DNA Sequence, 2002 Vol. 13 (2), pp. 123–127

Taylor & Francis
healthsciences

Short Communication

Complete Mitochondrial DNA Sequence and Amino Acid Analysis of the Cytochrome C Oxidase Subunit I (*COI*) from *Aedes aegypti*

I. MORLAIS and D.W. SEVERSON*

Department of Biological Sciences, Center for Tropical Disease Research and Training, University of Notre Dame, Notre Dame, IN 46556, USA

(Received 7 November 2001)

The complete sequence of the yellow fever mosquito, *Aedes aegypti*, mitochondrial cytochrome c oxidase subunit 1 gene has been identified. The nucleotide sequence codes for a 512 amino acid peptide. The *AeCOI* sequence is A + T rich (68.6%) and the codon usage is highly biased toward a preference for A- or T-ending triplets. The *A. aegypti COI* peptide shows high homology, up to 93% identity, with several other insect sequences and a phylogenetic analysis indicates that the *A. aegypti* sequence is closely related to two other mosquito species, *Anopheles gambiae* and *A. quadrimaculatus*. Comparisons of the nucleotide sequence for four *A. aegypti* laboratory strains revealed single nucleotide polymorphisms, with 25 nucleotide sites showing SNPs between strains. All SNPs occurred as synonymous transitions such that the peptide sequence is conserved among *A. aegypti* strains. RT-PCR analysis showed that *COI* is expressed at similar levels in all developmental stages and tissues.

Keywords: *Aedes aegypti*; *COI*; mtDNA; SNP; Insect

The nucleotide sequences reported in this paper have been submitted to GenBank and have been assigned the accession numbers: Red-eye *AeCOI*, AF390098; Liverpool *AeCOI*, AY056596; Formosus *AeCOI*, AY056597 and Moyo-R *AeCOI*, AF380835.

Comparative studies of mitochondrial DNA (mtDNA) among different groups have revealed an overall well conserved organization across metazoa but significant differences also exist. For example, compared to vertebrate mtDNA, insect mtDNA is much more A + T rich (Clary and Wolstenholme, 1985; Crozier and Crozier, 1993; Navajas *et al.*, 1996). The study of mtDNA also has led to the discovery that several genetic codes, differing from the universal code, are used in animal mitochondria that vary among vertebrates, echinoderms, insects, nematodes and cnideria (Wolstenholme, 1992). Mitochondrial genes have frequently been used as molecular markers in evolutionary studies (Lunt *et al.*, 1996; Garcia-Machado *et al.*, 1999; Stahls and Nyblom, 2000; Wu *et al.*, 2000). Indeed, mitochondrial genes are often sequences of choice for phylogenetic studies as they are (1) highly conserved among phyla, (2) maternally inherited, (3) present in high copy number and (4) because mtDNA evolves faster than nuclear DNA (Moriyama and Powell, 1997). Cytochrome c oxidase subunit 1 (*COI*) is the terminal catalyst in the mitochondrial respiratory chain and is involved in electron transport and proton translocation across the membrane (Saraste, 1990; Gennis, 1992). *COI* gene is the mitochondrial marker often used for evolutionary study because (1) it is the largest of the three mitochondrial-encoded cytochrome oxidase subunits (Clary and Wolstenholme, 1985; Beard *et al.*, 1993) and (2) the protein sequence contains highly conserved functional domains and variable regions (Saraste, 1990; Gennis, 1992).

We report here the isolation of the complete mitochondrial DNA sequence of the yellow fever mosquito, *Aedes aegypti*, cytochrome c oxidase subunit 1 gene. The *COI* gene was identified

*Corresponding author. Tel.: +1-574-631-3826. Fax: +1-574-631-7413. E-mail: david.w.severson.1@nd.edu

