Ornis Fennica 73:145–156. 1996

Evolutionary patterns and phylogeny of tits and chickadees (genus *Parus*) based on the sequence of the mitochondrial cytochrome *b* gene

Laura Kvist, Minna Ruokonen, Markku Orell & Jaakko Lumme

Kvist, L., Ruokonen, M., Orell, M. & Lumme, J., Department of Biology, University of Oulu, P. O. Box 333, FIN-90571 Oulu, Finland
E-mail address: laura.kvist@csc.fi

Received 27 September 1996, accepted 11 November 1996

The phylogenetic relationships of seven European and three American *Parus* (Aves: Paridae) species were examined by sequencing 439 nucleotides of the mitochondrial cytochrome *b* gene. There were 60 phylogenetically informative sites out of 109 variable sites. The transition/transversion ratios varied from 0.91 to 6.5 between the species pairs and the Kimura's two parameter distances from 0.02 to 0.13. Subgenus *Poecile* consists of two clades, one with *Parus cinctus, P. carolinensis* and *P. atricapillus* and the other with *P. palustris* and *P. montanus*. *P. ater* and *P. cristatus* are grouped together in a DNA-parsimony tree but not in the neighbor-joining tree. These two species share an unusual deletion of glycine 167. *P. major* and *P. caeruleus* are the closest species pair in this study. Our results confirm the recent proposition of placing them into one subgenus, instead of two. A new nuclear copy of the cytochrome *b* gene (sharing 63% homology to the mitochondrial cytochrome *b* gene) was found from *P. montanus* and *P. atricapillus*. This pseudogene is of very ancient origin.

1. Introduction

Harrap and Quinn (1996) have recently published a revision of the genus *Parus*, which is divided into 10 subgenera containing 55 species. Eck (1988) divided the genus into 12 subgroups including 65 species complexes. This difference is caused by the problems in defining subspecies and species. A classical example of superspecies is the highly polytypic *Parus major* complex, which is sometimes divided into four species (Rassenkreis of *major, minor, cinereus, bokharensis*) and sometimes into only two species, *P. major* and *P. bokharensis*.

The main interest in *Parus* comes from the fact that several species of the genus are ecologically well studied (Cramp & Perrins 1993). The box-nesting habit facilitates research. This leads to investigations on family relations, on breeding structure of the populations, to questions about migration, etc. Molecular methods will be very useful in answering these questions, and the molecular phylogeny makes a necessary framework for this kind of research. No wonder that several attempts have been made.

Braun and Robbins (1986) studied protein relationships of North American *P. atricapillus, P. carolinensis* and *P. gambeli*. Gill et al. (1989) used allozyme comparison to determine the relationships of nine North American and six Eurasian species. Eight North American and nine Eurasian *Parus* species were included in the DNA–DNA hybridization-based phylogeny by Sheldon

et al. (1992) and Slikas et al. (1996). Restriction fragment length polymorphism of the mitochondrial DNA was used to construct a phylogeny of six North American species and several of their geographical races (Mack et al. 1986, Gill et al. 1993). These phylogenies already cover most of the subgenera or species groups, but the relations between the subgenera are not yet completely clarified (Slikas et al. 1996).

The mitochondrial cytochrome *b* gene has been widely used for constructing phylogenetic trees of closely related species or subspecies (e.g. Blackbirds *Agelaius*, Lanyon 1994; Blue and Great tits *Parus caeruleus* and *P. major*, Taberlet et al. 1994; Warblers *Phylloscopus*, Helbig et al. 1995; Cranes, Krajewski & King 1996; Cardueline finches, Fehrer 1996). It is probably the best known gene in the mitochondria and its structure and function have been studied intensively (Esposti et al. 1993). Mitochondrial DNA is maternally inherited, it does not recombine and evolves quite rapidly. The cytochrome *b* gene, however, is under strong evolutionary constraints and certain parts of the gene are more conserved than others due to functional restrictions. First and second positions of the codons evolve slowly. The silent base substitutions at the third codon positions are abundant though the amino acid substitutions are not (Meyer 1994).

Because it is predicted that modern molecular methods will create a useful perspective for an active ecological research, we present here our contribution to the phylogenetic analysis of European *Parus*, using "the beginners choice of molecule" (Meyer 1994), the mitochondrial cytochrome *b* gene. The main advantage of using DNA sequence data is that they are connectible (Avise 1994). Certain parts of mtDNA are used to create a universal metric for phylogenetic studies (Meyer 1994). Until now, the cytochrome *b* gene sequence has not been used for phylogenetic purposes within Paridae, but three of the sequences used in this report were gathered from the GenBank (see also Taberlet et al. 1992). So, this study can be complemented easily in both directions. New species can be implemented, and longer segments of the mtDNA molecule can be analyzed, if the phylogenetic resolution is not adequate.

Nuclear pseudogenes sometimes interfere in the analysis of mtDNA, when PCR amplification is made from total DNA extract (Arctander 1995, Collura & Stewart 1995, Zischler et al. 1995, Zhang & Hewitt 1996). We demonstrate here that ancient pseudogenes exist in *Parus*, too.

2. Material and methods

2.1. Samples

DNA of the Coal Tit *Parus ater,* the Willow Tit *P. montanus,* the Marsh Tit *P. palustris,* the Crested Tit *P. cristatus*, the Black-capped Chickadee *P. atricapillus* and the Carolina Chickadee *P. carolinensis* was obtained from blood samples of living birds or from newly laid or slightly incubated eggs. The sample of the Siberian Tit (*P. cinctus*) originates from pectoral muscle tissue of a dead individual in the collections of the Zoological Museum at the University of Oulu. In addition, four cytochrome *b* sequences were fetched from GenBank (Plain Titmouse *P. inornatus,* Great Tit *P. major* and Blue Tit *P. caeruleus,* and an extreme outgroup, Greater Flamingo *Phoenicopterus ruber*). The list of collection localities, GenBank access numbers and other relevant information is given in Table 1.

