A Microsatellite-Based Linkage Map of the Honeybee, *Apis mellifera* L.

Michel Solignac,*[,1] Dominique Vautrin,* Emmanuelle Baudry,*[,2] Florence Mougel,* Anne Loiseau[†] and Jean-Marie Cornuet[†]

**Laboratoire Populations, Génétique et Evolution, Centre National de la Recherche Scientifique, F91198 Gif-sur-Yvette Cedex, France and [†]Centre de Biologie et de Gestion des Populations, F34988 Saint-Gely-du-Fesc Cedex, France*

Manuscript received October 7, 2003
Accepted for publication January 30, 2004

ABSTRACT

A linkage map for the honeybee (*Apis mellifera*) was constructed mainly from the progeny of two hybrid queens (*A. m. ligustica* × *A. m. mellifera*). A total of 541 loci were mapped; 474 were microsatellite loci; a few were additional bands produced during PCRs, one of the two rDNA loci (using ITS), the MDH locus, and three sex-linked markers (Q and FB loci and one RAPD band). Twenty-four linkage groups were estimated of which 5 were minute (between 7.1 and 22.8 cM) and 19 were major groups (>76.5 cM). The number of major linkage groups exceeded by three the number of chromosomes of the complement ($n = 16$). The sum of the lengths of all linkage groups amounts to 4061 cM to which must be added at least 320 cM to link groups in excess, making a total of at least 4381 cM. The length of the largest linkage group I was 630 cM. The average density of markers was 7.5 cM and the average resolution was about one marker every 300 kb. For most of the large groups, the centromeric region was determined genetically, as described in Baudry *et al.* (2004, accompanying article in this issue), using half-tetrad analysis of thelytokous parthenogens in which diploid restoration occurs through central fusion. Several cases of segregation distortion that appeared to result from deleterious recessives were discovered. A low positive interference was also detected.

THE development of DNA marker technologies has allowed us to envisage linkage mapping in species, which would otherwise remain inaccessible because the number of classical mutants (visible, biochemical, physiological, and enzyme polymorphism) was too low to permit genotyping of loci through the entire genome. Consequently, the genome of an increasing number of animal and plant species can now be mapped using DNA markers.

A rather large variety of markers exists, each possessing its own advantages and drawbacks. The development of arbitrary markers of the genome, such as randomly amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP), has not yielded reproducible markers, unless they are developed into sequence-tagged markers. On the other hand, design of specific markers such as microsatellites is time consuming and expensive (for the laboratory that develops them as well as for subsequent users) but they are easy to reproduce for the exploitation of the information contained in a genetic map and may be of great utility to anchor a genetic map on a physical map.

A large number of microsatellites have been developed mainly for the human genome, model species (mouse, rat, and derio), domestic animals (mammals, poultries, and salmonids), and a number of cultivated plants. Linkage maps in most of the other species have been developed mainly using arbitrary primers.

Among the invertebrates, almost all the maps have been developed for insect species, with the notable exception of the nematode *Caenorhabditis elegans*. *Drosophila melanogaster* (and a few other Drosophila species; O'Brien 1993) was the pioneer species for mapping (Morgan *et al.* 1925) and several hundred markers of all types are positioned (FlyBase Consortium 1998). In Lepidoptera, maps have been built for the silk worm *Bombyx mori* using restriction fragment length polymorphism (RFLP; Shi *et al.* 1995), RAPD (Promboon *et al.* 1995), RAPD with double primer pairs (Yasukochi 1998), and AFLP markers (Tan *et al.* 2001). Diptera, mainly mosquitoes, have been investigated more: *Anopheles gambiae* with microsatellites (Zheng *et al.* 1996), RAPD (Dimopoulos *et al.* 1996), and RFLP (Mori *et al.* 1999); *Aedes aegypti* using RFLP (Severson *et al.* 1993), cDNA/single-strand conformation polymorphism (SSCP; Fulton *et al.* 2001), and RAPD/SSCP (Antolin *et al.* 1996); *Culex pipens* using RFLP (Mori *et al.* 1999); and *A. albopictus* using RFLP (Severson *et al.* 1995) and RAPD-SSCP (Mutebi *et al.* 1997). In Coleoptera, a linkage map of *Tribolium castaneum* was developed using RAPD markers (Beeman and Brown 1999). In Hymenoptera, linkage maps were built for the honeybee *Apis mellifera* (Hunt and Page 1995) and several wasps, *Trichogramma brassicae* (Laurent *et al.* 1998) and Nasonia

[1]*Corresponding author:* Laboratoire Populations, Génétique et Evolution, Centre National de la Recherche Scientifique, F 91198 Gif-sur-Yvette Cedex, France. E-mail: solignac@pge.cnrs-gif.fr

[2]*Present address:* Evolution et Systématique Végétales, Université Paris Sud, 91405 Orsay, France.

(Gadau *et al.* 1999), using RAPD and for *Bracon hebetor* using RAPD-SSCP (Antolin *et al.* 1996). As can be seen from this nonexhaustive list, RAPDs are the favorite markers of the entomologists.

Thus, it is not surprising that the first genetic map for the honeybee was constructed using RAPDs (Hunt and Page 1995). It revealed the very large genetic (recombination) size of the honeybee genome and hence the large number of markers necessary to saturate it. A first screening for microsatellites in a genomic library, developed for population genetics purposes (Estoup *et al.* 1993), revealed a high density of these markers in the honeybee genome and prompted us to isolate additional markers for mapping. The markers polymorphic in *A. mellifera* have recently been published (Solignac *et al.* 2003) and they were used to genotype the progeny of several hybrid honeybee queens. In this article we present a map that is composed of a total of 541 loci distributed in 19 major and 5 minor linkage groups. Its estimated genetic length is 4380 cM, corresponding to an average density of one marker every 7.5 cM. Several analyses have been done in parallel on other meioses, including those of a queen belonging to another subspecies (nonhybrid African queen) and arrhenotokous workers (see Baudry *et al.* 2004, accompanying article this issue). These observations generalize the results obtained on hybrid queens to the whole species. In addition, the meioses of the thelytokous parthenogens of the Cape honeybee (*A. m. capensis*) have been used to genetically map the centromeric regions using half-tetrad analysis (see Baudry *et al.* 2004, accompanying article) and almost all the linkage groups of the microsatellite map are presented with the location of their centromeric regions.

