Folia Zool. – 55(2): 131–144 (2006)

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

A b s t r a c t. 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 (E l i a s et al. 1999).

In the last decade many references to *Numts* have appeared, both in invertebrates (S c h n e i d e r - B r o u s s a r d & N e i g e l 1997, N g u y e n et al. 2002) and vertebrates (L o p e z et al. 1997, D e W o o d y et al. 1999, M o u r i e r et al. 2001, H a y n e s et al. 2003, C l i f f o r d 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 (A r c t a n d e r 1995, N i e l s e n & A r c t a n d e r 2001, S a t o et al. 2001), Anseriformes (S o r e n s o n & F l e i s c h e r 1996) and Charadriiformes (K i d d & F r i e s e n 1998). Recently, a pseudo-control region was also discovered in Old World eagles *Aquila* (Falconiformes) (V ä l i 2002). P e r e i r a & B a k e r (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 (M o u r i e r et al. 2001).

The undetected co-amplification of *Numts* during phylogenetic studies using mitochondrial DNA (mtDNA) can lead to apparently credible, but erroneous, results

(Arctander 1995, Zischler et al. 1995). 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: 94ºC for 40 s, 50ºC for 45 s, and 72ºC for 1 min, preceeded by 5 min at 95ºC and followed by 10 min at 72ºC. PCR reactions contained 500 ng of template DNA,

Table 1. Primers for PCR amplification and sequencing.

Name	Sequence (5'→ 3')	Source
L14841	AAAAAGCTTCCATCCAACATCTCAGCATGATGAAA	Kocher et al. 1989
H4a	AAGTGGTAAGTCTTCAGTCTTTGGTTTACAAGACC	Harshman 1996
H156446	GGNGTRAARTTTTCTGGGTCYCC	Modified from Sorenson et al. 1999
H15149	AAACTGCAGCCCCTCAGAATGATATTTGTCCTCA	Kocher et al. 1989
L14764	CAAAAAAATAGGMCCVGARGG	Modified from Sorenson et al. 1999

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.

1.25 U of *Taq* DNA polymerase (Fermentas), 1x buffer (Fermentas), 3 mM $MgCl_2$, 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

