Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 39 (2006) 335–346
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Lusitania revisited: A phylogeographic analysis of the natterjack toad *Bufo calamita* across its entire biogeographical range

Graham Rowe [a,b], D. James Harris [c], Trevor J.C. Beebee [a,*]

[a] *School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK*
[b] *School of Biological Sciences, University of Wales, Bangor, Gwynedd LL57 2UW, UK*
[c] *CIBIO/UP, Campus Agrário de Vairão, R. Monte-Crasto, 4485-661 Vila do Conde, Portugal*

Received 4 July 2005; revised 18 August 2005; accepted 23 August 2005
Available online 17 October 2005

Abstract

Attempts to understand the current distributions of plants and animals require both historical and ecological information. Phylogeography has proved highly effective in elucidating historical events such as postglacial colonisations in north temperate zones. However, interesting questions still await resolution. Lusitanian distributions of fauna and flora in western Europe, for example, have puzzled biogeographers for more than 150 years. Lusitanian species have highly disjunct distributions in Ireland and in Iberia, often with few or no other populations inbetween. Despite much debate, no agreed explanation for Lusitanian distributions has yet emerged. We investigated the phylogeographic structure of one Lusitanian species, the natterjack toad *Bufo calamita*, using mitochondrial DNA control region sequences and allelic variation at eight microsatellite loci. Our results show that this amphibian must have survived in north European refugia, as well as in Iberia, during and since the last (Weichselian) glacial maximum around 20,000 years before present (BP). Subsequent local recolonisation after the Younger Dryas cooling around 11,000 years BP best explains the Lusitanian aspect of natterjack toad distribution.
© 2005 Elsevier Inc. All rights reserved.

Keywords: Lusitanian hypothesis; Natterjack toad; Mitochondrial DNA; Microsatellites

1. Introduction

It is widely recognised that historical events, as well as ecological constraints, play key roles in the determination of biogeographical ranges. A major challenge in evolutionary biology is to understand the interplay of these factors. Phylogeographical analysis (Avise, 2000) has been remarkably successful in revealing patterns of postglacial recolonisation by many north-temperate zone animals and plants (Hewitt, 2000, 2001; Taberlet et al., 1998). However, there has as yet been no satisfactory explanation of the so-called Lusitanian fauna and flora of southwestern Ireland and Iberia (Praeger, 1939; Webb, 1983). These organisms (perhaps 15–20 species, mostly plants and invertebrates; see Vincent, 1990) occur in southwest Ireland and Iberia but are often absent, or demonstrate highly disjunct distributions, elsewhere in Ireland and in countries between them such as England and France. Well-known examples include the Kerry slug *Geomalacus maculosus* and the strawberry tree *Arbutus unedo*, and the Lusitanian issue remains as one of the most intriguing unresolved questions in biogeography. Four main hypotheses have been proposed to account for this curious distribution: (1) A glacial refugium situated in or near southwest Ireland on what was then dry land (Forbes, 1846). This refuge would have sustained relicts from pre-glacial continuous distributions in western Europe. (2) Introduction to Ireland by humans or birds from Iberia in ancient times (>2000 years BP) when there is evidence of human trade between the two localities (Corbet, 1961, 1962). (3) Introduction to Ireland by humans much more recently, from an unknown source (O'Rourke, 1970). (4) Post-glacial colonisations from an Iberian refuge through France and by land bridges via Britain, with some

* Corresponding author. Fax: +44 1273 678433.
E-mail address: t.j.c.Beebee@sussex.ac.uk (T.J.C. Beebee).

1055-7903/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2005.08.021

or all intermediate populations subsequently going extinct (Preece et al., 1986).

We tested these hypotheses by investigating the phylogeographic structure of one Lusitanian species, the natterjack toad *Bufo calamita*. Natterjacks occur in southwestern Ireland and Iberia, and also in two disjunct regions of England (the northwest coast and the southeast) as well as in much of north-central Europe. A preliminary phylogeographic study using only microsatellite loci and a small sample of populations indicated that genetic diversity was highest in Iberia, the only Mediterranean peninsula in which the species occurs (Beebee and Rowe, 2000). Here, we report a more extensive analysis, using both mitochondrial DNA control region sequences and allelic variation at eight microsatellite loci across a much larger population sample. Our primary goal in this study was to distinguish between the various explanations previously proposed to account for Lusitanian distributions.

2. Materials and methods

2.1. Sample collection

We sampled 10–40 natterjack toads from each of 22 separate sites spread across the species' entire biogeographical range (Fig. 1). Adult tissues (toe clips) were obtained in two cases (sites 3 and 4). In all the other sites, larval tailfins were the source of DNA. Every effort was made to minimise sibship sampling at these sites by visiting multiple different ponds and by catching larvae at many places around each pond. DNA was extracted from *B. calamita* tissues using a Chelex-based procedure (Walsh et al., 1991) and stored at −20 °C.

2.2. Mitochondrial DNA amplification and sequencing

For mitochondrial DNA analysis we used material from 120 individuals in 12 of the sampling sites (numbers from each site are given in Fig. 1), concentrating mostly on localities in western Europe. A section of mitochondrial DNA control region was sequenced following PCR amplification with forward (5′-AGCCTCTCCTTGAATTAAGAGC) and reverse (5′-CGGGATTAAGGTACGATAGTGG) primers. The region amplified corresponded to base pairs 15,701–16,082 of the *Bufo melanosticus* mitochondrial DNA genome (GenBank NC005794; Arnason et al., 2004). It contained no repeat motifs and was situated 3′ of the centre of the control region. Approximately 100 ng of DNA were amplified in 50 mM Tris–HCl, pH 8, 16 mM ammonium sulphate, 3.5 mM $MgCl_2$, 0.1 mM dNTPs, 0.2 μM of each primer, and 0.5 units *Taq* Express DNA polymerase (Genetix, UK). An initial denaturation step at 94 °C for 4 min was followed by 30 cycles of 94 °C for 1 min, 55 °C for 1 min, and 70 °C for 1 min.