MATERIALS AND METHODS

Biological material and DNAs

To maximize the number of informative loci, we have analyzed the meioses of queens that were hybrids between different subspecies of *A. mellifera*. These were the progeny of two F_1 hybrid queens named B and V (Figure 1). Virgin *A. m. ligustica* queens (the grandmothers) were imported from Italy (Bologna and Perugia) and instrumentally inseminated with the mixed sperm from 12 drones (the grandfathers) from four *A. m. mellifera* colonies from France (Charente). The 12 inseminated queens were allowed to found as many colonies. Adult hybrid queens were obtained in several of these colonies and we retained one colony that provided 9 queens. Every one of these 9 F_1 hybrid queens (the mothers) was instrumentally inseminated with the mixed sperm from two males (the fathers), one belonging to the subspecies *A. m. ligustica* and the other one to the subspecies *A. m. mellifera* (*i.e.*, to both parental subspecies). The pair of males was different for each inseminated queen. The female progeny (workers) of these backcrosses were collected for two of the nine families (queens B and V). The bodies of the males used for insemination (grandfathers and fathers) were preserved in alcohol.

The male (haploid) progeny of a third hybrid queen (M) were used, mainly to map the MDH locus. The queen was a "triple hybrid": her mother was the daughter of an *A. m. ligustica* queen artificially inseminated by an *A. m. caucasica* male and this hybrid was naturally fecundated by *A. m. mellifera* drones. Laying of unfertilized eggs was obtained after carbon dioxide treatment of the queen.

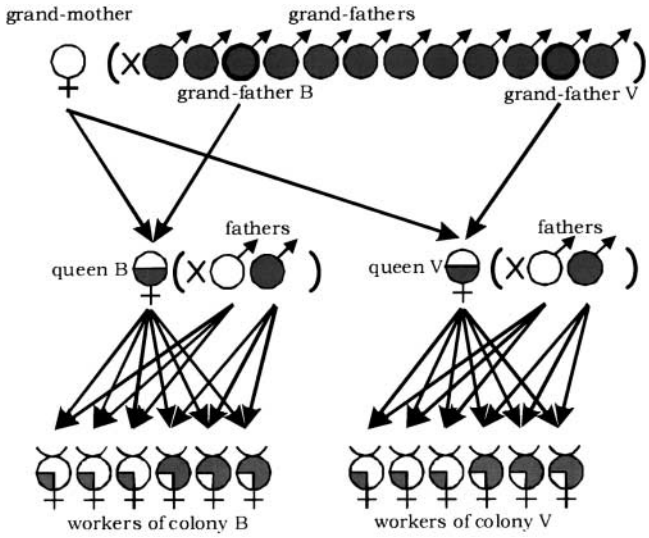

Figure 1.—Crosses performed to obtain workers for map construction. The grandmother was *Apis mellifera ligustica* (white) and it was instrumentally inseminated with the sperm admixture from 12 drones (grandfathers) of *A. m. mellifera* (black). Two hybrid queens (mothers, black/white), B and V, were backcrossed to two drones, one for each parental subspecies (fathers). The two families (workers) were composed of an admixture of two backcross subfamilies. The two grandfathers among the 12 drones were determined on the basis of their genetic profile (see text); their genotypes were used to establish the allelic phase of the B and V queens.

For each of the three families, DNA extracts from 96 workers (B and V) or drones (M) were prepared from the heads and later from the thoraces. DNAs were diluted to 1/40 and conserved at −20° in 96-well microtitration plates. We used only 92 workers for the B family. The remaining 4 wells were used to test DNA from three other Apis species and to genotype the B grandfather for establishing the allelic phase of the B hybrid queen. A total of 95 workers for the V family (the remaining sample being the V grandfather) and 96 drone progeny were analyzed. For the two families in which workers were studied (B and V), the two patrilines corresponding to the two different fathers are hereafter called subfamilies.

Markers

We used 552 markers described in Solignac *et al.* (2003), comprising mainly microsatellites and a few long repeated regions (ITS of the rDNA locus, royal jelly gene), and adopted the nomenclature proposed in this article: *Am* followed by a four-digit number.

A total of 82 markers were not mapped because the mothers were homozygous. Conversely, 67 additional bands (not corresponding to microsatellite loci) were mapped. In Figure 2, the numbers appearing in loci names correspond to bands showing the size of the sequenced allele. Additional bands that were associated with a specified microsatellite are labeled with an "*s*" (for "supplementary" bands) plus a number if several bands were generated. A total of 31 supplementary bands for 21 primer pairs were detected.

In addition, one RAPD marker (P13, Operon), the locus Q (Hunt and Page 1995) using *Sau*3A restriction of the PCR products, and the locus FB, all three known to be linked to the sex locus, were mapped. The malate dehydrogenase (MDH) locus was genotyped as described in Cornuet (1979) and was positioned on the linkage map only with the male progeny.

PCR and genotyping

Radioactive PCRs were performed following Solignac *et al.* (2003). PCRs were generally performed in multiplexes and the remaining loci were amplified separately but two or three PCR products were loaded on the same gel.

Initially, genotyping of the backcross progeny of queens B and V was established without preliminary tests of heterozygosity of the queens. Later, heterozygosity of the mothers was tested before analyzing the whole progeny: the tests were performed either on the queen itself for family V or on eight worker descendants for family B in which the queen was lost.

The grandfathers B and V were run in parallel with the workers to determine allelic phases of the B and V queens. Because the *ligustica* queens had been inseminated by a mixture of sperm from 12 *mellifera* males, it was first necessary to determine which ones fathered the hybrid queens. When a reasonable number of mapping gels had been completed, the grandfathers were easily determined on the basis of their genetic profiles: for each locus they must share their unique allele with their daughter. The fathers were used only to assist genotype determination and to confirm null alleles.

Statistics

Map construction: A series of preliminary checks and computations were performed. In two female progenies (B and V), the two patrilines had to be separated to determine the maternal allele. This was determined in the first 12 loci. Because of male haploidy, all workers of the same patriline share the same paternal alleles. The assignment of the workers to their respective patriline allowed designation of individual genotypes and tests for segregation distortion. The following checks were performed on every locus. Within a patriline, there are two possible diploid genotypes in the progeny of a heterozygous queen. Additional genotypes are due to either a mutation or a mistyping error. Two patrilines arising from the same queen must provide the same maternal genotype. Within each patriline, the two maternal alleles should be distributed equally among workers in the absence of segregation distortion or mistyping errors.

Once all detectable mistyping errors had been ruled out at individual loci, linkage of a locus was tested against previously mapped loci. The phase of the alleles was determined as each locus was entered. Eventually, the input file was edited by distinguishing the two maternal alleles (*e.g.*, A, grandpaternal; H, grandmaternal). This setting corresponds to the usual "F_2 backcross" data type analyzed by genetic mapping software.

