

The Old World Sparrows (Genus *Passer*) Phylogeography and Their Relative Abundance of Nuclear mtDNA Pseudogenes

Luis M. Allende, Isabel Rubio, Valentin Ruíz-del-Valle, Jesus Guillén, Jorge Martínez-Laso, Ernesto Lowy, Pilar Varela, Jorge Zamora, Antonio Arnaiz-Villena

Department of Immunology and Molecular Biology, Hospital 12 de Octubre, Universidad Complutense, Madrid, 28041 Spain

Received: 29 November 2000 / Accepted: 22 March 2001

Abstract. The phylogenetic relationships of genus *Passer* (Old World sparrows) have been studied with species covering their complete world living range. Mitochondrial (mt) cyt b genes and pseudogenes have been analyzed, the latter being strikingly abundant in genus *Passer* compared with other studied songbirds. The significance of these *Passer* pseudogenes is presently unclear. The mechanisms by which mt cyt b genes become pseudogenes after nuclear translocation are discussed together with their mode of evolution, i.e., transition/transversion mitochondrial ratio is decreased in the nucleus, as is the constraint for variability at the three codon positions. However, the skewed base composition according to codon position (in 1st position the percentage is very similar for the four bases, in 2nd position there are fewer percentage of A and G and more percentage of T, and in 3rd codon position fewer percentage of G and T and is very rich in A and C) is maintained in the translocated nuclear pseudogenes. Different nuclear internal mechanisms and/or selective pressures must exist for explaining this nuclear/mitochondrial differential DNA base evolutive variability. Also, the phylogenetic usefulness of pseudogenes for defining relationships between closely related lineages is stressed. The analyses suggest that the primitive genus *Passer* species comes from Africa, the Cape sparrow being the oldest: *P. hispaniolensis italiae* is more likely conspecific to *P. domesticus* than to *P. hispaniolensis*. Also, *Passer* spe-

cies are not included within weavers or *Estrildinae* or *Emberizinae*, as previously suggested. European and American *Emberizinae* sparrows are closely related to each other and seem to be the earliest species that radiated among the studied songbirds (all in the Miocene Epoch).

Key words: Mitochondrial DNA — Nuclear pseudogenes — Sparrow — Passerines — *Emberiza* — *Petronia*

Introduction

The Old World sparrows (genus *Passer*) are probably the most familiar group of bird species, since their worldwide distribution reflects that of humans. House sparrow (*Passer domesticus*), Spanish sparrow (*P. hispaniolensis*), tree sparrow (*P. montanus*), and others can be regarded as human commensals; also, most *Passer* species nest in human-made constructions. The relationships among the sparrow species within genus *Passer* and to other finches (*Passeridae* and New World sparrows) have not been fully resolved. On the other hand, it is widely believed that Pleistocene temperature variations (glaciations) and subsequent isolation are the most important factors provoking the appearance of new extant bird species (Gill 1995). Recent contradictory evidence suggests that speciation of some genera and orders may have already occurred a long time ago (Chiappe 1995; Feduccia 1995; Hackett 1996; Hedges et al. 1996; Härlid et al. 1997), particularly in *Passerines* (Klicka and Zink 1997) and in *Carduelinae* (Marten and Johnson 1986; Fehrer 1996; Arnaiz-Villena et al. 1998, 1999a). In the

Table 1. List of species, origin, mt cytochrome b, and nuclear pseudogenes sequence identification

Species	Mt cyt b sequence	Nuclear pseudogenes sequence	Sample region
1 House Sparrow (<i>Passer domesticus domesticus</i>)	AF230906	AF230916	Madrid, Spain
2 Sudan Golden-Sparrow (<i>Passer luteus</i>)	L76714	AF230919	Dakar, Senegal
3 Eurasian Tree Sparrow (<i>Passer montanus montanus</i>)	AF230911	—	Madrid, Spain
4 Pegu Sparrow (<i>Passer flaveolus</i>)	AF230907	—	Ho-Chi-Minh city, Vietnam
5 Cape Sparrow (<i>Passer melanurus melanurus</i>)	AF230910	—	Capetown, South Africa
6 Spanish Sparrow (<i>Passer hispaniolensis hispaniolensis</i>)	AF230909	AF230918	Gran Canaria, Canary Islands, Spain
7 Spanish Sparrow (<i>Passer hispaniolensis italiae</i>)	AF255708	AF255707	Piacenza, Italy
8 Russet Sparrow (<i>Passer rutilans rutilans</i>)	AF230912	—	Beijing, China
9 Grey headed Sparrow (<i>Passer griseus griseus</i>)	AF230908	AF230917	Dakar, Senegal
10 Saxaul Sparrow (<i>Passer ammodendri ammodendri</i>)	AF230905	AF230915	Almaty (Kazakhstan), Russia
11 Rock Sparrow [†] (<i>Petronia petronia petronia</i>)	AF230914	—	Madrid, Spain
12 Bronze Mannikin [#] (<i>Lonchura cucullata cucullata</i>)	AF255705	—	Kimshasa, Zaire
13 Cardinal Quelea ^{*#} (<i>Quelea cardinalis</i>)	AF255709	—	Kimshasa, Zaire
14 Chaffinch [§] (<i>Fringilla coelebs coelebs</i>)	L76609	—	Madrid, Spain
15 Corn Bunting ^{&} (<i>Miliaria calandra calandra</i>)	AF255706	—	Madrid, Spain
16 Lark Sparrow [‡] (<i>Chondestes grammacus grammacus</i>)	AF255704	—	Mexico DF, Mexico
17 Chipping Sparrow [‡] (<i>Spizella passerina passerina</i>)	AF255710	—	Mexico DF, Mexico

All specimens studied are males and belong to the Families *Passeridae* or *Fringillidae*. **Passeridae:** Old World sparrows (*Passerinae*) Genus *Passer* [†]Rock-sparrows (*Passerinae*), [#]African Estrildine (*Estrildinae*). ^{*}Weaver (*Ploecinae*), **Fringillidae:** [§](*Fringilinae*), [&]Old World Emberizid. [‡]New World Emberizid (Sibley and Monroe 1990).

present work, we have collected *Passer* and other related species samples from around the world (*Passeridae*; Sibley and Monroe 1990) in order to sequence an orthologous gene from each of them: the mt cyt b (924 bp). Mitochondrial DNA has proven to be helpful for defining the evolutionary relationships among relatively distant and closely related birds (Arnason and Gullberg 1994; Seutin et al. 1994); thus, we have also aimed to study the relatedness of these bird species in the context of the paleogeography and molecular clock timing, in order to get an overall picture of the evolution and the relative time of origin of these birds (Paturi 1991, pp. 284–496; Smith et al. 1994; pp. 24–39; Cox and Moore 1995, pp. 134–276). In addition, the genetic relationships between some of the *Passeridae* subfamilies (Sibley and Monroe 1990; i.e., *Passerinae*, *Ploecinae*, *Estrildinae*) have been studied. New and Old World *Emberizinae* sparrow-like finches are also used for determining the corresponding phylogenetic relationships. Finally, the

very striking observation that many mitochondrial cyt b (probably nuclear) pseudogenes were only found at very high frequency in *genus Passer* species among the *Passerine* birds is discussed in the light of the evolutionary significance and phylogenetic usefulness of this observation.

