Mitochondrial DNA variability in honeybees and its phylogeographic implications

J.M. Cornuet, L. Garnery

Original article

Mitochondrial DNA variability in honeybees and its phylogeographic implications

JM Cornuet, L Garnery

INRA-CNRS, Laboratoire de Neurobiologie Comparée des Invertébrés, URA1190, La Guyonnerie 91440 Bures-sur-Yvette, France

Summary — A physical map of mitochondrial DNA (mtDNA) of the honeybee (*Apis mellifera* L) has been established with 17 restriction enzymes (46 sites). Elements of the genic map have been inferred from sequence data. The superimposition of both maps indicate that the gene order is quite similar between honeybee and *Drosophila*. The total length of the mitochondrial genome falls between 16 500 and 17 000 bp. This range is due to several regions exhibiting length polymorphisms. Two of them overlap with the control region, but a third one is unexpectedly located between the CO-I and CO-II genes. This last polymorphism is explained by the occurrence of variable numbers of 2 related sequences, called P and Q, which arose through tandem duplication. Sequence data from 3 regions of the mtDNA genome can be used to infer a phylogenetic tree for 4 *Apis* species: the resulting tree topology, *(florea(dorsata(cerana,mellifera)))*, confirms the phylogeny based on morphometry and behavior.

The mtDNA variability of *Apis mellifera* indicates 3 major lineages: African colonies (lineage A) including *intermissa, adansonii, scutellata, capensis* and *monticola* subspecies; *mellifera* colonies (lineage M); *ligustica, carnica* and *caucasica* colonies (lineage C). This distribution is very similar to the 3 evolutionary branches inferred from morphometric analysis by Ruttner. The main difference concerns the branch M which, according to Ruttner, includes also *intermissa* and *iberica*. From an Asian origin, 3 evolutionary branches colonized respectively northern Europe (M), the north-Mediterranean region (C) and Africa (A). Based on the *Drosophila* evolutionary rate, this divergence would have occurred between 300 000 and 1300 000 bp.

Apis mellifera / **mitochondrial DNA** / **variability** / **phylogeny** / **evolution**

INTRODUCTION

Mitochondrial DNA (mtDNA) of animals is a duplex, small, circular, covalently closed molecule. Its gene content is remarkably conserved across taxa (Moritz *et al*, 1987) and comprises 2 ribosomal RNA genes (rRNA, large and small subunits), 22 transfer RNA genes (tRNA) and 13 protein genes (12 in nematodes where the gene for ATPase 8 is lacking; Okimoto *et al*, 1990). Intervening sequences within genes are unknown, spacers between genes are absent or very short and some instances of slightly overlapping genes are known. Consequently, animal mtDNA is a lesson in economy (Attardi, 1985). Molecules contain a single non-coding region (D-loop in vertebrates, A+T-rich region in *Drosophila*), also called the control region, in which replication and transcription are initiated.

This high gene conservation combined with the relatively small variation of the non-coding region means that most mitochondrial genomes comprise between 15 000 and 20 000 nucleotides (Brown, 1983). The order in which genes are arranged along the molecule is generally rather conserved within phyla, but some rearrangements occur between phyla.

The particular interest of mtDNA in phylogenetic studies is mainly due to 2 features. The first is the apparent lack of recombination. Therefore, genetic markers are completely linked and substitutions are serially accumulated on the molecule (Harrison, 1989). The second is its maternal transmission (at least in animals). Because of these characteristics, the evolution of mtDNA is clonal, even in sexual species, and the relationship between maternal lineages is perfectly depicted by the phylogenetic tree of the mtDNA molecules.

MtDNA has become an important marker in evolutionary studies of natural populations (Avise *et al*, 1987). It has proved to be particularly useful for *Apis mellifera*, a species in which few nuclear markers are available (Sheppard and Huettel, 1988). Furthermore, the biology of this species makes it especially suited since all the colony members share the same mtDNA (mtDNA is a colony level marker), and colonization of new territories is performed by swarms, *ie*, females.

Given its sequestration in mitochondria and its physical properties, mtDNA offers several possibilities for purification (reviewed in Solignac, 1991). MtDNA can be separated from nuclear DNA on ethidium bromide–CsCl gradients (Lansman *et al*, 1981). Another method involves purification of mitochondria by differential centrifugation on sucrose gradients. The mtDNA extracts obtained through these methods have been analyzed with restriction enzymes or cloned for sequencing.

Previously 10–50 workers per colony, alive or frozen, were necessary to obtain sufficient mtDNA to complete experiments (*eg*, to obtain restriction profiles for one colony or to clone restriction fragments). This constraint no longer exists : a single worker per colony, stored in pure ethanol for several weeks at room temperature, is now sufficient for mtDNA analysis. For mtDNA restriction studies, total DNA is extracted according to a simple procedure (Kocher *et al*, 1989) (24 samples can be routinely treated in one day) and digested with various restriction enzymes (one worker provides enough DNA for up to 12 digestions). After electrophoresis, fragments are transferred to membranes and revealed with a labeled probe as already described in, for instance, Sheppard *et al* (1991). Previous methods are only necessary to prepare pure mtDNA probes.

Polymerase chain reaction (PCR) is also a powerful technique that can produce large amounts of DNA from minute quanti-

ties of material. Some "universal" or "insect" primers can be used as well as specific primers designed with the help of sequences available from the literature. PCR-amplified DNA can be used for direct analysis (length variation), restriction (Hall and Smith, 1991) or sequence (Cornuet *et al*, 1991; Garnery *et al*, 1991) analysis.

ORGANIZATION OF THE MITOCHONDRIAL GENOME IN *APIS*

In *Apis mellifera,* the genetic map, *ie,* the nature and position of genes on the molecule, is only partially known. The available information has been deduced from a physical map (restriction map), hybridization experiments and sequence data. Vlasak *et al* (1987) published the nucleotide sequence of 95% of the large ribosomal RNA gene (1270 bp). Crozier *et al* (1989) sequenced 2 950 bp in a region in which they recognized 4 tRNA genes (tryptophan, leucine, aspartate and lysine), and 2 cytochrome-oxidase subunit genes (CO-I and CO-II). In this region, they also found an unassigned 194 bp sequence, the nature of which will be discussed in another section. In the first 185 bp of their sequence, one can also find a tRNA gene (tyrosine), extending from position 56 to 124 and transcribed in the opposite direction. Furthermore, between the tRNA[tyr] and tRNA[trp] genes lies a tRNA-like sequence (with a GGG in the anticodon position). Other sequences will soon be available: cytochrome *b* and ATPases 6 and 8 subunit genes (Crozier and Crozier, in press), part of the CO-III gene (Koulianos and Crozier, 1991).

