

# Exceptionally high levels of recombination across the honey bee genome

Martin Beye,<sup>1,9</sup> Irene Gattermeier,<sup>1</sup> Martin Hasselmann,<sup>1</sup> Tanja Gempe,<sup>1</sup> Morten Schioett,<sup>1</sup> John F. Baines,<sup>2</sup> David Schlipalius,<sup>3</sup> Florence Mougél,<sup>4</sup> Christine Emore,<sup>3</sup> Olav Rueppell,<sup>5</sup> Anu Sirviö,<sup>6</sup> Ernesto Guzmán-Novoa,<sup>7</sup> Greg Hunt,<sup>3</sup> Michel Solignac,<sup>4</sup> Robert E. Page Jr.<sup>8</sup>

<sup>1</sup>Institute of Genetics, Heinrich Heine Universität Duesseldorf, Duesseldorf 40225, Germany; <sup>2</sup>Section of Evolutionary Biology, Department of Biology II, University of Munich, Planegg-Martinsried 82152, Germany; <sup>3</sup>Department of Entomology, Purdue University, West Lafayette, Indiana 47907, USA; <sup>4</sup>Populations, Génétique et Evolution, Centre National de la Recherche Scientifique, F91198 Gif-sur-Yvette, France; <sup>5</sup>Department of Biology, University of North Carolina, Greensboro, North Carolina 27403, USA; <sup>6</sup>Department of Biology, University of Oulu 90014, Finland; <sup>7</sup>Department of Environmental Biology, University of Guelph, Ontario N1G 2W1, Canada; <sup>8</sup>School of Life Sciences, Arizona State University, Tempe, Arizona 85287, USA

The first draft of the honey bee genome sequence and improved genetic maps are utilized to analyze a genome displaying 10 times higher levels of recombination (19 cM/Mb) than previously analyzed genomes of higher eukaryotes. The exceptionally high recombination rate is distributed genome-wide, but varies by two orders of magnitude. Analysis of chromosome, sequence, and gene parameters with respect to recombination showed that local recombination rate is associated with distance to the telomere, GC content, and the number of simple repeats as described for low-recombining genomes. Recombination rate does not decrease with chromosome size. On average 5.7 recombination events per chromosome pair per meiosis are found in the honey bee genome. This contrasts with a wide range of taxa that have a uniform recombination frequency of about 1.6 per chromosome pair. The excess of recombination activity does not support a mechanistic role of recombination in stabilizing pairs of homologous chromosome during chromosome pairing. Recombination rate is associated with gene size, suggesting that introns are larger in regions of low recombination and may improve the efficacy of selection in these regions. Very few transposons and no retrotransposons are present in the high-recombining genome. We propose evolutionary explanations for the exceptionally high genome-wide recombination rate.

[Supplemental material is available online at [www.genome.org](http://www.genome.org). The sequence data from this study have been submitted to dbGSS under accession nos. 15028937–15029063.]

Sex is believed to have evolved as a mechanism to break apart and recombine genetic material (Michod and Levin 1988). The exchange of DNA between homologous chromosomes by meiotic recombination proves to be an ancient and ubiquitous process among eukaryotes, suggesting the fundamental importance of recombinant sex to many groups of organisms. Although recombination is a ubiquitous process, the evolutionary advantages are elusive (Otto and Lenormand 2002). Mixing genes with another individual is a risky endeavor that involves costs of mating and sexual reproduction when compared with asexual reproduction (Michod and Levin 1988). The advantages of meiotic recombination are difficult to understand, as they disrupt allelic associations already tested and proven by natural selection.

Mechanistic explanations of meiotic recombination posit that the narrow range of chiasmata per chromosome per meiosis (on average 1.56) over a wide range of different taxa is consistent with an important role in stabilizing pairs of chromosomes during meiosis (Baker et al. 1976). In addition, multiple sequence characteristics are known to be associated with recombination that could be a cause or consequence of recombination

(Jensen-Seaman et al. 2004). Alternatively, evolutionary explanations consider the selective advantages of recombination (Otto and Lenormand 2002). For example, selection is more effective when the fates of alleles at different loci are uncoupled by recombination (Hill and Robertson 1966; Barton and Otto 2005).

Two characteristics that may shed more light on the mechanistic or evolutionary advantage(s) of recombination are that recombination rates (recombination events per physical distance) vary across species and within genomes (Nachman 2002; Jensen-Seaman et al. 2004). Linkage mapping in the honey bee (*Apis mellifera*) showed that this species has an unusually high meiotic recombination rate, on the order of 10-fold higher than that of *Drosophila* and other genetic model systems (Supplemental Table 1; Hunt and Page Jr. 1995; Solignac et al. 2004). The genome-wide recombination rate of the bee is several-fold higher than that reported for any other higher eukaryote, but in the range of some fungi.