ISSN 1042-5179
DOI: 10.1080/10425170290030051

```
TCG CGA CAA TGG TTA TTT TCA ACA AAT CAT AAA GAT ATT GGA ACT TTA TAT TTC ATT TTT     60
 S   R   Q   W   L   F   S   T   N   H   K   D   I   G   T   L   Y   F   I   F     20

GGA GTA TGA TCT GGA ATA GTC GGA ACT TCT CTA AGA ATT TTA ATT CGT GCT GAA CTT AGC    120
 G   V   W   S   G   M   V   G   T   S   L   S   I   L   I   R   A   E   L   S     40

CAC CCT GGT ATA TTT ATT GGG AAT GAC CAA ATT TAT AAT GTA ATT GTA ACA GCT CAT GCA    180
 H   P   G   M   F   I   G   N   D   Q   I   Y   N   V   I   V   T   A   H   A     60

TTT ATT ATA ATT TTC TTT ATA GTA ATG CCA ATT ATA ATT GGA GGA TTT GGA AAT TGA TTA    240
 F   I   M   I   F   F   M   V   M   P   I   M   I   G   G   F   G   N   W   L     80

GTT CCT TTA ATA TTA GGA GCC CCT GAT ATA GCT TTC CCT CGA ATG AAT AAT ATA AGT TTT    300
 V   P   L   M   L   G   A   P   D   M   A   F   P   R   M   N   N   M   S   F    100

TGA ATA CTA CCT CCT TCA TTG ACT CTT CTA TTA TCA AGC TCA ATA GTA GAA AAT GGG GCA    360
 W   M   L   P   P   S   L   T   L   L   L   S   S   S   M   V   E   N   G   A    120

GGA ACT GGG TGA ACA GTT TAT CCT CCT CTC TCT TCA GGA ACA GCT CAT GCT GGA GCT TCT    420
 G   T   G   W   T   V   Y   P   P   L   S   S   G   T   A   H   A   G   A   S    140

GTT GAT TTA GCT ATT TTT TCT CTT CAT TTA GCT GGA ATT TCC TCA ATT TTA GGG GCA GTA    480
 V   D   L   A   I   F   S   L   H   L   A   G   I   S   S   I   L   G   A   V    160

AAT TTT ATT ACA ACT GTG ATT AAT ATG CGA TCG TCA GGG ATT ACT TTA GAT CGA CTA CCC    540
 N   F   I   T   T   V   I   N   M   R   S   S   G   I   T   L   D   R   L   P    180

TTA TTT GTT TGA TCT GTA GTT ATT ACA GCT ATC TTA TTA CTT CTT TCT CTT CCT GTT TTA    600
 L   F   V   W   S   V   V   I   T   A   I   L   L   L   L   S   L   P   V   L    200

GCT GGA GCT ATT ACT ATA TTA TTA ACA GAC CGA AAC TTA AAT ACA TCT TTC TTT GAT CCA    660
 A   G   A   I   T   M   L   L   T   D   R   N   L   N   T   S   F   F   D   P    220

ATC GGA GGG GGA GAC CCT ATT TTA TAC CAA CAC TTA TTT TGA TTC TTT GGA CAC CCA GAA    720
 I   G   G   G   D   P   I   L   Y   Q   H   L   F   W   F   F   G   H   P   E    240

GTT TAT ATT TTA ATT TTA CCC GGA TTT GGA ATA ATT TCT CAT ATT ATT ACT CAA GAA AGC    780
 V   Y   I   L   I   L   P   G   F   G   M   I   S   H   I   I   T   Q   E   S    260

GGA AAA AAG GAA ACA TTT GGA ACT TTA GGA ATA ATT TAT GCT ATA TTA ACA ATT GGA TTA    840
 G   K   K   E   T   F   G   T   L   G   M   I   Y   A   M   L   T   I   G   L    280

TTG GGA TTT ATT GTT TGA GCT CAT CAT ATA TTT ACA GTA GGT ATA GAC GTA GAT ACT CGA    900
 L   G   F   I   V   W   A   H   H   M   F   T   V   G   M   D   V   D   T   R    300

GCT TAT TTT ACT TCA GCA ACT ATA ATT ATT GCT GTT CCT ACA GGA ATT AAA ATT TTT AGT    960
 A   Y   F   T   S   A   T   M   I   I   A   V   P   T   G   I   K   I   F   S    320

TGA TTA GCA ACT TTA CAC GGA ACT CAA TTA ACA TAT AGT CCA GCC CTT CTA TGA TCA TTA   1020
 W   L   A   T   L   H   G   T   Q   L   T   Y   S   P   A   L   L   W   S   L    340

GGA TTT GTA TTT TTA TTT ACA GTT GGA GGT TTA ACA GGA GTA GTA TTA GCT AAT TCT TCA   1080
 G   F   V   F   L   F   T   V   G   G   L   T   G   V   V   L   A   N   S   S    360

ATT GAT ATT GTT CTT CAT GAT ACT TAT TAC GTA GTT GCC CAT TTT CAT TAC GTT TTA TCT   1140
 I   D   I   V   L   H   D   T   Y   Y   V   V   A   H   F   H   Y   V   L   S    380

ATA GGA GCT GTA TTT GCT ATT ATA GCA GGA TTT ATT CAT TGA TAC CCT TTA TTA ACA GGA   1200
 M   G   A   V   F   A   I   M   A   G   F   I   H   W   Y   P   L   L   T   G    400

ATA GTT ATA AAC CCT TCA TGA TTA AAG GCT CAA TTT AGT ATA ATA TTT ATT GGA GTA AAT   1260
 M   V   M   N   P   S   W   L   K   A   Q   F   S   M   M   F   I   G   V   N    420

CTA ACT TTC TTT CCT CAA CAT TTT TTA GGG TTA GCT GGA ATA CCT CGA CGA TAC TCA GAT   1320
 L   T   F   F   P   Q   H   F   L   G   L   A   G   M   P   R   R   Y   S   D    440

TTT CCT GAT AGC TAC TTA ACT TGA AAT ATT ATT TCT TCT TTA GGA AGA ACA ATT TCA CTA   1380
 F   P   D   S   Y   L   T   W   N   I   I   S   S   L   G   S   T   I   S   L    460

TTT GCC GTT ATT TTC TTT TTA TTT ATT ATT TGA GAA AGT ATA ATT ACT CAA CGA ACA CCT   1440
 F   A   V   I   F   F   L   F   I   I   W   E   S   M   I   T   Q   R   T   P    480

TCT TTC CCT ATA CAA TTA TCT TCA TCT ATT GAA TGA TAT CAT ACA CTT CCT CCT GCA GAA   1500
 S   F   P   M   Q   L   S   S   S   I   E   W   Y   H   T   L   P   P   A   E    500

CAT ACT TAT TCA GAA CTA CCA CTA CTT TCT TCT AAT T                                 1537
 H   T   Y   S   E   L   P   L   L   S   S   N                                     512
```

