Insect Molecular Biology (1996) 5(4), 281–285

SHORT NOTE

Mitochondrial COI sequences in mites: evidence for variations in base composition

M. Navajas,[1,2] D. Fournier,[3] J. Lagnel,[1,2] J. Gutierrez[1] and P. Boursot[4]

[1]*Laboratoire de Zoologie, INRA-ENSAM-ORSTOM, Montpellier,* [2]*Institut des Sciences de l'Evolution (CNRS, UMR-5554), Laboratoire Génétique et Environnement, Université Montpellier II,* [3]*Laboratoire d'Entomologie, Université Paul Sabatier, Toulouse,* [4]*Laboratoire Génome et Populations (CNRS, UPR 9060), Université Montpellier II, France*

Abstract

Studies of mitochondrial DNA sequences in a variety of animals have shown important differences between phyla, including differences in the genetic codes used, and varying constraints on base composition. In that respect, little is known of mites, an important and diversified group. We sequenced a portion (340 nt) of the cytochrome oxidase subunit I (COI) encoding gene in twenty species of phytophagous mites belonging to nine genera of the two families Tetranychidae and Tenuipalpidae. The mitochondrial genetic code used in mites appeared to be the same as in insects. As is generally also the case in insects, the mite sequences were very rich in A + T (75% on average), especially at the third codon position (94%). However, important variations of base composition were observed among mite species, one of them showing as little as 69% A + T. Variations of base composition occur mostly through synonymous transitions, and do not have detectable effects on polypeptide evolution in this group.

Keywords: COI, mitochondrial genetic code, Tetranychidae, spider mites, base composition.

Received 25 March 1996; accepted 10 June 1996. Correspondence: Dr Maria Navajas, Laboratoire de Zoologie, INRA, 2 Place P. Viala, 34060 Montpellier, France.

Introduction

There has been considerable interest in mitochondrial DNA (mtDNA) in evolutionary studies and its organization and sequence have been determined in a number of animal species (Wolstenholme, 1992, and references therein). This has led to the discovery that several genetic codes differing from the so-called universal code are used in animal mitochondria, such as the code of vertebrates, echinoderms, insects, nematodes, platyhelminths and cnidaria (Osawa *et al.*, 1992; Wolstenholme, 1992). Comparative studies of different species have revealed some major trends of the mode of evolution of mtDNA in different groups (Moritz *et al.*, 1987). Although base composition can vary from one gene to another, gross overall differences exist between groups of organisms. For instance, insect mtDNA is overall much more A + T rich than vertebrate mtDNA (Clary & Wolstenholme, 1985; Crozier & Crozier, 1993; Irwin *et al.*, 1991; Jermiin & Crozier, 1994). It is not yet clear how frequent variations in base composition can be, and at what time scale they occur, and additional data are needed at various evolutionary scales.

Although mites are a highly diversified group, with 388 families described, and include many species of medical and agronomic importance, little is known of the organization and mode of evolution of their mtDNA. It is a general feature of mtDNA in animals to be extremely compact, and we have previously found the mitochondrial chromosome of the mite *Tetranychus urticae* to be among the shortest (Fournier *et al.*, 1994). In the present study, part of the mitochondrial Cytochrome Oxidase subunit I (COI) encoding gene was sequenced in twenty species of phytophagous mites belonging to the families Tetranychidae and Tenuipalpidae. These data were used to infer the mitochondrial genetic code in mites and reveal variations of their mtDNA sequence composition.

Results and Discussion

Sequence variation

A total of 340–390 base pair (bp) of the central part of the mitochondrial COI gene was sequenced from the twenty mite species listed in Table 1 (EMBL database accession numbers X74571, X79901 and X80856–X80873). The number of nucleotide differences in pairwise comparisons of species ranged from eighteen (5%) to eighty-six (25%) over the 340 nucleotide region sequenced in all species.

Base composition

Table 1 shows the base composition of the sense strand at the three codon positions in all species. The sequences are overall extremely rich in A+T (average 75%). The first codon positions have more A than T, whereas the second have more T than A. Base composition bias is particularly pronounced at the third codon position, which contains an average 94% A+T. Using Fisher's exact probability test, significant heterogeneity in base composition among species is detected at the third codon position ($P = 0.0003$), but not at the first and second positions. In fact it can be seen in Table 1 that *Eurytetranychus buxi* and, to a lesser extent, *Bryobia kissophila*, have more G+C at the third codon position than the other species. When *E. buxi* is removed from the analysis, the third codon position is no longer significantly heterogenous ($P = 0.285$), so that the higher G+C content of this species is significant, whereas that of *B. kissophila* is not.