Materials and Methods

Bird samples come from species and places that are described in Table 1. GenBank sequence accession numbers are also given. Blood from living birds was drawn after photographing and cutting the claws locally anesthetized with a lidocaine ointment. Blood was collected in EDTA cooled at 4°C and frozen until use. DNA was obtained and mitochondrial cytochrome b gene (mt cyt b) (924 DNA bases) was amplified with primers L14841 5'-AAAAGCTTCCATCCAA-CATCTC AGCATGATGAAA-3' and H15767 5'-ATGAAGGGAT-GTTCTACTGGTTG-3' as detailed by Edwards et al. (1991). At least two birds per species were sequenced in order to discard any variation

Table 2. Matrix of pairwise Kimura two-parameter distances (%) in cyt b genes (below diagonal) and cyt b nuclear pseudogenes (above diagonal)

Species ^a	1	2	6	7	9	10
House sparrow(1)		3.81	9.34	0.98	4.30	4.16
Sudan golden sparrow(2)	7.76		9.61	3.92	3.72	3.81
Spanish sparrow "hispaniolensis"(6)	3.01	6.76		9.46	9.67	9.48
Spanish sparrow "italiae"(7)	0.54	7.63	2.66		4.41	4.27
Grey headed sparrow(9)	8.61	7.74	8.34	8.61		3.02
Saxaul sparrow(10)	8.74	8.48	7.97	8.48	7.85	

^a Numbers in parentheses after species names correspond to column numbers to the right.

among individuals. Polymerase chain reaction (PCR), cloning, and automatic DNA sequencing were performed as previously described (Edwards et al. 1991; Arnaiz-Villena et al. 1992). At least four clones from each of two different PCRs were sequenced from each species in order to assess both PCR and DNA sequencing quality. For comparing the specific existence of mitochondrial nuclear pseudogenes in *genus Passer*, 120 other species belonging to *Order Passeriformes* (including 20 different *genera*), and the following families/subfamilies: *Corvidae*, *Dicrurinae*, *Vangini*, *Turdinae*, *Muscicapinae*, *Sturnidae*, *Poliophtinidae*, *Paridae*, *Alaudidae*, *Passerinae*, *Ploceinae*, *Estrildinae*, and *Fringillidae* were tested (Arnaiz-Villena et al. 2001). Eight mt cyt b clones were analyzed for each bird and an average of five clones per individual were suitable to be fully analyzed in the automatic sequencer because an average of one to three clones showed PCR mistakes and were discarded. Three different phylogenetic tree-constructing methodologies were used to confirm independently the robustness of the topologies (Härlid et al. 1997): unweighted parsimony, neighbor joining (NJ), and unweighted pair group with arithmetic mean (UPGMA). The matrix of genetic distances for the NJ tree was obtained by the maximum-likelihood method, and Kimura two-parameter distances were used for the UPGMA dendrogram. The UPGMA tree was also obtained for estimating relative coalescence times from known outgroups' divergence times (Vincek et al. 1997). Times of species divergence are only a rough and relative estimate. Maximum parsimony (MP), NJ, and UPGMA dendrograms and their corresponding biparametric distance matrices were obtained with the PAUP*4.0b2 program, kindly provided by D.L. Swofford (1996). The following statistical calculations were carried out: (1) number of substitutions of the two group of sequences, mitochondrial and nuclear, (2) their transitions and transversions ratios, and (3) the base composition according to codon position in both genes and pseudogenes. Bootstrap values were calculated to test the topology robustness of trees (Felsenstein 1985), and low-bootstrap value branches are shown because a similar tree branch topology is obtained by at least two different tree construction methodologies, as suggested by Edwards 1995. The LINTRE test was also performed in order to test the molecular clock homogeneity among all lineages used in our analyses (Takezaki et al. 1995); thus, linearized trees were obtained, which reestimate the branch lengths under the assumption of a constant rate of evolution. The computer software used for these calculations can be obtained from the web site: <ftp://ftp.bio.indiana.edu/molbio/evolve/lintr/> or <http://cib.nig.ac.jp/dda/ntakezak.html>.

The universal criterion for defining a *Passer* pseudogene as such is the existence of a stop codon at position 77 (nuclear reading code). In addition, different anomalies occur, such as deletions (see below).

In order to show the evolutionary difference between mt and nuclear pseudogenes, pairwise comparisons of species represented by both the mt sequence and the nuclear counterpart were done (Table 2). The mean sequence divergence was also calculated for each data par-

tion by adding the respective pairwise divergences (%) and divided by the total number of comparisons done (see Table 2).

Results and Discussion

Patterns of Base Substitution

Saturation was considered to have occurred in any of the data partitions if the scatter of points showed a levelling off of mutations as sequence divergence increased. Saturation plots for cyt b DNA (Fig. 1) indicated that only third position transitions showed a clear levelling off associated with saturation; this occurred at about 13% uncorrected total sequence divergence (*Passer/Petronia*) and at sequence divergences of more than 13% in the comparisons of *Passer/Gallus* (see Fig. 1). Assessment of saturation gave similar results by using Kimura's two parameter distances. Also, the number of variable and phylogenetically informative sites (403 and 278, respectively, out of 924 mt cyt b DNA bases), between the birds described in Table 1, was appropriate to establish sound phylogenetic comparisons (Hillis et al. 1994).

The nucleotide distribution pattern of the mt cyt b gene of the *Passer* birds under study was similar to the pattern found in a previous analysis of this gene in birds and mammals (Kornegay et al. 1993). At the first codon positions, the four bases were equally distributed; at the second position, fewer G residues and a higher amount of T were seen. At the third codon position, the bias against G and T was strong, as previously found by others (Edwards et al. 1991; Kornegay et al. 1993). This bias in base composition was similar in all species studied (results not shown). Thus, our parsimony methodology seems to be adequate for all species studied (Lockhart et al. 1994). Our cyt b gene variability was theoretically sufficient to establish phylogenetic relationships according to the number of phylogenetically informative sites (Hillis et al. 1994); most of the differences were silent substitutions, as expected for a protein-coding gene, particularly for close relatives, such as species within a single *genus*, i.e., *Passer*, (Kocher et al. 1989). Within species variability of the mt cyt b sequence was very low in the *Passer* species tested (between 0.0 and 0.1%). Our data are concordant with those already published (Avisé and Ball 1991). Therefore, within species variability was not likely to interfere with interspecific comparisons which were in a higher range (see Fig. 2).

As expected for gene evolving relatively rapidly under strong functional constraints, most (53%) of the third codon positions among *Passer* species, in which substitutions are often silent, were variable. By contrast, relatively few of the first and second positions, (9% and 1%, respectively) were variable. Therefore, more than three-quarters (84%) of the variable sites occurred in third positions of codons. A similar evolution rate for the mt cyt b DNA was found in the subfamilies *Estrildinae*, *Ploceinae*, and *Emberizinae* (56% of variable sites in the third position of codons, and 15% and 6% of the first and

