

Phylogenetic relationships in the genus *Apis* inferred from mitochondrial DNA sequence data

L Garnery¹, D Vautrin² JM Cornuet^{1*}, M Solignac²

¹INRA-CNRS (URA 1190), Laboratoire de Neurobiologie,

Comparée des Invertébrés, 91440, Bures-sur-Yvette, France;

²CNRS, Laboratoire de Biologie et Génétique Evolutive, Avenue de la Terrasse,
91198 Gif-sur-Yvette, France

(Received 11 January 1991; accepted 13 February 1991)

Summary — The nucleotide sequence of the 5' end of the mitochondrial cytochrome-oxidase subunit II gene was obtained through direct sequencing of PCR (Polymerase Chain Reaction) product from 4 *Apis* species (*mellifera*, *cerana*, *dorsata* and *florea*) and *Bombus lucorum* (Apidae) and *Xylocopa violacea* (Anthophoridae) as outgroups. Phylogenetic trees were built using Neighbor-Joining and parsimony methods. Three branches, *dorsata*, *florea* and *cerana-mellifera* diverged almost simultaneously and *cerana* separated from *mellifera* later but not as early as generally thought.

Apis/Phylogeny/molecular systematics/mitochondrial DNA/nucleotide sequence/cytochrome-oxidase

INTRODUCTION

To-date, phylogenetic relationships between the species of the genus *Apis* are mainly based on morphological, eco-ethological and biogeographic considerations (Ruttner, 1988). Recently, a phylogenetic tree has been obtained from allozyme data (Sheppard and Berlocher, 1989). In addition, preliminary reports of molecular systematics of *Apis* species have been presented by MacPherson (1989) and Sheppard (1989). In this rapid communication, we provide sequence data of the mitochondrial subunit II of the cytochrome-oxidase gene (CO-II) which

allow to infer phylogenetic relationships between *Apis cerana*, *A dorsata*, *A florea* and *A mellifera*.

MATERIALS AND METHODS

Origin of samples

Individuals of *A cerana*, *A dorsata* and *A florea* were sampled in Faridpur (Bangladesh) during January 1990. *A cerana* and *A dorsata* were taken from feral nests and workers of *A florea* were caught when they were feeding on palm sugar. *A mellifera* (subspecies *mellifera*) is represented by a colony from France. A second

* Correspondence and reprints.

sequence was taken from the literature (Crozier *et al*, 1989) for *A m ligustica*. Two other species have been used as an outgroup: *Bombus lucorum* (Apidae, Bombinae) and *Xylocopa violacea* (Anthophoridae, Xylocopinae). One individual of each species was caught in Bures.

Extraction of DNA

mtDNA from *Apis mellifera* was cloned in pUC8. For the other species, total DNA was extracted from single individuals (one individual per species) according to Kocher *et al* (1989). The head and thorax for the small *A florea* and the head alone for the other species were ground in 750 μ l of extraction buffer (Tris-HCl 100mmol.l⁻¹, pH 8.0, EDTA 1mmol.l⁻¹, NaCl 100mmol.l⁻¹, SDS 0.1% and DTT 50mmol.l⁻¹) containing 0.1mg of proteinase K. The homogenate was incubated for 2-4 h at 37°C. After centrifugation (5 min at 1 500 rpm), the supernatant was deproteinated by 2 successive phenol-chloroform extractions. Nucleic acids were subsequently ethanol precipitated and dissolved in 1 ml of TE (10 mmol.l⁻¹ Tris, 1 mmol.l⁻¹ EDTA, pH 8.0).

Sequencing

The DNA template was produced through polymerase chain reaction (PCR) with the primers: E2: 5'-GGCAGAATAAGTGCATTG-3' located in the tRNA^{leu} gene and H2: 5'-CAATATCATTGATGACC-3', 300 bp downstream from the 5' end of the CO-II gene (Cornuet *et al*, in press). Single-stranded DNA was obtained according to Higuchi and Ochman (1989): amplification was achieved with one kinased primer, the other being unchanged. After amplification, the strand with the 5'-terminal phosphate was specifically digested with lambda exonuclease. Sequencing was performed through the dideoxy chain-termination method (Sanger *et al*, 1977) with the Pharmacia sequencing kit. Both strands were sequenced for all species.