The biases in base composition appear very similar in mites and insects, with a high proportion of A+T, specially at the third codon position. Similarly, the A+T content of mitochondrial 16S sequences is 74% in ixodid mites (Black & Piesman, 1994). A related shared characteristic between mites and insects is the rarity of C in first codon position and of G in third. However, variations around this general tendency are observed in insects. Based on known cytochrome oxydase II sequences, Jermiin & Crozier (1994) show that in insects the Endopterygota have a generally higher A+T content (~75%) than other insect divisions (~69%). However, inside the Endopterygota the honeybee *Apis mellifera* (Hymenoptera, Apidae) was found to have a higher A+T content than average (80.0%), whereas the reverse was found in one Coleoptera, Curculionidae (*Sitophilus granarius* 69.8%). On the basis of cytochrome b sequences, one ant, *Tetraponera rufonigra* (Hymenoptera, Formicidae), was found to have reduced A+T content (69.9%) when compared to the other insects studied (73.9–80.7%). Thus, provided these two different genes represent the same general trends on the whole mtDNA molecule, it appears that, among Endopterygota insects, the pressure toward high A+T content

Table 1. Base composition in a *COI* fragment of twenty mite species at the three codon positions.

Species	First				Second				Third			
	A	T	G	C	A	T	G	C	A	T	G	C
Tetranychus urticae	40.7	28.3	24.8	6.2	19.5	44.2	15.9	20.4	37.2	59.3	0.9	2.7
Tetranychus kanzawai	42.5	26.5	24.8	6.2	19.5	44.2	15.9	20.4	38.9	54.9	0.9	5.3
Tetranychus mcdanieli	41.6	26.5	24.8	7.1	19.5	44.2	15.9	20.4	41.6	57.5	0	0.9
Tetranychus pacificus	42.5	25.7	23.9	8	20.4	44.2	15.9	19.5	38.9	54.9	1.8	4.4
Tetranychus gloveri	41.6	26.5	24.8	7.1	17.7	44.2	17.7	20.4	38.9	56.6	2.7	1.8
Tetranychus neocaledonicus	41.6	26.5	24.8	7.1	19.5	43.4	15.9	21.2	42.5	52.2	3.5	1.8
Tetranychus viennensis	40.7	26.5	25.7	7.1	18.6	45.1	15.9	20.4	38.9	60.2	0.9	0
Oligonychus gossypii	40.7	27.4	25.7	6.2	18.6	45.1	15.9	20.4	44.2	50.5	1.8	3.5
Oligonychus platani	41.6	28.3	23.9	6.2	18.6	44.2	16.8	20.4	36.3	54.9	1.8	7.1
Oligonychus ununguis	41.6	27.4	23.9	7.1	18.6	44.2	16.8	20.4	31.9	61.9	3.5	2.7
Eotetranychus carpini	41.6	27.4	23.9	7.1	17.7	45.2	15.9	21.2	36.3	60.2	1.8	1.8
Eotetranychus coryli	42.5	26.5	23	8	17.7	45.2	15.9	21.2	32.7	62.8	1.8	2.7
Eotetranychus tiliarium	43.4	26.5	23.9	6.2	16.8	45.2	16.8	21.2	38.1	56.6	0.9	4.4
Panonychus ulmi	41.6	25.7	25.7	7.1	17.7	45.1	16.8	20.4	38.1	56.6	1.8	3.5
Panonychus citri	40.7	26.5	25.7	7.1	18.6	44.2	15.9	21.3	38.1	55.8	3.5	2.7
Mononychellus progresivus	40.7	29.2	23.9	6.2	18.6	44.2	15.9	21.3	35.4	59.3	0.9	4.4
Eurytetranychus buxi	38.1	29.2	26.5	6.2	15.9	46	16.8	21.3	25.7	52.2	9.7	12.4
Petrobia harti	39.8	25.7	28.3	6.2	15.9	46	15.9	22.2	43.4	55.8	0	0.9
Bryobia kissophila	33.6	30.1	30.1	6.2	15.9	45.1	16.8	22.2	31.9	55.8	5.3	7.1
Cenopalpus pulcher	33.6	35.4	23	8	15	46	16.8	22.2	29.2	68.1	1.8	0.9
Mean	40.5	27.6	25	6.9	18	44.8	16.3	20.9	36.9	57.3	2.3	3.5
Standard deviation	2.5	2.2	1.7	0.6	1.4	0.7	0.5	0.7	4.6	4.0	2.1	2.8
Base-compositional bias		0.24					0.26			0.59		