Restriction maps of honeybee mtDNA have been independently elaborated by Smith (1988), Crozier *et al* (1989) and Cornuet (1989). The restriction sites of the enzymes used to establish the physical map can be searched in available partial sequences of the genome. The pattern obtained for the sequenced region can be unambiguously superimposed onto the corresponding region of the physical map, and then the position of the gene or of the block of genes assigned to a precise region of the genome. A partial genetic map has been obtained by this procedure for the honeybee. The relative position of the genes and their distances on the map correspond almost exactly to their counterparts in the *Drosophila yakuba* map (Clary and Wolstenholme, 1985) (fig 1). A similar conclusion has been obtained through Southern blotting (Garnery, unpublished data). Considering that the genetic map is generally conserved across taxa and that the genome sizes in *Apis mellifera* and *Drosophila yakuba* are not very different, it seems very likely that both species have similar genetic maps, keeping in mind that some t-RNA genes, at least, have slightly different locations (tryptophan, aspartate, lysine and, perhaps, some others). Also, the intergenic sequence between the t-RNA leucine and the CO-II genes has no counterpart in *D yakuba* mtDNA. More sequence data are needed in order to obtain a definite map of honeybee mtDNA.

Another feature for which honeybee mtDNA is like that of *Drosophila* is in the high proportion of A+T. In the genes so far sequenced, the A+T content is even higher than in *Drosophila*: 76% compared to 68% and 80% compared to 75%, in CO-I and CO-II respectively (Crozier *et al,* 1989).

LENGTH POLYMORPHISMS

The total length of honeybee mtDNA is variable, approximately ranging between

JM Cornuet, L Garnery
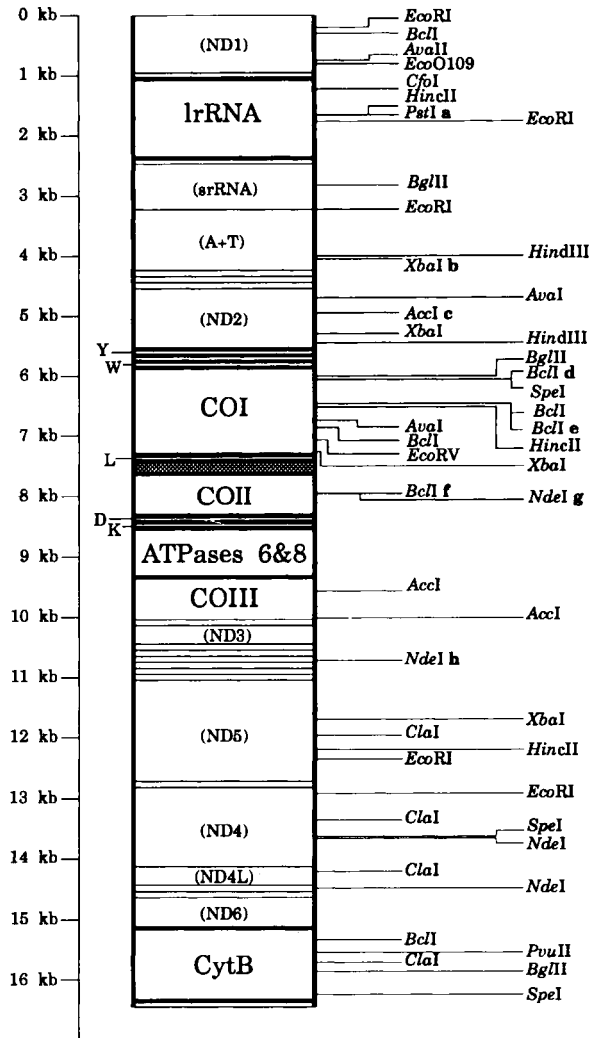
Fig 1. Restriction map of *Apis mellifera* mtDNA. From left to right are given : 1) a scale in kb (1 kb = 1 000 nucleotides); 2) the location of *tRNA* genes (Y for tyrosine, W for tryptophan, L for leucine, D for aspartate and K for lysine) deduced from nucleotide sequence; 3) the gene map of *Drosophila ya-kuba* (Clary and Wolstenholme, 1985). Genes noted in larger characters have been found also in *Apis mellifera* at the same location. For other genes, marked between parentheses, the exact location in honey bee mtDNA has still to be verified. The shaded area between the tRNA[leu] and COII genes is a variable length, intergenic sequence which is found in *Apis* but not in *Drosophila*. Its origin and puta-tive function are discussed in the text: a list of restriction sites which have been found in every *Apis mellifera* colony (constant sites); a list of restriction sites which have not been found in every *Apis mellifera* colony (variable sites). Some of the above restriction sites, followed by a letter (a to h), are refered to in the text. These notations are taken after Smith and Brown (1990).

16 500 and 17 000 bp . This range results from length variability in several regions of the molecule. Smith and Brown (1990) have detected 6 locations with length variation. Two of them are of small size but others are close to or over 100 bp.

Actually, the 2 adjacent small size (20 bp) variations, both detected on *BclI* digests, could be the result of one site loss and one site gain. The sequence of the CO-I gene (Crozier *et al*, 1989), where the three concerned *BclI* sites, noted d, e and f in figure 1, are located, provides a possible clue: 20 bp upstream of *BclI* e, there is a motif (TGACCA) which is one transition away from a *BclI* site (TGATCA), this transition being located at the unconstrained third position of the codon. Combined with the loss of *BclI* e, the gain of this potential site would then explain these 2 size variations.

Most large scale size variations encountered in various animal species have been found in or near the control region of the mtDNA molecule. This seems to be the case for 2 of the larger variations in honeybee mtDNA. They are respectively located between *PstI* a and *XbaI* b and between *XbaI* b and *AccI* c. According to our putative genetic map, these 2 regions include a portion of the A+T rich region. For the *XbaI* b - *AccI* c region, Smith and Brown (1990) propose that a variable number of tandem repeats 80–100 bp in length may have caused the size variation.

The size variation located between *BclI* f and *BclI* g is the largest and best characterized (Cornuet *et al*, 1991). In the intergenic region between the tRNA[leu] and CO-II genes, 2 types of sequences have been identified: P and Q which are 54 and 194–196 bp long, respectively. The first one, P, is optional. When it is present, it can take 2 forms respectively called P and P_o which differ by a 15-bp deletion in the middle of the sequence. The size variation

in this mtDNA region results from the occurrence/absence of P (or P_o) and a variable number of Q. The following genomes have been found: Q, PQ, PQQ, PQQQ, P_oQ, P_oQQ, P_oQQQ, and P_oQQQQ (the last one was not observed by us, but its existence is inferred from the data of Hall and Smith, 1991).

