**Columba palumbus** Cyt b-like Numt sequence: comparison with functional homologue and the use of universal primers

Ana R. GROSSO, Cristiane BASTOS-SILVEIRA, Maria M. COELHO and Deodália DIAS

Universidade de Lisboa, Faculdade de Ciências, Centro de Biologia Ambiental/Departamento de Biologia Animal, 1749-016 Lisboa, Portugal; e-mail: argrosso@fc.ul.pt; cbsilveira@fc.ul.pt; mmcoelho@fc.ul.pt; dmdias@fc.ul.pt

Received 10 November 2005; Accepted 24 May 2006

**Abstract.** The woodpigeon *Columba palumbus* has a wide distribution from the Mediterranean region to western Asia. During the amplification of mitochondrial cytochrome *b* gene to assess the degree of differentiation among *C. palumbus* populations a fragment containing an insertion of 5bp was amplified. The non-functional copy of cytochrome *b* gene amplified was interpreted as most likely the result of a past transposition into the nuclear genome of a mitochondrial DNA fragment, Numt. To overcome the technical problems, in terms of both PCR co-amplification and sequence analysis, different sets of primers were used and several criteria used for detection of pseudogene amplification were also discussed. The divergence time estimated between the Numt and its functional homologue indicate that the transfer may have occurred between 4.9 and 5.2 MY ago. Data from cytochrome *b* sequence did not support the existence of a geographically-based divergence among *C. palumbus* populations, which may indicate the persistence of gene flow.

**Key words:** *Columba palumbus*, cytochrome *b* gene, Numt, universal primers, molecular clock

**Introduction**

The woodpigeon *Columba palumbus* Linnaeus, 1758 is one of the main game birds of Europe and it can be distinguished from other pigeon species (Columbiformes, Columbidae) by its larger size, a white patch on both sides of the neck and white marks at the bend of the wings. It is a species with a wide distribution, from the Mediterranean region to western Asia. The populations from northern and central Europe are migratory, with movements to west, while the populations from the south are essentially sedentary (Elias et al. 1999).

In the last decade many references to Numts have appeared, both in invertebrates (Schneider-Brousard & Neigel 1997, Nguyen et al. 2002) and vertebrates (Lopez et al. 1997, DeWoody et al. 1999, Mouri et al. 2001, Haynes et al. 2003, Clifford et al. 2004), adding now up to 82 different species. Although Numts are widely distributed across taxa, they have been reported for only five families of birds belonging to: Passeriformes (Arctander 1995, Nielsen & Arctander 2001, Sato et al. 2001), Anseriformes (Sorenson & Fleischer 1996) and Charadriiformes (Kid & Friesen 1998). Recently, a pseudo-control region was also discovered in Old World eagles *Aquila* (Falconiformes) (Vali 2002). Pereira & Baker (2004) performed BLAST searches for Numt sequences in the chicken genome recently published and they found at least 13 Numts ranging from 131 to 1.733 nucleotides. This fact reveals that the number of Numts in chicken genome is low compared to the 296 Numts detected in the human genome (Mouri et al. 2001).

The undetected co-amplification of Numts during phylogenetic studies using mitochondrial DNA (mtDNA) can lead to apparently credible, but erroneous, results.
Nevertheless, due to a slower rate and random nature of nucleotide substitution in nuclear pseudogenes, when compared to their mitochondrial counterparts, *Numts* can be detected through a fine phylogenetic analysis, since they behave in terms of molecular evolution as damaged molecular fossils (Collura & Stewart 1995). These nuclear inserts do not necessarily correspond to the direct ancestors of current mtDNA haplotypes, and may have had origin in mtDNA lineages that subsequently became extinct (Sorenson & Fleischer 1996). It is possible to estimate a date for the translocation event by comparing the *Numt* sequence with the functional mitochondrial sequence, using local molecular clocks for each region (Lopez et al. 1997).

