot account for the overall increase of genome-wide recombination rates. Moreover, very little is known about the actual mechanisms that can lead to the high recombination rates (Wallberg et al., 2015). A deeper mechanistic understanding of the extraordinarily high recombination rate of honey bees is needed to evaluate these hypotheses and lead to a better understanding of the evolution of recombination rates in general.

Recombination ensues from the repair of a DNA double-strand break during meiosis (Pâques & Haber, 1999; Szostak et al., 1983). Different mechanisms of double-strand break repair have been postulated based on mutational screens and consensus between different studies and model species has been incorporated into a general model of recombination (Kohl & Sekelsky, 2013). However, it is unclear how widely applicable the specific pathways of this model are because many taxa deviate from the consensus model (e.g., (Voelkel-Meiman et al., 2015)). The initial choice of double-strand break repair pathway seems to depend largely on the length of homology between DNA strands (Piazza & Heyer, 2019). The double-strand break is repaired through non-homologous end joining when there is no or very short (0–4 base pair) homology between DNA strands, alternative end joining with microhomology (2–20 bp), single strand annealing with >50 bp homology and homologous recombination with >100 bp homology.

While all DNA repair pathways can lead to some genetic exchange, only homologous recombination leads to a large-scale swap of DNA between chromatids. After end resection of each DNA strand, the initial double-strand break is processed into an extended displacement loop (D-loop) during homologous recombination, resulting from successful DNA strand-invasion guided by RAD51 helicase and likely mediated by BRCA1-BARD1 (Zhang et al., 2019). In the presence of anti-crossover helicase, the D-loop is unwound to produce a non-crossover. This double-strand break repair is achieved via synthesis-dependent strand annealing (McMahill et al., 2007), resulting in the transfer of a relatively short segment of DNA and is a form of gene conversion (Figure 1a). Gene conversion appears to be more mutagenic than canonical homologous recombination mechanisms (Rodgers & McVey, 2016). In addition, the repair can be overtaken by other mechanisms under different circumstances, such as break-induced repairs or multi-invasion-induced rearrangements (reviewed in (Piazza & Heyer, 2019)). In the absence of anti-crossover helicases,

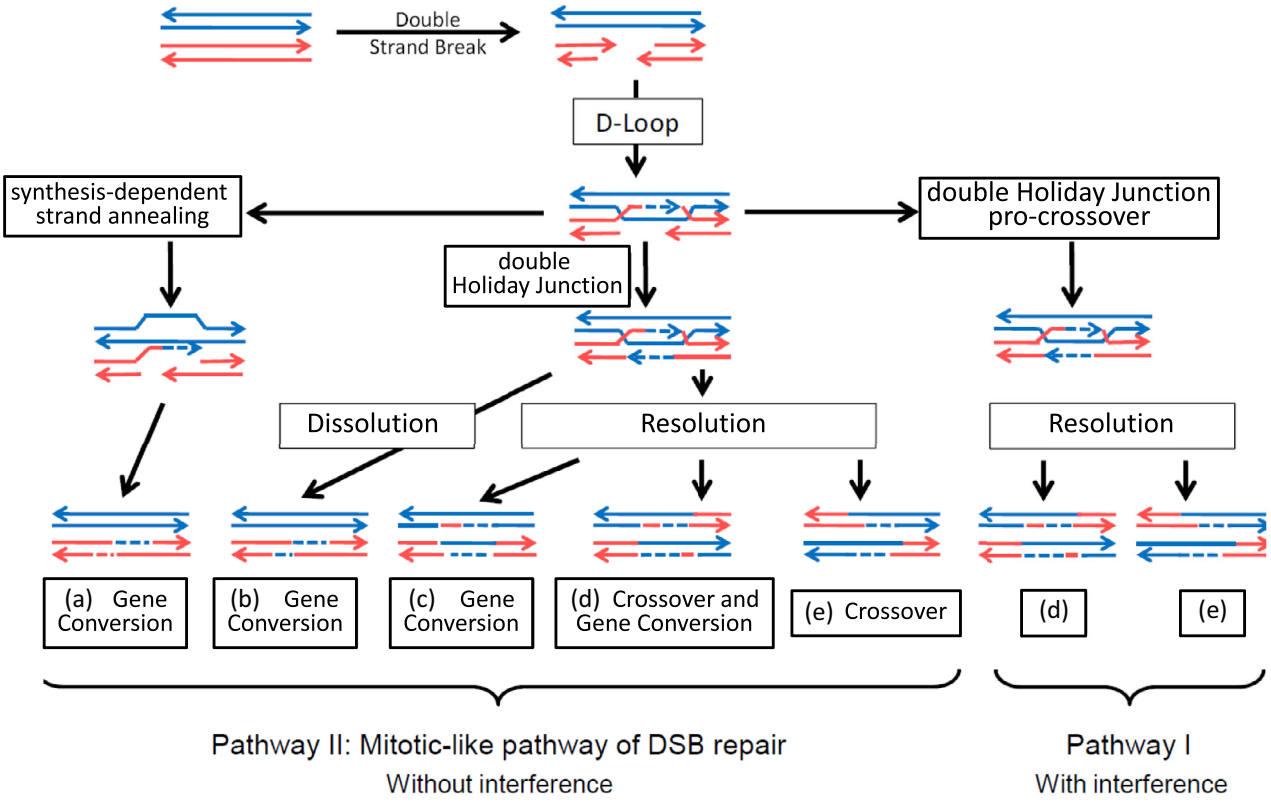

FIGURE 1 Mechanisms of homologous recombination are represented in a schematic view along with their respective outcomes, based on a general model (Kohl & Sekelsky, 2013). An initial double-strand break is processed into a displacement loop (D-loop). In the presence of anti-crossover helicase, the D-loop is processed by strand displacement strand annealing resulting in a gene conversion. When anti-crossover helicase is absent, the D-loop is processed into a double Holliday junction. The double Holliday junction can be resolved via resolvase, which leads to either a gene conversion, a crossover or a gene conversion that is linked to a crossover event. Instead of being resolved, it can be dissolved, which invariably causes a gene conversion. These mechanisms are part of the pathway II of meiotic recombination, which is similar to mitotic recombination processes. As part of the pathway I of meiotic recombination that is specific to meiosis, the D-loop is processed into a double Holliday junction, which will be resolved by specific pro-crossover resolvases, leading invariably to crossovers.

canonical homologous recombination occurs with the D-loop being proceed to a double Holliday junction (Pâques & Haber, 1999) via reciprocal strand exchange. In the presence of dissolvases, the double Holliday junction is dissolved, which leads to a simple gene conversion event (Figure 1b). In the presence of resolvases, the double Holliday junction can be resolved in three different ways: a simple gene conversion (Figure 1c), a crossover associated with a gene conversion (Figure 1d; (Allers & Lichten, 2001)) or a simple crossover, leading to recombination of the flanking markers (Figure 1e). In contrast to synthesis-dependent strand annealing, such double Holliday junction repair is biased towards incorporating G/C nucleotides (Lesecque et al., 2013; Marais, 2003). This G/C-biased gene conversion is thus typically associated with Crossovers (Lesecque et al., 2013). These mechanisms bear similarities to double-strand break repair in mitosis and presumably result in non-interfering crossovers (Figure 1; (Kohl & Sekelsky, 2013)). In the presence of an anti-anti-crossover (=pro-crossover) resolvase, a meiosis-specific repair pathway of double-strand break (which is likely induced during meiosis by Spo11 (reviewed in (Lam & Keeney, 2015)) or homologues (McKim & Hayashi-Hagihara, 1998) may occur that exclusively proceeds from the double Holliday junction to crossovers (Figure 1e), leading to interfering crossovers (Figure 1; (Kohl & Sekelsky, 2013)).

Despite a broad functional conservation of double-strand break repair pathways (van de Bosch et al., 2002), the regulation of this repair is variable among species (Kohl & Sekelsky, 2013). Preference of crossover-inducing pathways over alternative repair pathways may provide mechanistic explanations for the strong up-regulation of genomic recombination in honey bees. However, it is not understood which of the alternative mechanistic pathways are employed or even how gene conversions and crossovers relate to each other (Bessoltane et al., 2012; Liu et al., 2015; Wallberg et al., 2015). Mutational studies of recombination pathways are problematic in honey bees (Schulte et al., 2014). Instead, analyses of the variety and frequencies of different meiotic products have been performed, which is particularly powerful in honey bees due to their high rates of meiotic recombination and haplo-diploidy. Such analyses (Bessoltane et al., 2012; Liu et al., 2015; Mougel et al., 2014; Ross, DeFelice, Hunt, Ihle, & Amdam, 2015) have yielded some insights into DNA sequence features that are associated with recombination rate in honey bees. Two classes of DNA motifs have been associated with recombination:

A/T-rich and G/C-rich motifs: While G/C-rich motifs are positively associated with recombination rate, A/T-rich motifs show negative correlations (Bessoltane et al., 2012). Two specific G/C-rich motifs (CGCA and GCCGC) have been found in high recombination regions of the honey bee genome (Bessoltane et al., 2012; Mougel et al., 2014). However, the correlations between G/C-rich motifs and recombination rates may be due to a positive effect of G/C content on recombination or the enrichment of G/C bases in regions of high recombination via G/C-biased gene conversion (Ross, DeFelice, Hunt, Ihle, & Amdam, 2015).

Here, a large data set is presented and analysed to characterise patterns of gene conversions and crossovers throughout the entire honey bee genome and identify which genomic features are associated with different outcomes of double-strand break repair during honey bee meiosis. The data support a synthesis-dependent strand annealing (non-canonical homologous recombination) repair mechanism during meiosis that seems to locally enrich the A/T content of the DNA. The data further show that double Holliday junction repair (canonical homologous recombination) leads to enrichment of G/C bases, consistent with the G/C-biased gene conversion hypothesis. Together these findings may help explain the paradox of the A/T-rich honey bee genome that exhibits an exceptionally high rate of recombination, which in turn is associated with G/C-biased gene conversion.

RESULTS

Genome-wide recombination patterns

The location of meiotic recombination events was estimated from complete genome resequencing data of 187 male offspring of a single queen. Since the haploid sons represent meiotic products of their mother, phase shifts in the 931,350 segregating single nucleotide polymorphisms (SNPs) indicated directly the location of genetic recombination between homologous chromosomes. The frequency distribution of tract lengths between phase shifts (Figure 2) was bimodally distributed (Hartigan's dip test for unimodality: $D = 0.036$, $n = 166{,}849$, $p < 10^{-15}$).

When initially separating gene conversions and crossovers at a 10 kbp tract length threshold (following (Liu et al., 2015)), 56,268 gene conversions (300.9 gene conversions per drone) with a median tract length of 77 bp and 9041 crossovers (48.3 crossovers per drone) with a median of tract length of 215,381 bp were identified. To separate gene conversions and crossovers more conservatively for subsequent analyses, only a part of the left distribution was defined as gene conversion events (indicated by yellow in Figure 2), ambiguous tracts of intermediate length were excluded from further analyses (coloured green) and the tracts belonging to the right tail of the distribution (blue colour) were defined as crossover events. With these stringent criteria, 7384 crossovers and 8285 gene conversions were identified, resulting in respective averages of 39.5 crossovers and 44.3 gene conversions per drone. These records included 32 cases of "crossover-associated gene conversion" events and 910 "non-crossover gene conversion" events. Crossover number ranged in individual drones from 21 to 56 (coefficient of variation = 0.17) and gene conversion number from 0 to 358 (CV = 1.16).

