LAURA KVIST

PHYLOGENY AND PHYLOGEOGRAPHY OF EUROPEAN PARIDS

Academic Dissertation to be presented with the assent of the Faculty of Science, University of Oulu, for public discussion in Kuusamonsali (Auditorium YB 210), on March 3rd, 2000, at 12 noon.
Mitochondrial DNA sequences were used to study the phylogeny, population structure and colonisation history of *Parus* species. The phylogenetic relationships of seven European and three American species were examined by sequencing a part of the cytochrome *b* gene. Phylogenetically the closest species were the great tit (*Parus major*) and the blue tit (*P. caeruleus*). Subgenus *Poecile* was divided into two clades, one consisting of the Siberian tit (*P. cinctus*), the Carolina chickadee (*P. carolinensis*) and the Black-capped chickadee (*P. atricapillus*) and the other consisting of the marsh tit (*P. palustris*) and the willow tit (*P. montanus*). The coal tit (*P. ater*) and the crested tit (*P. cristatus*) did not group with any of the species studied. The population structure and the colonisation history of the willow tit, the great tit and the blue tit were examined by using the control region sequences. Two sampled populations of the willow tit were intermingled in the minimum spanning network. All haplotypes were unique and the nucleotide diversity was relatively high. The results suggest that the historical effective population size in the willow tit has been large and not contracted by the last ice age. Current gene flow must also be extensive. In contrast, one major haplotype was found in almost every great tit population representing approximately one third of all the haplotypes. The nucleotide diversity was clearly lower than in the willow tits. No population structuring was evident and the populations showed distinctive signs of a recent population expansion. The patterns of genetic variation probably reflect a population bottleneck during the ice age, and a recolonisation of the European continent thereafter, presumably from a refugium situated in the Balkans. Two maternal lineages were found in the blue tit. The southern lineage was restricted to the Iberian peninsula whereas the northern lineage was detected from all the populations. The minimum spanning network and genetic variability in the northern lineage was similar to that of the great tit, suggesting a similar colonisation history. The southern lineage, however, may have survived the ice age in a different refugium in the Iberian peninsula and was not as successful as the northern lineage in colonising available regions when the ice retreated. Both, the blue tit and the great tit have continued to expand their distribution northwards during this century and gene flow plays an important role in homogenising the populations.

**Keywords**: mitochondrial DNA, *Parus*, colonisation history, population structure
Acknowledgements

This research was carried out at the Department of Biology, Division of Genetics, in the University of Oulu. I am grateful to my supervisors Professor Markku Orell and Dr. Jaakko Lumme for the guidance and support during the work. I thank Professor Seppo Lakovaara, the former Head of the Department for providing excellent working facilities. I am indebted to the former and present members of the Bird Genetics Group, Leena Uimaniemi, Tuija Liukkonen-Anttila, Päivi Kuuva, Keijo Viiri and especially Minna Ruokonen for all the inspiring discussions we had. These discussions helped me to learn a great deal more than I otherwise would ever have been able to. I also thank the ‘Bird Boys’ Dr. Seppo Rytkönen, Dr. Kari Koivula, Dr. Kimmo Lahti, Dr. Petteri Welling and Jouko Inkeröinen for providing me with samples and ecological information on the species studied. I thank Hannele Parkkinen and Miia Makkonen for their technical assistance in parts of this work. I would like to extend my thanks to everybody in the Department for making the working days pleasant.

I am grateful to Professor Trevor Price and Professor Robert Zink for the helpful comments on my thesis and to Professor Robert Podolsky for revising the language.

This work was supported by the Academy of Finland, University of Oulu, Thule Institute at the University of Oulu and Eemil Aaltonen Foundation.

The love and support and the help especially on computers and biocomputing given by my husband Ari-Pekka has been essential all the way through. Our two daughters, Reetta and Paula, were the ones that reminded me what life really is about. Last but not the least I thank my parents Marja-Liisa Tschek and Esa Mikonsaari.
List of original publications

This thesis is based on the following publications, which are referred to in the text by their Roman numerals.


Abstract
Acknowledgments
List of original publications
1. Introduction: Evolution and mitochondrial DNA in birds........................................11
   1.1. Evolutionary forces and processes .................................................................11
   1.2. Molecular markers and neutrality .................................................................11
       1.2.1. Molecular clock .........................................................................................12
       1.2.2. Genetic structure of populations, phylogeography and gene flow ............12
       1.2.3. Effective population size ...........................................................................13
       1.2.4. Population growth or decline .....................................................................14
       1.2.5. Phylogeny ...................................................................................................14
   1.3. Mitochondrial DNA .........................................................................................15
       1.3.1. Avian mitochondrial DNA .........................................................................17
       1.3.2. Cytochrome b ...........................................................................................18
       1.3.3. Control region of mtDNA ..........................................................................19
2. Outlines of the present study .................................................................................22
3. Materials and methods ..........................................................................................23
   3.1. The bird species and populations studied .........................................................23
   3.2. Molecular methods ..........................................................................................25
   3.3. DNA Sequence analysis ...................................................................................25
       3.3.1. Genetic distances and phylogenetic methods .............................................25
       3.3.2. Population diversity indices and population divergence ............................26
4. Results ....................................................................................................................29
   4.1. Structure of the control region of the tits ................................................................29
   4.2. Taxonomy and diversity at the species level .......................................................30
   4.3. Diversity indices and coalescence times within populations ............................31
   4.4. Population structure and gene flow ....................................................................32
5. Discussion ...............................................................................................................33
   5.1. Sequence variation in the cytochrome b and in the control region and their suitability to phylogenetic and population genetic studies ..........................33
5.2. Phylogeography and spatial population structure..............................................36
5.3. Effective population size..................................................................................37
5.4. Post-glacial history............................................................................................38
5.5. Present or past gene flow ................................................................................40
6. Concluding remarks and some future aspects .....................................................42
7. References............................................................................................................44
Original papers
1. Introduction: Evolution and mitochondrial DNA in birds

1.1. Evolutionary forces and processes

Darwin (1859) was first to observe that evolution is ‘descent with modification’, though the mechanisms of both the inheritance (descent) and nature of variation (modification) were unknown to him at that time.

Genetic variation is introduced into a population by either mutation or immigration. Once variation has been introduced, the total array of genotypes may be considerably increased by recombination (Lewontin 1974). Existing variation in the population can be affected by natural selection or random fluctuations of allele frequencies. Selection can promote constancy (normalising selection), direct continuous change (directional selection) or diversification (diversifying selection; Dobzhansky et al. 1977).

Variation in finite populations may be reduced by random loss of alleles. The evolutionary change can be more rapid if a large population breaks up into isolated local strains. The rate of evolutionary change depends on the effective size of the local populations, the amount of interchange of individuals between populations, and is limited by mutation rates (Wright 1930).

1.2. Molecular markers and neutrality

The advent of molecular techniques in the mid 1960’s enabled evolutionary researchers to first detect genetic variation in proteins (Margoliash 1963, Kimura 1969) and later directly in DNA. Molecular variation could then be measured quantitatively and empirical results interpreted using population genetics theory (Fisher 1930, Wright 1931, Haldane 1932).

Many studies provided evidence that molecular levels of variation are high in most natural populations (Lewontin & Hubby 1966, Harris 1966, Kimura & Ohta 1971a). The neutral theory proposed by Kimura (1968a, b) asserts that a ‘great majority of evolutionary changes at the molecular level are caused not by Darwinian selection but by
random drift of selectively neutral mutants’. The theory, originally applied to protein variation, does not deny the role of natural selection in determining the course of adaptive evolution, but it assumes that only a minute fraction of DNA changes in evolution are adaptive in nature, while the great majority of phenotypically silent molecular substitutions exert no significant influence on survival and reproduction and drift randomly through the species (Kimura 1987). Neutral (or closely so) molecular variation together with population genetics theory offers powerful tools for ecological research. Dynamics of allele frequency change is determined by demographic parameters such as population size, population subdivision, breeding structure and family relations. All these parameters are ecological by definition.

