MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 4123–4139 doi: 10.1111/j.1365-294X.2011.05259.x

A multi-gene approach reveals a complex evolutionary history in the *Cyanistes* species group

JUAN CARLOS ILLERA,*† KARI KOIVULA,‡ JULI BROGGI,§ MARTIN PÄCKERT,¶ JOCHEN MARTENS** and LAURA KVIST‡

**Research Unit of Biodiversity (UO-CSIC-PA), C/ Catedrático Rodrigo Uría, s/n; Oviedo University, Campus del Cristo, 33006 Oviedo, Spain, †Island Ecology and Evolution Research Group, IPNA-CSIC, 38206 La Laguna, Tenerife, Canary Islands, Spain, ‡Department of Biology, Univ. of Oulu, PO Box 3000, FI–90014 University of Oulu, Finland, §Estación Biológica Doñana, CSIC, Avda. Américo Vespucio s/n, 41092 Sevilla, Spain, ¶Senckenberg Nat Hist Sammlungen, Museum Tierkunde, D-01109 Dresden, Germany, **Johannes Gutenberg Univ Mainz, Institut for Zoology, D-55099 Mainz, Germany*

Abstract

Quaternary climatic oscillations have been considered decisive in shaping much of the phylogeographic structure around the Mediterranean Basin. Within this paradigm, peripheral islands are usually considered as the endpoints of the colonization processes. Here, we use nuclear and mitochondrial markers to investigate the phylogeography of the blue tit complex (blue tit *Cyanistes caeruleus*, Canary blue tit *C. teneriffae* and azure tit *C. cyanus*), and assess the role of the Canary Islands for the geographic structuring of genetic variation. The Canary blue tit exhibits strong genetic differentiation within the Canary Islands and, in combination with other related continental species, provides an ideal model in which to examine recent differentiation within a closely related group of continental and oceanic island avian species. We analysed DNA sequences from 51 breeding populations and more than 400 individuals in the blue tit complex. Discrepancies in the nuclear and mitochondrial gene trees provided evidence of a complex evolutionary process around the Mediterranean Basin. Coalescent analyses revealed gene flow between *C. caeruleus* and *C. teneriffae* suggesting a dynamic process with multiple phases of colonization and geographic overlapping ranges. Microsatellite data indicated strong genetic differentiation among the Canary Islands and between the Canary archipelago and the close continental areas, indicating limited contemporary gene flow. Diversification of the blue tit complex is estimated to have started during the early Pliocene (≈ 5 Ma), coincident with the end of Messinian salinity crisis. Phylogenetic analyses indicated that the North African blue tit is derived from the Canary blue tits, a pattern is avian 'back colonization' that contrasts with more traditionally held views of islands being sinks rather than sources.

Keywords: blue tit, *Cyanistes*, Canary islands, gene flow, Mediterranean basin, reverse colonization, speciation

Received 4 April 2011; revision received 14 July 2011; accepted 21 July 2011

Introduction

The Mediterranean region is considered one of the main hotspots of biodiversity in the world (Myers *et al.* 2000; Blondel & Aronson 2010), and much of this diversity is considered to have been created during the Quaternary climatic oscillations (Hewitt 1999, 2000; Weiss & Ferrand 2006). Pleistocene glacial periods produced several range contractions of European biota around the Mediterranean Basin with postglacial expansions after the retreat of the ice sheets. Such contraction and expansion events have been shown to have effects on genetic structure, including differentiation and speciation processes between the glacial refuges during pleniglacials and postglacial expansions (e.g. Hewitt 1996, 2004;

Correspondence: Juan Carlos Illera, Fax: +34 98 5104777; E-mail: illerajuan@uniovi.es

Griswold & Baker 2002; Guillaumet *et al.* 2006; Hansson *et al.* 2008; Hourlay *et al.* 2008). Islands are usually considered as the endpoints of colonization processes (MacArthur & Wilson 1967), and therefore their impact on genetic structures in a wider geographical scale is often disregarded. For example, from the influence of the peripheral North Atlantic islands on the genetic structure of the Mediterranean taxa has scarcely been evaluated. However, islands are not necessarily the endpoints of the colonization processes (Bellemain & Ricklefs 2008), instead they might also serve as sources of new elements enriching the mainland biotas. With regard to the Macaronesian islands this has been demonstrated in the plant genera *Aeonium, Lotus,* and *Convolvulus* (Mort *et al.* 2002; Allan *et al.* 2004; Carine *et al.* 2004) and in the land snail genus *Theba* (Greve *et al.* 2010) in Macaronesia. In birds, phylogeographic studies have consistently described a sink role for populations inhabiting North Atlantic islands, without no back colonization events to the mainland (e.g. Marshall & Baker 1999; Pérez-Tris *et al.* 2004; Päckert *et al.* 2006; Illera *et al.* 2007). Nevertheless, recent avian phylogenetic studies have revealed several examples of colonization from islands to continents suggesting a role for islands in contributing to continental biodiversity (Filardi & Moyle 2005; Sheldon *et al.* 2009; Jønsson *et al.* 2011).

A critical evaluation of the possible role of island biotas in reverse colonization requires extensive sampling of extant taxa throughout their ranges, and use of multilocus genetic approaches. Here we used mitochondrial and nuclear genetic data to characterize the phylogeography and population structure of the blue tit complex (blue tit *Cyanistes caeruleus,* Canary blue tit *C. teneriffae* and azure tit *C. cyanus*) across their geographic range. The blue tit and Canary blue tit are small (≈10 g) Palaearctic passerines with breeding populations in North Africa, Europe and western Asia, meanwhile the azure tit has its main distribution range in Asia with small breeding populations in eastern Europe (Cramp & Perrins 1993; Del Hoyo *et al.* 2007). The blue tit and Canary blue tit were recently classified into distinct species based on mitochondrial DNA studies providing evidence of strong genetic structure between populations of Europe, the Canary Islands, and populations from North Africa (Salzburger *et al.* 2002; Kvist *et al.* 2005; Päckert *et al.* 2007). Indeed, European blue tit populations are phylogenetically closer to the azure tit than to the Canary blue tit, showing that the speciation process between the three taxa has been largely hidden by the stronger phenotypic divergence of the azure tit from its sister lineage, the European blue tit (Salzburger *et al.* 2002). Contractions of its range into southern European refugia due to Pleistocene glacial cycles have been well documented (Kvist *et al.* 1999, 2004). However, despite the proximity of the African mainland, there is no evidence of mixture of mitochondrial lineages between North African and European populations (Salzburger *et al.* 2002; Kvist *et al.* 2005). These results suggest that the Mediterranean Sea forms an effective barrier limiting the gene flow between Europe and Africa. Although our understanding of the genetic structure of this species complex is improving, we know little about the timing, colonization pathways and diversification within this group and knowledge about the effects of Quaternary glacial and interglacial periods is limited.

In addition to climatic events in continental setting, colonization of islands such as the Canary Islands also promote diversification and speciation. Traditionally, four endemic subspecies based on differences in morphology, plumage and song have been recognized within the Canary Islands (Martín & Lorenzo 2001), however the genetic structure of this species is far from simple. Kvist *et al.* (2005) studied the colonization and diversification of the blue tits in the Canary Islands and found unexpected divergences of Gran Canaria (central Canary Islands) and La Palma (western Canary Islands; Fig. 1) from the other islands. The Gran Canaria population exhibited a cryptic lineage (Kvist *et al.* 2005) and was consecutively described as a distinct subspecies (Dietzen *et al.* 2008). However, the most striking result was found in La Palma. Although all Canary blue tits share several common substitutions suggesting a common ancestor, birds inhabiting La Palma also share a 12 base pairs (bp) fragment with the European blue tit, a mutation that is absent within the remaining Canary Island and North African populations. These findings suggest some form of connection between the European and North African lineages, which could support a diversification route from the Canary Islands to continents (Kvist *et al.* 2005).

The aims of this study were to: (i) quantify genetic differentiation within the blue tit complex, (ii) discern pathways and times of colonization and diversification among geographically structured groupings; and (iii) place this history within a framework of past climatic events. Despite the fact that the blue tit has been a popular study bird model for testing a plethora of evolutionary, physiological and ecological studies (see Cramp & Perrins 1993, and references there in), a global phylogeographic analysis has, until this point, not been undertaken. Here we apply phylogenetic and coalescent analyses to mitochondrial and nuclear sequence data, and microsatellite data to address these issues.

Materials and methods

Sample collection and molecular procedures

We collected feather and blood samples during 1997–2009 from 51 breeding populations throughout the