FIGURE 1 Nucleotide sequence of the *Aedes aegypti* Red-eye strain cytochrome c oxidase subunit 1 gene (*AeCOI*) (GenBank accession number AF390098). The predicted amino acid sequence is shown below the nucleotide sequence; translation follows the invertebrate mitochondrial genetic code (5).

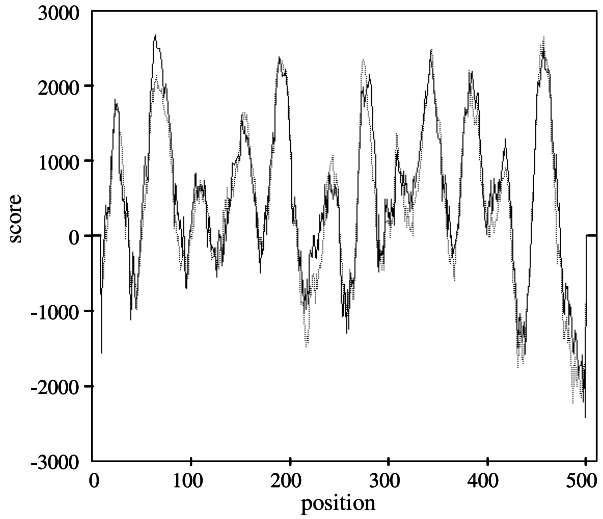

FIGURE 2 Hydropathy plot of *AeCOI*. The plot was obtained from TMPred (available at www.ch.embnet.org).

following random nucleotide sequencing of cDNA inserts from a mosquito midgut cDNA library. The preparation of the cDNA library was described elsewhere (Morlais and Severson, 2001). *A. aegypti* cytochrome c oxidase subunit 1 (*AeCOI*) is 1537 nucleotides in length and encodes a 512 amino acid peptide. The complete nucleotide sequence and its deduced amino acid sequence are shown in Fig. 1. The amino acid sequence comprises twelve transmembrane helices joined by six external loops and five internal loops (Fig. 2), in accordance with the topographical model of the *COI* protein (Saraste, 1990). In the *A. aegypti COI* gene, the putative initiation codon is TCG and the termination codon is a single T at position 1537. These suggested codons are consistent with those reported elsewhere (Beard *et al.*, 1993; Lunt *et al.*, 1996). The translation start codon in insects remains unclear, although in *Drosophila* species, the tetranucleotide ATAA is usually recognized as the start codon (de Bruijn, 1983). In *A. aegypti*, as for *Anopheles* species, the TCG is preceded by the hexanucleotide ATTTAA and it is unlikely that it serves as the initiation start. Therefore, our results agree with the suggestion by Beard *et al.* (1993), that the TCG (Ser) is used as an initiation codon in dipterans, and it seems likely that the exact start codon may vary across insect species.