The sequence Q is actually composed of 3 subunits, Q_1, Q_2 and Q_3, which show strong similarities in the primary and secondary structures with the 3' end of CO-I, the tRNA[leu] gene and P_o, respectively. The duplication followed by the divergence of this block can explain the appearance of the Q sequence but the origin of the P sequence and hence of the Q_3 sequence remained unexplained. In order to elucidate their origin, we examined the region between the tRNA[leu] and CO-II genes in closely related species (*Apis cerana*, *A dorsata* and *A florea*) and in two related Apidae : *Bombus lucorum* and *Xylocopa violacea*. These *Apis* species have an intergenic sequence of 89, 24 and 32 bp, respectively, whereas the last 2 species do not have a single nucleotide between these genes. Two possible scenarios for the origin of these sequences have been given (Cornuet *et al*, 1991), one of which is illustrated in figure 2. It has been suggested by Cornuet *et al* (1991) that this intergenic region might function as an origin of replication.

HIGH EVOLUTIONARY RATES OF MITO-CHONDRIALLY ENCODED PROTEINS

Crozier *et al* (1989) have compared the subunits CO-I and CO-II of *Drosophila* and *Apis* using *Xenopus* and *Mus* as outgroups. The number of amino acid replacements was respectively 3.6 and 1.9 times higher in the lineage leading to *Apis* than of that leading to *Drosophila*. However,
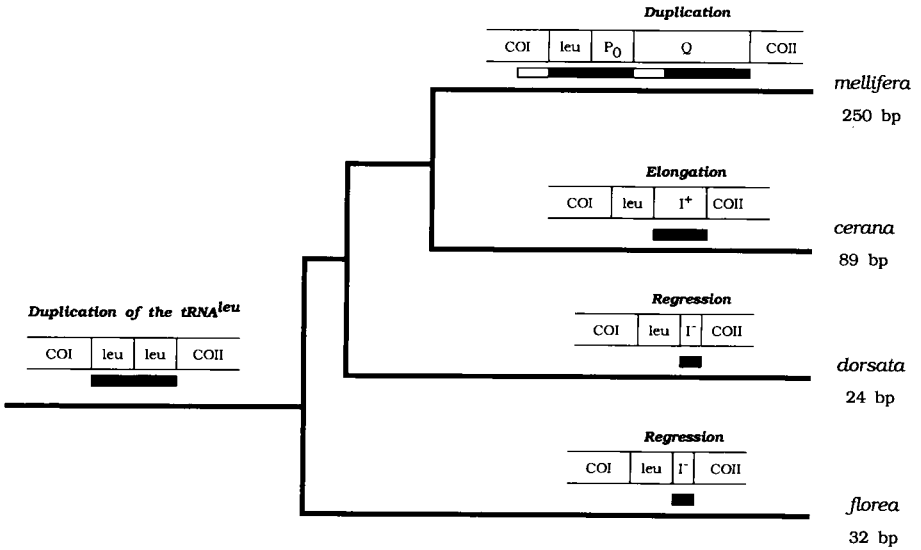
Fig 2. Putative origin of the COI-COII intergenic region of *Apis* mtDNA (Cornuet *et al*, 1991). In the lineage common to all *Apis* species, the tRNA^{leu} gene has been tandemly duplicated. Then, in *Apis dorsata* and *Apis florea,* this duplicated tRNA gene has been progressively deleted, leaving a short sequence similar to the tRNA^{leu} gene stem (noted I⁻ for "intergenic deleted"). In *Apis cerana*, the duplicated tRNA gene evolved toward a A+T rich clover-leaf sequence (noted I⁺ for "intergenic elongated"). In *Apis mellifera*, in the ancestor common to all current subspecies, a second tandem duplication encompassing the 3' end of COI subunit gene, the original tRNA^{leu} gene and its duplicate, occured. P_0 corresponds to the primarily duplicated tRNA and Q to the whole secondarily duplicated sequence. This P_0Q, 250 bp long, intergenic sequence would be ancestral to all other combinations found in the various subspecies.

these values are averages for the 280 Myr of divergence between the 2 species and may not necessarily reflect the present day ratio between the amino acid replacement rates of these species.

In short-term evolution, 90% of nucleotide substitutions are silent mutations. Therefore, there is no direct relationship between amino acid replacement and nucleotide substitution rates when considering related species and populations of the same species.

Additional sequences (ATPase 6, ATPase 8 and cytochrome *b* genes (Crozier and Crozier, personal communication)

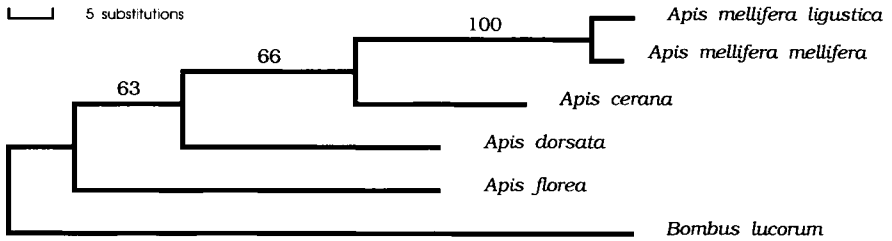confirm the higher amino acid replacement rate of mtDNA encoded proteins along the *Apis* lineage.

EVOLUTION OF THE GENUS *APIS*

The phylogeny of *Apis* species is still an open question, especially given the recent description of *A andreniformis* (Wu and Kuang, 1987) and *A koschevnikovi* (Koeniger *et al*, 1988; Tingek *et al*, 1988). A comprehensive review, based on morphology and behavior, has been given by Alexander (1991).
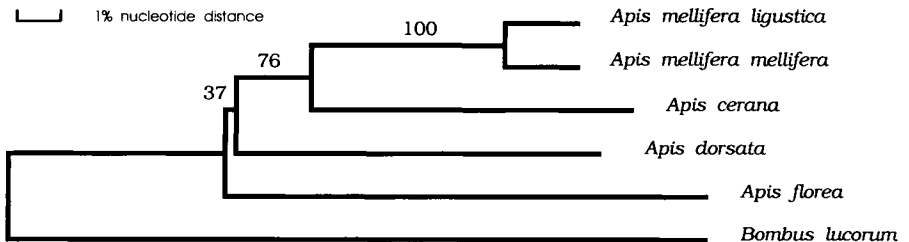
Large ribosomic subunit

```
A. mel. lig.   ATTTGATTGGGAGGATTGGTAAATTTAATAAACTTTATTTAAGAT-TTAATTTTGATATAAAGA-TAATTTTTGA-ATTATAAATTGTAATTA
A. mel. mel.   ..................................................A..-................-..........-....C..........
A. cerana      .....................A...AT........C...T.A-.........A..GA.....-.TTA......-C.....T..A.G....
A. dorsata     ...................A.......A.T..........A...T..-A................-..GA.A....AT..T...T..A.....
A. florea      ................AA....................T.A..-.........TAG....-.TAAAA...TT.AT...T..A...A..
B. lucorum     T....-..........A...........T.N.........T.A..A....C..A..TA.....A..TA......-TC.T.T.G.AAA...G.

A. mel. lig.   AAAGAATAAATTACCTTAGGGATAACAGCGTAATATCTTTTGATAGACCATATAGATAAAGATGTTTGCGACCTCGATGTTGAATTAAGATG
A. mel. mel.   ...........................................................T............................
A. cerana      ....................................................C..T............................A
A. dorsata     TT...G...............................................A......................NNN
A. florea      T..................................................GT.....T.........A.....G.................A
B. lucorum     .....G...............................T.A..T.T...T..A.......A.......................A
```