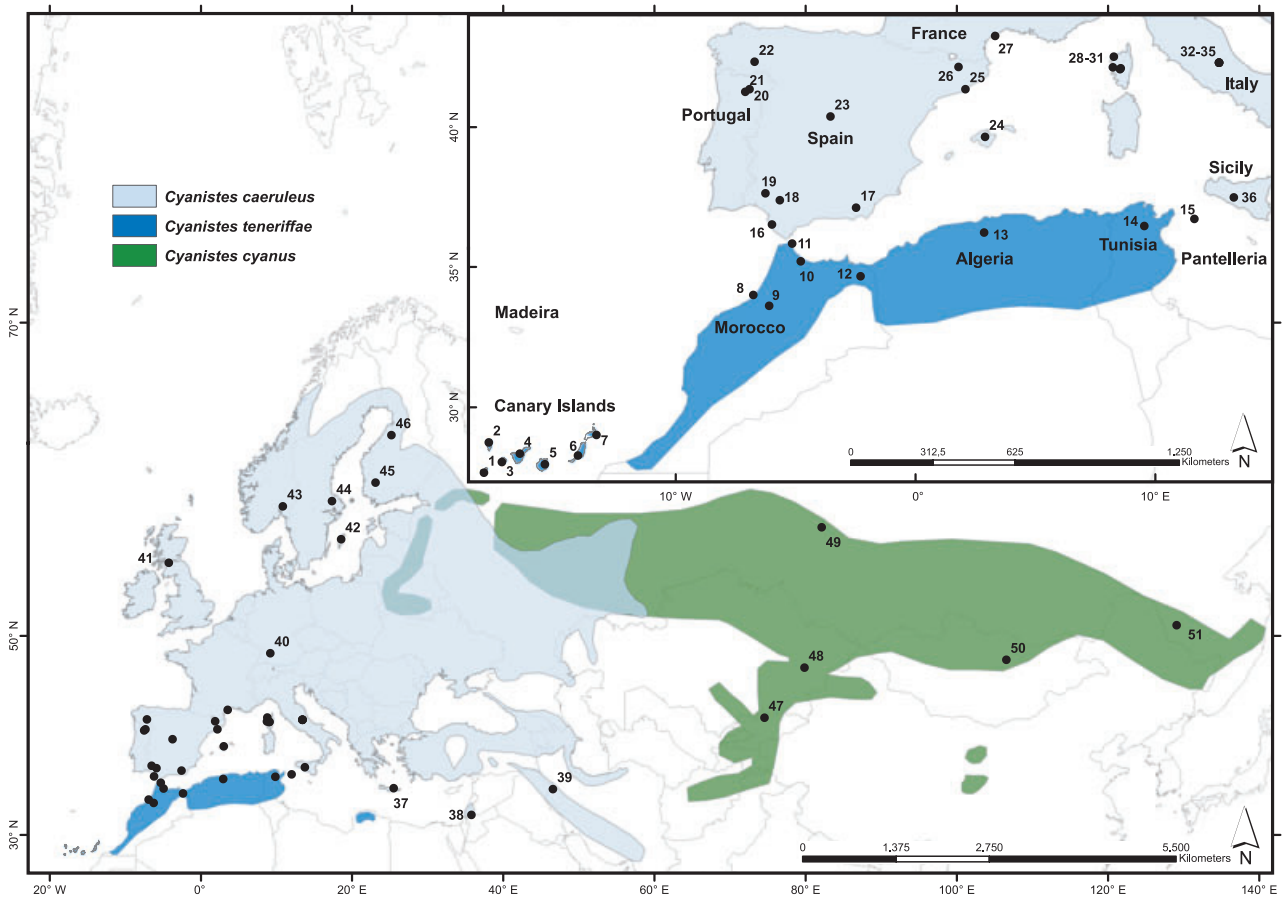

Fig. 1 Breeding distribution of the blue tit complex (i.e. *Cyanistes caeruleus*, *C. teneriffae* and *C. cyanus*). Localities (numbers) where tits were sampled are shown with dots. See S-Table 1 for further information.

species' breeding range (Supplementary Table S1, Fig. 1), with the exception of the most eastern parts such as Russian, Ukrainian and Byelorussian populations. Additionally, we obtained blood and feather samples from the azure tit (*Cyanistes cyanus*), great tit (*Parus major*), willow tit (*Poecile montanus*) and marsh tit (*Poecile palustris*), and four sequences of the 12884 nuclear gene of the Corsican finch (*Carduelis corsicana*) and greenfinch (*C. chloris*) available from Genbank as outgroups. DNA was extracted from blood and feathers using either standard phenol-chloroform method or the UltraClean™ BloodSpin kit (MO BIO Laboratories, Inc). The mitochondrial control region I and II was amplified and sequenced using the primers TL16700 and TH590 following the procedures and conditions in Kvist *et al.* (1999). We also amplified the nuclear marker locus 12884, located on chromosome 1 (Backström *et al.* 2008). PCR reactions were performed in a 50 μL volume containing around 50 ng of template DNA, 1.0 μM of each primer, 0.2 mM of each dNTP, 5 μL of 10 × PCR buffer (2.5 mM $MgCl_2$) and 1.0 unit of Biotools™ DNA polymerasa. The amplification profile was 95°C for 5 min followed by 40 cycles of 95°C for 30 s, 50°C for 30 s and 72°C for 1 min and a final extension in 72°C for 5 min. Sequencing reactions were performed for both strands using forward and reverse primers with Big Dye Terminator Cycle Sequencing Kit v. 2.0 and run with ABI 3730 automatic sequencer.

Canary, North African (Morocco and Ceuta populations) and Iberian Peninsula individuals were genotyped for six polymorphic microsatellites: PCA3; PCA7; PCA9; (Dawson *et al.* 2000); PMAC25 (Saladin *et al.* 2003); ESC 6 (Hanotte *et al.* 1994) and PK12 (Tanner SM, Richner H & Schuemperli D, unpublished data; GenBank accession no. AF041466). The PCR contained 1 μL of 10 × buffer (1.5 mM MgCl), 0.2 mM each dNTP, 0.2 μM of fluorescent-labelled forward primer, 0.2 μM of reverse primer, 1 unit of BioTools DNA polymerase in a final volume of 10 μL. PCR cycles started with an initial denaturing at 92°C for 3 min followed by 35 cycles of denaturing at 92°C for 30 s, with an annealing temperature varying from 50 to 65 °C, depending on the locus. Fragment size was determined using GeneScan-500 LiZ size standard in an ABI PRISM 3730 and scored with GeneMapper software version 3.7.

Sequence analyses

Sequences were aligned by eye using BioEdit version 7.0.9 (Hall 1999). Number of haplotypes, haplotype (h) and nucleotide (π) diversities with their standard deviations, and theta (θ) values were obtained with DnaSP version 5.10.01 (Librado & Rozas 2009). We inferred the nuclear haplotypes using the algorithm PHASE implemented in DnaSP. We used different methods implemented in the software Recombination Detection Program (RDP) version Beta 4.5 (Martin *et al.* 2010): RDP method (Martin & Rybicki 2000), GENECONV (Padidam *et al.* 1999), Maximum Chi-Square (Maynard Smith 1992), Chimaera (Posada & Crandall 2001) and 3Seq (Boni *et al.* 2007), in order to detect recombination in the 12884 nuclear marker. We calculated the degree of genetic differentiation among populations by computing pairwise Φ_{ST}s and performing molecular variance analyses (using the Tamura-Nei distances Kvist *et al.* 1999, 2004, 2005) with the program ARLEQUIN v. 3.5.1.2 (Excoffier & Lischer 2010). The analyses of molecular variance allowed us to partition the genetic variation among and within groups, and to obtain the significance of the different levels of genetic structure in the blue tit complex (Excoffier *et al.* 1992). We considered 26 populations within six groups (including the azure tit group), which represented geographic distribution and known genetic structure of the blue tit complex. The statistical significance was obtained after performing 10,000 random permutations of the data. We built statistical parsimony networks with the control region and 12884 sequences using the program TCS (Clement *et al.* 2000). We used the default limit of 5% implemented in the software.

Markov Chain Monte Carlo (MCMC) based Bayesian inferences were performed to infer the phylogenetic relationships among blue tit complex populations both mitochondrial and nuclear sequences using Mr. Bayes 3.1.2 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003). The best model of nucleotide substitution was chosen by using the Bayesian Information Criteria model selection implemented in the program jModelTest version 0.1.1 (Posada 2008). Mr. Bayes was run for 10,000,000 iterations, using one cold and three heated independent MCMCs. The temperature parameter for heating the chains was 0.5 for each run, and trees and model parameters were sampled every 1,000 generations. The convergence of the MCMCs was assessed with TRACER v. 1.5 (Rambaut & Drummond 2007). Trees produced before the stationary phase was reached were discarded with a burn-in of 2,500,000 generations. Two independent runs were performed in order to ensure the posterior probabilities were stable.

Gene flow and divergence time

Gene flow and incomplete lineage sorting may result in similar genetic signals, causing difficulties for the inference of evolutionary relationships among closely related taxa. We applied the Isolation and Migration coalescent model implemented in the IMa2 software (Hey 2010) to distinguish between both events. The program assumes no recombination in the loci analysed, and we tested this possibility with the nuclear gene using the software RDP. We performed a combined analysis using de mitochondrial (CR) and nuclear loci (12884) to detect gene flow. We performed coalescent analyses considering three populations, that is, one per species.

IMa2 excludes indels from the data set before beginning the analysis. Because it could remove both variation in the nuclear and mitochondrial gene, we changed each unique indel to a single nucleotide polymorphism (Turmelle *et al.* 2011). For mitochondrial CR we used a sequence divergence rate of 4% (±2%) between mitochondrial lineages per Ma estimated within genus *Cyanistes* (Päckert *et al.* 2007), which was converted to 1.016×10^{-5} (5.08×10^{-6} to 1.52×10^{-5}) substitutions/lineage/year by multiplying by the number of base pairs of the locus (508), and transforming from millions years to years. For the nuclear gene we used a standard mutation rate of 3.6×10^{-9} substitution/site/year estimated from 33 autosomal avian loci (Axelsson *et al.* 2004). We performed 10 independent runs of one million steps, with a burn-in period of 100,000 steps per run, using different seeds with 40 Metropolis coupled chains and medium heating parameters ($a = 0.975$, $b = 0.75$) under the HKY model of sequence evolution.

To estimate divergence times between and within clades we used the program BEAST (Drummond *et al.* 2006; Drummond & Rambaut 2007) and a multilocus data set of own sequences and others available from Genbank of three genes (cytochrome-*b*, CR and b-fibrinogen intron 7). In S-Table 4 Genbank accession numbers of each taxa and locus used are provided. A Yule process prior was implemented, assuming a constant rate of speciation. BEAST was run with 10 million generations, and we applied three partitions corresponding to the three gene fragments with different model settings. For the three marker data set the tree was calibrated at two nodes: (i) split between Nearctic and Palearctic *Poecile* species with a uniform prior range of 4.8 to 14.0 Ma assuming a possible Holarctic faunal interchange across a closed boreal forest belt until the Pliocene opening of the Bering Strait (Sanmartín *et al.* 2001; Gladenkov *et al.* 2002) and (ii) initial separation of Mediterranean *Periparus ater* (North Africa and Cyprus)

from continental clades with a uniform upper TMRCA prior set to the beginning of the Messinian salinity crisis (5.95 Ma). For details concerning the choice of calibration points and the respective age estimates for palaeogeographic events see Päckert *et al.* (2006, 2007). Results were examined using Tracer v1.04 (Rambaut & Drummond 2007) to evaluate stationarity, and the first three million generations were discarded as burn-in. Linearized consensus trees including posterior probabilities were inferred from the tree output files (concatenated sequence data sets) using TreeAnnotator v. 1.4.8 (as implemented in the BEAST package) with node heights set to 'mean'. We performed two independent runs with the cytochrome-*b* data set alone the first on without any fixed node age assigned but using an assumed fixed substitution rate of 0.0105 according to a recently re-evaluated estimate by Weir & Schluter (2008). In the second run no rate was fixed and the same node ages were used for calibration as described above.

Demography

To detect possible departures from a constant population size that could be interpreted as a result of a past demographic expansion we calculated the statistics F_s (Fu 1997), $R2$ (Ramos-Onsins & Rozas 2002) and Tajima's D (Tajima 1989). Coalescent simulations have shown that F_s and $R2$ tests are the best and most robust analyses for detecting population growth, with F_s being the best test for large sample sizes and $R2$ for small sample sizes (Ramos-Onsins & Rozas 2002). F_s test has also been shown to be the most powerful analysis for detecting expansions on non-recombining genes, whereas the use of Tajima's D and $R2$ is suggested when recombination levels are unknown (Ramírez-Soriano *et al.* 2008). Low positive values of $R2$ and significantly negative F_s values suggest demographic expansion. Positive values of D suggest a recently bottlenecked population or diversifying selection, and negative values provide evidence of a population expansion. All these statistics, both for mitochondrial and nuclear genes, were calculated with the software DnaSP 5.10.01, and significance was determined using the coalescent process implemented in DnaSP (1000 replicates).