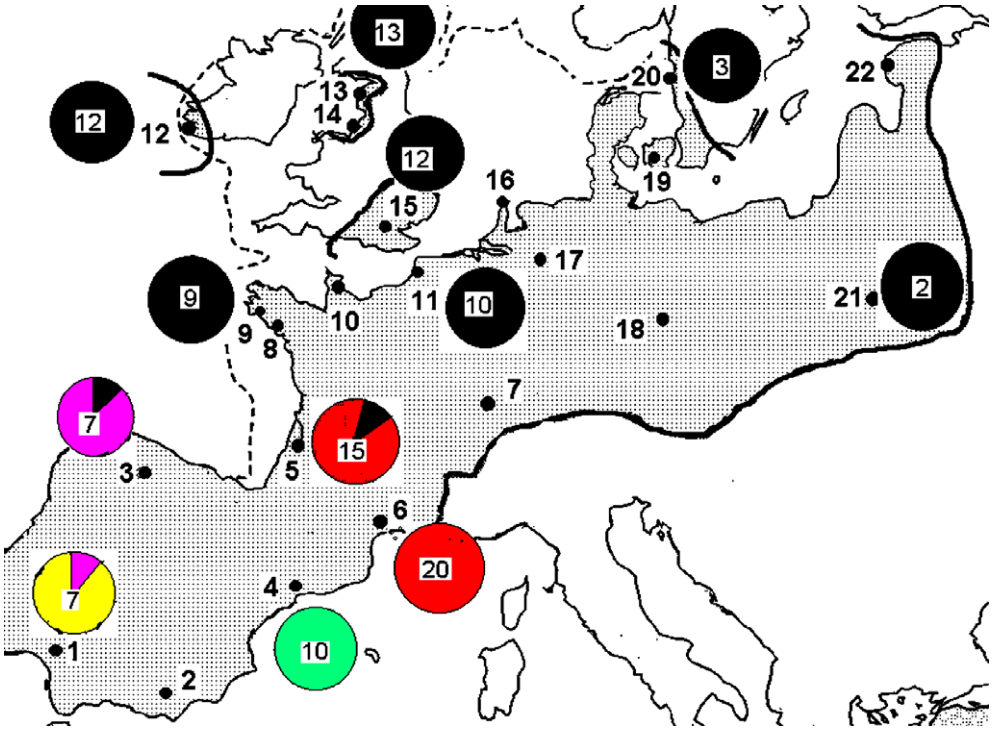

Fig. 1. Natterjack toad distribution, sampling sites and mitochondrial DNA clade frequencies. (shaded area) Approximate distribution limits. (- - - -) Coast line ca. 15,000 years ago. Sampling sites (area or nearest large town): 1, Seville, Spain; 2, Almeria, Spain; 3, Leon, Spain; 4, Barcelona, Spain; 5, Bordeaux, France; 6, Carmargue, France; 7, Zurich, Switzerland; 8, South Brittany, France; 9, West Brittany, France; 10, Cherbourg, France; 11, Boulogne, France; 12, Kerry, Ireland; 13, Duddon, England; 14, Merseyside, England; 15, East Anglia, England; 16, Texel, Netherlands; 17, Nijmegen, Netherlands; 18, Halle, Germany; 19, Zealand, Denmark; 20, Uddevalla, Sweden; 21, Bielowieza, Poland; 22, Parnu, Estonia. Pie charts show the frequencies of the five major clades as shown in Fig. 2, with total sample sizes for each site: clade A (black), clade B (red), clade C (green), clade D (purple), and clade E (yellow).

There was a final elongation cycle of 70 °C for 5 min. The PCR products were purified using Qiagen kit columns and all samples were tested for the amplification of a single band of the expected size by electrophoresis on a 1.5% agarose gel. The remaining amplification products were sequenced in forward and reverse directions by Genetic Research Instrumentation. Sequence data were edited and aligned using the DNASTAR (Applied Biosystems) EDITSEQ and MEGALIGN programs.

2.3. Microsatellite genotyping

Eight polymorphic microsatellite loci previously characterised for *B. calamita* were amplified by the PCR including [α-^{33}P]dATP as a substrate, and alleles were scored after gel electrophoresis and autoradiography all as described elsewhere (Rowe et al., 1997). Twenty-one populations out of the 22 sampled altogether (Fig. 1) were analysed for microsatellite genotypes. Population 4 (northeast Spain) was omitted (though included in the mtDNA analysis) because of insufficient material. Microsatellite data were already available prior to this study from the British and Irish populations (nos. 12–15), and from seven others (nos. 2, 9, 11, 16, 17, 20, and 21) in mainland Europe (Beebee and Rowe, 2000). For this work, however, we obtained data from a further 10 populations on the continent (see Fig. 1 for site numbers), notably southwest Spain (1), north Spain (3), southwest France (5), southeast France (6), Switzerland (7), Brittany (8, the second population sampled in this area), northern France (10), east Germany (18), Denmark (19) and Estonia (22). Sample sizes were 40 individuals from each population except for populations 2 ($n = 11$), 3 ($n = 13$), 5 ($n = 16$), 6 ($n = 22$), 9 ($n = 32$), and 11 ($n = 34$).

2.4. Mitochondrial DNA sequence analysis

Gene (haplotype) diversity (H) and nucleotide diversity (π_n) of mitochondrial DNA sequences were estimated using ARELQUIN 2.0 (Schneider et al., 1997).

$$\mathrm{H} = n/n - 1(1 - \sum_{i=1}^{k} p_i^2) \quad \text{and} \quad \pi_n = \sum_{i=1}^{k} \sum_{j<\mathrm{I}} p_i p_j d_{ij},$$

where n is the number of haplotype copies in the sample, k is the number of haplotypes, p_i and p_j are the sample frequencies of the ith and jth haplotypes, respectively, and d_{ij} is an estimate of the number of mutations since divergence of haplotypes i and j. Tajima's test for neutrality was also performed using ARLEQUIN. Phylogenetic analysis of mitochondrial DNA sequences used PAUP 4b10* (Swofford, 2002) with neighbour-joining (NJ), maximum parsimony (MP) and maximum likelihood (ML) methods. For NJ we tested both the Jukes–Cantor and HKY85 substitution models for consistency of results. MP using all equal weights was performed under the heuristic search option with 50 replicate searches and random addition of taxa. One thousand bootstrap replicates were carried out for the NJ and MP analyses. The optimal model of molecular evolution for the ML analysis was determined using MODELTEST 3.6 (Posada and Crandall, 1998). The ML analysis used a heuristic search with random sequence addition and 200 bootstraps. Bayesian analysis was implemented using MRBAYES (Huelsenbeck and Ronquist, 2001) with random starting trees, run for 5×10^5 generations and sampled every 100 generations using a general-time-reversible model of evolution. Burn-in data were discarded and all data collected at stationarity were used to estimate posterior nodal probabilities.

Genetic (P) distances between clades were estimated using PAUP. Every pairwise distance between all the populations in pairs of clades were computed, and means and standard deviations of these pairwise data sets were then used as inter-clade distance estimates. We also constructed haplotype networks of the *B. calamita* control region sequences using the program TCS 1.15 (Clement et al., 2000), using gaps as a fifth state. The program LINTRE (Takezaki et al., 1995) was used with all 25 haplotype sequences to test for rate constancy of evolution in the NJ trees. We applied the two-cluster test using the Jukes–Cantor model and Kimura's 2 parameter model (K2P) to assess rate constancy in these analyses.