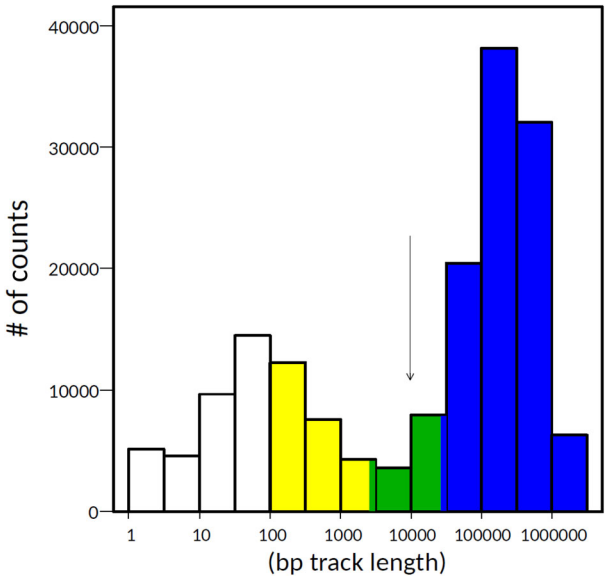

FIGURE 2 Distribution of phase tract lengths. The distribution of the frequency of log-transformed tract lengths showed a bimodal distribution (Hartigan's dip test: $p < 10^{-15}$), revealing a distinction between shorter tracts generated by gene conversion and longer tracts that indicate crossovers. The arrow indicates the 10-kbp threshold that has been used previously to distinguish between gene conversion and crossover events (Liu et al., 2015). However, both underlying distributions had a wide range and were partially overlapping, which complicates unambiguous categorization of a particular phase change as gene conversion or crossover. Consequently, intermediate data points (green) were omitted and only tracts of >25,000 bps were considered as indicating crossovers (blue) and tracts between 100 and 2500 bps were analysed as gene conversion (yellow).

The weighted average of the crossover rates of all 16 chromosomes resulted in genome-wide recombination rate estimates of 18.0 cM/Mbp (Table 1), while the initial, more liberal crossover count would increase this estimate to 21.9 cM/Mbp. Regression of the recombination rate on the length of each chromosome indicated a negative but not significant relationship ($r^2 = 0.24$, $F_{(1,14)} = 4.5$, $p = 0.053$). Crossover events were heterogeneously distributed across the genome (Exact Poisson test: $p < 10^{-15}$). The distributions of crossover and gene conversion counts per individual reveal that crossover counts follow a normal distribution indicating they results from controlled process, while gene conversion counts follow a Poisson distribution (Figure 3).

A gamma distribution fit to our inter-crossover distances (Figure 4) indicated an overall crossover interference parameter of 2.16 (Broman & Weber, 2000). Some of the chromosomes were experiencing a non-negligible proportion of non-interfering crossovers (Table 1 and Figure S1; (Housworth & Stahl, 2003)). In the entire genome only 18 hotspots were detected according to our definition (see Methods). In contrast, 531 windows of 100 kbp length were found without crossover and considered crossover deserts and 422 windows were considered gene conversion deserts because they

TABLE 1 Summary of overall chromosomal recombination statistics and evidence for crossover interference following (Housworth & Stahl, 2003).

Chromosome	Crossover rate	Number of gene conversions	Genetic length (cM)	Physical length (Mbp)	Interference index	Proportion of non-interfering crossovers
LG01	17.78	1146	531.55	29.89	2.1	0
LG02	18.74	569	291.44	15.55	3.34	0.15
LG03	19.85	539	262.57	13.23	5.05	<0.001
LG04	18.92	539	240.64	12.72	4.77	<0.001
LG05	17.13	589	246.00	14.36	2.44	0.01
LG06	17.11	747	316.04	18.47	2.44	0.05
LG07	17.64	602	233.16	13.22	6.15	0.24
LG08	15.75	416	213.37	13.55	3.12	0
LG09	18.42	441	204.81	11.12	6.76	0.01
LG10	16.99	364	220.32	12.97	2.26	0
LG11	16.99	463	250.27	14.73	3.72	0
LG12	19.46	391	231.55	11.9	8.03	0.07
LG13	18.03	478	185.56	10.29	4.83	0.16
LG14	19.04	395	195.19	10.25	5.85	0.13
LG15	18.93	285	192.51	10.17	4.17	0
LG16	18.62	322	134.22	7.21	3.07	0.009
Total	17.98	8286	3949.20	219.63	2.16	0

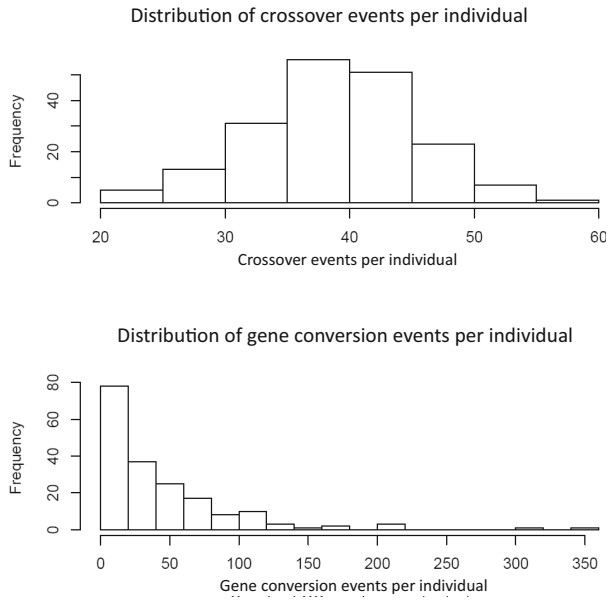

FIGURE 3 Comparison of the distribution of crossover and gene conversion counts per individual. While crossover counts are normally distributed, gene conversion counts are more consistent with a Poisson distribution, indicating regulated and random processes, respectively.

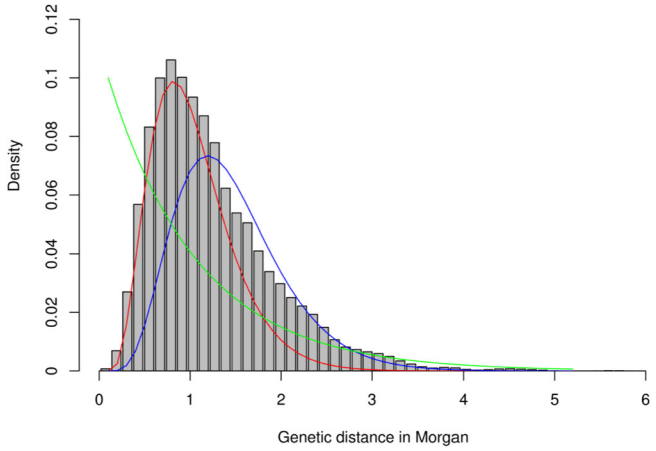

FIGURE 4 Observed distribution of inter-crossover distances compared with the theoretical gamma distributions with and without interference. The red line represents the gamma function fitted to the observed distribution through a maximum likelihood approach. The blue line is the gamma (v, 2v) fitted to the observed data. The green line represents the gamma function without interference ($m = 1$) fitted to the observed distribution.

were devoid of gene conversion events. Marey maps (Figure 5) showed an approximately linear relationship between physical and genetic distances. The most profound deviations from collinearity of physical and genetic maps were found at the sites of the acrocentric centromeres.

Crossovers and gene conversions in genome regions with segregating genetic variation

The combined genomic regions for pollen hoarding, worker ovary size, age of first foraging Quantitative Trait Loci (=QTL) and the Complementary Sex Determination region, that were selected due to their important segregating genetic variation, exhibited significantly increased rates of crossover events compared with the rest of the

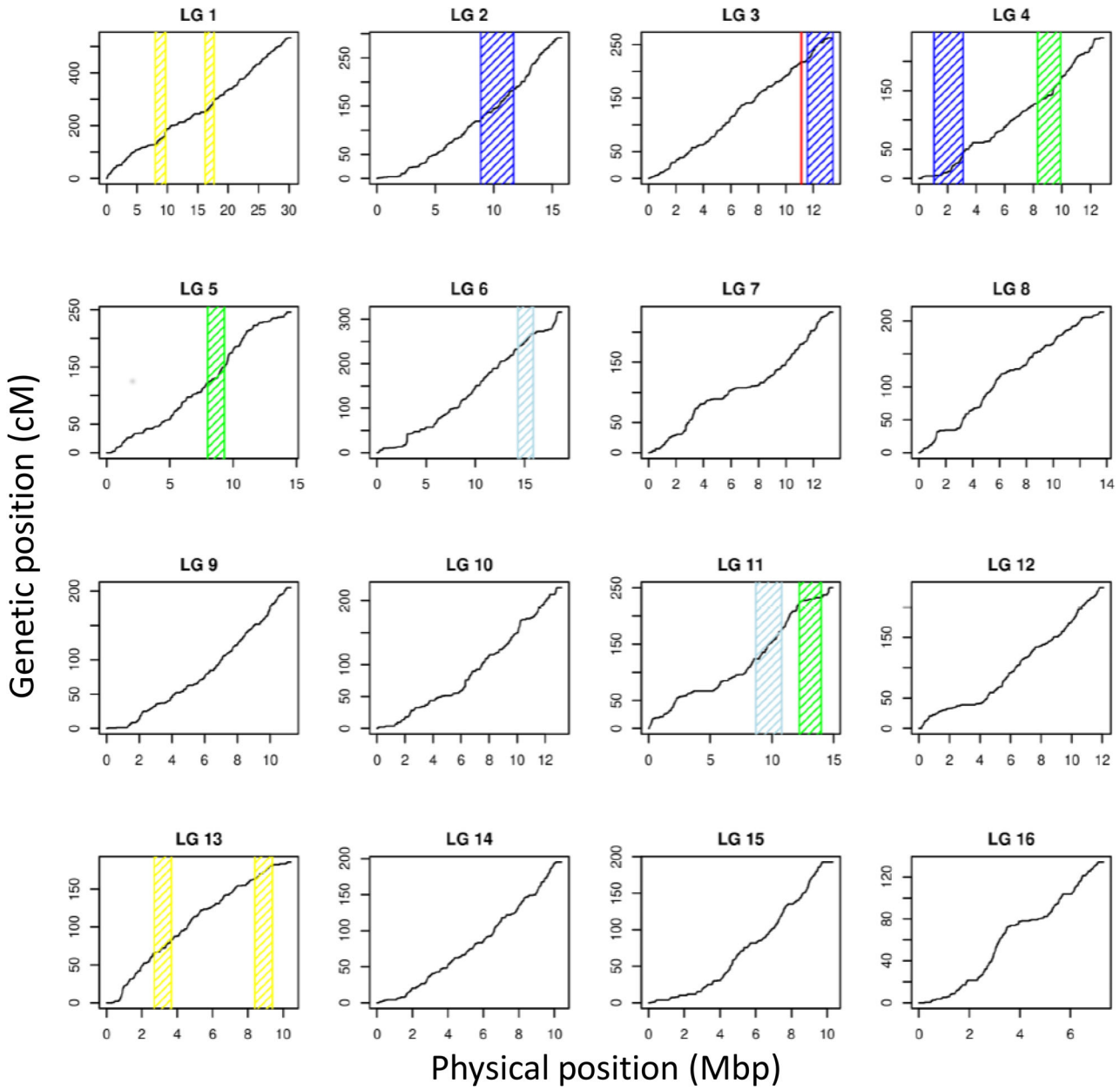

FIGURE 5 Marey maps of genetic map length in centi-Morgans over the physical map length for each honey bee chromosome. Across most of the chromosomes, the relationship between genetic and physical distances is close to linear. However, a reduction of recombination is apparent at the beginning of most chromosomes, corresponding to the acrocentric honey bee chromosomes (#2–16). Indicated in colour are the regions with important segregating genetic variation that exhibit higher than average gene conversion and crossover frequencies (red = sex determination region, yellow = pollen hoarding QTL, green = Age of First Foraging QTL, blue = Worker Ovary Size QTL as determined in an interpopulation cross: light blue and in an intrapopulation cross: dark blue).

genome (glm: $p = 0.028$). Simultaneously, gene density ($p = 0.029$) and the interaction between both variables ($p = 0.042$) influenced the rate of crossovers. Specifically, crossover events were increased by gene density in the general genome, whereas this relation was inverted in the selected genome regions. Similarly, the rate of gene conversion events was significantly affected by genome region ($p = 0.002$), gene density ($p = 0.012$) and their interaction ($p = 0.012$).