Cytochrome oxidase I subunit

```
A. mel. lig.   AGATTCTTATTACTGTTGAAATTCAATTTCATCTATAGGATCAATAATTTCATTAAATAGAATAATTTTTTTAATTTTTATTATTTTAGAAAGA
A. mel. mel.   .......................................C..................................................
A. cerana      T...........T...................AT...............T..C...............................
A. dorsata     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN.AT.N........................A..............A.............
A. florea      N...T..T..NT.................T..A..T.......CTG.A............T....................CA.......
B. lucorum     NNNNNNNNNNNNNNNNNN...C.T...A..T..A..T.N.G...........A...........T.A.........C........T......

A. mel. lig.   TTAATTTCTAAACGAATATTATTATTTAAATTCAACCAATCATCACTTGAATGATTAAATTTTTTACCACCTCTAGATCATTCACATTTAGAA
A. mel. mel.   ...............................................................C.........T................
A. cerana      ........A......T..A.......C.....T..T........T.......AT....AT.......C...........
A. dorsata     ........A.......A...............T..T.....T..TT..AG.........A..C..G....AT....C........A.....
A. florea      A....AG.A......T................T................CAAC......AA.T..A..C......G.....
B. lucorum     ...............T..A.T...........C.T....................AA..AT..T...TAT.....C...TTAA.T...
```

Fig 3. Nucleotide sequences of 4 *Apis* species and *Bombus lucorum* (one individual per species) in 2 different mtDNA regions: 185 bp in the middle of the large rRNA gene and 187 bp in the 3' end of the CO-I subunit gene. Dots indicate nucleotide identity with the first line *(Apis mellifera ligustica),* dashes indicate deleted nucleotides and N are for undetermined nucleotides.

Molecular data are especially suited to address phylogenetic questions (Hillis and Moritz, 1990). Recently, nucleotide sequence data of mtDNA have been used to produce phylogenetic trees of the 4 "old" *Apis* species : *cerana, dorsata, florea* and *mellifera* (Garnery *et al,* 1991). The analysis was based upon the sequence of 269 nucleotides located at the 5' end of the CO-II gene. In the consensus tree, it was found that *A mellifera* and *A cerana* separated after the divergence of *A florea* and *A dorsata,* but the *(cerana-mellifera)-dorsata-florea* trifurcation remained unsolved.

In order to solve this polytomy, we have extended the sequences from the 4 species, and that of *Bombus lucorum,* taken as outgroup: to the 269 previously reported nucleotides, we have added 187 nucleotides in the 3' end of CO-I and 185 nucleotides in the large rRNA subunit (fig 3). This provides a total of 641 sites of which 61 are informative. Two methods of tree reconstruction, Parsimony (PAUP v3.0) and Neighbor-Joining (Saitou and Nei, 1987) have been applied. Bootstrap values (percentages computed on 1 000 replicates) are given for every internal branch. Both methods give the same topology (fig 4).

The present results confirm the closer relationship between *A mellifera* and *A cerana.* In addition, the trifurcation *(mellifera-*

Parsimony

Neighbor-Joining

Fig 4. Phylogenetic trees based on mtDNA sequence data from 4 *Apis* species, *Bombus lucorum* being taken as an outgroup. Sequences are 641 bp long (187 bp in the 3' end of COI, 269 bp in the 5' end of COII and 185 bp in the middle of the large rRNA gene). They have been obtained by direct sequencing of PCR amplified DNA as described in Garnery *et al* (1991). The bootstrap values above the internal branches are percentages computed from 1 000 replicates, for both trees. The most parsimonious tree has been obtained through PAUP v3.0 software, taking only the 61 informatives sites and considering them as unordered characters. The Neighbor-Joining algorithm of Saitou and Nei (1987) has been performed on nucleotide distances computed by Kimura's (1980) formula through a personal program which also performs the bootstrap procedure. In the computation of nucleotide distances, deletions have been counted as transversions.

cerana, dorsata, florea) found by Garnery *et al* (1991) is partially resolved: both trees support the early divergence of *A florea*. Although bootstrap values are not high enough to consider that this study is conclusive, the consistency of these mtDNA analyses with morphometric, behavioral (Alexander, 1991) and allozymic (Sheppard and Berlocher, 1989) characters is noticeable.

The mtDNA evolutionary rate has not been calibrated in *Hymenoptera*. The most closely related group of insects for which evolutionary rates have been estimated is *Drosophila*, which exhibits a rate of divergence between lineages of 2% per Myr for the main bulk of mtDNA (DeSalle *et al,* 1987; Monnerot *et al,* 1990), a value comparable to that observed in mammalian species (Brown *et al,* 1982). This value

must be accepted only with caution for *Apis*. The divergence between *A mellifera* and *A cerana,* based on the above 643 bp, is equal to 8.5%. However, the 3 sequenced fragments do not necessarily provide a representative value of the divergence for the whole molecule. The comparison of restriction maps of both species (53 restriction sites) gives a 13.7% divergence (Garnery *et al,* unpublished data) which, after deduction of the nucleotide diversity in the ancestral species, provides an estimate of 5.9 Myr, in perfect agreement with the former estimate (Garnery *et al,* 1991).

EXISTENCE OF THREE MAJOR mtDNA LINEAGES IN *APIS MELLIFERA*

MtDNA variability has been investigated within and between subspecies by restriction enzymes (Moritz *et al,* 1986; Smith, 1988; Smith and Brown, 1988, 1990; Cornuet, 1989; Hall and Smith, 1991) and sequencing (Cornuet *et al,* 1991; Koulianos and Crozier, in press; Garnery *et al,* unpublished data). Sampling is greater with restriction enzymes (more colonies analysed), but the data are more accurate with sequences (more nucleotides assayed). The 2 approaches are largely congruent.

Restriction site variability in *Apis mellifera* is large: 22 out of 46 restriction sites mapped in honeybee mtDNA are absent in at least one colony. In a sample of 68 colonies from 10 different subspecies (Garnery *et al,* unpublished observations), 20 different haplotypes have been scored and the average nucleotide diversity (*ie,* the average difference between 2 mitochondrial genomes taken at random in the species) is equal to 1.9%. The largest nucleotide distance reaches 4% and has been observed between a colony of *A m mellifera* and a colony of *A m carnica* (Smith and Brown, 1990).