2.5. Microsatellite data analysis

Tests for concordance of microsatellite data with Hardy–Weinberg equilibrium and linkage equilibrium were performed using GENEPOP 3.4 (Raymond and Rousset, 1995). Genetic diversities (expected heterozygosities) and allelic richness were estimated using FSTAT 2.9.3 (Goudet, 1995). All sample sizes were standardised at $n = 11$ (the smallest total sample we obtained, from southeast Spain) for allelic richness estimates by random sampling of the larger data sets from the other populations. Phylogenetic analysis of microsatellite data used the NJ method with Cavalli-Sforza chord distances and 1000 bootstrap replicates, and the ML method with 45 bootstrap replicates, in PHYLIP 3.5c (Felsenstein, 1993). Principal component analysis of microsatellite allele frequencies using the SPSS 1.5 statistics package (SPSS, Chicago, USA) was followed by production of interpolation maps using the spatial analyst extension of ARCVIEW 3.2 (ESRI, Aylesbury, UK), using the inverse distance weighted (IDW) option with a power of two. IDW assumes that each input point has a local influence that diminishes with distance. The sampling sites and six nearest neighbours were used for the calculations, and the interpolation surfaces were divided into three equal classes (Hanotte et al., 2000).

The microsatellite data were also used to assess the relationships between *B. calamita* populations in Ireland, northwest England, southeast England, and northern Spain using the IM program (Hey and Nielsen, 2004). This method incorporates a maximum likelihood approach to estimate the divergence times, interpopulation migration rates and effective sizes of population pairs. We used the

stepwise mutation model for all eight microsatellite loci in four pairwise comparisons: Ireland (site 12) × northwest England (site 14), Ireland × southeast England (site 15), Ireland × northern Spain (site 3), and northwest England × southeast England. Forty genotypes were analysed from all populations except northern Spain, for which only 13 were available. Because each run took a long time with our data (48 h for 10^6 steps), we repeated the analysis five times with 10^6 steps for each pairwise comparison simultaneously on several PCs, rather than using individual long runs of $>10^7$ steps as recommended by the program authors but which were impractical for our data. We used a burn-in of 100,000 steps, three Markov chains and three chain swaps for each run. Thus convergence of the chain simulations upon the true stationary distribution was assessed by using multiple independent chains starting at different points and including a procedure for swapping among the chains. Values of scalars for θ_1 ($=4N_e\mu$ of population 1) maxima and maximum times of population splitting were determined empirically during short preliminary runs as suggested by the program authors. In all cases we set the inter-site migration rates (m1 and m2, migrations in both directions) to zero, because in most cases the population divergences we were investigating were generated by sea level rises that must have precluded subsequent gene flow among these amphibians. This was probably an oversimplification for the Ireland × northern Spain comparison (see Section 4), but should suffice to generate a minimum divergence time. In one case (southeast England × northwest England) there is no intervening sea, but we also excluded the possibility of migration because forest cover between these regions was complete by about 9000 BP, within 1000 years of post-Younger Dryas colonisations of Britain (Vincent, 1990). *Bufo calamita* has an absolute requirement for open, unforested habitats in northern latitudes and will not migrate across even small areas of forest (Beebee, 1983). We were interested primarily in estimating the different divergence times of the population pairs, though we also obtained estimates of effective population sizes. Peaks of the resulting distributions were interpreted as estimates of these parameters.

2.6. Statistical tests

General statistical tests including estimation of confidence intervals, rank correlation, Wilcoxon signed rank tests and Kruskal–Wallis one-way analysis of variance (ANOVA) were performed using the STATISTIX package (Tallahassee, USA). The data were always used in non-parametric procedures. Geographical distances between sampling sites were estimated (in km) as great circle (taking account of earth curvature) connecting lines, except those entering and leaving Iberia which were all routed through a coastal point immediately east of the Pyrenees (and therefore were each the sum of two straight lines). Rank correlations were then made against the geographical distances using mean expected heterozygosity, mean allelic richness and mean allele repeat number (all across the full eight microsatellite loci).

3. Results

3.1. Mitochondrial DNA analysis

Three hundred and seventy-two base pairs of natterjack toad mitochondrial DNA control region sequences were compared among the 120 samples from 12 sampling sites (Fig. 1). Twenty-five haplotypes were identified (GenBank Nos. AY631244–68), with a total of 72 polymorphic sites. Fig. 2 shows the phylogenetic relationships of these control region haplotypes, in an unrooted tree generated using the NJ method. Trees generated with both substitution models (Jukes–Cantor is shown), or using ML or MP methods, were essentially identical with five robustly supported clades. Sixty-five individuals clustered in clade A, which contained six haplotypes, and 64 of these were from north of the Pyrenees. All British and Irish toads ($n = 37$, including 12 from Ireland) were fixed for the same ''North Europe'' haplotype that was also present in three Swedish toads and in nine out of 10 samples from northeast France. A single individual with this haplotype was also found in northern Spain, while the remaining Spanish samples ($n = 23$) had a total of 14 divergent haplotypes in clades C, D, and E, all of which were unique to Iberia. The five clade B haplotypes were all identified in 31 individuals from southern France. Using *Bufo viridis* (the closest relative of *B. calamita*) as an outgroup, all trees for *B. calamita* rooted between clades D and E (data not shown).

Haplotypes could not be connected into a single network using the TCS program with 95% confidence, essentially because of the large numbers of mutational steps (>10) between some of the clades, and especially between clade A and the rest (data not shown). For this reason we did not attempt nested clade analysis. However, a robust sub-network for clade A was obtained and confirmed that the north European haplotype fixed in the British Isles and elsewhere in northern Europe differed by only single nucleotide substitutions from the derived haplotype found in northeast France, and from the derived haplotypes fixed in northwest France and Poland. The two derived clade A haplotypes found in southwest France also varied by just one or two substitutions from the main north European haplotype.