Local rates of gene conversions and crossovers

Local crossover rates of each chromosome were analysed in more detail, using a 100-kbp window sliding at 10 kbp intervals across the genome (chromosome 1 in Figure 6; all other chromosomes in Figure S2). Local crossover rates at this scale were variable, and 221 crossover peaks were higher than the empirically determined 1% Poisson threshold. These crossover peaks were distributed evenly

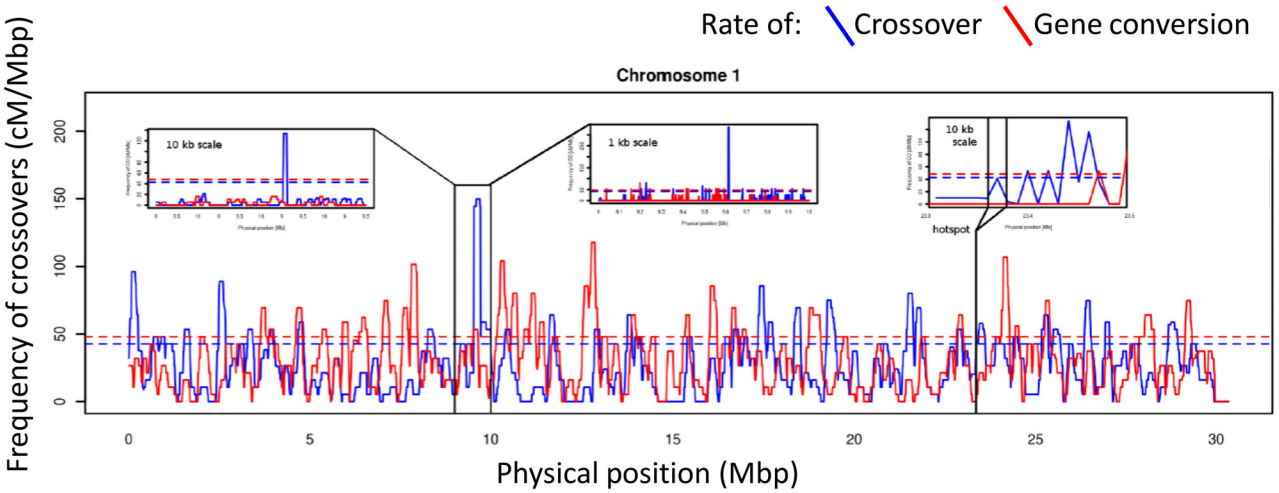

FIGURE 6 Distribution of crossover and gene conversion rate along the first chromosome. Rates averaged across 100-kbp window sliding every 10 kbp were plotted. The horizontal, dashed threshold lines represent the Poisson estimate that identifies outliers ($p < 0.01$). Most of the crossover and gene conversion rate peaks are above the respective thresholds, which illustrates the spatial heterogeneity of the distribution of these rates. However, the peaks appear throughout the chromosome and are not clustered into a small part of the genome, which contrasts with recombination rate distributions in previously constructed fine-scale recombination maps in other species.

TABLE 2 Best-fit multiple regression models of the influence of sequence characteristics on local crossover and gene conversion rates.

Crossover model	B (±S.E.)	Significance
$(G/C)^2$	0.010 ± 0.001	$t = 11.4$, $p < 0.001$
CpG × Low complexity	−61.6 ± 12.4	$t = -4.97$, $p < 0.001$
G/C × CpG × Low complexity	1.95 ± 0.40	$t = 4.82$, $p < 0.001$
Overall Model	$r^2 = 0.13$, $F_{(3,2193)} = 110.8$, $p < 0.001$	
Gene Conversion model	**B (±S.E.)**	**Significance**
CpG	4.73 ± 0.88	$t = 5.38$, $p < 0.001$
Low complexity	4.62 ± 1.20	$t = 3.84$, P < 0.001
$(CpG)^2$	−0.54 ± 0.10	$t = -5.30$, $p < 0.001$
$(Low\ complexity)^2$	−0.56 ± 0.16	$t = -3.39$, $p < 0.001$
CpG × Low compexity	−5.63 ± 2.67	$t = -2.11$, $p < 0.05$
Overall Model	$r^2 = 0.01$, $F_{(5,2191)} = 6.46$, $p < 0.001$	

along each chromosome and no small genome region experienced a pronounced increase of crossover rates compared with the rest of the genome. At the 100 kbp windows scale, gene conversion rates were significantly correlated to crossover rates (rho = 0.22, $n = 2184$, $p < 10^{15}$).

Despite our saturated linkage map with a median interval between markers of only 314 bp, the numbers of detected crossovers and gene conversions were correlated with markers density at the 100-kbp scale (rho = 0.06, $n = 2197$, $p < 10^{-15}$ and rho = 0.61, $n = 2197$, $p < 10^{-15}$, respectively; Figure S3). Independent of marker density, local crossover and gene conversion rates were related to CpG frequency and abundance of low complexity sequences. Local crossover rate was also related to G/C content (Table 2). Despite some interactions among predictor variables, the best-fitting general models of these relations remained relatively simple and bivariate relations could be reasonably approximated by linear functions (Figure S4).

Sequence features associated with crossover and gene conversion

At a smaller scale, the target sequences for gene conversion and crossover events were contrasted with the adjacent genome sequences. The average G/C content of crossover target sequences was 41.0 ± 0.2%, (mean ± 95% CI) and thus significantly higher ($t = 14.7$, df = 12,333, $p < 10^{15}$) than the corresponding flanking regions that contained 39.2 ± 0.1% G/C bases (Figure S5). In contrast, gene conversion target sequences showed with 32.9 ± 0.2% a significantly lower G/C content ($t = 11.5$, df = 15,528, $p < 10^{-15}$) than the corresponding flanking regions (34.4 ± 0.1%). Furthermore, G/C bases were significantly more frequent in "crossover-associated gene conversion" sequences (39.6 ± 2.5%) than in "non-crossover gene conversion" sequences (27.4 ± 0.6%; $t = 9.24$, df = 34.0, $p < 10^{-10}$). In addition, the percentage of G/C was lower in the flanking sequences of gene conversions (34.4 ± 0.1%) and "non-crossover gene conversion" events (28.6 ± 0.4%) than in the flanking regions of crossovers (39.2 ± 0.1%) and "crossover-associated gene conversion" events (38.5 ± 2.0%), respectively (gene conversion vs. crossover: $t = 52.5$, df = 31,214, $p < 10^{-15}$; "crossover-associated gene conversion" vs. "non-crossover gene conversion": $t = 9.54$, df = 67.7, $p < 10^{-13}$).

Gene conversion and crossover events also differed in their physical relation to genic features. On average, crossovers were located significantly more distant from the nearest exon, 5′- and 3′-

TABLE 3 Comparisons between crossover, gene conversion and corresponding random control sequences with respect to their distance to nearest gene, exon and UTRs.

Genome feature	Control for crossover (mean ± S.D.)	Difference crossover versus control	Crossover (mean ± S.D.)	Difference crossover versus gene conversion	Gene conversion (mean ± S.D.)	Difference gene conversion versus control	Control for gene conversion (mean ± S.D.)
Gene	8.8 ± 0.57 kbp	$t = 0.36$, df = 14,023, $p = 0.720$	8.9 ± 0.46 kbp	$t = 1.68$, df = 15,158, $p = 0.092$	8.4 ± 0.41 kbp	$t = -1.25$, df = 14,792, $p = 0.211$	8.8 ± 0.56 kbp
Exon	11.3 ± 0.57 kbp	$t = 3.56$, df = 14,116, $p < 0.001$	12.7 ± 0.48 kbp	$t = 2.22$, df = 15,189, $p = 0.026$	11.9 ± 0.43 kbp	$t = 1.1$, df = 14,983, $p = 0.256$	11.5 ± 0.57 kbp
5′ UTR	58.2 ± 2.0 kbp	$t = 5.82$, df = 13,813, $p < 10^{-8}$	66.4 ± 1.9 kbp	$t = -2.13$, df = 15,186, $p = 0.033$	69.5 ± 2.0 kbp	$t = 9.4$, df = 15,332, $p < 10^{-15}$	56.2 ± 1.9 kbp
3′ UTR	47.8 ± 1.7 kbp	$t = 5.71$, df = 13,959, $p < 10^{-7}$	54.7 ± 1.6 kbp	$t = -0.48$, df = 15,317, $p = 0.630$	55.2 ± 1.6 kbp	$t = 9.6$, df = 15,630, $p < 10^{-15}$	44.7 ± 1.5 kbp

	Crossover-associated gene conversion (mean ± S.D.)	Non-crossover gene conversion (mean ± S.D.)	Difference
Gene	6.5 ± 4.0 kbp	6.2 ± 1.0 kbp	$t = 0.14$, df = 35.5, $p = 0.887$
Exon	9.7 ± 4.3 kbp	8.0 ± 1.1 kbp	$t = 0.72$, df = 35.1, $p = 0.476$
5′ UTR	79.2 ± 29.4 kbp	50.5 ± 5.5 kbp	$t = 1.9$, df = 33.2, $p = 0.069$
3′ UTR	65.1 ± 29.0 kbp	38.0 ± 4.3 kbp	$t = 1.8$, df = 32.4, $p = 0.080$

untranslated region (=UTR) than length-matched random control sequences (Table 3). The average distance between a gene conversion event and the nearest 5′UTR, and 3′UTR was also higher than corresponding control sequences. Gene conversion events were significantly closer to exons and farther from 5′ UTRs than crossover events (Table 3). Furthermore, "non-crossover gene conversions" were on average closer to the nearest gene, exon, 5′- and 3′UTR than "crossover-associated gene conversion" events, but these differences were not significant (Table 3).

Motif search for crossover and gene conversion

The search for DNA motifs of 3, 4, 5, 6, 7 and 8 nucleotide length that were overrepresented in the crossover and gene conversion target sequences compared with the respective flanking sequences resulted in a total of 434 (crossover) and 26 (gene conversion) motifs (Table S6). G/C bases were commonly found in crossover-associated motifs, while motifs that were enriched in gene conversion target sequences typically exhibited a high frequency of A/T bases. The motifs enriched in crossover regions included the previously identified GCCGC motif (Bessoltane et al., 2012). The A/T-rich motifs found to be negatively associated with crossovers, such as AAAGA (Bessoltane et al., 2012), were found to be enriched in gene conversion regions. Corresponding differences in motif enrichment were found between "crossover-associated gene conversion" sequences and "non-crossover gene conversion" sequences (Table S6). Additionally, 20 motifs were found enriched in the 18 crossover hotspots compared with the background but these were less G/C-biased (Table S6).

DISCUSSION

The highly recombining honey bee genome is increasingly studied as a model for understanding the ultimate and proximate causation of meiotic recombination in general. Despite significant progress, a comprehensive understanding of the causes for the excessive recombination rate is lacking, particularly with regard to the mechanisms involved. Here, we extended previous research by generating a large, high-resolution data set that allowed a direct analysis of the fine-scale landscape of recombination in one mapping population. We used these data to better understand how different double-strand break repair mechanisms shape the genomic landscape and which cis-regulatory elements may influence the outcome of double-strand break repair. Our results suggest a mechanistic explanation for the exceptionally high recombination rate of honey bees that complements existing evolutionary hypotheses.

High recombination rates in honey bees: A consequence of high double-strand break rates?