Here we report how technical hurdles placed by the existence of a *Numt* were solved. An estimation of the time of translocation for the nuclear genome was calculated and additionally, the degree of differentiation of cyt *b* among *C. palumbus* populations was also addressed.

**Methods**

**Sample collection and DNA extraction**

Feather and liver tissue samples of *C. palumbus* were obtained from different regions of Europe (Portugal, Spain, France, Germany, Poland, Hungary, Lithuania, Latvia, Finland and Sweden). Liver tissue samples were preserved in 96% ethanol. DNA was extracted from feather samples using a NucleoSpin C+T Extraction Kit (Macherey Nagel) and from liver tissue using standard protocols of incubation with SDS and digestion with proteinase K, followed by phenol – chloroform extraction (Sambrook et al. 1989).

**Primer selection and amplification of cyt *b* gene sequence and corresponding Numt**

Initially an attempt was made to amplify a 1042bp fragment of mitochondrial cyt *b* gene using the primers L14841 and H4a (Table 1), previously used by Johnson et al. (2001) for the woodpigeon. Since the amplification was not successful, the primer H4a was replaced by H15646 (Table 1) modified from Sorenson et al. (1999). This new pair of primers amplified a 572bp fragment. The alignment of this fragment with homologous published woodpigeon sequences (Acc. No. AF375960; AF353411 in Johnson et al. 2001; AF483335 in Shapiro et al. 2002) allowed us to identify an insertion of 5bp (Fig. 1). The translation of such fragment would be disrupted much earlier than expected as the frameshift was confirmed to originate several stop codons downstream, and consequently led to a smaller and likely to be non-functional cyt *b* protein. Therefore, it was suspected that these primers were amplifying a *Numt*. To overcome this problem, different sets of primers for the mitochondrial cyt *b* gene were tested (Table 1).

**PCR amplification and sequencing**

PCR amplifications were performed using a Minicycler™ MJ Research with 35 cycles of the following steps: 94°C for 40 s, 50°C for 45 s, and 72°C for 1 min, preceded by 5 min at 95°C and followed by 10 min at 72°C. PCR reactions contained 500 ng of template DNA,
1.25 U of *Taq* DNA polymerase (Fermentas), 1x buffer (Fermentas), 3 mM MgCl₂, 0.2 mM of each dNTP, 0.3 µM of each primer in a final reaction volume of 25 µl. PCR products were purified by precipitation with sodium acetate and ethanol for subsequent sequencing. Amplified fragments were sequenced using the same primers, labelled with fluorescent dye (IRD800; Li-Cor), and the CycleReader™ Auto DNA Sequencing Kit (Fermentas). Products were run on a Li-Cor 4200 automated DNA sequencer. Cycle sequencing was carried out

Table 1. Primers for PCR amplification and sequencing.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5’→3’)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>L14841</td>
<td>AAAAACGTCCTACATCAGCATCTACGATGATGAAA</td>
<td>Koehler et al. 1989</td>
</tr>
<tr>
<td>H4a</td>
<td>AAGTGGTAAGCTTCCAGATCTGAGATGATGACCC</td>
<td>Harsman 1996</td>
</tr>
<tr>
<td>H156446</td>
<td>GGNGTAAATTTCCTGGTTCYCC</td>
<td>Modified from Sorenson et al. 1999</td>
</tr>
<tr>
<td>H15149</td>
<td>AAAACTGCAAGGCCCTAGAATGATTTGTCCTCA</td>
<td>Koehler et al. 1989</td>
</tr>
<tr>
<td>L14764</td>
<td>CAAAAAATAGGMCVARGG</td>
<td>Modified from Sorenson et al. 1999</td>
</tr>
</tbody>
</table>

N = Adenine (A), Cytosine (C), Guanine (G) or Thiamine (T); R = A or G; Y = C or T; M = C or A; V = A, C or G.