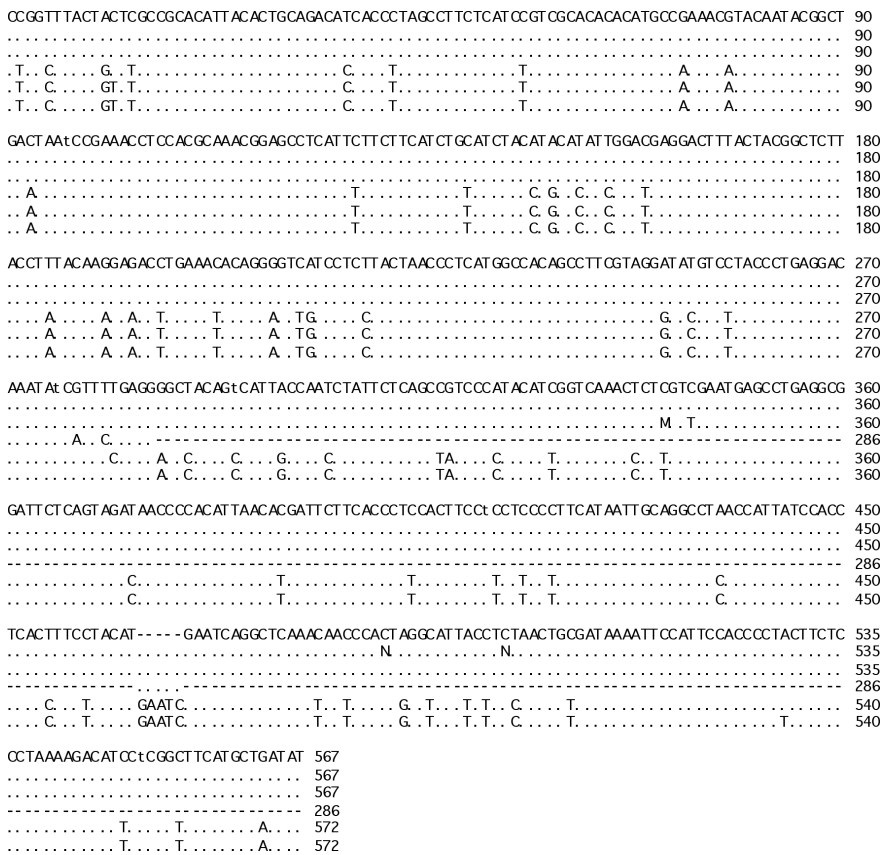

Fig. 1. Aligned sequences of *Columba palumbus* cytochrome *b* gene (Mt1) and *Numt* (Nuc1 and Nuc2), along with published woodpigeon sequences (identified by their respective accession number). Dots indicate sequence identity and gaps are indicated by a dash.

over 35 cycles of the following steps: 94ºC for 30 s, 50ºC for 30 s, and 72ºC for 1 min 30 s, preceded by 3 min at 94ºC and followed by 8 min at 72ºC.

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 (Maddison & Maddison 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 δ is the genetic divergence, λ is the absolute rate of divergence in each different lineage (nuclear – N cytoplasmatic – C) and T is time (Lopez et al. 1997). For λ_C the value of 2.0×10^{-8} substitutions/site/year was used, based on a molecular clock calibration for mitochondrial genes of birds (Klicka & Zink 1997) and for λ_N the value of 4.6×10^{-9} substitutions/site/year, as estimated for pseudogenes (Li et al. 1981). The second approach

was an adaptation from studies in birds (Sorenson & 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.

Genus Species	GeneBank Acc No.	Source
Columba		
Columba arquatrix	AF353412	Johnson et al. (2001)
Columba guinea	AF279708	Johnson & Clayton (2001)
Columba livia	AF182694	Johnson & Clayton (2000a)
Columba oenas	AF375961	
Columba palumbus	AF353411	Johnson et al. (2001)
Columba pulchrichollis	AF353413	Johnson et al. (2001)
Columba rupestris	AF353410	Johnson et al. (2001)
Macropygia		
Macropygia phasianella	AF483339	Shapiro et al. (2002)
Nesoenas		
Nesoenas mayeri	AF483322	Shapiro et al. (2002)
Reinwardtoena		
Reinwardtoena browni	AF353417	Johnson et al. (2001)
Streptopelia		
Streptopelia bitorquata	AF353406	Johnson et al. (2001)
Streptopelia capicola	AF279709	Johnson & Clayton (2001)
Streptopelia chinensis	AF483341	Shapiro et al. (2002)
Streptopelia decaocto	AF353398	Johnson et al. (2001)
Streptopelia decipiens	AF353400	Johnson et al. (2001)
Streptopelia hypopyrrha	AF353403	Johnson et al. (2001)
Streptopelia mayeri	AF353408	Johnson et al. (2001)
Streptopelia orientalis	AF353405	Johnson et al. (2001)
Streptopelia picturata	AF353409	Johnson et al. (2001)
Streptopelia roseogrisea	AF353399	Johnson et al. (2001)
Streptopelia semitorquata	AF353401	Johnson et al. (2001)
Streptopelia senegalensis	AF279710	Johnson et al. (2001)
Streptopelia tranquebarica	AF353407	Johnson et al. (2001)
Streptopelia turtur	AF353404	Johnson et al. (2001)
Streptopelia vinacea	AF353402	Johnson et al. (2001)

(GTR, Rodriguez 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 defficiency 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.

Geographic Origin	Number of specimens for each fragment		Source of DNA
	572bp	782bp	
Portugal	4	6	Liver
Spain	2	2	Feathers
France	3	2	Feathers
Germany	1	2	Liver
Poland	1	1	Liver
Hungary	2	4	Liver
Lithuania	–	4	Liver
Latvia	2	2	Liver
Finland	2	3	Liver
Sweden	2	3	Liver
Total	19	29	