Our results reconfirm the exceptionally high rate of meiotic recombination in the Western Honey Bee, *Apis mellifera*, although the genome-wide estimate of 18 cM/Mbp is lower than previous studies (Beye et al., 2006; Hunt & Page, 1995; Liu et al., 2015; Mougel et al., 2014; Ross, DeFelice, Hunt, Ihle, & Amdam, 2015; Solignac et al., 2007; Wallberg et al., 2015), presumably due to our conservative scoring of crossovers. The extreme recombination rates systematically observed in honey bees (DeLory et al., 2020; Kawakami

et al., 2019; Rueppell et al., 2016), but they are not necessarily the result of adaptive evolution. An adaptive explanation of the high recombination rate would be supported at the molecular level by a clear signal of increased preferential repair of double-strand breaks to crossovers. Our study estimates positive crossover interference to be higher than previous studies (2.16 compared with ~1.5), but even our estimate is lower than the corresponding value (>3) of bumble bees, which have a lower recombination rate (Kawakami et al., 2019). Crossover interference can be attributed to the specific meiotic pathway for double-strand break repair leading preferentially to crossover events (Otto & Payseur, 2019; Von & Rog, 2021), but it cannot explain the high recombination rates of honey bees compared with bumble bees. The lack of evidence of preferential double-strand break repair to crossover events is also demonstrated in our study by the equal number of crossover and gene conversion events per drone. Hence, the high recombination rate observed in honey bees does not seem to be related to a preferential use of the double-strand break repair pathway specific to meiosis.

Instead, the extreme recombination rates observed in honey bees seem to be the consequence of a high double-strand break formation rate in honey bee species. Such high double-strand break rate in honey bees, compared with other species, could be selected for by a strong generation time effect, due to a large colony size, coupled with the A/T-rich genome (Fouks et al., 2021). The production of sexuals in honey bees occurs after the production of thousands of workers, which translates into an increased number of replicative events for those eggs due to the ongoing mitoses (Fouks et al., 2021). This increased number of replication coupled with the A/T-rich genome of honey bees may lead to an increased double-strand break rate, since DNA replication in A/T-rich sites of the genome have been reported to promote double-strand breaks (Wang et al., 2014; Zhang & Freudenreich, 2007). A/T-rich sequences are prone to form secondary structures during replication resulting in the formation of double-strand breaks (Lukusa & Fryns, 2008; Wang et al., 2018). This phenomenon has been well documented in humans, mice and yeasts, where it is related to AT-rich chromosome fragile sites (Durkin & Glover, 2007). Furthermore, double-strand breaks in A/T-rich chromosome fragile sites are typically repaired through the mitotic recombination pathway (Wang et al., 2018). A similar mechanism could explain high recombination rates coupled with low levels of crossover interference throughout the honey bee genome.

Our results indicate that the different double-strand break repair pathways involved during meiosis might be important because they lead to different chromosomal tract exchange. Double-strand break repair leading to gene conversion may result from a random process because our gene conversion count per meiosis followed a Poisson distribution. In contrast, the crossover count per meiosis followed a normal distribution, suggesting that this process is regulated. The latter argument fits well with a model of meiotic recombination, where DNA-binding proteins and specific DNA motifs (Lam & Keeney, 2015) induce double-strand breaks which are then repaired with canonical homologous recombination specific to meiosis leading to a crossover. The argument for different mechanisms leading to either gene conversion or crossover is also supported by our bimodal distribution of tract lengths. Finally, the different base composition of DNA near gene conversions versus crossovers also suggest that different double-strand break repair pathways are involved in each type of chromosomal exchange: While gene conversion occurs within A/T-rich regions, crossovers are mostly present in G/C-rich regions. Generally, A/T-rich regions lead to secondary structure during replication (Lukusa & Fryns, 2008), which makes double-strand breaks more likely in A/T-rich than G/C-rich regions. This strengthens the evidence that most double-strand breaks leading to gene conversion events are the results of a random process, while most double-strand breaks leading to crossover events have evolved to be induced and regulated by the organism to occur at a particular rate.

Our study corroborates previous distinctions between crossover and gene conversion events in honey bees (Bessoltane et al., 2012; Liu et al., 2015), paralleling reports from other species (Comeron et al., 2012; Keeney, 2001; McMahill et al., 2007; Pâques & Haber, 1999). The findings are compatible with the general model of recombination (Kohl & Sekelsky, 2013). Gene conversions may result from synthesis-dependent strand annealing, which is more mutagenic than the double Holliday junction-dependent double-strand break repair (van de Bosch et al., 2002). Synthesis-dependent strand annealing mutations are generally biased towards A/T bases (Birdsell, 2002) because the DNA polymerase involved has a low fidelity (Lercher & Hurst, 2002), which may induce A/T-biased mutations (Birdsell, 2002; Strathern et al., 1995). In contrast, double Holiday junction repair is associated with G/C-biased gene conversion, which may enrich the target sequences of crossovers (Lesecque et al., 2013; Marais, 2003). Thus, our results provide the first evidence that G/C-biased gene conversion could explain the general correlations between recombination and G/C content in the honey bee genome (Beye et al., 2006; Ross, DeFelice, Hunt, Ihle, & Amdam, 2015). Furthermore, our results indicate that G/C-biased gene conversion associated with crossover events may counteract the strong A/T bias of the honey bee genome. Thus, the high recombination rate of the honey bee genome may have been partly selected for mechanistic reasons. Selection for increasing the G/C content and balance the base composition of the genome may lead to up-regulating the formation of double Holliday junctions and meiosis-specific double-strand break repair. In contrast, the alternative synthesis-dependent strand annealing mechanism of double-strand break repair does not increase G/C content and consequently may not be selected for, although we have presented evidence that it regularly occurs in the honey bee genome and the high recombination rate of honey bees cannot be explained by an up-regulation of pro-crossover pathways alone. Comparative support for the hypothesis that high recombination rates have evolved in response to A/T-rich genomes comes from the relatively A/T-rich genomes of the ants *Acromyrmex echinatior* (Nygaard et al., 2011) and *Pogonomyrmex barbatus* (Smith et al., 2011), which also display exceptionally high recombination rates (DeLory et al., 2024; Sirviö et al., 2006). This mechanistic hypothesis complements the hypothesis that selection for novel phenotypic diversity during social evolution has increased recombination near female or worker biased genes, which occur in

genome regions with higher G/C content (Kent et al., 2012; Liu et al., 2015; Wallberg et al., 2015).

Comparison of overall crossover and gene conversion rates in honey bees

Our study has used the previous version of the honey bee genome (Amel_4.5), which differs from the enhanced current chromosome-level assembly (HAv3.1, RefSeq: GCF_003254395.2, GenBank: GCA_003254395.2). Thus, some erroneous assemblies could have inflated our recombination estimates. However, we applied a stringent quality filtering of SNPs, deleting data that would have resulted in unrealistically high recombination rates, as would occur from misassemblies. Additionally, we excluded phase shifts at scaffold boundaries, limiting the errors that could have occurred due to misassemblies. It is likely that our overall estimate is too low due to our conservative definition of crossover events. The stringent criteria were chosen based on the bimodal distribution of phase tract lengths to clearly distinguish between crossover and gene conversion events, the primary objective of our analyses. Based on a 10-kbp threshold to define crossovers and gene conversions, our recombination rate estimate would increase to 22 cM/Mbp. However, neither of our two values agrees with the 37 cM/Mbp reported by the only comparable previous study (Liu et al., 2015). While we cannot exclude biological reasons for the difference, such as age of the sample individuals or the sampled population, overall recombination rate estimates have been relatively constant across different maps (Ross et al., 2015). Our findings suggest that marker density alone does not necessarily increase the recombination rate estimate of the honey bee genome (Liu et al., 2015), but other methodological differences may explain the discrepancies. First, our large sample size of 187 individuals allowed the identification and exclusion of five individuals as statistical outliers. Secondly, we identified many phase changes at some scaffold boundaries, suggesting assembly problems. Consequently, we excluded scaffold boundaries from our analyses in contrast to (Liu et al., 2015). A similar correction, excluding non-unique crossover sites in a data set of 43 individuals, previously reduced Liu et al.'s estimate to 24.5 cM/Mbp (Kent et al., 2012), which is reasonably close to the 22 cM/Mb.

The gene conversion frequency differed more drastically between our study, Liu et al. (2015) and Kawakami et al. (2019). Even with our more stringent size definition for gene conversions, we identified 44.3 gene conversion events per drone, compared with the previously identified 5.8 (Liu et al., 2015) and 20 (Kawakami et al., 2019). In contrast to Liu et al., we did not automatically exclude multi-copy regions, but our size filter presumably prevented most false discoveries of gene conversions (Qi et al., 2014). Furthermore, we excluded markers that exhibited more than two alleles and markers that led to >5% map expansion. Our estimate of gene conversion frequency is only slightly higher than that for the *Drosophila* genome (Comeron et al., 2012). Thus, it is unlikely that we have erroneously overestimated gene conversion frequency (Bessoltane et al., 2012), but our positive correlation between marker density and gene conversion frequency suggests that our high marker density may have allowed a better detection of gene conversions. Despite some variation in the ratio of crossovers to gene conversions in our study when different data filters and definitions are employed (compare main text with supplemental results S1 and S2), our study invariably places the crossover/gene conversion ratio between the two existing previous estimates (Bessoltane et al., 2012; Liu et al., 2015). Likewise, we found some evidence of crossover interference but not to the extent of (Solignac et al., 2007).

Local recombination rate variation

Following the classic definition of recombination hotspots (McVean et al., 2004; Myers et al., 2005), only 18 hotspots were identified overall. None of these hotspots were identified on chromosomes 15 and 16, consistent with (Mougel et al., 2014). Our low number of hotspots could be partially explained by our data filtering and restrictive method to discover hotspots. Furthermore, ours and other linkage map-based studies suffer from a lack of statistical power: Even though we employed a substantially larger sample size than any previous genetic linkage analysis of the honey bee genome, many 10-kbp intervals did not contain any crossover event. The adjacent intervals were automatically excluded from the search for hotspots because any ratio between a crossover value in the adjacent interval and zero is mathematically undefined. This mathematical problem also prevented a more fine-scale analysis, while analyses at larger scales (e.g. 100 kbp) are less meaningful in the context of hotspot identification. Linkage disequilibrium patterns may thus be better suited for hotspot identification (Coop & Przeworski, 2007), but the only study in honey bees to conduct such an analysis also failed to identify pronounced recombination hotspots (Wallberg et al., 2015). Thus, the cumulative empirical evidence suggests that recombinational hotspots are not common or pronounced in the honey bee genome.

Despite the apparent paucity of true hotspots, the recombination rate in the honey bee genome was heterogeneous, with 50% of crossovers found in 19% of the genome. This result is comparable with the finding that 25% of the crossovers are restricted to 5% of the honey bee genome (Liu et al., 2015). Thus, the skew in honey bees is quite low compare with humans, where 71% of crossovers were found in 12% of the genome (Tiemann-Boege et al., 2006) and previous studies in the honey bee reported an even lower skew (Mougel et al., 2014; Wallberg et al., 2015). At the scale of 100 kbp, our recombination estimate regularly crossed the threshold defined by our Poisson estimate, indicating a skewed distribution of local recombination rates. Thus, a large part of the genome experienced recombination rates that deviated significantly from the average recombination rate. However, based on the visual inspection of the data and the cumulative Marey maps, we found no evidence of large-scale differences in recombination rates along chromosomes other than small regions of reduced recombination close to centromeres (Beye et al., 2006).