Microsatellite analyses

We tested departures from Hardy–Weinberg equilibrium and linkage disequilibrium for each locus using the web version of GENEPOP (http://genepop.curtin.edu.au/) (Raymond & Rousset 1995; Rousset 2008). We quantified genetic diversity in each population and locus by estimating the observed and expected heterozygosity, mean number of alleles and number of private alleles using the excel microsatellite TOOLKIT v. 3.1.1 (Park 2001). In order to estimate genetic structure among populations within the blue tit complex, we estimated pairwise F_{ST} values with ARLEQUIN v. 3.5.1.2, and pairwise R_{HO} values (an unbiased estimator of Slatkin's R_{ST}; Slatkin 1995), using the program RST-CALC (Goodman 1997). In addition, we calculated Jost's D, which is an estimator of divergence based on the effective number of alleles rather than on the expected heterozygosity (Jost 2008) using SMOGD (Crawford 2010).

Results

Mitochondrial DNA

We obtained a fragment of 504 bp from the control region for 394 blue and Canary blue tits and 14 azure tits. Haplotype numbers, haplotype and nucleotide diversities and theta-values for each population are shown in Table 1. We also obtained one sequence from great tit, willow tit and marsh tit. Sequences corresponding to unique haplotypes have been deposited to the GenBank and accession numbers are listed in Supplementary Table S1. The nucleotide diversity (π) for all blue tit samples analysed was 0.03641, and the haplotype diversity ($\hat{h}$) was 0.9646 (Table 1).

The best-fit model of evolution determined by jModeltest was the Hasegawa Kishino Yano model including rate variation among sites (HKY + G). Bayesian inference revealed a topology supporting two geographically distinct clades: one Eurasian clade including all European and Western Asian populations plus the azure tit and the Canary-North-African clade (including all North African populations grouped as a terminal lineage within this clade) with high nodal support (Fig. 2, node A). The azure tit is firmly nested in the Eurasian clade with the maximum posterior probability ($P = 1.0$). Within *C. teneriffae* La Palma population represented a divergent lineage supported with a posterior probability of $P = 0.77$, and was sister to a strongly supported clade uniting the remaining *C. teneriffae* populations (node C, $P = 0.99$) with a posterior probability of $P = 0.92$ (node B). The Gran Canaria and El Hierro populations are reciprocally monophyletic with high posterior probability representing two independent lineages. Finally, the eastern Canary Islands (i.e. Fuerteventura and Lanzarote) were not grouped with the other Canary populations, but were firmly nested within a monophyletic group including all North African haplotypes and those from Pantelleria with high nodal support ($P = 1.0$, node D).

We treated the large gaps found in the statistical parsimony networks as 5th state and the whole gap as one

Table 1 Sample sizes (N) and diversity estimates for the study populations of the blue tit. Haplotype (H) and nucleotide (π) diversities are given with their standard deviations (SD)

	Control region							Nuclear gene (12884)						
Populations	*N*	*K*	*H*	SD	π	SD	θ	*N*	*K*	*H*	SD	π	SD	θ
Total	394	125	0.9646	0.0040	0.03641	0.00027	0.03587	196	57	0.906	0.008	0.016	0.00043	0.02012
Nothern and Central Europe	45	16	0.668	0.081	0.00265	0.00053	0.00778	24	14	0.836	0.040	0.00535	0.00075	0.00841
Nordic Countries	30	9	0.600	0.103	0.00157	0.00037	0.00454	9	6	0.778	0.082	0.00345	0.00067	0.00388
Germany	8	3	0.464	0.200	0.00100	0.00047	0.00154	8	8	0.900	0.046	0.00658	0.00131	0.00804
Scotland	7	6	0.952	0.096	0.00514	0.00095	0.00490	7	7	0.846	0.074	0.00583	0.00126	0.00671
Southern Europe	127	44	0.933	0.012	0.01060	0.00027	0.01228	32	14	0.805	0.039	0.00436	0.00044	0.00780
Spain	43	18	0.847	0.049	0.01087	0.00075	0.00973	9	7	0.882	0.039	0.00601	0.00092	0.00698
Mallorca	11	3	0.636	0.089	0.00146	0.00031	0.00137	3	4	0.800	0.172	0.00338	0.00085	0.00350
France	15	6	0.705	0.114	0.00202	0.00053	0.00369	4	4	0.643	0.184	0.00267	0.00104	0.00411
Corsica	33	11	0.902	0.024	0.00794	0.00091	0.00690	7	4	0.648	0.116	0.00399	0.00077	0.00335
Italy	11	11	1.000	0.039	0.00982	0.00187	0.01161	1	2	1.000	0.500	0.00267	0.00133	0.00267
Sicily	7	3	0.667	0.160	0.00400	0.00138	0.00408	9	7	0.784	0.085	0.00322	0.00063	0.00465
Greece	7	3	0.667	0.160	0.00229	0.00088	0.00246	5	1	0.000	0.000	0.00000	0.00000	0.00000
Western Asia	20	9	0.789	0.086	0.00485	0.00096	0.00676	10	7	0.811	0.069	0.00411	0.00044	0.00451
Iran	9	6	0.889	0.091	0.00522	0.00131	0.00662	3	3	0.733	0.155	0.00302	0.00066	0.00234
Jordan	11	4	0.491	0.175	0.00305	0.00136	0.00410	7	5	0.670	0.126	0.00346	0.00062	0.00419
North Africa	96	21	0.804	0.035	0.00601	0.00054	0.00959	51	20	0.805	0.028	0.00953	0.00127	0.01293
Morocco	22	6	0.797	0.052	0.00480	0.00073	0.00338	10	4	0.605	0.101	0.01023	0.00301	0.00909
Algeria	16	8	0.825	0.076	0,00281	0.00055	0.00433	3	5	0.933	0.122	0.00693	0.00156	0.00701
Pantelleria	18	6	0.490	0.142	0.00160	0.00059	0.00418	8	4	0.525	0.137	0.00158	0.00049	0.00241
Fuerteventura	20	5	0.716	0.069	0.00411	0.00058	0.00347	16	9	0.817	0.051	0.01299	0.00194	0.01068
Lanzarote	20	1	0.000	0.000	0.00000	0.00000	0.00000	14	3	0.204	0.098	0.00075	0.00040	0.00206
Canary Islands	106	40	0.924	0.014	0.02156	0.00113	0.02382	73	13	0.754	0.023	0.01628	0.00031	0.01113
Gran Canaria	22	9	0.861	0.044	0.00449	0.00063	0.00508	15	2	0.129	0.079	0.00034	0.00021	0.00067
Tenerife	22	14	0.926	0.039	0.00672	0.00113	0.01018	16	3	0.179	0.088	0.00049	0.00025	0.00132
La Gomera	20	6	0.447	0.137	0.00123	0.00046	0.00347	17	2	0.299	0.085	0.00077	0.00022	0.00063
La Palma	22	10	0.814	0.064	0.00389	0.00061	0.00550	17	4	0.619	0.056	0.00192	0.00027	0.00189
El Hierro	20	5	0.368	0.135	0.00174	0.00071	0.00296	8	4	0.442	0.145	0.00305	0.00111	0.00544
Cyanistes cyanus	14	10	0.945	0.045	0.00760	0.00122	0.00818	11	2	0.173	0.101	0.00045	0.00026	0.00071

K number of haplotypes; θ theta.

event. Geographical association of haplotypes visualized with the parsimonious network identified a similar structure than the Bayesian inference. Interestingly, the link between the Eurasian clade and the Canary-North African clade suggests a connection through La Palma haplotypes (Fig. 3). La Palma individuals share common substitutions with the other Canary blue tits, but they also have an indel of 12 bp, which is shared with the Eurasian clade, and which is not present in the haplotypes in the rest of the Canary-North-African clade.

Molecular variance analysis showed significant genetic variation among the six groups (Φ_{ST} = 0.92, $P < 0.0001$; pairwise Φ_{ST} values between populations are shown in S-Table 2), among populations within groups (Φ_{SC} = 0.71, $P < 0.0001$) and within populations among groups (Φ_{CT} = 0.73, $P < 0.0001$). Most variation was attributable among groups (72.85%) and among populations within groups (19.24), but also a small amount of the variation was attributable within populations (7.91%).

Nuclear DNA

12884 gene. We obtained sequences from 196 blue and Canary blue tits, 11 azure tits, three willow tits and two great tits. The nuclear locus yielded a fragment of 375–389 bp (depending on the number and length of indels). Recombination was not detected by any of the tests performed. The overall nucleotide diversity (π = 0.016) and haplotype diversity ($\hat{h}$ = 0.906) values were lower than those obtained with the mitochondrial sequence. The highest diversity values were obtained from Fuerteventura (π = 0.01299; $\hat{h}$ = 0.817) and Morocco (π = 0.01023; $\hat{h}$ = 0.605). All diversity values for the study populations are shown in Table 1.

jModelTest selected the Hasegawa Kishino Yano model (HKY + G) as the best model of evolution for the