We investigated genetic diversity among haplotypes after compartmentalising the sampling sites into four geographical regions. These were zones in which historical separation was considered likely on the basis of topography (mostly mountain ranges) and the phylogeny shown in Fig. 2. The first two zones, southern and northern Spain, corresponded with sampling sites 1 and 3 (and thus mostly with clades E and D). The third zone included northeast Spain and southern France (mostly clades C and B). Clade C was not analysed separately because there were only two haplotypes present, and both were unique to the single sample site. The fourth zone was the rest of Europe, and

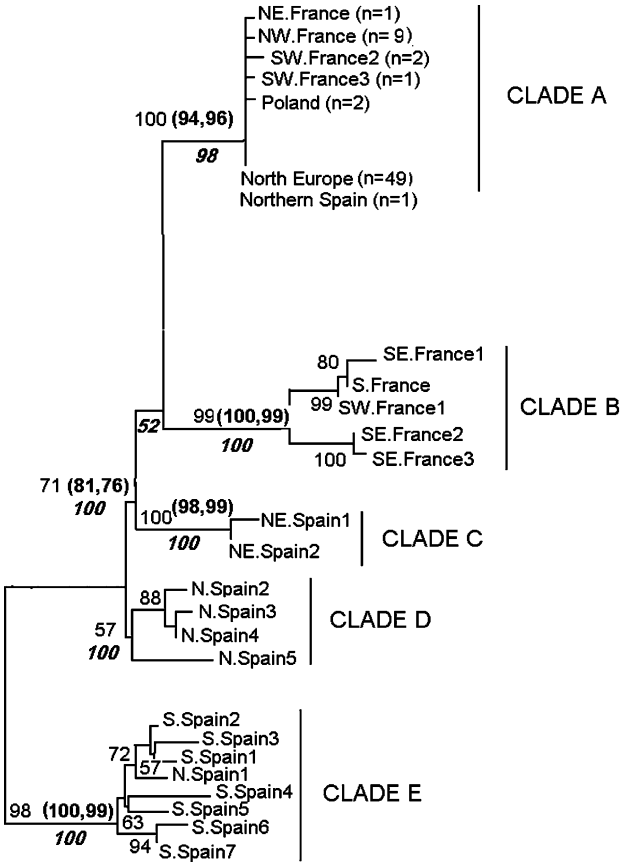

Fig. 2. Unrooted neighbour-joining tree showing relationships of the 25 haplotypes and clades A–E. Numbers show percentage bootstrap support (where >50%) with values for maximum likelihood and maximum parsimony given in parentheses. Nodal probabilities from Bayesian analysis are italicised beneath nodes, and numbers (n=) of individual sequences found in clade A are also shown.

thus entirely clade A. Only one haplotype was found in more than one of these regional divisions (an individual with the north Europe haplotype in northern Spain). Haplotype diversity ranged between 0.70 and 1.00 in Iberia and southern France but averaged only 0.34 in the rest of Europe (Table 1). Nucleotide diversity (π_n) was more than 30-fold higher in Iberia and southern France than in the rest of Europe. Tajima's test (Tajima, 1989) confirmed that the control region sequence was acting as a neutral marker (overall D statistic $= -0.316$, $P = 0.399$).

Table 2
Approximate mtDNA clade divergence times

Clades	Divergence times in Myr BP (SD)
Northern Europe × southern France + northeast Spain ($A \times [B + C]$)	1.22 (0.25)
Southern France + northeast Spain × northern + southern Spain ($[B + C] \times [D + E]$)	1.81 (0.38)
Northern Spain × southern Spain ($D \times E$)	1.47 (0.10)

Using two different evolution models (Jukes–Cantor and K2P), the program LINTRE (Takezaki et al., 1995) found no nodes inconsistent with a linear rate of evolution. To estimate divergence times between clades we used an inter-population nucleotide substitution rate for the *Bufo* control region of about 3.5% per million years (Liu et al., 2000; Macey et al., 1998). Average distances between haplotypes in the north European clade A and those in clade B from southern France and clade C from northeast Spain indicated a divergence time for clade A of around 1.22 Myr BP, well before the Weichselian glacial maximum of ca. 20,000 years ago (Table 2).

3.2. Microsatellite analysis

We analysed polymorphism across eight microsatellite loci in samples from every population in Fig. 1 except northeast Spain. Our expectation was that microsatellite analysis should reveal more recent and finer-scale events than those demonstrable with mitochondrial DNA. A total of 150 alleles was detected in the 21 population samples, and mean expected heterozygosities across all loci ranged from 0.092 in Denmark to 0.807 in southwest Spain. Mean allelic richness varied from 1.25 in Denmark to 7.05 in southwest Spain. After Bonferroni corrections, there were no cases of linkage disequilibrium among the loci, but 16 out of 168 population × loci comparisons showed significant deviations from Hardy–Weinberg equilibrium. All but one of these deviations were of homozygote excess, and 12 of the 16 were in just three populations (six in southwest Spain, three in northern France (Cherbourg) and three in Germany (Halle)). Despite our best efforts, it may be that sampling at these sites was biased in favour

Table 1
Mitochondrial DNA genetic diversity estimates

	Southern Spain	Northern Spain	Northeast Spain + southern France	Northern Europe	Total (all sites)
Sample size	7	7	45	61	120
No. haplotypes found	7	6	9	4	25
Gene (haplotype)	1.000	0.952	0.706	0.337	0.781
diversity	(0.076)	(0.100)	(0.068)	(0.070)	(0.033)
Nucleotide diversity	0.028	0.032	0.027	0.001	0.031
	(0.017)	(0.019)	(0.014)	(0.001)	(0.155)
Tajima's D	0.311	−0.506	0.878	−0.888	−0.316
P value of D (random < observed)	0.619	0.355	0.803	0.199	0.399

Southern Spain = site 1; northern Spain = site 2; NE Spain and Southern France = sites 4, 5, and 6; northern Europe = sites 9, 11, 12, 13, 15, 20, and 21. Standard deviations for diversity estimates are given in parentheses.

of excess sibs. As shown in Fig. 3A, genetic diversity as measured by mean expected heterozygosity across loci, or as allelic richness, decreased with geographic distance from southwest Spain. Geographical distances were direct (linear) measurements between sampling sites, except that distances into and out of Iberia were all pivoted through a point on the Mediterranean coast immediately east of the Pyrenees. This is the only feasible migration route in and out of Iberia for natterjack toads, which cannot survive at the high altitudes that span the entire length of the Pyrenean mountain range. Apart from this, no other account was taken of likely postglacial colonisation routes, which could only be guessed. The Irish population, highlighted in Fig. 3A, was very close to expectations according to the regression line. Allelic richness correlated strongly with expected heterozygosity ($r_s = 0.962$, $P < 0.0001$). Both measures of genetic diversity correlated strongly, and inversely, with geographical distance measured from southwest Spain (for heterozygosity, $r_s = -0.906$, $P < 0.0001$; and for allelic diversity $r_s = -0.856$, $P < 0.0001$).