2.2. DNA extraction

Blood samples (50–100 µl) and muscle tissue (cut into small pieces) were mixed with an equal volume of lysis buffer (0.01 M Tris, 0.01 M NaCl, 0.01 M Na_2EDTA, 1% N-Lauroylsarcosine, pH 8.0). To these mixtures, 800 µl of SET buffer (0.15 M NaCl, 1 mM Na_2EDTA 50 mM maleic acid, pH 8.0), 28 µl of 10% sodium dodecyl sulphate and 8 µl of proteinase K (20 mg/ml) were added. The mixtures were incubated overnight at 55°C. The samples were then extracted with phenol and chloroform to remove proteins. DNA was precipitated with ethanol, washed, dried and diluted into 200 µl of sterile water.

From the eggs (preincubation or early), the embryonic plate was picked with a pipette tip. The mitochondrial DNA was isolated according to the alkaline lysis procedure by Tamura and Aotsuka (1988), except that the first centrifugation step was omitted.

2.3. Polymerase chain reaction

PCR primers were designed based on conserved regions located by comparing *Parus inornatus*, *P. major* and *P. caeruleus* cytochrome *b* sequences available from GenBank. Only partial cytochrome *b* sequences were available from these species, therefore we designed primers to amplify a 600 bp-long fragment instead of the whole gene. The names and sequences of the primers were:

L15036: 5′GATACTTCCTAGGCATCTGCCT3′
and
H15636: 5′GGGTTTGCTGGGGTGAAGTTTT3′.

Double-stranded PCR amplifications were performed to a reaction volume of 100 µl, containing 5–10 µl of the total DNA or 2 µl of mtDNA, 1 µM primers (L15036 and L15636), 10 mM Tris-HCl, pH 8.8, 1.5 mM $MgCl_2$, 50 mM KCl, 0.1% Triton X-100, 200 µM each dNTP and 5U of Dynazyme (Finnzymes). The amplification profile was the following: denaturation at 94°C (1 min), annealing at 50°C (1 min 30 s) and synthesis at 72°C (2 min). This was repeated for 30 cycles.

All the PCR products were purified from 0.8% agarose gel to eliminate excess primers and dNTPs. The PCR product was then cut from the gel and then further cut into small pieces. To these pieces, 200 µl of TE buffer (1mM Na_2EDTA, 10 mM Tris-HCl, pH 7.8) and 200 µl of phenol were added. After brief vortexing, the samples were frozen in liquid nitrogen, then centrifuged for 15 min at 11 000 g and 400 µl of chloroform-isoamylalcohol (24:1) was added to the aqueous phase. The samples were then vortexed and centrifuged for 2 min again at 11 000 g. The DNA was precipitated from the aqueous phase using 40 µl of 3M sodium acetate and 800 µl ethanol at −20°C for one hour. After centrifugation (11 000 g, 25 min) the DNA pellets were washed with 70% ethanol, vacuum dried and diluted in 20 µl of sterile water.

From three species, namely *montanus*, *atricapillus* and *carolinensis*, the primers amplified two fragments, one of the correct size (600 bp) and one shorter, about 550 bp, which we cloned into the TA vector (Invitrogen).

2.4. Sequencing

Double-stranded DNA sequencing of the PCR products was performed according to Bernatchez et al. (1992) with the Sequenase kit (Version 2.0, US Biochemical) using L15036 (sequence above) or L15273 (ACCTGAAACATCGGAGT) primers. The cloned products were sequenced with T7 or sp6 universal primers by the standard dye terminator reaction procedure for ABI 377 DNA Sequencer.

2.5. Analysis of the DNA sequences

The methods used for inferring phylogenies from molecular data have been discussed intensively

Table 1. Summary of source localities of the species under study.

Species	Locality	Tissue type	GenBank accession number/Author
Parus ater 1	Liminka, Finland	Blood	U60773/This work
Parus ater 2	Liminka, Finland	Blood	U60774/This work
Parus atricapillus	Ohio, USA	Blood	U60770/This work
Parus caeruleus	Grenoble, France	Liver, Heart	M88168/Taberlet et al. 1992
Parus carolinensis	Ohio,USA	Blood	U60771/This work
Parus cinctus	Pulkkila, Finland	Muscle	U60768/This work
Parus cristatus	Oulu, Finland	Blood	U60772/This work
Parus inornatus	California, USA	Blood	X60944/Edwards et al. 1991
Parus major	Grenoble, France	Liver, Heart	M88590/Taberlet et al. 1992
Parus montanus 1	Oulu, Finland	Embryonal plate	U60775/This work
Parus montanus 2	Oulu, Finland	Blood	U60776/This work
Parus palustris 1	Tovetorp, Sweden	Blood	U60769/This work
Parus palustris 2	Oxford, UK	Blood	Sequence similar to the one above
Phoenicopterus ruber	–	Blood	U08940/Avise et al. 1994

since they were created. The three main classes of methods are based on maximum parsimony, maximum likelihood and distance analysis (reviewed in Felsenstein 1988). The consistency of these methods has been discussed by many authors, but no consensus has been reached so far (e.g. Saitou & Imanishi 1989, Tateno et al. 1994). Sampling the data by bootstrapping and related methods have been used to estimate the true tree topology with increasing reliability (Felsenstein 1985, Lapointe et al. 1994, Cummings et al. 1995). In this paper, we present the phylogenetic trees obtained by neighbor-joining (Saitou & Nei 1987) method. The differences to the trees obtained by maximum parsimony or maximum likelihood methods are mentioned.

The sequence alignment was done using the PILEUP program of the Wisconsin Sequence Analysis Package (Genetics Computer Group 1994). The phylogenetic analyses were carried out by the PHYLIP programs (Phylogeny Inference Package version 3.5c by Joseph Felsenstein, the University of Washington 1993). DNApars was used for maximum parsimony analysis, NEIGHBOR for neighbor-joining analysis. Distances were Kimura's two-parameter distances and the transition-transversion ratio was set to 2.0. FastDNAml (Olsen et al. 1994) was used for maximum likelihood analysis with the transition-transversion ratio set also to 2.0 and global option. For each analysis, the sequences were bootstrapped to get 100 or 1 000 replicates of the data. To test if the phylogenetic information varies in different parts of the sequence, the 439-bp sequence was cut into small overlapping fragments (50–200 bp long) and maximum parsimony analysis was performed using 100 bootstrap replicates from all these fragments separately.