The linkage map was built with Carthagene mapping software (Schiex and Gaspin 1997) version 0.5. This program computes genetic maps using a range of different algorithms. At any time, the best maps are recorded in a heap structure and any new map is compared to the previous maps of the heap. The different algorithms can be chained in a single command. Typically, the command that we used included the following:

i. Starting from each locus (on a linkage group), the algorithm gradually builds the best neighborhoods maps using a two-point LOD criterion (command *nicemapl*).
ii. Starting from two loci, the algorithm adds a new marker by testing all possible positions, until all markers are mapped. The criterion also uses the two-point LOD criterion and the best three orders (maps) are kept at each step (command *build*).
iii. A Tabu search (Glover 1989a,b) is performed on the multipoint LOD score criterion. Starting from the best map of the heap, it explores the *two-change* neighborhood of maps, *i.e.*, the set of maps obtained by inverting all possible subsections of the map (command *greedy*).
iv. A simulated annealing algorithm based on the multipoint LOD score is used. Two and three changes are randomly applied to the current map (command *annealing*).
v. Maps are next considered as individuals having a selective fitness proportional to their multipoint LOD score. Starting from an initial population, maps evolve along generations through recombination and mutation. Only the best-fit maps are conserved to produce the next generation (command *algogen*).
vi. Starting from the best map of the heap, the algorithm tests systematically all permutation of markers within a window of a given size (command *flips*).
vii. Starting from the best map of the heap, the algorithm takes one locus at a time and tests all its possible locations (command *polish*).

Linkage groups were established at the two-point LOD score threshold of 3.3.

Checkpoints during map construction: Additional precautions were taken to prevent errors in identification of loci and individuals and in the determination of genotypes.

Locus: The number given to the locus was confirmed by comparison of the mapping gels with the gels for the test of heterozygosity and the gels made for the completion or correction of data.

Individuals: PCR products of half-sisters from the two subfamilies were loaded on the gels' intermingling subfamilies but always in the same order. Once paternal alleles had been identified, this allowed for an easier detection of any shift when reading genotypes. When the two male parents had the same allele, the preceding control was inefficient and one or several series of individuals with characteristic genetic profiles were reamplified. A similar map location between the two patrilines implied that individuals were correctly assigned to the subfamily.

Genotypes: The films were read independently by at least two people. Personal programs were used to check the data for the presence of only two genotypes for each of the two subfamilies and for possible segregation distortion. Ambiguous or missing genotypes were systematically reamplified and run again in the presence of controls. Once the first version of the map had been constructed, every individual showing two recombination events surrounding a single locus was genotyped again for this locus.

Maps: The last control consisted of comparing maps constructed with the progenies of the B and V queens separately before combining them.

Chiasma interference: The occurrence of a crossover in one genetic region frequently decreases the probability of a concomitant crossover in an adjacent region. This phenomenon, called positive interference, has been observed in many species (Zhao *et al.* 1995b). Investigations of interference have classically compared observed frequencies of multiple recombination events in adjacent intervals to their expected frequency (Zhao *et al.* 1995b). However, due to the rarity of such events, these analyses require a huge number of meioses (*i.e.*, several thousands). We have therefore made use of the approach proposed by Broman *et al.* (2002) and Broman and

Weber (2000) that considers the distribution of the estimated distances between crossovers.

In this approach, it is important to identify all recombination events that occurred along a chromosome. We have therefore considered only the data from the B progeny, which were genotyped for the largest set of markers. All recombination events were identified and their positions were assumed to be the midpoint of the interval between the two flanking markers. We computed the intercrossover distance distribution for all linkage groups showing at least 40 intercrossover distances, *i.e.*, groups I–XI. We then fitted the observed distribution of the distance between crossovers to that expected under the gamma model with an integer parameter, using the correlation coefficient plot technique (Filliben 1975) with the software Dataplot (http://www.itl.nist.gov/div898/software/dataplot/).

The gamma model assumes that distances between chiasmata are independent and follow a gamma distribution with shape and rate parameters of ν and 2ν (see Zhao *et al.* 1995a for a review; Broman and Weber 2000). Several studies have found that the gamma model provides an excellent fit to recombination data (Zhao *et al.* 1995b; Lin and Speed 1996; Broman and Weber 2000; Lin *et al.* 2001; Broman *et al.* 2002). We focused on a special case of the gamma model where the shape parameter is $\nu = m + 1$ with m a nonnegative integer. This model, also called the chi-square model, explains well certain empirical observations concerning recombination and gene conversion (Foss *et al.* 1993).

Segregation distortion: In the three progenies, we observed the segregation of the maternal genes only. At each locus, the two alleles were expected to be represented in equal frequencies in the progeny. However, alleles at some loci segregated in proportions significantly different from the expected 1:1 ratio. This may be due to random type I errors, but this may also be due to specific segregation distortion loci (SDL). These SDL were detected by examining observed ratios in adjacent loci.

A graphical approach was used in which the segregation distortion measured by the chi-square criterion was plotted against the genetic distance on each linkage group. This was applied to only the first two progenies (of queens B and V) because the third (drone) progeny did not provide a sufficiently dense map to allow this analysis. The two subfamilies in each progeny were analyzed separately because an SDL paternal allele can interfere in different ways with each of the two maternal alleles.

An additional analysis was performed to detect potential interactions between loci exhibiting segregation distortions. This was done to distinguish between loci that have deleterious alleles and loci that have alleles that are deleterious only in association with alleles at another locus. Linkage does induce such an association, so that the analysis was reduced to all pairs of unlinked loci in which both loci shared significant segregation distortion. In these pairs, we tested whether or not the association of the maternal alleles at each locus was random through a homogeneity chi-square analysis.

RESULTS

Mapped markers: Among the 556 markers assayed, 474 principal loci as well as 67 supplementary bands were mapped. This was due to the high heterozygosity of the hybrid queens: $H = 0.75$ for queen B and 0.73 for queen V. The heterozygosity of queen M was low and the few DNA markers used with this family were analyzed to detect the linkage of a DNA marker with the MDH locus or to increase the power of statistics for loose or doubtful linkages.

The total number of individual genotypes collected per locus varied from 34 to 281, with an average of 132.2. The total number of genotypes established for this work was 71,648.

In Figure 2, a letter added to the locus name indicates the progeny genotyped for this locus: b for B queen (201 loci), v for V queen (48 loci), d for both (156 loci), m for males (5 loci), t for all (49 loci), c for B and M (9 loci), w for V and M (6 loci), *i.e.*, 415 for B, 259 for V, and 69 for M. The number of individuals used to connect some pairs of loci justifies the inclusion of large genetic distances observed without intervening markers, for instance, a distance of 46.0 cM on group I.