Base-compositional bias is calculated as: $B = (2/3)\sum_{i=1}^{4}|b_i - 0.25|$ where B is the compositional bias and b_i is the frequency of the ith base.

has been relaxed in at least two lineages leading to one weevil (Coleoptera) and one ant (Hymenoptera), respectively. Our data present an additional case of such a phenomenon in mites, with the A+T content of *Eurytetranychus buxi* being markedly lower (68.7%) than that of the other mites studied (74.6–77.0%). *Bryobia kissophila* also shows a lower, although not significantly so, A+T content (70.8%).

We must, however, eliminate the possibility that the change in base composition of *Eurytetranychus buxi* is due to a change, or loss, of function of the gene we sequenced. We checked that the region sequenced is not an open reading frame in the two other phases or on the other strand. Because pieces of mtDNA have been found in the nucleus of some species (Perna & Kocher, 1996, and references therein) we tried to find evidence that our sequence was not that of a (nuclear) pseudogene, by doing the following observations: (1) the sequence clearly excludes the possibility that we sequenced a mixture of two divergent copies as reported by Sunnucks & Hales (1996) in the case of cytochrome oxidase I–II in Aphids; (2) the sequence does not contain any stop codon; (3) under relaxed mutation pressure, different base positions should have their G+C content increase proportionally to their A+T content prior to the relaxation. By comparing *Eurytetranychus buxi* to the most closely related species in our survey (*Tetranychus kanzawai*), we found four α(A/T)$\rightarrow\gamma$(G/C) differences at codon positions 1 or 2 (which have 150 α pairs in *T. kanzawai*), and twenty-seven at codon position 3 (which has 106 α pairs). These unequal proportions ($P<10^{-5}$, Fisher's exact probability) can be explained if the gene is still active, so that base positions 1 and 2 are constrained because most $\alpha\rightarrow\gamma$ mutations at these positions are non-synonymous. It therefore seems unlikely that we sequenced a pseudogene in *Eurytetranychus buxi*.

It thus appears that base composition can change substantially during evolution. This has important bearings for phylogeny inference and requires the development of specific methods taking into account such variation (Galtier & Gouy, 1995). Furthermore, variations of base composition could induce variations in the rates of evolution of proteins coded by these sequences. Figure 1 enables us to examine these questions in mites. We plotted the number of nucleotide differences between *Cenopalpus pulcher*, and the other mite species against the G+C content of these latter species at the third codon position, which reflects the intensity of the mutation pressure to which the sequences are submitted. As can be seen in Fig. 1, the number of non-synonymous substitutions separating species from the outgroup (*Cenopalpus pulcher* belongs to a different family than the other species) is

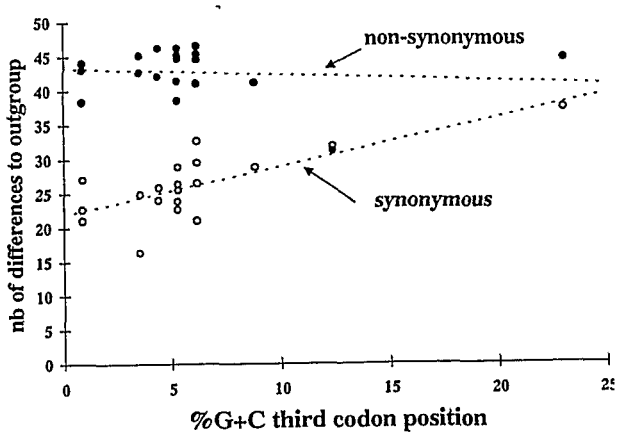


Figure 1. Plot of the number of synonymous and non-synonymous nucleotide differences between the outgroup, *Cenopalpus pulcher*, and the other studied mite species (listed in Table 1), against the G+C content of the species at the third codon position.

not correlated to their G+C % (Spearman's rank correlation coefficient $r^2=0.0022$, ns), whereas the number of synonymous substitution is ($r^2=0.645$, $P<0.01$). Therefore it appears that the relaxation of the mutation pressure that occurred in some lineages (at least that of *Eurytetranychus buxi*, cf. above) has not lasted long enough to have a detectable effect either on base composition at base positions 1 and 2 (Table 1), or on polypeptide divergence (Fig. 1). Jermiin & Crozier (1994) showed that the increased A+T content in bees as compared to *Drosophila* was accompanied by a more important polypeptide divergence from a common ancestor. However, in a broader survey, they could not detect any significant correlation between A+T content and amount of polypeptide evolution.