With the first draft of the honey bee genome (The Honey Bee Genome Sequencing Consortium 2006) and the increased number of genetic markers used in this study, we are now in the position to analyze this exceptionally high rate in greater detail. In this study, we analyze the distribution of recombination across the genome to evaluate whether this high rate is a genome-wide phenomenon or limited to specific regions associated with structural properties of chromosomes, or possibly the result

## <sup>9</sup>Corresponding author.

E-mail [Martin.Beye@uni.duesseldorf.de](mailto:Martin.Beye@uni.duesseldorf.de); fax 49-211-8112279.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.5680406>. Freely available online through the *Genome Research* Open Access option.

of selection for recombination in specific regions of the genome. We analyze chromosome, DNA sequence, gene characteristics, and nucleotide polymorphism in a natural population as potentially informative identifiers of the causes and consequences of the heterogeneity in recombination. We discuss these results with respect to patterns found in other studied organisms that have much lower genome-wide recombination.

## Results

### Genome-wide distribution

With the availability of improved genetic maps, we first analyzed the distribution of recombination rates throughout the genome. The genetic map used for genome assembly 3.1 spans 4553 cM, with a physical size of 238 Mb. Thus, the average genome-wide recombination rate is 19 cM/Mb, which is several times higher than that of *Drosophila* (1.59 cM/Mb), mouse (0.56 cM/Mb), human (1.22 cM/Mb), and various other animal species (Supplemental Table 1). Local recombination rates along chromosomes were estimated by comparing genetic distances (cM) between markers to physical distance (Mb) in 250- and 125-kb non-overlapping windows. Estimates of genetic distance were integrated from 1500 new genetic markers and two independent genetic maps, one based on 1100 AFLP and RAPD markers (of which 300 were cloned and anchored to the sequence), and a microsatellite map consisting of 2000 markers. The genetic maps are relatively marker dense, with an average of around one marker per 100 kb. The resolution varies between maps and regions and is based on an average of more than 100 meioses from single females. The analyses of windows required relatively long stretches of sequence, with the requirements of (1) using scaffolds over 125 kb long, (2) containing at least two genetic markers, and (3) containing no more than 10% undetermined nucleotide sequence in the 125-kb window. Half of the sequence falls into scaffolds. The sequence analyzed spans 73 Mb and comprises about one-third of the genome.

Bees display significantly more within genome variation in local recombination rate (average  $28.6 \pm 19$  [kb/cM] for 125-kb windows) than previously reported for human (Yu et al. 2001), *Drosophila melanogaster*, and *Caenorhabditis elegans* (Table 1). The proportion of the surveyed region with relatively low recombination (0–5 cM/Mb) is about 9%, similar to the amount found within the genome of flies and humans, and similar to recombination activity in the majority of the worm genome. We identified 12% of the sequences in which the recombination rate is >50 cM/Mb, the highest estimate being 143 cM/Mb. This part of the genome has 10-fold higher recombination than the highest estimates in humans or the fly. The very large differences of local recombination rates in the honey bee genome have important

**Table 1.** Distribution of recombination rates of *A. mellifera*, *D. melanogaster*, *C. elegans*, and humans when compared among windows of different scales

| Species                        | Scale (kb) | Range | Mean            | Median |
|--------------------------------|------------|-------|-----------------|--------|
| <i>Apis mellifera</i>          | 125        | 0–143 | $28.6 \pm 19.3$ | 25.5   |
|                                | 250        | 0–94  | $28.7 \pm 15$   | 27.1   |
| <i>Drosophila melanogaster</i> | 250        | 0–4   | $2.3 \pm 1.3$   | 2.7    |
| <i>Caenorhabditis elegans</i>  | 250        | 0–15  | $3 \pm 2.7$     | 2.4    |
| Human                          | 250        | 0–6   | $1.3 \pm 0.8$   |        |

**Table 2.** Physical and genetic length of chromosomes and chromosome-wide recombination rates

| Chromosome    | cM     | Mb    | cM/Mb |
|---------------|--------|-------|-------|
| 1             | 613.1  | 25.7  | 23.9  |
| 2             | 339.0  | 13.8  | 24.6  |
| 3             | 287.2  | 11.9  | 24.1  |
| 4             | 297.4  | 10.8  | 27.6  |
| 5             | 273.8  | 12.8  | 21.4  |
| 6             | 311.4  | 14.7  | 21.2  |
| 7             | 240.6  | 10.3  | 23.4  |
| 8             | 240.2  | 11.5  | 20.9  |
| 9             | 232.3  | 9.5   | 24.6  |
| 10            | 253.1  | 10.2  | 24.8  |
| 11            | 248.7  | 12.2  | 20.3  |
| 12            | 240.1  | 11.3  | 21.2  |
| 13            | 214.4  | 9.2   | 23.4  |
| 14            | 220.6  | 9.0   | 24.6  |
| 15            | 195.8  | 8.9   | 22.1  |
| 16            | 140.7  | 6.2   | 22.8  |
| Total/average | 3735.3 | 162.1 | 23.2  |

Chromosome 1 is a metacentric chromosome. Twelve of the remaining ones are acrocentric chromosomes while two of them are metacentric chromosomes (Beye and Moritz 1995). Despite these differences in structure, the chromosomes all share a similar average recombination rate.

consequences for defining genomic regions of naturally occurring traits by co-segregation or association studies.