The corresponding values obtained with sequence data are lower, essentially because sequences have been obtained for regions evolving more slowly than the whole genome. Nevertheless, the 600-bp sequences of 11 colonies from 8 subspecies appeared different, even those of the same geographic origin.

A clear picture emerges from mtDNA variability in *Apis mellifera:* colonies and subspecies are clustered into 3 major lineages A, M and C, which correspond respectively to the following (Smith, 1991; Garnery *et al,* unpublished data):

- an African lineage (A) including all the mtDNA types detected in subspecies *intermissa, adansonii, scutellata, capensis* and *monticola;*

- a west-Mediterranean lineage (M) corresponding to *Am mellifera;*

- a north-Mediterranean and Caucasian lineage (C) including the subspecies *ligustica, carnica* and *caucasica.*

The status of the subspecies *iberica* will be discussed in the section *Evolutionary history* of *Apis mellifera.*

The nucleotide diversity within the species is equal to 1.9%. The values of this average nucleotide distance between the different lineages *(ie,* one haplotype taken at random in one lineage and the second in the other lineage) are as follows:

- lineage A - lineage C = 2.1%;

- lineage A - lineage M = 2.8%;

- lineage C - lineage M = 3.3%.

WITHIN LINEAGE mtDNA VARIABILITY

The within lineage diversity is much lower than between lineages: 0.25, 0.30 and 0.07% for lineages A, C and M, respectively. The lower diversity found in lineage M may not be significant. This may be due to the fact that the proportion of the M lineage sampled is smaller than for other lineages.

Variability has been found within every subspecies in which several colonies have been examined. Because of the variability within subspecies, it is often impossible to clearly discriminate between some sub-species of the same lineage. For instance, the same haplotype (according to the re-striction map) has been found in *intermis-sa, iberica, adansonii* and *scutellata*.

The number of nucleotides sequenced is still too small to figure out if it will be possible to unequivocally discriminate be-tween all the subspecies within a lineage. Direct sequencing of DNA amplified through the polymerase chain reaction will certainly be the most powerful tool for such a discriminative objective. It is none-theless necessary to apply this technique to a sequence with an appropriate evolu-tionary rate. The mtDNA regions that have

been sequenced so far evolve too slowly for this purpose.

Current data on honeybee mtDNA are then insufficient to discriminate between all the recognized subspecies, but as soon as sequences for the most variable regions (*eg,* the A+T rich region) are known, it is likely that subspecies and possibly popula-tions will be discriminated between using this method.

EVOLUTIONARY HISTORY OF *APIS MELLIFERA*

Figure 5 provides a Neighbor-Joining phy-logenetic tree of 8 subspecies (1 colony per subspecies) based on sequence data from the CO-I – CO-II region. It clearly shows (very high bootstrap values) the 3 distinct mtDNA lineages.
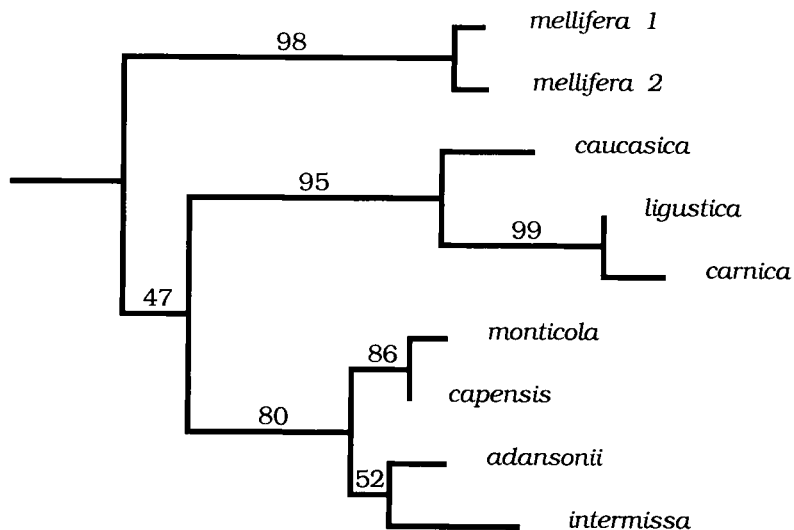
Fig 5. Neighbor-Joining tree of 9 colonies of *Apis mellifera* refered to by their subspecies names and rooted by one colony of *Apis cerana* (not shown). The distances used are nucleotide distances com-puted by Kimura's formula from a sequence of 652 bp including 187 bp in the 3' end of COI, 269 bp in the 5' end of COII and 196 bp in the intergenic sequence Q (Garnery *et al,* unpublished data). The bootstrap values above the internal branches are percentages computed from 1,000 replicates.

These lineages correspond to a racial and geographical division of the species which strongly corresponds to Ruttner's hypotheses. Based on morphometrical data, Ruttner *et al* (1978) recognized 3 "evolutionary branches" :

- the branch "M" goes from *intermissa* to *mellifera* through *iberica*;

- the branch "C" encompasses *cecropia*, *carnica* and *ligustica*;

- the branch "A" is made up of most of the African subspecies.

This subdivision of the species is supported by a phylogenetic analysis of allozyme data of Sheppard and Huettel (1988). However, the number of races was too small to fully validate the existence of all 3 branches.

The 3 major lineages of the mtDNA phylogenetic tree in *Apis mellifera* are in general agreement with the branches recognized by Ruttner *et al* (1978). The congruence is so high that the same symbols (M, C and A) have been given to the mtDNA lineages. However, some discrepancies exist.

The main discrepancy concerns the Spanish transition between *intermissa* and *mellifera*. Although Ruttner proposed a progressive step between 2 subspecies of the same branch, mtDNA reveals a secondary contact between 2 distantly related lineages (Smith *et al*, 1991). Spain appears as a hybrid zone between the subspecies *mellifera* and *intermissa*, the latter unambiguously exhibiting the African mitochondrial type. Both mitochondrial types coexist, but the *mellifera* type is dominant in the northern half of the Iberian peninsula whereas the African type is dominant in the south. The north to south clines found for morphometrical (Cornuet and Fresnaye, 1989) as well as allozymic characters (Cornuet, 1982), which were formerly considered as arguments in favor of Ruttner's view, can be alternatively interpreted

as the result of a hybridization process between evolutionary branches. Actually, a secondary contact between 2 distant branches (M and C) was the explanation given for the pattern of allozymic variability found in the Ligurian coast (Cornuet, 1983) or in the Aosta valley (Badino *et al*, 1982). The only difference is the geographic extension of the hybrid zone which is much larger in Spain than in Italy. This difference can be explained by the lack of natural obstacles in Spain and also perhaps by beekeepers' influence.

