Molecular Ecology (2001) **10**, 2187–2198

DNA footprints of European hedgehogs, *Erinaceus europaeus* and *E. concolor*: Pleistocene refugia, postglacial expansion and colonization routes

J. M. SEDDON,*§ F. SANTUCCI,*†¶ , N. J. REEVE,‡¶ and G. M. HEWITT*
School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK, †Smithsonian Tropical Research Institute, Unit 0948, APO AA34002, Miami, USA, ‡School of Life Sciences, University of Surrey Roehampton, London SW15 3SN, UK

Abstract

European hedgehogs, *Erinaceus europaeus* and *E. concolor*, are among the many European plant and animal taxa that have been subjected to cyclical restriction to glacial refugia and interglacial expansion. An analysis of 95 mitotypes, comprising partial cytochrome *b* and control region sequences, shows deep divergence between the two hedgehog species. Three *europaeus* and two *concolor* clades are clearly identified and are consistent with previously identified refugia for Europe: the Iberian peninsula, Italy, and the Balkans. The degree of mitochondrial divergence among these clades suggests pre-Pleistocene separation of the refugial populations. In contrast, analysis of two nuclear introns clearly separates the two *concolor* clades, as in the mitochondrial data, but cannot discriminate the three *europaeus* clades. This discrepancy between nuclear and mitochondrial data is attributed to historical differences in the refugial population size of *europaeus* and *concolor*. The geographical distribution of mitotypes is analysed using nested clade analysis. This method, by including unobserved ('missing') mitotypes, can identify mitotype groupings that remain undetected in conventional analyses. However, the application of nested clade analysis to the study of refugial populations may be hampered by such factors as the loss of haplotypes from the refugial areas by repeated contractions of the population and the recent time scale of colonization relative to mutation rate.

Keywords: European phylogeography, hedgehogs, postglacial colonization

Received 29 January 2001; revision received 23 May 2001; accepted 23 May 2001

Introduction

Climatic fluctuations have dramatically influenced the distribution of many flora and fauna taxa. In particular, there have been repeated glacial and interglacial cycles in which species are forced to retreat into refugia by advancing ice and tundra but are able to expand from refugia during interglacial warming (Hewitt 1996). The genetic consequences of these forced movements have been modelled theoretically (Nichols & Hewitt 1994; Ibrahim *et al*. 1996), showing how differing dispersal conditions may influence the consequent genetic diversity. For example, rapid colonization by leptokurtic dispersal results in successive bottlenecking and a consequent loss of genetic diversity.

The cycles of restriction and colonization induced by glacial oscillations have led to significant structuring in the distribution of genomes across Europe. The resultant genomic distributions have been described in several species of plants and animals in Europe (reviewed by Taberlet *et al*. 1998; Hewitt 1999). These genetic patterns, together with climate and pollen data, have consistently identified three regions — the Iberian peninsula, Italy and the Balkans — as potential refugia for a wide variety of European flora and fauna.

Improved methods for the analysis of phylogeography data are being developed. Methods such as Templeton's nested clade analysis (NCA; Templeton 1998) allow a statistical interpretation of the geographical associations of haplotypes. Furthermore, it incorporates information

Correspondence: J. M. Seddon. §Present address: Evolutionary Biology Centre, Uppsala University, Norbyvagen 18D, Uppsala SE752 36, Sweden. Fax: + 46 (0) 18–471 6310; E-mail: jennnifer.seddon@ebc.uu.se
¶These authors have contributed equally and are considered joint second authors.

about the phylogenetic relationships of haplotypes, including unobserved haplotypes, to interpret past evolutionary and population events that have shaped the current genetic structure.

This study uses such methods to explore the phylogeography of the European hedgehog. Two species of hedgehog are currently found in Europe. The brown-breasted hedgehog, *Erinaceus europaeus* (Erinaceidae: Erinaceinae), is found in western Europe, the UK, Ireland, southern Scandinavia, and into Estonia and northern Russia. The white-breasted hedgehog, *E. concolor*, is found in eastern Europe, Turkey and Israel, and extends eastwards into Russia and Asia. The distribution of the two species shows limited overlap, meeting at approximately 15°E, with little suggestion of interbreeding in the wild (Corbet 1988; Reeve 1994).

A preliminary study was undertaken to assess the refugial regions and colonization patterns of *E. europaeus* and *E. concolor* (Santucci *et al*. 1998) using partial cytochrome *b* sequence data from 56 hedgehogs. Refugial regions in the Iberian peninsula, Italy, and the Balkans were postulated with northwards expansion from each of the three refugia. Although this study was able to describe the colonization routes, the geographical coverage of Europe was limited.

Three main issues are addressed in this paper. First, the sampling effort is increased to determine whether the patterns of colonization proposed initially prove applicable with greater geographical coverage. Second, the more sophisticated analytical methods of Templeton's NCA are applied to identify factors which have shaped the current genetic variation. Third, a comparison is made between the phylogeography based on mitochondrial DNA (mtDNA) and that determined from nuclear DNA to provide a better estimate of the species' evolutionary history.

Materials and methods

A total of 228 *Erinaceus europaeus* and 45 *E. concolor* samples from 23 countries across Europe have been analysed (Fig. 1, Table 1). Tissue or hair samples were taken, primarily from roadkills and hedgehogs in animal rescue centres. Samples used in previous studies (Filippucci & Simson 1996; Santucci *et al*. 1998; Suchentrunk *et al*. 1998) have been included here.

Total genomic DNA (gDNA) was extracted by tissue digestion in 500 μL extraction buffer (20 mM Tris, 10 mM EDTA, 0.5% SDS) with 30–75 μg Proteinase K, followed by purification using Wizard DNA Clean Up (Promega) mini-columns. Alternative procedures were occasionally used, including a Chelex (BioRad) extraction (Zhang & Hewitt 1998) for hairs and a CTAB-based method (Yang *et al*. 1997) for museum specimens.

Target fragments were amplified by polymerase chain reaction (PCR) using BIOTAQ (Bioline) *Taq* DNA polymerase (1 unit/50 μL reaction) and buffer, with primers at 1.2–2.0 μM. Approximately 100 ng of gDNA was added to a 50-μL reaction. A negative control without DNA was included in each set of reactions.

Two mitochondrial fragments were amplified. A partial cytochrome *b* fragment was amplified with primers (L14724 and H15149) and conditions previously described (Santucci *et al*. 1998). The 5′ end of the control region was amplified with the primers ProL-He (5′-ATACTCCTACCATCAACACCCAAAG-3′) and DLH-He (5′-GTCCTGAAGAAAGAACCAGATGTC-3′), modified from the primers of Wilkinson & Chapman (1991). Control region reactions contained 6 mM $MgCl_2$ and were amplified with an annealing temperature of 60 °C.

In a subset of 23 samples, chosen to represent the main mitochondrial clades and from widespread locations, two nuclear loci were amplified. Intron 7 of the beta-fibrinogen gene (*bfibr*) was amplified using primers BFIBR1 (5′-ATTCACAACGGCATGTTCTTCAG-3′) and BFIBR2 (5′-AANGKCCACCCCAGTAGTATCTG-3′) located in exons 7 and 8, respectively. Reactions contained 6 mM $MgCl_2$ and were performed at an annealing temperature of 60 °C. The second myoglobin intron (*myo*) was amplified initially using MYO2 and MYO3 (Slade *et al*. 1993) with primers subsequently designed to amplify a variable section of this intron for sequencing. These hedgehog specific primers, MYOintCC (5′-GGCTTTCAGGGCTGGTCTAGC-3′) and MYOintNCB (5′-AAGGAAGGACAAGARACTGGCARC-3′), were used with 1 mM $MgCl_2$ and an annealing temperature of 56 °C.