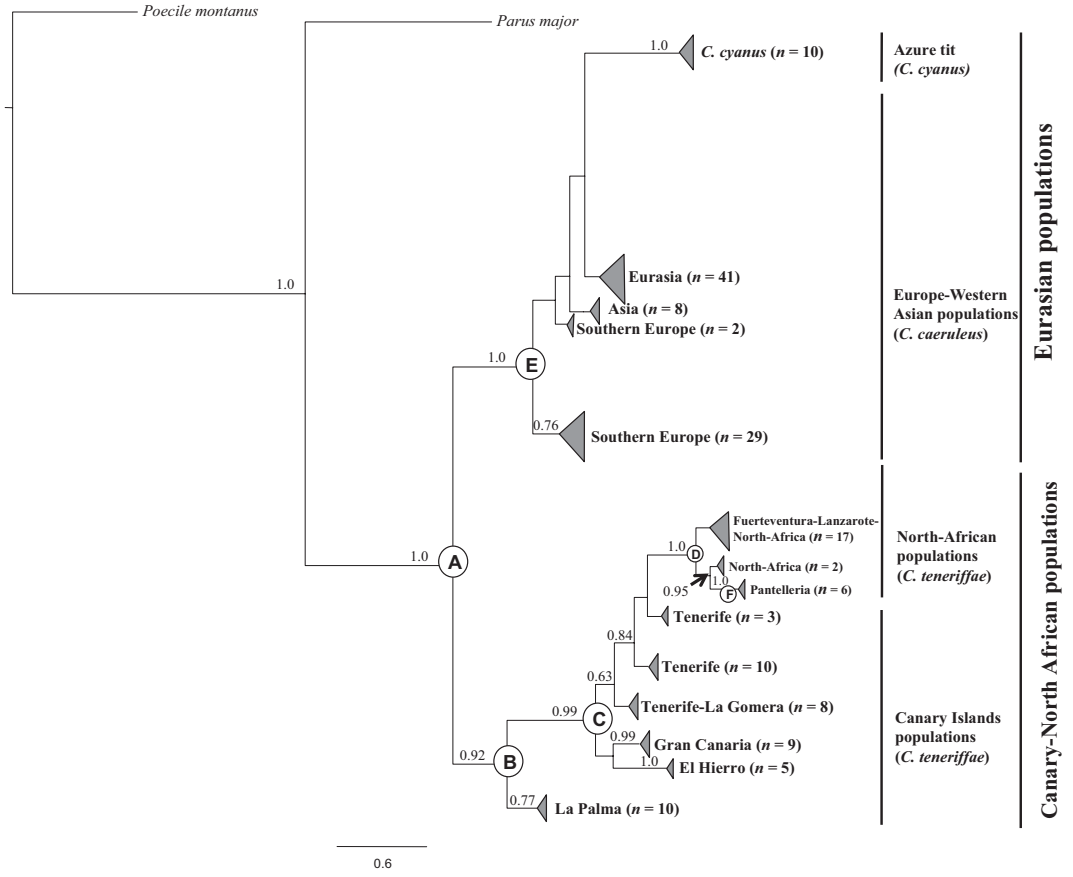

Fig. 2 Bayesian consensus tree based on mtDNA (control region) sequences. Numbers above branches are Bayesian posterior probabilities. Triangle areas represent the rough number of haplotypes included in each branch. Precise number of haplotypes is shown in brackets.

12284 marker. The phylogenetic tree shows a monophyletic relationship in the genus *Cyanistes* but unlike in the mitochondrial tree *C. teneriffae* does not result as a monophyletic unit (Fig. 4, nodes B and C). Despite the nuclear tree resolved two sister clades (nodes B and C) in the blue tit complex, there are a number of topological conflicts between the nuclear and mitochondrial tree. Thus, from the Canarian clade shown in the control region tree (Fig. 2, node B), only the western Canary Islands (i.e. La Palma, El Hierro and La Gomera) grouped together. This clade included also two Moroccan and four Fuerteventura individuals (node D. Fig. 4) and was sister to the azure tits with a posterior probability of $P = 0.93$ (node C, Fig. 4). The remaining individuals and populations of *C. teneriffae* and *C. caeruleus* formed a separate clade with maximum posterior probability support, but with internal unresolved phylogenetic relationships (node B). Within this clade most European haplotypes grouped together, but two European individuals grouped with the North African populations and the remaining Canary Island haplotypes with the highest posterior probability support (node E). The phylogenetic relationships among the latter two groups and a further Eurasian linage remained unresolved in the nuclear tree.

The nuclear parsimonious network shows the same topology as the Bayesian phylogenetic tree. Two informative indels (the most notable was a 16 bp indel) divided the blue tit complex in two big groups as with the Bayesian tree. Within the Eurasian and North African group it is still possible to distinguish two groups: one Eurasian and other Canarian (except La Palma, El Hierro and La Gomera islands)-North African group (Fig. 5).

Molecular variance analysis of the 12884 nuclear gene showed again significant genetic variation among the six groups ($\Phi_{ST} = 0.86$, $P < 0.0001$; albeit the number of non significant pairwise values was higher in the nuclear than in the mitochondrial gene, S-Table 3), among populations within groups ($\Phi_{SC} = 0.79$, $P < 0.0001$) and within populations among groups ($\Phi_{CT} = 0.35$, $P < 0.05$). However, the amount of variation accounted by each association changed. Thus, the most variation was attributable to among populations within

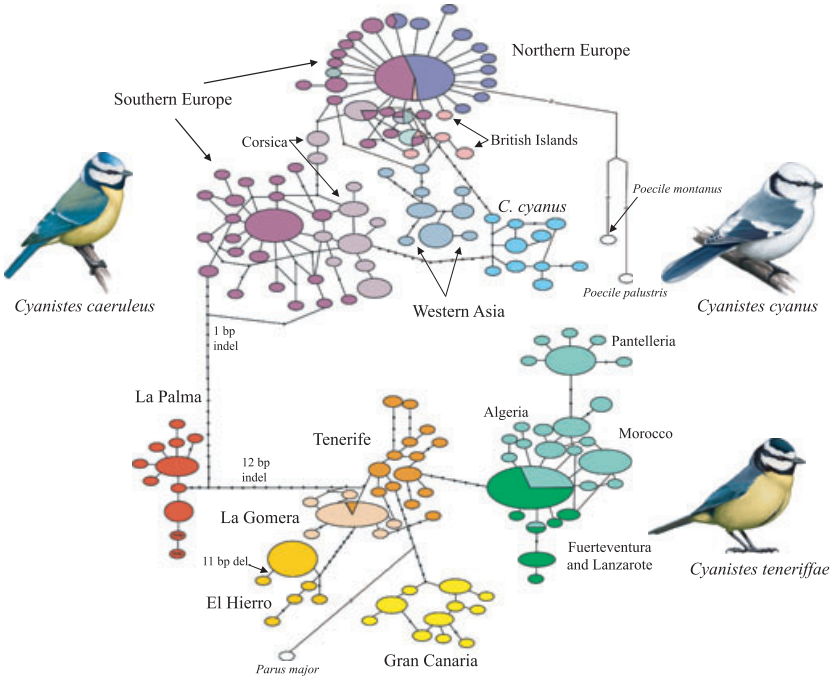

Fig. 3 Minimum spanning network of the blue tit complex (i.e. *Cyanistes caeruleus*, *C. teneriffae* and *C. cyanus*) based on mtDNA (control region) sequences. Ellipses represent haplotypes whose areas are proportional to the number of individuals bearing these haplotypes. Numbers of substitutions are depicted with bars, except connections with outgroups (ellipses without colours, *Parus major*, *Poecile palustris* and *P. montanus*) that due to the high number of substitutions (>50) are denoted with a double backslash. Colours indicate either sampled geographical areas in the blue and Canary tits or the azure tit species.

groups (51.62), more than variation among groups (34.92%) and variation within populations (13.46%).

Microsatellites. A total of 292 blue and Canary blue tits from nine populations; the seven Canary Islands plus North Africa (Morocco and Ceuta) and the Iberian Peninsula were genotyped. Of the six loci used, three (PCA3, Esc6, PCA9) showed departure from Hardy–Weinberg equilibrium after controlling for multiple comparisons. After excluding these loci from the analyses, results did not significantly change; therefore we used all six loci throughout genetic analyses in order to maximize the statistical power of tests. Genotypic linkage disequilibrium analyses did not detect significant linkage disequilibrium.

The highest number of alleles and number of private alleles per locus was found in the Iberian Peninsula (except in locus Esc6). Within the Canary Islands, Tenerife supported the highest number of alleles, but La Palma and La Gomera showed the highest number of private alleles (Table 3). All pairwise comparison using F_{ST}, D_{EST} and R_{HO} showed significant differences among populations ($P < 0.001$), after correction for multiple tests. The nine populations analysed showed a high genetic differentiation ($F_{ST} = 0.35$) with the lowest pairwise F_{ST} value obtained between Tenerife and Gran Canaria ($F_{ST} = 0.12$) and the highest between El Hierro and La Palma ($F_{ST} = 0.59$). A similar pattern was found with the pairwise Jost' D values ($D_{EST} = 0.73$, with a range $0.99 < D_{EST} > 0.12$), and with the R_{HO} values ($0.65 < R_{HO} > 0.12$) (Table 4).

Population demography

Mitochondrial and nuclear data sets from the three *Cyanistes* species showed significant deviation from neutrality according to the Fu's F_s tests suggesting a past demographic expansion event (Table 2). With the mitochondrial gene this result is consistent in the European groups of *C. caeruleus* analysed, however, in the Western-Asian group Fu's F_s value, albeit negative, was not significant. Results for Fu's F_s tests in the two *C. teneriffae* groups were also significant. The nuclear sequences showed a significant departure from the neutrality (significant negative values of Fu's F_s tests) in the European groups of *C. caeruleus*, and in the Tenerife population in *C. teneriffae* (Table 2). Tajima's *D* tests showed significant values with the mitochondrial gene in the Nordic countries and France within *C. caeruleus*, and in two *C. teneriffae* groups (Pantelleria and La Gomera). Ramons-Onsins and Rozas' R_2 tests reached similar results that Tajima's *D* tests although with three

Table 2 Ramons-Onsins and Rozas R_2, Fu's F_S, and Tajima's D values for the study populations

	Control Region			Nuclear gene 12884		
Populations	R_2	Fu's F_S	Tajima's D	R_2	Fu's F_S	Tajima's D
Total	0.089	**−53.527**	0.044	0.064	**−23.273**	−0.664
Nothern and Central Europe	**0.038**	**−13.270**	**−2.096**	0.073	**−6.015**	−1.112
Nordic Countries	**0.057**	**−6.684**	**−2.032**	0.127	−1.678	−0.347
Germany	0.216	−0.999	−1.310	0.116	−2.120	−0.677
Scotland	0.183	**−2.613**	0.594	0.139	−1.801	−0.493
Southern Europe	0.077	**−24.362**	−0.410	0.059	**−6.320**	−1.298
Spain	0.124	−4.070	0.205	0.122	−1.072	−0.490
Mallorca	0.198	−0.046	0.222	0.186	−1.350	−0.185
France	**0.103**	**−2.782**	**−1.585**	0.177	−1.236	**−1.535**
Corsica	0.137	−0.681	0.258	0.181	0.480	0.620
Italy	0.122	**−7.447**	−0.719	0.500	n.a.	n.a.
Sicily	0.239	0.668	−0.597	0.117	**−3.094**	−1.010
Greece	0.252	0.263	−0.302	n.a.	n.a.	0.000
Western Asia	0.098	−2.536	−1.015	0.147	−2.023	−0.281
Iran	0.148	−1.496	−0.973	0.283	0.020	1.392
Jordan	**0.131**	0.164	−1.007	0.154	−0.986	−0.607
North Africa	0.050	**−7.411**	−1.115	0.073	−4.399	−0.782
Morocco	0.195	0.346	1.302	0.155	4.205	0.449
Algeria	**0.087**	**−3.300**	−1.072	0.184	−1.565	−0.060
Pantelleria	**0.097**	**−3.256**	**−2.096**	**0.112**	−1.415	−1.002
Fuerteventura	0.162	0.689	0.582	0.159	1.211	0.725
Lanzarote	n.a.	n.a.	0.000	0.126	−1.059	**−1.527**
Canary Islands	0.093	**−8.140**	−0.303	0.143	4.106	1.308
Gran Canaria	0.112	**−3.976**	−0.068	0.064	−0.439	−0.764
Tenerife	**0.081**	**−6.681**	−1.140	0.092	**−1.747**	**−1.267**
La Gomera	**0.093**	**−3.952**	**−2.056**	0.150	0.785	0.336
La Palma	0.091	**−2.707**	−1.142	0.127	−0.198	0.030
El Hierro	0.109	−0.016	−1.260	0.149	0.120	**−1.547**
Cyanistes cyanus	0.138	**−3.560**	−0.281	**0.087**	−0.176	−0.641

n.a. = not applicable; significant values are denoted in bold.