Phylogeny reconstruction

The Neighbor-Joining (N-J) algorithm (Saitou and Nei, 1987) applied to corrected nucleotide

distances (Kimura, 1980) and the parsimony method (program DNAPARS from the package PHYLIP v3.01) were used to infer phylogenetic trees.

RESULTS AND DISCUSSION

In the CO-II gene of each species, a region of 269 nucleotides was sequenced of which 93 were polymorphic and 41 informative (fig 1). The number of substitutions observed in pairwise comparisons and the corresponding distance are given in table I. The tree obtained through the N-J method indicates that the ancestor species diverged in 2 lineages giving rise to single comb open air species (*dorsata* and *florea*) and multiple comb cavity nesting species (*mellifera* and *cerana*) respectively (fig 2A). *A dorsata* and *A florea* then diverged rapidly, the separation of *A mellifera* and *A cerana* occurring later. The DNAPARS program provides 2 equiparsimonious trees. The first one has the same topology as above (fig 2A) and the second one infers a phylogeny in which *florea* diverged first, followed by *dorsata* and finally *cerana* and *mellifera* (fig 2B). Because of this discrepancy, a consensus tree is presented (fig 2C) in which *dorsata* and *florea* branch at the same level as the *mellifera-cerana* lineage. Solving this trifurcation would require far larger sequences because the level of homoplasy is high in mtDNA owing to the low number of sites able to substitute, the number of sites able to make multiple substitutions at these sites and to the simplified pattern of substitutions dominated by transitions and A<->T transversions.

The early separation of *A florea*, hypothesized by Michener (1974), is supported by a cladistic analysis of 19 morphological characters (Alexander, personal communication) and by the allozyme study of Sheppard and Berlocher (1989). Our sequence data, which are not opposed to this possi-

```

A. mel. lig.  ATTGCCACATGATTATATTTATATTTCAAGAATCAAATTCATATTATGCTGATAAATTTAATTTTCATATTCATAATATAGTTATAAATAAT
A. mel. mel.  ...T.C.....TTC.....TA.A.....G.A..A..T..ATAT..TG.T.....T.AA.....T.....A..GTT....T...C
A. cerana    ...T.T.....TTC.....TA.A.....G.A..A..T..ATAT..CG.T.....T.AA.....C.....A..GTA....T...T
A. dorsata   ...T.A.....ATA.....AA.A.....G.A..A..T..ATTT..TG.A.....T.AA.....T.....A..GTA....C...A
A. florea    ...T.T.....ATA.....TA.A.....G.A..A..C..ATTT..TG.T.....T.AG.....C.....T..GTA....T...T
B. lucorum   ...T.T.....AAT.....TT.A.....G.T..A..T..TTTT..TT.T.....C.TA.....T.....T..ACA....T...A
X. violacea  ...G.A.....CAA.....TT.T.....A.T..T..T..TCCA..TT.A.....T.AA.....T.....T..GTT....T...T

```



```

A. mel. lig.  ATTATTATAAATTTCAACATTAACCTGATATATTATTTTAGATTATTATAAACAATTCCTCAAATTTATTTTATTAATAAAATCATAAT
A. mel. mel.  ..TA.T.....TCA.CAT.A.CTG.ATATA.T..TT.AG..T.A.TTA...C..A..TTCA...TT.TT.T.A.....AT
A. cerana    ..TA.T.....TCT.CAT.A.CAG.TTATA.T..TA.AG..C.T.TTT...T..A..TTCT...TT.TT.T.A.....AT
A. dorsata   ..TG.A.....TCA.CAT.A.CAA.TTATA.C..TA.AG..C.A.TCA...T..A..TTCA...TT.TT.T.A.....AT
A. florea    ..TA.T.....TCA.CAC.T.CAA.TTATA.T..TT.TG..T.A.TCA...T..A..CTCC...TT.TT.T.A.....AT
B. lucorum   ..AA.A.....ATT.CTT.A.CTA.ATTTT.C..TA.AG..T.T.CAA...T..T..TTTA...TT.AA.C.T.....CT
X. violacea  ..AA.T.....ATA.TTA.A.TTA.TATAA.T..AT.AA..T.A.TTA...T..A..TACA...CG.TT.T.A.....CA