The B progeny were analyzed for all loci for which the queen was heterozygous. The queen V was also systematically analyzed at the beginning of the map construction. However, when the map began to become relatively dense, it was used only when the queen B was homozygous, when the genetic distance between consecutive markers was large, or when the marker was located at the tip of a linkage group or was still unlinked.

We reanalyzed 625 potential double-recombination events: 114 were found to be wrong. This proportion could seem enormous but among 71,648 genotypes, it represents only 0.15% mistyping.

Linkage groups and genetic length: The present genetic map has 24 linkage groups and consequently is unsaturated, the complement of the honeybee composed of only 16 chromosomes. During the construction of the map, the number of linkage groups decreased: with 297 markers, the map possessed 32 groups and 9 unlinked loci; with 440 markers, these numbers were, respectively, 27 and 5; and with the 541 markers included in the present map, they were 24 and 0. We estimate that several hundred additional markers would be necessary to saturate the map.

Among these 24 groups, 5 are very short (7.1–22.8 cM) and encompass 2–7 markers. The 19 other groups are longer (76.5–630.3 cM and 11–73 markers). There are no further statistical criteria for concatenating groups.

The length of this genetic map is 4061.2 cM, still higher than a previous estimate (3450 cM; Hunt and Page 1995). At least 320 cM must be added for the groups in excess. A few errors in genotyping can create a significant increase in the length of the map (Brzusto-

Figure 2.—Microsatellite genetic map for the honeybee. The map is composed of 541 markers and 24 linkage groups (I–XXIV), 19 major groups and 5 minute ones. The number of loci and the genetic length are indicated above every group. See text for nomenclature of markers. The Kosambi distance is indicated between two adjacent markers. The shaded section in the top part of most of the groups indicates the position of the centromeric region.


I
650.3 cM
73 loci
II
349.5 cM
42 loci
III
309.8 cM
46 loci
IV
286.9 cM
26 loci
V
242.7 cM
24 loci
VI
233.5 cM
46 loci
VII
225.6 cM
36 loci
VIII
223.6 cM
28 loci
IX
210.3 cM
26 loci

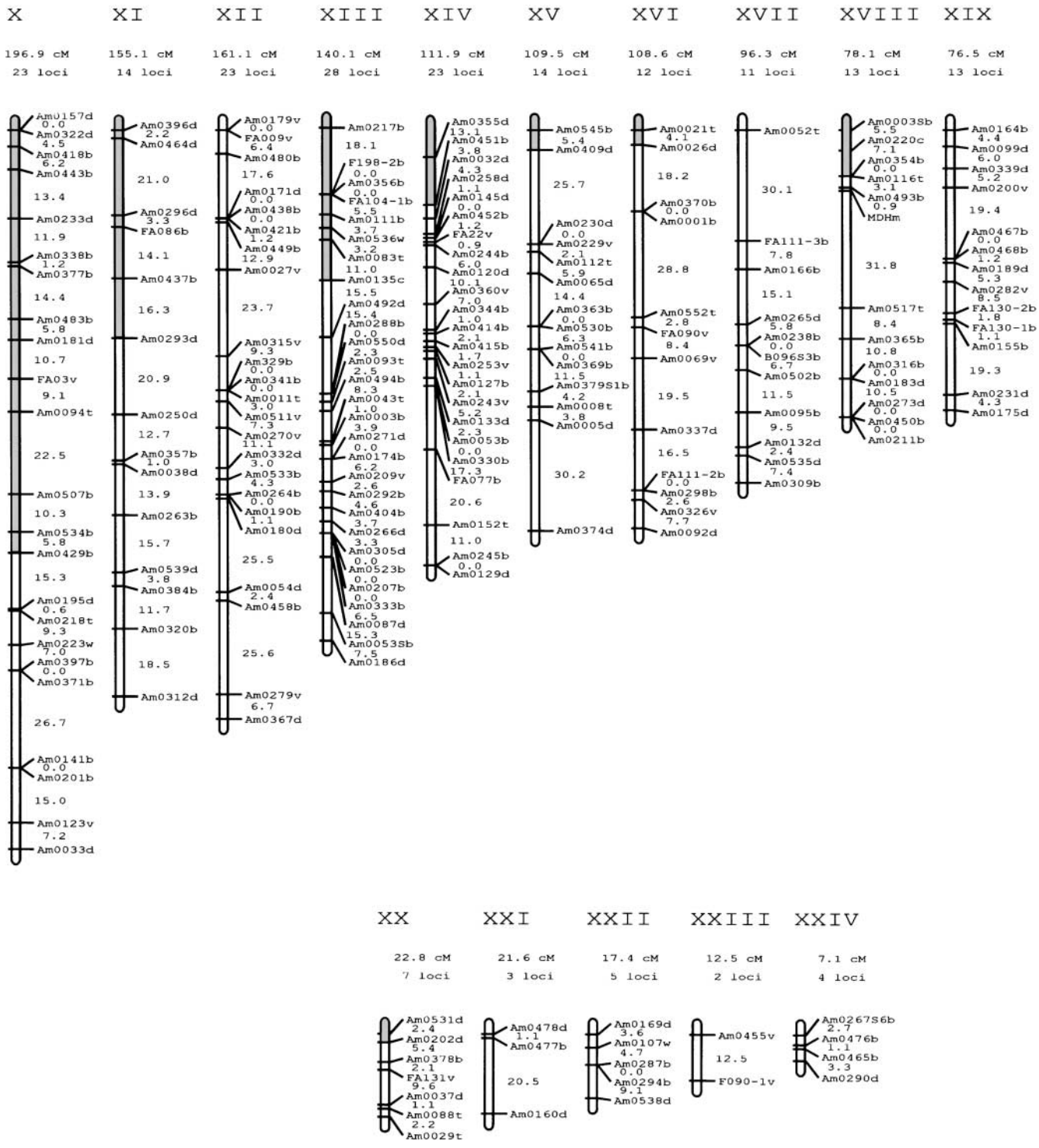

FIGURE 2.—*Continued.*

WICZ *et al.* 1993) but we have probably removed most of the mistypings.

The average density of markers on the map is about one every 7.5 cM. However, a higher density was observed for groups VI (one marker every 5.1 cM) and XIII (5.0 cM) whereas groups IV and XI have a lower density (11.0 and 11.1 cM, respectively). Moreover, a low or a high density characterizes different regions on the groups VIII and IX.