There was also an interesting pattern of variation in mean allele (repeat number) size in the *B. calamita* microsatellites across the biogeographical range. As shown in Fig. 3B, mean allele size (repeat numbers) across the eight loci tended to increase with distance from Iberia. For this calculation, repeat numbers for all loci were adjusted by appropriate correction factors to a common mean, in order to remove bias from loci with relatively large or small numbers of alleles or repeats. The overall correlation was significant ($r_s = 0.483$, $P = 0.027$), but this ascertainment bias was not consistent among all eight loci. Four loci (*Bcalμ*1, 2, 3, and 8) showed strong positive correlations, two (*Bcalμ*5 and 6) showed negative correlations, and two (*Bcalμ*4 and 7) exhibited no significant trend. There was a positive correlation between the total number of alleles detected at a locus, which varied from 8 in *Bcalμ*2 to 30 in *Bcalμ*8, and the mean number of repeats per allele ($r_s = 0.890$, $P = 0.007$). However, there was no clear relationship between ascertainment bias and either mean allele size or the number of alleles detected at a locus.

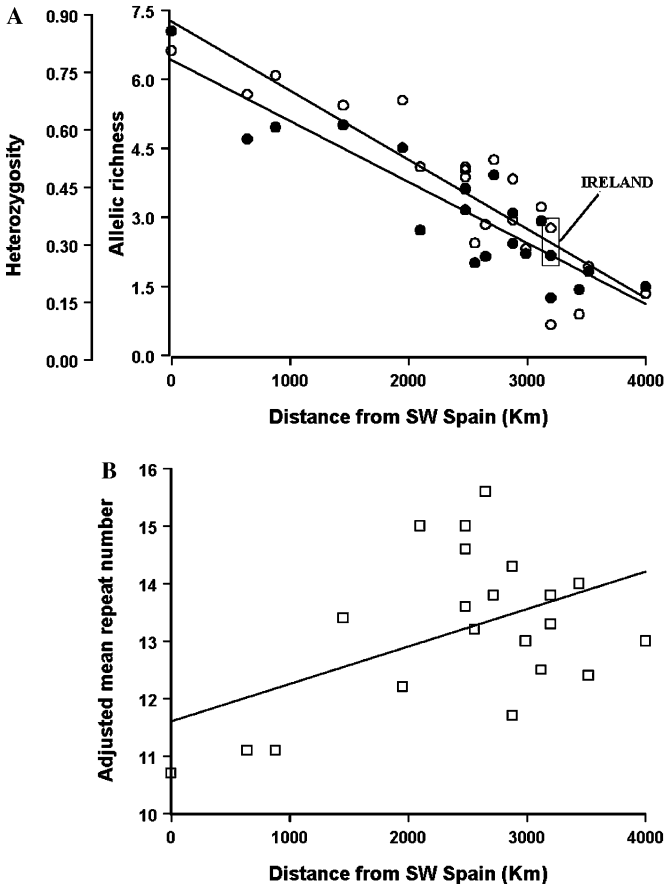

Fig. 3. (A) Genetic diversity and distance from the Iberian refugium. (○) mean heterozygosity; (●) mean allelic richness. (B) Mean allele repeat number increases with distance from Iberia. Each point represents a population sample.

Fig. 4 shows a phylogenetic tree derived from the microsatellite allele frequencies. Although many internal nodes were only weakly supported by bootstrap analysis, some groups were robust and in general there was strong overall similarity with the mtDNA phylogeny. In particular, the trees agreed in demonstrating lineages in Spain and southern France that were distinct from those in most of northern Europe. The robust branch points included a Spain, southeast France and Switzerland clade, a northwest England and Ireland clade, and a Scandinavian clade. Natterjacks from southeast England were not closely linked to those in the northwest England–Ireland clade. In a principal component analysis (PCA) of the allele frequency data, 111 alleles contributed significantly to 12 principal components, the first three of which accounted for 55.8% of the total variance. Interpolation maps of principal components reflect historical events superimposed upon geography (Cavalli-Sforza et al., 2000; Hanotte et al., 2000). Fig. 5 shows an interpolation map of the first PC, accounting for 27.2% of total variance. The second PC (not shown), accounting for 16.3% of total variance, also produced a pattern radiating from Iberia towards northern and eastern Europe.

The results of the microsatellite IM analysis, with maximum likelihood estimates of population parameters, are shown in Table 3. From the t estimates it is possible to obtain estimates of the divergence time (T), in generations, from the relationship $t = T\mu$ where μ is the mutation rate per generation (Hey and Nielsen, 2004). Similarly, mean effective population size (N_e) over the intervening time period since divergence can be estimated from $N_e = \theta/4\mu$. The generation time of *B. calamita* has been estimated from demographic data to be around 3.5–4 years (Rowe and Beebee, 2004). However, we also needed a calibrated estimate of μ, which was obtained using the divergence between populations in Ireland and northwest England. This must have happened either in the narrow time window of about 1000 years, centred approximately 10,000 years ago between the end of the Younger Dryas and the sea level rise which irreversibly separated the two localities; or

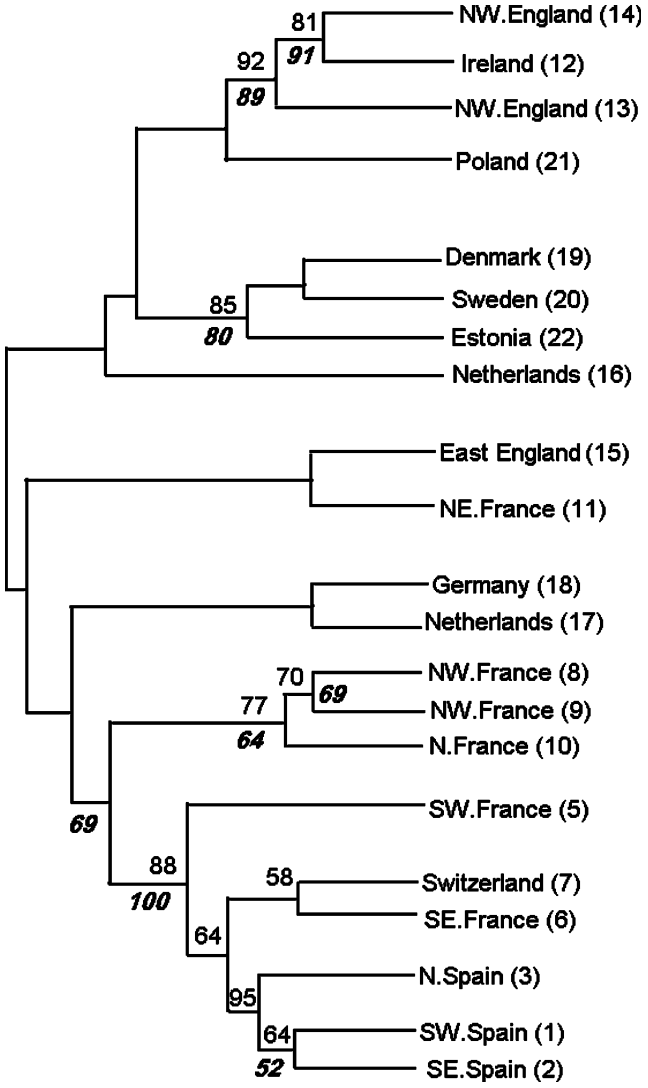

Fig. 4. Phylogenetic tree based on microsatellite data. Bootstrap values (1000 replicates) >50% are shown for the neighbour-joining method (above nodes) and for maximum likelihood analysis (45 replicates) in italics below nodes. Numbers refer to sampling sites shown in Fig. 1.