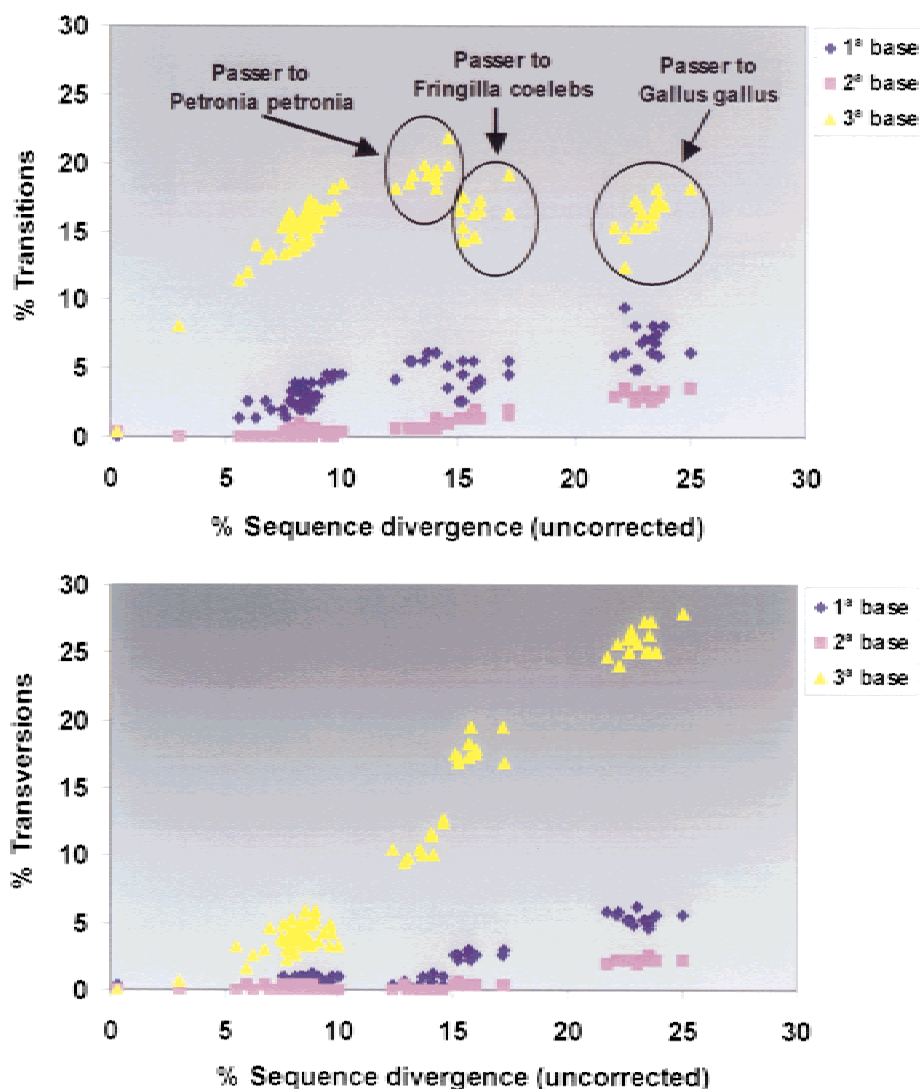


Fig. 1. Saturation plots for the cytochrome b gene that relate uncorrected sequence divergence to changes due to transitions (top) and transversions (bottom) at first, second, and third codon positions.

second positions, respectively, were variable; results not shown).

Also, an attempt to estimate the relative tempo of evolution of *Passer* and other sparrows and finches was undertaken: in order to estimate the tempo of evolution, the calculations done by Takahata, Klein, and ourselves (Vincek et al. 1997, Arnaiz-Villena et al. 1998, 1999a) to assess the time of appearance of Darwin's finches, canaries, and goldfinches were followed. An UPGMA dendrogram (Fig. 2) was constructed because this type of phylogenetic tree is more suitable for estimating coalescence times than other methods, particularly when a molecular clock exists (shown in this case by the LINTRE test, see below) (Nei 1987). The UPGMA tree was obtained by comparing the mt cyt b DNA sequences of the pheasant (Wittzell et al. 1994) and the chicken (Zoorob et al. 1990), two species that have diverged around 18 MYA (Helm-Bychowski and Wilson 1986), not shown in Fig. 2. The mt cyt b comparison of genus *Passer* yields

an evolutionary rate (per lineage) of 0.68×10^{-9} non-synonymous substitutions per nonsynonymous site per year and 1.41×10^{-8} synonymous substitutions per synonymous site per year, so that the overall rate is $4.0 \times 10^{-9} \pm 0.26$ (Vincek et al. 1997). This results in a substitution rate of 0.4% per million years, which leads to a rough approximation to the 4% of nucleotide substitution per lineage between the most distant *Passer* species. Taking into account that the *Passer* mt DNA substitution rate was found between the most distant species (approximates to 4%), the *Passer* radiation seems to have started shortly before that of *Serinus* and *Carduelis* genera (Arnaiz-Villena et al. 1999a and unpublished), about 11 MYA (not shown in Fig. 2). However, the substitution rate found by us (0.4% per million years) differs from the standard 2% per million years. Also, cranes show a faster rate of nucleotide substitution, closer to our results found in *Carduelis*, *Serinus*, and *Passer*; (about 1%) (Krajewski and King 1996).