### Chromosome parameters

Recombination rate is not related to the physical size of chromosomes. Little variation in rate is seen among the 16 honey bee chromosomes (Table 2). Chromosome 1 is twice the size of other chromosomes (25.7 Mb) and has an intermediate rate (23.9 cM/Mb), while the medium-sized chromosomes 4 (10.8 Mb) and 8 (11.5 Mb) have the highest and lowest rates, 27.6 cM/Mb and 20.9 cM/Mb, respectively. Chromosome 16 is about half the size of medium-sized chromosomes (6.2 Mb), yet has an intermediate recombination rate (22.8 cM/Mb). There is no correlation between chromosome size and chromosome recombination rate ( $r = -0.24$ ;  $n = 16$ ;  $P > 0.37$ ). In many species including mice, rats, and humans, recombination rates vary substantially between chromosomes and there is a strong negative relationship between chromosome size and chromosome recombination rate (Hunt and Page Jr. 1995; Jensen-Seaman et al. 2004). Honey bees do not show this relationship. Moreover, honey bees show no relationship between recombination and position of the centromere (Table 2). Metacentric, acrocentric, and submetacentric chromosomes (Beye and Moritz 1995) all have the same average recombination rate. The rather uniform average frequency of chiasmata per chromosome pair (average = 1.56) over a wide variety of taxa suggests (Otto and Lenormand 2002) that recombination plays an important role in stabilizing pairs of homologous chromosomes during meiosis (Baker et al. 1976). However, bees exhibit substantially more recombination events per chromosome per meiosis (5.7) compared with the average.

Recombination rates are reduced near the centromere (Table 3) but increase toward the telomere, a finding consistent with previous studies of human, mouse, rat, *C. elegans*, and *D. melanogaster*, which display comparatively lower genome-wide recombination (Hey and Kliman 2002; Jensen-Seaman et al. 2004; Prachumwat et al. 2004). Scatter plots of local recombination rate versus physical distance for each of the 16 chromosomes (Supplemental Fig. 1) support the above notion of a relationship between local recombination activity and chromosome position.

**Table 3.** Non-parametric correlation (Spearman's  $\rho$ ) of recombination rates with chromosome and sequence parameters that were estimated from 125-kb sequence windows

| Parameter                              | Relationship           | $r$    | $P$    |
|--|------------------------|--------|--------|
| Distance to the telomere               | Positive               | 0.2    | <0.001 |
| GC content                             | Positive               | 0.43   | <0.001 |
| Proportion of simple repeats           | Positive               | 0.19   | <0.001 |
| Proportion of low-complexity sequences | Negative               | -0.21  | <0.001 |
| Number of genes                        |                        | 0.01   | >0.7   |
| Distance between genes                 | Positive               | 0.13   | <0.014 |
| Proportion of most diverged genes      |                        | -0.01  | >0.7   |
| Intron length                          |                        | -0.065 | =0.19  |
| Exon length                            |                        | 0.079  | =0.1   |
| Transposons                            | Too few to be analyzed |        |        |

Moreover, the scatter plots show the high heterogeneity of recombination rate across these chromosomes.

### Sequence parameters

In order to shed light on the factors affecting the patterns of recombination, we looked for correlations between local recombination rates and sequence and gene-specific parameters (Table 3).

Among the sequence variables, GC content, the proportion of simple repeats, and the proportion of low-complexity sequences display a correlation with recombination, with GC content being the strongest correlate. Regions of higher recombination have higher GC content and exhibit more simple repeats, but have a lower proportion of low-complexity sequences that are more often found in AT-rich, low-recombining regions. GC content, the proportion of simple repeats, and the proportion of low-complexity sequences all co-vary. Partial correlation analysis indicates that the relationship between recombination and GC content still holds even if the proportion of simple repeats ( $r = 0.37$ ,  $P < 0.01$ ) and low-complexity sequences ( $r = 0.32$ ,  $P < 0.01$ ) are controlled for. This relationship is consistent with that found in the fly, mouse, and human (Hey and Kliman 2002; Jensen-Seaman et al. 2004). Another correlate of recombination is the fraction of windows containing simple repeats. The proportion of simple repeats is also associated with higher recombination in the human, rat, and mouse genomes (Yu et al. 2001; Jensen-Seaman et al. 2004).

### Gene parameters

Among gene parameters, the distance between genes positively correlates with recombination rate (Table 3). However, no correlation was found with gene density. Thus, genes in regions of low recombination appear to be on average larger. This could be due to an increased intron length, an increase in the length of coding regions, or both. Intron length displays a negative trend with recombination rate while exon length shows the opposite, but both relationships are nonsignificant (Table 3). Thus, larger genes in regions of low recombination may be due to larger introns. A negative correlation between intron length and recombination rate has been reported for *Drosophila* and humans (Comeron and Kreitman 2000, 2002). However, analysis of intron length in *Caenorhabditis* showed an opposite pattern, mainly due to an increased presence of transposable element insertions in regions of high recombination (Prachumwat et al. 2004). Although the genes of the bee genome are confident pre-

dictions, future annotations will improve the analysis and reduce the variance created by incorrect intron and exon definitions.