TABLE II Base composition in the *AeCOI* gene at the three codon positions

Position	A	T	C	G
1st	28.52	29.69	16.41	25.39
2nd	16.99	41.60	25.78	15.63
3rd	42.58	46.29	7.62	3.52

On the nucleotide level, the A + T percentage for the *A. aegypti COI* gene is 68.6%. This high A + T content falls within the mean values of 68.1–70.7% reported for six species of nematoceran and brachyceran Diptera (Lunt *et al.*, 1996). The codon usage in *AeCOI*, as seen in Table I, clearly shows a preference for A- or T-ending codons. The base composition of mtDNA is highly correlated with codon usage, because insect mitochondrial protein genes exhibit a preference for using A + T rich codons (Crozier and Crozier, 1993). This phenomena seems characteristic of the Insecta, while the A + T content is significantly lower in crustaceans than in insects (Garcia-Machado *et al.*, 1999). Because of the codon preference, the base composition in *AeCOI* is particularly biased at the third codon position, which totaled 88.9% A + T (Table II). The A + T content at the first and second positions are 58.2 and 58.6%, respectively. These values are similar to those observed for other insects (Lunt *et al.*, 1996; Navajas *et al.*, 1996).

The *AeCOI* gene was isolated for four different laboratory strains, Red-eye, Formosus, Liverpool and Moyo-R using primers flanking the 5′- and 3′-ends (AeCOIF 5′-TATCGCCTAAACTTCAGCC-3′, AeCOIR 5′-CCTAAATTTGCTCATGTTGCC-3′). The alignment of the four nucleotide sequences revealed a total of 25 single nucleotide polymorphisms (SNPs)

TABLE I Frequency of codons in *COI* from *Aedes aegypti*. Data derived from the nucleotide sequence from the Red strain (Acc No AF390098) and follow the invertebrate mitochondrial translation table

CTT(L)	10.0(0.94)	CCT(P)	20.0(2.96)	CAT(H)	14.0(1.56)	CGT(R)	1.0(0.40)
CTC(L)	1.0(0.09)	CCC(P)	2.0(0.30)	CAC(H)	4.0(0.44)	CGC(R)	0.0(0.00)
CTA(L)	9.0(0.84)	CCA(P)	5.0(0.74)	CAA(Q)	9.0(2.00)	CGA(R)	9.0(3.60)
CTG(L)	0.0(0.00)	CCG(P)	0.0(0.00)	CAG(Q)	0.0(0.00)	CGG(R)	0.0(0.00)
ATT(I)	45.0(1.91)	ACT(T)	19.0(2.00)	AAT(N)	14.0(1.75)	AGT(S)	5.0(0.85)
ATC(I)	2.0(0.09)	ACC(T)	0.0(0.00)	AAC(N)	2.0(0.25)	AGC(S)	4.0(0.68)
ATA(M)	26.0(1.79)	ACA(T)	19.0(2.00)	AAA(K)	3.0(1.20)	AGA(S)	2.0(0.34)
ATG(M)	3.0(0.21)	ACG(T)	0.0(0.00)	AAG(K)	2.0(0.80)	AGG(S)	0.0(0.00)
GTT(V)	15.0(1.88)	GCT(A)	20.0(2.58)	GAT(D)	10.0(1.43)	GGT(G)	3.0(0.27)
GTC(V)	1.0(0.13)	GCC(A)	4.0(0.52)	GAC(D)	4.0(0.57)	GGC(G)	0.0(0.00)
GTA(V)	15.0(1.88)	GCA(A)	7.0(0.90)	GAA(E)	9.0(2.00)	GGA(G)	34.0(3.09)
GTG(V)	1.0(0.13)	GCG(A)	0.0(0.00)	GAG(E)	0.0(0.00)	GGG(G)	7.0(0.64)

Total number of codons = 512. The relative synonymous codon usage (RSCU) is given in parentheses.