Overall, our local estimates of recombination rates did not match previous findings in detail, although the values in our study fall well within the range of previous studies (Liu et al., 2015; Solignac et al., 2007). Much of the variation could result from different analysis methods and scale (Ross et al., 2015; Stevison & Noor, 2010), but natural intra-specific variation of local recombination rates in honey bee populations is also a likely explanation (DeLory et al., 2020). Such variation has been suggested previously in honey bees (Ross et al., 2015) and occurs in other species (Brooks & Marks, 1986; Comeron et al., 2012). Fine-scale variation reveals the evolutionary potential of local recombination rates, which are presumably driven by polymorphic local sequence features, such as cis-acting regulatory motifs or chromosome structure (Zickler & Kleckner, 2023).

While intra-specific variation of recombination rate has been studied among individuals and among different genome regions (Chowdhury et al., 2009; Comeron et al., 2012), another dimension of intra-specific variation in recombination occurs within individuals as variation among separate meioses. At this level, variation is assumed to be minor and it is rarely investigated but recombination rates may be more plastic than anticipated (Singh et al., 2015) and quantification of intra-individual plasticity is important for evaluating the evolutionary potential of recombination rates. Our data showed substantial variation even among offspring that were produced in a time-frame of 24 h in the homeostatic environment of a social insect colony. The meioses leading to the production of these offspring presumably occurred under highly similar conditions in close proximity to each other in the germarium of the queen's ovary but produced very different outcomes. This variation, as well as the low degree of preservation of local recombination rates between our and other studies highlights the need for more in-depth comparative studies to gain a deeper understanding of the evolutionary dynamics of the high recombination rates in the honey bee genome.

Local sequence features associated with recombination rate

Local recombination rate may be influenced by specific motifs (Berg et al., 2010), DNA accessibility differences due to transcriptional activity or base composition (Lichten, 2008), and by selection for reduced linkage disequilibrium (Kent et al., 2012; Liu et al., 2015). On average, crossovers and gene conversions were found more distant to exons and UTRs than expected by chance. The distances between recombination events and functional genetic elements reduce intragenic recombination events and/or a reduction in double-strand break-induced deleterious mutations in important sequences (Webster & Hurst, 2012). Gene conversions appeared closer than crossovers to genes, exons, 5′ and 3′ UTRs. This difference suggests also that long DNA sequence exchange by crossover may be selected against in the immediate vicinity of genes. However, analysis of genome regions that harbour important segregating allelic variation (Beye et al., 1996; Graham et al., 2011; Hunt et al., 2007; Rueppell, 2009; Rueppell et al., 2011) demonstrated a higher proportion of both gene conversions and crossovers compared with the rest of the genome, in accordance with general models for adaptive increases in local recombination rate to facilitate independent evolution of genes under selection (Agrawal, 2006; Gandon & Otto, 2007; Hill & Robertson, 1966; Roze, 2012). The contrast between the higher than average gene conversion and crossover rates in these specific genome regions and the localization of gene conversions and crossovers away from genes can be explained by the different scale of these analyses. In parallel to our study, a detailed analysis of the sex determination region found little evidence for recombination (Liu et al., 2015), while earlier and broader-scale analyses identified this genomic region as highly recombining (Beye et al., 1999; Hasselmann & Beye, 2006). Thus, selection seems to have increased recombination rates in important genome regions (Kent et al., 2012) while decreasing its potentially mutagenic impact in and close to genic sequences.

In the honey bee genome, genes are predominantly located in G/C poor regions (Elsik et al., 2014) and consistent with previous studies, our data at the 100 kbp scale indicated that crossovers occur in more G/C-rich regions of the genome (Beye et al., 2006; Kent et al., 2012;Liu et al., 2015; Ross, DeFelice, Hunt, Ihle, & Amdam, 2015). Crossover and gene conversion rates showed a positive correlation with CpG frequency. CpG frequency is naturally related to G/C content but the overall statistical evaluation of the joint effects of these two variables on crossover and gene conversion frequency was complicated by higher-order interactions, including the abundance of low complexity DNA. CpG and DNA methylation are believed to inhibit high recombination rates in honey bees (Mougel et al., 2014; Wallberg et al., 2015). Contrary to this notion, a simple positive correlation between crossover rate and CpG occurrence existed in our data, comparable with previous results derived from multiple low-density recombination maps (Ross et al., 2015).

In the absence of PRDM9, no specific factor responsible for the initial double-strand break has been identified in insects but certain DNA motifs have been repeatedly found to correlate with recombination events in honey bees (Bessoltane et al., 2012; Mougel et al., 2014), especially the 5-mer motif GCCGC. This short sequence could be part of a longer motif (Mougel et al., 2014), which makes a specific initiation of recombination comparable with the PRDM9 mechanism (Berg et al., 2010) conceivable. However, the identified motifs may also be a consequence of a general association between recombination and G/C-enriched sequences (Liu et al., 2015; Ross, DeFelice, Hunt, Ihle, & Amdam, 2015), a notion that is further supported by the identification of numerous recombination-associated motifs in our study.

Overall, gene conversion-associated motifs differed fundamentally from those associated with crossovers, which may be due to different factors binding to induce gene conversions or crossovers. Alternatively, the divergent results could be due to a different mutagenic bias of gene conversion versus crossover events. Nevertheless, one motif was similar between the crossover- and gene conversion-enriched motifs: GGGAC. Thus, this motif could represent or be contained in recognition sites for an initial, general double-

strand break-inducing factor in honey bees. Our study indicates the need for further research that includes crossovers and gene conversions in the search for DNA motifs that guide recombination.

Conclusions

Our study of recombination patterns across 187 genomes of the honey bee have led to several new insights but also reconfirms that the genomic recombination rate of *Apis mellifera* is exceptionally high, supporting earlier estimates of slightly above 20 cM/Mbp. Most gene conversions are the result of non-crossover events. Non-crossovers and crossovers are associated with different motifs, appear in genome regions with opposed nucleotide content, and seem to lead respectively to A/T or G/C bias. This influence on genome composition may select for increased crossover rates in eusocial Hymenoptera. Thus, our results do not support the "genotypic diversity hypothesis" or the "social innovation hypothesis" but point to a mechanistic explanation instead. Our results suggest that extreme recombination rates observed in honey bees compared with other eusocial hymenopterans are the result of increased double-strand break formations, potentially due to an increase in colony size, leading to a stronger generation time effect, and their A/T-rich genome, translating in numerous A/T fragile sites. Our study identifies differences in genome sequence correlates of crossover, gene conversion and "crossover-associated gene conversion" events (Elbakry & Löbrich, 2021), suggesting different cis-regulation of these processes. This evidence complements mutational analyses of recombination pathways that have only been possible in established model organisms. The dichotomy of nucleotide content and bias between gene conversion and crossover, mechanistically explained by alternative consequences of synthesis-dependent strand annealing and of double Holiday junction repair, contributes to our understanding of several specific genome features of the honey bee. To confirm our findings, further studies are needed on other honey bee populations because our study is limited to the progeny of a single female from the general North American commercial population The honey bee with its excessive rate of recombination, haplo-diploid sex determination and large offspring number, provides a valuable model for understanding the evolution of recombination and genome evolution, and the advent of genetic tools that allow identification and manipulation of the recombination machinery in this species further its potential.

EXPERIMENTAL PROCEDURES

Sample collection and extraction of genomic DNA

In honey bees, a single female reproductive (the queen) is responsible for all egg-production in a colony. The queen can lay up to 2000 eggs per day and she can be induced to produce haploid eggs that invariably develop into males (=drones). To induce a high number of male eggs, empty drone comb was introduced into several mature colonies of *Apis mellifera* (L) colonies at the University of North Carolina at Greensboro honey bee yard in Greensboro, NC. After 18 days, the colony with the highest drone brood production in the experimental frame was selected and all pupae were collected. Each drone was individually stored in a 1.5-mL micro-centrifuge tube at −20°C until DNA extraction.

Genomic DNA was extracted from 800 of the collected drone pupae using the Puregene® kit and protocol (Qiagen, Germany). Frozen thoraces were separated with sterile forceps from the abdomen and head and ground on ice until liquid. An aliquot of 100 μL of the liquefied tissue was mixed with 300-μL cell lysis buffer and 1.5-μL Proteinase K (20 mg/mL) and processed according to the manufacturer's recommendation. Extracted DNA was rehydrated in Tris-EDTA (TE) buffer and stored at −20°C. Quality and quantity of the extracted DNA was assessed by spectrophotometry (Nanodrop 1000, Thermo Scientific) and agarose gel electrophoresis. The 192 drone samples with the highest molecular weight DNA were selected for subsequent genotyping and sequencing since all samples exceeded our internal quantity and purity standards.

Confirmation of maternal lineage with microsatellite genotyping

To verify all samples were haploid and derived from the same mother queen, three highly polymorphic microsatellite loci (K0907, K05128 and AP174) were genotyped in all 192 individuals (Solignac et al., 2007). One microliter of sample (100 ng/μL) was mixed with 14 μL of PCR master mix, containing 8.95 μL of molecular grade water, 1.5 μL of 10× buffer, 1.5 μL of 200 μM dNTP solution, 0.35 μL of 0.25 μM forward primer, 0.75 μL of 0.5 μM reverse primer, 0.75 μL of LI-COR 700 IRDye and 0.20 μL of 0.2 μM Taq polymerase. The reaction followed a previously established touch-down protocol (Hayworth et al., 2009) and products were frozen until determination of allele sizes by gel electrophoresis using a LI-COR 4300 DNA analyser (Licor Inc., Lincoln, NB) with a 25 cm 6% polyacrylamide gel (Dixon et al., 2012). All drones were confirmed as true brothers by having one of two maternal alleles at each locus.

DNA library preparation and genome sequencing

Genomic DNA from the selected 192 samples was processed at the High Throughput Sequencing Facility (HTSF) at UNC-Chapel Hill. Genomic DNA was sheared with an E220 focused-ultrasonicator (Covaris Inc., Woburn, MA), quantified and libraries were created using KapaBiosystems® (Woburn, MA) high throughput preparation kit following manufacturer's recommendations, using polyethylene glycol and magnetic beads for all clean-up and size selection (to 250–450 bp) steps. Each library was tagged with dual adapter (2D) indexed adaptors. The 192 samples were separated into two pools with 96 samples per batch. Each batch was sequenced on an Illumina™ HiSeq 2000 (San Diego, CA) machine in one lane of a 100-bp single-end run and two lanes of a 100-bp paired-end run.

Alignment to reference genome and SNP calling

The genome sequence data for the 16 honey bee chromosomes were downloaded from the NCBI website (RefSeq assembly accession: GCF_000002195.4, GenBank: GCA_000002195.1) and concatenated to form one genome file. Sequencing reads were de-multiplexed, and the data from all three lanes were combined for each of the 192 samples. The Burrows–Wheeler Alignment (BWA) tool (Li & Durbin, 2009) was used to align the reads of each of the sequenced samples to this reference genome. Subsequently, SAMtools (Li et al., 2009) utilities were used to call SNPs with minimal quality filtering. The SNP data from the samples were separated by chromosome resulting in chromosome-specific matrices. To allow for the detection of SNPs that segregated in our mapping population but were identical to the reference genome, an alternative reference genome was created, substituting the original sequence with that of the alternative allele at each SNP position. The data were combined into one file for each chromosome that contained for each individual and each SNP either specific SNP genotypes or missing data.

Quality filtering, crossover and gene conversion event identification

Each SNP was evaluated for the distribution of its allele frequencies and amount of missing data. SNPs with a minor allele frequency of <20% or allele counts of <20 were removed. Low-quality markers were also removed based on the occurrence of a third allele and based on an unrealistic expansion of the linkage map: we assumed a recombination frequency of >5% as unrealistic between adjacent markers on the same scaffold, given an average physical marker spacing of <250 bp on each chromosome. After these filtering steps, a total of 931,350 markers remained in the analysis, while 313,140 were dropped. Scaffold boundaries were excluded from the analysis due to the uncertainty of physical size distance between adjacent scaffolds. Five individual samples were removed altogether from the data set due to their high overall number (>10× the average) number of phase shifts, indicating sample contamination or systematic sequencing error. We therefore adopted a conservative approach and analysed the remaining 187 samples without these five outliers.