3. Results

We analyzed a 439 bp fragment of the cytochrome *b* gene from all the species under study. The fragment includes the coding sequence from the end of the transmembrane helix A to the end of the helix D. There are four heme ligands in helices B and D, and altogether 18 conserved amino acids, which tend to remain invariant from yeast to human (Esposti et al. 1993).

In this region of the Paridae cytochrome *b* gene, we found 109 variable nucleotide sites (Fig. 1), from which 15 are at the first codon position, 6 at the second and 88 at the third. There are 60 phylogenetically informative sites, 7 at the first position, 1 at the second and 52 at the third. The number well exceeds the number of species studied (Stewart 1993). The assumptions for constructing a phylogenetic tree are therefore met.

The overall base composition is biased in this part of the cytochrome *b* gene. Guanines are rare (15.5%) and cytosines are frequent (34.1%). The first codon position is fairly evenly occupied by all four nucleotides. The second position lacks guanine and adenine (17.1% and 19.1%, respectively) and thymine is in excess. The largest bias is at the third codon position, which is occupied by cytosine (53.5%), and lacks guanine (2.4%) and thymine (8.3%) (Fig. 2).

The overall transition/transversion ratio estimated from pairwise comparisons of the species is 2.04. The ratio varies from 0.91 between *cristatus* and *atricapillus* to 6.5 between *inornatus* and *ater* 2 (Table 2). In the first codon position, the AG transitions are the most abundant, in the second position the AC transversions and the CT transitions, and in the third, the CT transitions. The amount of AC transversions over all the codon positions is quite high (20.7% from all the substitutions). The first codon position lacks the CG and GT transversions and the second lacks the GT transversions (Table 3). As a summary, when all the codon positions are pooled together, there is twofold excess of pyrimidine transitions over purine transitions and twofold excess of AC transversions over other types of transversions.

Amino acid substitutions are quite rare and altogether there are 18 variable sites (8%) in the region analyzed. Using the nomenclature of Esposti et al. (1993) one polymorphic amino acid site was located in the first intramembrane domain (A), two polymorphic sites at the first extramembrane domain (ab), five at the third transmembrane domain (C), five at the second extramembrane domain (cd) and five at the fourth transmembrane domain (D) (Fig. 3).

The greatest number of amino acid substitutions was found from *montanus* 2, which had six substitutions compared with the consensus sequence. The smallest number of amino acid substi-

```
                 1
                 5
                 0              1                    2
                 5566777788999001122234455567888900111123334445555556778
                 3439235815039281437924703621039217367986780672356890036 5

Paridae          TAACCACACACACCACCCCCCCACACCCCCCCCCACCAAACACCAACCCAGCATCA

P. cinctus       C.G.T.............A.....................A...........
P. palustris 1, 2 ....T.....T........AT..T.T..............T..T.....ACC.G
P. atricapillus  ...TT...........A.A...........T.........AG..T........
P. carolinensis  ...TT.T..........TA....................AG..T........
P. montanus 1    .G..T...T....T.....A........T.....T...GCG....CAT....C...
P. montanus 2    .G..T...T....T.....A........T.....T...GCG....CAT....C...
P. major         .G...T.G.TTT....T.............T.TC........T.....GA...A.
P. caeruleus     .G...T.G...T....T.............A.TT........T.C...GA...AC
P. cristatus     .G.....G............T.TC.....T.T....G...G.T............
P. ater 1        .....T....T.T..........G..T.........G......G...T....C..
P. ater 2        .....T....T.T..........G..T.........G......G...T....C..
P. inornatus     .....T......TT.T....T.T..TT..TT.T...T...T.........A..C..
```

```
                 1
                 5
                 2   3                        4
                 8999000001122334456677899999901112223334555556677777788 8
                 8147035691417360770358701236924780392487023672812345603 9

Paridae          CCACAAAAAACCACACCACCATCAGGTACAAACACCCCCCACCCGCCCACACTCCC

P. cinctus       ...A.............C...C.G..G..CT.............A.T.........
P. palustris 1, 2 ..G.........C....C.....................T.C...A........T..
P. atricapillus  ..G....G...AT....C...A.......T.............AT....G.....
P. carolinensis  ..............G..C...C.......T....T.........ATT.........
P. montanus 1    T.......G......TT..A..............T.A..T....TA.GG.T....
P. montanus 2    T.......G........C.................T....T....TA.GG.T....
P. major         .T........TT...T..........A...C.T...T.A.......A.......TT
P. caeruleus     ..........TT...T..........G...C.T...T.G.......A...G...TT
P. cristatus     .TGACGG.................---C.G.C......AAGA.T............
P. ater 1        ........G....TG...T.....---C..............A.A.......A.T.
P. ater 2        ........G....T....T.....---C..............A.A.........T.
P. inornatus     TT.A.....G.T.......TGCT...CCT.CT.G..T.........TAT.....T..
```

Fig. 1. The nucleotide substitutions in the partial cytochrome *b* gene of the Paridae. The numbers refer to the chicken mitochondrial nucleotide positions (Desjardins & Morais 1990) and the nucleotides below the numbers to the consensus of the variable sites.

tutions was found from *palustris*, which had only one substitution compared with the consensus sequence. An interesting feature is that though *caeruleus* and *inornatus* have altogether 48 nucleotide substitutions between them, their amino acid sequence is identical.