A variety of neutral molecular markers, including mini- and microsatellites, different kinds of restriction fragments, PCR-markers and DNA sequences are available today for studying ecology and evolution (Moritz & Hillis 1996). The markers can be applied to evaluate evolutionary rates, processes and constraints on molecular change through time or to make inferences about population processes, systematics and phylogeny.

1.2.1. Molecular clock

Since the idea of the molecular clock was proposed by Zuckerkandl and Pauling (1965), dating of past evolutionary events has been attempted based on the assumption that the molecular evolutionary rate is constant and equal to the mutation rate of neutral markers (Kimura 1987). However, molecular evolutionary rate does not appear to change linearly with time. Different DNA sequences, for example, evolve at different rates. This observation holds true across nucleotide positions, among nonhomologous genes within a lineage, among classes of DNA within a genome, and among genomes within an organismal lineage (Avise 1994). Separate calibrations of the molecular clock are needed for the specific organism and DNA region of interest. One example of the use of a molecular clock was dating the roots of human mitochondrial DNA tree to 200 000 years ago, and suggesting an African origin for the modern humans (Cann et al. 1987).

1.2.2. Genetic structure of populations, phylogeography and gene flow

At least some degree of genetic differentiation of populations is usually found at some geographic scale. This differentiation may be caused by a number of factors, including social structure, mating system, dispersal ability, association of parents with their offspring, and habitat fragmentation, which in turn, lead to certain patterns of gene flow, genetic recombination, natural selection and random drift. When the haplotypes and phylogeny are superimposed over geographical locations the term ‘phylogeography’ has been introduced (Avise 1994).

Studying gene flow (genetically effective dispersal) in relation to the magnitude and spatial scale over which populations differ genetically offers a possibility to establish a
link between the ecology and evolution of species. Understanding the microevolutionary forces throughout the history of a species depends on quantification of how gene flow interacts with genetic drift, mutation and natural selection in forming spatial or temporal population structures (Bohonak 1999).

Molecular analyses of population structure and gene flow between subpopulations have been conducted for hundreds of species at a variety of temporal and spatial scales. Currently the most commonly used neutral or nearly neutral markers for studying population structure are microsatellites (e.g., subdivision of South African buffalo populations, Syncerus caffer, caused by habitat fragmentation; O’Ryan et al. 1998, philopatry and genetic differentiation in European harbor seals, Phoca vitulina vitulina; Goodman 1998) and mitochondrial DNA in animals (e.g., male-biased gene flow between rookeries of Australian green turtles, Chelonia mydas; FitzSimmons et al. 1997, lack of phylogeographic structure in Australian red kangaroo, Macropus rufus; Clegg et al. 1998, uniformity of the snapping turtle, Chelydra serpentina, populations; Walker et al. 1998) and chloroplast DNA in plants (e.g., recognition of three refugial sources of European oaks Quercus robur and Quercus petrea; Ferris et al. 1998).

1.2.3. Effective population size

The concept of effective population size (Ne) was first introduced by Wright (1931) referring to the size of an idealised model population that has the same genetic properties as observed for the real population. It is one of the most important parameters in population genetics. Under strict neutral theory, the molecular variability of a population is a function of the effective population size, mutation rate and gene flow (Kimura & Ohta 1971b). Usually the effective population size is much smaller than the census size. The ratio of effective population size to census size average only 0.1 when fluctuations in population size, variance in family size and unequal sex ratios are included (reviewed in Frankham 1995). Several factors affect the prediction of the effective population size, such as sex ratio, mating system, selection, pattern of inheritance, changes of the population size over generations, and population subdivision (Caballero 1994). The estimation of the effective population size has important applications in evolutionary studies, in conservation biology and in breeding programs (Wang & Caballero, 1999). Mitochondrial DNA polymorphism can be used to determine the long-term female effective population size when the mutation rate is known. In Avise et al. (1988) all the three species studied (American eel Anguilla rostrata; hardhead catfish, Arius felis; red-winged blackbird Agelaius phoeniceus) had smaller long-term effective population sizes than present-day census sizes.
1.2.4. Population growth or decline

Change in population size can be detected by examining the patterns of genetic polymorphism. The number of segregating sites (number of polymorphic nucleotides per nucleotide site) is influenced by the current population size more strongly than is the average number of nucleotide differences, while the average number of differences is affected by the historical population size more than the number of segregating sites (Tajima 1989a). The distribution of pairwise genetic differences can also be used for detecting population expansion or decline (Rogers & Harpending 1992, Rogers 1995), though it is possible that this method is insensitive to further changes in population size. Lavery et al. (1996) studied the sequence divergence in coconut crabs (*Birgus largo*) and noted that the effects of more recent events may not be detectable if the population is not in genetic equilibrium due to past growth events.

In addition to detecting growth or decline of a population, the size and length of a fairly recent population bottleneck can be estimated using the genetic data obtained from natural populations as was done with the endangered elephant seals, *Mirounga angustirostrus* (Hedrick 1995).

1.2.5. Phylogeny

Phylogeny is a stream of gene transmission that flows from generation to generation and continues through conspecific populations, subspecies and species to higher taxonomic levels. Based on this idea, the microevolutionary processes described above can to some extent be applied to explain macroevolutionary differences of higher taxa (Avise et al. 1987).

The coalescent theory is a population genetic model of neutral evolution applied usually to interbreeding populations, instead to different taxa like the phylogenetic methods (Tajima 1983). The coalescent approach is based on modelling mutational processes that occur along lineages. As long as the mutation rate is not zero, more closely related sequences will tend to be more similar. A backward temporal perspective is adopted, thus the name coalescent theory (Harvey & Nee 1996). Coalescence occurs at some point in the past between all pairs of individual lineages and in each generation that a coalescence occurs the number of ancestral stages is reduced by one, generating a bifurcating coalescence tree (Harding 1996).

The phylogenetic approach can be used to help find answers to a variety of evolutionary biology questions. In addition to resolving the taxonomic relationships of species, phylogenetics can be used to study the evolution of gene families (e.g. Zhang & Nei 1996, Johnston et al. 1998), evaluate evolutionary rates in different lineages (e.g. Kocher et al. 1989, Pamilo & O’Neill 1997), date past historical events (e.g. Cann et al. 1987, Klicka & Zink 1997), study the coevolution of host-parasite relationships (Hafner et al. 1994, Page et al. 1998), and clarify the sources of epidemic diseases (e.g. Zhu et al. 1998, Taubenberger et al. 1997).
1.3. Mitochondrial DNA

Animal mitochondrial DNA (mtDNA) is a small (15-20 kb) circular molecule, composed of about 37 genes coding for 22 tRNAs, two rRNAs and 13 mRNAs, the latter coding for proteins mainly involved in the electron transport and oxidative phosphorylation of the mitochondria. The mitochondrial genome is arranged very efficiently. It lacks introns, has small intergenic spacers where the reading frames even sometimes overlap. The control region is the primary non-coding region, and is responsible for the regulation of heavy (H) and light (L) strand transcription and of H-strand replication (Figure 1).

As a molecular marker, mitochondrial DNA has many advantages. It evolves faster than nuclear DNA (Brown et al. 1982), probably due to inefficient replication repair (Clayton 1984). Different regions of the mitochondrial genome evolve at different rates (Saccone et al. 1991) allowing suitable regions to be chosen for the question under study. Mitochondrial DNA is maternally inherited in most species (exceptions with paternal leakage including mice, Gyllensten et al. 1991; biparental inheritance in marine mussels, Zouros et al. 1992). Mitochondrial DNA does not recombine (Hayashi et al. 1985), though some evidence of recombination events has recently been reported (Eyre-Walker et al. 1999, Hagelberg et al. 1999). Individuals are usually homoplasmic for one mitochondrial haplotype though heteroplasmic conditions have been reported in many species (e.g., perches, Nesbo et al. 1998; Drosophila, Volz-Lingenhöhl et al. 1992; bats, Wilkinson & Chapman 1991). These features mean that each molecule as a whole usually has a single genealogical history through maternal lineages.