Fig. 1. Aligned sequences of *Columba palumbus* cytochrome *b* gene (Mt1) and *Numt* (Nuc1 and Nuc2), along with published woodpigeon sequences (identified by their respective accession number). Dots indicate sequence identity and gaps are indicated by a dash.
over 35 cycles of the following steps: 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min 30 s, preceded by 3 min at 94°C and followed by 8 min at 72°C.

**Genetic differentiation**

The sequences were aligned manually using BioEdit version 5.0.9 (Hall 1999) and the identification of the sequenced fragments was confirmed, as referred above, by their alignment with published sequences of *C. palumbus* cyt *b* gene (Acc. No. AF375960; AF353411 in Johnson et al. 2001; AF483335 in Shapiro et al. 2002), *Ciconia ciconia* (Acc. No. NC_002197) and *Gallus gallus* (Acc. No. X52392 in Desjardins & Morais 1990). MacClade version 3.08a (Madison & Madison 1999) was used to define the haplotypes.

The haplotype (*h*) and nucleotide diversity (*π*) (Nei 1987) were estimated with Arlequin version 2.0 (Schneider et al. 2000).

A network was constructed by departing from the median-joining network (Bandelt et al. 1999) as implemented in Network version 4.1.0.0 (available at www.fluxus-engineering.com). Phylogenetic trees were generated using maximum-parsimony, maximum likelihood and neighbour-joining methods, as implemented in PAUP version 4.0d (Swofford 1998). Cyt *b* gene sequences for *C. livia* (Acc. No. AF182694 in Johnson & Clayton 2000a) and *C. guinea* (Acc. No. AF279708 in Johnson & Clayton 2001) were used as outgroup for tree rooting. Maximum-parsimony analysis used heuristic search, random stepwise addition and tree bisection-reconnection methods. For the maximum-likelihood and neighbour-joining distance trees, sequence divergence was calculated according to a nucleotide substitution model determined by Modeltest version 3.0 (Posada & Crandall 1998). The Felsenstein model (1981) was the selected, with a gamma value of 0.440869. Support for nodes was assessed by bootstrap resampling using 1000 replicates (Felsenstein 1985). The genetic distances among samples including uncorrected p-distances and distances corrected using the model described above were calculated with PAUP version 4.0d (Swofford 1998).

A hierarchical analysis of genetic structure was performed using the analysis of molecular variance (AMOVA) (Excoffier et al. 1992) as implemented in Arlequin. Populations were grouped in alternative testable arrangements according to geographic distribution and the migratory routes of the woodpigeon, resulting in two groups: 1) all populations and 2) Iberian Peninsula (Portugal and Spain) versus remaining populations. The groups were tested using three molecular distances: Kimura’s two-parameters (1980), Tamura & Nei (1993), both with gamma correction, and simple uncorrected pairwise divergence.

**Dating the translocation of the mitochondrial cyt *b* gene copy to the nucleus**

To estimate the time of divergence between the mitochondrial cyt *b* sequence and the *Numt* sequence, two different approaches were used. First, the equation: \( \delta = (\lambda_n + \lambda_c)T \) was used, where \( \delta \) is the genetic divergence, \( \lambda \) is the absolute rate of divergence in each different lineage (nuclear – N cytoplasmatic – C) and T is time (Lopez et al. 1997). For \( \lambda_c \) the value of 2.0x10^{-8} substitutions/site/year was used, based on a molecular clock calibration for mitochondrial genes of birds (Klicka & Zink 1997) and for \( \lambda_n \), the value of 4.6x10^{-9} substitutions/site/year, as estimated for pseudogenes (Li et al. 1981). The second approach
was an adaptation from studies in birds (Sorensen & Fleischer 1996) and mammals (Collura & Stewart 1995, Cracraft et al. 1998). This methodology implied the phylogenetic reconstruction of some Columbiformes species (from genera Columba, Nesoenas and Streptopelia) using published cyt b and our Numt sequence (Table 2), followed by the application of a molecular clock. Maximum-parsimony, maximum-likelihood and neighbour-joining methods were used to reconstruct the phylogenetic trees. In this analysis the 5bp insertion, was removed from the Numt sequences, assuming that this mutation occurred after the moment of translocation. The Columbiformes species Macropygia phasianella and Reinwardtoena browni (Table 2) were used as outgroups. Given that transition-to-transversion ratios differed across codons positions, a weighted (4:1) maximum parsimony analysis was made. The most appropriate model for neighbour-joining analysis, determined by Modeltest version 3.0 (Posada & Crandall 1998), was the General Time Reversible model.