Too few transposons are present (11 that map into scaffolds) (Honey Bee Genome Sequencing Consortium 2006) to adequately assess any putative effect on recombination.

To assess the relationship of recombination with evolutionary rate on a broad scale, the best homology score of an annotated gene to the nonredundant protein database at the National Center for Biotechnology Information (NCBI) was taken as an estimate of divergence. The genes showing the greatest divergence (top third) demonstrated no pattern with respect to recombination rate (Table 3), suggesting that recombination rate and evolutionary rates of this class of genes (typically representing evolutionary distances of >300 million years [Myr]) are not correlated.

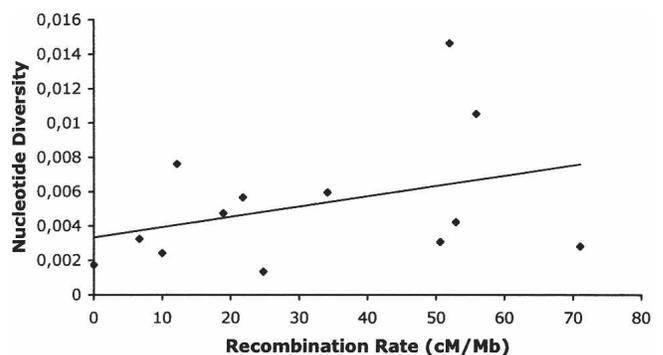
### Nucleotide polymorphism

Next, we investigated the relationship between nucleotide diversity and recombination in a natural population by surveying polymorphism at loci that vary by nearly two orders of magnitude in their local recombination rate. There is a positive trend between nucleotide diversity and recombination at 13 loci (introns) spanning rates of recombination from zero to >70 cM/Mb, but no significant correlation (Fig. 1;  $r = 0.32$ ,  $P = 0.28$ ). Divergence between *A. mellifera* and a closely related outgroup, *A. cerana*, at these same loci (average = 6.4%) displays no correlation ( $r = 0.06$ ,  $P = 0.86$ ).

### Discussion

The analysis of the honey bee genome shows that the high recombination rate is a genome-wide phenomenon and is not restricted to certain regions of chromosomes or specific chromosomes (Tables 2, 3). The recombination rate, however, varies substantially more (by two orders of magnitude) than that in humans, *D. melanogaster*, and *C. elegans* (Table 1).

The sequence analysis shows that GC content, simple repeats, and low-complexity sequences are associated with recombination rate. The relationship between GC content and recombination is consistent with that found in the fly, mouse, and human (Hey and Kliman 2002; Jensen-Seaman et al. 2004), suggesting a common cause or consequence of recombination. The basis of the relationship between GC content and recombination rate remains unresolved, but accumulating evidence suggests that recombination leads to GC-biased gene conversion (Birdsell 2002; Montoya-Burgos et al. 2003). Bees have a lower genome-

**Figure 1.** Scatter plot of nucleotide diversity,  $\theta$ , vs. recombination rate (Spearman's  $\rho = 0.324$ ;  $P = 0.28$ ).

wide GC content relative to *Drosophila* and *Anopheles* (The Honey Bee Genome Sequencing Consortium 2006). Higher GC content correlates with higher recombination, thus the lower GC content is unlikely to be responsible for higher recombination in the bee. The relationship between the proportion of simple repeats and higher recombination rate in the bee genome has been documented in several mammalian species as well (Yu et al. 2001; Jensen-Seaman et al. 2004), with the exception of some sequence-specific repeats. Simple repeats could mechanistically facilitate recombination, or, alternatively, recombination itself may increase their abundance by non-homologous recombination.

Too few transposons could be included in the analysis to assess their effect on recombination. Although transposons are underrepresented in this analysis (they are less often assembled into larger sequences and are thus more often excluded in the recombination analysis), there are very few transposons in the honey bee genome as a whole when compared with other genomes (The Honey Bee Genome Sequencing Consortium 2006). Almost all transposons identified in the bee genome are members of the *mariner* family, with ~360 copies in the genome. Other common types of transposons and retrotransposons are largely absent from the honey bee genome. It is possible that the high recombination rate enables a more efficient removal of deleterious insertions. Alternatively, deleterious insertions that are recessive in nature may be more efficiently purged when exposed to selection in haploid males, though these two explanations are not mutually exclusive.