more populations being significant (Tenerife and Algeria within *C. teneriffae* and Jordan within *C. caeruleus*). Tajima's *D* and Ramons-Onsins and Rozas' R_2 values were not significant with *C. cyanus*. Finally, with the nuclear gene Tajima's *D* tests showed significant values in one *C. caeruleus* population (France), and three in *C. teneriffae* (Lanzarote, Tenerife and El Hierro). Ramons-Onsins and Rozas' R_2 tests were only significant in Pantelleria (*C. teneriffae*, Table 2). *C. cyanus* reached significant values in Ramons-Onsins and Rozas' test, however, although negative was not significantly different from zero in Tajima's *D* test.

Coalescent analyses and divergence times

Coalescent analyses provided non-zero estimates of migration parameters among the three populations analysed. Significant migration rates were produced between *C. caeruleus* and *C. teneriffae* (LLR tests ≥ 5.14), but analyses did not detect significant migration between *C. cyanus* and the other two species. This result was consistent in the ten replicates performed suggesting a scenario of gene flow between *C. caeruleus* and *C. teneriffae* after their divergence more than a situation of incomplete lineage sorting.

The cyt-*b* and multilocus calibrations showed similar results estimating a basal split in the blue tit complex being dated to the early Pliocene, that is, slightly younger or older than 5 Ma depending on the dating approach performed (Fig. 6). 95% highest posterior density intervals from two independent runs with the cyt-*b* data set (fixed rate and fixed node ages) largely overlapped at most nodes. The Canary-North-African clade diverged approximately 3.1 Ma into a linage from La Palma and another lineage formed by the remaining Canary Islands and the North African populations (means: 3.12 for cyt-*b* data set with fixed rate and 3.18 for three marker data set with fixed node ages). The

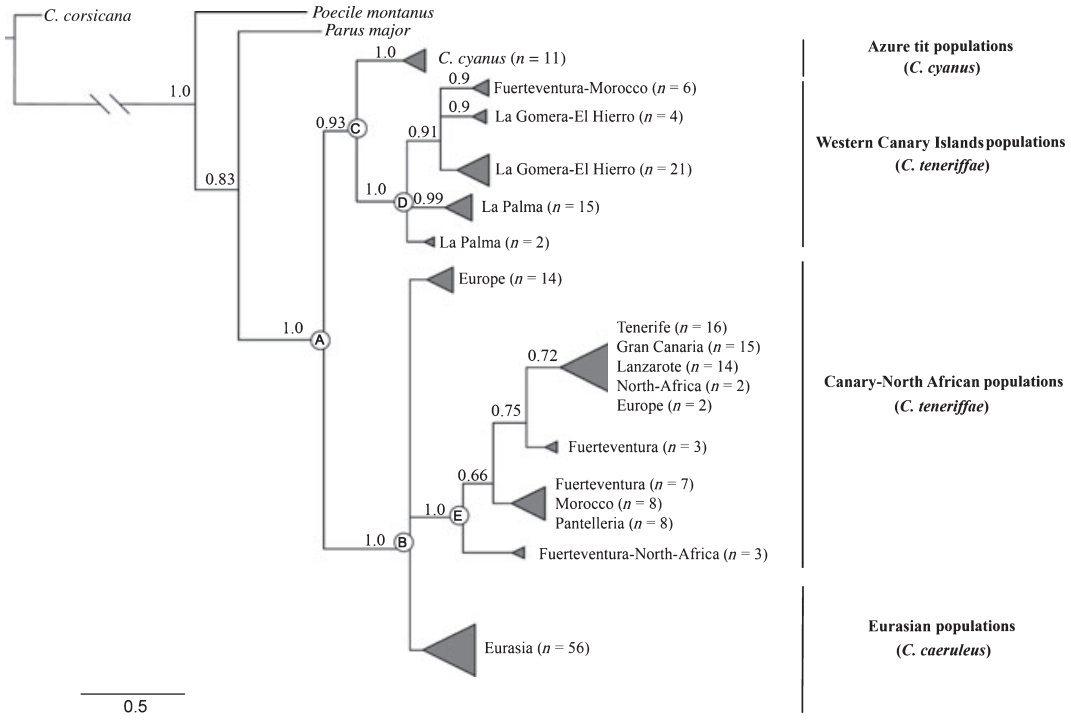

Fig. 4 Bayesian consensus tree based on nDNA (12884) sequences. Numbers above branches are Bayesian posterior probabilities. Triangle areas represent the rough number of haplotypes included in each branch. Precise number of haplotypes is shown in brackets.

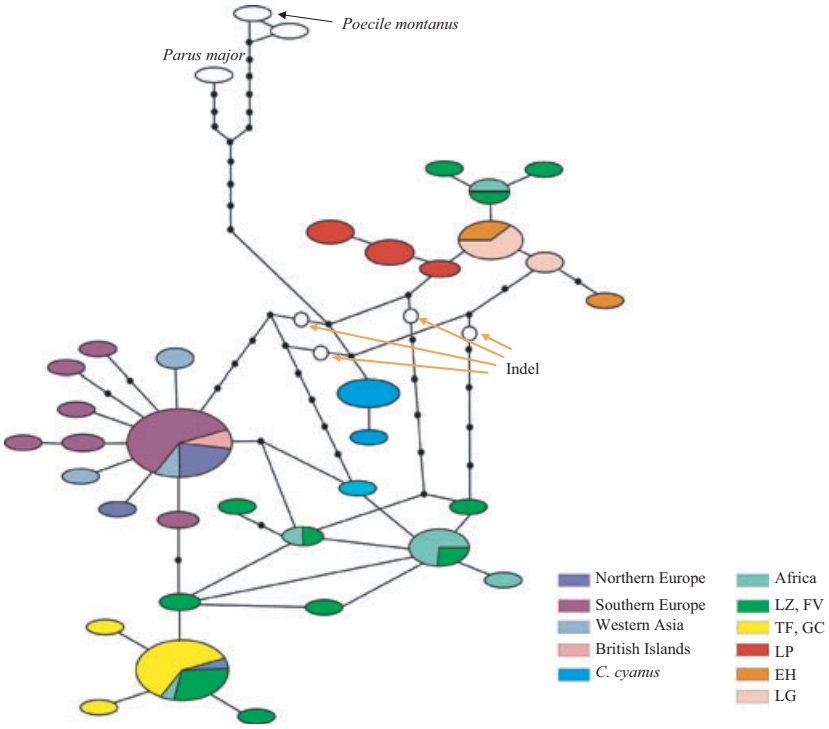

Fig. 5 Minimum spanning network of the blue tit complex (i.e. *Cyanistes caeruleus*, *C. teneriffae* and *C. cyanus*) based on nDNA (12884) sequences. Ellipses represent haplotypes whose areas are proportional to the number of individuals bearing these haplotypes. Numbers of substitutions are depicted with bars. Main indels are indicated with white circles.

Table 3 Sample sizes (N), mean number of alleles (*a*) and observed (H_O) and expected (H_E) heterozygosities from microsatellite data estimated for each population (standard deviations in parentheses). Allele numbers (*A*), number of private alleles (*P*) and observed and expected heterozygosities are given also for each locus. Initials of first row correspond with loci used (see text for further details)

					PCA7				Esc6				Pk12				PmC25				Pca3				Pca9			
	N	*a* (±SD)	*Ho* (±SD)	*He* (±SD)	*A*	*P*	*Ho*	*He*	*A*	*P*	*Ho*	*He*	*A*	*P*	*Ho*	*He*	*A*	*P*	*Ho*	*He*	*A*	*P*	*Ho*	*He*	*A*	*P*	*Ho*	*He*
Canary Islands																												
El Hierro	34	3.67 ± 1.51	0.40 ± 0.03	0.37 ± 0.11	5	0	0.12	0.14	3	0	0.23	0.28	1	0	0	0	4	0	0.59	0.54	4	1	0.65	0.55	5	0	0.82	0.72
Fuerteventura	28	4.00 ± 1.10	0.45 ± 0.04	0.56 ± 0.05	5	1	0.61	0.61	3	0	0.11	0.50	3	0	0.36	0.49	5	0	0.61	0.58	3	0	0.36	0.41	5	0	0.68	0.77
Gran Canaria	35	5.50 ± 3.89	0.46 ± 0.03	0.55 ± 0.16	2	0	0.17	0.16	10	2	0.91	0.86	1	0	0	0	4	0	0.37	0.58	6	1	0.48	0.82	10	0	0.8	0.88
La Gomera	36	3.67 ± 2.16	0.44 ± 0.03	0.43 ± 0.13	5	1	0.55	0.48	3	0	0.55	0.57	2	0	0.05	0.05	1	0	0	0	4	3	0.64	0.67	7	0	0.86	0.81
La Palma	29	3.83 ± 1.72	0.48 ± 0.04	0.44 ± 0.10	3	0	0.55	0.67	6	2	0.62	0.65	3	2	0.07	0.07	2	0	0.55	0.48	3	0	0.21	0.19	6	0	0.90	0.61
Lanzarote	22	2.83 ± 0.98	0.43 ± 0.04	0.50 ± 0.07	4	0	0.68	0.59	2	0	0.59	0.50	2	0	0.24	0.21	3	0	0.45	0.64	4	0	0.18	0.68	2	0	0.41	0.38
Tenerife	47	7.17 ± 1.94	0.67 ± 0.03	0.68 ± 0.08	6	0	0.34	0.32	7	0	0.85	0.76	5	0	0.64	0.58	6	0	0.72	0.68	10	2	0.83	0.86	9	1	0.66	0.85
Iberian Peninsula																												
Spain	30	11.17 ± 4.62	0.65 ± 0.04	0.78 ± 0.06	12	6	0.90	0.91	4	1	0.01	0.53	15	8	0.77	0.87	9	1	0.90	0.83	17	5	0.77	0.90	10	2	0.50	0.67
North Africa																												
Morocco	31	8.83 ± 3.71	0.75 ± 0.03	0.81 ± 0.02	5	0	0.64	0.75	6	0	0.77	0.79	7	2	0.83	0.83	11	2	0.90	0.88	15	2	0.68	0.78	9	0	0.64	0.83

divergence of the North African populations from eastern Canarian populations was estimated to have commenced approximately 0.15–0.09 Ma (node D, means for cyt-*b* and three marker data sets, respectively). Within the Eurasian clade, the azure tits appear to have split from their common ancestors with European blue tit populations around 1.5–2.1 Ma.