**Table 2.** Published cyt b gene sequences of Columbiformes species used in the present study.

<table>
<thead>
<tr>
<th>Genus</th>
<th>Species</th>
<th>GeneBank Acc No.</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Columba</td>
<td>Columba arquatrix</td>
<td>AF353412</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Columba guinea</td>
<td>AF279708</td>
<td>Johnson &amp; Clayton (2001)</td>
</tr>
<tr>
<td></td>
<td>Columba livia</td>
<td>AF182694</td>
<td>Johnson &amp; Clayton (2000a)</td>
</tr>
<tr>
<td></td>
<td>Columba oenas</td>
<td>AF375961</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Columba palumbus</td>
<td>AF353411</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Columba pulchrichollis</td>
<td>AF353413</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Columba rupestris</td>
<td>AF353410</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Macropygia</td>
<td>Macropygia phasianella</td>
<td>AF483339</td>
<td>Shapiro et al. (2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nesoenas</td>
<td>Nesoenas mayeri</td>
<td>AF483322</td>
<td>Shapiro et al. (2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reinwardtoena</td>
<td>Reinwardtoena browni</td>
<td>AF353417</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Streptopelia</td>
<td>Streptopelia bitorquata</td>
<td>AF353406</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia capicola</td>
<td>AF279709</td>
<td>Johnson &amp; Clayton (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia chinensis</td>
<td>AF483341</td>
<td>Shapiro et al. (2002)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia decaocto</td>
<td>AF353398</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia decipiens</td>
<td>AF353400</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia hypopyrrha</td>
<td>AF353403</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia mayeri</td>
<td>AF353408</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia orientalis</td>
<td>AF353405</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia picturata</td>
<td>AF353409</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia roseogrisea</td>
<td>AF353399</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia semitorquata</td>
<td>AF353401</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia senegalensis</td>
<td>AF279710</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia tranquebarica</td>
<td>AF353407</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia turtur</td>
<td>AF353404</td>
<td>Johnson et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Streptopelia vinacea</td>
<td>AF353402</td>
<td>Johnson et al. (2001)</td>
</tr>
</tbody>
</table>
(GTR, Rodríguez et al. 1990, Yang et al. 1994), with the underlying parameters of the model (unequal nucleotide frequencies, substitution rate matrix, gamma distribution and proportion of invariant sites) estimated from the data. The heterogeneity rate of evolutionary change along the lineages was analyzed by comparing the likelihood scores of neighbour-joining trees with and without an enforced molecular clock.

Results

Sequence identification and nucleotide composition

The primers L14841 – H15646 and L14841 – H15149 (Table 1) amplified a 572bp fragment with a 5bp insertion, corresponding to a duplication of the nucleotide motif GAATC (Fig. 1). The primers L14764 – H15646 (Table 1) amplified a 782bp fragment that did not have the insertion or stop codons. When both fragments were compared with the three published woodpigeon sequences it was observed that besides the insertion, the 572bp fragment was more similar to the woodpigeon sequence AF375960 (99%) relatively to the other published sequences (AF353411 and AF483335 – 88%) (Fig. 1). The sequence AF375960 has only 356bp, and did not include the region where of the insertion occurs. In contrast, the 782bp fragment was more similar to the woodpigeon sequences AF353411 and AF483335 (100%) relatively to the sequence AF375960 (89%) (Fig. 1). A total of 19 specimens of C. palumbus were sequenced for the Numt sequence and 29 samples for the mitochondrial fragment (Table 3). Fifteen specimens were common for both fragments. The DNA was extracted from feather and liver tissues and differences between them were not found. The mitochondrial fragment included 86bp of NADH dehydrogenase subunit V (corresponding to 4.7% of the total fragment sequenced), 11bp intergenic and 685bp of cyt b gene (60% of the total fragment).