The two *montanus* individuals studied show that there exists a relatively high amount of intraspecific variation in this region, because within the 439 nucleotides there are five nucleotide substitutions and three amino acid substitutions. The sequence divergence is 1.1% between these two individuals, which were sampled from the same population. From these five nucleotide substitutions, three are transversions and two are transitions (Table 2). There are two nucleotide substi-

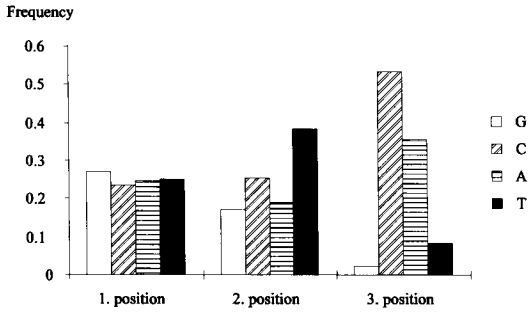

Fig. 2. The base composition of the Paridae cytochrome *b* gene.

tutions between the two *ater* individuals, one transition and one transversion leading to one amino acid substitution. The sequence divergence is 0.45%. Also these two birds were from one population. The two *palustris* individuals are identical, although one was from Sweden, and the other from England.

The phylogenetic analyses with different methods gave rather similar tree topologies. We first used *Emberiza rustica* (GenBank accession number D38326) as an outgroup. However, the tree topology was similar when we used *Parus inornatus* (subgenus Baelophus) as an outgroup. It obviously is the most distant species among those studied. The phylogenetic tree constructed by the neighbor-joining method is shown in Fig. 4.

Parus cinctus joins the American Poecile species *carolinensis* and *atricapillus*. This clade occurs 814 times out of 1 000 bootstrap repeats. *P. palustris* and *montanus* form a separate clade, but the bootstrap value is only 326/1 000. Subgenus *Poecile* is separated from others quite well (637/1 000). The topology of maximum parsimony and maximum likelihood trees was identical with neighbor-joining tree in *Poecile*.

P. ater and *cristatus* were grouped together in a DNA-parsimony tree (tree not shown, 456/1 000). In a neighbor-joining tree, based on Kimura's two parameter distance, and in a maximum likelihood tree, *cristatus* remains alone, when *ater* still has some affinity with the Poecile group. *P. cristatus* and *ater* do share an unusual deletion of glycine 167 (triplet GGT missing at position 15391-3, Fig. 1).

Table 2. The amount of transitions and transversions and the transition/transversion ratio (below the diagonal), and Kimura's two parameter distance (above the diagonal) in pairwise comparisons of the Paridae.

Species	mon1	mon2	atri	caro	palu	cinc	ater1	ater2	cris	inor	majo	caer
P. montanus		0.0115	0.0899	0.0850	0.0959	0.0874	0.1140	0.1086	0.1273	0.1311	0.1167	0.1112
		0.0000	0.0002	0.0002	0.0003	0.0002	0.0003	0.0003	0.0003	0.0004	0.0003	0.0003
P. montanus 2		2/3	0.0771	0.0723	0.0830	0.0746	0.1063	0.1010	0.1191	0.1256	0.1141	0.1086
		0.67	0.0002	0.0002	0.0002	0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
P. atricapillus		20/17	18/14	0.0328	0.0696	0.0570	0.0950	0.0898	0.1081	0.1277	0.1136	0.1055
		1.18	1.29	0.0001	0.0002	0.0001	0.0002	0.0002	0.0003	0.0003	0.0003	0.0003
P. carolinensis		22/13	20/10	10/4	0.0699	0.0353	0.0751	0.0752	0.1012	0.1123	0.0983	0.0981
		1.69	2.00	2.50	0.0002	0.0001	0.0002	0.0002	0.0003	0.0003	0.0003	0.0003
P. palustris		28/11	26/8	17/12	21/8	0.0620	0.0851	0.0800	0.1003	0.1172	0.1059	0.1059
		2.55	3.25	1.42	2.63	0.0002	0.0002	0.0002	0.0003	0.0003	0.0003	0.0003
P. cinctus		20/16	18/13	13/11	12/3	15/11	0.0775	0.0724	0.0878	0.1042	0.0878	0.0849
		1.25	1.38	1.18	4.00	1.36	0.0002	0.0002	0.0002	0.0003	0.0002	0.0002
P. ater 1		28/18	29/14	20/19	23/8	23/12	22/10	0.0046	0.0907	0.1188	0.0988	0.1039
		1.56	2.07	1.05	2.88	1.92	2.20	0.0000	0.0002	0.0003	0.0003	0.0003
P. ater 2		27/17	28/13	19/18	24/7	22/11	21/9	1/1	0.0855	0.1134	0.0935	0.0986
		1.59	2.15	1.05	3.43	2.00	2.33	1.00	0.0002	0.0003	0.0002	0.0003
P. cristatus		29/22	27/21	21/23	29/12	22/19	24/12	27/10	26/9	0.1269	0.1036	0.1115
		1.32	1.28	0.91	2.42	1.16	2.00	2.70	2.89	0.0004	0.0003	0.0003
P. inornatus		36/16	35/15	32/19	34/11	32/15	32/10	40/7	39/6	41/9	0.1153	0.1205
		2.25	2.33	1.68	3.09	2.13	3.20	5.71	6.50	4.56	0.0003	0.0003
P. major		29/18	29/17	25/21	27/13	26/17	24/12	30/10	29/9	28/14	36/10	0.0232
		1.61	1.71	1.19	2.08	1.53	2.00	3.00	3.22	2.00	3.60	0.0001
P. caeruleus		27/18	27/17	22/21	25/15	26/17	21/14	30/12	29/11	29/16	36/12	6/4
		1.50	1.59	1.05	1.67	1.53	1.50	2.50	2.64	1.81	3.00	1.50

Location:	5 2	5 9	6 3	1 5	1 6	1 9	1 2	1 3	1 6	1 8	1 2	1 7	1 7	1 8	1 7	1 9	1 4	1 5
Amino acid:	M	T	T	N	I	I	L	A	W	Q	S	I	G	A	P	V	T	L
Paridae																		
P. cristatus					V				C	R			-		H			
P. ater		S				V							-			I		Q
P. atricapillus						V										I		
P. carolinensis						V										I		
P. cinctus	T															I		
P. palustris																I		
P. montanus 1				T			I										G	
P. montanus 2				T			I				L	M		D			G	
P. caeruleus		S						T										
P. major		S	S					T										
P. inornatus		S						T										

Fig. 3. Amino acid differences in the partial cytochrome *b* gene in Paridae. The location in the mitochondrial genome refers to the amino acid numbers in Esposti et al. (1993) and the amino acids below to the consensus of the species studied.