Discussion

Discrepancies between mitochondrial and nuclear gene trees suggest a less than simple history within the genus *Cyanistes* around the Mediterranean Sea, with several phases of colonization resulting in overlapping ranges. Our results reveal a novel role for the Canary Islands, with North African blue tits being derived from *C. teneriffae*, an island to continental colonization that contrast to the typically held view of islands being end points of colonization processes (but see, Bellemain & Ricklefs 2008; Sheldon *et al.* 2009).

mtDNA phylogeography

Our phylogenetic results describe long-standing isolation and divergence between the Eurasian and the Canary-North African clades in allopatry, after the timing of which is consistent with the refilling of the Mediterranean, at the end of the Messinian salinity crisis, approximately 5.33 Ma (Garcia-Castellanos *et al.* 2009). Despite their capacity for dispersal, the complete mitochondrial lineage sorting found between these two clades and among several Canary and North African lineages suggest that both the Mediterranean Sea and the Atlantic Ocean surrounding the Canary Islands have formed effective barriers for blue and Canary blue tit populations, facilitating their diversification and speciation. AMOVA and parsimony network analyses support this interpretation. Previous phylogenetic studies have highlighted an important role for Pleistocene glacial events on the genetic structure of the European blue tit (Kvist *et al.* 1999, 2004), a feature documented in many other taxa with similar distributions (Hewitt 1999, 2000). Fu's F_s values and the higher haplotype and nucleotide diversity found in the southern Europe compared to northern and central Europe (Tables 1 and 2) support a past demographic expansion event, and are compatible with one or more southern European glacial refugia with subsequent range expansion to northern and central Europe, and further east (Kvist *et al.* 1999, 2004; Khoury *et al.* 2007).

Mitochondrial lineage divergence among the Canary Islands and North African populations provide evidence of long and independent evolutionary histories

Table 4 F_{ST} and Jost's D values (below the diagonal) and R_{HO} (above diagonal) estimated from the microsatellite. All pairwise values were significant ($P < 0.0001$). EH, El Hierro; LG, La Gomera; LP, La Palma; GC, Gran Canaria; TF, Tenerife; Fuerteventura, FV; LZ, Lanzarote; MO, Morocco; IP, Iberian Peninsula (Spain)

Canary Islands								Morocco	Spain
	EH	LG	LP	GC	TF	FV	LZ	MO	IP
EH	–	0.850	0.840	0.613	0.221	0.734	0.833	0.529	0.784
LG	0.565 (0.726)	–	0.746	0.646	0.711	0.737	0.884	0.639	0.755
LP	0.590 (0.989)	0.462 (0.556)	–	0.650	0.711	0.751	0.878	0.659	0.734
GC	0.503 (0.690)	0.292 (0.280)	0.372 (0.548)	–	0.243	0.625	0.789	0.485	0.752
TF	0.333 (0.452)	0.268 (0.316)	0.358 (0.697)	0.120 (0.196)	–	0.584	0.691	0.406	0.735
FV	0.530 (0.951)	0.441 (0.575)	0.475 (0.872)	0.335 (0.444)	0.304 (0.660)	–	0.211	0.122	0.513
LZ	0.551 (0.889)	0.465 (0.569)	0.506 (0.874)	0.366 (0.519)	0.327 (0.687)	0.186 (0.102)	–	0.214	0.618
MO	0.363 (0.798)	0.344 (0.651)	0.335 (0.743)	0.264 (0.615)	0.189 (0.602)	0.138 (0.253)	0.199 (0.412)	–	0.411
IP	0.418 (0.973)	0.360 (0.884)	0.353 (0.848)	0.303 (0.873)	0.243 (0.903)	0.285 (0.822)	0.294 (0.636)	0.155 (0.747)	–

and complement previous work describing strong phylogeographic structure within the Canary Islands (Kvist *et al.* 2005). Fu's F_s values and the higher haplotype and nucleotide diversity found on the Canary Islands, compared to North African populations (Tables 1 and 2), support a past demographic expansion event with Canary Islands origin for this clade (Fig. 2).

The mitochondrial analysis suggests that La Palma represents a distinct lineage from the other Canary Islands, a result supported by the multilocus analysis performed with BEAST. The TMRCA of La Palma and the remaining Canary blue tits of the Canary-North African clade was estimated to be approximately 3.1 Ma, with the most recent divergence event being that between the eastern Canarian and North African populations during the late Pleistocene. The broad pattern of diversification is thus one of colonization and diversification from west to east. This is a surprising result and contrasts with other phylogeographical studies published with vertebrates in Macaronesia, which present continental origins and east to west patterns colonization within the Canary Islands (Marshall & Baker 1999; Juste *et al.* 2004; Cox *et al.* 2010).

Although the island of La Palma is clearly divergent, the remaining Canary Islands also indicate substructure among. The parsimony network and the Bayesian phylogenetic tree reveal three largely allopatric lineages within the Canary Islands: El Hierro, Gran Canaria and a group composed by Tenerife-La Gomera (Figs 2 and 3). The allopatric lineage found in Gran Canaria, but also detected in other avian species such as robins (Dietzen *et al.* 2003) and chaffinches (Suárez *et al.* 2009), indicating that a lack of morphological differentiation within this island obscures the evolutionary uniqueness of some of its avian fauna. The lack of monophyly found within the mitochondrial sequences of Tenerife and La Gomera indicate either very recent isolation, and thus incomplete lineage sorting, or limited gene flow between these two close islands, or perhaps some combination of the two. However, microsatellite data indicate high population differentiation (Table 4) between these two islands supporting a recent differentiation, with limited gene flow, supporting the hypothesis of incomplete lineage sorting within the mtDNA sequence data, rather than gene flow.

The North African lineage (Fig. 2, node D) includes haplotypes from North Africa and eastern Canary Islands (Fuerteventura and Lanzarote). The TMCRA of this group was estimated to be a very recent event (90,000–150,000 years). The presence of North African haplotypes in each of the eastern islands suggests either a very recent colonization to the mainland from these islands, a back colonization from the continent or extensive gene flow between Africa and the eastern islands, which are located less than 100 km from the continent (Fig. 1). A similar genetic pattern has also been found between the eastern island populations of the houbara bustard (*Chlamydotis undulate fuertaventurae*) and North African populations (*C. u. undulata*) (Idaghdour *et al.* 2004). It was suggested that colonization of the Canary Islands by North African houbaras occurred during the last glacial maximum, followed by a subsequent secondary contact with a final period of isolation. Although we cannot discard any colonization hypotheses, microsatellite data for Cyanistes reveals significant population differentiation between North Africa and the eastern islands, as well as among Islands of the Canary archipelago (Table 4) indicating limited contemporary gene flow.

MtDNA data also provided evidence also for differentiation of the North African Pantelleria population (Fig. 2, node F; and Fig. 3). Pantelleria is a volcanic island, with the older parts dated to mid Pleistocene (≈300,000 years, Mahood & Hildreth 1986; Civetta *et al.* 1984). With an area of 83 km^2 Pantelleria is located in the strait of Sicily (Mediterranean Sea) between the coasts of Sicily (100 km) and Tunisia (70 km) (Fig. 1). Mitochon-

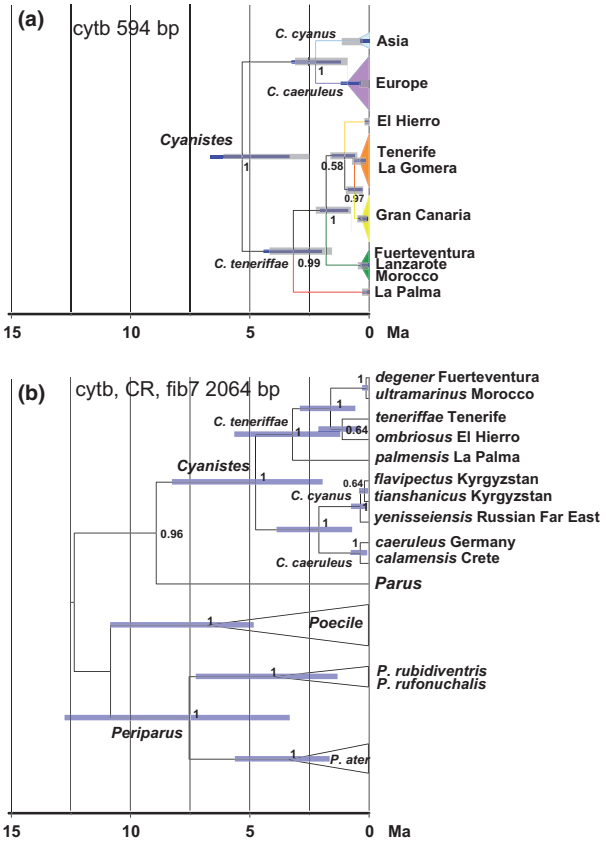

Fig. 6 Single marker (cytochrome b) and multilocus (cytochrome b, control region and b-fibrinogen intron 7) dated phylogenies reconstructed with BEAST; Markov chain length = 10,000,000 generations, tree prior = speciation (Yule process), relaxed uncorrelated lognormal clock model; bars indicate 95% highest posterior density (HPD) intervals, for cytochrome b: blue bars = dating based on fixed substitution rate, grey bars = dating based on fixed node ages (uniform TMRCA prior distribution assigned to two nodes basal *Poecile* and *Periparus ater*).

drial haplotypes from Pantelleria are unambiguously African, despite its vicinity to Sicily, where only Eurasian haplotypes are present (Figs 2 and 3), suggesting a recent colonization from the African continent. No significant mitochondrial genetic structure was found among North African continental populations, thus we cannot infer the specific origin of the Pantelleria populations. However, it seems plausible to hypothesize that the neighbouring Tunisian coast may have served as the source of this colonization, as has been shown for the greater white-toothed shrew (*Crocidura russula;* Cosson *et al.* 2005).