The sequences generated by the present work were deposited in GeneBank under Accession Nos. AY251463 – AY251472 and AY251473 – AY251474.

Nucleotide composition was estimated for C. palumbus cyt b gene sequence, for the Numt sequence and for the cyt b gene sequence of Columbiformes (see species in Table 2). The obtained values revealed, for all the sequences, a similar content of adenosine (~26.4%) and a similar deficiency of guanosine (~14.7%). However, in the Numt sequence the content of

Table 3. Geographic origin, source of DNA and number of specimens of Columba palumbus analysed in the present study, for each of the amplified fragments.

<table>
<thead>
<tr>
<th>Geographic Origin</th>
<th>Number of specimens for each fragment</th>
<th>Source of DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>572 bp</td>
<td>782 bp</td>
</tr>
<tr>
<td>Portugal</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Spain</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>France</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Germany</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Poland</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Hungary</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Lithuania</td>
<td>–</td>
<td>4</td>
</tr>
<tr>
<td>Latvia</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Finland</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Sweden</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>19</strong></td>
<td><strong>29</strong></td>
</tr>
</tbody>
</table>
cytosine was lower (30.8%) and the content of thymine higher (28.2%) relatively to the cyt b gene sequences of Columbiformes and *C. palumbus* (averages of 33.4% and 25.6% respectively).

**Sequence divergence and phylogenetic relationships of cyt b gene haplotypes**

In this analysis only 685bp of the cyt b gene were used and this fragment corresponds to the region between the positions 14893 and 15577 from the chicken mtDNA reference sequence Acc. No. X52392. Eleven variable base changes (1.6%) were observed, corresponding to six transitions and five transversions (Table 4). The majority of the variable sites were found in the third codon positions (six), three were found in the second codon positions and two in the first codon positions. Sequences were translated into amino acids and it was not detected any deletion or insertion. The total nucleotide variation induced five amino acids changes.

The 29 *C. palumbus* sequences comprised 10 different haplotypes defined by 11 variable sites and differed by one to two substitutions (Table 4). Nine unique haplotypes were found and only the Mt1 haplotype was common to all populations (Table 4). Low to high levels of haplotypic diversity (0.000–1.000) and remarkably low nucleotide diversity (0.00000–0.00293) were observed over all populations (Table 4).

The median-joining network showed that the node correspondent to the common haplotype (Mt1), with the highest haplotype frequency, was connected by one mutational step in most cases to all the unique haplotypes (Fig. 2). Equally weighted maximum parsimony

---

**Fig. 2.** Median-joining haplotype network for the observed *Columba palumbus* cytochrome *b* gene sequence haplotypes. Nodes are labelled with the haplotype; node sizes are proportional to the haplotype frequency. Mutating sites are noted along the branches.
Table 4. Cyt b gene sequence haplotypes (Haplot) and variable sites in the 685bp sequenced region of *Columba palumbus* (dots indicate equality with sequence Mt1). Haplotype frequencies, haplotype diversity and nucleotide diversity within the ten populations (P, Portugal; Sp, Spain; Fr, France; G, Germany; Pol, Poland; H, Hungary; Li, Lithuania; La, Latvia; Fi, Finland; S, Sweden) are shown.