Cytochrome *b* sequences were sequenced directly to give continuity with the data set of Santucci *et al*. (1998). However, the control region fragments were screened for variation using a DGGE system (CBS Scientific) following manufacturer's directions. Samples were combined to give a *europaeus*/*concolor* heteroduplex to improve resolution (Campbell *et al*. 1995). Samples were electrophoresed through a gel with a gradient from 7.5 to 70% denaturant for 5 h at 150 V in 1× TAE (40 mM Tris, 1 mM EDTA) buffer heated to 40 °C, and visualized by silver staining. Variants identified on each gel were sequenced directly. Sequencing of samples showing an identical DGGE pattern confirmed the efficiency of mutation detection. Nuclear loci were sequenced directly.

Double-stranded PCR products were purified with the Qiaquick (Qiagen), Wizard (Promega) or Concert (Life Technologies) PCR purification columns prior to sequencing. Direct sequencing in both directions was performed with Big Dye Terminator RR (Applied Biosystems) and an ABI 377 sequencer. The primers described above were used for sequencing reactions, although two additional internal primers (BFIBR3 5′-GGTCTTTGCTCCACTGGGTGA-3′ and BFIBR4 5′-CCAAAACTGTGTTCTAATTCCA-3′) were used for sequencing the *bfibr* intron. Sequences have been deposited in GenBank (accession numbers AF379703–AF379853).

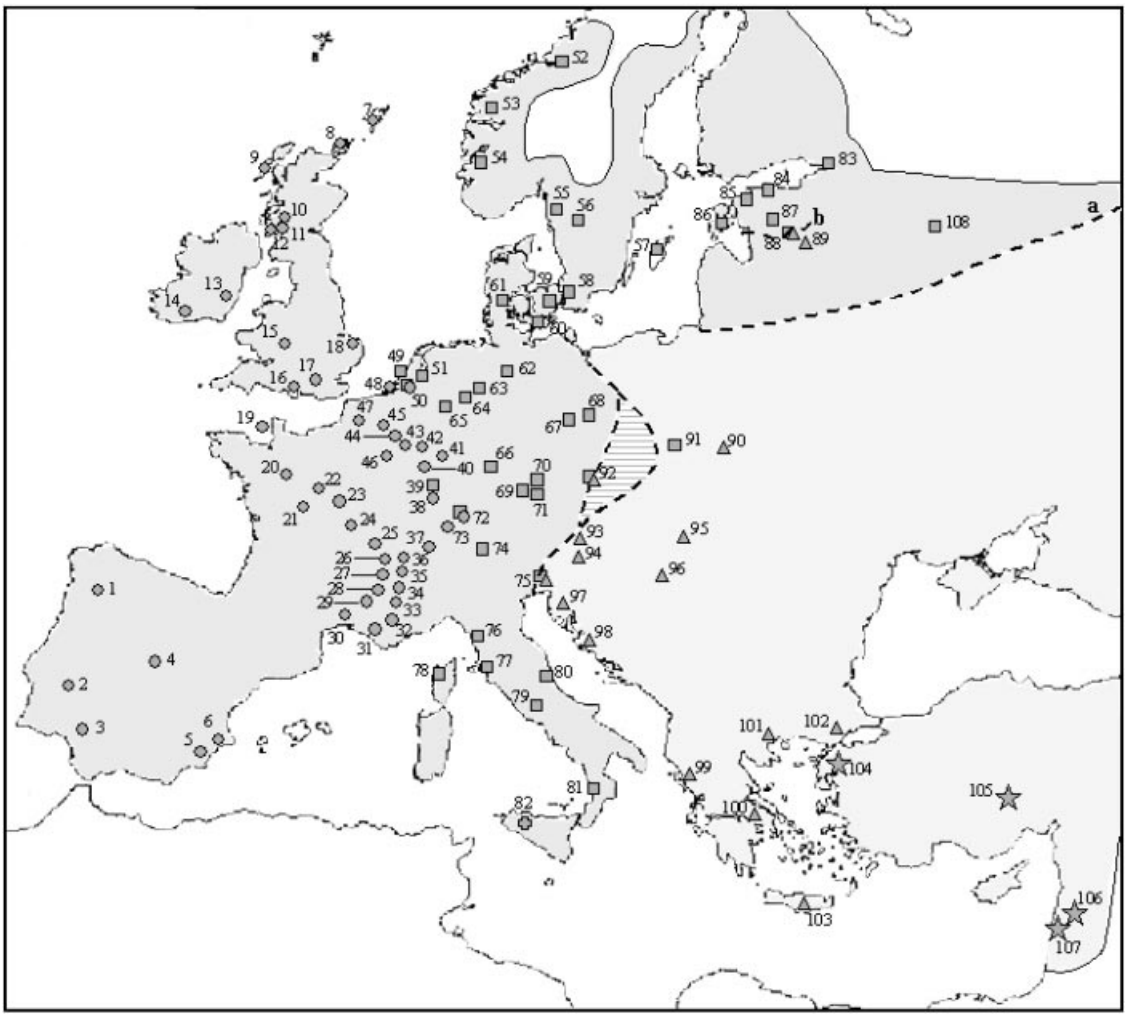

Fig. 1 Map of sampling locations and distribution of *Erinaceus europaeus* and *E. concolor*. Circle, *E. europaeus* sample of E2 mitotype group; square, E1 *E. europaeus* sample; cross, E3 *E. europaeus* sample; triangle, C1 *E. concolor* sample; star, C2 *E. concolor* sample. Location numbers correspond to those shown in Table 1. Unbroken lines show the northern extent of *E. europaeus* (dark shading) and the southern extent of *E. concolor* (light shading). The broken lines approximately depict the postulated boundary between the species, marking a zone of overlap in central Europe (horizontal lines), the northern boundary described by Reeve (1994; marked **a**), and the boundary within Estonia (Masing 1999; marked **b**).

Analysis

For each individual, the cytochrome *b* and the control region sequences were combined to give a 'mitotype', named by species name, clade name, and the cytochrome *b*/control region haplotype. Cytochrome *b* sequences were aligned by eye and translation into protein sequences did not reveal any stop codons, suggesting functional sequences were obtained. Alignment of the control region sequences was assisted by MegAlign (version 1.05 DNASTAR) using the clustal method. The sequences obtained for nuclear loci were aligned by eye. The identity of the nuclear intron was confirmed by comparing the flanking exon sequence with GenBank sequences.

Phylogenetic relationships were constructed with the neighbour joining (NJ) algorithm in PAUP* (version 4.0b3a, Swofford 2000). This method was based on Tamura–Nei distances for the mitochondrial data to account for an observed nucleotide bias and on Kimura's 2-parameter distances for the nuclear loci. Gaps were treated as missing data. Bootstrap resampling was used to assess support for tree nodes with 1000 replicates. Within-clade nucleotide diversity and, for each pairwise comparison of clades, both nucleotide diversity and nucleotide divergence were calculated with a Jukes–Cantor correction in DnaSP (Rozas & Rozas 1999).