Whether the mitochondrial DNA can be considered a strictly neutral marker has been controversial (e.g. Rand & Kann 1996 and references therein). Though support for neutrality comes from the high evolutionary rate of the molecule (Brown et al. 1979, Brown et al. 1982, Vawter & Brown 1986), the assumed uniformity in the substitution rates of the mtDNA and the relaxed translation of mitochondrial mRNAs (Cann et al. 1984), mitochondrial DNA evolution more likely follows the mildly deleterious model or the nearly neutral model. Based on neutrality tests of molecular data from 14 studies Fry (1999) suggested that within species there is an excess of rare haplotypes and these haplotypes carry mildly deleterious mutations. Ballard and Kreitman (1995) point out in their review that selection on any part of the mtDNA has an influence on polymorphism in the whole molecule in the population because the lack of recombination makes mitochondrial genomes particularly susceptible to genetic hitchhiking. The heterogeneous substitution rates along lineages and the relative excess of replacement polymorphism (=substitutions leading to nonsynonymous amino acid changes) also support the idea that selection has a role in mtDNA polymorphism. Yet, Ballard and Kreitman (1995) state that genetic drift may be the prevailing force in mitochondrial evolution.
Fig. 1. Mitochondrial genomes of birds (a, b) and mammals and *Xenopus* (c). tRNA genes are identified by their 1-letter amino acid codes. Outer circle represents the heavy (H) strand and the inner circle the light (L) strand. Polarity of transcription and the transcribed strand is shown with the arrowheads. When no arrowhead is marked the gene is transcribed from the H-strand with clockwise polarity. The regions used in this study are marked with dark grey. The genomes are redrawn from Desjardins and Morais (1990) and Mindell et al. (1998).
However, the high value of using mtDNA in phylogenetic studies is not severely diminished by the uncertainty of mitochondrial neutrality. Even if molecules differ in fitness, properly identified synapomorphies still allow the recognition of monophyletic clades. Knowledge of mtDNA neutrality is essential for analyses involving genetic distance estimates and molecular clock (Avise et al. 1987).

Whether or not mtDNA is strictly neutral, it is a sensitive indicator of population level processes. Analysis of mtDNA divergence can be used to reveal geographic clusters of related molecules (individuals) or matrilineal relationships within populations. It can be used also to trace historical events like bottlenecks, or to analyse hybrid zones. MtDNA can also be very useful in resolving phylogenetic relationships between closely related taxa (Moritz et al. 1987).

### 1.3.1. Avian mitochondrial DNA

The first complete sequence of an avian mitochondrial genome was published from chicken by Desjardins and Morais (1990). It showed highly conserved features when compared to other vertebrate mtDNAs. The protein genes are very similar to the homologous genes in mammals and amphibians and they are translated using the same genetic code. Guanine is relatively infrequent at the third position of codons. Several genes overlap and several end with an incomplete stop codon that is completed by polyadenylation (Quinn 1997).

Though many features are the same in all the vertebrate mtDNAs, the avian genomes have some remarkable differences. First, the avian gene order is novel compared to mammalian and amphibian mitochondrial genomes (Figure 1). The ND5 gene (nicotinamide adenine dinucleotide dehydrogenase subunit 5) is followed by cytochrome b, tRNA^{Trp} and tRNA^{Val}, ND6 and tRNA^{Glu} in the 5’ → 3’ direction of the avian L-strand (Desjardins & Morais 1990, Desjardins & Morais 1991, Quinn & Wilson 1993). This rearrangement could have arisen through duplication via replication slippage followed by at least two independent deletion events (Quinn & Wilson 1993). Second, the L-strand replication origin that is found between tRNA^{Cys} and tRNA^{Asn} in other vertebrates is absent in the avian genome (Desjardins & Morais 1990, Desjardins & Morais 1991). In addition, COI (cytochrome oxidase I) has an unusual initiation codon GTG (instead of ATG) and there is evidence of low incidence of thymine at silent positions within coding regions (Quinn 1997). Recently, Mindell et al. (1998) found another gene order in avian mtDNA, found among four bird orders (Picidae, Cuculidae, suboscine Passeriformes and Falconiformes; figure 1). This gene arrangement probably has multiple independent origins because it is found in quite divergent taxa.
Cytochrome $b$ is one of the cytochromes involved in the electron transport in the respiratory chain of mitochondria. It contains eight transmembrane helices connected by intramembrane or extramembrane domains (Figure 2, Esposti et al. 1993). It is the only cytochrome coded by mitochondrial DNA.

Fig. 2. Structure of the cytochrome $b$ protein. The gene region used in this study corresponds the shaded parts of the protein.

The cytochrome $b$ gene is the most widely used gene for phylogenetic work for several reasons. Although it evolves slowly in terms of non-synonymous substitutions, the rate of evolution in silent positions is relatively fast (Irwin et al. 1991). The wide use of cytochrome $b$ has created a status as a universal metric, in the sense that studies can be easily compared. Cytochrome $b$ is thought to be variable enough for population level questions, and conserved enough for clarifying deeper phylogenetic relationships. However, the cytochrome $b$ gene is under strong evolutionary constraints because some parts of the gene are more conserved than others due to functional restrictions (Meyer 1994). Most of the variable positions seem to be located within the coding regions for transmembrane domains or for the amino- and carboxy-terminal ends (Irwin et al. 1991).

So far, cytochrome $b$ has been the most prevalent source of sequence data in avian studies. Although use of cytochrome $b$ has some pitfalls, Moore and DeFilipps (1997) argue that it could anyhow be the best choice for resolving relatively recent evolutionary history. The tendency of birds to have low sequence divergence rates at high taxonomic levels compared to other vertebrates makes cytochrome $b$ a good choice as a marker.

Helm-Bychowski and Cracraft (1993) provided a good example of low divergence in a phylogenetic study of corvine passerine birds. Although some monophyletic groups could
be identified at the family level, some remained unsolved, possibly because of a rapid radiation of lineages over a relatively short period of time. Similar results are obtained also from cardueline finches, where all the branches of the phylogenetic tree could not be convincingly resolved (Fehrer 1996). However, cytochrome \( b \) gene divergence suggested that some subspecies should be elevated to species status in Tanagers (*Ramphocelus*; Hackett 1996), and the taxonomy of swiflets (Apodidae) based on the presence or absence of echolocating ability has been called into question (Lee *et al.* 1996). The cytochrome \( b \) sequences have successfully been used to identify taxonomic groups even at subspecies level, for example in bluethroats (*Luscinia svecica svecica* and *L. s. namnetum*; Questiau *et al.* 1998) and common guillemots (*Uria aalge*; Friesen *et al.* 1996).

### 1.3.3. Control region of mtDNA

The mtDNA control region is the only large non-coding region in avian mitochondria varying from 1044 bp in *Cairina moschata* (Liu *et al.* 1996) to 1227 bp in *Gallus domesticus* (Desjardins & Morais 1990). It contains the heavy-strand replication origin and the promoters for both the L- and H-strand transcription (L’abbe *et al.* 1991). This region is divided into three domains (Figure 3) according to the criteria of Baker and Marshall (1997). The first domain at the 5’ end of the control region contains a C-stretch and the putative termination associated sequence (TAS). The C-stretch is characteristic for the 5’ terminus of the avian control region being present in various forms at least in Anatidae, Phasianidae and Paridae (Quinn & Wilson 1993, Desjardins & Morais 1990, Marshall & Baker 1997, works II-IV). A C-stretch was also found from the platypus (*Ornithorhynchus anatinus*; Janke *et al.* 1996) and the frog (*Rana catesbeiana*; Yoneyama 1987) control region, but not at the 5’ terminus.

The central domain is the most conserved. It contains several structural elements that can be readily aligned even between different bird families. Three of these elements (F-, D- and C-boxes) were identified by Desjardins and Morais (1991) from *Gallus domesticus* and by Quinn and Wilson (1993) from *Anser caerulescens*. In addition, there are highly homologous regions among bird families upstream from F- and C-boxes designated as E- and B-boxes after Southern *et al.* (1988). In the middle of the E-box and the D-box there are short regions (rebox and Mt-3, respectively) that are homologous to sequences able to bind nuclear transcription factors connected to the regulation of oxidative phosphorylation (Wallace 1993, Suzuki *et al.* 1991).