Parus major and *caeruleus* are the closest relatives (0.0232) among the species studied. They group together with all the methods used and in each bootstrap replicate.

3.1. Nuclear pseudogene

In addition to the cytochrome *b* gene, the PCR-primers amplified a shorter fragment, approximately 550 bp, from *montanus*, *carolinensis* and *atricapillus*. This short fragment was detected only when the total DNA was used as a template in the PCR. When purified mitochondrial DNA was used, no extra fragments were seen. Altogether we screened 35 willow tit samples, of which 22 were total DNA extractions from blood (of these, 19 had double bands) and 13 were mitochondrial DNA extractions (only one band detected). In addition we had the total DNA from muscle tissue from five individuals, from which mtDNA also was isolated and screened, and these templates produced only one band in PCR. Because only one band was repeatedly amplified from mtDNA extractions, the extra bands are due to nuclear contamination. The sequences of these short fragments were compared with the cytochrome *b* gene sequences. The homologies between *atricapillus* and *montanus* cytochrome *b* sequences and the sequence of the short fragments were approximately 63% and the homology between the short fragments of these species was

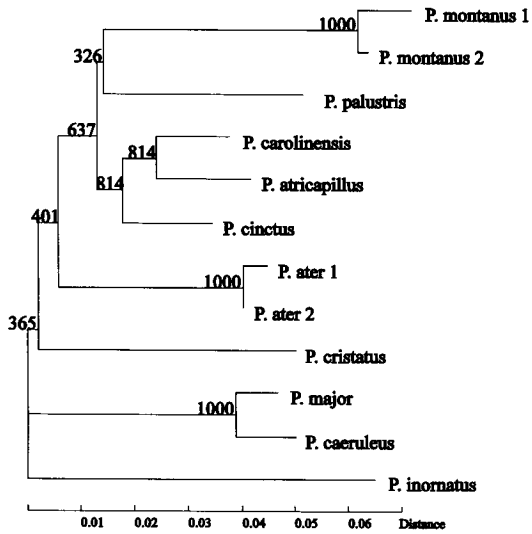

Fig. 4. The neighbor-joining tree from the cytochrome *b* sequences of Paridae. The numbers are the bootstrap replicates that support the branching.

Table 3. The frequency of nucleotide substitution types in the 439 bp-long fragment of the cytochrome *b* gene of the Paridae.

Substitution type	Positions 1st	2nd	3rd	All
AG	0.412	0.180	0.176	0.206
CT	0.370	0.311	0.486	0.465
AC	0.052	0.328	0.225	0.207
AT	0.166	0.033	0.089	0.096
CG	–	0.148	0.016	0.019
GT	–	–	0.008	0.007

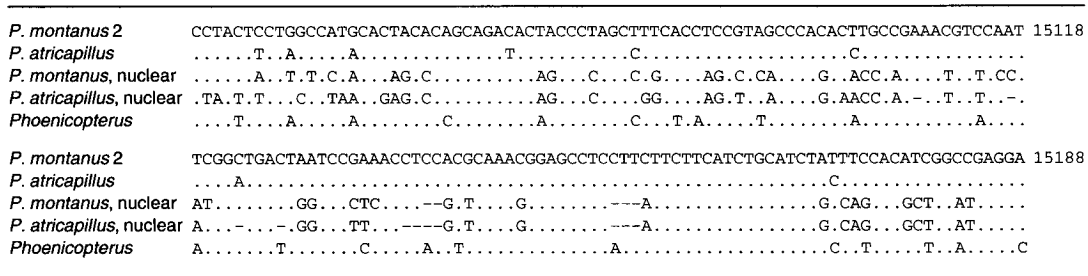

Fig. 5. Alignment of the nuclear pseudogenes with the respective mitochondrial cytochrome *b* sequences of *Parus montanus* and *P. atricapillus*, with Greater Flamingo, *Phoenicopterus ruber*, as an outgroup.

94% (Fig. 5). The short fragment of *P. carolinensis* had no resemblance to the cytochrome *b* gene. These fragments had plenty of stop codons in each three reading frames. We therefore conclude that these short fragments from *atricapillus* and *montanus* are nuclear pseudogenes.

Blast search from GenBank with the pseudogene of *atricapillus* gave the highest scores to the Greater Flamingo *Phoenicopterus ruber*, the Crowned Crane *Balearica pavonina*, the Japanese Great Bunting *Emberiza variabilis* and the Whiskered Tree Swift *Hemiprocne mystacea*. We aligned the sequence from the *Phoenicopterus* with the pseudogene and cytochrome *b* sequences and found 19 sites where one or both pseudogenes and *Phoenicopterus* cytochrome *b* sequences shared the same nucleotides but differed from the *Parus* cytochrome *b* sequences (Fig. 5). A neighbor-joining tree of these five sequences is presented in Fig. 6.

4. Discussion

Nucleotide distribution at different codon positions within the genus *Parus* is very similar to the nucleotide distribution of the entire cytochrome *b* gene in other birds and mammals (Irwin et al. 1991, Kornagay et al. 1993). This indicates that the part of the cytochrome *b* gene studied is a good representative of the whole gene. Still, when compared with cardueline and crane distribution, the distribution within tits resembles that of carduelines (Fehrer 1996) more than cranes (Krajewski & King 1996). The four nucleotides are evenly distributed in the first codon position within carduelines and titmice but in cranes there are cytosines in excess. In the second position, the lack of guanines is stronger in cranes than in carduelines or tits.

The proportion of transitions in the cranes is much higher than the frequency of tit cytochrome

b (0.84 and 0.67, respectively). This is caused mainly by the higher rate of CT transitions in the crane sequences. The transition/transversion ratio varies from 0.91 to 6.5 between species pairs among the Paridae. In carduelines the rate varies between 2.0 and 23.0 and in blackbirds between 1.1 and 8.9 (Lanyon 1994). The small transition/transversion rate in the Paridae indicates that some species pairs are close to saturation and the divergence of these species has occurred a long time ago.