Centromere mapping: Centromeric regions were genetically mapped using half-tetrad analysis in thelytokous-laying workers (or pseudo-queens) of the subspecies *A. m. capensis* (Cape bees). Diploidy is restored through the fusion of the two central products of meiosis, producing in the presence of crossovers a gradient of homozygosity from the centromeres toward the telomeres (BAUDRY *et al.* 2004, accompanying article). These centromeric regions are shaded and placed at

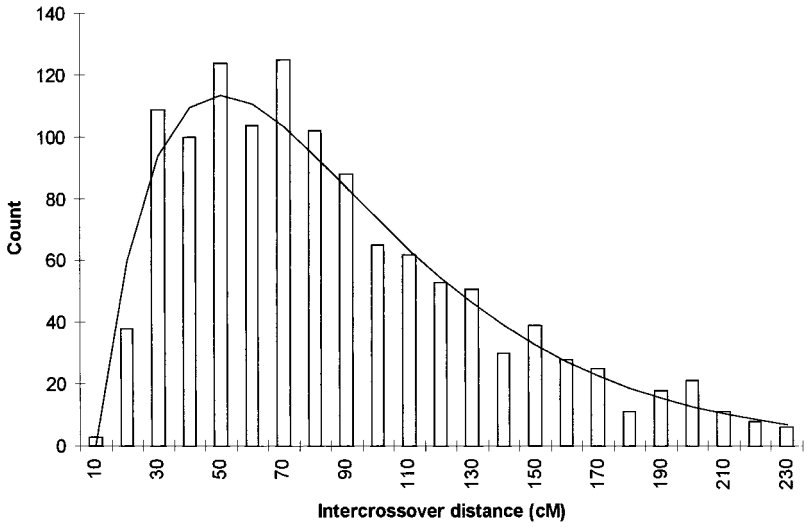

FIGURE 3.—Histogram showing the distribution of distances between crossovers for linkage groups I–XI (n = 1236). The curve corresponds to a gamma distribution with shape parameter of two and a scale parameter of four.

the top part of the group in Figure 2. The delimitation of centromeric regions depended on the density of markers available, heterozygosity of the pseudo-queens, and the number of chiasmata. Centromeres were highly localized in some linkage groups (groups I and II, for instance) but in other cases they covered a large part of the group (groups IX, X, and XIII). Some linkage groups may correspond to the extremities of chromosomes yet are unlinked to the arm to which they belong. These chromosome fragments obviously have no centromeric regions.

We have tried to place some additional molecular markers. Two repeated sequences, *Ava* (BEYE and MORITZ 1994) and *Alu* (TARÈS *et al.* 1993), have been characterized in the honeybee genome and BEYE and MORITZ (1995) have established their cytological distribution on the chromosomes. *Ava* is a marker of centromeric regions whereas *Alu* is mostly limited to the short arms of the subtelocentric elements. We have screened a bacterial artificial chromosome (BAC) library for clones containing *Ava* or *Alu* sequences and prepared microsatellites from these BACs. Unfortunately, the positive clones contained only small clusters of dispersed *Alu* or *Ava* sequences and did not occur in the expected location.

Crossover interference: We tested whether the distributions of intercrossover distances were homogeneous on linkage groups I–XI. The accuracy of the calculations was greatly improved by detecting and removing false double recombinants. A chi-square test of homogeneity indicated no significant difference among linkage groups (P = 0.26; χ^2 = 77.2; d.f. = 70) in spite of the moderate heterogeneity in marker density. We therefore pooled all data and fit the pooled distribution to a gamma model of interference. We found that the best fit was for a shape parameter of two. Figure 3 shows the histograms of the distances between crossovers for linkage groups I–XI and the curve for the gamma model with a shape parameter of two. A chi-square goodness-of-fit test showed that the observed pooled distribution does not deviate significantly from the gamma model (P = 0.24; χ^2 = 21.8; d.f. = 18). According to Foss *et al.*'s (1993) interpretation of this interference model, this indicates that in honeybee, crossovers need to be separated by one potential conversion event without associated crossover.

We have also used linkage group I, the only one corresponding to a metacentric chromosome, to determine whether the centromere plays a special role in the recombination process. For linkage group I, we have compared the distribution of intercrossover distances when the two arms of the chromosome are treated independently or together. We found no significant differences between the two distributions, suggesting that the level of interference across the chromosome is not influenced by the centromere.

Segregation distortion: We have observed several cases of very strong segregation distortion. All these were limited to one subfamily and their effect decreases to zero at ~50 cM from the peak (Figure 4).

We found no pairs of unlinked loci with segregation distortions and significant associations of maternal alleles, which would suggest that deleterious effect is produced by unlinked interacting loci. This suggests that segregation distortions result from specific loci having deleterious alleles.

Neomutations: We observed novel mutations on markers *Am*0043 [$(CT)_{37} \rightarrow (CT)_{38}$], *Am*0085 [$T_{27} \rightarrow T_{28}$], *Am*0129 [$(TA)_{23} \rightarrow (TA)_{22}$], *Am*0195 [$(CT)_{42} \rightarrow (CT)_{44}$], and *Am*0292 [$(TC)_{33} \rightarrow (TC)_{32}$]. Four mutations corresponded to the change of a single motif (mono- and dinucleotide) and one to the change of two motifs (dinucleotide). Three mutation events increased the size of the resulting allele and two decreased it. In all cases the mutations occurred in a large array of repeats. All of them were attributable to the gamete transmitted by the father. The observation of five mutations among 71,648 genotypes corresponds to a rate of $\sim 3.5 \times 10^{-5}$

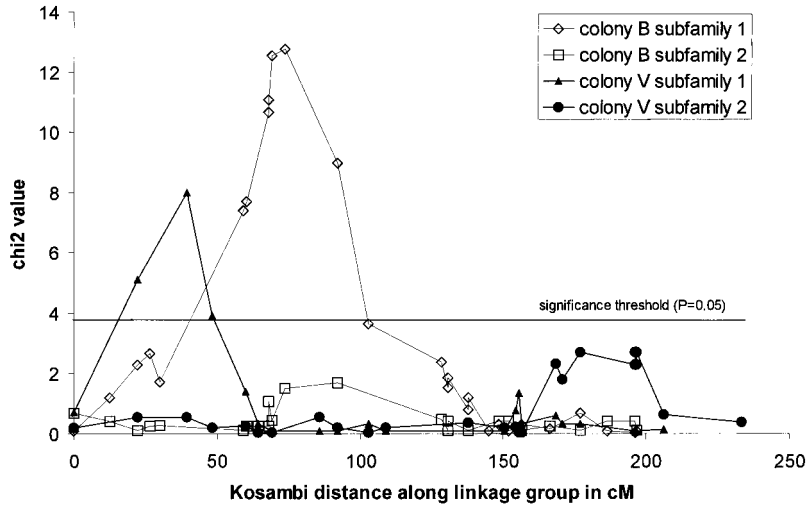

FIGURE 4.—An example of segregation distortion for linkage group VI. A peak of distortion is observed for subfamily 1 of family B and another one for subfamily 1 of family V. Distortion, measured by the χ^2 statistics, decreases regularly on both sides of the peak to insignificant values ($P > 0.05$) at a distance of ~50 cM.