Usually the most variable part is the third domain at the 3’ end of the control region. This domain begins with conserved sequence block 1 (CSB-1). The mitochondrial transcription factor (mtTFA) is probably bound to CSB-1, and mediates the transition from transcription to replication when another factor is bound to the mtTFA-CSB-1-complex (Ghivizzani *et al.* 1994). The rest of the third domain seems to be free from functional constraints and large interspecific insertions or deletions, as well as intraspecific tandem repeats (loggerhead shrike *Lanius ludovicianus*; Mundy *et al.* 1996, Ciconiiformes; Berg *et al.* 1995) are concentrated in this region.
### Fig 3

Control region alignment of one individual from each of four *Parus* species and the structural elements in the region. Identical nucleotide sites are shown by stars.

<table>
<thead>
<tr>
<th>Species</th>
<th>Structural Elements</th>
<th>80</th>
<th>480</th>
<th>490</th>
<th>580</th>
</tr>
</thead>
<tbody>
<tr>
<td>Willow tit</td>
<td>C-stretch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blue tit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Great tit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siberian tit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structural Elements</th>
<th>80</th>
<th>480</th>
<th>490</th>
<th>580</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-stretch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAB?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSB-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III Insertion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Note: The diagram includes sequences and structural elements such as C-box, B-box, E-box, and Rebox, with specific nucleotide sites marked by stars.
The control region has been shown to often evolve faster than the rest of the mitochondrial genome (e.g. hominoids; Horai et al. 1995) and to be highly variable in birds (Wenink et al. 1993). This variability has lead to the expanding usage of control region sequences to examine questions ranging from population structure to phylogenetic relationship. This region has already proven to be quite a powerful tool in elucidating the global population structure in shorebirds (Wenink et al. 1993, 1994, 1996) and fringilline finches (Marshall & Baker 1997, 1998), in revealing recent mixing of maternal lineages in snow geese (Quinn 1992), and in evaluating gene flow between social groups and populations in babblers (Edwards 1993). However, in some avian genera (e.g. gnatcatchers, Polioptila; Zink & Blackwell 1998 and towhees, Pipilo; Zink et al. 1998) mitochondrial coding genes have recently been shown to evolve as rapidly as the control region.

Fewer studies have dealt with control region sequences at higher taxonomic levels. Kidd and Friesen (1998) succeeded in resolving the phylogenetic relationships between three species of guillemots, but failed to solve the branching order of the subspecies. Between closely related goose species (Anser), control region sequences have been shown to resolve phylogenetic relationships quite reliably (Ruokonen 2000, in press).
2. Outlines of the present study

The work presented here is closely connected to studies on the evolution and ecology of Palearctic bird populations. My contribution to this field was to study genetic variation of different levels of phylogenetic hierarchy in Parids.

The first aim was to set a phylogenetic framework for the Parids using DNA sequences of the mitochondrial cytochrome b gene. Such information provides a background for conducting more detailed population level studies.

The second aim of the work was to examine genetic variability and population structure in three species, the willow tit (*Parus montanus*), the great tit (*P. major*) and the blue tit (*P. caeruleus*) by using mitochondrial control region sequences. One of the main reasons for choosing these species was that they are well studied ecologically, which is essential when conclusions are made from genetic data. I addressed the following specific questions:

1. are the populations structured geographically?
2. what is the amount of gene flow between the sampled populations?
3. what is the amount of genetic variation within the populations and between the populations?
4. how did the post-glacial colonisation affect the present population structures?
3. Materials and methods

3.1. The bird species and populations studied

The tits and chickadees are small cavity-nesting passerine birds forming the family Paridae, genus *Parus*. They are mostly found in deciduous, evergreen and mixed woodlands, including parks and gardens (Cramp & Perrins 1993). The genus *Parus* contains 30-40 species, of which many occur in wide geographical ranges. Many species share similar habitats and their distribution ranges coincide (Snow 1954). The genus is divided into 5-12 groups depending on authors (eg. Snow 1954, Eck 1988, Cramp & Perrins 1993). The Fennoscandian tits can be allocated into five of these groups following Snow (1954): *Major* (great tit, *P. major*), *Cyanistes* (blue tit, *P. caeruleus*), *Lophophanes* (crested tit, *P. cristatus*), *Periparus* (coal tit, *P. ater*) and *Poecile* (marsh tit, *P. palustris*, willow tit, *P. montanus* and Siberian tit, *P. cinctus*). The phylogeny of these species was examined in study I, added with three American chickadees, of which two belong to the *Poecile* group: *P. atricapillus* (black-capped chickadee) and *P. carolinensis* (Carolina chickadee). The third, *P. inornatus* (plain titmouse) belong to the *Baeolophus* group and was used as an outgroup.

The population genetic structure of three of these species was studied more closely (the willow tit, *P. montanus* in study II, the great tit, *P. major* in study III and the blue tit, *P. caeruleus* in study IV). These species are territorial during the breeding season but form flocks outside the season. The adults are highly sedentary after natal dispersal but juveniles may perform irruptive autumn movements (Tiainen 1980, Koivula & Orell 1988, Gossler 1993, Cramp & Perrins 1993, Orell *et al.* 1994). All three species are mainly monogamous but occasional extra-pair fertilisations may occur (Gullberg *et al.* 1992, Kempenaers *et al.* 1992, Orell *et al.* 1997).

The distribution area of the willow tit covers a zone from north-eastern France through the central and northern Europe and Russian Taiga to the Pacific coast and over the Bering strait into Alaska (Cramp & Perrins 1993). The great tit’s range is more southern, reaching the Mediterranean Sea, the north coast of Africa, India and Indonesia in the south and northern Norway through Sweden and Finland to southern Russia and northern China in the north, restricted by the Pacific Ocean in the west and Atlantic
Fig. 4. Distribution areas of the willow, the great and the blue tit and their sampling sites.
Ocean in the east (Gosler 1993). The blue tit is the most southern and western of these three species. In the south, it occurs in the African north coast and on the Canary Islands. This species reaches the Atlantic coast in the west and the Caspian Sea in the east and the northern limit is in central Fennoscandia and southern Russia (Cramp & Perrins 1993).

The two willow tit populations sampled in this work were located at the western side of the species’ distribution area, representing the subspecies *P. montanus borealis*. Similarly, the eight populations sampled from the great tit were all from the western edge of the species range and all belonged to the nominate subspecies *P. major major*. Five blue tit populations belonged to the nominate subspecies *P. caeruleus caeruleus* and the sixth, most south-western population was from an area where the nominate subspecies occurs together with another subspecies *P. caeruleus oeliastra* (Figure 4).

3.2. Molecular methods

Material for obtaining DNA was collected from a variety of sources during years 1995-1997. Total DNA was isolated mainly from blood, pectoral muscle (Sambrook 1989) or feathers (Walsh *et al.* 1991) and mitochondrial DNA was isolated from heart and embryonic plates or embryos from eggs (Tamura & Aotsuka 1988). The amplification of the mitochondrial cytochrome *b* gene or the control region was performed by polymerase chain reaction, and if necessary the amplified fragments were cloned into the TA vector (Invitrogen). The PCR fragments were purified from agarose gels prior to sequencing (Glenn & Glenn 1994). Initially, the PCR fragments were sequenced manually from the vector or using the protocols for double-stranded sequencing (Bernatchez *et al.* 1992) of the PCR products. Later, all sequencing was performed with the ABI 377 DNA Sequencer using double-stranded PCR-products directly as templates.

3.3. DNA Sequence analysis

3.3.1. Genetic distances and phylogenetic methods

All sequences were aligned prior to any analysis. The correct alignment of the sequences is fundamental to identifying homologous characters. The alignment often includes the creation of gaps in the sequences. Once the alignment has been made, variation can be treated in different ways according to the substitution type or the way to consider the gaps (Thompson *et al.* 1994).