We find no relationship between recombination and the evolutionary rate of proteins. Previous studies have shown that protein adaptation is influenced by the local recombination rate (Betancourt and Presgraves 2002; Presgraves 2005). However, these studies compare close relatives of *Drosophila*. In our analysis the evolutionary distances are likely too large (evolutionary distances of >300 Myr) to detect such a pattern (i.e., small differences in protein evolution that have occurred in the last 60 Myr in which the high recombination has evolved are masked). Comparative analysis suggests that the high recombination rate of the honey bee has evolved over the last 60 Myr after the split from a common ancestor of the bumble bee (see below).

On average we find 5.7 chiasmata per chromosome pair during meiosis, setting the honey bee genome apart from previously studied, low-recombining genomes. The constant number of 1.56 chiasmata over a wide range of taxa suggests that recombination plays an important role in stabilizing pairs of chromosomes during meiosis (Baker et al. 1976). It is notable that the 5.7 chiasmata per chromosome is the lower bound estimate of the honey bee genome as not all chiasmata lead to DNA exchange. This difference extends to closer relatives such as the bumble bee (1.2) and fruit fly (1.4) (Supplemental Table 1). The substantial increase above 1.56 chiasmata suggests that at least a portion of this activity evolved independently of a possible stabilizing role of chiasmata during homologous chromosome pairing.

Genes in regions of low recombination appear to be on average larger, which could, at least partially, be explained by an increased intron length, although this trend is not significant in our analysis. A negative correlation between intron size and recombination rate has been reported for *Drosophila* and humans (Comeron and Kreitman 2000, 2002). Carvalho and Clark (1999) posit that insertions in introns are slightly deleterious, and the efficacy of selection to remove these deleterious insertions is reduced in regions of low recombination. Alternatively, Comeron

and Kreitman (2000) argue that longer intervening sequences such as introns may serve as targets of selection (so-called recombination “modifiers”) that increase the frequency of recombination (Otto and Barton 1997). This finding suggests that similar forces could operate in the honey bee genome even in the presence of an order of magnitude higher than rate of recombination.

In *Drosophila* and various other organisms, nucleotide diversity is reduced in regions of low recombination (Begun and Aquadro 1992; Nachman et al. 1998; Stephan and Langley 1998). Positive selection for rare beneficial mutations resulting in genetic hitchhiking (Smith and Haigh 1974) or negative selection against frequent deleterious mutations causing background selection (Charlesworth et al. 1993) are both expected to reduce the level of linked, neutral polymorphism. These effects are expected to be more extreme in regions of reduced recombination. Although a positive trend exists among our surveyed loci, there is no significant correlation. This may be expected given that few loci from regions of very low recombination were included (these regions are also underrepresented in the genome sequence). Future work including more such regions may shed light on the influence of selection on linked neutral variation in a haplodiploid system.

A survey of genome-wide recombination rates in numerous multicellular social and nonsocial animals indicates (Supplemental Table 1) that the high recombination is not a general feature of the hymenopteran lineage as a whole or a result of its haplodiploid mode of reproduction. The bumble bee (*Bombus*) (Gadau et al. 2001) and the parasitic wasp (*Nasonia*) (Gadau et al. 1999) have very similar recombination rates as the *D. melanogaster*, human, and *C. elegans* (1–4 cM/Mb). The observation that the harvester ant (*Pogonomyrmex*), which is phylogenetically more distant to the bee complex (bumble bee, honey bee), has a genome-wide recombination rate of 14 cM/Mb (Sirviö et al. 2006) suggests that high recombination has evolved independently and more recently in ants and bees. This also excludes the possibility that breeding of bees (which is not strong in honey bees) has substantially influenced recombination rates (Burt and Bell 1987; Korol and Iliadi 1994).

Taken together, these results suggest that the high recombination of the honey bee is a genome-wide phenomenon and cannot be explained by solely mechanistic reasons. The strong relationship of recombination rate to GC content and number of simple repeats is consistent with observations in low-recombining genomes. Thus, these findings do not set the high-recombining bee genome apart from previous analyzed genomes. Over a wide range of taxa, a single crossover is sufficient to stabilize a bivalent, often two for metacentric chromosomes (Baker et al. 1976). Many organisms show a total interference (a single crossover per arm) and display normal meiosis. The number of chiasmata in the bee is substantially above what is mechanistically required to stabilize chromosome pairs during meiosis. The idea that honey bee chromosomes are inherently more unstable (for unknown reasons), and are thus necessarily stabilized by more chiasmata per chromosome arm (or per unit of chromosome length), is not supported by our analysis. We find no relationship between recombination and chromosome length or the number of chromosome arms per chromosome (Table 2).

