Heredity **84** (2000) 538–547 *Received 28 June 1999, accepted 25 October 1999*

Multiple nuclear pseudogenes of mitochondrial cytochrome *b* in *Ctenomys* (Caviomorpha, Rodentia) with either great similarity to or high divergence from the true mitochondrial sequence

PATRICIA M. MIROL†, SILVIA MASCHERETTI‡ & JEREMY B. SEARLE*

Department of Biology, University of York, PO Box 373, York YO10 5YW, U.K.

A fragment of the mitochondrial cytochrome *b* gene was studied in 13 species of the South American fossorial rodent *Ctenomys* using PCR with 'universal' primers and DNA sequencing after cloning. Five different groups of sequences were found, one of which corresponds to the functional mitochondrial gene (mt). The other four groups (A, B, C and D) were believed to be nuclear pseudogenes. Sequences A–C were highly divergent from the mt sequences and included substitutions, deletions and insertions such that they could not possibly have coded a functional protein. They all shared a common insertion between positions 15055 and 15056 suggestive of a common origin, although the A, B and C sequences otherwise differed greatly from each other. The D sequences also could not have been functional on the basis of nucleotide sequence, but the differences with the mt sequences were far more subtle and in a more limited study the D sequences could easily have been classified as a true mtDNA sequence. It is suggested that there were two transfers of the cytochrome *b* gene from the mitochondrion to the nucleus; the first leading to sequences A–C and the second to the D sequence. Subsequent to transfer, a sequence of duplications within the nucleus appears to have generated the full range of pseudogenes that are observed. This study adds to other recent observations suggesting the frequent transfer of mtDNA sequences to the nucleus and reinforces the necessity of great care in interpreting PCR-generated sequences, particularly those produced with universal primers. There are now data from several species of mammals and birds relating to PCR-generated nuclear copies of cytochrome *b*, which we review.

Keywords: *Ctenomys*, cytochrome *b*, mitochondrial DNA, nuclear pseudogenes, numtDNA.

Introduction

In recent years it has become accepted that during the evolution of the eukaryotic cell, many genes were transferred from the mitochondrion (originally a free-living bacterium) to the nucleus, in accordance with the serial endosymbiosis theory (Margulis, 1970; Douglas, 1994). What has also become clear is that the transfer of genetic material between the mitochondrion and the nucleus may still occur in a diverse array of eukaryotic lineages, as demonstrated in yeast, fungi, flowering plants, sea urchins, crabs, locusts, aphids, birds, rodents, bovids, cetaceans, proboscideans, felids, primates and humans (Van Den Boogaart *et al*., 1982; Gellissen & Michaelis, 1987; Blanchard & Schmidt, 1995, 1996; Collura *et al*., 1996; Kvist *et al*., 1996; Zhang & Hewitt, 1996a; Schneider-Broussard & Neigel, 1997; Cracraft *et al*., 1998; Kidd & Friesen, 1998; Tiedemann & von Kistowski, 1998; De Woody *et al*., 1999; Greenwood & Pääbo, 1999; and references therein). With regards these recent transfers, most of the mitochondrial-like sequences within the nucleus are believed to be nonfunctional and are appropriately classified as pseudogenes (Gellissen & Michaelis, 1987; Perna & Kocher, 1996). Although

*Correspondence. E-mail: jbs3@york.ac.uk

†Present address: CIGEBA, Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, 60 y 118, CP 1900, La Plata, Buenos Aires, Argentina.

‡Present address: Mucosal Immunology Group, 1st Department of Medicine, Christian-Albrechts University, Schittenhelmstr. 12, 24105 Kiel, Germany.

examples of reverse transcriptase-derivatives of mitochondrial RNA have been documented in the nucleus (Blanchard & Schmidt, 1995, 1996), the majority of the nuclear mitochondrial-like sequences appear to derive from direct insertions of mitochondrial DNA. Presence of nontranscribed mitochondrial-like sequences in the nuclear genome have been well-demonstrated (e.g. Zullo *et al*., 1991) and in the cat there has been a single nuclear insertion of about half of the mitochondrial genome (Lopez *et al*., 1994). The escape of DNA from the mitochondrion could be caused by transient breaches of the mitochondrial membrane, either spontaneously or during organelle division (Thorsness & Fox, 1993). Mitochondria can also disintegrate sometimes (e.g. as a result of bacterial endotoxin attack) allowing the release of genetic material into the cytoplasm (Shay & Werbin, 1992). The mtDNA fragments may insert into the nuclear genome by end-joining at the position of chromosome breaks (Blanchard & Schmidt, 1996).

Many of the most recent discoveries of nuclear copies of mitochondrial genes have been made by population and evolutionary geneticists employing 'universal' primers at the PCR stage of DNA sequencing studies. These universal primers, first suggested by Kocher *et al*. (1989), will amplify mitochondrial genes in a previously unstudied species. However, because of their low specificity to the particular mitochondrial sequence being studied they may also amplify nuclear copies of the same gene (Smith *et al*., 1992; Arctander, 1995).

For our mtDNA studies investigating the phylogenetic relationship among the numerous species of the South American fossorial rodent genus *Ctenomys* (Mascheretti *et al*., in press), we used universal PCR primers to amplify a portion of the cytochrome *b* gene. These primers generated several different anomalous sequences that we believe to be nuclear pseudogenes. Our studies reinforce a concern already expressed (Smith *et al*., 1992; Arctander, 1995; Collura *et al*., 1996; Zhang & Hewitt, 1996a,b) that it is easy to confuse nuclear and mitochondrial copies of the same gene. Our studies also illustrate the potential of universal PCR primers to find nuclear pseudogenes, thereby helping in the investigation of mtDNA transfer to the nucleus and revealing genetic markers that can be useful for evolutionary studies (Perna & Kocher, 1996; Zhang & Hewitt, 1996a).

Materials and methods

Specimens

Details of the *Ctenomys* that we screened for both functional and anomalous mtDNA sequences are reported elsewhere (Mascheretti *et al*., in press). They were collected from the central-east part of the distribution of the genus, i.e. in Argentina, Paraguay and Bolivia. The functional mtDNA sequences themselves are also described in Mascheretti *et al*. (in press). In this paper we consider the anomalous sequences, making comparisons with data from the functional sequences as appropriate. The anomalous sequences have been submitted to the GenBank Data Libraries (Accession Numbers: AF210389 – AF210397, AF221718 – AF221745).