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

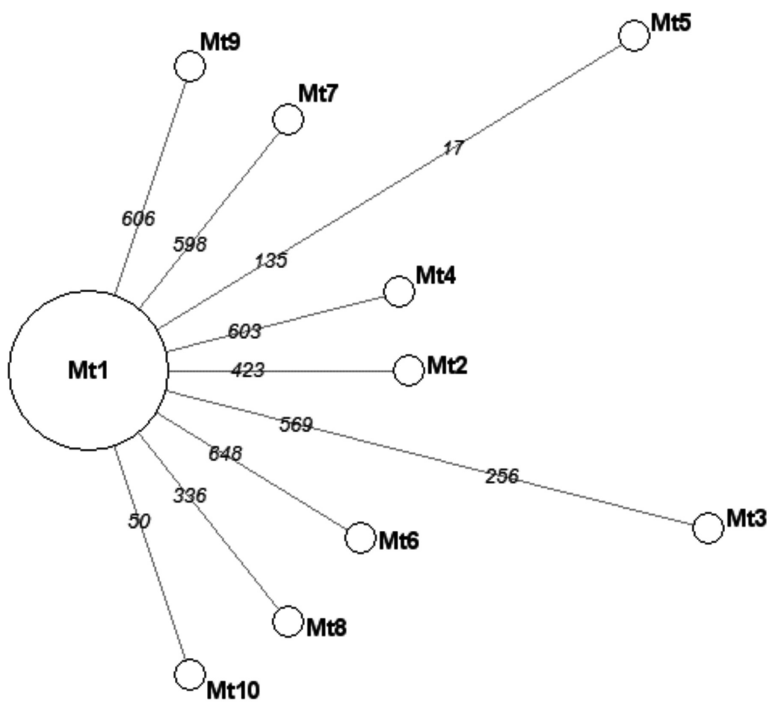

Fig. 2. Median-joining haplotype network for the observed *Columba palumbus* cytochrome *b* gene sequence haplotypes. Nodes are labelled with the haplotype; node sizes are proportional to the haplotype frequency. Mutating sites are noted along the branches.

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.

		Variable sites											Populations									
	Haplot	017	050	135	256	336	423	569	598	603	606	648	P	Sp	Fr	G	Pol	H	Li	La	Fi	S
Haplotype frequencies	Mt1	G	A	A	A	G	T	T	T	A	T	C	4	2	1	1	1	3	3	1	1	3
	Mt2	•	•	•	•	•	C	•	•	•	•	•				1						
	Mt3	•	•	•	C	•	•	G	•	•	•	•									1	
	Mt4	.	.	.	•	•	•	•	•	T	.	.									1	
	Mt5	C	•	G	•	•	•	•	•	•	•	•			1							
	Mt6	•	•	•	•	•	•	•	•	•	•	T						1				
	Mt7	•	•	•	•	•	•	•	C	•	•	•								1		
	Mt8	.	.	.	•	A	•	•	•	•	•	•							1			
	Mt9	.	.	.	.	.	.	.	.	•	C	•	1									
	Mt10	•	T	•	•	•	•	•	•	•	•	•	1									
Haplotype Diversity													0.500 ± 0.265	0.000 ± 0.000	1.000 ±0.500	1.000 ±0.500	1.000 ±0.000	0.500 ±0.265	0.500 ±0.265	1.000 ±0.500	1.000 ±0.272	0.000 ±0.000
Nucleotide Diversity													0.00073 ±0.00207	0.00000 ±0.00000	0.00292 ±0.00358	0.00146 ±0.00207	0.00000 ±0.00000	0.00073 ±0.00091	0.00073 ±0.00091	0.00146 ±0.00207	0.00293 ±0.00276	0.00000 ±0.00000