Genetic code

Using the genetic code of *Drosophila* mtDNA (Clary & Wolstenholme, 1985), we were able to translate our sequences without incompatibilities. However, a comparative method illustrated in Table 2 was used to check for the plausibility of this hypothesis. The table shows the tentative assignment of four codons, the significance of which varies between the different known genetic codes. It shows the nucleotide sequences found in the twenty mite species for a certain number of amino acid positions and the amino acid inferred in various other organisms. Codons AGG and CGC were not found in our sequences. The same code is also used by Crustaceans (Perez *et al.*, 1994). The genetic code of the Eu-Arthropoda (Mandibulata and Chelicerata) thus appears to be conserved.

Table 2. Inference of the mitochondrial genetic code used in mites. Numbers correspond to the nucleotide position in the sequence of the mite *Tetranychus urticae* (EMBL accession number 74571).

	ATA						TGA			AGA		AAA		
	85	97	142	157	196	307	130	235	286	43	232	58	61	223
Tetranychus urticae	ATA	ATA	ATA	ATA	ATA	ATA	TGA	TGA	TGA	AGT	AGT	AAA	AAA	AAA
Tetranychus kanzawai	ATA	ATA	ATA	ATA	ATA	ATA	TGA	TGA	TGA	AGT	AGT	AAA	AAA	AAA
Tetranychus mcdanieli	ATA	ATA	ATA	ATA	ATA	ATA	TGA	TGA	TGA	AGA	AGA	AAA	AAA	AAA
Tetranychus pacificus	ATA	ATA	ATA	ATA	ATA	ATG	TGA	TGA	TGA	AGA	AGT	AAA	AAA	AAA
Tetranychus gloveri	ATA	ATA	ATA	ATA	ATA	ATA	TGA	TGA	TGA	AGT	AGA	AAA	AAA	AAA
Tetranychus neocaledonicus	ATA	ATA	ATG	ATA	ATA	ATA	TGA	TGG	TGA	AGA	AGT	AAA	AAA	AAA
Tetranychus viennensis	ATA	ATA	ATA	ATA	ATA	ATA	TGA	TGA	TGA	AGT	AGA	AAA	AAA	AAA
Oligonychus gossypii	ATA	ATA	ATA	ATA	ATA	ATA	TGG	TGA	TGA	AGA	AGT	AAA	AAA	AAG
Oligonychus platani	ATA	ATA	ATA	ATA	ATA	ATA	TGA	TGG	TGA	AGC	AGA	AAA	AAA	AAA
Oligonychus ununguis	ATA	ATA	ATA	ATA	ATA	ATA	TGA	TGA	TGA	AGA	AGA	AAA	AAA	AAG
Eotetranychus carpini	ATA	ATA	ATA	ATA	ATA	ATA	TGG	TGA	TGA	AGA	AGT	AAA	AAA	AAA
Eotetranychus coryli	ATA	ATA	ATA	ATA	ATA	ATA	TGA	TGA	TGA	AGA	AGT	AAA	AAA	AAA
Eotetranychus tiliarium	ATA	ATA	ATA	ATA	ATA	ATA	TGA	TGA	TGA	AGT	AGT	AAA	AAA	AAA
Panonychys ulmi	ATA	ATA	ATA	ATA	ATA	ATA	TGA	TGA	TGA	AGA	AGA	AAA	AAA	AAA
Panonychs citri	ATA	ATA	ATA	ATA	ATA	ATG	TGA	TGA	TGA	AGA	AGT	AAA	AAA	AAA
Mononychellus progresivus	ATA	ATG	ATA	ATA	ATA	ATA	TGA	TGA	TGA	AGT	AGA	AAA	AAA	AAA
Eurytetranychus buxi	ATA	ATA	ATA	ATA	ATA	ATA	TGG	TGG	TGA	AGT	AGT	AAA	AAA	AAA
Petrobia harti	ATA	ATA	ATA	ATA	ATA	ATA	TGA	TGA	TGA	AGA	AGT	AAA	AAA	AAA
Bryobia kissophila	ATA	ATA	ATA	ATG	ATA	ATA	TGA	TGG	TGA	AGT	AGT	AAA	AAA	AAA
Cenopalpus pucher	ATA	ATT	ATA	TTA	ATA	ATA	TGA	TGA	TGA	AGA	AGT	AAA	AAA	AAA
Honeybee	M	M	M	L	M	L	W	W	W	M	S	K	K	K
Drosophila	M	M	M	M	M	L	W	W	W	S	S	K	K	K
Xenopus	M	M	M	M	M	L	W	W	W	T	S	K	K	K
Mouse	M	M	M	M	M	L	W	W	W	T	S	K	K	K
Caenorhabditis elegans	M	I	M	M	M	L	W	W	W	L	S	K	K	K