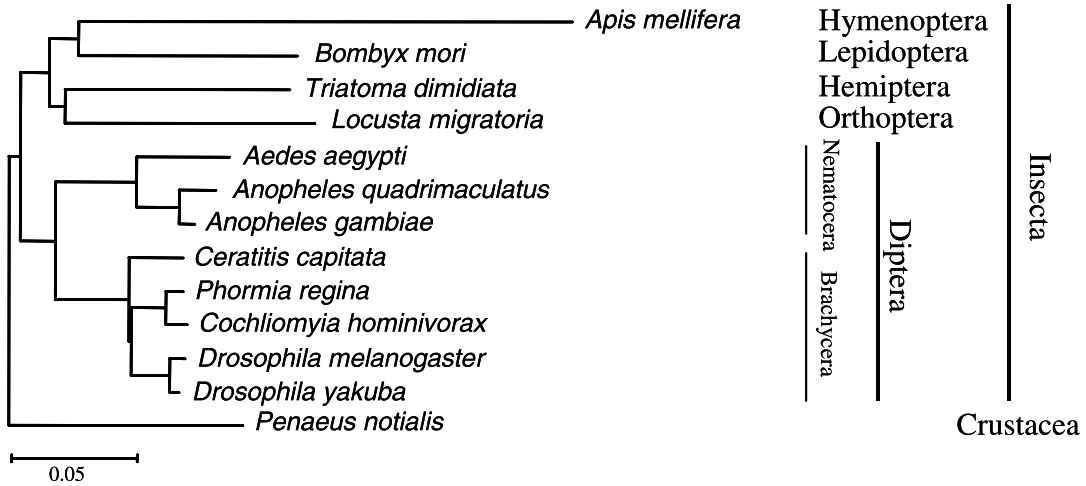

FIGURE 3 Dendrogram based on the alignment of the *COI* sequences of different insect species, constructed using Clustal W (Thompson *et al.*, 1994). Accession numbers for the different sequences are as follow: *Anopheles gambiae*, NP_008070; *Anopheles quadrimaculatus*, NP_008687; *Ceratitis capitata*, NP_008648; *Drosophila yakuba*, NP_006903; *Drosophila melanogaster*, NP_008278; Cochliomyia hominivorax, NP_075449; *Phormia regina*, AAG34088; *Locusta migratoria*, CAA56527; *Triatoma dimidiata*, AAG31621; *Apis mellifera*, AAB96799; *Bombyx mori*, NP_059476. The sequence from the crustacean *Penaeus notialis* (CAB40364) was used as the out group.

between the strains, although the nucleotide sequences of the Red-eye and the Liverpool strains are identical. All nucleotide substitutions are base transitions and occur as synonymous substitutions and, therefore, the *AeCOI* peptide is identical for the four *A. aegypti* strains (data not shown). These results are consistent with others. Higher levels of variation at synonymous sites are commonly observed, because most of these mutations are not functionally constrained (Navajas *et al.*, 1996; Stahls and Nyblom, 2000).

The *AeCOI* peptide sequence was submitted to the BLAST program (http://www.ncbi.nlm.nih.gov/) for homology searches. *AeCOI* is 93% identical to *Anopheles gambiae COI* (NP_008070), 92% to *Anopheles quadrimaculatus COI* (NP_008687) and 88% to *Drosophila yakuba COI* (NP_006903). *COI* sequences reported in the literature are numerous; therefore several sequences were selected that provide a broad representation of the different insect genera and these sequences were aligned with *AeCOI* using ClustalW (Thompson *et al.*, 1994). Sequence alignment reveals that 12 amino acid residues are unique to the *A. aegypti COI* (data not shown), 6 of which are located at the COOH-terminal region, the most variable region (Lunt *et al.*, 1996). A dendrogram was constructed based on the aligned sequences and is shown in Fig. 3. All dipteran *COI* cluster together and the phylogeny of the nematoceran and the brachyceran groups is well supported. The results are in agreement with previous studies (Lunt *et al.*, 1996; Garcia-Machado *et al.*, 1999).

AeCOI RNA expression levels were verified by RT-PCR analysis on RNA from different developmental stages (pupal, larval and adult) and from female guts and carcasses (whole body minus gut). As seen in Fig. 4, no differential expression was observed and *AeCOI* is expressed at similar levels across all stages and tissues. *A. aegypti* is an important disease vector, transmitting the yellow fever and dengue fever viruses, and has also been well studied as a model for lymphatic filariasis and malaria transmission. Investigations of *AeCOI* expression levels following pathogen infections seem warranted, as recent studies have shown that mitochondrial cytochrome c participates in the apoptotic process (Bossy-Wetzel *et al.*, 1998) and it is known that, for example, *Plasmodium* parasites, the causal agent of malaria, induce severe damage to the mosquito midgut epithelium (Han *et al.*, 2000). Indeed, *COI* enzyme activity is overexpressed in praziquantel-resistant strain of *Schistosoma mansoni* (Pereira *et al.*, 1998) and *COI* expression levels increase in the late stages of infection of *Bombyx mori* cells infected with nucleopolyhedovirus (BmNPV) (Okano *et al.*, 2001).