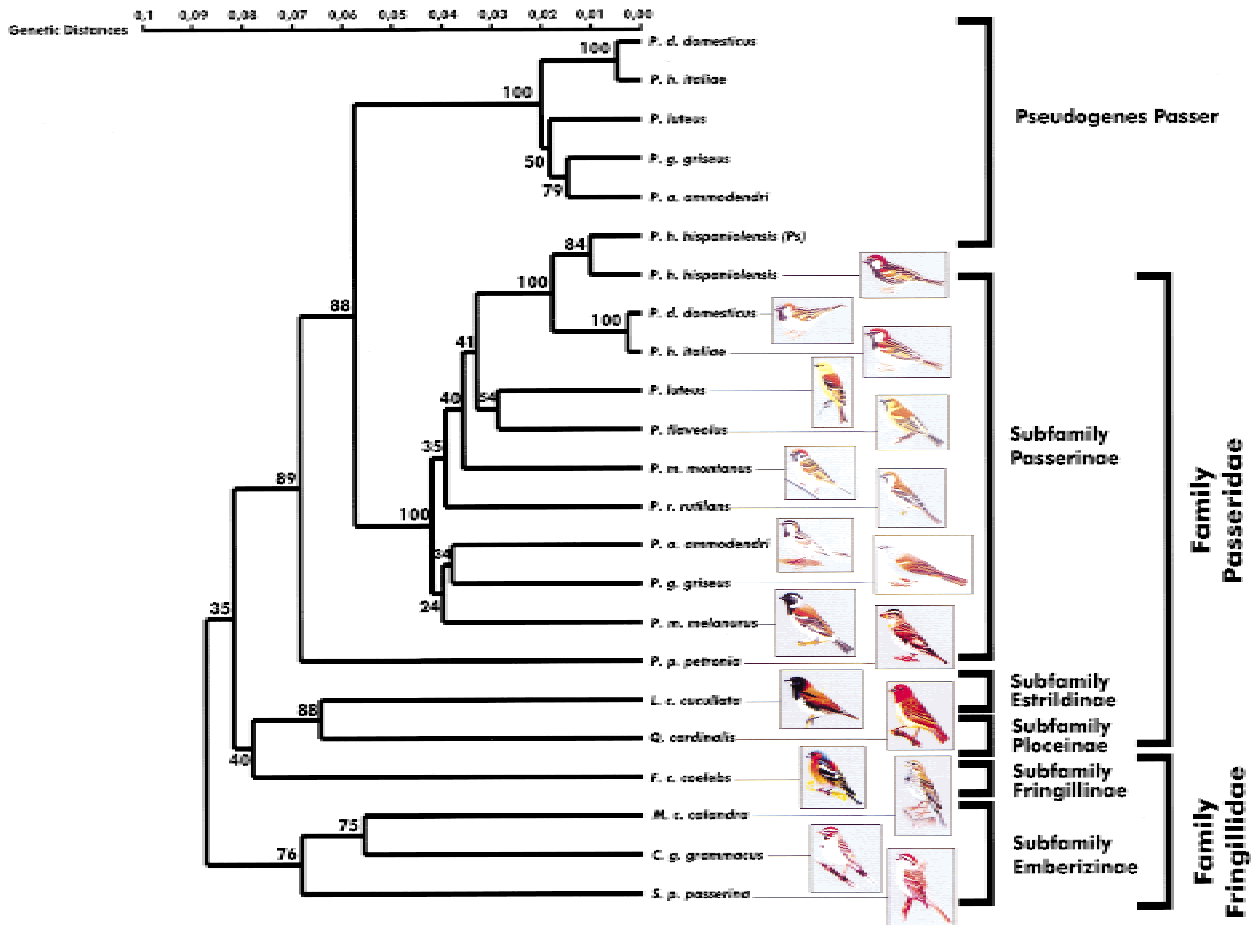


Fig. 2. Relative time of appearance of *genus Passer* and other sparrows lineages. UPGMA methodology tends to perform poorly if the assumption of equal rate of *cyt b* evolution among species does not reflect their actual evolution; however, LINTRE test shows that a molecular clock exists (see text). Also, it seems to perform correctly in the closely related bird species used for this work since groups of taxa are

Notwithstanding, the calculation of species divergence times needs to be confirmed by other methodologies and with additional species in this particular study and others (i.e., Härlid et al. 1997). In the later study, *Passerines* are found to be older than paleognathous birds; also, the relative close-to-*Passer Carduelini* (*Carduelis*, *Serinus*, and others) species are considered nearly as old as *Passer* (Marten and Johnson 1986; Fehrer 1996; Arnaiz-Villena et al. 1998, 1999a, 2001).

The use of either Chaffinch or chicken and pheasant as outgroups would seem to be correct, because only third position transitions appear to be saturated (Fig. 1). Notwithstanding, the LINTRE computer program (Takezaki et al. 1995) was used to assess constancy of evolutionary rates among the used bird lineages. The “two cluster test” shows that the pseudogene cluster (with the exception of *P. hispaniolensis* pseudogene which clusters with the genes, see Fig. 2 and Fig. 3) evolves with a significantly different rate (>1%). Other analyzed sequences do not show a significant heterogeneity of evolutionary rate. Two linearized trees were ob-

similar to those obtained in NJ and parsimony dendrograms (see Fig. 3). Bird species used are detailed in Table 1; Ps: note that the *P. h. hispaniolensis* pseudogene cluster with the genes (see text). The absolute timing is uncertain; however, our data support the Miocene/Pliocene species appearance.

tained (one for the genes and one for the pseudogenes) by also using the LINTRE software; these trees do not show significant differences from our UPGMA tree in branch length (Fig. 2) and their topology was also identical. This supports the fact that a constant evolutionary rate there exists among different gene sequences and the validity of our time calculations.

Phylogeography of Passer species

In the present work, the complete geographical range of Old World sparrows distribution is covered with the studied species (i.e., Africa and Eurasia) (Table 1). The African grey-headed (*P. griseus*) and black-headed (*P. melanurus*) sparrows together with the Saxaul sparrow (*P. ammodendri*) seem to form a different clade in relation to the other sparrows in all the trees (Figs. 2, 3a, 3b). This clade seems to have originated earlier (Fig. 2) and this grouping suggests that (1) Saxaul sparrow may also have had its living range in the African deserts at

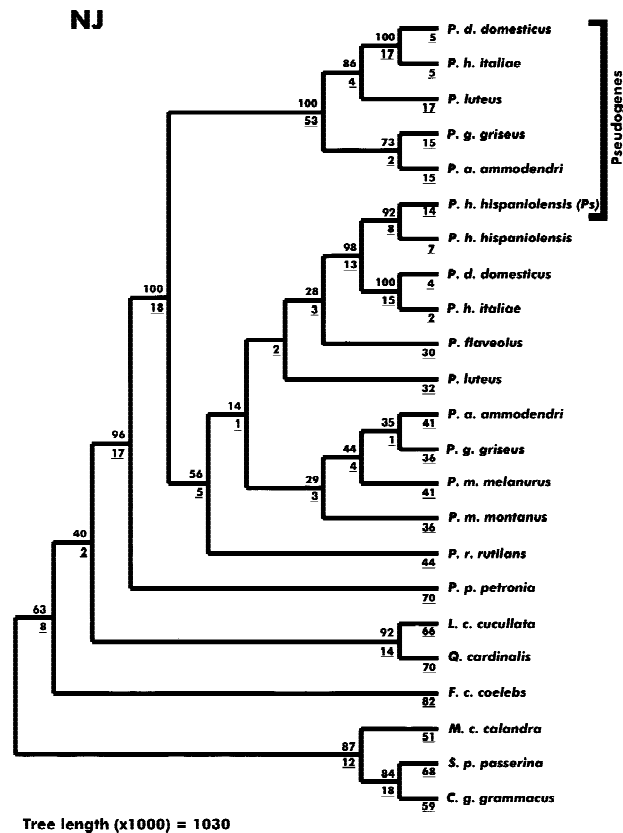
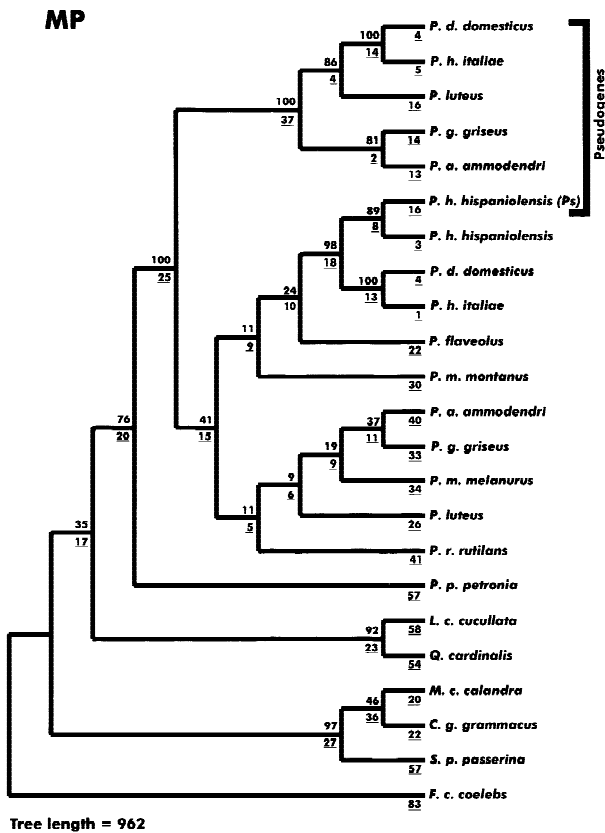


Fig. 3. Maximum unweighted parsimony tree. Heuristic search was used (PAUP). Only one tree was the best parsimonious. Consistency and retention indexes were 0.57 and 0.57, respectively. The majority rule bootstrap consensus tree of mt *cyt b* genes and nuclear pseudogenes is shown. First, second, and third codon bases were used unweighted. Parsimony was used unweighted because weighting is only recommended for greater amounts of evolutionary divergence, which are not expected among the relatively closely related species analyzed (Hillis et al. 1994). Parsimony bootstrap analysis was done with 1,000 replications, and values (in percent) shown above branches. The num-

ber of events is shown underlined below branches. Bird species used are detailed in Table 1. Ps: Note that the *P. h. hispaniolensis* pseudogene clusters with the genes (see text). b. Neighbor-joining bootstrap tree (1,000 replications) based on 924 bases of *cyt b* genes and nuclear pseudogenes. Bootstrap values are shown above branches; branch lengths ($\times 1,000$) are shown underlined below branches. The evolutionary model used was “the minimum evolution.” Distance matrices were calculated based on maximum likelihood analysis (Swofford 1996). Bird species used are detailed in Table 1. Note that the *P. h. hispaniolensis* pseudogene clusters with the genes (see text).

one time (Sahara, Arabian Peninsula) and (2) Old World sparrows originated in Africa. The five species of African grey sparrows may all belong to one species (Summers-Smith 1988). The species studied by us (*P. griseus*) is related to the Cape Sparrow (*P. melanurus*) (Figs. 2, 3a, 3b), the latter living only in the southern Sahara African range where grey sparrows do not thrive (except *P. diffusus*).