```



```

A. mel. lig.  ATTGAAATATTGGAACAATTTCCAATTATTATTTCTATTAATTATTGTTTTCCATCAITAAAAATTTATATTAAATGATGAAAT
A. mel. mel.  .....G.TATT..A..TA.TA..C.AT..A.T.....T.....A.....TA.....TGA...
A. cerana    .....G.AATC..A..TA.TA..T.AT..A.T.....T.....A.....TA.....TGA...
A. dorsata   .....A.TATT..T..TT.TG..C.TT..A.A.....T.....A.....TA.....TGA...
A. florea    .....A.TGTT..T..TG.AA..T.AT..A.T.....T.....A.....TA.....TGA...
B. lucorum   .....T.AACT..C..AA.AA..T.AA..A.T.....C.....T.....TT.....ATT...
X. violacea  .....A.TATT..T..AA.TA..T.AA..T.A.....T.....A.....AT.....TGA...

```

Fig 1. Nucleotide sequence of the 5' end of the mitochondrial DNA cytochrome-oxidase subunit II (COII) gene for four *Apis* species (Apidae), *Bombus lucorum* (Apidae) and *Xylocopa violacea* (Anthophoridae). The first line is the reference sequence of *Apis mellifera linguistica* (Crozier *et al*, 1989). Dots represent nucleotides common to all species.

Table I. Distances between taxa: number of observed substitutions between sequence and nucleotide distances with Kimura's correction (between parentheses).

A m ligustica	A m mellifera	A cerana	A dorsata	A florea	B lucorum
<i>A mellifera mellifera</i>					
4					
(0.0151 ± 0.0076)					
<i>A cerana</i>					
19	17				
(0.0744 ± 0.0173)		(0.00661 ± 0.0162)			
<i>A dorsata</i>					
25	25	27			
(0.0992 ± 0.0202)		(0.0992 ± 0.0202)		(0.1121 ± 0.0216)	
<i>A florea</i>					
23	26	26	26		
(0.0909 ± 0.0193)		(0.1039 ± 0.0208)		(0.1037 ± 0.0208) 0.1036 ± 0.0207)	
<i>B lucorum</i>					
52	52	52	54	55	
(0.2236 ± 0.0325)		(0.2236 ± 0.325)		(0.2236 ± 0.0324) (0.2338 ± 0.0334) (0.2387 ± 0.0338)	
<i>X violacea</i>					
46	47	55	52	51	56
(0.1943 ± 0.0298)		(0.1990 ± 0.0302)		(0.2391 ± 0.0339) (0.2186 ± 0.0320) (0.2186 ± 0.0320) (0.2439 ± 0.0343)	