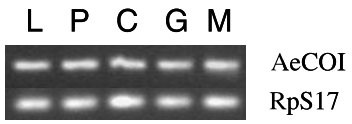

FIGURE 4 RT-PCR analysis comparing the expression level of *AeCOI* in different mosquito developmental stages and tissues; L, larvae; P, pupae; F, adult female carcasses; G, adult female guts; M, adult males. The ribosomal protein RpS17 RNA level was used as control.

References

Beard, C.B., Hamm, D.M. and Collins, F.H. (1993) "The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects", *Insect Molecular Biology* **2**, 103–124.

Bossy-Wetzel, E., Newmeyer, D.D. and Green, D.R. (1998) "Mitochondrial cytochrome c release in apoptosis occurs upstream of DEVD-specific caspase activation and independently of mitochondrial transmembrane depolarization", *EMBO Journal* **17**, 37–49.

de Bruijn, M.H. (1983) "*Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code", *Nature* **304**, 234–241.

Clary, D.O. and Wolstenholme, D.R. (1985) "The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code", *Journal of Molecular Evolution* **22**, 252–271.

Crozier, R.H. and Crozier, Y.C. (1993) "The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization", *Genetics* **133**, 97–117.

Garcia-Machado, E., Pempera, M., Dennebouy, N., Oliva-Suarez, M., Mounolou, J.C. and Monnerot, M. (1999) "Mitochondrial genes collectively suggest the paraphyly of Crustacea with respect to Insecta", *Journal of Molecular Evolution* **49**, 142–149.

Gennis, R.B. (1992) "Site-directed mutagenesis studies on subunit I of the aa3-type cytochrome c oxidase of *Rhodobacter sphaeroides*: a brief review of progress to date", *Biochimica et Biophysica Acta* **1101**, 184–187.

Han, Y.S., Thompson, J., Kafatos, F.C. and Barillas-Mury, C. (2000) "Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: the time bomb theory of ookinete invasion of mosquitoes", *EMBO Journal* **19**, 6030–6040.

Lunt, D.H., Zhang, D.X., Szymura, J.M. and Hewitt, G.M. (1996) "The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies", *Insect Molecular Biology* **5**, 153–165.

Moriyama, E.N. and Powell, J.R. (1997) "Synonymous substitution rates in Drosophila: mitochondrial versus nuclear genes", *Journal of Molecular Evolution* **45**, 378–391.

Morlais, I. and Severson, D.W. (2001) "Identification of a polymorphic mucin-like gene expressed in the midgut of the mosquito, *Aedes aegypti*, using an integrated bulked segregant and differential display analysis", *Genetics* **158**, 1125–1136.

Navajas, M., Fournier, D., Lagnel, J., Gutierrez, J. and Boursot, P. (1996) "Mitochondrial COI sequences in mites: evidence for variations in base composition", *Insect Molecular Biology* **5**, 281–285.

Okano, K., Shimada, T., Mita, K. and Maeda, S. (2001) "Comparative expressed-sequence-tag analysis of differential gene expression profiles in BmNPV-infected BmN cells", *Virology* **282**, 348–356.

Pereira, C., Fallon, P.G., Cornette, J., Capron, A., Doenhoff, M.J. and Pierce, R.J. (1998) "Alterations in cytochrome-c oxidase expression between praziquantel-resistant and susceptible strains of *Schistosoma mansoni*", *Parasitology* **117**, 63–73.

Saraste, M. (1990) "Structural features of cytochrome oxidase", *Quaterly Reviews of Biophysics* **23**, 331–366.

Stahls, G. and Nyblom, K. (2000) "Phylogenetic analysis of the genus Cheilosia (Diptera, Syrphidae) using mitochondrial COI sequence data", *Molecular Phylogenetics and Evolution* **15**, 235–241.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice", *Nucleic Acids Reseach* **22**, 4673–4680.

Wolstenholme, D.R. (1992) "Animal mitochondrial DNA: structure and evolution", *International Review of Cytology* **141**, 173–216.

Wu, W., Schmidt, T.R., Goodman, M. and Grossman, L.I. (2000) "Molecular evolution of cytochrome c oxidase subunit I in primates: is there coevolution between mitochondrial and nuclear genomes?", *Molecular Phylogenetics and Evolution* **17**, 294–304.