The most common way to evaluate the degree of sequence dissimilarity is to calculate pairwise genetic distances. In its simplest form, the distance is the estimate of the number of nucleotide substitutions per nucleotide site between two sequences. Usually the distance is ‘corrected’ in some way due to factors such as different substitution rates of transitions versus transversions and multiple substitutions at one site. The larger the
genetic distance, the more efficient the correction has to be to take into account the possibility of ‘multiple hits’ at each nucleotide site. The two distance methods used in this study are the Jukes-Cantor distance (II, II, IV; Jukes & Cantor 1969) and Kimura’s two-parameter distance (I; Kimura 1980).

Three basic methods were used to construct the phylogeny of tits: 1) neighbor-joining method (a distance based method); 2) maximum parsimony and 3) maximum likelihood. The neighbor-joining method identifies the closest pairs of taxonomic units by the distances between them. A pair of these neighbors is defined to be two units connected through a single node in an unrooted bifurcating tree, where two branches join at each interior node. This method continues by successive clustering of the lineages, setting branch lengths as the lineages join. The tree does not assume a constant substitution rate i.e., an evolutionary clock (Saitou & Nei 1987).

Parsimony methods are based on character values observed for each species, rather than the distances between the sequences. Branch lengths are generally not obtained. For each topology, the sequences at each node are inferred to be those that require the least number of changes to give each of the two immediately descendant lineages. The total number of changes required to traverse the whole tree is then found, and the tree with the minimum total is the most parsimonious (Edwards & Caffalli-Sforza 1963, Fitch 1971).

The parsimony approach is used also in describing intraspecific phylogeny of mtDNA molecules. When the number of informative characters is too low and much homoplasy exists to construct traditional parsimonious phylogenies, the median networks or minimum spanning networks can be used. With this method, it is possible to illustrate the most parsimonious tree, polytomies and homoplasmic sites in a single network that relates the sequence based haplotypes (Bandelt et al. 1995).

Maximum likelihood methods do not reduce the data into distances, and use all the data instead just the phylogenetically informative sites. It differs from parsimony also by employing standard statistical methods for a probabilistic model of evolution. Maximum likelihood methods consider that changes are more likely along long branches than short ones, and estimation of branch lengths is an important component of this method (Felsenstein 1981, Olsen et al. 1994).

3.3.2. Population diversity indices and population divergence

Most analyses of genetic variability in a population assume it to be in an equilibrium with respect to mutations, gene flow and population size (Kimura & Ohta 1971b). However, this assumption hardly ever holds. For instance, measures of genetic variability may increase in a spatially structured population (Amos & Harwood 1998). On the other hand, the measures may show quite a uniform structure, if the population has rapidly expanded in size following a bottleneck, even if the population has some structuring (Walker et al. 1998). Thus, the knowledge of the ecology of the species under study, as well as the processes affecting the population enriches conclusions based on population parameters.

The diversity indices used throughout this study describe the DNA sequence polymorphism, and thus the amount of genetic variation in the population.
Nucleotide diversity $\pi$ is the average number of nucleotide differences per site between sequences

$$\pi = n/(n - 1) \Sigma x_i x_j \pi_{ij}$$  \hspace{1cm} (Nei 1987, equation 10.5)

or

$$\pi = \Sigma x_i / n_c$$  \hspace{1cm} (Nei 1987, equation 10.6)

where $n$ is the number of sequences examined, $x_i$ is the frequency of the $i$th type of DNA sequence in the sample and $n_c$ is the total number of sequence comparisons.

The number of polymorphic nucleotide sites per nucleotide site is estimated with $\theta$

$$\theta = s/a_1, \quad a_1 = \Sigma i (1/i)$$  \hspace{1cm} (Tajima 1996, equation 3)

where $s$ is the proportion of segregating sites among a sample of DNA sequences and $a_1$ is a correction factor to account for the number of sampled sequences.

Given several assumptions, including neutrality, the nucleotide diversity and number of polymorphic nucleotides per site should be equal. If this is not the case, the marker might be affected by selection, but also by other processes including population expansion, bottleneck, or mutation rate heterogeneity (Tajima 1989b, Aris-Brosou & Excoffier 1996). The haplotype diversity, $h$, describes the number and frequency of different haplotypes in the sample

$$h=1-\Sigma f_i^2$$  \hspace{1cm} (Nei & Tajima 1981, equation 6).

The coalescence time (the time in the past at which two ‘genes’ share a common ancestor) and the effective population size also describe the diversity of the population. The more diverse the individuals are in the population, the longer the coalescence time and the larger is the effective population size. Both of these estimates are approximated according to Avise et al. (1988):
\[ t = (0.5 \times 10^8 \times p) \]

\[ Ne_f = (0.5 \times 10^8 \times p)/g \quad (\text{Avise et al. 1988, equation 2}) \]

where the \( p \) is the mean pairwise genetic distance, \( g \) is the generation length and the conventional mitochondrial divergence rate of 2%/Myr for birds is assumed (the basis for this assumption is explained in III, see also Tarr & Fleischer 1993, Arbogast & Slowinski 1998, Klicka & Zink 1998, Avise & Walker 1998, Klicka & Zink 1999).

The distribution of the pairwise genetic distances, ‘the mismatch distribution’ was compared to the expected distributions of a population that has recently grown in size (the expansion population) or has been stationary over a long time period (Rogers & Harpending 1992, Rogers 1995).

The divergence of the populations was estimated mainly by the analysis of molecular variance (AMOVA; Excoffier et al. 1992). The variance components were estimated from the molecular data (as pairwise genetic distances and haplotype frequencies). The total genetic variance was partitioned into that caused by the differences between individuals within populations and that caused by the differences among populations. Pairwise differences of the populations were calculated as \( F_{st} \) values, which can also be used to measure the short-term genetic distances between the populations (Reynolds et al. 1983). If possible, the number of migrants between the populations (\( N_{mf} \)) was estimated for haploid organisms

\[ N_{mf} = \frac{1}{2} \left( 1/F_{st} - 1 \right) \quad (\text{Nei, 1987, equation 13.25}) \]

or

\[ N_{mf} = \frac{1}{2} H_w/(H_b-H_w) \quad (\text{Hudson et al. 1992, equation 4}) \]

where \( H_w \) is the mean number of differences between different sequences sampled from the same subpopulation and \( H_b \) is the mean number of differences between sequences sampled from the two different subpopulations.
4. Results

4.1. Structure of the control region of the tits

The sequence alignment and the conserved elements of the control region of one individual of each, the willow (II), great (III) and blue (IV) tits, are shown in figure 3. In addition, the Siberian tit, a close relative to the willow tit, was included in the alignment.

The length of the control region was quite constant between the species studied: 1205-1207 bp in willow tit (25 birds), 1188-1189 bp in great tit (68 birds) and 1191 bp in blue tit (43 birds). Most of the insertions or deletions causing length differences between the species are located within the third domain of the control region. This portion of the control region seems to be quite free from functional constraints and has very low similarity among the species.

Conserved elements were identified on the basis of sequence similarity with other vertebrate sequences (Yoneyama 1987, Southern et al. 1988, Desjardins & Morais 1990 and 1991, Suzuki et al. 1991, Quinn & Wilson 1993, Janke et al. 1996, Marshall & Baker 1997). The first domain contained a C-stretch, which is located near to the 5’ end of the control region and is identical among the tit species studied. A possible TAS element, identified on the basis of the consensus sequence from eleven vertebrate species aligned by Foran et al. (1988), was also located in the first domain.

Most of the conservative boxes were within the second domain (F-, E-, D-, C- and B-boxes and part of the CSB-1). The E-box showed the least amount of sequence similarity (55 %). The other boxes had 83-96 % similarity over the four species. The CSB-1, beginning at the end of the second domain and ending at the third domain, contained an insertion that is also found in some other Passerines (e. g. in the common chaffinch, Fringilla coelebs). This insertion has 85 % similarity with an avian infectious bronchitis virus (Penzes et al. 1994).
4.2. Taxonomy and diversity at the species level

The phylogenies constructed from the cytochrome $b$ sequences by different methods were quite similar in overall topology. The *Poecile* group was distinguished from the others, and the great tit and blue tit grouped together. The coal tit and the crested tit were placed somewhere in between the former two groups (I). However, if more species could have been included, the topology of the phylogenetic tree could change also with respect to some of the currently included species.