Another discrepancy relates to an additional branch, later proposed by Ruttner (1988). This branch, named "O", includes *meda, anatoliaca* and *caucasica*. Although *meda* and *anatoliaca* are missing in our sample, this branch is not supported by mtDNA data which show that *caucasica* is close to *ligustica* and *carnica* and belongs unequivocally to the same lineage.

In the general classification of geographic distribution of mtDNA variability made by Avise *et al* (1987), *Apis mellifera* belongs to category I, which shows a discontinuous divergence pattern, geographically structured and characterized by levels of divergence much higher between than within regions. According to these authors, this pattern, which is quite common, is explained by "long term, extrinsic (*ie* zoogeographic) barriers to gene flow, such that conspecific populations occupy easily recognizable branches on an intraspecific evolutionary tree" and/or by "extinction of intermediate genotypes in widely distributed species with limited dispersal and gene flow capabilities".

After the remarkable spread of Africanized bees in America, one cannot consider that *Apis mellifera* has "limited dispersal capabilities". Therefore, the first explanation appears correct for honeybees: the current pattern results from a long-term separation of 3 branches which diverged allopatrically.

In *Apis mellifera,* the mitochondrial genome is structured not only at the level of the 3 major lineages but also, although to a lesser extent, within each of these 3 lineages, *ie* at the level of subspecies. This suggests that the divergence of the mtDNA types is not the result of independent genetic drift on previously highly polymorphic populations but, more probably, was mainly established in their respective present-day geographic areas, *ie* in allopatry. Consequently the divergence of their mtDNA can be used to estimate the age of the populations.

The period of installation of the 3 main mtDNA lineages was probably comprised between −1.3 and −0.3 Myr. The first value corresponds to the age of divergence of their mtDNA, allowing for the extent of polymorphism within lineages (see Wilson *et al,* 1985 for the principle). The second date corresponds to the beginning of the differentiation of the genomes within the main lineages; as the variability within these lineages is also structured, the 3 ancestral populations were already in place at this time. Both calculations imply the acceptance of a 2% divergence per Myr. These estimates can be compared to allozyme based estimates taken from Sheppard and Berlocher's article (1989). According to their figure 1, Nei's distance between *mellifera* and *carnica–ligustica* is approximately equal to 0.04. Applying the formula $t = 5D$ (Nei, 1987), this leads to a value of 200 000 years which agrees well with the above estimates.

Résumé — La variabilité de l'ADN mitochondrial chez l'abeille domestique et ses implications phylogéographiques. L'ADN mitochondrial (ADNmt) de l'abeille domestique (*Apis mellifera* L) est une molécule circulaire dont la taille varie de 16 500 à 17 000 paires de bases. Sa carte physique, établie à l'aide de 17 enzymes de restriction, comprend 46 sites. La connaissance de la séquence d'un certain nombre de gènes codant pour la grande sous-unité de l'ARN ribosomique (lrRNA), trois sous-unités de la cytochrome oxydase (CO-I, CO-II et CO-III), les ATPases 6 et 8, le cytochrome *b* et 5 ARN de transfert (tyrosine, tryptophane, leucine, aspartate et lysine) a permis de superposer la carte physique et ces quelques éléments de la carte génétique. Ces résultats, encore incomplets, mettent en évidence une similitude remarquable de la position des gènes chez l'abeille et drosophile. L'étendue de variation indiquée plus haut pour la longueur totale de la molécule provient de l'existence de plusieurs régions présentant des variations de longueurs. Deux de ces régions incluent vraisemblablement la région de contrôle, dite «riche en A+T» chez la drosophile ou «D-loop» chez les vertébrés. Cette région, où s'initient réplication et transcription de l'ADNmt, est le siège privilégié de variations de longueur chez de nombreuses espèces animales. À la différence des autres insectes connus, la région comprise entre CO-I et CO-II présente un important polymorphisme de taille chez l'abeille, dû à l'existence d'un nombre variable de deux séquences notées P et Q. La première peut manquer; lorsqu'elle est présente, toujours en un seul exemplaire, elle peut comporter ou non une insertion centrale de 15 pb. La séquence Q, longue de 194 à 196 bp, est présente à un ou plusieurs (2, 3 ou 4) exemplaires répétés en tandem. En raison des similitudes de séquence et de structure secondaire entre Q et la séquence située en 5', il est très probable que la séquence Q résulte d'une duplication. L'examen de la séquence correspondante chez d'autres espèces *d'Apis* indique que la séquence P proviendrait, elle aussi, d'une duplication plus ancienne et limitée au gène voisin du tRNA leucine. Des analyses réalisées par Crozier *et al* sur le taux de remplacement des

acides aminés dans l'évolution des gènes CO-I et CO-II indiquent que ce taux serait de 2 à 3 fois plus élevé dans la lignée conduisant à *Apis* que dans celle conduisant à la drosophile. Toutefois, ces différences sont des moyennes sur les 280 millions d'années qui séparent ces deux lignées de leur ancêtre commun; elles n'impliquent pas nécessairement que cette évolution accélérée soit encore à l'œuvre actuellement. Un arbre phylogénétique a été construit à partir d'une séquence de 651 bp établie chez 4 espèces d'*Apis (mellifera, cerana, dorsata et flora)* et chez *Bombus lucorum* considéré comme référence externe. La topologie de l'arbre confirme celle trouvée par Alexander (1991) à partir de caractères morphologiques et comportementaux: *(florea(dorsata (cerana,mellifera)))*. En supposant que le taux d'évolution de l'abeille au cours des derniers millions d'années est identique à celui de la drosophile, soit 1% par million d'années, les espèces *A cerana* et *A mellifera* auraient divergé il y a environ 6 millions d'années. L'analyse de la variabilité de l'ADNmt de l'abeille domestique, qu'elle concerne les données de restriction ou de séquence, met en évidence l'existence de 3 lignées majeures, notées A, M et C; elles correspondent respectivement : aux colonies africaines (lignée A) des races *intermissa, adansonii, scutellata, capensis* et *monticola;* aux colonies de race *mellifera* (lignée M); aux colonies de race *ligustica, carnica* et *caucasica* (lignée C).

Cette distribution rappelle étonnamment les trois branches évolutives que Ruttner avait inférées à partir d'analyses morphométriques. La principale différence concerne la branche M qui, selon Ruttner, s'étendait de l'Afrique du Nord (*intermissa*) jusque dans le nord de l'Europe (*mellifera*) en passant par la péninsule ibérique (*iberica*). D'un point de vue mitochondrial, la race *intermissa* est typiquement africaine

et la race *iberica* est un mélange graduel des types africain (dominant au sud) et *mellifera* (dominant au nord).