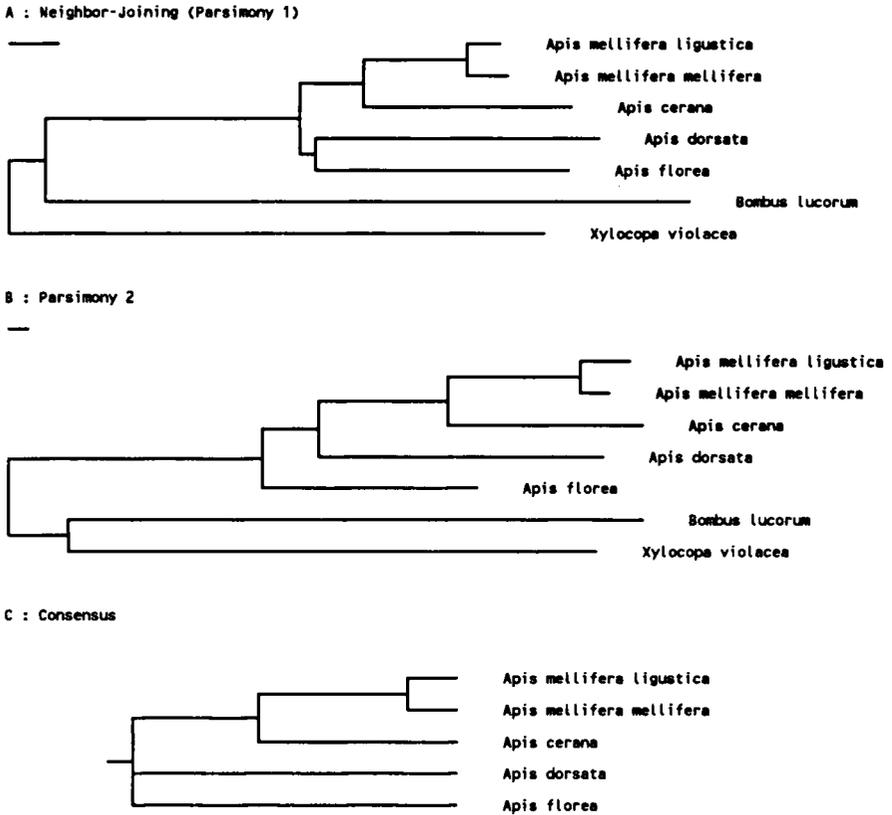


Fig 2. Phylogenetic trees obtained through the Neighbor-Joining algorithm (A) and parsimony method (B) and a consensus tree (C). In A and B, the scale is indicated under the title: it corresponds to 1% divergence (A) and 1 substitution (B) respectively. The parsimony method gave two equiparsimonious trees, one of which had the same topology as the tree shown in A.

bility, indicate that *A florea* and *A dorsata* diverged almost simultaneously of the cluster *A mellifera* - *A cerana*, as considered by Ruttner (1988, fig 3.11).

Considering now the separation between *A cerana* and *A mellifera*, the allozyme study of Sheppard and Berlocher, by dating this event at the time of divergence of *dorsata* or slightly later casts doubt on the old and still widely accepted hypothe-

sis that both species separated quite recently (Culliney, 1983; Ruttner, 1988). An intermediate situation is obtained with our present data: the divergence of *A cerana* and *A mellifera* is clearly posterior to the *dorsata*-(*mellifera/cerana*) node but the nucleotide distance between them is hardly compatible with a quaternary speciation.

Nucleotide distances between the different molecules can be used to calculate dis-

tance between species and then the absolute time of divergence, provided the rate of evolution is known (Wilson *et al.*, 1985). This rate is not known in *Apis*, but it has been found close to 2% per million year (mya) for the whole mtDNA genome in Diptera as well as in vertebrates (De Salle *et al.*, 1987). The rate of evolution of the sequenced CO-II region is about half that of the whole genome (a similar ratio is found in *Drosophila* sequenced mtDNA).

The nucleotide distance between the CO-II region of *mellifera* and *cerana* is 7%; the corresponding value for the whole genome would be 14%; the nucleotide diversity of the ancestral species is assumed to be the same as in *A. mellifera*, ie 2% (unpublished results) and the rate of divergence 2% per mya. Their absolute time of speciation is then: $(14-2)/2 = 6$ mya. A similar computation would date the trifurcation *florea-dorsata-cerana/mellifera* around -9 mya. These time values should be considered as underestimates, owing to a rapid saturation effect for substitutions in mtDNA (De Salle *et al.*, 1987).