Phylogenetic history revealed by nuclear and mitochondrial data

Nuclear sequences are expected to record older demographic events than mitochondrial DNA due to the slower lineage sorting, influenced by lower mutation rate and larger effective population size for nuclear DNA. Indeed, these different properties of nuclear and mtDNA sequences can prove advantageous for gaining a more general understanding of intraspecific history, as they may provide complementary insights into phylogeographic patterns and the processes underlying these (Edwards & Bensch 2009). Despite general agreement between mtDNA and nuclear sequences regarding the monophyly of the blue tit complex (Fig. 4, node A), our data reveal several instances of genetic admixture. The first of these is between three Canary populations and the azure tit, within the Canary-North African clade (node C). There is also evidence for admixture between the remaining Canary Islands and North African populations, and the European-Western Asian populations. Differences in the rate of lineage sorting between the mitochondrial and nuclear loci cannot explain this discrepancy alone, because all azure tits, which were unambiguously grouped in the mitochondrial analysis with the Eurasian blue tits, are nested in the nuclear tree with the three western Canary Islands, and several individuals from Fuerteventura and Morocco (node C). Alternative hypotheses to explain these patterns are: (i) male mediated gene flow results in the mixing of nuclear sequences, with no consequence for structure within the maternally inherited mtDNA genome; (ii) non-sexed biased gene flow, followed by selective sweeps or drift related with a significant decrease in female effective population size, resulting in differentiation of maternally inherited loci but not in biparentally inherited markers; or (iii) haplotypes of the western Canary Islands and the azure tits represent the retention of an ancient polymorphism from the ancestor of the blue and azure tit. The disjunct distributions of the Canary Islands and azure tits (Fig. 1) preclude the two first hypotheses of gene flow, because then also more Eurasian and African haplotypes were expected to be nested in the same clade. Nuclear phylogenetic and coalescent analyses strongly support the retention of an ancient polymorphism from the ancestor of the blue and azure tit. However, the sharing of alleles/lineages in the Eurasian-Canarian clade (node B) could be accounted also by gene flow in two non-exclusive ways: (i) historical or recent gene flow, perhaps facilitated by Pleistocene glacial periods, between Eurasian and African populations together with the erosion of signals of ancient divergence as found in herring gulls of the *Larus argentatus* complex (Liebers *et al.* 2004; Sternkopf *et al.* 2010); (ii) back colonization from the African mainland to the Canary Islands. Our coalescent analyses strongly indicate gene flow between *C. caeruleus* and *C. teneriffae,* although based on our phylogenetic results the western Canary Islands were excluded of such

events. Because a signal of gene flow was absent from mitochondrial DNA, our results suggest gene flow to be mediated by male dispersal. Two non-exclusive hypotheses might explain this phenomenon: (i) male-biased dispersal during the post-fledgling movements; or (ii) hybrid sterility in the heterogametic (females) sex (Haldane 1922; McCormack *et al.* 2010).

The hypothesis of colonization from southern Europe to the Canary Islands via La Palma gains additional support by the presence of a 12 bp mtDNA fragment shared by individuals from La Palma and Europe. However, an alternative explanation is that the La Palma individuals represent a relict population of an earlier east-west invasion from the African mainland. This is based on: (i) the four nuclear haplotypes from Fuerteventura and two from Morocco grouped with the western Canary Islands, which could represent relict haplotypes from this first colonization wave, and (ii) *Cyanistes* are absent from the other northern Macaronesian archipelagos (i.e. Madeira and Azores).

Overall, our results reveal a less than simple phylogeographic history within the genus *Cyanistes*, specifically within *C. caeruleus* and *C. teneriffae*, with a novel role for the Canary Islands as a probable source for the colonization of North Africa. These results provide the first example of an avian colonization from the oceanic islands of Macaronesia to the neighbouring continental region. This demonstrates that although the sea is an effective barrier that promotes the structuring of genetic variation within oceanic archipelagos, by means of limiting the movements of individuals among islands, it cannot be assumed to be an impediment to the successful colonization of continental areas.

Acknowledgements

Eulalia García, Oscar García, Jordi Figuerola, José Cabot, Roger Jovani, Paula Días, Jon Fjeldså, Barboutis Christos, Javier Pérez-Tris, Álvaro Ramírez, Felipe Rodríguez, Ángel Moreno, Guillermo López, Lluís Brotons, Reija Dufva, Seppo Rytkönen, Gernot Segelbacher, Tapio Eeva, Marcel Lambrechts, Vicente Polo, Manuela de Lucas, Eliseo Strinella, Augusto De Sanctis, Tayebeh Arbabi, Fares Khoury and CHAGRA ringing group provided samples. Juan Carlos Atienza, José Manuel Herranz, Cristina Rabadán and Ana Íñigo assisted with the sampling and many friends provided accommodation in the Canary Islands and Morocco. Martí Rodríguez kindly provided the blue, Canarian and azure tit drawings. We are very much indebted to Fernando Pacios who provides us with the map. We are grateful to CESGA (Centro de Supercomputación de Galicia), where Isolation and Migration analyses were performed. The University of Oulu provided all kind of facilities during the lab work. We thank Brent C. Emerson and four anonymous referees for invaluable comments on this study. This work was supported by a Spanish postdoctoral fellowship (subprogram Ramón y Cajal) and a Spanish mobility grant (José Castillejo) from the Spanish Ministry of Education and Science to J.C.I. Birds were trapped and ringed with permission of the countries cited.

References

Allan GJ, Francisco-Ortega J, Santos-Guerra A, Boerner E, Zimmer EA (2004) Molecular phylogenetic evidence for the geographic origin and classification of Canary Island *Lotus* (Fabaceae: Loteae). *Molecular Phylogenetics and Evolution*, **32**, 123–138.

Axelsson E, Smith NGC, Sundström H, Berlin S, Ellegren H (2004) Male-biased mutation rate and divergence in autosomal, Z-linked introns of chicken and turkey. *Molecular Biology and Evolution*, **21**, 1538–1547.

Backström N, Fagerberg S, Ellegren H (2008) Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Molecular Ecology*, **17**, 964–980.

Bellemain E, Ricklefs RE (2008) Are islands the end of the colonization road? *Trends in Ecology and Evolution*, **23**, 461–468.

Blondel J, Aronson JC (2010) *The Mediterranean Region: Biological Diversity Through Time and Space*. Oxford University Press, Oxford.

Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*, **176**, 1035–1047.

Carine MA, Russell SJ, Santos-Guerra A, Francisco-Ortega J (2004) Relationships of the Macaronesian and Mediterranean floras: molecular evidence for multiple colonizations into Macaronesia and back-colonization of the continent in *Colvolvulus* (Convolvulaceae). *American Journal of Botany*, **91**, 1070–1085.

Civetta L, Cornette Y, Crisci G, Gillot PY, Orsi G, Requejo CS (1984) Geology, geochronology and chemical evolution of the island of Pantelleria. *Geological Magazine*, **121**, 541–562.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Cosson JF, Hutterer R, Libois R, Sara M, Taberlet P, Vogel P (2005) Phylogeographical footprints of the Strait of Gibraltar and Quaternary climatic fluctuations in the western Mediterranean: a case study with the greater white-toothed shrew, *Crocidura russula* (Mammalia : Soricidae). *Molecular Ecology*, **14**, 1151–1162.

Cox SC, Carranza S, Brown RP (2010) Divergence times and colonization of the Canary Islands by *Gallotia* lizards. *Molecular Phylogenetics and Evolution*, **56**, 747–757.

Cramp S, Perrins CM (1993) *Handbook of the Birds of Europe the Middle East and North Africa. The birds of the Western Paleaertic, vol. VII*. Oxford University Press, New York.

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources*, **10**, 556–557.

Dawson DA, Hanotte O, Greig C, Stewart IRK, Burke T (2000) Polymorphic microsatellites in the blue tit, *Parus caeruleus*, and their cross-species utility in 20 songbird families. *Molecular Ecology*, **9**, 1941–1944.

Del Hoyo J, Elliott A, Christie DA (2007) *Handbook of the Birds of the World (HBW), Band 12: Picathartes to Tits and Chickadees*. Lynx Edicions, Barcelona.

Dietzen C, Witt H-H, Wink M (2003) The phylogeographic differentiation of the European robin, *Erithacus rubecula*, on the Canary Islands revealed by mitochondrial DNA sequence data and morphometrics: evidence for a new robin taxon on Gran Canaria? *Avian Science*, **2 & 3**, 115–131.

Dietzen C, García E, Delgado G, Wink M (2008) Phylogeography of the blue tit (*Parus teneriffae*-group) on the Canary Islands based on mitochondrial DNA sequence data and morphometrics. *Journal of Ornithology*, **149**, 1–12.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed Phylogenetics and Dating with Confidence. *PLoS Biology*, **4**, e88. doi:10.1371/journal.pbio.0040088.

Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology*, **18**, 2930–2933.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Filardi CE, Moyle RG (2005) Single origin of a pan-Pacific bird group and upstream colonization of Australasia. *Nature*, **438**, 216–219.

Fu YX (1997) Statistical test of neutrality of mutations against population growth, hitchhiking, and background selection. *Genetics*, **147**, 915–925.

Garcia-Castellanos D, Estrada F, Jiménez-Munt I *et al.* (2009) Catastrophic flood of the Mediterranean after the Messinian salinity crisis. *Nature*, **462**, 778–782.

Gladenkov AY, Oleinik AE, Marincovich L Jr, Barinov KB (2002) A refined age for the earliest opening of Bering Strait. *Palaeogeography, Palaeoclimatology and Palaeoecology*, **183**, 321–328.

Goodman SJ (1997) Rst calc: a collection of computer programs for calculating unbiased estimates of genetic differentiation and determining their significance for microsatellite data. *Molecular Ecology*, **6**, 881–885.

Greve C, Hutterer R, Groh K, Haase M, Misof B (2010) Evolutionary diversification of the genus *Theba* (Gastropoda: Helicidae) in space and time: A land snail conquering islands and continents. *Molecular Phylogenetics and Evolution*, **57**, 572–584.

Griswold CK, Baker AJ (2002) Time to the most recent common ancestor and divergence times of populations of common chaffinches (*Fringilla coelebs*) in Europe and North Africa: insights into Pleistocene refugia and current levels of migration. *Evolution*, **56**, 143–153.