(or more precisely to 7×10^{-5} for the males and 0 for the females), well in the range reported for other organisms (SCHUG *et al.* 1998).

Null alleles: The following statistics are limited to the principal loci themselves. A null allele was detected at 28 loci (5%). The B and V progenies allowed examination of eight haploid genomes (the diploid genome of the queen and the haploid genomes of the two drones, for each of the two families). Between 1 and 5 null gene copies were observed among these 8 copies, of which 18 occurred in a single subspecies. A total of 52 copies of null alleles have been observed for these 28 markers, 15 in *mellifera* and 37 in *ligustica*, an amount expected because the primers have been prepared from DNA extracts of the *mellifera* subspecies.

DISCUSSION

In spite of its biological importance in physiology, olfaction, and social behavior, genetic analysis in honeybees has been poorly developed. Mutant alleles are rather difficult to conserve and controlled crosses were not possible until the development of instrumental insemination. Enzyme polymorphisms are not abundant in Hymenoptera in general and in the honeybee in particular (PAMILO *et al.* 1978). For a long time, molecular investigations were limited to a determination of genome size (JORDAN and BROSEMER 1974), the sequence of a few genes and a few repeated elements (CRAIN *et al.* 1976; TARÈS *et al.* 1993; BEYE and MORITZ 1994), and a RAPD map (HUNT and PAGE 1995). This situation is rapidly changing, beginning with large-scale genomic analyses (EVANS and WHEELER 1999; WHITFIELD *et al.* 2002) and culminating with the sequence of the whole genome in progress at the National Center for Biotechnology Information (R. Gibbs and G. Weinstock). The development of a microsatellite linkage map will be of great use in future genomics in *Apis*.

Of 552 markers placed in the map, 153 have successfully amplified in three Apis species (*A. cerana*, *A. dorsata*, and *A. florea*; SOLIGNAC *et al.* 2003), opening the possibility of comparative synteny in the genus. They have also been assayed in five Hymenopteran species (the bumble bee *Bombus terrestris*, the ant *Gnamptogenys striatula*, the ichneumonid *Agrothereutes parvulus*, the chalcidian *Megastigmus rafni*, and the sawfly *Diprion pini*, unpublished results) and 54, 12, 14, 24, and 29 markers have amplified, respectively. However, preliminary tests have detected multiband profiles or lack of polymorphism. The only species for which some markers can be used is the bumblebee with ~10 polymorphic markers.

Genetic length: The length of our linkage map was 4061 cM, 18% longer than that of HUNT and PAGE (1995). Linkage group I alone is 650.3 cM, more than two times the total length of the *D. melanogaster* genome (~280 cM). An average of 13 chiasmata occur at each meiosis on the two arms of this chromosome.

This large linkage length has been confirmed in other meiotic contexts, *i.e.*, in queens belonging to *A. m. capensis* subspecies (BAUDRY *et al.* 2004, accompanying article). This observation suggests that the hybrid nature of the queens is not responsible for an increase in the linkage length of the map. In addition, recombination rates are similar in arrhenotokous *A. m. mellifera* workers (BAUDRY *et al.* 2004, accompanying article). Consequently, it can be considered as the "normal" recombination rate in the species.

From the genome size (178 Mb) and the genetic length of the map (4061 cM), it can be estimated that 1 cM is equivalent to ~44 kb, an amount to be compared to ~1 Mb in humans (DIB *et al.* 1996) and 2.5 Mb in mice (DIETRICH *et al.* 1996). Hence the average density of the map of 7.5 cM corresponds to only 300 kb between two consecutive markers, a favorable situation for positional cloning or to identify candidate genes once the physical map is completed.

Assignment of linkage groups to chromosomes: The number of chromosomes in the honeybee is $n = 16$

($2n = 32$). These chromosomes are small, the *C* value for the honeybee being 178 Mb, and the karyotype is highly symmetrical. Consequently, they are cytologically difficult to characterize and molecular markers are necessary to identify chromosomes individually (Beye and Moritz 1995). In the current map, the large linkage group (group I) corresponds to chromosome 1, the only metacentric chromosome and the only one in which the centromeric region maps in the middle. Linkage group IV corresponds to chromosome 8, carrying the sex locus. Finally, linkage group XVI where one ITS has been mapped corresponds to chromosome 8 or 11, each of which bears an rDNA locus.

Interference: Among the first 11 linkage groups of the honeybee genetic map, we found strong evidence for a low level of positive crossover interference. Our data fit well to a gamma model with shape parameter $\nu = 2$. For comparison, $\nu = 1$ under no interference and the estimated level of interference of the human, the Drosophila, and the mouse genomes are, respectively, $\nu = 4.3$, $\nu = 4.9$, and $\nu = 11.3$. The low value of the interference parameter observed in honeybee may be related to its high recombination rate. It has been proposed that interference is a biological mechanism, allowing each chromosome to have at least one chiasma, which is necessary for proper segregation (Egel 1995; Bascom-Slack *et al.* 1997; Moore and Orr-Weaver 1998; Broman *et al.* 2002). The total genetic length of the honeybee map indicates an average of about five chiasmata per chromosome. Broman *et al.* (2002) and Sym and Roeder (1994) have proposed that organisms with a high ratio of chiasmata to chromosome number should have low interference.

Zhao and Speed (1996) have shown that gamma models with $\nu = 1$ (no interference) and $\nu = 2.6$ are equivalent to Haldane and Kosambi map functions, respectively. We found a value of $\nu = 2$, which means that in honeybee, the appropriate map function is intermediate between the Haldane and the Kosambi functions, although much closer to the latter (see Figure 5), hence justifying its choice in this analysis.

Genetic load: Segregation distortion observed in several points of the map was limited to a single subfamily. All these distortions were independent. This indicates that in spite of the noticeable genetic divergence between the two subspecies used for crosses, we found no evidence for nascent Dobzhansky-Muller incompatibilities (Orr and Turelli 2001). Instead the distortion appears to be due to homozygosity for deleterious alleles. This means that, contrary to what is generally thought for Hymenoptera (Pamilo *et al.* 1978), the elimination of these deleterious alleles is far from perfect in haploid males, perhaps because those we have detected do not affect the male physiology.