DNA extraction and sequencing

DNA was extracted from tail tissue preserved in absolute ethanol following the standard phenol/chloroform procedure (Sambrook *et al*., 1989). A fragment of the mitochondrial cytochrome *b* gene was amplified in 50 μL reactions with standard concentrations of primers (L 14841 and H 15149: Kocher *et al*., 1989), buffer, dNTPs and *Taq* polymerase. Each PCR cycle consisted of 93°C for 1 min, 50°C for 1 min and 72°C for 2 min. The cycle was repeated 35 times. PCR products were purified using QIAquick PCR purification kit (QIAGEN, Crawley, West Sussex, UK) following manufacturer's recommendations.

Following an initial demonstration that multiple sequences were produced by PCR amplification, a strategy of cloning the PCR products was adopted. They were blunt-end-cloned in a pGem vector (Promega, Madison, Wisconsin, USA), the plasmids were purified with Wizard minipreps (Promega) and sequences generated with the T7 primer. Routinely, DNA sequences were obtained from three bacterial colonies for each individual on an ABI 373 automated sequencer. If the functional mtDNA sequence was not detected in those three colonies, additional colonies were examined until it was obtained. Sequence traces were directly downloaded from the sequencer to a Macintosh LC475 and checked using Analysis software (ABI).

Phylogenetic analyses

The sequences that were produced were aligned and cut to equal lengths using SeqEd (ABI). They were then compared with the EMBL database using the FASTA routine in the Genetics Computer Group package (Wisconsin Package, 1994), to confirm their origin. Phylogenies were reconstructed using maximum parsimony (PAUP: Swofford, 1993) and distance (PHYLIP: Felsenstein, 1991) methods.

For maximum parsimony, the shortest tree(s) were found in an heuristic search with 10 replicates of 100 random stepwise addition of taxa. Where more than one minimum-length tree was produced, a strict consensus tree was calculated. Bootstrap values were based on 100 pseudoreplicates, using 10 random orders of input taxa.

A variety of weighting schemes was used: the three codon positions were unweighted or there was a 2:5:1 ratio for the first, second and third positions; transversions were unweighted or allocated five, 10 or 20 times more than transitions. Indels were unweighted.

Pairwise distances between taxa were calculated using DNADIST in the PHYLIP package, under the assumptions of the Kimura 2-parameter model. Trees were reconstructed using the neighbour-joining method (Saitou & Nei, 1987) and the SEQBOOT program was used to produce bootstrap values from 1000 pseudoreplicates of the data set. The data were either unweighted or transversions were weighted five, 10 or 20 times more than transitions.

Results

Sequence variability

The functional 260-bp cytochrome *b* sequence was detected in all 47 individuals belonging to the 21 species of *Ctenomys* that we examined (Mascheretti *et al*., in press). Studies with alternative PCR primers and a detailed phylogenetic analysis confirmed these to be the standard sequences from the mitochondrion (Mascheretti *et al*., in press). Additional anomalous sequences were obtained in 22 individuals belonging to 13 species: *C. argentinus*, *C. azarae*, *C. boliviensis*, *C. d'orbignyi*, *C. juris*, *C. latro*, *C. nattereri*, *C. occultus*, *C. opimus*, *C. pilarensis*, *C. scagliai*, *C. tuconax* and *C. tucumanus*. In the remaining 25 individuals, only the functional mtDNA sequence was found in the first three bacterial clones examined; this does not necessarily imply an absence of anomalous sequences in the individuals concerned, only that those sequences did not insert into the bacteria screened.

Four different groups of anomalous sequences were found. These showed the highest homology to the mitochondrial cytochrome *b* when compared with sequences in the EMBL data bank, but they were very peculiar in a number of aspects. They will be referred to as the A, B, C and D sequences, as opposed to the functional copy, referred to as the mt sequence. There were two types of variation among individuals: in the range of anomalous sequence groups that they were shown to carry (A, B, C or D or some combination) and in the precise nucleotide sequences recorded for each sequence group. The between-individual variation in the A, B, C and D nucleotide sequences, respectively, will be considered in the phylogenetic analysis below; it is important to note here, however, that no within-individual variation in nucleotide sequence was recorded for any of these sequence groups. With regards the range of sequence groups found in the 22 specimens for which anomalous sequences were demonstrated: 11 individuals displayed one type of anomalous sequence only; *C. latro* no. 2 had both an A and a D sequence; *C. nattereri* no. 1 and *C. tuconax* no. 2 had both a C and a D sequence; *C. pilarensis* no. 1 and *C. scagliai* no. 2 had an A, a B and a C sequence; *C. occultus* no. 2 and *C. opimus* had an A, a C and a D sequence; and *C. boliviensis*, *C. scagliai* no. 1, *C. tuconax* no. 1 and *C. tucumanus* no. 1 had a B, a C and a D sequence.

There were numerous nucleotide differences among the anomalous sequence groups and between the anomalous sequence groups and the functional mtDNA gene; these will be indicated with reference to the numbered nucleotide positions of the functional mtDNA gene (Mascheretti *et al*., in press). All anomalous sequences groups A, B and C shared the insertion of a T between positions 15055 and 15056. Sequence A was found in 11 individuals belonging to nine species: *C. argentinus*, *C. azarae*, *C. d'orbignyi*, *C. latro*, *C. nattereri*, *C. occultus*, *C. opimus*, *C. pilarensis* and *C. scagliai*. This sequence had a 10-bp deletion producing a frameshift between positions 14932 and 14941 and a 4-bp deletion between positions 15125 and 15128. Comparing all the A sequences with all the mt sequences available from the same nine species (Mascheretti *et al*., in press), and without considering nucleotides that were variable among the mt sequences, an additional 51 base substitutions were recorded in the A sequences. The nature of these substitutions was certainly unexpected from the perspective of a mitochondrial sequence: 23 changes were transitions and 28 were transversions; 13 changes occurred in the first position, nine in the second and 29 in the third.

Sequence B was found in six individuals belonging to five species: *C. boliviensis*, *C. pilarensis*, *C. scagliai*, *C. tuconax* and *C. tucumanus*. These sequences had a 10-bp deletion between positions 14946 and 14955, producing a frameshift, and a 3-bp deletion between positions 15034 and 15036. The number of additional substitutions relative to the mt sequences was 65, 30 of which were transitions and 35 were transversions. Nineteen replacements occurred in the first position of codons, 18 in the second and 28 in the third.