<table>
<thead>
<tr>
<th>Variable sites</th>
<th>Populations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplot</td>
<td>P</td>
</tr>
<tr>
<td>Mt1</td>
<td>G</td>
</tr>
<tr>
<td>Mt2</td>
<td>•</td>
</tr>
<tr>
<td>Mt3</td>
<td>•</td>
</tr>
<tr>
<td>Mt4</td>
<td>.</td>
</tr>
<tr>
<td>Mt5</td>
<td>C</td>
</tr>
<tr>
<td>Mt6</td>
<td>•</td>
</tr>
<tr>
<td>Mt7</td>
<td>•</td>
</tr>
<tr>
<td>Mt8</td>
<td>.</td>
</tr>
<tr>
<td>Mt9</td>
<td>.</td>
</tr>
<tr>
<td>Mt10</td>
<td>•</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Haplotype Diversity</th>
<th>0.500</th>
<th>0.000</th>
<th>1.000</th>
<th>1.000</th>
<th>1.000</th>
<th>1.000</th>
<th>0.500</th>
<th>0.500</th>
<th>1.000</th>
<th>1.000</th>
<th>0.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>±0.265</td>
<td>±0.000</td>
<td>±0.500</td>
<td>±0.500</td>
<td>±0.000</td>
<td>±0.265</td>
<td>±0.265</td>
<td>±0.500</td>
<td>±0.272</td>
<td>±0.000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nucleotide Diversity</th>
<th>0.00073</th>
<th>0.00000</th>
<th>0.00292</th>
<th>0.00146</th>
<th>0.00000</th>
<th>0.00073</th>
<th>0.00073</th>
<th>0.00146</th>
<th>0.00293</th>
<th>0.00000</th>
</tr>
</thead>
<tbody>
<tr>
<td>±0.00207</td>
<td>±0.00000</td>
<td>±0.00358</td>
<td>±0.00207</td>
<td>±0.00000</td>
<td>±0.00091</td>
<td>±0.00091</td>
<td>±0.00207</td>
<td>±0.00276</td>
<td>±0.00000</td>
<td></td>
</tr>
</tbody>
</table>
analysis of haplotypes revealed seven most parsimonious trees consisting of 86 steps. These topologies showed a major group including all the woodpigeon haplotypes, which was strongly supported (100%). However, the analysis did not support the existence of any population structure and phylogenetic tree was a comb. The trees were similar, except in the basal position where the unique haplotypes alternated. The same topologies were obtained with the maximum-likelihood and neighbour-joining methods.

Analysis of molecular variance (AMOVA) confirmed the absence of genetic structure in all the hierarchical arrangements tested and this was constant regardless of the genetic distances employed (data not shown).

Numt sequences

In a total of 19 samples two Numt haplotypes, Nuc1 and Nuc2, differing by two substitutions (one transition and one transversion) were detected (Table 3). Haplotype Nuc2 was only found in two samples of Portugal, while the remaining 17 samples had haplotype Nuc1. The divergence found between the cyt b gene sequences and the Numt sequences presented values

![Phylogenetic tree of Columbiformes cytochrome b gene sequences and Columba palumbus Numt sequence. Maximum parsimony tree with a 4:1 weighting. Percentage bootstrap supports according to the maximum parsimony method (above branches) and the neighbour-joining (GTR distance) method (below branches) are given. Bootstrap values below 50% are not shown.](image)
between 12.01% and 12.76%. The replacement of these values by $\delta$, indicates that the Numt was translocated to the nuclear genome approximately between 4.9 and 5.2 MYA.

Phylogenetic analysis of cyt $b$ gene sequences of Columbiformes and Numt sequences, using weighted (4:1) maximum parsimony, maximum-likelihood and neighbour-joining analyses, produced phylogenetic trees with identical topologies (data not shown). The maximum parsimony analysis generated only one parsimonious tree consisting of 1432 steps (Fig. 3). All analysis showed that the Numt sequences formed a reciprocally monophyletic clade with Streptopelia and Nesoenas sequences, although this node was not supported by bootstrap (<50%). Based on nuclear and mitochondrial gene sequences, Johnson et al. (2001) recommended Nesoenas mayeri to be transferred to the genus Streptopelia, concept that was incorporated in our analysis. The exclusion of the Numt sequences did not have any impact on the topology of the trees. Based on the molecular clock calibration for the mtDNA in birds (Klicka & Zink 1997), divergence time between the genera Columba and Streptopelia was estimated. The neighbour-joining trees obtained without assuming a constant rate among lineages (log \(-5966.24\) and enforcing a molecular clock (log \(-5985.38\)) had the same topology. The likelihood ratio test between these trees did not reject the hypothesis of rate constancy ($x^2 = 38.28$, df = 25, $P < 0.01$). The maximum and minimum values of genetic distance found between the two genera were 26.5% and 11.5% respectively. Consequently, the divergence time obtained ranged from 2.9 to 6.6 MYA.