To further complete our understanding of the high recombination rate, studies of nucleotide polymorphism on a more genome-wide scale and analysis of protein adaptation (Betancourt and Presgraves 2002) over shorter evolutionary distances (represented by closely related species) are necessary. These stud-

ies could provide further evidence that the efficacy of selection is limited even in the presence of a high recombination rate. Recent advances in evolutionary theories on recombination have predicted that genetic drift is a key force that broadens conditions for the evolution of higher recombination (Otto and Lenormand 2002; Barton and Otto 2005). Recombination improves the efficacy of selection in the presence of genetic drift. The magnitude of genetic drift is dependent on the effective size of a population. The honey bee has a small breeding population and an increased recombination rate, consistent with theoretical expectations. The number of colonies that contribute to the pool of males available to mate with honey bee queens is estimated to be no more than 238 (Baudry et al. 1998), resulting in an estimated effective population size of ~500. Even if the estimate does not represent the effective population size of the bee, but rather the size of the local population (deme), this would suggest a highly spatially structured population, which also influences recombination (Barton and Otto 2005). Alternatively or in addition, selective forces could operate stronger under social complexity. Higher levels of genetic variation and genetic associations could slow the spread of parasites and pathogens in colonies or provide the basis for more task specialization, leading to an increase in colony performance and fitness (Gadau et al. 2000). The large differences in recombination within the annotated and sequenced honey bee genome and across various related organisms represent splendid evolutionary examples to test theoretical predictions of the evolution of recombination rates.

## Methods

### Recombination rates

Each marker sequence was assigned to the sequence assembly version 2 that includes the fully annotated genome for gene parameter analysis (<http://www.ncbi.nlm.nih.gov/genome/guide/bee/>). The derivatives of the three- to five-term polynomials describing the relationship between the genetic and physical maps of markers were estimated in 125-kb ( $n = 615$ ) and 250-kb ( $n = 252$ ) windows. The model of higher-order polynomials adjusts for local differences in recombination over several tens of kilobases of distance that characterize at least some parts of the honey bee genome (Beye et al. 1999). When only two markers were anchored to the sequence ( $n = 53$ ), a straight line was calculated. Plots combine genetic map distances of two revised maps, the "Solignac" and the "Hunt" map, on which the estimates were obtained. The "Solignac" genetic map is based on the analysis of microsatellite markers through 50–200 meioses from two queens (both hybrids between two honey bee subspecies). The orientation of the chromosomes (centromere to telomere region) was previously determined using half-tetrad analysis, made possible by the central fusion during meiosis of the lytokous workers of the Cape bees (Baudry et al. 2004). We have used the map version Solignac\_1700 ([http://www.ncbi.nlm.nih.gov/mapview/map\\_search.cgi?taxid=7460](http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=7460)) recently extended to 2000 markers (4553 cM total map size). The average distance between two adjacent markers is 2.2 cM (CI: 0.55, 5.04); only 18 distances are longer than 10 cM, but they do not exceed 20 cM, most of them being within—not between—scaffolds. The maps were separately established for each queen, were both saturated, and provide the same general structure. This map was also used as an aid in assembling the genomic sequence into scaffolds. The "Hunt" linkage map is constructed using 1152 markers (4996 cM total map size), consisting of about two-thirds AFLP, one-third RAPD, plus 45 STS, and 48 microsatellite markers. MapManager-

QTX software was initially used to make a preliminary linkage map and to test the quality of markers by testing for map expansion and scoring errors for each marker. JoinMap software was used to determine the final orders and distances of the markers because it uses multipoint linkage algorithms rather than two-point algorithms. A total of 465 AFLP marker fragments were sequenced and verified as being correct if their linkage map position was congruent with the genome assembly. Results of a BLASTn search of the genome assembly were used to position markers in the assembly. In some cases, markers aligned with sequence scaffolds that had not been assembled into the 16 chromosomes. In these latter cases, marker clones were only confirmed as correct if markers that mapped close to each other on the linkage map hit the same sequence scaffold. Recombination rates for individual chromosomes were calculated by dividing the genetic length (cM) of the denser Solignac map by the sequence length (Mb) between the first and last marker placed on each chromosome. Recombination rates of *D. melanogaster* and *C. elegans* were obtained from 250-kb window analysis. Derivatives of the six-term polynomials (one for each chromosome) of genetic and physical map relationship were estimated in 250-kb windows. Original data sets were used (Marais et al. 2001; Hey and Kliman 2002).

### Chromosome and sequence parameters for 125-kb windows

For each window we calculated the proportional distance from the centromere region as defined by the "Solignac" map of the chromosome to the center of the window. The proportion of simple repeats (di- to pentameric and some hexameric repeats larger than 20 bp), G or C and low-complexity sequences (100-bp sequence that contains >87% AT or >89% GC or a 30-bp sequence that contains 29 A/T [or GC]) were calculated as the relative fraction of this sequence in the window by the program RepeatMasker (<http://www.repeatmasker.org>). As an estimator of gene divergence we used *E*-values resulting from a tBLASTx search of all Gnomon-predicted genes against the GenBank nr database (<http://www.ncbi.nlm.nih.gov/genome/guide/bee/>). The *E*-value gives an indication of the statistical significance of a given pairwise alignment. Hence, the higher the *E*-value, the less significant the hit, and the more diverged is the gene in question. The proportion of most diverged genes was obtained from the number of genes with a higher *E*-value than the 67th percentile of the genome-wide *E*-value distribution, divided by the total number of genes that were annotated in that window. The 67th percentile of gene annotations predicted by Gnomon was  $3.5 \times 10^{-23}$  (<http://www.ncbi.nlm.nih.gov/genome/guide/bee/AmeGnomon.html>). Genes without any BLAST hits were included in this analysis and were assigned the value of 1. Distance between genes is the average distance between annotated genes (ab initio models) within each window. This estimate is based on distances between coding regions, as 5' and 3' untranslated regions are typically not included in the annotation process. The number of genes includes the number of genes within a window. Genes exceeding the size of the window were weighted as half a gene in the calculation. Average intron length was calculated for each window. Introns were taken from Genomon gene models from the NCBI bee map viewer Web site (Build 2.1) that are displayed as GenesOnSequence. Introns exceeding the size of windows were weighted as half an intron in the calculation. A rough estimate of the average exon length per window was obtained on the basis of intron length and the distances between genes.