alternatively, as a result of human translocation which must have occurred, if at all, within the past 1000 years (and probably within the last 500 years). There is no evidence of significant human movements between these two regions earlier than 1000 years BP. However, assuming (as a control) that the divergence time between southeast England and Ireland must have been at least 9500 years BP, before intervening sea level rises, the recent translocation hypothesis would require a t ratio (southeast England–Ireland:northwest England–Ireland of $\geqslant$10:1). The actual ratio of about 1.9:1 (Table 3) is not compatible with this scenario, so we calibrated the mutation rate on the basis of a natural divergence around 9500 years BP. From $t = T\mu$ with $T = 9500/4$ generations, $\mu = \approx 1 \times 10^{-5}$. The other estimates of divergence times in years, and of effective population sizes, were calculated on this basis as shown in Table 3.

These analyses suggest that Ireland and southeast England probably last shared a common ancestral population some 15,000–20,000 years BP, far longer ago than the common ancestors of the Ireland and northwest England populations. Irish connections with northern Spain are at least 30,000 years BP. The time estimates for the three population splits of primary interest (northwest England × Ireland, southeast England × Ireland and Spain × Ireland) were significantly different (Kruskal–Wallis one-way ANOVA statistic = 12.02, $P = 0.0025$). By contrast, there was no significant difference in the split times of southeast England × Ireland and southeast England × northwest England (Wilcoxon signed-rank test, $P = 0.2807$), confirming that southeast England natterjacks are equally different from those in Ireland and in northwest England. Estimates of effective population sizes were in the low thousands for all sites in the British Isles, in the order northwest England > southeast England > Ireland.

4. Discussion

This study is the first to apply phylogeographical analysis to the problem of Lusitanian distributions. The results presented in this paper also expand substantially on a preliminary study of *B. calamita* phylogeography that used only a relatively small set of microsatellite data and no mtDNA sequences (Beebee and Rowe, 2000). As far as we know, our analysis also provides the first estimate of microsatellite mutation rates in amphibians. Of course our results only relate to one Lusitanian species, and it will be important to make comparative studies with others that share this curious biogeographic pattern. The mtDNA analysis provided information about the relatively distant past, up to around 1 Myr BP. The large divergence from other clades, wide distribution, and low nucleotide diversity of clade A across northern and central Europe, together with its "star-like" phylogeny of derived haplotypes, infer: (1) a mid-Pleistocene origin, from Iberian stock, and continuous survival north of the Pyrenees and (2) the occurrence of bottlenecking and radiation events prior to the last glacial maximum, indicated by the patterns of recently derived clade A haplotypes in different parts of northern Europe. The distribution of clade A haplotypes also suggests some relatively recent gene flow into southern France and Iberia.

The microsatellite data give indications of more recent events, especially local radiations within the past 10–50,000 years. However, the inference of common ancestry for Ireland and northern Spain around 30–40,000 years BP based on microsatellite data is much more recent than the divergence of the different mtDNA clades now dominant in these regions. Substantial discord between mtDNA and microsatellite inferences has been reported previously, including with amphibians (e.g., Lu et al., 2001; Monsen and Blouin, 2003), and can have several causes. Sex-biased dispersal, in which females are much more philopatric than males, can sometimes provide a convincing explanation

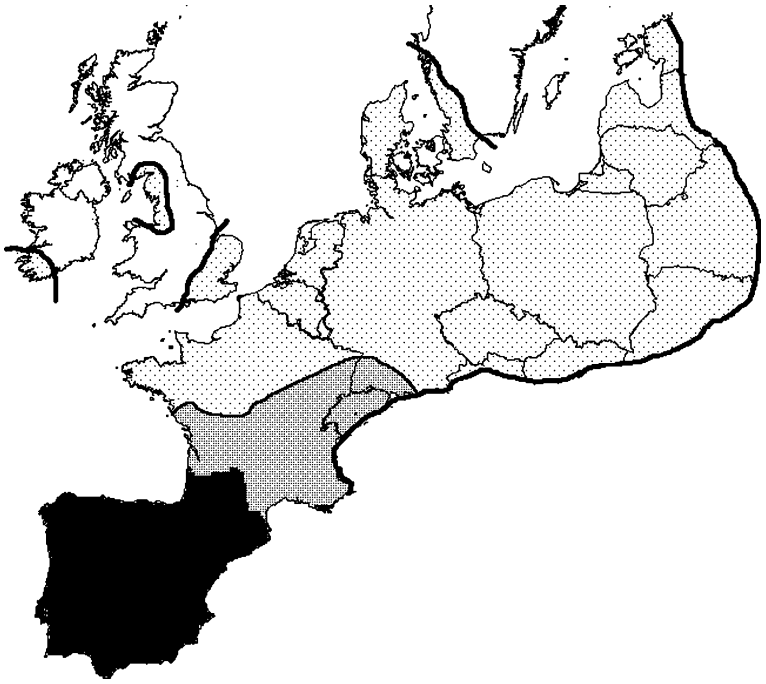

Fig. 5. Interpolation map showing geographic variation of the first principal component of microsatellite allele frequency variation. The different shadings represent three equally sized classes of interpolation surfaces, ranging from the region of maximum effect (dark shading) to the region of minimum effect (light stippling) of this component. Bold lines indicate approximate range limits (thick) and borders between the interpolation surfaces (thin).