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

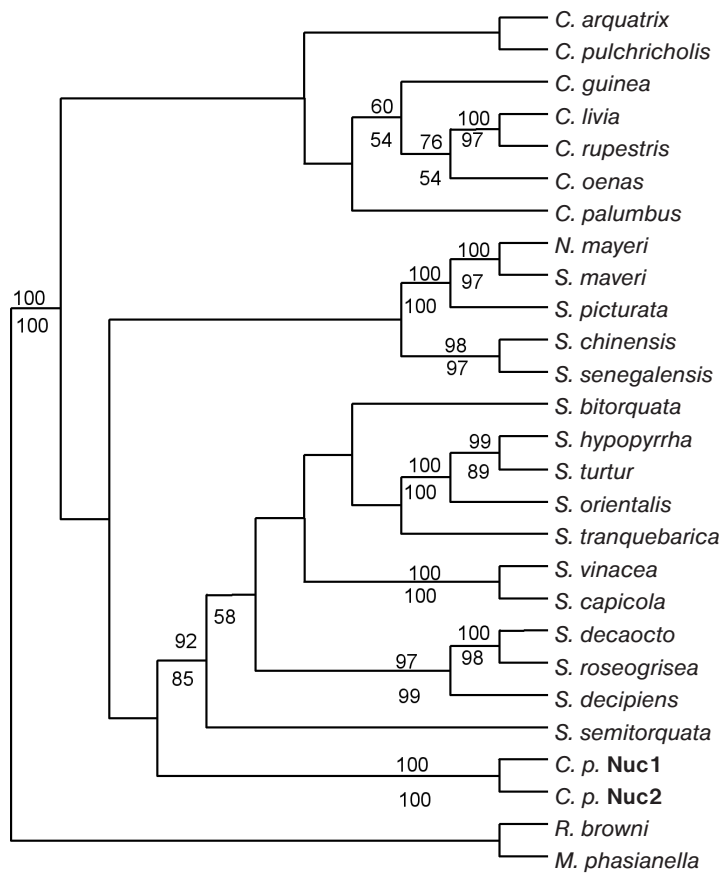

Fig. 3. Phylogentic tree of Columbiformes cytochrome *b* gene sequences and *Columba palumbus Numt* sequence. Maximum parsimony tree with a 4:1 weighting. Percentage bootstrap supports according to the maximum parsimony method (above branches) and the neighbour-joining (GTR distance) method (below branches) are given. Bootstrap values below 50% are not shown.

between 12.01% and 12.76%. The replacement of these values by δ, 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 Wenink 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 discovery of insertions of bird mitochondrial genes into the nuclear genome (Quinn 1992, Sorenson & Fleischer 1996, Kidd & Friesen 1998, Nielsen & Arctander 2001) including the cyt *b* gene (Arctander 1995, Sato et al. 2001) has gradually increased during the last decade.

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 (Mindell et al. 1998, Lovette et al. 1999, Johnson & Clayton 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 Mindell 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 (Kumazawa et al. 1998, Haring et al. 1999, Brito 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* (Bourset 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. Morever, in another study on bird species from *Scytalopus* genus (Passeriformes Family) (Arctander 1995), using the same universal primers (Kocher 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 Actander & Fjeldsa (1994) a phylogenetic reconstruction revealed an inexplicable phylogenetic relation of two taxa of *Scytalopus* genus, posteriorly explained by the probably existence of a pseudogene (Arctander 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 Patrícia Salgueiro for the support in the informatics programmes. We also thanks to Direcção-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

Arctander P. 1995: Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. *Proc. R. Soc. Lond. B. Biol. Sci. 262: 13–19.*

Arctander P. & Fjeldsa J. 1994: Andean tapaculos of the genus *Scytalopus* (Aves, Rhinocryptidae): a study of speciation patterns using DNA sequence data. In: Loeschcke V., Tomiuk J. & Jain S.K. (eds), Conservation Genetics. *Basel, Birkhäuser.*

Bandelt H.-J., Forster P. & Röhl A. 1999: Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol. 16 (1): 37–48.*

Bensasson D., Zhang D.Z., Hartl D.L. & Hewitt G.M. 2001: Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol. 16(6): 314–321.*

Boursot P., Yonekawa H. & Bonhomme F. 1987: Heteroplasmy in Mice with deletion of a large coding region of Mitochondrial DNA. *Mol. Biol. Evol. 4(1): 46–55.*

Brito P.H. 2005: The influence of Pleistocene glacial refugia on tawny owl genetic diversity and phylogeography in western Europe. *Mol. Ecol. 14: 3077–3094.*

Clifford S.L., Anthony N.M., Bawe-Johnson M., Abernethy K.A., Tutin C.E.G., White L.J.T., Bermejo M., Goldsmith M.L., McFarland K., Jeffery K.J., Brufor M.W. & Wickings E.J. 2004: Mitochondrial DNA phylogeography of western lowland gorillas (*Gorilla gorilla gorilla*). *Mol. Ecol. 13: 1551–1565*.