Sequence C was found in 10 specimens belonging to eight species: *C. boliviensis*, *C. nattereri*, *C. occultus*, *C. opimus*, *C. pilarensis*, *C. scagliai*, *C. tuconax* and *C. tucumanus*. This was the most distinct of all the sequence types found. A 7-bp deletion between positions 15031 and 15037, another 1-bp deletion in position 15087 and a 30-bp deletion between positions 14914 and 14943 were observed. *Ctenomys occultus* no. 2 also had a 3-bp deletion between positions 14987 and 14989. There was a total of 77 additional substitutions relative to the mt sequences, 32 of which were transitions and 45

transversions. The distribution of the substitutions by codon position was 24 in the first position, 19 in the second and 34 in the third.

And finally, sequence D was found in 14 specimens from nine species: *C. boliviensis*, *C. juris*, *C. latro*, *C. nattereri*, *C. occultus*, *C. opimus*, *C. scagliai*, *C. tuconax* and *C. tucumanus*. This type of sequence was particularly interesting because of its high degree of homology with the functional mt sequence, in stark contrast to sequences A–C. The 260-bp sequence did not have any insertions or deletions. However, one individual (*C. latro* no.1) had a stop codon at position 15002, which indicates that the sequence could not possibly code for a functional protein. Further evidence suggesting that sequence D is nonfunctional was obtained when one individual (*C. tucumanus* no.1) was amplified with primers L14841 (Kocher *et al.*, 1989) and H15548 (MVZ16 in Smith & Patton, 1993) to produce a longer fragment of 706 bp. This generated a standard D sequence for the 260 bp already examined, but elsewhere in the amplified fragment there were three short deletions of 4 bp (positions 15477–15480), 9 bp (15483–15491) and 2 bp (15529–15530).

There were, in general, 19 additional substitutions when sequence D was compared with the mt sequences: five transversions and 14 transitions, four in the first, one in the second and 14 in the third codon position. However, sequence D in *C. boliviensis* was rather different. It shared only 11 out of the 19 additional substitutions, and displayed three unique mutations. There were no stop codons, insertions or deletions to indicate that the sequence was functionless. Nevertheless, in the phylogenetic trees this sequence appeared closely related to all the other D sequences with high bootstrap support (see next section).

The data on additional substitutions are summarized in Fig. 1 and show the distinctiveness of the D sequences relative to the A, B and C sequences, in terms of more transitions and third-position substitutions.

Sequences A–C all contained stop codons when translated with the mitochondrial genetic code. The A and C sequences both had two stop codons, whereas sequence B had three. Sequence C also lacked a heme-ligating histidine residue at position 82, crucial for the normal function of the cytochrome *b* protein (Howell, 1989). The D sequences also coded some amino acids that would render the cytochrome *b* dysfunctional. The arginine at position 80 is mutated to cysteine in all the D sequences, and the arginine at position 100 is mutated to tryptophan in *C. occultus* and *C. juris*. The number of amino acid changes that would be produced if the anomalous sequences were translated is incredibly high: 43 for the A sequences, 48 for the B sequences, 60 for the C sequences and 11 for the D sequences (50%, 56%, 70% and 13% of the residues, respectively). Evidently, the proteins coded by these sequences could not be functional.

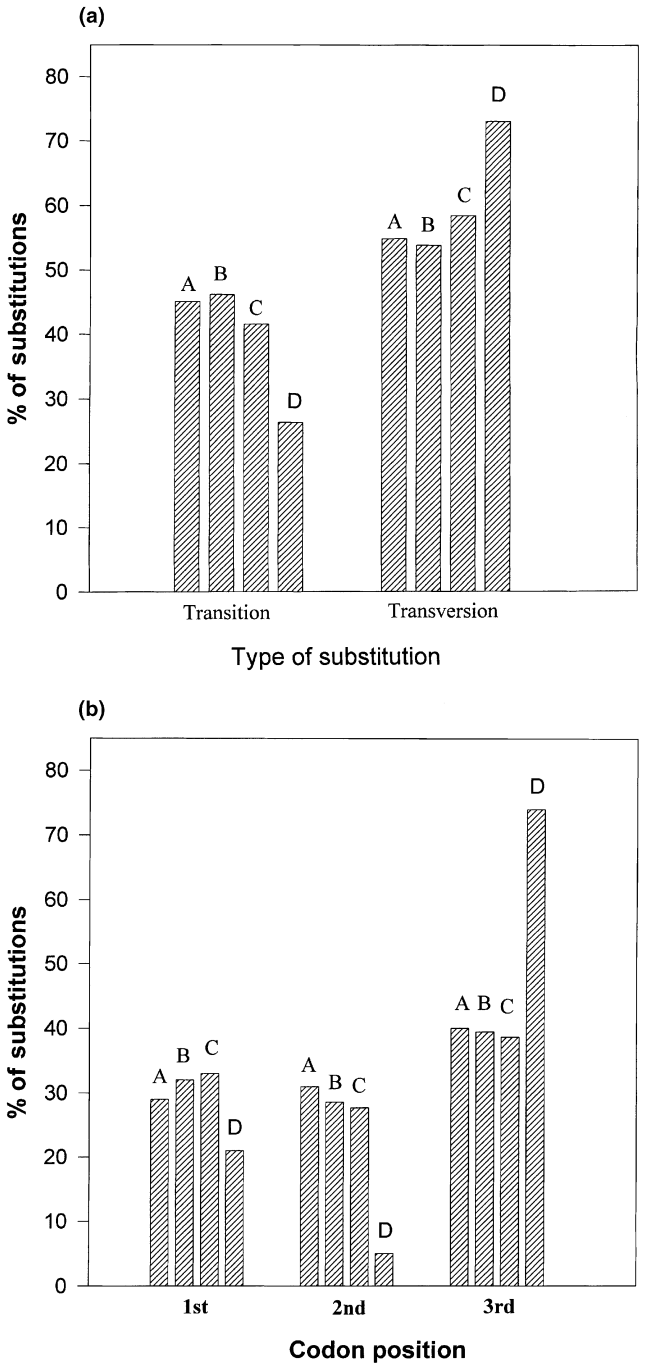

Fig. 1 Additional substitutions observed in anomalous sequences A–D of *Ctenomys* spp. after their comparison with mt sequences (see text). Details of the relative frequency of these substitutions (a) as transitions and transversions and (b) in first vs. second vs. third codon position.