La distribution géographique des espèces actuelles du genre *Apis* conduit à penser qu'*Apis mellifera* est originaire d'Asie. De là, seraient parties 3 grandes nappes, l'une peuplant l'Europe par le nord (correspondant à la lignée M), l'autre longeant le nord de la Méditerranée (lignée C) et la troisième envahissant l'Afrique (lignée A). L'ampleur des distances nucléotidiques entre lignées est difficilement compatible avec une séparation des nappes postérieure à la dernière glaciation. En se fondant sur le taux d'évolution de l'ADNmt de la drosophile, cet événement se serait produit entre −300 000 et −1 300 000 ans.

Apis mellifera / ADN mitochondrial / variabilité / phylogénie / évolution

Zusammenfassung — Die Variabilität der mitochondrialen DNA der Honigbienen und die phylogeographischen Folgerungen. Die mitochondriale DNA (mtDNA) der Biene ist ein zirkuläres Molekül mit einer Größe von 16 500 bis 17 000 Basenpaaren. Seine physische Karte, erstellt mit Hilfe von 17 Restriktionsenzymen, umfaßt 46 Stellen (sites). Die Kenntnis der Sequenz für eine gewisse Anzahl von Genen, welche die große Untereinheit der ribosomalen DNA (lrRNA), drei Untereinheiten für die Zytochromoxydase (CO-I, CO-II und CO-III), die ATP-asen 6 und 8, das Zytochrom *b* und 5 Transfer-RNAs (Tyrosin, Tryptophan, Leucin, Aspartat und Lysin) hat es ermöglicht, die physische Karte und diese Elemente der genetischen Karte miteinander zu vergleichen. Die Resultate, obgleich noch unvollständig, ergeben eine bemerkenswerte Ähnlichkeit der Position der Gene bei der Biene und bei *Drosophila*. Das oben angegebene

Ausmaß der Variation der Gesamtlänge des Moleküls stammt aus mehreren Regionen, die eine Längenvariation aufweisen. Zwei von diesen Regionen umfassen wahrscheinlich die Kontrollregion, bezeichnet als 'Reich an A+T' bei *Drosophila* oder 'D-loop' bei Wirbeltieren. Diese Region, wo die Replikation und Transkription der mtDNA beginnt, ist bei zahlreichen Tierarten der bevorzugte Sitz von Längenvariationen. Zum Unterschied zu anderen bekannten Insekten zeigt die Region zwischen CO-I und CO-II bei der Biene einen beträchtlichen Polymorphismus der Größe. Er kommt von einer variablen Anzahl von zwei Sequenzen, die als P und Q benannt werden. Die erstere kann fehlen; wenn sie aber vorhanden ist (immer nur in einem Exemplar), kann sie eine zentrale Einfügung von 15 bp enthalten – oder auch nicht. Die Sequenz Q, 194–196 pb lang, ist in einem oder in mehreren (2, 3 oder 4) Examplaren als Tandemduplikation vorhanden. Auf Grund der Sequenzähnlichkeiten und der sekundären Struktur von Q und der Sequenz bei 5', ist es sehr wahrscheinlich, daß die Sequenz Q aus einer Duplikation entsteht. Die Untersuchung der entsprechenden Sequenz bei anderen *Apis*-Arten weist darauf hin, daß auch die Sequenz P aus einer älteren Duplikation entstanden, und auf ein Gen in der Nachbarschaft der Leucin-tRNA beschränkt ist. Die von Crozier und Mitarbeitern durchgeführten Analysen über die Rate des Austauschs der Aminosäuren in der Evolution der Gene CO-I und CO-II weisen darauf hin, daß diese Rate in der zu *Apis* führenden Linie 2- bis 3-mal so hoch ist wie in der zu *Drosophila* führenden Linie. Aber diese Unterschiede sind Mittelwerte, die eine Zeitspanne von 280 Millionen Jahren betreffen, welche diese beiden Linien von ihrem gemeinsamen Vorfahren trennen; das bedeutet also nicht, daß diese beschleunigte Evolution notwendigerweise auch gegenwärtig stattfindet. Ein phylogenetischer Baum wurde auf grund einer Sequenz von 651 bp für 4 *Apis*-Arten *(mellifera, cerana, dorsata und florea)* konstruiert, mit *Bombus lucorum* als externe Referenz. Die räumliche Anordnung des Baumes stimmt gut mit der von Alexander (1991) nach morphologischen und Verhaltens-Merkmalen gefundenen überein *(florea(dorsata(cerana, mellifera)))*. In der Annahme, daß die Evolutionsrate der Biene im Verlaufe der letzten Jahrmillionen mit der von *Drosophila* identisch ist, das heißt 1% pro Million Jahre, hätten sich die beiden Arten *A cerana* und *A mellifera* ungefähr vor 6 Millionen Jahren getrennt. Die Analyse der mtDNA-Variabilität der Westlichen Honigbiene ergibt sowohl nach den Daten der Restriktion oder der Sequenzen die Existenz von drei Hauptlinien, die als A, M und C bezeichnet werden; es sind dies: i) afrikanische Völker (Linie A) mit den Rassen *intermissa, adansonii, scutellata, capensis* und *monticola*; ii) Völker der Rasse *mellifera* (Linie M); iii) Völker der Rassen *ligustica, carnica* und *caucasica* (Linie C). Diese Einteilung erinnert in überraschender Weise an die drei evolutiven Zweige, die Ruttner auf grund morphometrischer Analysen abgeleitet hat. Der Hauptunterschied betrifft den Zweig M, der sich nach Ruttner von Nordafrika *(intermissa)* über die Iberische Halbinsel *(iberica)* bis nach Nordeuropa *(mellifera)* erstreckt. Nach den Befunden an den Mitochondrien ist die Rasse *intermissa* eine typisch afrikanische Rasse, und die Rasse iberica eine graduelle Mischung von afrikanischen Typen (dominierend im Süden) und *mellifera*-Typen vorherrschend im Norden. Die geographische Verteilung der gegenwärtigen Apis-Arten führt zu dem Gedanken, daß Apis *mellifera* aus Asien stammt. Von dort wären drei große Ströme ausgegangen; der eine bevölkerte Europa von Norden (Linie M), ein zweiter zog entlang der Nordküste des Mittelmeers (Linie C) und der dritte drang nach Afrika vor (Linie

A). Die Größe der Distanz zwischen den Linien nach ihren Nukleotiden läßt sich mit der Annahme einer Trennung dieser Ströme nach der letzten Vereisung schwer vereinbaren. Ausgehend von der Evolutionsrate der mtDNA von *Drosophila*, könnte sich dieses Ereignis vor 300 000 bis 1300 000 Jahren zugetragen haben.