The analysis of phylogeographic patterns was performed using NCA (Templeton 1998) for the two main *europaeus* clades. Other clades were not included in the analysis because of limited sampling and restricted genetic variation. A cladogram was constructed (Templeton *et al*. 1987) using TCS (alpha 1.01, Clement *et al*. 2000), assigning

Table 1 Sampling locations and mitotype distribution. *Location numbers correspond to those shown in Fig. 1. The number of mitotypes (*n*) is shown in brackets after each location number and mitotype

Country	Location* (*n*)	Mitotypes (*n*)	Location* (*n*)	Mitotypes (*n*)
Portugal	1 (4)	E2-10/17 (4)	2 (1)	E2-13/11 (1)
Spain	3 (2)	E2-10/11 (2)	5 (2)	E2-12/12 (2)
	4 (5)	E2-10/11 (2), E2-11/10 (3)	6 (1)	E2-12/12 (1)
UK	7 (1)	E2-01/01 (1)	12 (3)	E2-01/01 (3)
	8 (1)	E2-01/01 (1)	15 (1)	E2-01/01 (1)
	9 (6)	E2-01/01 (6)	16 (3)	E2-01/01 (1), E2-02/01 (2)
	10 (1)	E2-01/01 (1)	17 (5)	E2-01/01 (5)
	11 (2)	E2-01/01 (2)	18 (5)	E2-01/01 (5)
Jersey	19 (5)	E2-01/01 (1), E2-01/02 (4)		
Ireland	13 (2)	E2-01/01 (2)	14 (6)	E2-01/01 (6)
France	20 (4)	E2-01/03 (3), E2-01/05 (1)	33 (1)	E2-05/06 (1)
	21 (1)	E2-01/03 (1)	34 (1)	E2-09/14 (1)
	22 (1)	E2-01/04 (1)	35 (1)	E2-08/15 (1)
	23 (4)	E2-05/06 (3), E2-06/05 (1)	36 (1)	E2-09/14 (1)
	24 (1)	E2-05/06 (1)	40 (1)	E2-01/01 (1)
	25 (4)	E2-05/06 (2), E2-07/05 (1), E2-08/15 (1)	41 (2)	E2-01/01 (1), E2-01/09 (1)
	26 (2)	E2-08/15 (2)	42 (3)	E2-01/01 (2), E2-08/15 (1)
	27 (1)	E2-08/15 (1)	43 (3)	E2-01/01 (2), E2-01/09 (1)
	28 (1)	E2-08/13 (1)	44 (1)	E2-01/01 (1)
	29 (2)	E2-05/06 (1), E2-08/15 (1)	45 (4)	E2-01/01 (3), E2-01/09 (1)
	30 (3)	E2-01/05 (2), E2-08/13 (1)	46 (1)	E2-01/09 (1)
	31 (1)	E2-05/06 (1)	47 (1)	E2-01/01 (1)
	32 (2)	E2-05/06 (2)		
Netherlands	48 (2)	E2-01/01 (1), E2-03/01 (1)	50 (3)	E2-01/01 (1), E1-11/12 (1), E1-11/16 (1)
	49 (10)	E1-11/16 (2), E1-12/06 (3), E1-12/08 (4), E1-12/25 (1)	51 (5)	E1-11/14 (2), E1-11/16 (3)
Switzerland	37 (5)	E2-04/01 (2), E2-04/16 (3)	73 (5)	E2-04/01 (1), E2-04/07 (3), E2-04/16 (1)
	72 (4)	E1-01/30 (2), E2-04/07 (2)	74 (4)	E1-05/29 (1), E1-13/03 (3)
Germany	38 (2)	E2-04/07 (1), E2-04/08 (1)	66 (1)	E1-13/03 (1)
	39 (1)	E1-13/03 (1)	67 (1)	E1-12/08 (1)
	62 (8)	E1-11/10 (1), E1-08/01 (2), E1-09/01 (1),	68 (2)	E1-11/10 (1), E1-13/03 (1)
		E1-10/12 (1), E1-11/02 (2), E1-11/08 (1)	69 (3)	E1-13/04 (2), E1-15/04 (1)
	63 (4)	E1-08/01 (2), E1-08/08 (1), E1-11/12 (1)	70 (3)	E1-13/04 (1), E1-15/04 (1), E1-17/04 (1)
	64 (1)	E1-18/12 (1)	71 (2)	E1-13/04 (2)
	65 (1)	E1-11/08 (1)		
Norway	52 (5)	E1-01/21 (5)	54 (2)	E1-01/21 (2)
	53 (1)	E1-01/21 (1)		
Sweden	55 (1)	E1-01/23 (1)	57 (1)	E1-01/22 (1)
	56 (1)	E1-01/23 (1)	58 (1)	E1-04/22 (1)
Denmark	59 (3)	E1-01/24 (3)	61 (1)	E1-11/13 (1)
	60 (4)	E1-01/21 (4)		
Italy	75 (10)	E1-05/19 (5), E1-06/19 (1), C1-01/07 (1)	79 (9)	E1-01/18 (1), E1-01/20 (7), E1-05/19 (1)
		C1-02/07 (1), C1-03/07 (1), C1-04/07 (1)	80 (2)	E1-07/19 (1), E1-07/20 (1)
	76 (1)	E1-01/20 (1)	81 (4)	E1-01/18 (3), E1-07/18 (1)
	77 (2)	E1-05/19 (2)	82 (3)	E3-01/01 (3)
	78 (1)	E1-01/17 (1)		
Russia	83 (1)	E1-11/15 (1)	108 (1)	E1-02/01 (1)
	89 (1)	C1-05/01 (1)		
Estonia	84 (1)	E1-02/09 (1)	87 (3)	E1-02/12 (1), E1-03/26 (2)
	85 (1)	E1-01/28 (1)	88 (7)	E1-02/26 (1), C1-05/01 (6)
	86 (2)	E1-02/27 (1), E1-03/26 (1)		
Poland	90 (1)	C1-09/11 (1)	91 (1)	E1-16/05 (1)
Austria	92 (4)	E1-14/05 (1), C1-05/10 (1), C1-07/04 (1)	93 (5)	C1-05/05 (4), C1-07/03 (1)
		C1-08/10 (1)	94 (1)	C1-05/03 (1)
Hungary	95 (2)	C1-07/02 (2)	96 (2)	C1-08/10 (1), C1-12/05 (1)
Serbia	97 (1)	C1-10/05 (1)		
Croatia	98 (1)	C1-05/12 (1)		
Greece	99 (1)	C1-05/08 (1)	101 (1)	C1-11/09 (1)
	100 (1)	C1-05/06 (1)	103 (2)	C1-06/13 (2)
Turkey	102 (3)	C1-05/10 (1), C1-05/13 (1), C1-11/13 (1)	105 (1)	C2-01/04 (1)
	104 (4)	C2-01/01 (4)		
Israel	106 (1)	C2-02/02 (1)	107 (4)	C2-02/02 (1), C2-03/03 (2), C2-04/02 (1)

gaps as a fifth state. The 95% connection limit was established at 12 steps and so the clades could not be parsimonously connected. For the E1 clade, three haplotypes showed an unresolvable network due to homoplasy and were removed from the analysis. There was more than one most parsimonious solution, as indicated by the presence of loops in the cladogram. The cladogram presented here represents the relationships found most frequently among the most parsimonious trees derived in PAUP*, although alternative solutions are possible. Nesting clades were identified following the rules of Templeton *et al*. (1987) and Templeton & Sing (1993). The null hypothesis of no geographical associations of clades and nesting clades was tested by assessing the Dc (clade distances, which measures the geographical range of clades) and Dn (nested clade distances) for tip and interior clades in the program GeoDis (Posada *et al*. 2000) with 1000 random permutations.