Table 1 shows the nucleotide diversity (π) within and between groups of sequences. The differences between

Table 1 Nucleotide diversity (π,%) within and between the mt, A, B, C and D clades of *Ctenomys* species

	mt	A	B	C	D
Within	5.67	1.94	1.23	1.69	2.98
mt	—	26.47	32.54	35.50	14.58
A	22.67	—	29.11	38.08	28.91
B	29.09	27.52	—	38.23	34.72
C	31.82	35.27	36.77	—	32.12
D	10.25	26.44	32.62	29.79	—

For the between-clade nucleotide diversity, the raw value is shown above the diagonal.
Below the diagonal, we present the nucleotide diversity corrected for within-clade diversity ($\pi_{ABcorrected} = \pi_{AB} - [(\pi_A + \pi_B)/2]$: Nei, 1987).

the A, B, C, D and mt sequences were very high at approximately 23% to 37%, the only exception being the comparison between the D sequences and mt, which was only 10%. The values within the groups of the anomalous sequences (1.23% to 2.98%) were considerably lower than the value for the mitochondrial gene (5.67%). Considering all within-lineage substitutions, the transition:transversion ratios for the A, B, C, D and mt sequences were 6:5, 4:1 plus one variable for three nucleotides, 6:4, 6:3, 103:22, respectively. Equivalent values for the ratio of 1st:2nd:3rd position substitutions were: 3:7:1, 3:3:0, 4:3:3, 4:2:3 and 24:6:95. Over the period of diversification of the genus *Ctenomys*, the anomalous sequences have clearly accumulated a higher proportion of transversions and first and second position changes than the functional mitochondrial gene.

Phylogenetic analyses

Preliminary phylogenetic analysis (not shown) using data from the EMBL database indicated that sequences A–D grouped with mtDNA sequences of the rodent suborder Caviomorpha, but there was considerable uncertainty in the position of sequences relative to other members of the superfamily Octodontoidea to which *Ctenomys* belongs. However, the sequence for the South American porcupine *Coendou bicolor* (Lara *et al*., 1996), a caviomorph distantly related to *Ctenomys*, was basal in all phylogenies in relation to the *Ctenomys* mt, A, B, C and D sequences. Therefore, the *Coendou* sequence was used as an outgroup in phylogenies examining the relationship among the different *Ctenomys* sequences.

Figure 2a shows the strict consensus tree produced with parsimony analysis of all anomalous sequences obtained and one representative mtDNA sequence for each species (equal weighting 1st:2nd:3rd codon positions and transitions : transversions, and ignoring deletions). A near-identical distance tree was obtained with the same dataset based on the Kimura 2-parameter model (Fig. 2b). Similar trees were obtained with different weighting schemes.

The most striking features of the phylogenies are the close relationship of the mt and D sequences and the somewhat less close relationship to each other of the A and B sequences. The C sequences are particularly divergent. It should be noted, however, that deletions and insertions were ignored in tree-construction. In general, this should have little effect on the phylogeny. The indels are usually specific to either the A, B or C sequences. However, the insertion between positions 15055 and 15056 is shared by sequences A–C. If this insertion is considered to have been a single event, then sequences A–C should be 'forced' together in a single clade by, for instance, giving a very high weighting to shared indels during tree-construction. Clearly, if a distance tree was forced in this way, the branches within the A–B–C lineage would be very long, indicating the lack of similarity among the A, B and C sequences apart from the presence of the common insertion.

Another interesting feature of the phylogenies, that is particularly clear in the distance tree (Fig. 2b) is that the D sequence found in *C. boliviensis* is distinctive from all other D sequences. In other respects the relationship of sequences from different species is difficult to resolve for any particular type of anomalous sequence; presumably reflecting the short length of sequence examined. With regards the relationships of the sequences found within species, there are five cases where exactly the same anomalous sequence has been found in different individuals of the same species. For sequences A, C and D there are also cases where small differences in sequence have been found between individuals of the same species.

Discussion

Evidence that the anomalous sequences are nuclear pseudogenes

Mitochondrial-like sequences additional to the normal mitochondrial sequence could represent heteroplasmy or intramitochondrial duplications instead of nuclear pseudogenes. It is important, therefore, to rule out these alternatives in the case of the anomalous sequences in *Ctenomys*.

There has been considerable recent interest in the medical implications of heteroplasmy (Larsson & Clayton, 1995; Liang & Wong, 1998; Chinnery *et al*., 1999). Humans with cells that include a mixture of both normal mitochondria and mitochondria with deleterious mutations may develop disease if the defective mitochondrial type becomes predominant in postmitotic tissue such as skeletal muscle and neurones (Chinnery &