The number of amino acid changes are similar to the pattern in carduelines. The 18 amino acids that are supposed to remain invariant in this region (Esposti et al. 1993) were the expected ones in all the species examined here. An exciting feature is that there is one codon for glycine (G167) missing in *cristatus* and *ater*. Deletions of one amino acid seem to be rare in cytochrome *b* and to our knowledge no such deletions have been reported from birds, though such deletions do exist in other animals (e.g. the Tassel-Eared Squirrel, *Sciurus aberti*; Wettenstein et al. 1995). In a comprehensive list of animals in Esposti et al. (1993), G167 is present in the human down to the roundworm, with no exception. The same deletion was found also in the pseudogene from *montanus* and *atricapillus*, but there are also several other deletions in them. In further studies the phylogenetic value of this deletion might be of great interest.

The nuclear pseudogene detected in this work in only two species, most probably may be present in other tits, too. Because we used PCR to detect it, small differences in the 3′ end of the primer can prevent amplification. Apparently the transfer from the mitochondria to the nucleus has occurred very early in the evolution of birds (Fig. 6). There exist many such similarities to distantly related species like *Phoenicopterus,* which have disappeared from the Paridae and the pseudogene has many deletions that cause frameshifts etc. The pseudogene found by Arctander (1995) from South American tapaculos (*Scythalopus*) is much more recent. Further studies are needed to evaluate the significance of this marker as a phylogenetic signal. At least, it is different enough not to badly disturb the use of genuine cytochrome *b* in the phylogenetic analysis of *Parus*. Interactions of the nuclear pseudogenes to the mitochondrial markers are discussed in Zhang and Hewitt (1996).

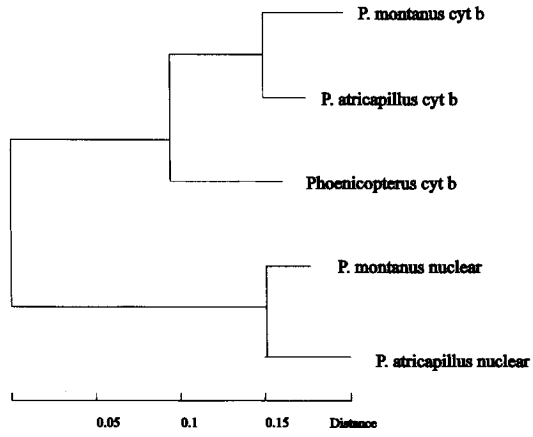

Fig. 6. The neighbor-joining tree from the cytochrome *b* gene and the nuclear pseudogene of *Parus montanus* and *P. atricapillus*. The Greater Flamingo, *Phoenicopterus ruber,* is included for dimensioning.

In this study, we analysed two specimens of *montanus, ater,* and *palustris*. The *montanus* sequences diverged by 1.1% (Kimura's two parameter distance 0.0115), even though the individuals were from the same population in Oulu, Finland. This high nucleotide variation has also been seen in a comprehensive analysis of the mtDNA control region within two *montanus* populations (Kvist et al. in preparation). Taberlet et al. (1992) reported a 1.0% sequence divergence in the same part of the cytochrome *b* gene between two lineages of *P. caeruleus*. The estimated amount of variation from mtDNA-restriction fragment length polymorphism is 1.23% between individuals of two lineages, and 0.12% between individuals of the same lineage (Taberlet et al. 1992). By RFLP analysis of mtDNA, Tegelström (1987) estimated the mean sequence divergence of 0.19% between individuals in three Swedish *P. major* populations. The average intraspecific diversity between haplotypes of North American *Parus* species varied from 0.46% to 0.71% (Gill et al. 1993). The coal tit (*P. ater*) sequence divergence of 0.45% in this study also falls within this range. Both *ater* specimens were sampled in the same location in Liminka, Finland. As a contrast, the two *P. palustris* specimens are identical in this region of the cytochrome *b* gene even though the samples were collected quite a distance from each other (Sweden and Great Britain).

The amount of intraspecific variation may differ between species of the same genera depending on the history of the species. In stable populations that have a large effective population size, the intraspecific variation is high, which seems to be the case with *montanus*. Because the mitochondrial DNA is maternally inherited, the amount of variation implies, in addition, a high female migration. The lack of sequence divergence in two *palustris* sampled from distant localities, indicates a reduction in the effective population size in the recent history of this species. Naturally, when only two individuals are compared, strong conclusions cannot be drawn. The intraspecific variation in Paridae certainly deserves further studies.

The Kimura's two parameter distances between the *Parus* species varies from 0.02 between *caeruleus* and *major,* to 0.13 between *inornatus* and *montanus* 1. The estimates of unweighted p (Nei 1987) obtained by RFLP of mtDNA from North American *Parus* species vary from 0.03 to 0.09 (Mack et al. 1986, Gill et al. 1993). The distance between *major* and *caeruleus* falls into the same range with the distance between eastern and western populations of *P. carolinensis* (Gill et al. 1993). Still these species have been placed into different sub-groups (Eck 1988, Harrap & Quinn 1996). Our data suggest the classification of *major* and *caeruleus* into one subgenus. This is supported also by the results of Slikas et al. (1996, p. 78), obtained from DNA–DNA hybridization.

The position of *P. cinctus* in the phylogeny has not been studied before with methods of molecular biology. It is placed in the Poecile subgroup and close to Boreal Chickadee *P. hudsonicus* and Chestnut-backed Chickadee *P. rufescens* by morphological characters (Snow 1954, Voous 1977). Unfortunately, we had no samples of these two species at our disposal, but *cinctus* groups rather close to the American *atricapillus* and *carolinensis*. In this clade, there is space for more species.

It is interesting that *montanus* and *atricapillus* are not very close to each other. These two species have earlier been considered as conspecific and both were under the name *P. atricapillus* (Voous 1977). The enzyme phylogeny of Gill et al. (1989, p. 187) is in accordance with our results.