The trees have been rooted by way of 2 external species (*Bombus lucorum* and *Xylocopa violacea*), the relative position of which is irrelevant.

Résumé - Relations phylogénétiques dans le genre *Apis* déduites de données de séquences d'ADN mitochondrial. La séquence nucléotidique de l'extrémité 5' du gène de la sous-unité II de la cytochrome-oxydase mitochondriale a été obtenue par séquençage direct du produit de PCR (Polymerase chain reaction) de 4 espèces d'*Apis* (*mellifera*, *cerana*, *dorsata* et *florea*) et de *Bombus lucorum* (Apidae) et *Xylocopa violacea* (Anthophoridae) comme références externes. Des arbres phylogénétiques ont été construits à l'aide de la méthode du Neighbor-Joining

et d'une méthode de parcimonie. L'arbre consensus indique que les 3 rameaux *florea*, *dorsata* et *mellifera-cerana* ont divergé à peu près à la même époque et que *mellifera* et *cerana* ont divergé plus tard, mais pas aussi récemment qu'on le croyait généralement.

***Apis* / phylogénèse / taxonomie moléculaire / ADN mitochondrial / séquence de nucléotides / cytochrome oxydase**

Zusammenfassung — Versuch der Konstruktion eines phylogenetischen Baums aus den Sequenz-Daten der mitochondrialen DNA von vier *Apis*-Arten. Die Nukleotid-Sequenz des 5' Endes des mitochondrialen Gens der Untereinheit II der Cytochromoxydase wurde durch direkte Sequenzierung des PCR-Produkts (Polymerase Chain Reaktion) von 4 *Apis*-Arten (*mellifera*, *cerana*, *dorsata* und *florea*) und von *Bombus lucorum* und *Xylocopa violacea* als Außengruppen gewonnen. Mit Hilfe der Neighbor-Joining und einer Parsimony-Methode wurden phylogenetische Bäume konstruiert. Die Verzweigung *dorsata*, *florea* und *cerana-mellifera* erfolgte offenbar beinahe gleichzeitig und *mellifera* trennte sich von *cerana* erst später, aber nicht erst in jüngster Vergangenheit wie allgemein angenommen wird.

***Apis* / Phylogenesis / molekulare Systematik / mitochondriale DNA / Nucleotid-Sequenz / Cytochromoxydase**

REFERENCES

- Cornuet JM, Garnery L, Solignac M (1991) Putative origin and function of the intergenic region between CO-I and CO-II of *Apis mellifera* L mitochondrial DNA. *Genetics* (in press)
- Crozier RH, Crozier YC, McKinlay AG (1989) The CO-I and CO-II region of honeybee mitochondrial DNA: evidence for variation in in-

- sect mitochondrial rates. *Mol Biol Evol* 6, 399-411
- Culliney TW (1983) Origin and evolutionary history of the honey bees *Apis*. *Bee World* 64, 29-38
- 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
- Higuchi RG, Ochman H (1989) Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. *Nucleic Acids Res* 17, 5865
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions. *J Mol Evol* 16, 111-120
- 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
- McPherson BA (1989) Insect molecular systematics based on sequence data from several genes. ESA Annual meeting Dec 10-14, San Antonio, Texas, USA
- Michener CD (1974) *The social behavior of the Bees*. Harvard University Press, Cambridge, Massachusetts, 404 p
- Ruttner F (1988) *Biogeography and Taxonomy of Honeybees*. Springer-Verlag, Berlin Heidelberg, 284 p
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminator inhibitors. *Proc Natl Acad Sci USA* 74, 5436-5437
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406-425
- Sheppard WS (1989) Ribosomal RNA diversity in *Apis*. ESA Annual meeting Dec 10-14, San Antonio, Texas, USA
- Sheppard WS, Berlocher SH (1989) Allozyme variation and differentiation among four *Apis* species. *Apidologie* 20, 419-431
- Wilson AC, Cann RL, Carr SM, George M, Gyllenstein UB, Helm-bychowski, 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