Marker genotypes were converted to parental phases based on their linkage patterns to the surrounding markers, starting with an arbitrary phase assignment for each chromosome. To estimate and control for the potential effects of erroneous genotype calls three slightly different data sets were generated from these data: The first set was left uncorrected; in the second set, phase shifts that were only supported by a single SNP genotype were replaced with missing data; and in the third set, any phase shift that was shorter than the length of a single read (100 bp) was replaced by missing data, regardless of how many SNPs supported the phase shift. While the second and third data set safeguard against sequencing errors, they also exclude the possibility of discovering short tracts of gene conversion. The second data set with only singleton removal is presented in the main text as the best compromise between excluding short tracts of gene conversion and excluding sequencing errors. The results of the two alternative data sets are presented as supplements ("Supplemental Results S7" represents results from the unfiltered data and "Supplemental Results S8" represents the results from the third, most stringently filtered data set).

Initially, the frequency distribution of all phase tract lengths was visually inspected and tested for deviation from unimodality (Hartigan's dip test for unimodality (Hartigan & Hartigan, 1985)) to ascertain the existence of two distinct phase tract distributions, suggesting distinct gene conversion and crossover processes. The distribution also justified the following categorization of phase shifts as gene conversions and crossovers with custom scripts (Supplemental Methods S9) that analysed the observed phase switches in each individual sample: Phase switches were categorised as gene conversions (Figure 1a–c) or crossovers (Figure 1e) based on phase tract lengths between successive phase switches. A gene conversion tract length was defined as >100 bp and <2.5 kbp, while a crossover event was defined by tract lengths of >25 kbp and >3 markers on each side of the phase shift (Supplemental Methods S9). While tract lengths can be estimated through maximum likelihood methods from population data (Yin et al., 2009), our offspring genotypes were not suitable for such an approach. Additionally, gene conversions that occurred in the immediate vicinity of crossovers were distinguished as "crossover-associated gene conversion" by identifying a tract length <2.5 kbp, followed by another <2.5 kbp tract, which was in turn followed by a tract length > 25 kbp (Supplemental Methods S9). Gene conversions were further split into events that were clearly unlinked to crossover events (non-crossover gene conversions: >50 kbp distance from nearest crossover) and events that were neither linked nor unlinked. All phase shifts that did not fit into those categories were omitted from the further analyses.

The locations for the specific recombination events were determined in the following manner: The beginning position of crossover, gene conversion and "crossover-associated gene conversion" events was equated to the position of the last SNP marker before the phase shift. The end position of crossover events corresponded to the position of the first sequence (marker) after the phase shift. In contrast, the end position of gene conversion events was defined to include the entire gene conversion event tract, including the sequences (markers) after the second phase shift (Supplemental Methods S9). In the same manner, the end position of the "crossover-associated gene conversion" events corresponded to the position of the first sequence (marker) after the third phase shift, thus including the two gene conversion tracts plus the phase shift leading to the longer tract that followed the crossover (Supplemental Methods S9).

The *Apis mellifera* genome Amel_4.5 (Elsik et al., 2014) was divided into 1 kbp windows and the frequency of gene conversions and crossovers was calculated for each window by summing their occurrences across all samples (Supplemental Methods S9). Recombination events that spanned multiple windows were evenly allocated across these windows. The recombination rate of each window was calculated by dividing the total number of its crossovers by the

number of individuals (187) and multiplying by 100,000 in order to obtain a recombination rate in cM/Mbp. Despite the fact that only crossovers determine the genome-wide recombination rate, the same calculation was applied to gene conversion events, allowing a direct comparison of both local rates. Furthermore, Marey maps were created, depicting the overall relation between physical and genetic length for each chromosome (Chakravarti, 1991). The average slope of these Marey maps, as an estimator of the chromosome's recombination rate, was regressed on chromosome size to investigate the relation between recombination rate and chromosome size. In addition, the distribution of crossover and gene conversion counts per individual were plotted to visually assess if their frequency distributions suggest a regulated or random process.

Crossover interference was calculated to determine whether the non-interfering mitotic or interfering meiotic pathway prevails in the repair of double Holiday junctions (Kohl & Sekelsky, 2013). First, crossover interference was inferred by fitting gamma distributions (v, 2v) to all inter-crossover genetic distances using maximum likelihood estimation (Broman & Weber, 2000). Secondly, we compared the distribution of observed number of crossovers among individuals to the corresponding distributions predicted with or without interference for each chromosome (Housworth & Stahl, 2003).

Analyses of local rates of gene conversions and crossovers

Despite our relatively large sample size, most of the 1-kbp windows were devoid of gene conversions and crossovers. Thus, the data were summarised into larger analysis units. According to the definition of recombination hotspots in mammals (McVean et al., 2004; Myers et al., 2005), crossover hotspots were defined as 10-kbp windows that exhibited at least a fivefold crossover rate increase relative to the adjacent 10-kbp windows. Windows that were bordered by regions of zero crossovers were omitted because no relative increase could be computed. Deserts of crossovers and gene conversions were defined as 100-kbp windows without any crossover and gene conversion events, respectively. Following previous studies (Comeron et al., 2012; Ross, DeFelice, Hunt, Ihle, & Amdam, 2015), subsequent analyses also employed a 100-kbp scale.

The frequencies of gene conversions and crossovers were correlated across 100-kbp windows. Secondly, the heterogeneity of crossover rates among 100-kbp windows across the genome was assessed by comparing each value to a 99% threshold that was empirically determined based on a Poisson distribution of our data. Third, we determined the dependency of crossover and gene conversion frequency on marker density by regression analysis and subsequently corrected the crossover and gene conversion frequency values accordingly. Fourth, the frequency of three sequence features (percent of G/C content, CpG and an index of low complexity sequence) in the 100-kbp windows was determined. Their combined effect on the corrected crossover and gene conversion rates was evaluated by multiple regression analysis, with model selection by pairwise AIC comparison (Crawley, 2005). Low complexity was calculated using a custom Python script that used the low complexity index defined previously (Beye et al., 2006).

Analysis of crossovers and gene conversions in select genome regions

To test whether genome regions with important segregating allelic variation demonstrate crossover or gene conversion frequencies that are different from the genomic background, the number of crossovers and gene conversions, and the occurrence of crossovers and gene conversions deserts (in 100 kbp windows) was calculated in 13 select genome regions in which previous studies had identified such allelic variation. These regions included the Pollen Hoarding QTL (Hunt et al., 2007), Worker Ovary Size QTL (Graham et al., 2011; Rueppell et al., 2011), Age of First Foraging QTL (Rueppell, 2009) and the Complementary Sex Determination region that contains the CSD and Fem genes (Beye et al., 1996; Hasselmann et al., 2008). The frequency of crossovers, gene conversions and deserts of gene conversions and crossovers in these regions was compared with the remainder of the genome.

Analyses of crossover and gene conversion sequences

The position of each crossover, gene conversion, "crossover-associated gene conversion" or "non-crossover gene conversion" event was used to retrieve sequences in which these events occurred (target sequences) and the corresponding adjacent sequences (flanking sequences) from the *Apis mellifera* genome (Amel v4.5) from Beebase (Elsik et al., 2014). We used the Bedtools utility "getFasta" (Quinlan, 2014). The length of the retrieved flanking sequences was set equal to the average length of the target event (crossover: 2150 bp, all forms of gene conversion: 950 bp). The G/C content of these sequences was determined with the Bedtools utility "nuc" (Quinlan, 2014) and compared between target and flanking sequences. The G/C content was also compared between crossover and gene conversion target sequences and between "crossover-associated gene conversion" and "non-crossover gene conversion" target sequences. The corresponding flanking sequences were compared in a similar fashion: The G/C content was compared between the flanking sequences of crossover and gene conversion events and between the flanking sequences of "crossover-associated gene conversion" and "non-crossover gene conversion" events.

Next, the retrieved sequences were purged of repeats (vmatch: (Bessoltane et al., 2012; van Helden, 2003)). The word frequencies of flanking regions were calculated and used as background model for DNA motif discovery of the target sequences (crossover, gene conversion, "crossover-associated gene conversion" and "non-crossover gene conversion") using RSAT (van Helden, 2003). In addition, the sequences in recombination hotspots were analysed. To complement these enrichment searches, deserts of crossovers and gene

conversions were also analysed and motifs that were found enriched in the deserts were eliminated from the lists of enriched motifs for crossover and gene conversion, respectively. These procedures were repeated for the analysis of motifs that were 3, 4, 5, 6, 7 and 8 base pairs long. DNA motifs with a higher occurrence in the target regions compared with the background models (occ-sig above 0) are presented.

The distances of crossover and gene conversion positions to the closest gene, exon, 5′ and 3′UTR were calculated and compared using the Bedtools utility "closest" (Quinlan, 2014) and the gene set of Amel 4.5 downloaded from Beebase (Elsik et al., 2014). Furthermore, crossover and gene conversion positions were compared with random positions (retrieved with Bedtools) with respect to their distance from the nearest gene, exon, 5′ and 3′UTR.

Statistical analyses

All statistical analyses were performed in R (v 3.1.2 & v 4.2.3). The test of the unimodality of the phase shifts distribution to, ascertain the existence of two distinct phase tract distributions, was performed using the dip. test() function from the diptest package (DOI: 10.32614/CRAN.package.diptest).

All analyses of local rates of crossovers and gene conversions at the 100 kb windows were performed with linear regression analyses, using the lm() function. To ensure the robustness of our linear models, *p*-values of the best model were bootstrapped ($n = 9999$) by resampling half of the original data set, with the package boot (Canty and Ripley 2017: Package "boot" - *Bootstrap Functions. CRAN R Proj*).

To assess if select regions underlying important traits in honey bees possesses higher crossover and gene conversion rates, we compared the crossover and gene conversion rates across 100-kb windows around those regions to the rest of the genome. We used a generalised linear model with a Poisson distribution as our sample size was limited by the number of 100-kb windows within select regions $n = 39$. The dependent variable was the number of either crossover or gene conversion events, and the independent variables were the genome region (selected or control genome region), the gene density and their interaction. We used one-sided tests for genome region and gene density to test the directional prediction that either variable increases the number of crossover or gene conversion events. We tested the robustness of our results by bootstrapping, omitting 100-kb windows without genes to avoid bias due to our zero-inflated data.

The comparison of G/C content between the different events (crossover, gene conversion, crossover-associated gene conversion and non-crossover gene conversion) was performed using a *t*-test with Bonferroni correction to account for multiple comparisons. In addition, the comparison of GC content of those events with their flanking regions were done using a paired *t*-test and Bonferroni correction.

The distances of several gene features were compared between crossovers, gene conversions or random genome positions obtained with Bedtools with *t*-tests. To assess the validity of our results, a bootstrap approach was used by resampling half the data set while maintaining even proportion between crossovers/gene conversions and the random sequences.

AUTHOR CONTRIBUTIONS

Bertrand Fouks: Investigation; methodology; visualization; writing – original draft; formal analysis; data curation; validation. **Katelyn J. Miller:** Methodology; data curation; investigation. **Caitlin Ross:** Investigation; methodology; software. **Corbin Jones:** Conceptualization; methodology; resources. **Olav Rueppell:** Funding acquisition; conceptualization; writing – review and editing; project administration; resources; supervision; visualization.