As reviewed in Tegelström and Gelter (1990), the mean mtDNA divergence between bird genera was 0.089 and within genera 0.052. The observed pairwise divergence between European *Parus* species does not differ much from those values, although they are from the upper end of the scale. Estimates obtained by DNA sequencing are in general slightly higher than those based on restriction site variation (Avise 1994).

Based on the estimated mtDNA divergence rate of 2% per one million years in birds (e.g. Shields & Wilson 1987), the amount of divergence measured here suggests that the speciation of existing *Parus* species has begun some three to five million years ago, during the Pliocene epoch. DNA–DNA hybridization results of Slikas et al. (1996) give slightly higher estimates, 4–7 Mya.

The nature of variation is the probable reason for low bootstrap replicate values. Some of the phylogenetically informative sites draw the species in question to another direction than some other sites. This can reflect the real history of the species, or rather the history of the cytochrome *b* gene. A phylogenetic tree constructed from one gene does not necessarily agree with the species tree (Pamilo & Nei 1988). According to Avise (1989), failure to obtain a correct tree might be caused by: 1) sampling errors attributable to the small number of nucleotides examined, 2) heterogeneity of the evolutionary rate across the lineages, 3) genetic polymorphism in the ancestral species or 4) introgressive hybridization. All these phenomena, except the first one, reflect the real evolution of the species. The case with Paridae seems to be quite similar to the cases with some other bird groups, e.g. carduelines (Fehrer 1996), having globally very large and stable populations (like redpoll finches *Carduelis flammea-hornemanni*; Seutin et al. 1995). On the other hand, in large species with a restricted population size, a single gene phylogeny may fit well to the species tree (cranes of Krajewski & King 1996). The observed high genetic variation within present-day tit populations tempts us to explain some controversial amino acid changes (Table 3) or nucleotide substitutions (Table 1) on the basis of polymorphic ancestors.

Acknowledgments. We are grateful to Jeremy Blakey, Cindy Bronson, Jouko Inkeröinen, Juha Ollila, Vladimir Pravosudov, Anna Thessing and Risto Tornberg for the sample collection. We also thank Miia Makkonen and Hannele Parkkinen for their skillful technical assistance. This work was supported by the Academy of Finland, University of Oulu and Emil Aaltonen Foundation.

Selostus: **Tiaisten sytokromi-*b* geenin evoluutio ja tiaislajien sukulaisuussuhteet**

Seitsemän eurooppalaisen ja kolmen amerikkalaisen tiaislajin sukulaisuussuhteita tutkittiin sekvensoimalla mitokondriaalista sytokromi *b*-geeniä 439 nukleotidia. *Poecile*-alasuku jakautui kahteen ryhmään, joista toiseen kuului lapintiainen (*Parus cinctus*), Amerikan hömötiainen (*P. atricapillus*) ja tammitiainen (*P. carolinensis*) ja toiseen viitatiainen (*P. palustris*) ja hömötiainen (*P. montanus*). Talitiainen (*P. major*) ja sinitiainen (*P. caeruleus*) olivat kaikkein läheisin lajipari. Tämä tukee ajatusta, jonka mukaan nämä lajit pitäisi sijoittaa samaan alasukuun. Tällä hetkellä ne kuuluvat kahteen eri alasukuun (*Parus* ja *Cyanistes*). Kuusitiainen (*P. ater*) ja töyhtötiainen (*P. cristatus*) sijoittuivat näiden kahden ryhmän väliin joko yhteiseen haaraan tai omiin haaroihinsa menetelmästä riippuen. Hömötiaiselta ja Amerikan hömötiaiselta löydettiin nukleaarinen kopio sytokromi *b*-geenistä. Tämän pseudogeenin todettiin siirtyneen tumaan jo ennen kuin lintuheimot ovat eronneet toisistaan.

References

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

Avise, J. C. 1989: Gene trees and organismal histories: A phylogenetic approach to population biology. — Evolution 43: 1192–1208.

Avise, J. C. 1994: Molecular markers, natural history and evolution. — Chapman & Hall, New York. 511 pp.

Avise, J. C., Nelson, W. S. & Sibley, C. G. 1994: DNA sequence support for a close phylogenetic relationship between some storks and New World vultures. — Proc. Natl Acad. Sci. U.S.A. 91: 5173–5177.

Bernatchez, L., Guyomard, R. & Bonhomme, F. 1992: DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout Salmo trutta populations. — Mol. Ecol. 1: 161–173.

Brown, D. G. B. & Robbins, M. B. 1986: Extensive protein similarity of the hydridizing chickadees Parus atricapillus and Parus carolinensis. — The Auk 103: 667–675.

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.

Cramp, S. & Perrins, C. M. 1993: The birds of the Western Palaearctic. Vol. VII Flycatchers to Shrikes: 145–281. — Oxford University Press, London.

Cummings, M. P., Otto, S. P. & Wakeley, J. 1995: Sampling properties of DNA sequence data in phylogenetic analysis. — Mol. Biol. Evol. 12: 814–822.

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: 599–634.

Eck, S. 1988: Gesichtspunkte zur Art-Systematik der Meisen (Paridae). — Zool. Abh. Mus. Tierkd. Dresden 43: 201–134.

Edwards, S. V., Arctander, P. & Wilson, A. C. 1991: Mitochondrial resolution of a deep branch in the genealogical tree for perching birds. — Proc. R. Soc. Lond. B 243: 99–107.

Esposti, M. D., de Vries, S., Crimi, M., Ghelli, A., Patarnello, T. & Meyer, A. 1993: Mitochondrial cytochrome b: evolution and structure of the protein. — Biochim. Biophys. Acta 1143: 243–271.

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 on phylogenies: An approach using the bootstrap. — Evolution 39: 783–791.

Felsenstein, J. 1988: Phylogenies from molecular sequences: Inference and reliability. — Ann. Rev. Genet. 22: 521–565.