Conclusion: The project of complete genome sequencing of the honeybee genome has convinced us that it would be useful to further increase the density of the present linkage map with the goal of saturating it. We plan to continue this work with microsatellite markers on the same individuals.

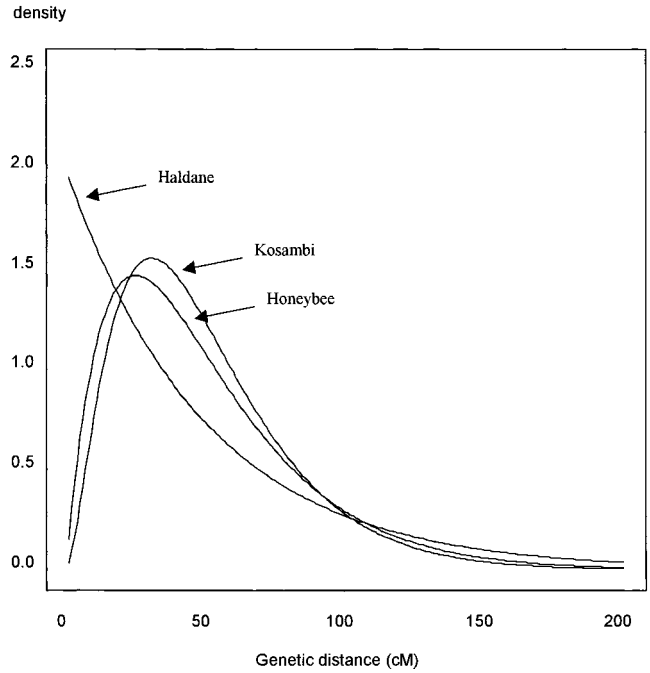

Figure 5.—Distributions of intercrossover distances under gamma models corresponding to Haldane and Kosambi distances and the gamma model.

The complete sequence of the genome, the large cDNA library from the brain, and the two genetic maps (the RAPD map and the present microsatellite ones), will probably promote original genetic investigations, particularly for complex behaviors in honeybees.

We thank, for their help in genotyping, Celia Chen, Philippe Dru, Lan Quan, Laure Fougeray, Sylvie Haze, Olivier Boutard, Robert Dorazi, Vanessa Rouaud, Julio Pereira, Maëlle Bilous, Stéphanie Gobin, Patricia Kries, Sylvain Brun, Claire Lions, and particularly Elsa Edline and Isabelle Leclainche. We are very grateful to anonymous referees for their helpful comments and suggestions on the manuscript. The *A. m. ligustica* queens were provided by Marcella Bernardini Battaglini (Istituto de Zootecnica generale, Università di Perugia) and Marco Lodesani (Istituto Nazionale di Apicoltura, Bologna). Jacques Kemp performed instrumental inseminations. Jean-François Odoux performed beekeeping experiments. Nikolas Koeniger provided the progeny of *capensis* laying workers and Per Kryger and Mike Allsopp provided samples of the *capensis* clone. Gérard Arnold, Lionel Garnery, Aurore Arnold, and Nadine Labet helped during the preparation of the biological material. Specimens of Hymenoptera were graciously provided by Claire Villemant, Chantal Poteaux, Jean-Yves Rasplus, and Jérôme Rousselet. Funding was provided by the Groupement de Recherches pour l'étude des Génomes and the Action Coordonnée Concertée (Sciences de la Vie 1).

LITERATURE CITED

Antolin, M. F., C. F. Bosio, J. Cotton, W. Sweeney, M. R. Strand *et al.*, 1996 Intensive linkage mapping in a wasp (*Bracon hebetor*) and a mosquito (*Aedes aegypti*) with single-strand conformation

polymorphism analysis of random amplified polymorphic DNA markers. Genetics **143:** 1727–1738.

Bascom-Slack, C., L. Ross and D. Dawson, 1997 Chiasmata, crossovers, and meiotic chromosome segregation. Adv. Genet. **35:** 253–284.

Baudry, E., P. Kryger, M. Allsopp, N. Koeniger, D. Vautrin *et al.*, 2004 Whole-genome scan in thelytokous-laying workers of the Cape honeybee (*Apis mellifera capensis*): central fusion, reduced recombination rates and centomere mapping using half-tetrad analysis. Genetics **167:** 243–252.

Beeman, R. W., and S. J. Brown, 1999 RAPD-based genetic linkage map of *Tribolium castaneum*. Genetics **153:** 333–338.

Beye, M., and R. F. A. Moritz, 1994 A centromere-specific probe for fluorescence in situ hybridization on chromosomes of *Apis mellifera*. Apidologie **25:** 322–326.

Beye, M., and R. F. A. Moritz, 1995 Characterization of honeybee (*Apis mellifera* L.) chromosomes using repetitive DNA probes and fluorescence in situ hybridization. J. Hered. **86:** 145–150.

Broman, K. W., and J. L. Weber, 2000 Characterization of human crossover interference. Am. J. Hum. Genet. **66:** 1911–1926.

Broman, K. W., L. B. Rowe, G. A. Churchill and K. Paigen, 2002 Crossover interference in the mouse. Genetics **160:** 1123–1131.

Brzustowicz, L. M., C. Merette, X. Xie, L. Townsend, T. C. Gilliam *et al.*, 1993 Molecular and statistical approaches to the detection and correction of errors in genotype databases. Am. J. Hum. Genet. **53:** 1137–1145.

Cornuet, J.-M., 1979 The MDH system in honeybees from Guadaloupe. J. Hered. **70:** 223–224.

Crain, W. R., E. H. Davidson and R. J. Britten, 1976 Contrasting patterns of DNA sequence arrangement in *Apis mellifera* (honeybee) and *Musca domestic*a (housefly). Chromosoma **59:** 1–12.

Dib, C., S. Faure, C. Fizames, D. Samson, N. Drouot *et al.*, 1996 A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature **380:** 152–154.

Dietrich, W., J. Miller, R. Steen, M. Merchan, D. Damron-Boles *et al.*, 1996 A comprehensive genetic map of the mouse genome. Nature **380:** 149–152.

Dimopoulos, G., L. Zheng, V. Kumar, A. della Torre, F. C. Kafatos *et al.*, 1996 Integrated genetic map of *Anopheles gambiae*: use of RAPD polymorphisms for genetic, cytogenetic and STS landmarks. Genetics **143:** 953–960.

Egel, R., 1995 The synaptonemal complex and the distribution of meiotic recombination events. Trends Genet. **11:** 206–208.

Estoup, A., M. Solignac, M. Harry and J.-M. Cornuet, 1993 Characterization of (GT)n and (CT)n microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. Nucleic Acids Res. **21:** 1427–1431.