ACKNOWLEDGEMENTS

We would like to acknowledge the support of Yan Shi and Alicia Brandt at the UNC high throughput sequencing facility for practical help in generating the sequencing data. Ed Vargo, then in the Entomology Department at NCSU, provided help for genotyping the drone offspring. Furthermore, we would like to thank David Remington, Malcolm Schug and all members of the UNCG Social Insect Lab for their support, encouragement and suggestions. This study would not have been possible without the financial support of the US National Institutes of Health (NIGMS, R15GM102753), Natural Sciences and Engineering Research Council of Canada (RGPIN-2022-03629), the US Department of Defence (W911NF1520045 and W911NF2210195).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Supplemental *information* on methods, including computational scripts, additional results and figures are available as supplemental electronic files (S1–S9), which are referenced throughout this text. The raw data of this study have been deposited for public access at NCBI (Bioproject PRJNA316380, Biosample SAMN04579688) in the short read archive (accession #: SRP083190).

REFERENCES

Agrawal, A.F. (2006) Evolution of sex: why do organisms shuffle their genotypes? *Current Biology*, 16, 696–704.

Allers, T. & Lichten, M. (2001) Differential timing and control of noncrossover and crossover recombination during meiosis. *Cell*, 106, 47–57.

Baker, B.S., Carpenter, A.T.C., Esposito, M.S., Esposito, R.E. & Sandler, L. (1976) The genetic control of meiosis. *Annual Review of Genetics*, 10, 53–134.

Baudat, F., Imai, Y. & de Massy, B. (2013) Meiotic recombination in mammals: localization and regulation. *Nature Reviews. Genetics*, 14, 794–806. Available from: https://doi.org/10.1038/nrg3573

Berg, I.L., Neumann, R., Lam, K.W., Sarbajna, S. & Odenthal-Hesse, L. (2010) PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nature Genetics*, 42, 859–863.

Bessoltane, N., Toffano-Nioche, C., Solignac, M. & Mougel, F. (2012) Fine scale analysis of crossover and non-crossover and detection of recombination sequence motifs in the honeybee (*Apis mellifera*). *PLoS One*, 7, 36229.

Beye, M., Gattermeier, I., Hasselmann, M., Gempe, T. & Schioett, M. (2006) Exceptionally high levels of recombination across the honey bee genome. *Genome Research*, 16, 1339–1344.

Beye, M., Hunt, G.J., Page, R.E., Fondrk, M.K. & Grohmann, L. (1999) Unusually high recombination rate detected in the sex locus region of the honey bee (*Apis mellifera*). *Genetics*, 153, 1701–1708.

Beye, M., Moritz, R., Crozier, R. & Crozier, Y. (1996) Mapping the sex locus of the honeybee (*Apis mellifera*). *Die Naturwissenschaften*, 83, 424–426.

Birdsell, J.A. (2002) Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Molecular Biology and Evolution*, 19, 1181–1197.

Broman, K.W. & Weber, J.L. (2000) Characterization of human crossover interference. *American Journal of Human Genetics*, 66, 1911–1926.

Brooks, L.D. & Marks, R.W. (1986) The organization of genetic variation for recombination in *Drosophila melanogaster*. *Genetics*, 114, 525–547.

Chakravarti, A. (1991) A graphical representation of genetic and physical maps: the Marey map. *Genomics*, 11, 219–222.

Chowdhury, R., Bois, P., Feingold, E., Sherman, S.L. & Cheung, V.G. (2009) Genetic analysis of variation in human meiotic recombination. *PLoS Genetics*, 5, 1000648.

Comeron, J.M., Ratnappan, R. & Bailin, S. (2012) The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genetics*, 8, 1002905.

Coop, G. & Przeworski, M. (2007) An evolutionary view of human recombination. *Nature Reviews. Genetics*, 8, 23–34.

Crawley, M.J. (2005) *Statistics: an introduction using R*. Hoboken, NJ: Wiley.

DeLory, T., Funderburk, K., Miller, K., Zuluaga-Smith, W., McPherson, S., Pirk, C.W. et al. (2020) Local variation in recombination rates of the honey bee (*Apis mellifera*) genome among samples from six disparate populations. *Insectes Sociaux*, 67, 127–138. Available from: https://doi.org/10.1007/s00040-019-00736-6

DeLory, T.J., Romiguier, J. & Rueppell, O. (2024) Recombination rate variation in social insects: an adaptive perspective. *Annual Review of Genetics*, 58, 7.1–7.23. Available from: https://doi.org/10.1146/annurev-genet-111523-102550

Dixon, L.R., McQuage, M.R., Lonon, E.J., Buehler, D. & Seck, O. (2012) Pleiotropy of segregating genetic variants that affect honey bee worker life expectancy. *Experimental Gerontology*, 47, 631–637.

Dumont, B.L. & Payseur, B.A. (2008) Evolution of the genomic rate of recombination in mammals. *Evolution*, 62, 276–294.

Durkin, S.G. & Glover, T.W. (2007) Chromosome fragile sites. *Annual Review of Genetics*, 41, 169–192. Available from: https://doi.org/10.1146/annurev.genet.41.042007.165900

Elbakry, A. & Löbrich, M. (2021) Homologous recombination sub-pathways: a tangle to resolve. *Frontiers in Genetics*, 12, 12. Available from: https://doi.org/10.3389/fgene.2021.723847

Elsik, C., Worley, K., Bennett, A., Beye, M. & Camara, F. (2014) Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics*, 15, 86.

Fouks, B., Brand, P., Nguyen, H.N., Herman, J., Camara, F., Ence, D. et al. (2021) The genomic basis of evolutionary differentiation among honey bees. *Genome Research*, 31, 1203–1215. Available from: https://doi.org/10.1101/gr.272310.120

Gandon, S. & Otto, S.P. (2007) The evolution of sex and recombination in response to abiotic or coevolutionary fluctuations in epistasis. *Genetics*, 175, 1835–1853.

Graham, A., Munday, M., Kaftanoglu, O., Page, R. & Amdam, G. (2011) Support for the reproductive ground plan hypothesis of social evolution and major QTL for ovary traits of Africanized worker honey bees (*Apis mellifera* L.). *BMC Evolutionary Biology*, 11, 95.

Hartigan, J.A. & Hartigan, P.M. (1985) The Dip test of unimodality. *The Annals of Statistics*, 13, 70–84.

Hasselmann, M. & Beye, M. (2006) Pronounced differences of recombination activity at the sex determination locos of the honeybee, a locus under strong balancing selection. *Genetics*, 174, 1469–1480. Available from: https://doi.org/10.1534/genetics.106.062018

Hasselmann, M., Gempe, T., Schiøtt, M., Nunes-Silva, C.G. & Otte, M. (2008) Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature*, 454, 519–522.

Hayworth, M.K., Johnson, N.G., Wilhelm, M.E., Gove, R.P. & Metheny, J.D. (2009) Added weights lead to reduced flight behavior and mating success in polyandrous honey bee queens (*Apis mellifera*). *Ethology*, 115, 698–706.

Hill, W.G. & Robertson, A. (1966) The effect of linkage on limits to artificial selection. *Genetical Research*, 8, 269–294.

Housworth, E.A. & Stahl, F.W. (2003) Crossover interference in humans. *American Journal of Human Genetics*, 73, 188–197.

Hunt, G.J., Amdam, G.V., Schlipalius, D., Emore, C. & Sardesai, N. (2007) Behavioral genomics of honeybee foraging and nest defense. *Die Naturwissenschaften*, 94, 247–267.

Hunt, G.J. & Page, R.E. (1995) Linkage map of the honey bee, *Apis mellifera*, based on RAPD markers. *Genetics*, 139, 1371–1382.

Hunter, C.M., Huang, W., Mackay, T.F.C. & Singh, N.D. (2016) The genetic architecture of natural variation in recombination rate in *Drosophila melanogaster*. *PLOS Genetics*, 12, e1005951. Available from: https://doi.org/10.1371/journal.pgen.1005951

John, B. (2005) *Meiosis*. New York: Cambridge University Press.

Kawakami, T., Wallberg, A., Olsson, A., Wintermantel, D., de Miranda, J.R., Allsopp, M. et al. (2019) Substantial heritable variation in recombination rate on multiple scales in honeybees and bumblebees. *Genetics*, 212, 1101–1119. Available from: https://doi.org/10.1534/genetics.119.302008

Keeney, S. (2001) Mechanism and control of meiotic recombination initiation. *Current Topics in Developmental Biology*, 52, 1–53.

Kent, C.F., Minaei, S., Harpur, B.A. & Zayed, A. (2012) Recombination is associated with the evolution of genome structure and worker behavior in honey bees. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 18012–18017.

Kent, C.F. & Zayed, A. (2013) Evolution of recombination and genome structure in eusocial insects. *Communicative & Integrative Biology*, 6, 22919.

Kohl, K.P. & Sekelsky, J. (2013) Meiotic and mitotic recombination in meiosis. *Genetics*, 194, 327–334. Available from: https://doi.org/10.1534/genetics.113.150581

Lam, I. & Keeney, S. (2015) Mechanism and regulation of meiotic recombination initiation. *Cold Spring Harbor Perspectives in Biology*, 7, a016634. Available from: https://doi.org/10.1101/cshperspect.a016634

Lercher, M.J. & Hurst, L.D. (2002) Human SNP variability and mutation rate are higher in regions of high recombination. *Trends in Genetics*, 18, 337–340.

Lesecque, Y., Mouchiroud, D. & Duret, L. (2013) GC-biased gene conversion in yeast is specifically associated with crossovers: molecular mechanisms and evolutionary significance. *Molecular Biology and Evolution*, 30, 1409–1419.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T. & Ruan, J. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Lichten, M. (2008) *Meiotic chromatin: the substrate for recombination initiation. Recombination and meiosis*. Berlin: Springer, pp. 165–193.

Liu, H., Zhang, X., Huang, J., Chen, J.-Q. & Tian, D. (2015) Causes and consequences of crossing-over evidenced via a high-resolution recombinational landscape of the honey bee. *Genome Biology*, 16, 15.

Louis, E.J. & Borts, R.H. (2003) Meiotic recombination: too much of a good thing? *Current Biology*, 13, 953–955.

Lukusa, T. & Fryns, J.P. (2008) Human chromosome fragility. *Biochimica et Biophysica Acta: Gene Regulatory Mechanisms*, 1779, 3–16. Available from: https://doi.org/10.1016/j.bbagrm.2007.10.005

Marais, G. (2003) Biased gene conversion: implications for genome and sex evolution. *Trends in Genetics*, 19, 330–338.

Mattila, H.R. & Seeley, T.D. (2007) Genetic diversity in honey bee colonies enhances productivity and fitness. *Science*, 317, 362–364.

Maynard-Smith, J. (1978) *The evolution of sex*. Cambridge: Cambridge University Press.

McKim, K.S. & Hayashi-Hagihara, A. (1998) mei-W68 in *Drosophila melanogaster* encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved. *Genes & Development*, 12, 2932–2942. Available from: https://doi.org/10.1101/gad.12.18.2932

McMahill, M.S., Sham, C.W. & Bishop, D.K. (2007) Synthesis-dependent strand annealing in meiosis. *PLoS Biology*, 5, 299.

McVean, G.A., Myers, S.R., Hunt, S., Deloukas, P. & Bentley, D.R. (2004) The fine-scale structure of recombination rate variation in the human genome. *Science*, 304, 581–584.

Mougel, F., Poursat, M.-A., Beaume, N., Vautrin, D. & Solignac, M. (2014) High-resolution linkage map for two honeybee chromosomes: the hotspot quest. *Molecular Genetics and Genomics*, 289, 11–24.

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, 321–324.

Nygaard, S., Zhang, G., Schiøtt, M., Li, C. & Wurm, Y. (2011) The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Research*, 21, 1339–1348.

Otto, S.P. & Lenormand, T. (2002) Evolution of sex: resolving the paradox of sex and recombination. *Nature Reviews. Genetics*, 3, 252–261.

Otto, S.P. & Payseur, B.A. (2019) Crossover interference: shedding light on the evolution of recombination. *Annual Review of Genetics*, 53, 19–44. Available from: https://doi.org/10.1146/annurev-genet-040119-093957

Pâques, F. & Haber, J.E. (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews*, 63, 349–404.