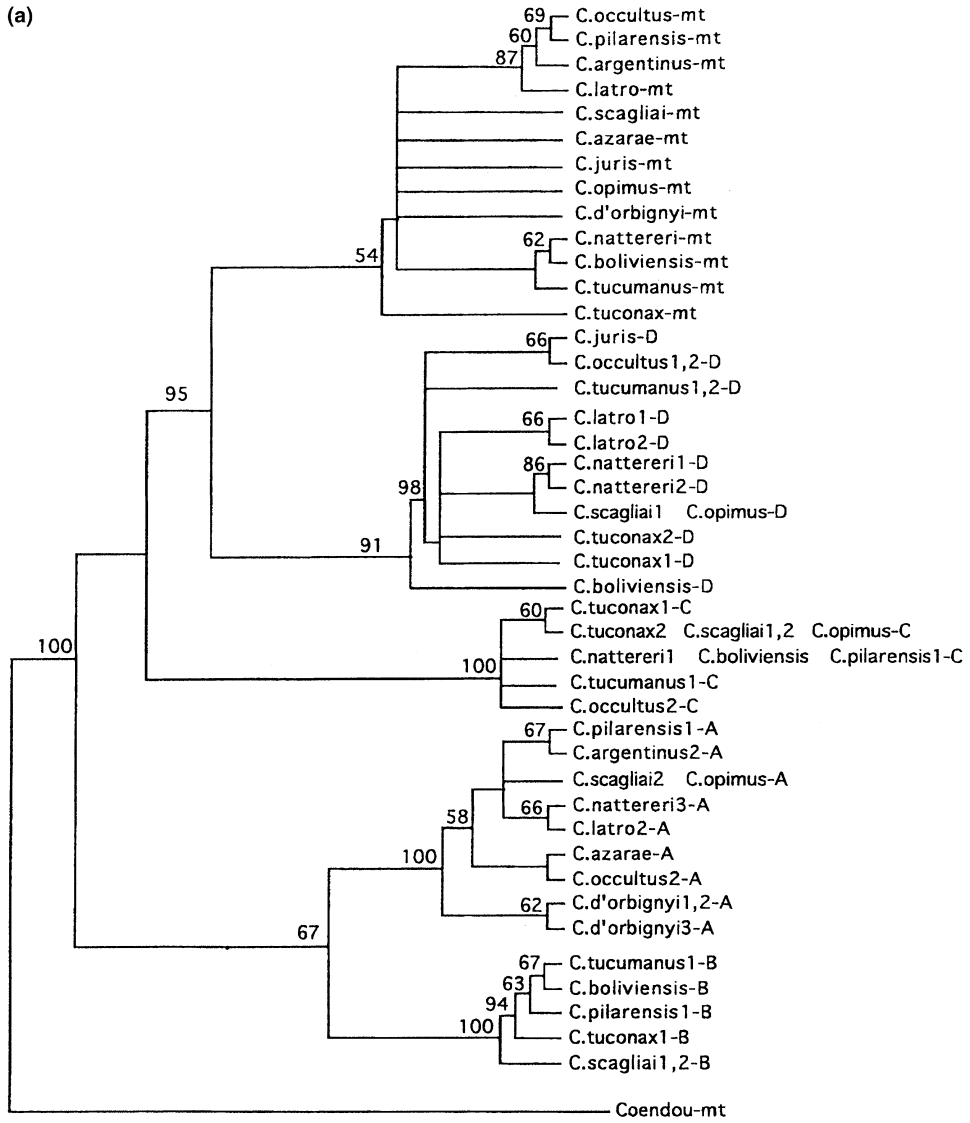

Fig. 2 Phylogenetic analysis of all 41 anomalous sequences obtained from 13 species of *Ctenomys* together with one representative mt sequence from each of those species. Note that the same sequence was sometimes found in two individuals (such that there were 30 *distinct* anomalous sequences). In construction of these trees indels were ignored and all nucleotide changes were weighted equally. The sequence for the South American porcupine *Coendou bicolor* (Lara *et al.*, 1996) was used as an outgroup. Bootstrap values (%) are indicated over branches. (a) Strict consensus of the 170 equally parsimonious trees (395 steps long, consistency index = 0.678). (b) Neighbour-joining tree. Note that in both trees all the mt sequences group into a single well-defined lineage, likewise the A, B, C and D sequences each form clear, separate lineages.

Samuels, 1999). In our studies we have also shown the presence of defective mitochondrial-like sequences in *Ctenomys*. However, we have no reason to believe that the *Ctenomys* that we examined were diseased or liable to disease. Even more pertinent, the substantial divergence between the functional and functionless cytochrome *b* sequences (e.g. Figure 2b) implies long-term survival of the anomalous sequences. This rules out heteroplasmy as an explanation for persistence of the functionless sequences as there is a general expectation that mitochondrial heteroplasmy should be short-lived (Avise, 1991; Bergstrom & Pritchard, 1998).

Persistence of the anomalous sequences as intramitochondrial duplications also appears unlikely. Three anomalous sequences have been found in each of eight different individuals. If the mitochondria of these individuals are holding those three copies plus the functional gene, the total amount of DNA would be increased considerably, which could violate constraints on the amount of DNA that the mitochondria can