Gill, F. B., Funk, D. H. & Silverin, B. 1989: Protein relationships among titmice. — Wilson Bull. 101: 182–197.

Gill, F. B., Mostrom, A. M. & Mack, A. L. 1993: Speciation in North American chickadees: I. Patterns of mtDNA genetic divergence. — Evolution 47: 195–212.

Harrap, S. & Quinn, D. 1996: Tits, Nuthatches & Treecreepers. — Chistopher Helm Publishers Ltd, London.

Helbig, A. J., Seibold, I., Martens, J. & Wink, M. 1995: Genetic differentiation and phylogenetic relationships of Bonelli's Warbler Phylloscopus bonelli and Green Warbler P. nitidus. — J. Avian Biol. 26: 139–153.

Irwin, D. M., Kocher, T. D. & Wilson, A. C. 1991: Evolution of the cytochrome b gene of mammals. — J. Mol. Evol. 32: 128–144.

Kornegay, J. R., Kocher, T. D., Williams, L. A. & Wilson, A. C. 1993: Pathways of lysozyme evolution inferred from the sequences of cytochrome b in birds. — J. Mol. Evol. 37: 367–379.

Krajewski, C. & King, D. G. 1996: Molecular divergence and phylogeny: Rates and patterns of cytochrome b evolution in cranes. — Mol. Biol. Evol. 13: 21–30.

Lanyon, S. M. 1994: Polyphyly of the blackbird genus Agelaius and the importance of assumptions of monophyly in comparative studies. — Evolution 48: 679–693.

Lapointe, F.-J., Kirch, J. A. & Bleiweiss, R. 1994: Jackknifing of weighted trees: Validation of phylogenies reconstructed from distance matrices. — Mol. Phyl. Evol. 3: 256–267.

Mack, A. L., Gill, F. B., Colburn, R. & Spolsky, C. 1986: Mitochondrial DNA: A source of genetic markers for studies of similar passerine bird species. — The Auk 103: 676–681.

Meyer, A. 1994: Shortcomings of the cytochrome b gene as a molecular marker. — TREE 9: 278–280.

Nei, M. 1987: Molecular evolutionary genetics. — Columbia University Press, New York. 255 pp.

Olsen, G. I., Matsuda, H., Hagström, R. & Overbeek, R. 1994: FastDNAml: A tool for construction of phylogenetic trees of DNA-sequences using maximum likelihood. — Comput. Applic. Biosci. 10: 41–48.

Pamilo, P. & Nei, M. 1988: Relationships between gene trees and species trees. — Mol. Biol. Evol. 5: 568–583.

Saitou, N. & Imanishi, T. 1989: Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum evolution and neighbour-joining methods of phylogenetic tree construction in obtaining the correct tree. — Mol. Biol. Evol. 6: 514–525.

Saitou, N. & Nei, M. 1987: The neighbor-joining method: A New method for reconstructing phylogenetic trees. — Mol. Biol. Evol. 4: 406–425.

Seutin, G., Ratcliffe, L. M. & Boag, P. T. 1995: Mitochondrial DNA homogeneity in the phenotypically diverse redpoll finch complex (Aves: Carduelinae: Carduelis flammea-hornemanni). — Evolution 49: 962–973.

Sheldon, F. H., Slikas, B., Kinnarney, M., Gill, F. B., Zhao, E. & Silverin, B. 1992: DNA-DNA hybridisation evidence of phylogenetic relationships among major lineages of Parus. — The Auk 109: 173–185.

Shields, G. F. & Wilson, A. 1987: Subspecies of Canada Goose (Branta canadensis) have distinct mitochondrial DNA's. — Evolution 41: 662–666.

Slikas, B., Sheldon, H. & Gill, F. G. 1996: Phylogeny of titmice (Paridae): I. Estimate of relationships among subgenera based on DNA-DNA hybridization. — J. Avian Biol. 27: 70–82.

Snow, D. W. 1954: Trends in geographical variation in palaearctic members of the genus Parus. — Evolution 8: 19–28.

Stewart, C.-B. 1993: The powers and pitfalls of parsimony. — Nature 361: 603–607.

Taberlet, P., Meyer, A. & Bouvet, J. 1992: Unusual mitochondrial DNA polymorphism in two local populations of blue tit Parus caeruleus. — Mol. Ecol. 1: 27–36.

Taberlet, P., Fumagalli, L. & Hausser, J. 1994: Chromosomal versus mitochondrial DNA evolution: Tracking the evolutionary history of the southwestern European populations of the Sorex araneus group (Mammalia, Insectivora). — Evolution 48: 623–636.

Tamura, K. & Aotsuka, T. 1988: Rapid isolation method of animal mitochondrial DNA by the alkaline lysis procedure. — Biochem. Genet. 26: 815–819.

Tateno, Y., Takezaki, N. & Nei, M. 1994: Relative efficiences of the maximum-likelihood, neighbor-joining and maximum parsimony methods whe substitution rate varies with site. — Mol. Biol. Evol. 11: 261–277.

Tegelström, H. 1987: Genetic variability in mitochondrial DNA in a regional population of the Great Tit (Parus major). — Biochem. Genet. 25: 95–110.

Tegelström, H. & Gelter, H. P. 1990: Haldane's rule and sex biased gene flow between two hybridizing flycatcher species (Ficedula albicollis and F. hypoleuca, Aves: Muscicapidae). — Evolution 44: 2012–2021.

Voous, K. H. 1977: List of recent holarctic bird species. Passerines. — Ibis 119: 376–406.

Wettenstein, P. J., Strausbauch, M., Lamb, T., States, J., Chakraborty, R., Jin, L. & Riblet, R. 1995: Phylogeny of six Sciurus aberti subspecies based on nucleotide sequences of cytochrome b. — Mol. Phyl. Evol. 4: 150–162.

Zhang, D.-X. & Hewitt, G. M. 1996: Nuclear integrations: challenges for mitochondrial DNA markers. — TREE 11: 247–251.

Zischler, H., Geisert, H., von Haeseler, A. & Pääbo, S. 1995: A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. — Nature 378: 489–492.