Results

Mitotype variation

A total of 95 mitotypes are identified among the 273 samples. The combined mitotype provides 845 bp, of which 151 sites are variable and 132 sites informative under parsimony. The mean transition to transversion ratio is 4.35 and there is a substantial nucleotide bias (A 32.6%, C 19.2%, G 10.9%, T 37.3%). The sequences were compared with that of the complete mitochondrial sequence for hedgehog (X88898, Krettek *et al*. 1995) and match positions 14208–14590 (cytochrome *b*) and 15434–15885 (control region).

Distinct specific and intraspecific clades

The mitotypes clearly separate into *europaeus* and *concolor* species groups in the NJ tree (Fig. 2). The species are deeply divided, with a nucleotide diversity of 9.9%. Within each of the species, the mitotypes show further deep divisions, each with strong bootstrap support. Within *europaeus*, three monophyletic clades are seen, termed E1, E2 and E3. The mitotypes forming E1 are found from Italy northwards through Austria, Switzerland, Germany, the Netherlands, Scandinavia, and Estonia (Table 1). In contrast, the E2 clade is found only in western Europe, from Spain northwards through France, the Netherlands and into the UK and Ireland. The third clade E3 shows a single mitotype and is restricted to Sicily. The three *europaeus* clades are deeply separated, with similar nucleotide diversities of 4.18–4.99% (Table 2) among the three clades, suggesting a trichotomy.

Similarly, *concolor* is further subdivided into clades, termed C1 and C2. Mitotypes of the C1 clade are found in eastern Europe from Turkish Thrace and Greece, northwards through Austria and Hungary to Estonia. The C2 clade is defined by mitotypes from a limited number of individuals and locations in Turkey (south of the Bosphorus) and Israel. These *concolor* clades show a deep intraspecific division, with a nucleotide diversity of 4.84% (Table 2). The level of divergence observed between the *concolor* clades is similar to that found among the *europaeus* clades.

Phylogeny vs. NCA cladograms

The *europaeus* E1 clade shows relatively little discrimination between haplotypes in the NJ tree. However, one group of six mitotypes is separated from the others with significant bootstrap support. These mitotypes are restricted to the region of southern Germany, northern Switzerland, Lower Austria, and Poland. In the cladogram constructed for NCA, this mitotype group is separated from other E1 mitotypes by seven mutations (Fig. 3a). However, nesting the cladogram provides five 3-step clades, with Clade 3–5 taking mitotypes from both NJ clades.

	n	1	2	3	4	5
(a) mtDNA						
1. *europaeus* 1	44	0.00857	0.04724	0.04993	0.09856	0.11022
2. *europaeus* 2	23	0.03591	0.01129	0.04175	0.09506	0.11111
3. *europaeus* 3	1	0.04585	0.03615	nd	0.09159	0.10497
4. *concolor* 1	22	0.09183	0.08118	0.08918	0.00491	0.04844
5. *concolor* 2	5	0.10249	0.10206	0.10158	0.04260	0.00674
(b) nuclear DNA						
1. *europaeus* 1	7	0.00492	0.00363	0.00381	0.00968	0.01192
2. *europaeus* 2	8	0.00024	0.00345	0.00398	0.00910	0.01101
3. *europaeus* 3	1	0.00135	0.00226	nd	0.01052	0.01238
4. *concolor* 1	5	0.00617	0.00568	0.00881	0.00343	0.01177
5. *concolor* 2	2	0.00858	0.00796	0.01026	0.00777	0.00425

Table 2 Nucleotide diversity and nucleotide divergence values for the *Erinaceus europaeus* and *E. concolor* clades. Upper matrix: nucleotide diversity between populations; lower matrix: nucleotide divergence (net nucleotide diversity) between populations; along diagonal: within population nucleotide diversity. *n*, number of sequences; nd, value not determined