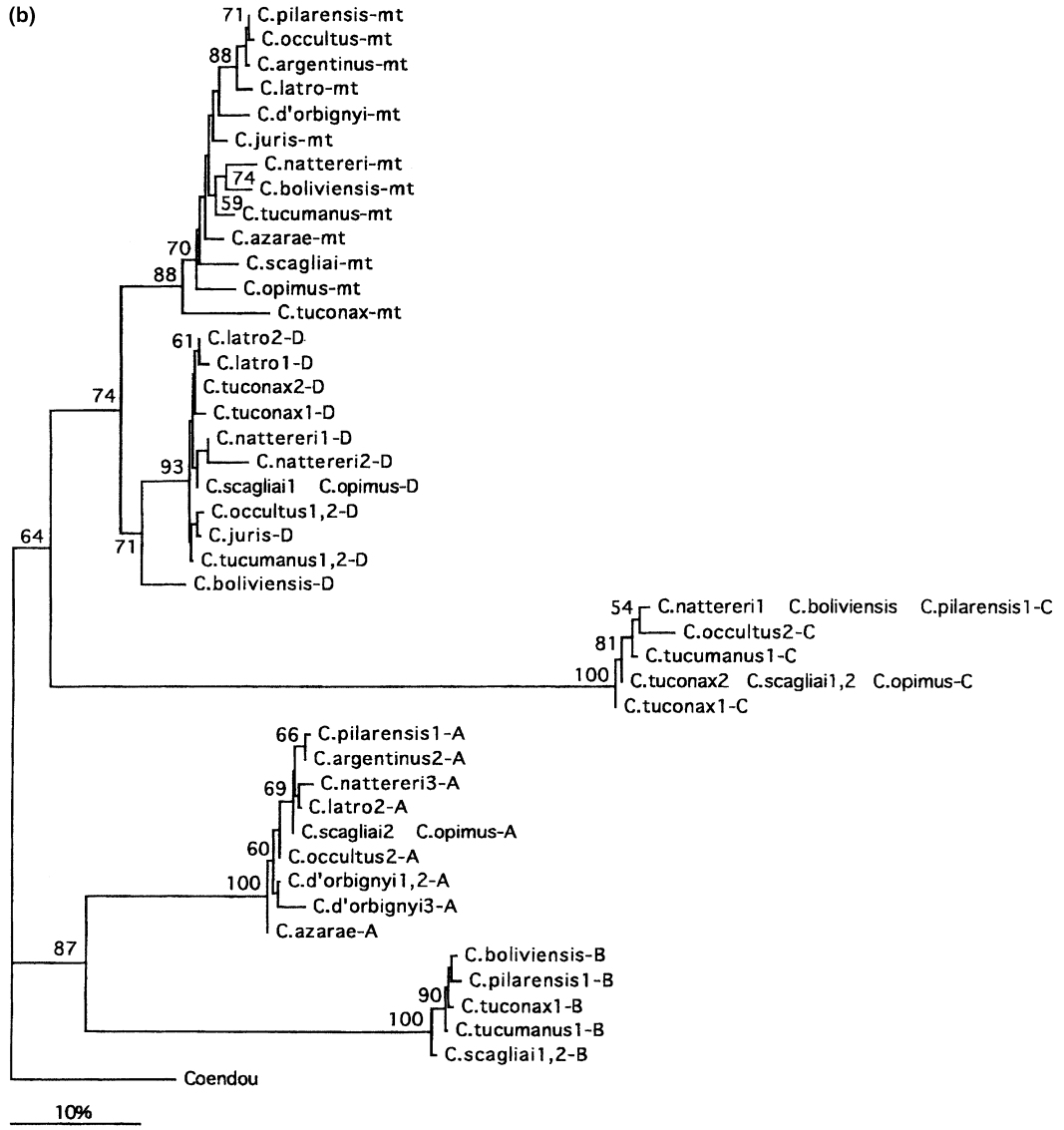

Fig. 2b (*Continued*)

tolerate to maintain efficient replication (Moritz *et al.*, 1987; Rand & Harrison, 1989).

The rate of evolution of the anomalous sequences also strongly suggests a nuclear rather than a mitochondrial localization. Within-lineage nucleotide diversity is lower for the anomalous sequences than for the mt sequence (Table 1), consistent with the expectation of a slower rate of evolution in the nucleus (Brown, 1985). A functionless gene in the mitochodrion may be expected to evolve faster than the functional version of the gene, not slower.

Therefore, there are several lines of evidence that suggest that the anomalous sequences A–D are nuclear pseudogenes, and that status will be assumed herewith.

Universal primers and the generation of pseudogenes

For the cytochrome *b* gene, there are now several examples in mammals and birds where nuclear pseudogenes have been detected using 'universal' primers. Specifically the primers L14724 and H15915 of Irwin *et al.* (1991) used to amplify the whole gene and the primers L14841 and H15149 of Kocher *et al.* (1989) for a small section of the gene have generated pseudogenes. Smith *et al.* (1992) and De Woody *et al.* (1999) detected single types of nuclear pseudogene in the rodents *Chroeomys jelskii* and *Microtus arvalis*, using L14841/H15149 and L14724/H15915, respectively. Collura & Stewart (1995) detected two nuclear versions of

cytochrome *b* in one orang-utan individual with L14724/H15915 and multiple pseudogenes were found with the same primers in proboscis monkeys (Collura *et al*., 1996). In their study of tigers, Cracraft *et al*. (1998) discovered a single nuclear version of a stretch of mtDNA that included the cytochrome *b* gene and part of the control region, again using L14724 as one of the primers.

In addition to these rather small-scale studies, Arctander (1995) used L14841/H15149 to generate a fragment of cytochrome *b* from seven individuals of three species of the South American bird genus *Scytalopus* and one individual of the closely related *Myornis senilis*, and found different anomalous sequences in each individual. However, the anomalous sequences were all closely related and formed a single class of pseudogene.

Therefore, our discovery of nuclear cytochrome *b* pseudogenes in *Ctenomys* using universal primers, and particularly the L14841/H15149 primers of Kocher *et al*. (1989), are by no means unprecedented. Using these primers we have documented four different classes of pseudogene which are all found throughout the genus *Ctenomys*. Although sequences A–D are likely to be the only common pseudogenes detectable under the particular PCR and cloning conditions that we used, this does not mean that other pseudogenes cannot be detected in *Ctenomys* with these particular universal primers. Indeed, we believe that the *C. boliviensis* class D pseudogene may in fact be a single representative of a new class of pseudogenes (see below).

Clearly, the great similarity of the D and mt sequences in *Ctenomys* provides a particularly striking illustration of the dangers of universal primers in generating nuclear pseudogenes, which may be mistaken for true mtDNA sequences. Over the 260 bp sequenced, the class D pseudogene of *C. boliviensis* showed no stop codons, insertions or deletions to indicate that the sequence was functionless. This builds on other examples from the literature. Collura *et al*. (1996) and Cracraft *et al*. (1998) provide examples in the proboscis monkey and the tiger of nuclear copies of the *whole* cytochrome *b* gene without indels or stop codons. Arctander (1995) describes a situation in birds where cytochrome *b* pseudogenes have been misinterpreted as functional mtDNA sequences.

Origin and evolution of the Ctenomys *pseudogenes*

Our study was an analysis of repeated copies of a short mitochondrial-like sequence present in the nuclear genome. The full extent of the mitochondrial integration that we sampled is unknown, as is the location of these mitochondrial-like sequences. Clearly, such information would be very informative (see Lopez *et al*., 1994), and could be used in a future study to help test some of the conjectures that we make below.

Because of the short length of sequence, we will not attempt to speculate on the timing of the transfer(s) of mtDNA sequences to the nucleus. However, given the wide range of *Ctenomys* species that carry each of the four classes of nuclear pseudogene, we propose that the transfer(s) took place either early in radiation of *Ctenomys* or previous to it. Our preliminary phylogenetic analyses indicated that transfer(s) took place within the caviomorph lineage of rodents.

To consider further the evolution of pseudogenes in *Ctenomys*, it is worth generalizing to data from other sources. Studies from a variety of species suggest that large stretches of mtDNA can integrate in an undamaged state into the nuclear genome (Blanchard & Schmidt, 1995, 1996; Collura *et al*., 1996; Cracraft *et al*., 1998). However, subsequent to integration these sequences accumulate indels and substitutions, consistent with their status as pseudogenes (Ophir & Graur, 1997). Taking together all available data on indels in cytochrome *b* nuclear pseudogenes (Table 2), it is clear that deletions are much more common than insertions and that the majority of indels observed are very short (1–4 bp). The preponderance of deletions over insertions has also been observed in nuclear-derived and -located processed pseudogenes (Ophir & Graur, 1997).