The shortest distance, corrected by the Kimura’s two-parameter method, among species was found between the great tit and the blue tit (2.32%). This result was surprising because the two species are traditionally placed into different species groups (*Major* and *Cyanistes*, respectively). The distance between even the closest relatives within the *Poecile* group (the black-capped chickadee and the Carolina chickadee) was larger (3.28%).

For comparison, we reconstructed small phylogenies using the control region sequences and cytochrome $b$ sequences of the great, blue, willow and Siberian tits (Figure 5). The distance estimated from the control region between the great tit and the blue tit was 10.57%, five times larger than when estimated from the cytochrome $b$ sequences. On the contrary, the distance between the Siberian and the willow tit was lower (5.34%) when estimated from the control region compared to the estimate from the cytochrome $b$ (8.1%).

The largest distance found by comparing the cytochrome $b$ sequences was 12.3% between the willow tit and the crested tit. The mean distance between the different species groups was 9.29% (2.32%-11.15%).

---

**Fig. 5.** Comparison of the phylogenetic trees built from the control region and cytochrome $b$ gene sequences. The numbers along branches denote percentages supporting the branching in 1000 bootstrap replicates.
4.3. Diversity indices and coalescence times within populations

Intraspecific diversity was studied in three species using the control region sequences: the willow tit (II), the great tit (III) and the blue tit (IV). The samples from the willow and the great tit showed no structuring; all the studied populations seemed to belong to one panmictic population. The blue tits, however, revealed two maternal lineages. The ‘southern’ lineage found in Barcelona differed from the northern lineage by a mean genetic distance of 1.3%.

The nucleotide diversity ($\pi$) estimated within lineages was 0.53% in the willow tit, 0.19% in the great tit and 0.19% in the northern lineage of the blue tits (diversity within southern lineage was 0.3%, but consisted of only two birds). The estimated number of segregating nucleotide sites ($\Theta$) was 1.10% in willow tits, 0.94% in great tits and 0.73% in northern blue tits. All the willow tits sampled from two Fennoscandian populations had unique haplotypes (haplotype diversity 0.96). In contrast, seven out of the eight European populations sampled from the great tit shared the same most common haplotype. Altogether, one third of the great tits had this common haplotype (haplotype diversity was 0.86). The pattern in the blue tit was quite similar to the great tits. Four of six populations shared the most common haplotype, which was found altogether in 37% from the birds (haplotype diversity was 0.84, table 1).

![Table 1. Nucleotide diversity (%$\pi$) and haplotype diversity (1-$\sum l^2$) of the great, blue and willow tit populations studied.](table1.png)

The estimate of the female effective population size also reflects the fact that diversity within the populations varied extensively. Willow tits had the largest estimate of effective population size (115 000) while great tits and the northern lineage of blue tits had lower estimates (47 000 and 49 000, respectively).

The estimate of coalescence time, which is just the female effective population size divided by the generation time, is as well dependent on the present diversity of the
populations. Within lineage coalescence times are the longest for the willow tit (260,000 years), falling into the period of Middle Pleistocene. The coalescence time of the great tits and the blue tits (93,500 years and 96,000-120,000 years, respectively) was more recent and fits to the occurrence of the latest glaciation. The divergence of the two blue tit lineages probably happened at the beginning of the Middle Pleistocene, 650,000 years ago.

4.4. Population structure and gene flow

The two Fennoscandian willow tit populations studied showed no phylogeographic structure (II). The same was also true for the eight great tit populations (III) and the five northernmost blue tit populations. On the contrary, the most southern blue tit population from Barcelona contained specimens widely different from the northern majority (IV).

The minimum spanning network constructed from the haplotypes of the willow tits formed a net-like structure, where the birds from the two populations were randomly distributed meaning that the populations are historically panmictic. This is supported by the observed immigration rates studied in the Oulu population and the recovery data from the Finnish Bird Ringing Center. The distribution of pairwise genetic distances was very close to the expected distribution under population expansion. The postulated expansion was further supported by the significantly negative Tajima’s D value (II).

The minimum spanning networks formed from the great tit and the ‘northern’ type of blue tit were more star-like having the central most common haplotypes. Tajima’s D values were significantly negative and the mismatch distributions followed those expected of expanding populations. This result and the presence of the common haplotypes in almost all the sampled populations support the population expansion of both species connected to the colonisation of the present distribution area (III, IV).

The blue tit population from Barcelona consisted of haplotypes that belonged to two maternal lineages. The other lineage, referred as the southern lineage, was inferred to represent the Iberian subspecies of the blue tit. Altogether, 13 nucleotide substitutions separated the northern central haplotype and the most distant southern haplotype (IV).
5. Discussion

5.1. Sequence variation in the cytochrome b and in the control region and their suitability to phylogenetic and population genetic studies

Two different regions of mtDNA were examined. Cytochrome b provided information about the phylogenetic relationship of the closely related Parus species and control region allowed for inferences about population structure within species to be made.

Cytochrome b sequences have been the most prevalent source of sequence data in avian studies (e.g., Richman & Price 1992, Helm-Bychowski & Cracraft 1993, Fehrer 1996, Hackett 1996, Lee et al. 1996, Friesen & Anderson 1997, Kimball et al. 1997, Groth 1998, Johnson & Sorenson 1998, Sheldon et al. 1999). Mostly this sequence has been used to analyse the phylogenetic relationships at the genus level, as in the case of Parus species presented in I. Generally the level of sequence variation within a genus is suitable for achieving good resolution in phylogenetic analyses. This is true for the genus Parus, where the average Kimura’s two-parameter distance was 0.095 (range 0.02-0.13). These results are comparable with the average distance found in the Sulidae (0.136, range 0.0075-0.234; Friesen & Anderson 1997), and in the Carduelinae (0.075, range 0.023-0.114; Fehrer 1996). Although there is enough variation, the nature of this variation can cause problems in phylogenetic analyses. In Parids, the low resolution of the phylogenetic tree (I) could be caused by the fairly high level of homoplasy in the sequences (e.g., Stewart 1993).

Setting the phylogenetic background to the Parids provided a scale where the variability within species could be placed. The knowledge of the relationships of the species later studied at the population level was essential when making comparisons between the species. The phylogenetic distance of the great tit and the blue tit, which showed similar features in population genetic studies, was very low. These two species were the closest relatives among all the species used for the phylogenetic reconstruction. The willow tit was quite distant from this species-pair.

In the control region the variable nucleotide sites were distributed quite uniformly along the whole region. In most bird species studied so far, intraspecific variation has been most commonly found in the flanking domains I and III, and the central domain II
has been the least variable (e.g. Wenink et al. 1993, Edwards 1993, Wenink et al. 1994, Marshall & Baker 1997, Zink & Blackwell 1998). Although the nucleotide substitutions within species were evenly distributed, the substitutions between the species were mostly concentrated in the third domain, while the central domain was quite conserved (Figure 6).

Fig. 6. Pairwise comparison of the nucleotide substitutions in the control region of four *Parus* species.
The nucleotide diversity measured from the DNA sequences of the whole control region of the blue and the great tit does not differ from the nucleotide diversity estimated over the entire mitochondrial genome using RFLP (no RFLP data are available from the willow tit). This observation suggests that the substitution rate in Parus is approximately equal in the control region and in the entire molecule. Several other studies on birds (e. g. the dunlin, *Calidris alpina*, Wenink *et al.* 1993; the lesser snow goose, *Anser caerulescens*, Quinn 1992) show considerably higher substitution rates in the control region than in the rest of the molecule. Lower rates have been also observed, in the bluethroat, (*Luscinia svecica*, Questiau *et al.* 1998) and in the loggerhead shrike (*Lanius ludovicianus*, Mundy *et al.* 1996). In gnatcatchers (*Polioptila*, Zink & Blackwell 1998) the variation in the control region was about the same as in third codon positions of coding genes. The substitution rate in the control region seems to vary between species and genera. Although the amount of variation in the Parus control region is not higher than that observed in the RFLP of the entire mitochondrial genome, there is enough variation for studies on the population level. Similar results are observed in the song sparrows (Emberizidae), but the control region sequences provide a more geographically structured haplotype tree than the RFLP data (Fry & Zink 1998). In addition, the phylogeny of haplotypes can be reconstructed when it is possible to identify the specific nucleotide substitutions.