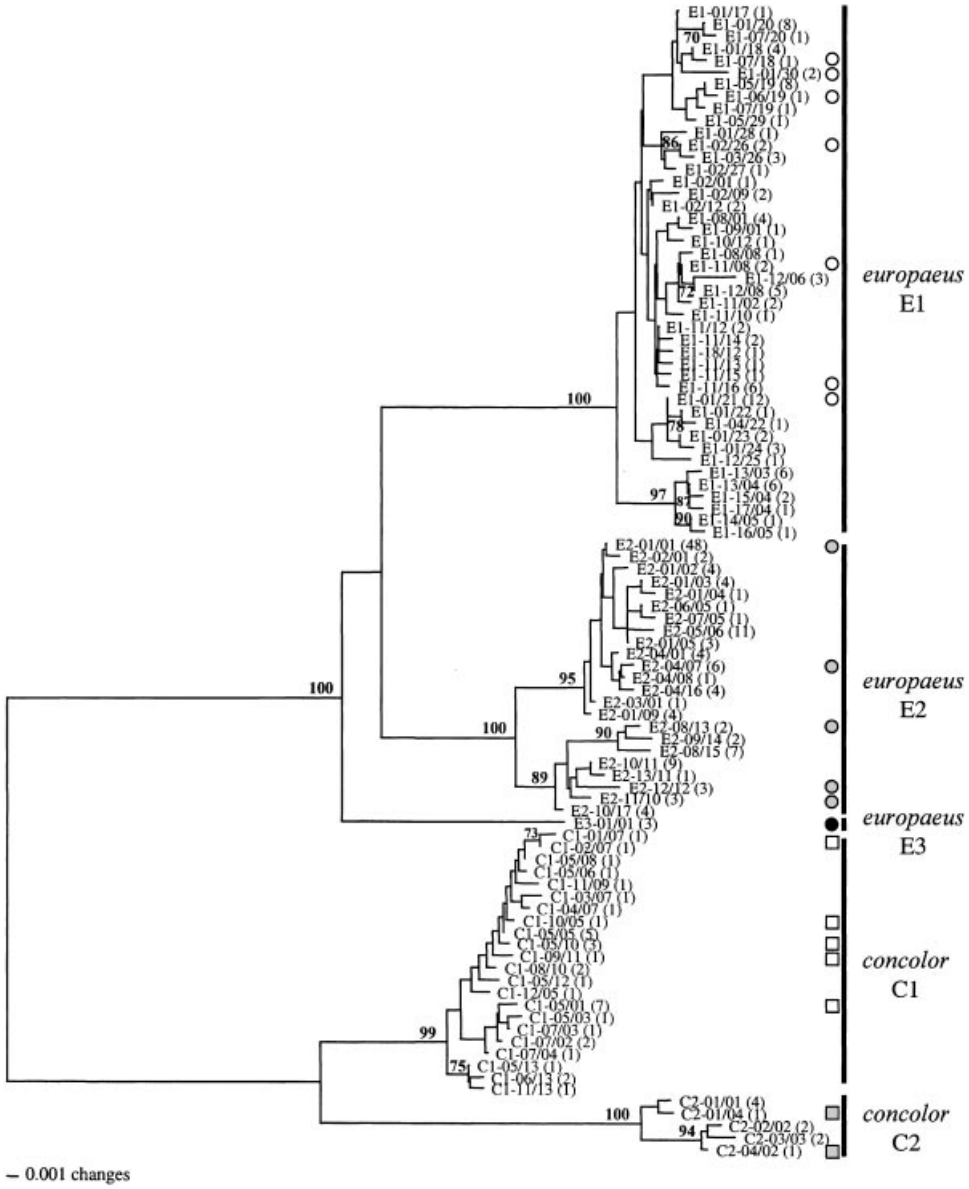

Fig. 2 Phylogenetic relationships of the mitotypes for *Erinaceus europaeus* and *E. concolor*. The neighbour joining tree is calculated on Tamura–Nei distances. Bootrap values from 1000 iterations are shown for values of 70 or greater. The tree is mid-point rooted. The number of mitotypes observed is shown in brackets following each mitotype. Circles and squares indicate samples sequenced for nuclear loci.

The E2 clade shows a substantial split in the NJ tree, with two clades each supported as monophyletic groupings with high bootstrap support. The upper clade (95% bootstrap support) contains the very frequent mitotype from the UK, Ireland, the Netherlands, and northern France together with others from England, France, Jersey, Switzerland, and southwest Germany. The lower clade (89% bootstrap support) contains mitotypes from Spain and Portugal, together with a monophyletic group (bootstrap support 90%) of mitotypes from southeast France. This strong division can be seen on the NCA cladogram (Fig. 3b), yet nesting of the E2 cladogram provides three 3-step clades, with Clade 3–2 taking mitotypes from both NJ clades. In both examples, the nested cladogram incorporates 'missing' mitotypes to detect otherwise cryptic groupings.

The *concolor* C1 group shows some subdivision in the NJ tree, with three mitotypes from Turkish Thrace and Crete forming a monophyletic clade (bootstrap support 75%). In the nested cladogram, this clade (Clade 2–6, Fig. 3c) is separated from other mitotypes by only two mutations. However, the distribution pattern shows two of the higher level clades represented in the Balkans, with northern extension to Poland (Clade 3–3) or to Austria (Clade 3–2). Clade 3–1 extends from Austria to Estonia and northwest Russia, with a

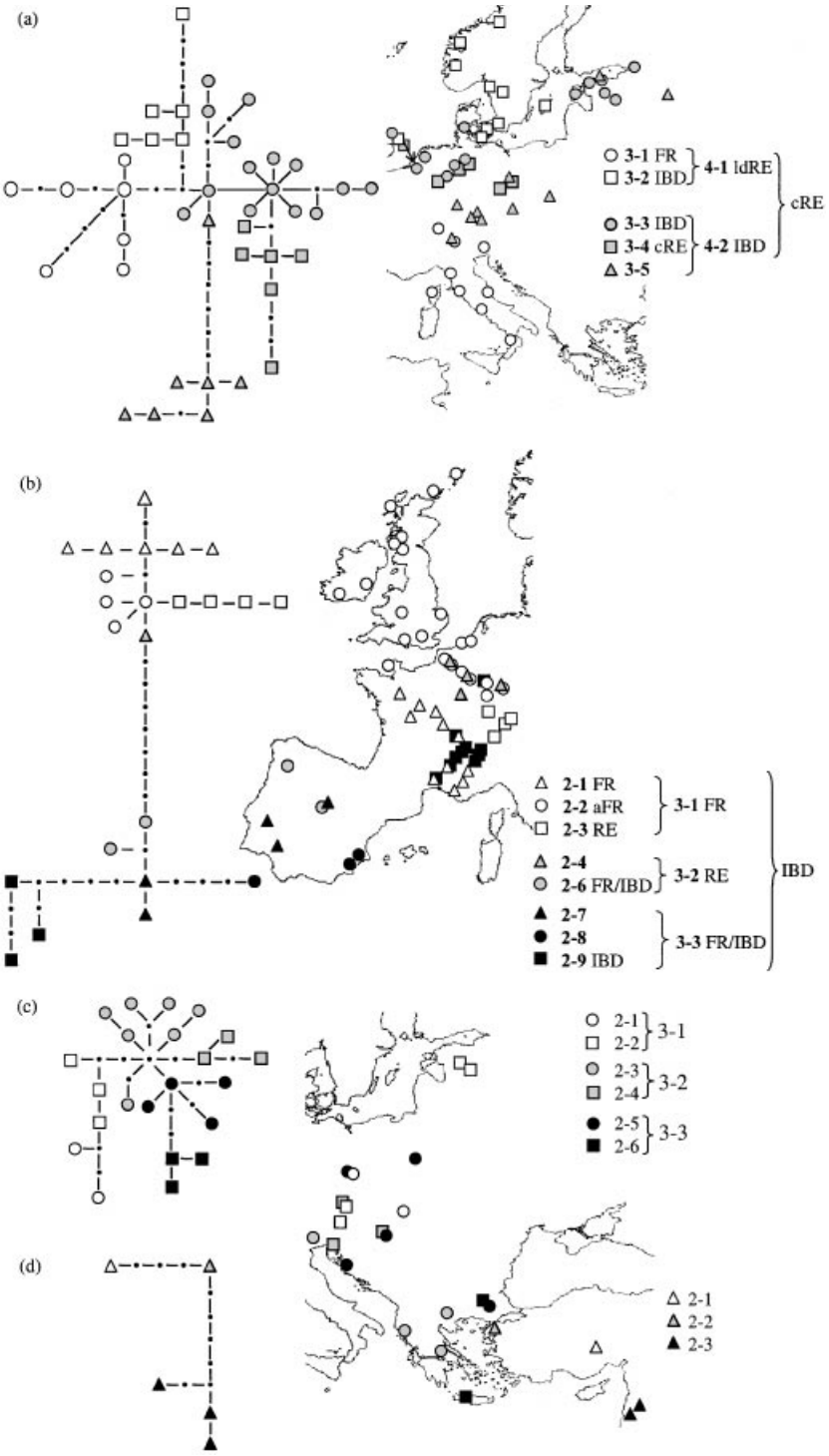

Fig. 3 Cladograms and Distribution maps for nested clade analysis (NCA). Results of the NCA analysis are shown for (a) *Erinaceus europaeus* group E1; (b) *E. europaeus* group E2; (c) *E. concolor* group C1; (d) *E. concolor* group C2. For each analysis, the higher level clades are mapped with corresponding designation in the cladogram. The results of the analysis of geographical distribution of nested clades is shown as IBD, isolation by distance; FR, fragmentation; aFR, allopatric fragmentation; RE, range expansion; cRE, contiguous range expansion; ldRE, long distance colonization. The complete nesting arrangement and the individual mitotype designation are not included.

single mitotype represented in the northern edge of the range.