Given the rarity of insertions in cytochrome *b* nuclear pseudogenes, the common occurrence of the insertion of T between positions 155055 and 155056 in the A, B and C sequences suggests common ancestry, i.e. that these sequences derive from a single integration of mtDNA into the nuclear genome. This integrated sequence then gained the insertion and, on the basis of the phylogenies

Table 2 Frequency of indels of different size classes recorded in cytochrome *b* nuclear pseudogenes in mammals and birds

Length (bp)	Deletions	Insertions
1	8	4
2	2	0
3	7	1
4	4	1
5	0	0
6–10	5	0
11–15	1	0
16–20	2	0
21–25	0	0
26–30	1	0

Data from the present study, Smith *et al*. (1992), Arctander (1995), Kvist *et al*. (1996) and De Woody *et al*. (1999). Deletions of different size in the same location within a particular lineage are treated as independent (see text).

(Fig. 2), it duplicated, thus generating class AB and C pseudogenes. Subsequently, the class AB pseudogene would have duplicated to generate the current situation of separate class A and B pseudogenes. Given the substantial divergence between the A, B and C sequences (Fig. 2b), it may be suggested that these duplications occurred in the distant past; more sequence is needed to estimate properly the timing of these events.

Although the presence of an insertion is a useful phylogenetic marker for the *Ctenomys* cytochrome *b* pseudogenes, interpretation of the deletions that are found in these pseudogenes is more problematic. All the deletions recorded in the A, B and C sequences have, in the discussion up to now, been considered independent. However, some of these deletions are overlapping. Thus, the 30-bp deletion in the C sequences overlaps the 10-bp deletion in the A sequences. Likewise, the 7-bp deletion in the C sequences overlaps the 3-bp deletion in the B sequence. The larger deletions could represent enlargement of the smaller deletions rather than independent events. This aspect of pseudogene evolution needs to be investigated further (see Ophir & Graur, 1997).

The occurrence of duplications in the evolution of the A, B and C pseudogenes in *Ctenomys* is in line with data from other species (e.g. Zullo *et al*., 1991; Lopez *et al*., 1994; Collura & Stewart, 1995). There have been various suggestions relating to amplification of the mitochondrial sequences before, during or after integration into the nuclear genome, but no clear consensus (Zullo *et al*., 1991; Lopez *et al*., 1994; Blanchard & Schmidt, 1995, 1996; Hu & Thilly, 1995) and, indeed, one might expect such duplications to occur in a variety of different ways.

Duplication subsequent to single transfers from the mitochondrion to the nucleus is one way in which numbers of mitochondrial-like sequences in the nucleus can increase. The other way is for there to be multiple transfers of mtDNA sequences. It has been suggested, for instance, that control region sequences have been transmitted from the mitochondrion to the nucleus at least six separate times in the evolution of pochard ducks of the genera *Aythya* and *Netta* (Sorenson & Fleischer, 1996). In *Ctenomys*, we believe that the class D pseudogene represents a second transfer of a cytochrome *b* sequence from the mitochondrion to the nucleus, that occurred more recently than that which led to the A, B and C pseudogenes. The rather divergent *C. boliviensis* D pseudogene sequence could represent either a duplication of the standard class D pseudogene or an independent integration event.

Clearly, further studies of D sequences would be worthwhile, not only to elucidate the issue of number of transfers and duplications involving the cytochrome *b* in *Ctenomys*, but also because these sequences are the most likely to be useful in phylogenetic studies of the species. They could be valuable 'molecular fossils' that could be used to establish the ancestral cytochrome *b* sequence for the genus (Perna & Kocher, 1996).

Acknowledgements

We are extremely grateful to Mabel Giménez and Drs Claudio Bidau and Julio Contreras for sending the samples of *Ctenomys*. Special thanks are due to Dr David Bilton for very useful comments and discussions. This study was supported by Fundación Antorchas (Argentina) and the British Council.

References

ARCTANDER, P. 1995. Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. *Proc. R. Soc. Lond.*, **B262**, 13–19.

AVISE, J. C. 1991. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Ann. Rev. Genet.*, **25**, 45–69.

BERGSTROM, C. T. AND PRITCHARD, J. 1998. Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. *Genetics*, **149**, 2135–2146.

BLANCHARD, J. L. AND SCHMIDT, G. W. 1995. Pervasive migration of organellar DNA to the nucleus in plants. *J. Mol. Evol.*, **41**, 397–406.

BLANCHARD, J. L. AND SCHMIDT, G. W. 1996. Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Mol. Biol. Evol.*, **13**, 537–548.

BROWN, W. M. 1985. The mitochondrial genome of animals. In: Macintyre, R. J. (ed.) *Molecular Evolutionary Genetics*, pp. 95–130. Plenum Press, New York.

CHINNERY, P. F. AND SAMUELS, D. C. 1999. Relaxed replication of mtDNA: a model with implications for the expression of disease. *Am. J. Hum. Genet.*, **64**, 1158–1165.

CHINNERY, P. F., ZWIJNENBURG, P. J. G., WALKER, M., HOWELL, N., TAYLOR, R. W., LIGHTOWLERS, R. N. *ET AL.* 1999. Nonrandom tissue distribution of mutant mtDNA. *Am. J. Med. Genet.*, **85**, 498–501.

COLLURA, R. V. AND STEWART, C.-B. 1995. Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature*, **378**, 485–489.

COLLURA, R. V., AUERBACH, M. R. AND STEWART, C.-B. 1996. A quick, direct method that can differentiate expressed mitochondrial genes from their nuclear pseudogenes. *Curr. Biol.*, **6**, 1337–1339.

CRACRAFT, J., FEINSTEIN, J., VAUGHN, J. AND HELM-BYCHOWSKI, K. 1998. Sorting out tigers (*Panthera tigris*): mitochondrial sequences, nuclear inserts, systematics, and conservation genetics. *Anim. Conserv.*, **1**, 139–150.

DE WOODY, J. A., CHESSER, R. K. AND BAKER, R. J. 1999. A translocated mitochondrial cytochrome *b* pseudogene in voles (Rodentia: *Microtus*). *J. Mol. Evol.*, **48**, 380–382.

DOUGLAS, A. E. 1994. *Symbiotic Interactions*. Oxford University Press, Oxford.

FELSENSTEIN, J. 1991. *Phylogeny Inference Package (PHYLIP)*, Version 3.4. University of Washington, Seattle, WA.

GELLISSEN, G. AND MICHAELIS, G. 1987. Gene transfer — mitochondria to nucleus. *Ann. N. Y. Acad. Sci.*, **503**, 391–401.