### Sequence and polymorphism analysis

Fragments of intronic sequences were amplified using standard protocols (Hasselmann et al. 2001) from a sample of 10 chromo-

somes (haploid drones) from a population of *Apis mellifera* sampled in Davis, CA. In addition, sequences were amplified from a single *Apis cerana* chromosome as an outgroup. Oligonucleotides were designed to amplify intronic sequences (sequences of oligonucleotides are available online as Supplemental data). Introns were assigned to sequences that have, to the left and to the right end, an EST match of a single-sequence read (<http://www.ncbi.nlm.nih.gov/genome/guide/bee/>). In addition, these sequences were checked by bioinformatic tools to test whether an obvious open reading frame could be deduced. The DnaSP program version 4.0 (Rozas et al. 2003) was used for inter- and intraspecific analyses. Nucleotide diversity,  $\theta$ , was estimated according to Watterson (1975). Sequences were initially aligned using ClustalX (Thompson et al. 1997) and then manually edited by eye. Gaps were excluded from the analysis. Amplified fragments were resolved on 1.3% agarose gels, eluted, and cloned into the pGEM-T vector (Promega). Sequencing was performed by MWG Biotech AG (Ebersberg). Local recombination rates of the 250-kb windows were assigned to these loci to improve confidence intervals.

## Acknowledgments

We thank Stefan Scroka, Marion Müller-Bork, Didina David, and Michel Dominique Vautrin for technical support and John Parsch and three anonymous reviewers for helpful comments on an earlier version of the manuscript. We thank the Baylor College of Medicine Human Genome Sequencing Center for making the *Apis mellifera* genome sequence publicly available prior to publication. This work was supported by DFG grants BE 2194/5, BE 2194/6, and DBU grant 21666 to M.B., NSF grants IBN 0110842 to G.H., and IBN 0090482 and IBN 0076811 to R.E.P.