The *concolor* C2 group is clearly divided, with the mitotypes from Israel separated as a monophyletic lineage on the NJ tree and in the NCA cladogram (Clade 2–3, Fig. 3d). Limited sampling in both *concolor* groups precludes geographical analysis of the nested mitotypes.

Population history suggested by NCA

Analysis of the geographical distribution of mitotypes incorporates the nested relationships of the cladogram to detect signals of population history events at different nesting levels and hence on a relative time scale. The substantial differentiation among the three *europaeus*

	myo	*bfibr*
	11111111222222223	1111112222223344455667
	11124647777889111446799	00012237788990067990126880136789480
	38973770458072128179153	24572477989291744797768232913299142
E. europaeus		
Italy 11	TTACCCCTGCCCACCTCCAGCTG	?GGTCCGGTCTTCACCCTTTATGCGAGAGTTACTC
Italy 1	CC.....?.G.TG.....	T.....T....-.....A.....Y..
Switzeraland 1	CC.....	ACT...T.....C.....
Germany 50	CC.....T..G.....	T.....TA.....--A.....
Netherlands 6	CC.....GTGTT.....G.....	T.....T.....T...C.....
Norway 23	?C.?.....?..G.....	A.....T....A...T.....
Estonia 9	CC.....-TC.G.....	T.....T.....T.....
Spain 4	CC.....??.....	???...T.....A.C.....A.
Spain 8	CC.....G....	T.....T.....MW.C.....W.
Switzerland 13	CC.....	T.....T.....C.....C.....
Switzerland 3	CC.....	T.....T.....T....T.....
France 18	..GTT.....	T.....T.....C.....
France 46	?.GT.....G.....	T.....T..?.....A.C.....
Netherlands 3	CC.....G.....	A.....T.....--AT.....
England 3	CC.....	T.....T.....C.....
Sicily 1	CC.....T....T...	T.....T.....
E. concolor		
Turkey 5	CC...TT.....T....C.	T...TTT..?...C.....C...A..A...A....
Italy 23	CC....T.....T....C.	T...T.T.....C.....C...A..A...A....
Serbia 1	CC....T....T.A..T....C.	???????.GTC..C.....C...A..A...A....
Poland 2	CC....T.....T....C.	T..CT.T.....C.....C.G.A..A...GC...
Estonia 10	CC....T.....AG.T....C.	???.T.T.....C.....C...A..A...A....
Turkey 8	CC....T.....A..T...TCR	T.....T.....AC....AC..T..C..T.....T
Israel 9	C.....T.....T...TC.	T.....T.....AC..T..C..TC....T

Fig. 4 Variable nucleotide sites for two nuclear introns in *Erinaceus europaeus* and *E. concolor*. Fifty-five variable sites and three nonvariable heterozygous sites for *bfibr* and *myo* sequences are shown for 23 samples of *E. europaeus* and *E. concolor*. Symbols: . = identity with first sequence; – = indel; ? = missing data.

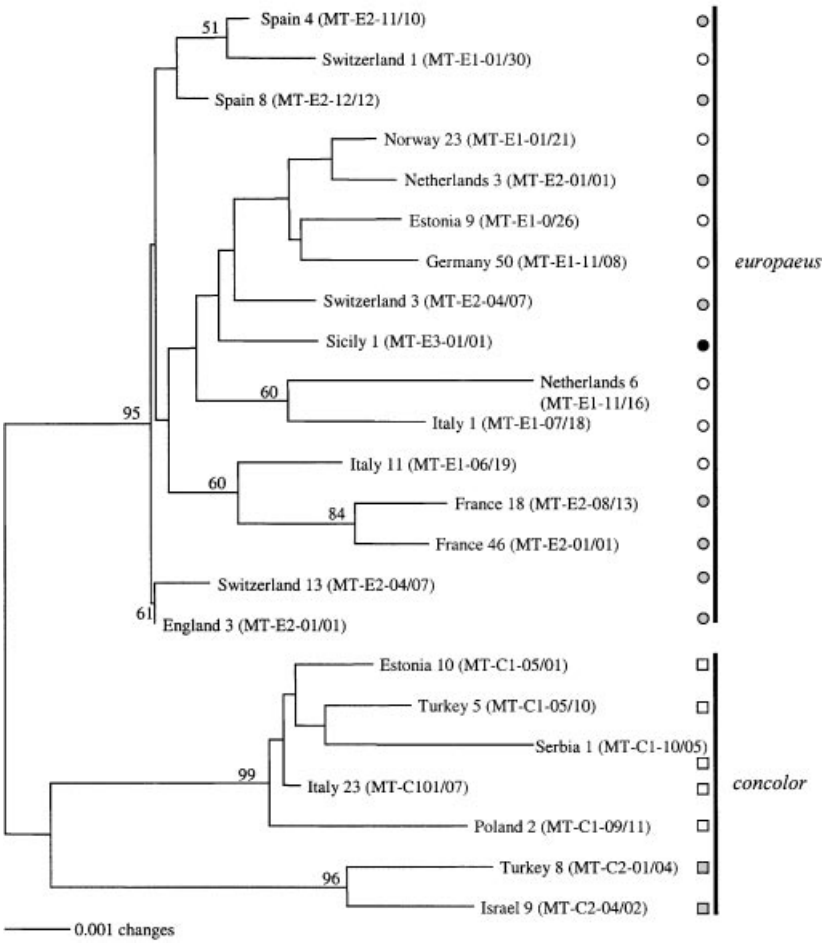

Fig. 5 Phylogenetic relationships of nuclear DNA sequences for *Erinaceus europaeus* and *E. concolor*. Neighbour joining tree of a combined data set of *bfibr* and *myo* intron sequences, calculated with Kimura's 2-parameter distances. Bootstrap values for 1000 iterations are shown where values exceed 60. The tree is mid-point rooted. Each sequence is labelled with the country of origin and sample number, with the sample's corresponding mitotype given in brackets. Circles and squares indicate *europaeus* and *concolor* clades and correspond to the symbols used in Fig. 2.

clades is indicative of past fragmentation but the high level of divergence precludes a combined analysis by NCA. The results for the geographical analysis of the E1 and E2 clades are given in Fig. 3. For the E1 clade, an overall contiguous range expansion is detected. Long distance colonization between a southern group in Italy and Switzerland and a northern group in Scandinavia and the Netherlands, is identified in Clade 4–1 (Fig. 3a). The

southern group (Clade 3–1) shows subsequent fragmentation between the Italian and Swiss mitotypes. More recent range expansion is identified in north-central Europe (Clade 3–4).

For the E2 clade, analysis of geographical variation reveals restricted gene overall. Range expansion is seen at the higher level clades only in Clade 3–2, which has mitotypes in Spain and Portugal and in northern France (Fig. 3b). A more complex pattern is found in the widespread Clade 3–1 with overall fragmentation, subsequent range expansion in the mitotypes of Switzerland and southwest Germany (Clade 2–3) and continued fragmentation in southern and central France (Clade 2–1). Separation between the Jersey mitotype and those in the UK, France, and the Netherlands may be attributed to allopatric fragmentation.