The oldest *genus Passer* species seems to be the South African *P. melanurus* (Cape sparrow) (Fig. 2); it also clusters together with the African species *P. griseus* (grey-headed sparrow). The latter is a representative of the grey-headed sparrows group, which thrive from the South Sahara down to South Africa. Other *Passer* species should also be tested, particularly African ones (although our study includes representatives for each African group, i.e., grey-headed, yellow ones). *P. luteus* lives in Africa in a land band just bellow the Sahara Desert (Sahel) from the East to the West Coast (Summers-Smith 1988); its position in the phylogenetic trees is ambiguous.

It may or may not be closely related to the other African sparrows. However, it is clear that the claims that *P. luteus* does not belong to *genus Passer* is not supported by our genetic data. It may have originated in Sudan and might have expanded from West to East following the hunter-gathered Cushitic people emigration (McEvedy 1980; Summers-Smith 1988). *P. ammodendri* (the single sparrow species, other than *P. melanurus*, that has a black nape) is tightly linked to *P. melanurus*; this could have given rise to mutation(s) that diluted melanin at various parts of the head and body and could have originated the grey-headed sparrows, including the *P. griseus* group—*P. ammodendri* and *P. domesticus* (grey-patched head). *P. ammodendri* is now restricted to the Asian deserts but may have also been originated in Africa and inhabited its deserts in the past, like *P. simplex* (with which it shares a common living range in Asia) (Summers-Smith 1988). *P. ammodendri* may have been displaced by a recent introduction of *P. simplex* in Africa.

The origin of *genus Passer* is likely to be confined to

Africa because of the highest number of extant species on this continent (Summers-Smith 1988). Our results agree with this assertion, *P. melanurus* being the oldest extant species: it would have been the origin of grey-headed African sparrows (including *P. domesticus*) and also yellow and other Palearctic bab-sparrows. It has been postulated that the bab-sparrows have probably arisen at the Nile or Rift Valley and followed the human expansion towards Eurasia, giving rise to all Eurasian bab-species (Summers-Smith 1988). That this could have happened would be more feasible if the human expansion had occurred much earlier than thought (5 million years ago. Not shown; see Fig. 2). *P. rutilans* (Asia) seems to be the oldest of the Eurasian sparrows although its phylogenetic placing within the genus is undefined (Figs. 2, 3a, 3b). The genus *Passer* is strongly associated with man, since most of species nest on man-made buildings.

Passer domesticus and *P. hispaniolensis* cannot be distinguished osteologically and their ancestor may have appeared 4 MYA (Fig. 2 and not shown). Fossil evidence for this precursor is found 350,000 years ago in Palestine (Summers-Smith 1988). However, the Pleistocene origin for sparrows (this paper) and other *Passerines* (including *Serinus* and *Carduelis*; Klicka and Zink 1997; Arnaiz-Villena et al. 1998, 1999a) could be placed much further back into the Miocene or Pliocene Epochs, Pleistocene being more important in subspeciation. Our Fig. 2 suggests that *P. domesticus* and *P. hispaniolensis* subspeciation may have occurred in the Pleistocene. However, *Passerines* appearance timing are still much debated and a Pleistocene origin for species is also suggested (Avice and Walker 1998).

The following conclusions about the much debated systematic uncertainties within Old World sparrows (Summers-Smith 1988) have been reached in the present study:

- 1) The origin of genus *Passer* seems to be African because the highest number of extant species is African and our phylogenetic results are also suggestive (Figs. 2, 3a, 3b); this is in accordance with Summers-Smith 1988. *P. melanurus* or other related extinct species might be the parental one.
- 2) *P. melanurus* (Cape sparrow) is found to be related to the grey-headed African sparrow and also to *P. ammodendri* (Saxaul sparrow) (Figs. 2, 3a, 3b). All three species males have melanic pigmentation in the head crown and nape (Figs. 2, 3a, 3b). Also, *P. domesticus* may be included in the grey-headed sparrows because of its grey head crown (Fig. 2, picture), although no precise placement is found in our phylogenetic trees, probably because more African species need to be studied.
- 3) *P. hispaniolensis italiae* is probably a *P. domesticus* (grey head crown) subspecies according to the re-

spective branch lengths observed in NJ and MP dendrograms (Fig. 3a and 3b) and not a *P. hispaniolensis* subspecies (brown head). This is in accordance with the classical view, but not with more recent opinions (Summers-Smith 1988). *P.h. italiae* (brown head) may have arisen from *P. domesticus* (grey head crown) by hybridization or divergence/speciation.

On the other hand, the relationships of Old World sparrows (*G. Passer*) to other *Passeridae*: Old and New World *Emberizinae* sparrows, weaver birds (*Ploceinae*), and *Estrildinae* have been studied. Rock sparrow (genus *Petronia*) is the closest relative studied to Old World sparrows within the Subfamily *Passerinae*; *Petronia* is more robust in appearance and has specialized habitat requirements (drier and more arid areas). Its distribution is African and Euroasiatic and according to Figs. 2, 3a, and 3b, seems to be the sister group to genus *Passer* but stands as a separate lineage (Clement et al. 1993, pp. 442–468). African *Passeridae*, *Lonchura cucullata* (*Estrildinae*), and *Quelea cardinalis* (a weaver-bird, *Ploceinae*) clade together corroborating their relatedness; however, genus *Passer* is not included within the *Ploceinae* (Figs. 2, 3a, 3b) in contrast to the suggestion of previous studies in which they were included within weavers by some authors based mainly on skeletal characters, nest building, and DNA hybridization (Summers-Smith 1988). *Emberizinae* are also separate from *Passer* in spite of suggestions to join both families according to egg-white protein studies (Summers-Smith 1988). *Fringilla coelebs* (the cropless Chaffinch) is closer to *Lonchura* and *Quelea* than to the *Emberizinae* in the UPGMA and NJ dendrograms; however, it stands as a separate lineage in the parsimony tree. It is clear from all the trees that Old and New world *Emberizinae* sparrows represent a separate (and single) radiation compared to both *Passeridae* family and Chaffinches (*Fringillidae*). This is concordant with the fact that other *Fringillidae* (Goldfinches and Canaries, Arnaiz-Villena et al. 1998, 1999a) do not cluster with any finch of the *Passeridae* family (Sibley and Monroe 1990), described in this paper (Arnaiz-Villena et al. 2001, and unpublished results). Also, the American species *Spizella passerina* may represent the oldest split within the *Emberizinae* (Fig. 2); this may suggest that *Emberizinae* finches originated in America, although the study of more species is needed.