## References

- Baker, B.S., Carpenter, A.T., Esposito, M.S., Esposito, R.E., and Sandler, L. 1976. The genetic control of meiosis. *Annu. Rev. Genet.* **10**: 53–134.
- Barton, N.H. and Otto, S.P. 2005. Evolution of recombination due to random drift. *Genetics* **169**: 2353–2370.
- Baudry, E., Solignac, M., Garnery, L., Gries, M., Cornuet, J.-M., and Koeniger, N. 1998. Relatedness among honeybees (*Apis mellifera*) of a drone congregation. *Proc. R. Soc. Lond. B. Biol. Sci.* **265**: 2009–2014.
- Baudry, E., Kryger, P., Allsopp, M., Koeniger, N., Vautrin, D., Mougél, F., Cornuet, J.M., and Solignac, M. 2004. Whole-genome scan in thelytokous-laying workers of the cape honeybee (*Apis mellifera capensis*): Central fusion, reduced recombination rates and centromere mapping using half-tetrad analysis. *Genetics* **167**: 243–252.
- Begun, D.J. and Aquadro, C.F. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519–520.
- Betancourt, A.J. and Presgraves, D.C. 2002. Linkage limits the power of natural selection in *Drosophila*. *Proc. Natl. Acad. Sci.* **99**: 13616–13620.
- Beye, M. and Moritz, R.F.A. 1995. Characterization of honeybee (*Apis mellifera* L.) chromosomes using repetitive DNA probes and fluorescence in situ hybridization. *J. Hered.* **86**: 145–150.
- Beye, M., Hunt, G.J., Page, R.E., Fondrk, M.K., Grohmann, L., and Moritz, R.F.A. 1999. Unusually high recombination rate detected in the sex locus region of the honey bee (*Apis mellifera*). *Genetics* **153**: 1701–1708.
- Birdsell, J.A. 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Mol. Biol. Evol.* **19**: 1181–1197.
- Burt, A. and Bell, G. 1987. Mammalian chiasma frequencies as a test of two theories of recombination. *Nature* **326**: 803–805.
- Carvalho, A.B. and Clark, A.G. 1999. Intron size and natural selection. *Nature* **401**: 344.
- Charlesworth, B., Morgan, M.T., and Charlesworth, D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- Cameron, J.M. and Kreitman, M. 2000. The correlation between intron length and recombination in *Drosophila*. Dynamic equilibrium between mutational and selective forces. *Genetics* **156**: 1175–1190.
- Cameron, J.M. and Kreitman, M. 2002. Population, evolutionary and genomic consequences of interference selection. *Genetics* **161**: 389–410.
- Gadau, J., Page Jr., R.E., and Werren, J.H. 1999. Mapping of hybrid incompatibility loci in *Nasonia*. *Genetics* **153**: 1731–1741.
- Gadau, J., Page Jr., R.E., Werren, J.H., and Schmid-Hempel, P. 2000. Genome organization and social evolution in Hymenoptera. *Naturwissenschaften* **87**: 87–89.
- Gadau, J., Gerloff, C.U., Kruger, N., Chan, H., Schmid-Hempel, P., Wille, A., and Page, R.E. 2001. A linkage analysis of sex determination in *Bombus terrestris* (L.) (Hymenoptera: Apidae). *Heredity* **87**: 234–242.
- Hasselmann, M., Fondrk, M.K., Page, R.E., and Beye, M. 2001. Fine scale mapping in the sex locus region of the honey bee (*Apis mellifera*). *Insect Mol. Biol.* **10**: 605–608.
- Hey, J. and Kliman, R.M. 2002. Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics* **160**: 595–608.
- Hill, W.G. and Robertson, A. 1966. The effect of linkage on limits to artificial selection. *Genet. Res.* **8**: 269–294.
- The Honey Bee Genome Sequencing Consortium. 2006. Insights into social insects from the genome of the honey bee *Apis mellifera*. *Nature* (in press).
- Hunt, G.J. and Page Jr., R.E. 1995. Linkage map of the honey bee, *Apis mellifera*, based on RAPD markers. *Genetics* **139**: 1371–1382.
- Jensen-Seaman, M.I., Furey, T.S., Payseur, B.A., Lu, Y., Roskin, K.M., Chen, C.F., Thomas, M.A., Haussler, D., and Jacob, H.J. 2004. Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res.* **14**: 528–538.
- Korol, A.B. and Iliadi, K.G. 1994. Increased recombination frequencies resulting from directional selection for geotaxis in *Drosophila*. *Heredity* **72**: 64–68.
- Marais, G., Mouchiroud, D., and Duret, L. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc. Natl. Acad. Sci.* **98**: 5688–5692.
- Michod, R.E. and Levin, B.R. 1988. *The evolution of sex: An examination of current ideas*. Sinauer Associates, Sunderland.
- Montoya-Burgos, J.I., Boursot, P., and Galtier, N. 2003. Recombination explains isochores in mammalian genomes. *Trends Genet.* **19**: 128–130.
- Nachman, M.W. 2002. Variation in recombination rate across the genome: Evidence and implications. *Curr. Opin. Genet. Dev.* **12**: 657–663.
- Nachman, M.W., Bauer, V.L., Crowell, S.L., and Aquadro, C.F. 1998. DNA variability and recombination rates at X-linked loci in humans. *Genetics* **150**: 1133–1141.
- Otto, S.P. and Barton, N.H. 1997. The evolution of recombination: Removing the limits to natural selection. *Genetics* **147**: 879–906.
- Otto, S.P. and Lenormand, T. 2002. Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* **3**: 252–261.
- Prachumwat, A., DeVincentis, L., and Palopoli, M.F. 2004. Intron size correlates positively with recombination rate in *Caenorhabditis elegans*. *Genetics* **166**: 1585–1590.
- Presgraves, D.C. 2005. Recombination enhances protein adaptation in *Drosophila melanogaster*. *Curr. Biol.* **15**: 1651–1656.
- Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- Sirviö, A., Gadau, J., Rueppell, O., Lamatsch, D., Boomsma, J.J., Pamilo, P., and Page, R.E.J. 2006. High recombination frequency creates genotypic diversity in colonies of the leaf-cutting ant *Acromyrmex echinatior*. *J. Evol. Biol.* **19**: 1475–1485.
- Smith, J.M. and Haigh, J. 1974. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**: 23–35.
- Solignac, M., Vautrin, D., Baudry, E., Mougél, F., Loiseau, A., and Cornuet, J.M. 2004. A microsatellite-based linkage map of the honeybee, *Apis mellifera* L. *Genetics* **167**: 253–262.
- Stephan, W. and Langley, C.H. 1998. DNA polymorphism in lycoposicon and crossing-over per physical length. *Genetics* **150**: 1585–1593.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL\_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876–4882.
- Watterson, G.A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- Yu, A., Zhao, C., Fan, Y., Jang, W., Mungall, A.J., Deloukas, P., Olsen, A., Doggett, N.A., Ghebranious, N., Broman, K.W., et al. 2001. Comparison of human genetic and sequence-based physical maps. *Nature* **409**: 951–953.

Received June 22, 2006; accepted in revised form August 18, 2006.