Piazza, A. & Heyer, W.-D. (2019) Homologous recombination and the formation of complex genomic rearrangements. *Trends in Cell Biology*, 29, 135–149. Available from: https://doi.org/10.1016/j.tcb.2018.10.006

Purandare, S.M. & Patel, P.I. (1997) Recombination hot spots and human disease. *Genome Research*, 7, 773–786.

Qi, J., Chen, Y., Copenhaver, G.P. & Ma, H. (2014) Detection of genomic variations and DNA polymorphisms and impact on analysis of meiotic recombination and genetic mapping. *Proceedings of the National Academy of Sciences*, 111, 10007–10012.

Quinlan, A.R. (2014) BEDTools: the Swiss-army tool for genome feature analysis. *Current Protocols in Bioinformatics*, 47, 11.12.1-34.

Rodgers, K. & McVey, M. (2016) Error-prone repair of DNA double-Strand breaks. *Journal of Cellular Physiology*, 231, 15–24. Available from: https://doi.org/10.1002/jcp.25053

Ross, C., DeFelice, D., Hunt, G., Ihle, K. & Rueppell, O. (2015) In: Collaborative Mathematics and Statistics Research (Ed.) *A comparison of multiple genome-wide recombination maps in Apis mellifera*. Switzerland: Springer International Publishing, pp. 91–98.

Ross, C.R., DeFelice, D.S., Hunt, G.J., Ihle, K.E. & Amdam, G.V. (2015) Genomic correlates of recombination rate and its variability across eight recombination maps in the western honey bee (*Apis mellifera* L.). *BMC Genomics*, 16, 107.

Roze, D. (2012) Disentangling the benefits of sex. *PLoS Biology*, 10, 1001321.

Rueppell, O. (2009) Characterization of quantitative trait loci for the age of first foraging in honey bee workers. *Behavior Genetics*, 39, 541–553.

Rueppell, O., Kuster, R., Miller, K., Fouks, B., Rubio Correa, S., Collazo, J. et al. (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 Biology and Evolution*, 8, evw269. Available from: https://doi.org/10.1093/gbe/evw269

Rueppell, O., Meier, S. & Deutsch, R. (2012) Multiple mating but not recombination causes quantitative increase in offspring genetic diversity for varying genetic architectures. *PLoS One*, 7, 47220.

Rueppell, O., Metheny, J.D., Linksvayer, T., Fondrk, M.K. & Page, R. (2011) Genetic architecture of ovary size and asymmetry in European honeybee workers. *Heredity*, 106, 894–903.

Samuk, K., Manzano-Winkler, B., Ritz, K.R. & Noor, M.A.F. (2020) Natural selection shapes variation in genome-wide recombination rate in *Drosophila pseudoobscura*. *Current Biology*, 30, 1517–1528. Available from: https://doi.org/10.1016/j.cub.2020.03.053

Schulte, C., Theilenberg, E., Müller-Borg, M., Gempe, T. & Beye, M. (2014) Highly efficient integration and expression of piggyBac-derived cassettes in the honeybee (*Apis mellifera*). *Proceedings of the National Academy of Sciences*, 111, 9003–9008.

Seeley, T.D. & Tarpy, D.R. (2007) Queen promiscuity lowers disease within honeybee colonies. *Proceedings of the Royal Society B*, 274, 67–72.

Sherman, P.W. (1979) Insect chromosome numbers and eusociality. *The American Naturalist*, 113, 925–935.

Shi, Q., Spriggs, E., Field, L.L., Ko, E. & Barclay, L. (2001) Single sperm typing demonstrates that reduced recombination is associated with the production of aneuploid 24, XY human sperm. *American Journal of Medical Genetics*, 99, 34–38.

Singh, N.D., Criscoe, D.R., Skolfield, S., Kohl, K.P. & Keebaugh, E.S. (2015) Fruit flies diversify their offspring in response to parasite infection. *Science*, 349, 747–750.

Sirviö, A., Gadau, J., Rueppell, O., Lamatsch, D. & Boomsma, J.J. (2006) High recombination frequency creates genotypic diversity in colonies of the leaf-cutting ant *Acromyrmex echinatior*. *Journal of Evolutionary Biology*, 19, 1475–1485.

Smith, C.R., Smith, C.D., Robertson, H.M., Helmkampf, M. & Zimin, A. (2011) Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 5667–5672.

Solignac, M., Mougel, F., Vautrin, D., Monnerot, M. & Cornuet, J.M. (2007) A third-generation microsatellite-based linkage map of the honey bee, *Apis mellifera*, and its comparison with the sequence-based physical map. *Genome Biology*, 8, 66.

Stapley, J., Feulner, P.G.D., Johnston, S.E., Santure, A.W. & Smadja, C.M. (2017) Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philosophical Transactions of the Royal Society B*, 372, 20160455. Available from: https://doi.org/10.1098/rstb.2016.0455

Stevison, L.S. & Noor, M.A.F. (2010) Genetic and evolutionary correlates of fine-scale recombination rate variation in *Drosophila persimilis*. *Journal of Molecular Evolution*, 71, 332–345.

Strathern, J.N., Shafer, B.K. & McGill, C.B. (1995) DNA synthesis errors associated with double-strand-break repair. *Genetics*, 140, 965–972.

Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. & Stahl, F.W. (1983) The double-strand-break repair model for recombination. *Cell*, 33, 25–35.

Tiemann-Boege, I., Calabrese, P., Cochran, D.M., Sokol, R. & Arnheim, N. (2006) High-resolution recombination patterns in a region of human chromosome 21 measured by sperm typing. *PLoS Genetics*, 2, 70.

van de Bosch, M., Lohman, P.H.M. & Pastink, A. (2002) DNA double-strand break repair by homologous recombination. *Biological Chemistry*, 383, 873–892.

van Helden, J. (2003) Regulatory sequence analysis tools. *Nucleic Acids Research*, 31, 3593–3596.

Voelkel-Meiman, K., Johnston, C., Thappeta, Y., Subramanian, V.V. & Hochwagen, A. (2015) Separable crossover-promoting and

crossover-constraining aspects of Zip1 activity during budding yeast meiosis. *PLoS Genetics*, 11, 1005335.
Von, D.L. & Rog, O. (2021) Let's get physical – mechanisms of crossover interference. *Journal of Cell Science*, 134, 255745. Available from: https://doi.org/10.1242/jcs.255745
Waiker, P., de Abreu, F.C.P., Luna-Lucena, D., Freitas, F.C.P., Simões, Z.L.P. & Rueppell, O. (2021) Recombination mapping of the Brazilian stingless bee *Frieseomelitta varia* confirms high recombination rates in social hymenoptera. *BMC Genomics*, 22, 673. Available from: https://doi.org/10.1186/s12864-021-07987-3
Wallberg, A., Glémin, S. & Webster, M.T. (2015) Extreme recombination frequencies shape genome variation and evolution in the honeybee, *Apis mellifera*. *PLOS Genetics*, 11, e1005189. Available from: https://doi.org/10.1371/journal.pgen.1005189
Wang, H., Li, S., Zhang, H., Wang, Y., Hao, S. & Wu, X. (2018) BLM prevents instability of structure-forming DNA sequences at common fragile sites. *PLOS Genetics*, 14, e1007816. Available from: https://doi.org/10.1371/journal.pgen.1007816
Wang, H., Li, Y., Truong, L.N., Shi, L.Z., Hwang, P.Y.-H., He, J. et al. (2014) CtIP maintains stability at common fragile sites and inverted repeats by end resection-independent endonuclease activity. *Molecular Cell*, 54, 1012–1021. Available from: https://doi.org/10.1016/j.molcel.2014.04.012
Webster, M.T. & Hurst, L.D. (2012) Direct and indirect consequences of meiotic recombination: implications for genome evolution. *Trends in Genetics*, 28, 101–109.
Wilfert, L., Gadau, J. & Schmid-Hempel, P. (2007) Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity*, 98, 189–197.
Yin, J., Jordan, M.I. & Song, Y.S. (2009) Joint estimation of gene conversion rates and mean conversion tract lengths from population SNP data. *Bioinformatics*, 25, 231–239.
Zhang, H. & Freudenreich, C.H. (2007) An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in *S. cerevisiae*. *Molecular Cell*, 27, 367–379. Available from: https://doi.org/10.1016/j.molcel.2007.06.012
Zhang, J., Fujiwara, Y., Yamamoto, S. & Shibuya, H. (2019) A meiosis-specific BRCA2 binding protein recruits recombinases to DNA double-strand breaks to ensure homologous recombination. *Nature Communications*, 10, 722. Available from: https://doi.org/10.1038/s41467-019-08676-2
Zickler, D. & Kleckner, N. (2023) Meiosis: dances between homologs. *Annual Review of Genetics*, 57, 1–63. Available from: https://doi.org/10.1146/annurev-genet-061323-044915

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. Following Housworth and Stahl 2003 (Housworth & Stahl, 2003), we determined some degree of crossover interference in the honey bee genome: The observed distributions of crossovers per chromosome (black bars) are intermediate between the distributions of crossovers predicted with crossover interference (light grey bars) and without interference (dark grey bars), with some variation among chromosomes.

Figure S2. Distribution of recombination rate along chromosomes. Recombination rate averaged into 100 kbp window sliding every 10 kbp was plotted along chromosome 2–16. The threshold line represents the Poisson estimate ($p < 0.01$). Most of the recombination rate peaks are above the threshold line, which demonstrates the heterogeneity of the distribution of the recombination rates. However, those high recombination peaks appear throughout the chromosome and are not clustered into a small part of the genome, which contrasts with recombination rate distribution in previous species.

Figure S3. Marker density increased the detection of crossovers and gene conversions, an effect that may explain some differences among the existing studies of recombination in honey bees, but also had to be statistically accounted for in subsequent analyses of this study. Due to the smaller tract length, gene conversion events are more dependent on marker density and our value thus represents an underestimate despite our very high marker density.

Figure S4. Scatterplots illustrate the positive relation of crossover (a) and gene conversion (b) frequencies to G/C content, CpG frequency, and low complexity indices across the genome in 100-kbp windows. Simultaneous analyses of all three factors with more complex model selection based on AIC indicated more complicated effects, including interactions among factors (Table 2). However, graphs represent intuitive relations that are likewise valid descriptions our findings.

Figure S5. In correspondence with previous studies, the G/C content was higher in genome regions where crossovers occurred than the genome average. Moreover, the target region of crossovers was significantly more enriched than the flanking regions. Crossover regions were also more G/C rich than genome regions in which gene conversions occurred. Gene conversion events were conversely leading to a lower G/C content in the target sequences compared with their flanking regions.

Table S6. DNA sequence motifs that are overrepresented in crossover and gene conversion regions, as well as in regions of crossover-associated gene conversion, non-crossover gene conversion, and recombination hotspots.

Data S7. The alternative SNP genotype set that included tracts that were only supported by a single SNP resulted in slightly more gene conversion and less crossover events. All major results were consistent with the main analysis described in the regular text.

Data S8. The alternative SNP genotype set that excluded all tracts that were smaller than 100 bp and thus potentially derived from a single sequencing read resulted in slightly less gene conversion and more crossover events. All major results were consistent with the main analysis described in the regular text.

Data S9. Explanation of methodological details for characterisation of crossover and gene conversion events. Scripts for custom coding, mostly in Visual Basics for Applications (VBA) format.

How to cite this article: Fouks, B., Miller, K.J., Ross, C., Jones, C. & Rueppell, O. (2025) Alternative double strand break repair pathways shape the evolution of high recombination in the honey bee, *Apis mellifera*. *Insect Molecular Biology*, 34(1), 185–202. Available from: https://doi.org/10.1111/imb.12961