Nuclear DNA analysis

There is a clear and relatively deep divergence among both interspecific and intraspecific clades in the mitochondrial sequences. Two nuclear intron regions, sequenced in a subset of 23 of the samples, show a different pattern. The level of variability for both loci is extremely low. For *myo*, 17 different sequences are identified and of the 402 bp, only 22 sites are variable (Fig. 4) and only 13 are informative under parsimony. Similarly, 22 *bfibr* sequences are identified and of the 796 bp, only 33 sites are variable (Fig. 4) and 15 informative under parsimony. There are only three heterozygous sequences (1 *myo* and 2 *bfibr*); the majority of sequences are homozygous, presumably a reflection of low within-site variation. Because of the low levels of observed variability and the low level of intraspecific nucleotide diversity (0.96%), further analysis is conducted on a combined data set.

As for the mtDNA data, the NJ tree (Fig. 5) shows clear separation of the *europaeus* and the *concolor* sequences. The *europaeus* sequences form a monophyletic clade with significant bootstrap support. However, the clear division of the *europaeus* sequences into three clades, as in the mitochondrial data, is not seen with the nuclear data. Indeed, the net nucleotide divergence between E1 and E2 is only 0.02% (Table 2). The removal of sequences from regions where differing mitochondrial clades contact, that is those with increased potential for introgression, or the removal of heterozygous sequences does not alter the resolution of the tree. The *concolor* sequences divide into two monophyletic groups, both with significant bootstrap support. These groups correspond to the two *concolor* clades of the mitochondrial data and are separated by a nucleotide diversity of 1.18%, greater than between the two species. There is no discordance in the species assignment between the mtDNA and nuclear DNA data, indicating that the results of any hybridization are not found among these samples.

Discussion

Species distribution

Both *europaeus* and *concolor* mitotypes are identified in Friuli, Italy, and Lower Austria confirming the known north–south boundary between the species (Fig. 1). However, a single specimen from Krakow, Poland (Location 91, Fig. 1) shows a *europaeus* mitotype, supporting a past karyological identification of *europaeus* in this area (Corbet 1988). In addition, a *concolor* mitotype is found in the Pskov region of Russia, and in Varska, Estonia. This finding concords with the recent morphological description of *concolor* in southeast and southwest Estonia (Masing 1999), suggesting that the described distribution of *Erinaceus concolor* in eastern Europe requires revision.

Potential refugia

Physical and fossil data of the climatic changes in Europe suggest that hedgehogs, along with other flora and fauna, were restricted to refugial areas in southern Europe during glacial periods. The distribution of the two main *europaeus* clades (E1 and E2) and one of the *concolor* clades (C1) is concordant with northwards expansion from each of the three refugial regions in the Iberian peninsula, Italy, and the Balkans, as identified by Santucci *et al*. (1998). The second *concolor* clade (C2) is clearly associated with a further refugium (Santucci *et al*. 1998) but the exact location is not yet discernable.

While the alignment of the north–south genomic division with postulated refugia demonstrates a postglacial colonization, the extent of the mtDNA divergence among populations indicates much older refugial separation. The time of divergence can be estimated by applying both an average mtDNA substitution rate (11.06×10^{-9} substitutions/site/year; Pesole *et al*. 1999) to nucleotide diversity values and a substitution rate for the control region (12.6×10^{-9} substitutions/site/year; Pesole *et al*. 1999) to the control region sequences only. The division between both the two main *europaeus* clades and between the two *concolor* clades approximates to 1.7–2.2 Myr and the two species of hedgehog, *E. europaeus* and *E. concolor*, have been separated for an estimated 3.2–4.5 Myr. These dates imply pre-Pleistocene divisions of the species and refugial populations, as suggested in Santucci *et al*. (1998).

Refugial history

The distinction between the various refugia with their colonized areas is clearly defined for the mtDNA, but for the nuclear DNA data there is a lack of discrimination for *europaeus*. This inconsistency provides information about

the history of the refugial populations. mtDNA is haploid and, with the consequent reduction in effective population size, is more sensitive to the effects of bottlenecks in population size. Therefore, a more severe bottleneck in the founding *concolor* population, relative to the *europaeus* population, may have resulted in structuring in the *concolor* population for both nuclear and mtDNA, yet only mtDNA structuring in *europaeus*. The alternative explanation, that there has been a lack of time for lineage sorting in the nuclear DNA of the *europaeus* clades, is unlikely because a similar time of divergence for the intra-*europaeus* and intra-*concolor* clades is suggested by the similarity of their mtDNA divergence. Male biased dispersal provides an adequate and obvious explanation for differences in the mitochondrial and nuclear patterns. However, there is currently no evidence to support substantial sex-based differences in the dispersal of hedgehogs. Furthermore, it is possible that the lack of concordance is an artefact of the limited sampling for the nuclear marker and requires confirmation with a more detailed study.

Colonization routes

Three basic patterns of expansion from the refugial areas of Europe have been identified (Hewitt 1999). The hedgehog's pattern of northwards expansion from all three refugial areas is similar to that seen with the oak (Ferris *et al*. 1993, 1998). This similarity is not unexpected. The current northern limit of *E. europaeus* lies approximately at the limit of deciduous woodland where the winter climatic conditions and restricted supply of macro-invertebrates limit their survival (Kristiansson 1981; Reeve 1994). Therefore, we could conclude that the expansion of the hedgehogs from southern refugia is likely to follow the establishment of deciduous woodland, resulting in the similar expansion patterns of oak and hedgehog.

An interesting distribution of mitotypes is observed in the E1 clade. The *europaeus* mitotypes in Estonia and Russia are closely related to those in north-central Europe, including the Netherlands, Germany and Denmark (Clade 3–3, Fig. 3a), yet a single clade dominates Scandinavia (Clade 3–2). This distribution suggests that Estonia and Russia were colonized from central Europe through the Baltic States, with subsequent displacement by *concolor*. An alternative hypothesis is a colonization route from central Europe, through Scandinavia and Finland to Estonia, with subsequent replacement of mitotypes in Scandinavia. However, this scenario is less likely because the dominance of a single clade throughout much of Scandinavia is suggestive of a long-distance colonization and such colonization events are unlikely to succeed where a few immigrants enter an occupied habitat (Hewitt 1996). Furthermore, suggestions of human-assisted movement of hedgehogs into Scandinavia (Kristiansson 1981) from Denmark and the Baltic States are not supported by our data.

Patterns of expansion

Although hedgehogs have been able to expand in interglacial periods, the climatic conditions were such that only the populations in the refugia would be able to survive a glacial period. The most recent expansion, therefore, post-dates the ending of the last glacial period approximately 16 000 years ago. This time scale is relatively recent in comparison with the mutation rate and hence the majority of the observed and inferred mitotypes (at a minimum, all but the tip mitotypes) must have survived in the refugium. Therefore, given the necessarily restricted range of a refugial area, we can conclude that it is most likely that the refugial populations exist as a series of smaller, isolated populations, maintaining overall population diversity. Such fragmented populations may result from spatial variability in climate as suggested by pollen data (Huntley 1999). Subsequent expansion from these smaller populations would carry different combinations of mitotypes to colonized areas (Hewitt 1996).

Certain patterns associated with the expansion of hedgehogs from the refugial regions can be expected in this study. Long distance colonization can result in a loss of diversity due to successive bottlenecking (Nichols & Hewitt 1994; Hewitt 1996). A loss of diversity in the northern geographical extremes of expansion is seen in the western *europaeus* clade (E2), with a single mitotype found in most of England, in Scotland, and Ireland and, similarly, a single mitotype of the *concolor* clade (C1) is found in Estonia.