Table 3
Maximum likelihood estimates (MLE) of demographic parameters

Comparison	θ_1	θ_2	θ_A	t	N_1	N_2	N_A	Divergence time (years)
North Spain × Ireland								
MLE	0.617	0.108	19.852	0.091	15,425	2700	496,300	36,400
95% CI L	0.516	0.059	18.394	0.072	12,900	1475	459,850	28,800
95% CI U	0.719	0.156	21.311	0.110	17,975	3900	532,775	44,000
Southeast England × Ireland								
MLE	0.120	0.106	14.980	0.044	3000	2650	374,500	17,600
95% CI L	0.086	0.080	11.398	0.034	2150	2000	284,950	13,600
95% CI U	0.154	0.131	18.562	0.055	3850	3275	464,050	22,000
Northwest England × Ireland								
MLE	0.231	0.072	4.294	0.023	5775	1800	107,350	(9500)
95% CI L	0.168	0.018	3.555	0.014	4200	450	88,875	(5600)
95% CI U	0.294	0.125	5.034	0.032	7350	3125	125,850	(12,800)
Northwest England × southeast England								
MLE	0.225	0.130	20.265	0.053	5625	3250	506,625	21,200
95% CI L	0.186	0.111	18.574	0.043	4650	2775	464,350	17,200
95% CI U	0.265	0.149	21.956	0.062	6625	3725	548,900	24,800

Lower (L) and upper (U) 95% confidence intervals (CI) of estimates from five independent runs of the IM program for each population pair comparison are shown. θ_1, θ_2, and θ_A are the ML estimates of population size parameters for the first and second named populations in each pair, and for the ancestral population, respectively. t is the ML estimator of the divergence time parameter. N_1, N_2, N_A, and T are the corresponding effective population size estimates and divergence time estimate (BP) in years. The divergence time for northwest England × Ireland is in parentheses because it was derived as the calibration point (see Section 3).

(e.g., Fitzsimmons et al., 1997), but this is unlikely in *B. calamita* because females of this species disperse more widely than males (Sinsch and Seidel, 1995). However, there is evidence of relatively recent but limited directional gene flow southwards from populations north of the Pyrenees as discussed above. Given such gene flow, differences in marker characteristics (Buonaccorsi et al., 2001), including N_e, mutation rates, and higher probabilities of homoplasy in microsatellite alleles relative to mtDNA haplotypes may all contribute to the differences in divergence times that we have estimated. The IM program also assumed no migration between populations due to intervening sea, but in the early stages of separation in mainland Europe between what were to become the Irish and

Table 4
Irish colonisation models

Model	Timescale	Origin	MtDNA data compatibility	Microsatellite divergence time compatibilities	
				Requirement	Observation
Ancient human translocation	2–5000 years BP	Spain	Highly unlikely	t Ireland/Spain $<$ t Ireland/SE England	t Ireland/Spain $\gg$ t Ireland/SE England. Incompatible
Recent human translocation	0–1000 years BP	NW England	Compatible	t Ireland/SE England $\geqslant$ $10 \times t$ Ireland/NW England	t Ireland/SE England $=$ c. $1.9 \times t$ Ireland/NW England Incompatible
Natural colonisation via France and England	9–10,000 years BP	Spain	Highly unlikely	t Ireland/SE England approximately $=$ t Ireland/NW England	t Ireland/SE England $=$ c. $1.9 \times t$ Ireland/NW England Incompatible
Natural colonisation from nearby refuge	9–10,000 years BP	Western approaches	Compatible	t Ireland/NW England $<$ t Ireland/SE England	t Ireland/NW England $<$ t Ireland/SW England Compatible

Spanish populations there must have been some unquantifiable opportunities for migration. Any such migration would set back the initial divergence time.

We can now evaluate with genetic data the relative probabilities of the four possible explanations of the origin of Irish natterjack toads listed in Table 4.

(1) Translocation of *B. calamita* from Iberia to Ireland in ancient historical times is a highly unlikely explanation of the current distribution pattern. Irish natterjacks are fixed for the same mitochondrial "north Europe" haplotype as those in Britain. This most abundant clade A haplotype is rare in Iberia and is highly divergent from the other clades (C, D, and E) found there (Fig. 2). The genealogical discontinuity found in southern France between the highly divergent mitochondrial clades A and B strongly suggests that these clades evolved in allopatry with recent secondary contact, as described above. ML estimates of divergence times based on nuclear locus (microsatellite) data also indicate that the common ancestral population of Spanish and Irish toads existed >30,000 years BP.

(2) The fixation of a single mtDNA haplotype (Figs. 1 and 2) throughout Britain and Ireland makes the mtDNA analysis uninformative with respect to the recent translocation hypothesis. The microsatellite phylogeny infers that northwest England would be the only feasible origin of such a translocation. However, recent translocation of natterjacks from northwest England to Ireland is incompatible with the divergence time of Irish and northwest England populations, relative to the divergence time of Irish and southeast England populations, based on the phylogenetic and ML analyses of microsatellite data (Fig. 4 and Table 3). These two divergences are relatively much too close together in time for the recent introduction hypothesis to be credible.

(3) Postglacial colonisation of southern Ireland from mainland Europe, via land bridges through southern Britain, is essentially discounted by both the mtDNA and the microsatellite data. A highly improbable lineage sorting would be required in this case to explain the dominance in northern Europe, including the British Isles, of a mtDNA clade apparently very rare in Iberia. Populations in northwest England and Ireland are closely related to each other (Fig. 4) but very different from those in southern and eastern England on the basis of the microsatellite data (see also Rowe et al., 1998). ML analysis confirmed a substantially earlier split between the southeast England × Ireland populations than for the northwest England × Ireland populations. However, the very short time window (probably <1000 years) available for natterjacks to reach Ireland from Britain in the post Younger Dryas warming period, before rising sea levels created a permanent physical separation, would require essentially simultaneous split times for these population pairs if colonisation of Ireland occurred via England.

(4) Although the mtDNA is not specifically informative in this case, our results do however support the fourth hypothesis as a possible explanation for the natterjack's Lusitanian distribution. ML estimates of divergence times, together with the microsatellite phylogeny robustly linking Irish populations to those in northwest England, are consistent with a common source for these two areas in the warming period immediately following the Younger Dryas around 10,000 years BP. The most likely geographical situation of such a source population is in the area between Ireland and southwest England, which was partly dry land at the time of the Younger Dryas (Devoy, 1985). Natterjacks in eastern England must have recolonised quite separately after the Younger Dryas, presumably from a southeasterly direction (see Fig. 6). Separate expansions are also indicated elsewhere (e.g., Scandinavia) from the microsatellite data.