The substitution rate in the Parus control region differs from the substitution rate in the cytochrome *b* gene. The overall trend is that the control region evolves faster than cytochrome *b* (table 2 and figure 5), which is expected when the function of the cytochrome *b* is taken into account. However, the opposite result is observed in the comparison between the Siberian tit and the willow tit. The difference between control region and cytochrome *b* distances is unexpectedly large between the great tit and the blue tit. The nucleotide composition of the control region and cytochrome *b* is also slightly different. Thymine is especially more abundant in the control region, and cytosine in cytochrome *b* (Figure 7).

![Fig. 7. Nucleotide compositions of the control region and the cytochrome b gene of the Parids. The mean frequency of nucleotides has been calculated from one individual per each of blue tit, great tit, willow tit and Siberian tit.](image_url)
To conclude, cytochrome \( b \) sequences are good tools for studying phylogenetics of closely related species. The structural and functional conservation of the cytochrome \( b \) gene mainly prevents its usage for intraspecific studies. Within species, control region sequences usually are a better choice, because more relaxed structural and functional constraints lead to a faster average substitution rate.

Table 2. Kimura’s two parameter distances of four Parus species. Above the diagonal are the mitochondrial control region distances. Below the diagonal are the cytochrome \( b \) gene distances based on a portion of the gene. The great tit and blue tit sequences are from Taberlet et al. (1992).

<table>
<thead>
<tr>
<th></th>
<th>Willow tit</th>
<th>Siberian tit</th>
<th>Blue tit</th>
<th>Great tit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Willow tit</td>
<td></td>
<td>0.0534</td>
<td>0.1309</td>
<td>0.1227</td>
</tr>
<tr>
<td>Siberian tit</td>
<td>0.0810</td>
<td></td>
<td>0.1335</td>
<td>0.1398</td>
</tr>
<tr>
<td>Blue tit</td>
<td>0.1090</td>
<td>0.0849</td>
<td></td>
<td>0.1057</td>
</tr>
<tr>
<td>Great tit</td>
<td>0.1154</td>
<td>0.0878</td>
<td>0.0232</td>
<td></td>
</tr>
</tbody>
</table>

5.2. Phylogeography and spatial population structure

Populations of a species usually exhibit at least some degree of genetic differentiation among geographic localities. This geographic structuring has several causes, such as social structure, mating system, dispersal capability, cohesion of parents with their offspring and habitat fragmentation. These processes lead to certain patterns of gene flow, genetic recombination, natural selection and random drift, which in turn have an impact on the structure (Avise 1994).

In bird studies, as in studies with other organisms, much research has focused on taxa that for some of the above mentioned reasons are \( a \ priori \) supposed to have geographic structure. This bias in species selection perhaps leads to a bias in the published results, where apparent geographic structure is frequently encountered (e.g. Zink 1996).

No phylogeographic structure was detected from the two willow tit populations, nor from the eight great tit populations, or among the five blue tit populations, excluding the Spanish population. Similar minimum spanning networks to the one constructed from the willow tits are rare in published avian data, but at least the network from European samples of the dunlin (Wenink et al. 1996) shows some resemblance. Still, the European dunlin seems to be developing population substructuring in some of the studied populations while the studied willow tit populations apparently have none.

Some examples of a starlike population structure, like the ones detected from the great tit and the northern lineage of the blue tit have been found among birds. Ball et al. (1988) reported a starlike network from the red-winged blackbird. Seutin et al. (1995) also found this pattern in the redpoll finches (Carduelis flammea-hornemanni-complex), as did Ellsworth et al. (1994) from the prairie grouse (Tympanicus-complex). These were all
detected by mtDNA RFLP. In addition Merilä et al. (1997) found similar population structure in the greenfinch (*Carduelis chloris*) using mitochondrial control region sequences. This type of population structure is best explained by rapid expansion after a bottleneck, when the molecular markers may show quite uniform structure over large areas. In such cases the haplotype and nucleotide diversities can be low, even if the population *de facto* has some breeding structuring (Walker et al. 1998). Another explanation for such a starlike population structure is a selective sweep, where one genotype is favoured by natural selection and that genotype has recently replaced all others (Rich et al. 1998).

The southern lineage of the blue tit was clearly diverged from the northern lineage, at least by 11 substitutions compared to the main northern haplotype. Whether this lineage has also gone through a bottleneck cannot yet be answered due to the few birds that were present in our sample.

### 5.3. Effective population size

Effective population size is one of the key parameters in population genetics. It is analogous to different measures of genetic variation within a population, which is a function of mutation rate, gene flow and population size (Kimura & Ohta 1971b). Several factors affect the prediction of effective population size, including sex ratio, mating system, selection, pattern of inheritance, changes in the population size over generations (e.g. population bottlenecks), and population subdivision (Caballero 1994).

In a spatially structured population, measures of genetic variability, such as effective population size may increase depending on migration patterns (Amos & Harwood 1998). Restricted gene flow, for example, can affect female effective population size, if the female dispersal is limited, as is found in many mammals (figure 1 in Chesser & Baker 1996). MtDNA is more sensitive to population subdivision and bottlenecks than nuclear genes because of its mode of inheritance (Wilson et al. 1985). The increase of female effective population size caused by population subdivision can be seen in the current study of blue tits, where $N_e$ estimated from the Barcelona population (consisting of two maternal lineages) was almost five times greater than in the five northernmost populations (including only one maternal lineage, IV). This result shows the effect of population subdivision, when the equilibrium in the mutation–migration balance is not yet reached, on the effective population size (Chesser & Baker 1996).

The estimates of effective population size are derivatives from the population diversity estimates, based on untested generalisations, and should therefore be taken with caution. However, these estimates can be used to make comparisons between closely related species more concrete. The northern blue tit and the great tit, for example, have no apparent substructure in the population; the haplotype and nucleotide diversities are low as is the effective population size. On the contrary, the willow tit populations, although not locally differentiated, are more variable, resulting in a large effective population size.
5.4. Post-glacial history

The last 700 000 years have been dominated by large changes in climate with roughly 100 000 year cycles of glacial-interglacial periods. Prior to this, the changes were less intense in the Pliocene, but the Northern Hemisphere ice sheets have varied in size for at least the last 2.5 Myr. Therefore, both the extant and extinct species have evolved in the face of huge changes in climate and biosphere (Webb & Bartlein 1992). The species must have gone through series of bottlenecks and expansions in population size and range according to the climatic conditions (Hewitt 1996, Hewitt 1999).

The last glaciation in Europe began some 120 000 years ago, being coldest about 20-18 000 years ago, when the ice covered the Europe to approximately 50°N. Warming began some 13 000 years ago, and the ice started to retreat. During the last 8 000 years, the climate has been much the same as today (Webb & Bartlein 1992). Even during the coldest periods, three regions in the Mediterranean area had temperate climates; the Iberian and Italian peninsulas, and the Balkans (Bennet et al. 1991). When the temperature decreased, the northern populations of many species became extinct, but the species may have survived during hostile times in these Mediterranean refugia. When the ice began to retreat, the refuge populations began to expand north (Taberlet et al. 1998).

In the European continent, several geographical features affected the expansion possibilities or routes. Europe itself is a large peninsula connected to Asia. The Mediterranean Sea in the south forms a barrier, as well as several mountains in Europe (the Alps, the Pyrenees and the Balkans). Taberlet et al. (1998) analysed 10 taxa to elucidate general colonisation routes in Europe. They found that the likely colonisation routes exhibited some similarities; northern regions were generally colonised from the Iberian and Balkan refugia, and Italian lineages were often isolated because of the Alpine barrier.