Evans, J., and D. E. Wheeler, 1999 Differential gene expression between developing queens and workers in the honey bee, *Apis mellifera*. Proc. Natl. Acad. Sci. USA **96:** 5348–5350.

Filliben, J. J., 1975 The probability plot correlation coefficient test for normality. Technometrics **17:** 111–117.

FlyBase Consortium, 1998 FlyBase: a drosophila database. Nucleic Acids Res. **26:** 85–88.

Foss, E., R. Lande, F. Stahl and C. Steinberg, 1993 Chiasma interference as a function of genetic distance. Genetics **133:** 681–691.

Fulton, R. E., M. L. Salasek, N. M. DuTeau and W. C. Black IV, 2001 SSCP analysis of cDNA markers provides a dense linkage map of the *Aedes aegypti* genome. Genetics **158:** 715–726.

Gadau, J., R. E. Page, Jr. and J. H. Werren, 1999 Mapping of hybrid incompatibility in *Nasonia*. Genetics **153:** 1731–1741.

Glover, F., 1989a Tabu search part I. ORSA J. Comp. **1:** 190–206.

Glover, F., 1989b Tabu search part II. ORSA J. Comp. **2:** 4–32.

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

Jordan, R. A., and R. W. Brosemer, 1974 Characterisation of DNA from three bee species. J. Insect. Physiol. **20:** 2513–2520.

Laurent, V., E. Wajnberg, B. Mangin, T. Schiex, C. Gaspin *et al.*, 1998 A composite genetic map of the parasitoid wasp *Trichogramma brassicae* based on RAPD markers. Genetics **150:** 275–282.

Lin, S., and T. Speed, 1996 Incorporating crossover interference into pedigree analysis using the chi-square model. Hum. Hered. **46:** 315–322.

Lin, S., R. Cheng and F. A. Wright, 2001 Genetic crossover interference in the human genome. Ann. Hum. Genet. **65:** 79–93.

Moore, D., and T. Orr-Weaver, 1998 Chromosome segregation during meiosis: building an unambivalent bivalent. Curr. Top. Dev. Biol. **37:** 263–299.

Morgan, T. C., C. B. Bridges and A. H. Sturtevant, 1925 The genetics of *Drosophila melanogaster*. Bibliog. Genet. **2:** 1–262.

Mori, A., D. W. Severson and B. M. Christensen, 1999 Comparative genetic maps for the mosquitoes (*Culex pipiens* and *Aedes aegypti*) based on common RFLP loci. J. Hered. **90:** 160–164.

Mutebi, J. P., W. C. Black, IV, C. F. Bosio, W. P. Sweeney, Jr. and C. B. Craig, Jr., 1997 Linkage map for the Asian tiger mosquito [*Aedes* (*Stegomyia*) *albopictus*] based on SSCP analysis and RAPD markers. J. Hered. **88:** 489–494.

O'Brien, S. J. (Editor), 1993 *Genetic Maps*, Ed. 6. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Orr, H. A., and M. Turelli, 2001 The evolution of post-zygotic isolation: accumulating Dobzhansky-Muller incompatibilities. Int. J. Org. Evol. **55:** 1085–1094.

Pamilo, P., S.-L. Varvio-Aho and A. Pekkarinen, 1978 Low enzyme gene variability in Hymenoptera as a consequence of haplodiploidy. Hereditas **88:** 93–99.

Promboon, A., T. Shimada, H. Fujiwara and M. Kobayashi, 1995 Linkage map of random amplified polymorphic DNAs (RAPDs) in the silkworm, *Bombyx mori*. Genet. Res. **66:** 1–7.

Schiex, T., and C. Gaspin, 1997 CARTHAGENE: constructing and joining maximum likelihood gene maps. Proc. Int. Conf. Intell. Syst. Mol. Biol. **5:** 258–267.

Schug, M. D., C. M. Hutter, A. Vetterstrand, M. S. Gaudette, T. F. C. Mckay *et al.*, 1998 The mutation rates of di-, tri- and tetranucleotides repeats in *Drosophila melanogaster*. Mol. Biol. Evol. **15:** 1751–1760.

Severson, D. W., A. Mori, Y. Zhang and B. M. Christensen, 1993 Linkage map for *Aedes aegypti* using restriction fragment length polymorphism. J. Hered. **84:** 241–247.

Severson, D. W., A. Mori, V. A. Kassner and B. M. Christensen, 1995 Comparative genetic map for the mosquitoes *Aedes albopictus* and *Aedes aegypti*, based on common RFLP loci. Insect Mol. Biol. **4:** 41–45.

Shi, J., D. G. Heckel and M. R. Goldsmith, 1995 A genetic linkage map for the domesticated silkworm, *Bombyx mori*, based on restriction fragment length polymorphism. Genet. Res. **66:** 109–126.

Solignac, M., D. Vautrin, A. Loiseau, F. Mougel, E. Baudry *et al.*, 2003 Five hundred and fifty microsatellites markers for the study of the honey bee genome (*Apis mellifera* L). Mol. Ecol. Notes **3:** 307–311.

Sym, M., and G. S. Roeder, 1994 Crossover interference is abolished in the absence of a synaptonemal complex protein. Cell **79:** 283–292.

Tan, Y.-D., C. Wan, Y. Zhu, C. Lu, Z. Xiang *et al.*, 2001 An amplified fragment length polymorphism map of the silkworm. Genetics **157:** 1277–1284.

Tarès, S., J.-M. Cornuet and P. Abad, 1993 Characterization of an unusually conserved *Alu*I highly reiterated DNA sequence family from the honeybee, *Apis mellifera*. Genetics **134:** 1195–1204.

Whitfield, C. W., M. R. Band, M. F. Bonaldo, C. G. Kumar, L. Liu *et al.*, 2002 Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. Genome Res. **12:** 555–566.

Yasukochi, Y., 1998 A dense genetic map of the silkworm, *Bombyx mori*, covering all chromosomes based on 1018 molecular markers. Genetics **150:** 1513–1525.

Zhao, H., M. S. McPeek and T. P. Speed, 1995a Statistical analysis of chromatid interference. Genetics **139:** 1057–1065.

Zhao, H., T. P. Speed and M. S. McPeek, 1995b Statistical analysis of crossover interference using the chi-square model. Genetics **139:** 1045–1056.

Zhao, H., and T. P. Speed, 1996 On genetic map functions. Genetics **142:** 1369–1377.

Zheng, L., M. Q. Benedict, A. J. Cornel, F. H. Collins and F. C. Kafatos, 1996 An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae*. Genetics **143:** 941–952.

Communicating editor: A. Nicolas