Guillaumet A, Pons J-M, Godelle B, Crochet P-A (2006) History of the crested lark in the Mediterranean región as revealed by mtDNA sequences and morphology. *Molecular Phylogenetics and Evolution*, **39**, 645–656.

Haldane JBS (1922) Sex ratio and unisex sterility in hybrid animals. *Journal of Genetics*, **12**, 101–109.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hanotte O, Zanon C, Pugh A, Greig C, Dixon A, Burke T (1994) Isolation and characterization of microsatellite loci in a passerine bird: the reed bunting *Emberiza schoeniclus*. *Molecular Ecology*, **3**, 529–530.

Hansson B, Hasselquist D, Tarka M, Zehtindjiev P, Bensch S (2008) Postglacial colonisation patterns and the role of isolation and expansion in driving diversification in a passerine bird. *PLoS ONE*, **3**, e2794. doi: 10.1371/journal.pone.0002794.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt GM (1999) Post-glacial recolonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hewitt GM (2004) Genetic consequences of climatic oscillation in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 183–195.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hourlay F, Libois R, D'Amico F, Sarà M, O'Halloran J, Michaux JR (2008) Evidence of a highly complex phylogeographic structure on a specialist river bird species, the dipper (*Cinclus cinclus*). *Molecular Phylogenetics and Evolution*, **49**, 435–444.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, **17**, 754–755.

Idaghdour Y, Broderick D, Korrida A, Chbel F (2004) Mitochondrial control region diversity of the houbara Bustard *Chlamydotis undulata* complex and genetic structure along the Atlantic seaboard of North Africa. *Molecular Ecology*, **13**, 43–54.

Illera JC, Emerson BC, Richardson DS (2007) Population history of Berthelot's pipit: colonisation, gene flow and morphological divergence in Macaronesia. *Molecular Ecology*, **16**, 4599–4612.

Jønsson KA, Fabre P-H, Ricklefs RE, Fjeldså J (2011) Major global radiation of corvoid birds originated in the proto-Papuan archipelago. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 2328–2333.

Jost L (2008) Gst and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Juste J, Ibañez C, Muñox J *et al.* (2004) Mitochondrial phylogeography of the long-eared bats (*Plecotus*) in the Mediterranean Palaearctic and Atlantic Islands. *Molecular Phylogenetics and Evolution*, **31**, 1114–1126.

Khoury F, Azar J, Kvist L (2007) On the phylogenetics and population differentiation of the Great Tit and Blue Tit in Jordan. *Journal of Ornithology*, **148**, 525–530.

Kvist L, Ruokonen M, Lumme J, Orell M (1999) Different population structures in northern and suthern populations of the European blue tit (*Parus caeruleus*). *Journal of Evolutionary Biology*, **12**, 798–805.

Kvist L, Viiri K, Dias PC, Rytkönen S, Orell M (2004) Glacial history and colonization of Europe by the blue tit *Parus caeruleus*. *Journal of Avian Biology*, **35**, 352–359.

Kvist L, Broggi J, Illera JC, Koivula K (2005) Colonisation and diversification of the blue tits (*Parus caeruleus teneriffae*-group) in the Canary Islands. *Molecular Phylogenetics and Evolution*, **34**, 501–511.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Liebers D, deKnijf P, Helbig AJ (2004) The herring gull complex is not a ring species. *Proceedings of the Royal Society London B*, **271**, 893–901.

MacArthur RH, Wilson EO (1967) *The theory of island biogeography*. Princeton University Press, Princeton, New Jersey.

Mahood GA, Hildreth W (1986) Geology of the peralkaline volcano at Pantelleria, Strait of Sicily. *Bulletin of Volcanology*, **48**, 143–172.

Marshall HD, Baker AJ (1999) Colonization history of Atlantic island common chaffinches (*Fringilla coelebs*) revealed by mitochondrial DNA. *Molecular Phylogenetics and Evolution*, **11**, 201–212.

Martín A, Lorenzo JA (2001) *Aves del Archipiélago Canario*. Francisco Lemus Editor, La Laguna.

Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, **16**, 562–563.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, **26**, 2462–2463.

Maynard Smith J (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, **34**, 126–129.

McCormack JE, Heled J, Delaney KS, Townsend Peterson A, Knowles LL (2010) Calibrating divergence times on species trees versus gene trees: implications for speciation history of *Aphelocoma* jays. *Evolution*, **65**, 184–202.

Mort ME, Soltis DE, Soltis PS, Francisco-Ortega J, Santos-Guerra A (2002) Phylogenetics and evolution of the Macaronesian clade of Crassulaceae inferred from nuclear and chloroplast sequence data. *Systematics Botany*, **27**, 271–288.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Päckert M, Dietzen C, Martens J, Wink M, Kvist L (2006) Radiation of Atlantic goldcrests *Regulus regulus* spp.: evidence of a new taxon from the Canary Islands. *Journal of Avian Biology*, **37**, 364–380.

Päckert M, Martens J, Tietze DT, Dietzen C, Wink M, Kvist L (2007) Calibration of a molecular clock in tits (Paridae). Do nucleotide substitution rates of mitochondrial genes deviate from the 2% rule? *Molecular Phylogenetics and Evolution*, **44**, 1–14.

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology*, **265**, 218–225.

Park SDE (2001) *Trypanotolerance in west African cattle and the population genetic effects of selection*. Ph.D.Thesis. University of Dublin.

Pérez-Tris J, Bench S, Carbonell R, Helbig AJ, Tellería JL (2004) Historical diversification of migration patterns in a passerine bird. *Evolution*, **58**, 1819–1832.

Posada D (2008) jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 13757–13762.

Rambaut A, Drummond AJ (2007) Tracer v1.4, Available from http//beast.bio.ed.ac.uk/Tracer

Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A (2008) Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics*, **179**, 555–567.

Ramos-Onsins S, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Saladin V, Bonfils D, Binz T, Richner H (2003) Isolation and characterization of 16 microsatellite loci in the Great Tit *Parus major*. *Molecular Ecology Notes*, **3**, 520–522.

Salzburger W, Marten J, Sturmbauer C (2002) Paraphyly of the Blue Tit (*Parus caeruleus*) suggested from cytochrome b sequences. *Molecular Phylogenetics and Evolution*, **24**, 19–25.

Sanmartín I, Enghoff H, Ronquist F (2001) Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society*, **73**, 345–390.

Sheldon FH, Lohman DJ, Lim HC *et al.* (2009) Phylogeography of the magpie-robin species complex (Aves: Turdidae: Copsychus) reveals a Philippine species, an interesting isolating barrier and unusual dispersal patterns in the Indian Ocean and Southeast Asia. *Journal of Biogeography*, **36**, 1070–1083.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Sternkopf V, Liebers-Helbig D, Ritz MS, Zhang J, Helbig AJ, deKnijf P (2010) Introgressive hybridization and the evolutionary history of the herring gull complex revealed by mitochondrial and nuclear DNA. *BMC Evolutionary Biology*, **10**, 348.

Suárez NM, Betancor E, Klasert TE, Almeida T, Hernández M, Pestano JJ (2009) Phylogeography and genetic structure of the Canarian common chaffinch (*Fringilla coelebs*) inferred with mtDNA and microsatellite loci. *Molecular Phylogenetics and Evolution*, **53**, 556–564.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Turmelle AS, Kunz TH, Sorenson MD (2011) A tale of two genomes: contrasting patterns of phylogeographic structure in a widely distributed bat. *Molecular Ecology*, **20**, 357–375.

Weir JT, Schluter D (2008) Calibrating the avian molecular clock. *Molecular Ecology*, **17**, 2321–2328.

Weiss S, Ferrand N (eds.) 2006. *Phylogeography of Southern European Refugia - Evolutionary perspectives on the origins and conservation of European biodiversity*. Springer, Dordrecht, The Netherlands.

J.C.I. is interested in studying both ecological and historical processes shaping bird distributions and patterns of genetic variation within and among species inhabiting oceanic islands

that are the result of colonization, adaptation and diversification. K.K. is a behavioural and conservation ecologist whose research is currently focused on dispersal linked ecological and genetic phenomena and their role in shaping bird distribution and population dynamics. J.B. does research on ecophysiology, immunoecology and phylogeography on several bird taxa. M.P. is working as the curator of birds at Senckenberg Natural History Collections Dresden with a research focus on the molecular systematics and bioacoustics of Eurasian passerines. J.M. works on speciation of Palaearctic and Indo-Malayan passerines by means of morphology, acoustics and molecular genetics. L.K. does research on population genetics, phylogeography and molecular ecology of several Palaearctic taxa.

Data accessibility

DNA sequences: Genbank accession numbers are available as online supplementary Tables S1 and S4.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Species, localities and GenBank accession numbers used in the analyses for the different loci. Map numbers correspond with numbers shown in Figure 1

Table S2 Pairwise Φ_{ST} values for the mtDNA (based on Tamura-Nei's distance method) among *Cyanistes* populations. Non significant pairwise values ($P > 0.01$) were marked in bold. NC, Nordic countries; GE, Germany; UK, United Kingdom; SP, Spain; MAL, Mallorca (Balearic Islands); FR, France; CO, Corsica; IT, Italy; SI, Sicily; GR, Greece; IR, Iran; JO, Jordan; MO, Morocco; AL, Algeria; PA, Pantelleria; FV, Fuerteventura; LZ, Lanzarote; GC, Gran Canaria; TF, Tenerife; LG, La Gomera; LP, La Palma; EH, El Hierro

Table S3 Pairwise Φ_{ST} values for the nDNA (12884, based on Tamura-Nei's distance method) among *Cyanistes* populations. Non significant pairwise values ($P > 0.01$) were marked in bold. NC, Nordic countries; GE, Germany; UK, United Kingdom; SP, Spain; MAL, Mallorca (Balearic Islands); FR, France; CO, Corsica; IT, Italy; SI, Sicily; GR, Greece; IR, Iran; JO, Jordan; MO, Morocco; AL, Algeria; PA, Pantelleria; FV, Fuerteventura; LZ, Lanzarote; GC, Gran Canaria; TF, Tenerife; LG, La Gomera; LP, La Palma; EH, El Hierro

Table S4 Taxa and sequences used for molecular dating. Cytochrome-b sequences by Dietzen *et al.* (2008) were added to our cyt-b data set for single locus dating with each haplotype represented by one sequence only (DQ473999-DQ474061; cyt-b sequences included in the multi-locus dating approach in bold; Gran Canarian population was not included). Cytb: cytochrome b; CR: control region; Fib7: b-fibrinogen intron 7

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.