Mitochondrial DNA sequence divergence at different taxonomic levels can be used to date the coalescence of haplotypes (e.g. Klicka & Zink 1997, Avise & Walker 1998, Avise et al. 1998, Klicka & Zink 1999). Coalescence times were estimated for willow-, great- and blue tits. The coalescence time estimates of the great tits (93 500 years) and the blue tits (within northern lineage 96 000 years) matched the period of the last glaciation. However, the divergence time of the two blue tit lineages (northern vs. southern), 650 000 years, indicated that the subspecies P. c. caeruleus (northern) and P. c. ogliastre (southern) had already diverged in the Middle Pleistocene. The fact that the minimum spanning network of these two species had a starlike structure indicates that they have gone through a severe bottleneck after the time of separation during the last ice age. When geographical haplotype distribution is taken into account, the European great tit has survived the last glaciation in the Balkan refugium and expanded thereafter rapidly to all of Europe. The same holds true for the nominate subspecies of the blue tit. The present distribution of the southern subspecies P. c. ogliastre instead, is restricted to the Iberian peninsula, with some leakage over the Pyrenees to southern France. This subspecies likely was trapped into the Iberian refuge during the last ice age. The colonisation routes in Europe used by the great and blue tits (Figure 8) follow the general rules proposed by Taberlet et al. (1998).
Hewitt (1996) presented an explanation for how the expansion of a species into a previously unoccupied region can result in a mainly homogenous population structure with one dominant genotype. Rapid expansion means that the long distance dispersers of the species would be able to establish colonies ahead of the main distribution area. The genotypes of these efficient dispersers would dominate in the leading populations. The later migrants would contribute little since they have to enter an already established population. Their reproduction possibilities would be low while that of the original colonisers would be high. This spreading from the ‘leading edge’ involves a series of bottlenecks for the colonising population and leads to a loss and homogenisation of genotypes (see also Hewitt 1999).

In contrast, the coalescence time of the willow tit populations (260 000 years) is vastly longer than the within-lineage coalescence of the great and blue tits. The minimum spanning network is more reminiscent of a net than a star. This observation indicates that
the last glaciation has not severely depressed the effective population size of willow tits. The present distribution of the species is quite northern and in the southern limits of its range this species prefers alpine environments (Cramp & Perrins 1993). This species could possibly have been able to stay in larger inhabitable areas during the glaciation than great and blue tits, probably in taiga east of the continental ice (Svendsen et al. 1999). Therefore, the contracting influence of the last ice-ages may not have been as pronounced in the willow tit as in the blue and great tit.

5.5. Present or past gene flow

With the methods provided by the theoretical framework of population genetics gene flow (migration) can be measured indirectly. This has mainly been done using Wright’s Fst (standardised measure of the genetic variation among populations) to calculate Nm (number of migrants successfully entering a population per generation). Whether there are any true relationships among population structure, indirect measure of gene flow and dispersal of a species has remained ambiguous (e.g. Templeton et al. 1995, Bohonak 1999). Whitlock and McCauley (1998) have suggested that ‘the measures of genetic structure are valuable in their own right, but that transformations of these measures to quantitative estimates of gene flow or dispersal are at best not needed and, at worst, misleading’. Recently, new methods for estimating gene flow from DNA sequence data using knowledge of the phylogenetic tree of the sequences have begun to be developed (e.g. Hudson et al. 1992).

Because indirect measures of gene flow are dependent on population structure that is a result of at least thousands of generations, gene flow estimates may contain information that is relevant to the history of the species rather than reveal current gene flow (Bohonak 1999). For example, when the neighbouring populations are large, only a small fraction of gene pairs are closely related, and only this fraction gives information about the current rates of gene flow (Barton & Wilson 1995). Gene flow measures from mtDNA sequences that were suggested to be estimated by coalescence methods (Slatkin & Maddison 1989) have been criticised. The main point underlying the criticism is that because the mtDNA lineages are generally not randomised geographically, the mtDNA distributions are primarily generated by historical processes and not by current gene flow (Neigel 1997).

Templeton et al. (1995) presented a nested clade analysis of haplotype trees to solve the problems of separating current and past events. They considered three major biological factors that may cause spatial or temporal associations of haplotype variation. The first, restricted gene flow (mainly due to isolation-by distance) may be difficult to separate from the second, past fragmentation events. The third major factor is range expansion. Recurrent gene flow is indicated, when no association of haplotypes can be detected (i.e., the haplotypes are randomly scattered spatially or temporally; Templeton 1998). The haplotype trees (minimum spanning networks) constructed from the tits can be linked with these major factors. In the network from the willow tit, the haplotypes are scattered, indicating recurrent gene flow, further supported by the observed immigration
rates studied in the Oulu population (Orell et al. 1999) and recovery data from the Finnish Bird Ringing Center (Saurola 1981).

The third factor, range expansion, seems to fit the networks from the great and blue tits, except that the southern lineage of the blue tit is a result of either restricted gene flow, past fragmentation events or both. Range expansion northwards is still occurring in the blue and great tits (Haftorn 1957, Veistola et al. 1994, Väisänen et al. 1998), but expansion is more recent in the blue tit. Ecological studies have shown that the northernmost populations are dependent on a continuous immigration from the south to persist. Thus present day gene flow most probably has a great impact in homogenising the populations, but the last ice-age has also had a profound contribution to the present population structure.

However, the coalescence method and nested clade analysis need further development before they can routinely be used to estimate the amount of current gene flow. Therefore, the number of migrants is still frequently estimated from Fst, but the estimates need to be treated with caution (Whitlock & McCauley 1998).
6. Concluding remarks and some future aspects

Recently, much effort has been put into understanding evolutionary processes important in birds. This effort includes molecular evolution and systematics as well as morphological, behavioural and ecological research. For many issues, the phylogenetic framework is of utmost importance for making interpretations and comparisons among and between species.

Evolutionary processes acting within populations (microevolution) have been studied using molecular methods for some time. These processes, like mutation, drift and selection can explain evolutionary phenomena at higher taxonomic levels (macroevolution). Although studies are focused at a certain level one can not ignore the evolutionary patterns occurring at the other levels (Edwards 1997).

Mitochondrial DNA has proven to be of great value in resolving phylogenetic relationships at different taxonomic levels. Although different regions of the mtDNA are used depending on the purpose, the whole molecule is inherited as one linkage group. This results in only one gene tree, no matter how many mitochondrial genes are examined. The gene tree may differ from the species tree, for example in the case of introgression or lineage sorting. The possibility of inferring a correct species tree from the gene tree increases in small populations and with population bottlenecks. This possibility is especially true for the mitochondrial DNA, which has approximately one fourth of the effective population size of nuclear genes. Moore (1995) estimated that it would take as many as 16 nuclear gene trees to provide a species tree as reliable as that obtained from a single mitochondrial gene tree.

With the use of mitochondrial DNA, I have tried to contribute to the growing field of evolutionary research in birds. The studies presented here focus on evolutionary processes at the population and species level in Parids. What still remains to be done with the species used here is to widen the sampling. The research on the willow tit especially needs more sampling sites from more distant locations within the subspecies. But the major task in the near future will be to fill the gap between the population level and species level. All three species, the willow, great and blue tit are known to consist of several subspecies. These subspecies in turn form groups of presumably closely related members. For example, the subspecies of the great tit have been classified into four groups consisting of 30 races altogether (as classified in Gosler 1993) and forming a geographical
‘Rassenkreis’. At least some of the geographically close subspecies hybridise in nature, while more distant subspecies do not (Kühn 1950). The hybridisation enables gene flow through the whole species. How closely related are these subspecies phylogenetically? What kind of an impact has hybridisation in nature on the subspecies? Are the geographically distant species also phylogenetically distant? These interesting questions are just a few of those that await an answer.
7. References