Abundance of Nuclear Mitochondrial cyt b Pseudogenes in genus Passer and Their Characteristic Mode of Evolution

Mitochondrial DNA fragments have been found in the nuclear DNA of yeast, locust, *Podospora*, sea urchin, maize, rat, human (Fukuda et al. 1985), and birds (Arctander 1995, for review see Blanchard and Lynch 2000). Although the mtDNA fragments seem to have

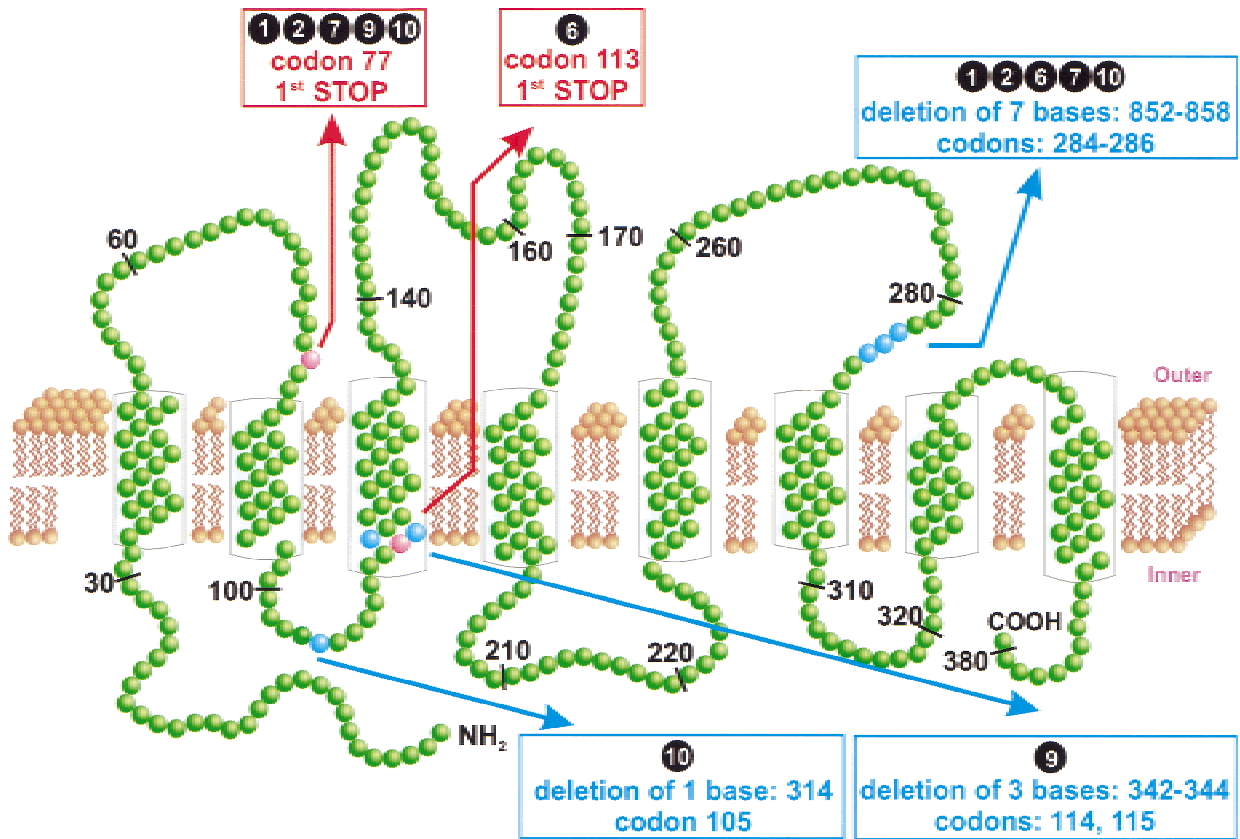


Fig. 4. Mitochondrial cyt b protein. Black circles refer to the species numbered in Table 1, i.e., 1 = *Passer domesticus*, 2 = *Passer luteus*, 6 = *Passer hispaniolensis hispaniolensis*, 7 = *Passer hispaniolensis italiae*, 9 = *Passer griseus griseus*, 10 = *Passer ammodendri ammodendri*. First stop codons (red square) and deletions (blue square) of nuclear pseudogenes are depicted in a cytochrome b molecule. More stop codons were found along the nuclear pseudogenes. The mechanism that generates the first stop codons is simply a translocation of mitochondrial genes to the nucleus (tr), and true base change (tc) may also occur; normal mt cyt b gene was used for comparisons. The type of observed stop codons and their order are the following: (a) In *P.*

domesticus, tr: 113, 141, 163, 165, 272 and tc: 77, 89, 107, 223, after codon 284 (deletion of 7 bases) the open reading frame was lost. (b) In *P. luteus*, tr: 77, 113, 141, 163, 165, 272 and no tc, after codon 284 (deletion of 7 bases), the open reading frame was lost. (c) In *P. h. hispaniolensis*, tr: 113, 135, 141, 163, 165, 272 and no tc, after codon 284 (deletion of 7 bases) the open reading frame was lost. (d) In *P. h. italiae*, tr: 113, 141, 163, 165, 272 and tc: 77 and 89; after codon 284 (deletion of 7 bases) the open reading frame was lost. (e) In *P. griseus*, tr: 77 and 113 and no tc, after codon 114 (deletion of 3 bases) the open reading frame was lost. (f) In *P. ammodendri*, tr: 77, and tc: 104, after codon 105 (deletion of 1 base) the open reading frame was lost.

been continuously integrated into nuclear DNA during evolution, the evolutionary significance and mechanisms of such phenomena are unclear. The present study describes for the first time the relative abundance of integrated mt cyt b pseudogenes among species of a single genus (*Passer*). The different types of alterations found in *Passer* mt cyt b pseudogenes are depicted in Fig. 4. About 610 *Passerine* mt cyt b DNA clones have been sequenced and studied in our laboratory during the last few years (see Materials and Methods) and we only found pseudogenes within genus *Passer* species. Not even a single pseudogene has been detected in other *Passerine* (or non-*Passerine*) species (Arnaiz-Villena et al. 2001), including *Carduelis*, *Serinus* (Arnaiz-Villena et al. 1998, 1999a), and in any other related-to-*Passer* genus, like the ones studied in the present paper. Also, a Fischer test was performed: 0 pseudogenes found in 610 clones belonging to non-*Passer* finches were compared with 39 pseudogenes found in 65 clones belonging to

Passer species; $P < 0.00001$ was obtained, supporting the fact that the pseudogenes finding in *Passer* species was not due to chance. In fact, the number of pseudogene and gene clones obtained were, respectively: *P. domesticus* (4 and 5), *P. luteus* (4 and 5), *P. h. hispaniolensis* (1 and 3), *P. griseus* (3 and 6), *P. h. italiae* (3 and 3), and *P. ammodendri* (24 and 4).