Apis mellifera / **mitochondriale DNA / Variabilität / Phylogenese / Evolution**

ACKNOWLEDGMENTS

REFERENCES

Alexander BA (1991) Phylogenetic analysis of the genus *Apis* (Hymenoptera: Apidae). *Ann Entomol Soc Am* 84(2), 137-149

Attardi (1985) Animal mitochondrial DNA: an extreme example genetic economy. *Int Rev Cytol* 93, 93-145

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu Rev Ecol Syst* 18, 489-522

Badino G, Celebrano G, Manino A (1982) Genetic variability of *Apis mellifera ligustica* Spin in a marginal area of its geographical distribution. *Experientia* 38, 540-541

Brown WM (1983) Evolution of animal mitochondrial DNA. *In: Evolution of Genes and Proteins* (M Nei, RK Koehn, eds) 62-88

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J Mol Evol* 18, 225-239

Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of *Drosophila yakuba:* nucleotide sequence, gene organization, and genetic code. *J Mol Evol* 22, 252-271

Cornuet JM (1982) The MDH polymorphism in some west-Mediterranean honeybee populations. IUSSI congr. Boulder, CO, USA

Cornuet JM (1983) Reproduction, génétique et sélection de l'abeille. *Bull Tech Apic* 10(1), 13-36

Cornuet JM (1989) Genetic diversity in *Apis mellifera*. ESA Annu Meet, Dec 10-14, San Antonio, TX, USA

Cornuet JM, Fresnaye J (1989) Etude biométrique de colonies d'abeilles d'Espagne et du Portugal. *Apidologie* 20, 93-101

Cornuet JM, Garnery L, Solignac M (1991) Putative origin and function of the intergenic region between COI and COII of *Apis mellifera* L mitochondrial DNA. *Genetics* 1128, 393-403

Crozier RH, Crozier YC, Mackinlay AG (1989) The COI-I and CO-II region of honeybee mitochondrial DNA: evidence for variation in insect mitochondrial evolutionary rates. *Mol Biol Evol* 6(4), 399-411

DeSalle R, Freedman T, Prager EM, Wilson AC (1987) Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. *J Mol Evol* 26, 157-164

Garnery L, Vautrin D, Cornuet JM, Solignac M (1991) Phylogenetic relationships in the genus *Apis* inferred from mitochondrial DNA sequence data. *Apidologie* 22, 87-92

Hall HG, Smith DR (1991) Distinguishing African and European honeybee matrilines using amplified mitochondrial DNA. *Proc Natl Acad Sci USA* 88, 4548-4552

Harrison RG (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends Ecol Evol* 4, 6-11

Hillis DM, Moritz C (eds) (1990) *Molecular Systematics*. Sinauer, Sunderland, USA, 588p

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions. *J Mol Evol* 16, 11-120

Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc Natl Acad Sci USA* 86, 6196-6200

Koeniger N , Koeniger G, Tingek S, Mardan M, Rinderer TE (1988) Reproductive isolation by

different time of drone flight between *Apis cerana* Fabricius, 1793 and *Apis vechti* (Maa, 1953). *Apidologie* 19, 103-106

Koulianos S, Crozier RH (1991) Two ancient mitochondrial alleles in Australian honeybees. *Apidologie* 22, 621-626

Lansman RA, Shade RO, Shapira JF, Avise JC (1981) The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. III. Techniques and potential applications. *J Mol Evol* 17, 214-226

Monnerot M, Solignac M, Wolstenholme DR (1990) Discrepancy in divergence of the mitochondrial and nuclear genomes of *Drosophila teissieri* and *Drosophila yakuba*. *J Mol Evol* 30, 500-508

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu Rev Ecol Syst* 18, 269-292

Moritz RFA, Hawkins CF, Crozier RH, Mackinlay AG (1986) A mitochondrial DNA polymorphism in honeybees (*Apis mellifera* L). *Experientia* 42, 322-324

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia Univ Press, New York 512 p

Okimoto R, MacFarlane JL, Wolstenholme DR (1990) Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, *Ascaris suum* and *Caenorhabditis elegans*. *Nucleic Acids Res* 18(20), 6113-6118

Ruttner F (1988) *Biogeography and Taxonomy of Honeybees*. Springer-Verlag, Berlin, 284 p

Ruttner F, Tassencourt L, Louveaux J (1978) Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. 1 Material and methods. *Apidologie* 9, 363-381

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406-425

Sheppard WS, Berlocher SH (1989) Allozyme variation and differentiation among four *Apis* species. *Apidologie* 20, 419-431

Sheppard WS, Huettel MD (1988) Biochemical genetic markers, intraspecific variation, and population genetics of the honey bee, *Apis mellifera*. *In: Africanized Honey Bees and Bee Mites* (Needham GR, Page RE Jr, Delfinado-Baker M, and Bowman CE, eds) Ellis Horwood, Chichester, 281-286

Sheppard WS, Rinderer TE, Mazzoli JA, Stelzer JA, Shimanuki H (1991) Gene flow between African- and European-derived honey bee populations in Argentina. *Nature (Lond)* 349, 782-784

Smith DR (1988) Mitochondrial DNA polymorphisms in five Old World subspecies of honey bees and in New World hybrids. *In: Africanized Honey Bees and Bee Mites* (Needham GR, Page RE Jr, Delfinado-Baker M, Bowman CE, eds) Ellis Horwood, Chichester, 303-312

Smith DR (1991) African bees in the Americas: insights from biogeography and genetics. *Trends Ecol Evol* 6(1), 17-21

Smith DR, Brown WM (1988) Polymorphisms in mitochondrial DNA of European and Africanized honeybees (*Apis mellifera*). *Experientia* 44, 257-260

Smith DR, Brown WM (1990) Restriction endonuclease cleavage site and length polymorphism in mitochondrial DNA of *Apis mellifera mellifera* and *A m carnica* (Hymenoptera: Apidae). *Ann Entomol Soc Am* 83(1), 81-88

Smith DR, Palopoli MF, Taylor BR, Garnery L, Cornuet JM, Solignac M, Brown WM (1991) Geographical overlap of two mitochondrial genomes in Spanish honeybees (*Apis mellifera iberica*). *J Hered* 82(2), 96-100

Solignac M (1991) Preparation and visualization of mitochondrial DNA for RFLP analysis. *In: Molecular Techniques in Taxonomy* (GM Hewitt *et al*, eds) NATO ASI ser, Springer-Verlag, Berlin, vol H57

Tingek SM, Mardan M, Rinderer TE, Koeniger N, Koeniger G (1988) Rediscovery of *Apis vechti* (Maa, 1953): the Saban honey bee. *Apidologie* 19, 97-102

Vlasak IS, Burgschaiger S, Kreil G (1987) Nucleotide sequence of the large ribosomal RNA of honeybee mitochondria. *Nucleic Acids Res* 15, 2388

Wilson AC, Cann R, Carr SM, George M, Gyllenstein UB, Helm-Bychovski KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD, Stoneking M (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol J Linn Soc* 26, 375-400

Wu Y, Kuang B (1987) Two species of small honeybee - a study of the genus *Micrapis*. *Bee World* 68, 153-154