Collura R.V. & Stewart C.B. 1995: Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature 378: 485–489*.

Cracraft J., Feinstein J., Vaughn J. & Helm-Bychowski K. 1998: Sorting out tigers (*Panthera tigris*): mitochondrial sequences, nuclear inserts, systematics, and conservation genetics. *Anim. Conserv. 1: 139–150*.

Desjardins P. & Morais R. 1990: Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. *J. Mol. Biol. 212(4): 599–634*.

DeWoody J.A., Chesser R.K. & Baker R.J. 1999: A translocated mitochondrial cytochrome *b* pseudogene in Voles (Rodentia: *Microtus*). *J. Mol. Evol. 48: 380–382*.

Elias G.L., Reino L.M., Silva T., Tomé R. & Geraldes P. 1999: Atlas das Aves Invernantes do Baixo Alentejo. Sociedade Portuguesa para o Estudo das Aves. *Lisbon, Portugal: 234–235 (in Portuguese)*.

Excoffier L., Smouse P.E. & Quattro J.M. 1992: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics 131: 479–491*.

Felsenstein J. 1981: Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol. 17:368–376*.

Felsentein J. 1985: Confidence limits on phylogenies: an approach using the bootstrap. *Evolution 39: 783–791*.

Hall T.A. 1999: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser. 41:95–98*.

Haring E., Riesing M.J., Pinsker W. & Gamauf A. 1999: Evolution of a pseudo-control region in the mitochondrial genome of palearctic buzzards (genus *Buteo*). *J. Zoolog. Syst. Evol. Res. 37:185–194*.

Harshman J. 1996: Phylogeny, Evolutionary Rates, and Ducks. *Ph. D. Thesis, University of Chicago*.

Haynes S., Jaarola M. & Searle J.B. 2003: Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. *Mol. Ecol. 12: 951–956*.

Johnson K.P. & Clayton D.H. 2000a: Nuclear and mitochondrial genes contain similar phylogenetic signal for pigeons and doves (Aves: Columbiformes). *Mol. Phylogenet. Evol. 14(1): 141–151*.

Johnson K.P. & Clayton D.H. 2000b: A molecular phylogeny of the dove genus *Zenaida*: Mitochondrial and nuclear DNA sequences. *Condor 102: 864–870*.

Johnson K.P. & Clayton D.H. 2001: Coevolutionary history of ecological replicates: Comparing phylogenies of wing and body lice to Columbiform hosts. Tangled Trees – Phylogeny, Cospeciation and Coevolution. *Chicago Univ. Press*.

Johnson K.P., Kort S., Dinwoodey K., Mateman A.C., Cate C.T., Lessels C.M. & Clayton D.H. 2001: A Molecular Phylogeny of the Dove Genera *Streptotelia* and *Columba*. *The Auk 118(4): 874–887*.

Kidd M.G. & Friesen V.L. 1998: Sequence variation in the Guillemot (Alcidae: *Cepphus*) mitochondrial control region and its nuclear homologue. *Mol. Biol. Evol. 15(1): 61–70*.

Kimura M. 1980: A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol. 16: 111–120*.

Klicka J.T. & Zink R.M. 1997: The importance of recent ice ages in speciation: A failed paradigm. *Science 277: 1666–1669*.

Kocher T.D., Thomas W.K., Meyer A., Edwards S.V., Pääbo S., Villablanca F.X. & 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*.

Kumazawa Y., Ota H., Nishida M. & Ozawa T. 1998: The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. *Genetics 150: 313–329*.

Li W.H., Gojobori T. & Nei M. 1981: Pseudogenes as a paradigm of neutral evolution. *Nature 292: 237–239*.

Lopez J.V., Culver M., Stephens J.C., Johnson W.E. & O'Brien S.J. 1997: Rates of Nuclear and Cytoplasmic Mitochondrial DNA Sequence Divergence in Mammals. *Mol. Biol. Evol. 14(3): 277–286*.