Our phylogeny suggests that the most ancient anomaly (Fig. 4) in pseudogenes is a stop signal at codon 77 seen in *P. griseus*, *P. ammodendri*, *P. luteus*, *P. domesticus*, and *P. hispaniolensis italiae*. This stop codon only appeared because of the mitochondrial-to-nuclear translocation itself in the first three species: it is a stop codon in the nuclear code, but not in the mitochondrial one (see Fig. 4 caption). It may have first occurred in *P. griseus* (or any of the African grey-headed group) after separation of the *P. melanurus* lineage or their extinct ancestor(s). Fourteen clones have been studied in *P. melanurus* and no pseudogene has been found, but

there is still the possibility that pseudogenes are present in all *Passer* lineages and they have not been found by us in some species by chance. The fact that both *P. hispaniolensis italiae* and *P. domesticus* have this first stop codon because of a true base change (TGG→TGA, see Fig. 4 caption) further corroborates that their relatedness is closer than that of *P. hispaniolensis hispaniolensis* and *P. hispaniolensis italiae* because the former has a different stop codon (Fig. 4). Thus, evolutive pressures at codon 77 exist to render it as a stop signal in both of these species. Position 77 is placed on the extramitochondrial region of the molecule (Fig. 4) and it shows a low change rate and probably is an integral part of the Qo redox core; most of the outer surface is implicated in the Qo redox center and this appears to be a major contributor to the reduced evolutionary rate for the outer surface (Howell 1989). Deletions may be secondary to the appearance of this 77 stop codon. A number of stop codons may only be generated because of the putative mitochondrial/nuclear translocation (Fig. 4, caption), once the codon 77 is read as “stop,” the translocated DNA may escape then to the translation and transcription gene machinery control.

The fact that *P. griseus* and *P. ammodendri* show the highest number of anomalies in pseudogenes also supports the fact that they belong to the oldest group of sparrows (Fig. 4). A stop codon at 113 has been found in the *P. h. hispaniolensis* pseudogene (Canary Islands) and codon 77 was not a stop codon either in mitochondrial or nuclear reading code in this particular bird. This pseudogene clusters with the normal mt cyt b genes in the dendrograms (Figs. 2, 3a, 3b), suggesting that the translocation has occurred independently and more recently in the relatively isolated Canary Islands sparrow populations and that the pseudogene is in the process of accumulating more differences with respect to its mitochondrial gene. This pseudogene has been detected in three unrelated birds.

Pseudogenes may also be useful to ascertain phylogenies in birds; because the nuclear copy has a slower substitution rate, it is more likely to be more similar to the ancestral form (Fukuda et al. 1985; Arctander 1995). In fact, original African sparrows (*P. griseus* and *P. ammodendri*) cluster together and are separated from *P. domesticus* and *P. h. italiae*, the latter being a subspecies of the former (Figs. 2, 3a, 3b). *P. luteus* is probably linked to the Mediterranean sparrows (in its origin) (Summers-Smith 1988), i. e., to *P. domesticus* which probably thrived close to the first urban nucleus around the once fertile Sahara (Arnaiz-Villena et al. 1999b) and/or the Nile Valley (Summers-Smith 1988). Also, higher bootstrap values are found in the pseudogene tree (as previously described by Arctander 1995); this may help to define phylogenetic relationships between closely related species, as in our case.

Evolutive constrictions has also been observed in the

Passer nuclear pseudogenes. First and second codon position substitutions in mitochondrial genes of six species amounted to 137, whereas third codon position substitutions amounted to 756 (ratio $756/137 = 5.5$). In the case of 6 nuclear pseudogenes (each belonging to a different species), first and second codon position substitutions were 323 and third codon position substitutions were 381 (ratio $381/323 = 1.18$). A constriction for third codon position substitution was clearly found in the nuclear mitochondrial pseudogenes and also a higher constriction at the first and second codon positions was observed, when compared with the mt genes. Arctander (1995) did not distinguish the substitution rate according to codon position because his data set did not contain several species within a single *genus*. In addition, and in order to gain a more complete picture for the mode of evolution of mitochondrial pseudogenes in birds, the nucleotide distribution patterns of the mt cyt b genes and pseudogenes were analyzed and found to be similar in both nuclear pseudogenes and mitochondrial genes to the patterns found by others for mt cyt b in birds and mammals (Kornegay et al. 1993; Hackett 1996; Arnaiz-Villena et al. 1998; 1999a). At the first codon positions, the four bases were equally distributed; at the second position, fewer G residues and a higher amount of T were seen. At the most variable third codon position, the bias against G and T was strong, as previously found by others (not shown, Edwards et al. 1991; Kornegay et al. 1993). Thus, it may be postulated that the evolutive mechanisms that maintain general constraints for both functional DNA and pseudogenes variation and reduce the DNA base change rate are different from those that preserve the specific skewed base variation; the latter, and not the former, are universally maintained (both in nucleus and mitochondria) in mammals and birds. This is confirmed by our results in *Passer* species (not shown).

The ts(498)/tv(232) ratio in six *Passer* pseudogenes was 2.14 and the same ts(738)/tv(155) ratio in six *Passer* genes was 4.80; this relative constriction for transitions (or the lack of bias against transversion) that occurs in the bird nucleus for a mt gene was also found by others (Arctander 1995).

The evolutionary difference (% mean sequence divergence) is lower in pseudogenes (5.6%) than in genes (6.9%) (Table 2); however, the pseudogene mean difference is much lower (3.6%) if *P. h. hispaniolensis* pseudogene is left out. This pseudogene clusters with genes in the dendrograms (Figs. 2 and 3) and is supposed to have recently been translocated into the nucleus, thus showing only a few anomalies (see above).

The evidences suggesting that we are dealing with nuclear pseudogenes are shown in Fig. 4. This is also shown by the distinctive mode of mitochondrial and nuclear gene evolution, i.e., ts/tv ratio, nuclear constraints for DNA base variability, and phylogenetic studies that show how they cluster together in dendrograms

(Figs. 2, 3a, 3b, Arctander 1995). Therefore, it may be postulated that translocation of mitochondrial genes to the nucleus is a common phenomenon in nature (Fukuda et al. 1985; Collura and Stewart, 1995; Blanchard and Lynch 2000), and may occur together with the generation of *genera* and species i.e., in our *genus Passer* it is not found in the postulated ancestor species (*Petronia*) and other species. The pathway of mtDNA integration into nuclear DNA is mediated by mechanisms similar to those seen in the case of Simian Virus 40 integration to nuclear DNAs, but not to those observed in retroviral integration (Fukuda et al. 1985). By studying closely related species like *genus Passer*, a possible sequence of events may be hypothesised for the evolutionary change of nuclear mt cyt b DNA:

- 1) The nuclear reading code detects stop codons (either caused only by a mt-to-nucleus translocation or by independent true base changes).
- 2) The new mtDNA integrated into the nucleus escapes the control of normal nuclear genes repairing machinery.
- 3) Scattered deletions are established.

It is possible that the mitochondrial genes may already arrive to the nucleus with promoter anomalies, but the possibility that the priming sites (for PCR amplification) in *genus Passer* are different from other *Passerines* may be ruled out since our primers attach to the mt cyt b coding DNA (Arnaiz-Villena et al. 1998, 1999a).

A more general inference may be drawn from the present work: mt pseudogenes bird phylogenies are more robust with pseudogenes (also found by Arctander 1995) and the conjoint evolutive analysis of genes and pseudogenes may help to unveil the mechanisms by which mt pseudogenes come into the nucleus and to follow their evolution thereafter.

Finally, it is not clear why *genus Passer* accumulates a high number of nuclear mtDNA pseudogenes and other closely related bird species do not (Arnaiz-Villena et al. 1998, 1999a, 2001). The only general characteristic of *genus Passer* is that they have become human commensals. This may have rendered these finches more exposed to xenobiotics. It might be hypothesized that pollutants may have driven the mtDNA integration to the nucleus through processes that damaged mitochondria in the first place. Indeed, the amount of mtDNA in human genome is high (Fukuda et al. 1985).