**Discussion**

The overall very low differentiation between sedentary (Iberian Peninsula) and migratory C. palumbus populations could be explained by the migratory movements to west allowing the contact with the sedentary populations. Other possible reason for such low level of genetic variability could be the size of the studied fragment together with the low mutation rate of the cytochrome $b$ gene. Nevertheless, in the studies of Weniak et al. (1993) and Marshall & Baker (1999) on Calidris alpina (Charadriiformes family) and Fringilla coelebs (Passeriformes family) respectively, the cytochrome $b$ gene and the control region were used simultaneously and the results indicated that although less variable, the first 685bp of the mitochondrial cyt $b$ gene contained sufficient information to differentiate the studied populations.

In the present study, the amplification of C. palumbus cyt $b$ gene produced two different fragments. Results from the alignment with published sequences and the presence of stop codons in the translation of one of the sequence, led us to conclude that the 572bp fragment corresponds to a copy of the cyt $b$ gene, which is likely to be inserted within the nuclear genome (Numt).


The accidental co-amplification of Numts can be influenced by the abundance of Numts in the taxon under study, DNA extraction method employed, type of the sample used, and primers selected for the PCR (Bensanson et al. 2001). For birds, PCR-amplifications of DNA extracts derived from blood samples can present a high number of nuclear insertions due to the presence of nucleated erythrocytes (Quinn 1992). Although it was not
possible to extract purified mtDNA, as it would require high-quality samples (that were not available), we extracted genomic DNA exclusively from feather and liver samples in order to avoid the potential problems with blood samples. Results generated from both types of samples showed no difference.

Concerning the assays with different primers used for the PCR reaction, the sets L14841 – H15646 and L14841 – H15149 amplified the Numt. When the sequence of the L14841 primer was compared with the sequence amplified using an upstream primer (L14764), the priming site in this mitochondrial gene sequence was different from the sequence of the L14841 primer. Therefore, primer L14841 was not hybridizing to the target sequence, which explains the absence of double peaks in the chromatograms. Since the absence of double peaks is often used as an argument for the absence of pseudogene amplification (M i n d e l l et al. 1998, L o v e t t e et al. 1999, J o h n s o n & C l a y t o n 2000b), our findings suggest that this observation should not constitute the sole criterion employed to rule out the presence of Numts in a dataset.

Other hypotheses could explain the identified Numt. It could be a gene with extra-nucleotides and a translation frameshifting mechanism, where during the translation of mRNA into amino acids would occur a change in the reading frame and the extra-nucleotides would not be read. This mechanism was referred by M i n d e l l et al. (1998) to explain the discovery of a single extra nucleotide in the mitochondrial NADH dehydrogenase subunit III of some birds and turtles. Also, in the present study, a frameshift with 5bp was found in the amplified fragment and if a translation frameshifting mechanism would occur, the translation would not yield stop codons. Meanwhile, the amplification of the 685bp sequence of cyt b without the insertion, discarded definitively this hypothesis.

The fragment now considered as a Numt, could be an intra-mitochondrial duplication, as advanced to explain two identical control regions found in some organisms (K u m a z a w a et al. 1998, H a r i n g et al. 1999, B r i t o 2005). Other hypothesis could be a heteroplasmy with more than one different molecular forms of mtDNA in the same individual, as described for the wild mice Mus mus musculus (B o u r s o t et al. 1987). Both hypotheses assume that the alternative fragments are located in the mtDNA as non-coding regions, and so a higher mutation rate should be expected, as reported for control region. These two hypotheses can not explain our findings. The genetic distance found among the 19 sequences obtained for the non-coding fragment of the present study ranged between 0.00000 and 0.00560.