Experimental procedures

Amplification and sequencing

A detailed description of the origin of the biological material (listed in Table 1) is presented in Navajas *et al.* (1996). PCR primers were designed from the sequence of a fragment of *Tetranychus urticae* mtDNA COI that we previously cloned (Fournier *et al.*, 1994). The primers were 5' TGATTTTTTGGTCACCCAGAAG 3' and 5' TACAGCTCCTATAGATAAAAC 3', and were used for both DNA amplification and direct sequencing of PCR products. Protocols are described in Navajas *et al.* (1996).

Data analysis

The number of synonymous and non-synonymous substitutions between pairs of sequences were estimated using the program MEGA (v. 1.0) (Kumar *et al.*, 1993). Fisher's exact probabilities were estimated using the Markov chain method described by Guo & Thompson (1992) implemented in the Genepop (v. 1.2) software (Raymond & Rousset, 1995).

Acknowledgements

We thank O. Bonato, D. Cotton and R. Putoa for providing mites from Africa and France. This is contribution No. 96.103 of the Institut des Sciences de l'Evolution.

References

Black, W.C. and Piesman, J. (1994) Phylogeny of hard- and soft-ticks taxa (Acari: Ixodida) based on mitochondrial 16S rDNA sequences. *Proc Natl Acad Sci USA* **91**: 10034–10038.

Clary, D.O. and Wolstenholme, D.R. (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J Mol Evol* **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.

Fournier, D., Bride, J.M. and Navajas, M. (1994) Mitochondrial DNA from spider mites: isolation, restriction map and partial sequence of the Cytochrome Oxidase Subunit I gene. *Genetica* **94**: 73–75.

Galtier, N. and Gouy, M. (1995) Inferring phylogenies from DNA sequences of unequal base composition. *Proc Natl Acad Sci USA* **92**: 11317–11321.

Guo, S.W. and Thompson E.A. (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. *Biometrics* **48**: 361–372.

Irwin, D.M., Kocher, T.D. and Wilson, A.C. (1991) Evolution of the Cytochrome b gene of mammals. *J Mol Evol* **32**: 128–144.

Jermiin, L.S., and Crozier, R.H. (1994) The Cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufonigra*: sequence divergence in Hymenoptera is associated with nucleotide content. *J Mol Evol* **38**: 282–294.

Kumar, S., Tamora, K. & Nei, M. (1993) *MEGA (v.1.0): Molecular Evolutionary Genetics Analysis.* The Pennsylvania State University.

Moritz, C., Dowling, T.E. and Brown, W.M. (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu Rev Ecol Syst* **18**: 269–292.

Navajas, M., Gutierrez, J., Lagnel, J. and Boursot, P. (1996) Mitochondrial cytochrome oxidase I in tetranychid mites: a comparison between molecular phylogeny and changes of morphological and life history traits. *Bull Entomol Res* **86**: 407–417.

Osawa, S., Jukes, T.H., Watanabe, K. and Muto, A. (1992) Recent evidence for evolution of the genetic code. *Microbiol Rev* **56**: 229–264.

Perez, M.L., Valverde, J.R., Batuecas, B., Amant, F., Marco, R. and Garesse, R. (1994) Speciation in the *Artemia* genus: mitochondrial DNA analysis of bisexual and parthenogenetic brine shrimps. *J Mol Evol* **38**: 156–168.

Perna, T. and Kocher, T. (1996) Mitochondrial DNA: molecular fossils in the nucleus. *Curr Biol* **6**: 128–129.

Raymond, M. and Rousset, F. (1995) GENEPOP (v. 1.2). Population genetics software for exact tests and ecumenicism. *J Hered* **86**: 248–249.