Acknowledgments. We are indebted to the following Spanish ornithologists: Bernardino Yebes, Gloria Gardó, Francisco Mira Chinchilla, Arturo Fernández Cagiao, and Alvaro Guillén. This work was supported in part by Ministerio de Educación grants (PM95-57, PM96-21 and PM99-23) and Comunidad de Madrid (06/70/97 and 8.3/14/98). L. Allende and I. Rubio contributed equally to this paper.

References

- Arctander P (1995) Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. *Proc R Soc Lond B* 262:13–19
- Arnaiz-Villena A, Timón M, Corell A, Pérez-Aciego P, Martín-Villa JM, Regueiro JR (1992) Primary immunodeficiency caused by mutations in the gene encoding the CD3 subunit of the T-lymphocyte receptor. *N Engl J Med* 327:529–533
- Arnaiz-Villena A, Alvarez-Tejado M, Ruiz-del-Valle V, García-de-la-Torre C, Varela P, Recio MJ, Ferre S, Martínez-Laso J (1998) Phylogeny and rapid Northern and Southern Hemisphere speciation of Goldfinches during the Miocene and Pliocene Epochs. *Cell Mol Life Sci* 54:1031–1041
- Arnaiz-Villena A, Alvarez-Tejado M, Ruiz-del-Valle V, García-de-la-Torre C, Varela P, Recio MJ, Ferre S, Martínez-Laso J (1999a) Rapid radiation of Canaries (*Genus Serinus*). *Mol Biol Evol* 16: 2–11
- Arnaiz-Villena A, Lliakis P, González-Hevilla M, Longás J, Gómez-Casado E, Sfyridaki K, Trápaga J, Silvera-Redondo C, Matsouka C, Martínez-Laso J (1999b) The origin of Cretan populations as determined by characterization of HLA alleles. *Tissue Antigens* 53: 213–226
- Arnaiz-Villena A, Guillen J, Ruiz-del-Valle V, Lowy E, Zamora J, Varela P, Stefan D, Allende LM (2001) Phylogeography of crossbills, bullfinches, grosbeaks, and rosefinches. *Cell Mol Life Sci* 58:1159–1166
- Arnason U, Gullberg A (1994) Relationship of baleen whales established by cytochrome b gene sequence comparison. *Nature* 36:726–728
- Avise JC, Ball Jr RM (1991) Mitochondrial DNA and avian microevolution. *Acta Congr Int Ornithol* 20:514–524
- Avise JC, Walker D (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proc R Soc Lond B* 265:457–463
- Blanchard JL, Lynch M (2000) Organellar genes: Why do they end up in the nucleus? *Trends Genet* 16:315–320
- Chiappe LM (1995) The first 85 million years of avian evolution. *Nature* 37:349–355
- Clement P, Harris A, Davies Y (1993) *Finches and sparrows*. C Helm, London
- Collura RV, Stewart CB (1995) Insertions and duplications of mt DNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* 38:485–489
- Cox CB, Moore PD (1995) *Biogeography*. Blackwell Humanity Press, Cambridge
- Edwards SV, Arctander P, Wilson AC (1991) Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. *Proc R Soc Lond B* 243:99–107
- Edwards AWF (1995) Assessing molecular phylogenies. *Science* 26: 253
- Feduccia A (1995) Explosive evolution in tertiary birds and mammals. *Science* 26:637–638
- Fehrer J (1996) Conflicting character distribution within different data sets on cardueline finches: artifact or history? *Mol Biol Evol* 13: 7–20
- Felsenstein J (1985) Confidence limits of phylogenies: an approach using the bootstrap. *Evolution* 39:783–795
- Fukuda M, Wakasugi S, Tsuzuki T, Nomiya H, Shimada K (1985) Mitochondrial DNA-like sequences in the human nuclear genome. *J Mol Biol* 186:257–266
- Gill FB (1995) *Ornithology*. Freeman, New York
- Hackett SJ (1996) Molecular phylogenetics and biogeography of Tanager in the *genus Ramphocelus* (Aves). *Mol Phy Evol* 5:368–382
- Härlid A, Janke A, Arnason U (1997) The mtDNA sequence of the Ostrich and the divergence between Palaeognathous and Neognathous birds. *Mol Biol Evol* 14:754–761
- Hedges SB, Parker PH, Sibley CG, Kumar S (1996) Continental

- breakup and the ordinal diversification of birds and mammals. *Nature* 381:226–229
- Helm-Bychowski KM, Wilson AC (1986) Rates of nuclear DNA evolution in pheasant-like birds: evidence from restriction maps. *Proc Natl Acad Sci USA* 83:688–692
- Hillis DM, Huelsenbeck JP, Cunningham CW (1994) Application and accuracy of molecular phylogenies. *Science* 264:671–677
- Howell N (1989) Evolutionary conservation of protein regions in the protonmotive cytochrome b and their possible roles in redox catalysis. *J Mol Evol* 29:157–169
- Klicka J, Zink RM (1997) The importance of recent ice ages in speciation: a failed paradigm. *Science* 277:1666–1669
- Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc Natl Acad Sci USA* 86:6196–6200
- Kornegay JR, Kocher TD, Williams LA, Wilson AC (1993) Pathways of lysozyme evolution inferred from the sequences of cytochrome b in birds. *J Mol Evol* 37:367–379
- Krajewski C, King DG (1996) Molecular divergence and phylogeny: rates and patterns of cytochrome b evolution in cranes. *Mol Biol Evol* 13:21–30
- Lockhart PJ, Steel MA, Hendy MD, Penny D (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol Biol Evol* 11:605–612
- Marten JA, Johnson NK (1986) Genetic relationships of North American Cardueline finches. *The Condor* 88:409–420
- McEvedy C (ed) (1980) *The penguin atlas of African history*. Penguin Books Ltd, London, England, UK
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Paturi FR (ed) (1991) *Die Chronik der Erde*. Harenberg Kommunikation, Dortmund, Germany
- Seutin G, Klein NK, Ricklefs E, Bermingham E (1994) Historical biogeography of the bananaquit (*Coereba flaveola*) in the caribbean region: a mitochondrial DNA assessment. *Evolution* 48:1041–1061
- Sibley CG, Monroe Jr BL (1990) *Distribution and taxonomy of birds of the world*. Yale University Press, London
- Smith AG, Smith DG, Funnell BM (1994) *Atlas of Mesozoic and Cenozoic Coastlines*. Cambridge University Press, Cambridge
- Summers-Smith JD (ed) (1988) *The Sparrows*. T AA Poyser Ltd, Calton, Staffordshire, England
- Swofford DL (1996) *PAUP (phylogenetic analysis using parsimony), Version 4.0b2*. Illinois Natural History Survey, Champaign
- Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of molecular clock and linearized trees. *Mol Biol Evol* 12:823–833
- Vincek V, O’Huigin C, Satta Y, Takahata N, Boag PT, Grant PR, Grant BR, Klein J (1997) How large was the founding population of Darwin’s finches? *Proc R Soc Lond B* 264:111–118
- Wittzell H, Von Schantz T, Zoorob R, Auffray C (1994) Molecular characterization of three Mhc class II B haplotypes in the ring-necked pheasant. *Immunogenetics* 39:395–403
- Zoorob R, Behar G, Kroemer G, Auffray C (1990) Organization of a functional chicken class II B gene. *Immunogenetics* 31:179–187