Consequently, the hypothesis of a Numt amplification seemed to be the most consistent. Moreover, in another study on bird species from Scytalopus genus (Passeriformes Family) (A r c t a n d e r 1995), using the same universal primers (K o c h e r et al. 1989) a Numt was also amplified.

Moreover, it was possible to verify that one of the sequences deposited on GenBank (Acc. No. AF375960) and identified as the mitochondrial cyt b gene of C. palumbus, was very similar to the Numt sequence obtained. This fact denounces the possible existence of other sequences in the database that could not really correspond to mitochondrial genes, which could lead to erroneous phylogenies. In the study referred by A c t a n d e r & F j e l d s a (1994) a phylogenetic reconstruction revealed an inexplicable phylogenetic relation of two taxa of Scytalopus genus, posteriorly explained by the probably existence of a pseudogene (A r c t a n d e r 1995). In the phylogeny of Columbiformes here presented, if the Numt had not been identified and the same sequence considered has part of cyt b gene, a monophyletic clade of C. palumbus and Streptopelia genera, separated from the genus Columba (Fig. 2) would be reconstructed.
In the present study the phylogenetic reconstruction of Columbiformes had the objective to determine the moment of translocation of the mitochondrial fragment to the nucleus. The clades formed in the phylogenetic tree (Fig. 3) were very similar to those described in the study of Johnson et al. 2001, although they were not supported by bootstrap values higher than 50%. The topology obtained in the phylogeny, did not show the Numt sequences as ancestors of the Columba genus or of the group formed by the Columba, Nesoenas and Streptopelia genera. However, the relationships obtained suggest that the nuclear insert is at least as old as the split between the genera Columba and Streptopelia. Moreover, if the estimated time of 4.9 and 5.2 MYA were correct, one should expect the appearance of 13–14 randomly substitutions in the Numt sequences. These substitutions, probably, would be the cause of the proximity between the C. palumbus Numt sequences and the Nesoenas and Streptopelia cyt b gene sequences. In a similar study on felines (Cracraft et al. 1998), the phylogenetic tree obtained placed the cyt b pseudogene sequence of the tiger close to the mitochondrial cyt b gene sequence of the lion. These authors inferred that the nuclear insert was as old as the evolution split between tigers and lions.

Our first approach for dating the translocation of the mitochondrial cyt b gene copy to the nucleus indicated a divergence time raging between 4.9 and 5.2 MYA, while the application of the molecular clock in the phylogenetic tree revealed an interval time between 2.9 and 6.6 MYA. In the second approach, the interval time had higher amplitude, due to the use of genetic distances values between many taxa.

The study of Numt sequences in the Columbiformes species, would allow a more accurate dating of the translocation moment. This could aid in the deepening of the knowledge on the time and pattern of the mitochondrial cyt b gene evolution of the Columbiformes species.

Acknowledgements

We are grateful to Michael W. Bruford and Carlos Fernandes for valuable comments on a previous version of the manuscript. We also thank Anabel Perdices, Carina Cunha, Cristina Luís, Natacha Mesquita and Patricia Salgueiro for the support in the informatics programmes. We also thanks to Direção-Geral das Florestas (Portugal), Conseil Régional de La Chasse d’Aquitaine (France) and Federación Guipuzcoana de Caza (Spain) for providing biological samples for this study and support financially this research.

C. Bastos-Silveira was supported by a Postdoctoral grant (SFRH/BPD/116/2002) from the Portuguese Foundation for Science and Technology (FCT/MCT).

Literature


Schneider S., Roessli D. & Laurent E. 2000: Arlequin ver. 2.000: A software for population genetics data analysis. *Genetics and Biometry Laboratory, University of Geneva, Switzerland.*