Expansion patterns are also expected in the NCA. The *europaeus* clade of central Europe (E1) shows results concordant with an initially slower rate of expansion from northern Italy, and subsequent long distance colonization from Italy to Scandinavia or northern Europe. In contrast, the western clade of *europaeus* (E2) shows only initial restricted gene flow in the NCA. This pattern is inconsistent with the restriction of hedgehogs to southern regions during glacial periods with expansion predicted to have occurred on a recent timescale relative to the mutation rate.

Several problems have been encountered that hamper the application of NCA to the analysis of refugial populations. First, mitotypes may be lost from the refugial region. If all mitotypes in an expanding population are lost from the ancestral (refugial) population, their observed distribution will be restricted to the colonized area (that is, the measured clade distances will be small) and expansion will not be detected. This situation is most obvious in the *europaeus* E1 clade, where a single clade was found in Italy, the site of the refugial population. Expansion detected within the E1 clade overall probably relates to expansion from central Europe not from the refugium.

Second, the refugial population may not remain in the same geographical location in glacial and interglacial periods. Movement to differing elevations within refugial areas and the north–south movement of refugia within Europe have been proposed (Hewitt 1996). The high level of diversity for the hedgehogs currently seen in Austria and southern Germany may reflect such movement. For NCA, the full historical clade distance will not be included in the analysis.

A third problem arises from the expansion from refugia occurring on a recent timescale compared with the mutation rate. The relative distributions of tip and interior clades used by NCA (Cruzan & Templeton 2000) will be hindered because all (or almost all) haplotypes 'originate' in the contracted population. Finally, the repeated contraction of the population into refugia and the consequent increased probability of haplotype loss results in a large number of 'missing' haplotypes with unknown distribution, weakening the analysis. NCA provides a useful method for determining factors affecting population structure, however, its application would benefit from investigating the behaviour of haplotypes in more complex biological situations.

Acknowledgements

We are grateful to the many people who have assisted this study by providing hedgehog samples from throughout Europe, without whom this work would not have been possible: P. Arntzen, A. Bannikova, S. Becher, J. Bestard, D. Burdon, P. Burri, G. Christie, G. Csorba, K. Fournais, J. Galian, N. Gietema, J. Gillies, J. Gomez-Zurita, E. Heller, M. G. Filippucci, M. Huijser, D. Jackson, B. Johansen, J. Kinsman, M-C. Lehmann, K. Lehmann, A. Loureiro, L. Lutsar, M. Masing, M. Masseti, F. Medeiros, F. Mielo, H. Navarro-Hahn, I. Olofsson, F. Palacios, H. Philipps, A. Schon, A. Servent, P. Sleeman, F. Suchentrunk, J. Szymura, T. Taylor, D. Wall and M. Zaitsev. Samples were also provided by the Harrison Museum and Wildlife Aid. Laboratory assistance was provided by Patricia Sourrouille and Andrea Ungaro. This project was funded by the Leverhulme Trust.

References

Campbell NJH, Harriss FC, Elphinstone MS, Baverstock PR (1995) Outgroup heteroduplex analysis using temperature gradient gel electrophoresis: High resolution, large scale, screening of DNA variation in the mitochondrial control region. *Molecular Ecology*, **4**, 407–418.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Corbet GB (1988) The family Erinaceidae: a synthesis of its taxonomy, phylogeny, ecology and zoogeography. *Mammal Review*, **18**, 117–172.

Cruzan MB, Templeton AR (2000) Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record. *Trends in Ecology and Evolution*, **15**, 491–496.

Ferris C, King RA, Vinola R, Hewitt GM (1998) Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. *Heredity*, **80**, 584–593.

Ferris C, Oliver RP, Davy AJ, Hewitt GM (1993) Native oak chloroplasts reveal an ancient divide across Europe. *Molecular Ecology*, **2**, 337–344.

Filippucci MG, Simson S (1996) Allozyme variation and divergence in Erinaceidae (Mammalia: Insectivora). *Israel Journal of Zoology*, **42**, 335–345.

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 re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Huntley B (1999) Climatic change and reconstruction. *Quaternary Proceedings*, **7**, 513–520.

Ibrahim K, Nicols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, **77**, 282–291.

Krettek A, Gulberg A, Arnason U (1995) Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europaeus*, and the phylogenetic position of the Lipotyphala. *Journal of Molecular Evolution*, **41**, 952–957.

Kristiansson H (1981) Distribution of the European hedgehog (*Erinaceus europaeus* L.) in Sweden and Finland. *Annales Zoologici Fennici*, **18**, 115–119.

Masing M (1999) *Taxonomy and status of wild mammals in Estonia 1945–94*. Sicista Development Centre, Tartu.

Nichols RA, Hewitt GM (1994) The genetic consequences of long distance dispersal during colonisation. *Heredity*, **72**, 312–317.

Pesole G, Gissi C, De Chirico A, Saccone C (1999) Nucleotide substitution rate of mammalian mitochondrial genomes. *Journal of Molecular Evolution*, **48**, 427–434.

Posada D, Crandall KA, Templeton AR (2000) GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

Reeve N (1994) *Hedgehogs*. T & AD Poyser (Natural History), London.

Rozas J, Rozas R (1999) DnaSP, Version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics*, **15**, 174–175.

Santucci F, Emerson BC, Hewitt GM (1998) Mitochondrial DNA phylogeography of European Hedgehogs. *Molecular Ecology*, **7**, 1163–1172.

Slade RW, Moritz C, Heideman A, Hale PT (1993) Rapid assessment of single-copy nuclear DNA variation in diverse species. *Molecular Ecology*, **2**, 359–373.

Suchentrunk F, Haiden A, Hartl GB (1998) On biochemical genetic variability and divergence of the two Hedgehog species *Erinaceus europaeus* and *E. concolor* in central Europe. *Zeitschrift fur Saugetierkunde*, **63**, 257–265.

Swofford DL (2000) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts.

Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. 1. Basic theory and an analysis

of alcohol dehydrogenase activity in Drosophila. *Genetics*, **117**, 343–351.

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV Nested analyses with cladogram uncertainty and recombination. *Genetics*, **134**, 659–669.

Wilkinson GS, Chapman AM (1991) Length and sequence variation in evening bat d-loop mtDNA. *Genetics*, **128**, 607–617.

Yang H, Golenberg EM, Shoshani J (1997) Proboscidean DNA from museum and fossil specimens: An assessment of ancient DNA extraction and amplification techniques. *Biochemical Genetics*, **35**, 165–179.

Zhang D-X, Hewitt GM (1998) Isolation of DNA from preserved specimens. In: *Molecular Tools for Screening Biodiversity: Plants and Animals* (eds Karp A, Isaac PG, Ingram DS). Chapman & Hall, London.

This work forms part of a larger study by Godfrey Hewitt examining the postglacial colonization of chosen taxa in Europe and the genetics of ice age refugial populations. Jennifer Seddon has wider interests in conservation biology and in the behaviour of genes within populations. Fiammetta Santucci is interested in historical biogeography and its consequences for the creation and maintenance of biodiversity, and she is now exploring this topic in the tropics. Nigel Reeve is a mammal ecologist with a special interest in hedgehogs.