Lovette I.J., Berminham E. & Ricklefs R.E. 1999: Mitochondrial DNA Phylogeography and the Conservation of Endangered Lesser Antillean *Icterius orioles*. *Conserv. Biol. 13(5): 1088–1096*.

Maddison W.P. & Maddison D.R. 1999: MacClade: Analysis of Phylogeny and Character Evolution, version 3.08a (Computer program for Macintosh). *Sinauer Associates, Sunderland, Massachusetts*.

Marshall H.D. & Baker A.J. 1999: Colonization history of Atlantic Island Common Chaffinches (*Fringilla coelebs*) revealed by mitochondria DNA. *Mol. Phylogenet. Evol. 11(2): 201–212*.

Mindell D.P., Sorenson M.D. & Dimcheff D.E. 1998: An extra nucleotide is not translated in mitochondrial ND3 of some birds and turtles. *Mol. Biol. Evol. 15(11): 1568–1571*.

Mourier T., Hansen A.J., Willerslev E. & Arctander P. 2001: The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol. Biol. Evol. 18(9): 1833–1837.*

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

Nguyen T.T.T., Murphy N.P. & Austin C.M. 2002: Amplification of multiple copies of mitochondrial Cytochrome *b* gene fragments in the Australian freshwater crayfish, *Cherax destructor* Clark (Parastacidae: Decapoda). *Anim. Genet. 33: 304–308.*

Nielsen K.K. & Arctander P. 2001: Recombination among multiple mitochondrial pseudogenes from a passerine genus. *Mol. Phylogenet. Evol. 18(3): 362–369.*

Pereira S.L. & Baker A.J. 2004: Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics, *BMC Evol. Biol. 4:17.*

Posada D. & Crandall K.A. 1998: Modeltest: testing the model of DNA substitution. *Bioinformatics 14(9): 817–818.*

Quinn T.W. 1992: The genetic legacy of Mother Goose – phylogeographic patterns of lesser snow goose *Chen caerulescens* maternal lineages. *Mol. Ecol. 1: 105–117.*

Rodriguez F., Oliver J.L., Marin A. & Medina J.R. 1990: The general stochastic model of nucleotide substitution. *J. Theor. Biol. 142: 485–501.*

Sambrook J., Fritsch E.F. & Maniatis T. 1989: Molecular Cloning. A Laboratory Manual. *Cold Spring Harber Laboratory Press, New York.*

Sato A., Tichy H., O'hUigin C., Grant P.R., Grant B.R. & Klein J. 2001: On the origin of Darwin's finches. *Mol. Biol. Evol. 18(3): 299–311.*

Schneider-Broussard R. & 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(2): 156–165.*

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.*

Shapiro B., Sibthorpe D., Rambaut A., Austin J., Wragg G.M., Bininda-Emonds O.R., Lee P.L. & Cooper A. 2002: Flight of the dodo. *Science 295: 1683.*

Sorenson M.D., Ast J.C., Dimcheff D.E., Yuri T. & Mindell D.P. 1999: Primers for a PCR-Based Approach to Mitochondrial Genome Sequencing in Birds and Other Vertebrates. *Mol. Phylogenet. Evol. 12: 105–114.*

Sorenson M.D. & 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. 1998: Phylogenetic Analysis Using Parsimony (PAUP). *Sinauer Associates, Sunderland, Massachusetts.*

Tamura K. & Nei M. 1993: Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol. 10: 512–526.*

Väli Ü. 2002: Mitochondrial pseudo-control region in old world eagles (genus *Aquila*). *Mol. Ecol. 11: 2189–2194.*

Wenink P.W., Baker A.J. & Tilanus M.G.J. 1993: Hypervariable-control-region sequences reveal global population structuring in a long-distance migrant shorebird, the Dunlin (*Calidris alpine*). *Proc. Natl. Acad. Sci. U S A 90: 94–98.*

Yang Z., Goldman N. & Friday A. 1994: Comparation of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Mol. Biol. Evol. 11: 316–324.*

Zischler H., Höss M., von Haeseler A., van der Kuyl A.C., Goudsmit J. & Pääbo S. 1995: Detecting dinosaur DNA. *Science 268: 1192–1193.*