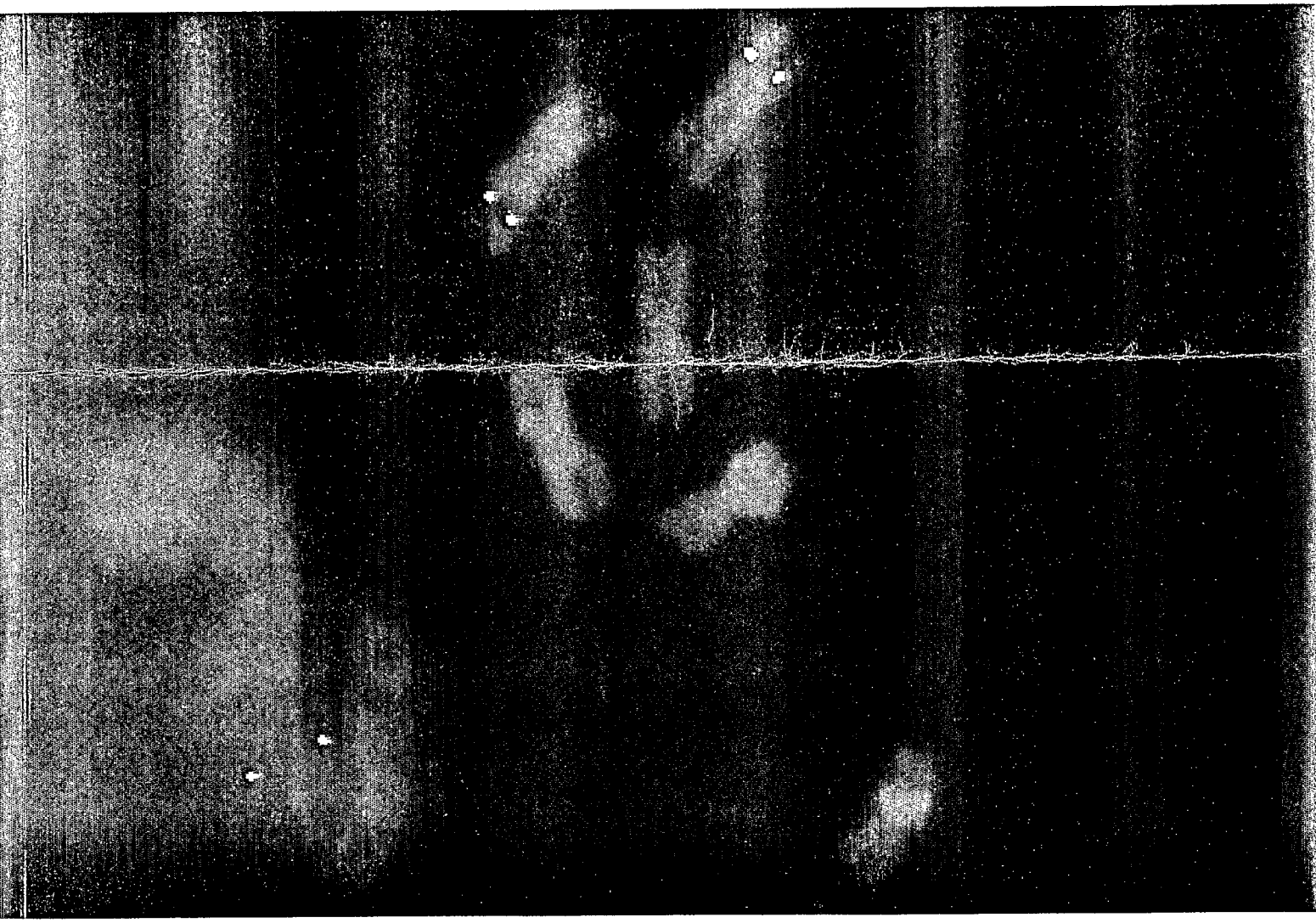
Sunnucks, P. and Hales, D.F. (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I–II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Mol Biol Evol* **13**: 510–524.

Wolstenholme, D.R. (1992) Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* **141**: 173–216.

Volume 5 • Number 4 • November 1996 CODEN IMBIE3 ISSN 0962-1075

Insect Molecular Biology

Editors: J. M. Crampton and A. A. James



Published for the Royal Entomological Society

Blackwell Science