GREENWOOD, A. D. AND PÄÄBO, S. 1999. Nuclear insertion sequences of mitochondrial DNA predominate in hair but not in blood of elephants. *Mol. Ecol.*, **8**, 133–137.

HOWELL, N. 1989. Evolutionary conservation of protein regions in the proton-motive cytochrome *b* and their possible roles in redox catalysis. *J. Mol. Evol.*, **29**, 157–169.

HU, G. AND THILLY, W. G. 1995. Multi-copy nuclear pseudogenes of mitochondrial DNA reveal recent acute genetic change in the human genome. *Curr. Genet.*, **28**, 410–414.

IRWIN, D. M., KOCHER, T. D. AND WILSON, A. C. 1991. Evolution of cytochrome *b* in mammals. *J. Mol. Evol.*, **32**, 128–144.

KIDD, M. G. AND FRIESEN, V. L. 1998. Sequence variation in the guillemot (Alcidae: *Cepphus*) mitochondrial control region and its nuclear homolog. *Mol. Biol. Evol.*, **15**, 61–70.

KOCHER, T. D., THOMAS, W. K., MEYER, A., EDWARDS, S. V., PÄÄBO, S., VILLABLANCA, F. X. AND WILSON, A. C. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 6196–6200.

KVIST, L., RUOKONEN, M., ORELL, M. AND LUMME, J. 1996. Evolutionary patterns and phylogeny of tits and chickadees (genus *Parus*) based on the sequence of the mitochondrial cytochrome *b* gene. *Ornis Fenn.*, **73**, 145–156.

LARA, M. C., PATTON, J. L. AND DA SILVA, M. N. F. 1996. The simultaneous diversification of South American echimyid rodents (Hystricognathi) based on complete cytochrome *b* sequences. *Mol. Phylogenet. Evol.*, **5**, 403–413.

LARSSON, N. G. AND CLAYTON, D. A. 1995. Molecular genetic aspects of human mitochondrial disorders. *Ann. Rev. Genet.*, **29**, 151–178.

LIANG, M. H. AND WONG, L. J. C. 1998. Yield of mtDNA mutation analysis in 2,000 patients. *Am. J. Med. Genet.*, **77**, 395–400.

LOPEZ, J. V., YUHKI, N., MASUDA, R., MODI, W. AND O'BRIEN, S. J. 1994. *Numt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.*, **39**, 174–190.

MARGULIS, L. 1970. *Origin of Eukaryotic Cells*. Yale University Press, New Haven, CT.

MASCHERETTI, S., MIROL, P. M., GIMÉNEZ, M. D., BIDAU, C. J., CONTRERAS, J. R. AND SEARLE, J. B. in press. Phylogenetics of the speciose and chromosomally variable rodent genus *Ctenomys* (Ctenomyidae, Octodontoidea), based on mitochondrial cytochrome *b* sequences. *Biol. J. Linn. Soc.*

MORITZ, C., DOWLING, T. E. AND BROWN, W. M. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Ann. Rev. Ecol. Syst.*, **18**, 269–292.

NEI, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.

OPHIR, R. AND GRAUR, D. 1997. Patterns and rates of indel evolution in processed pseudogenes from humans and murids. *Gene*, **205**, 191–202.

PERNA, N. T. AND KOCHER, T. D. 1996. Mitochondrial DNA: molecular fossils in the nucleus. *Curr. Biol.*, **6**, 128–129.

RAND, D. M. AND HARRISON, R. G. 1989. Molecular population genetics of mtDNA size variation in crickets. *Genetics*, **121**, 551–569.

SAITOU, N. AND NEI, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.

SAMBROOK, J., FRITSCH, E. F. AND MANIATIS, T. 1989. *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SCHNEIDER-BROUSSARD, R. AND NEIGEL, J. E. 1997. A large-subunit mitochondrial ribosomal DNA sequence translocated to the nuclear genome of two stone crabs (*Menippe*). *Mol. Biol. Evol.*, **14**, 156–165.

SHAY, J. W. AND WERBIN, H. 1992. New evidence for the insertion of mitochondrial DNA into the human genome: significance for cancer and aging. *Mutat. Res.*, **275**, 227–235.

SMITH, M. F. AND PATTON, J. L. 1993. The diversification of South American murid rodents: evidence from mitochondrial DNA sequence data for the akodontine tribe. *Biol. J. Linn. Soc.*, **50**, 149–177.

SMITH, M. F., THOMAS, W. K. AND PATTON, J. L. 1992. Mitochondrial DNA-like sequences in the nuclear genome of an akodontine rodent. *Mol. Biol. Evol.*, **9**, 204–215.

SORENSON, M. D. AND FLEISCHER, R. C. 1996. Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 15239–15243.

SWOFFORD, D. L. 1993. *PAUP: Phylogenetic Analysis Using Parsimony*, Version 3.1. Illinois Natural History Survey, Champaign, IL.

THORSNESS, P. E. AND FOX, T. D. 1993. Nuclear mutations in *Saccharomyces cerevisiae* that affect the escape of DNA from mitochondria to the nucleus. *Genetics*, **134**, 21–28.

TIEDEMANN, R. AND VON KISTOWSKI, K. G. 1998. Novel primers for the mitochondrial control region and its homologous nuclear pseudogene in the eider duck *Somateria mollissima*. *Anim. Genet.*, **29**, 468.

VAN DEN BOOGAART, P., SAMALLO, J. AND AGSTERIBBE, E. 1982. Similar genes for a mitochondrial ATPase subunit in the nuclear and mitochondrial genomes of *Neurospora crassa*. *Nature*, **298**, 187–189.

WISCONSIN PACKAGE 1994. Version 8. Genetics Computer Group, 575 Science Drive, Madison, WI.

ZHANG, D.-X. AND HEWITT, G. M. 1996a. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol. Evol.*, **11**, 247–251.

ZHANG, D.-X. AND HEWITT, G. M. 1996b. Highly conserved nuclear copies of the mitochondrial control region in the desert locust *Schistocerca gregaria*: some implications for population studies. *Mol. Ecol.*, **5**, 295–300.

ZULLO, S., SIEU, L. C., SLIGHTOM, J. L., HADLER, H. I. AND EISENSTADT, J. M. 1991. Mitochondrial D-loop sequences are integrated in the rat nuclear genome. *J. Mol. Biol.*, **221**, 1223–1235.