More generally, our results can be interpreted as shown in Fig. 6 with respect to the phylogeographic history of *B. calamita* in Europe. The high level of genetic diversity

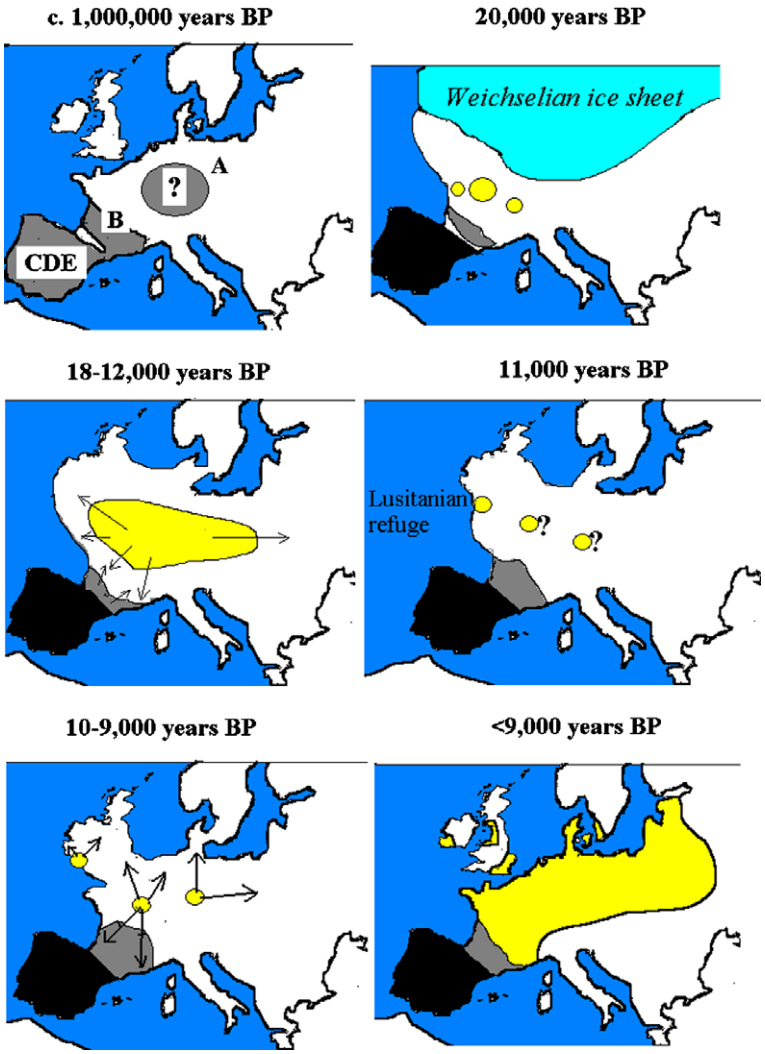

Fig. 6. Proposed phylogeographic history of *B. calamita* in Europe. (A–E) mtDNA clades; (black shading) Iberian refuge (clades C, D, and E); (grey shading) clade B; (yellow shading) clade A. Arrows indicate arbitrary colonisation patterns and question marks (?) indicate major uncertainties in details of distributions.

in Iberia as well as the low mean microsatellite allele sizes there indicate a long term (probably millions of years) refuge in this region, in which the "stepwise increase and occasional collapse" cycle of mutational change in microsatellite alleles (Schlotterer, 2000) is close to equilibrium. Low mtDNA haplotype and nucleotide diversity, decreasing microsatellite heterozygosity, increased microsatellite allele size and the patterns of allele frequency variations revealed by principal components analysis in the rest of Europe all support a pattern of colonisation originating in Iberia. Decreasing genetic diversity as a function of distance from a refugium is a well-known expectation of the sequential generation of leptokurtic distributions during colonisation processes (Ibrahim et al., 1996). Directional evolution for increased microsatellite size in geographically derived regions, generating intraspecific ascertainment bias, has also been shown for maize (Vigouroux et al., 2003), but most studies of ascertainment bias have been on interspecific comparisons. It will be interesting to see whether this is a common signature of ancient population expansions in other species.

However, the data do not support a simple, single recent expansion from Iberia following the end of the last (Weichselian) glacial maximum around 20,000 years BP. The divergence date for mitochondrial clades A and B was approximately mid-Pleistocene (ca. 1.22 Myr BP, Table 2), and it seems very likely that clade A haplotypes remained isolated in at least one north European refuge, perhaps in France, throughout the whole of the Weichselian glacial maximum. Rapid warming after this glacial maximum around 20,000 years BP would have permitted population expansions from these non-Iberian refugia, interrupted by the Younger Dryas climatic cooling (ca. 11,000 years ago), which was much less severe than the previous glacial maximum. The microsatellite data clearly indicate such local expansions (Fig. 4). The Lusitanian aspect of the natterjack toad's distribution very likely relates to one of these expansions into Ireland and

northwest England. A long-term Lusitanian refuge near Ireland throughout all the Pleistocene glaciations, however, is highly improbable. The "north Europe" mitochondrial clade A diverged only within the last 1.0–1.5 Myr, while the Pleistocene cooling commenced between 2.0 and 2.5 Myr ago (Vincent, 1990). Furthermore, the extreme cold conditions at the glacial maxima would not have permitted continuous survival of *B. calamita* in or near Ireland.

The natterjack toad, an ectothermic terrestrial vertebrate, therefore must have survived north of the Iberian peninsula throughout the period of the last glacial maximum with local population expansions in the early postglacial. In North America recent evidence suggests that endotherms (chipmunks, Tamias striatus) also survived much further north during the recent glaciation than previously thought (Rowe et al., 2004). Even during the Weichselian glacial maximum, permafrost was discontinuous throughout most of northern France (Renssen and Vandenberghe, 2003) and communities surviving in this region included species of freshwater fish (Hanfling et al., 2002). A similar pattern of northern clade survival is also evident in another amphibian, the pool frog *Rana lessonae*, though in this case not with an Irish connection (Zeisset and Beebee, 2001). In the case of *B. calamita* it appears that a modified version of the oldest Lusitanian hypothesis (Forbes, 1846), a local but relatively recent and short-lived refuge, is the most likely to be correct.

Acknowledgments

We thank H. Burton and M. Cross for labwork assistance, C. Andren, J. Arntzen, H. Bringsoe, J. Buckley, M. Busko, B. Le Garff, W. Grose, P. Pappel, P. Ramseier, M. Stoeck, and H. Strijbosch for help with sample collection, T. Browne for assistance with ARCVIEW mapping, R. Thorpe for helpful suggestions, J. Hey and M. Beaumont for advice concerning the IM program, and the Leverhulme Trust and Natural Environment Research Council for financial support.

References

Arnason, U., Gullberg, A., Janke, A., Joss, J., Elmerot, C., 2004. Mitogenomic analyses of deep gnathostome divergences: a fish is a fish. Gene 333, 61–70.

Avise, J.C., 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, MA.

Beebee, T.J.C., 1983. The Natterjack Toad. Oxford University Press, Oxford.

Beebee, T.J.C., Rowe, G., 2000. Microsatellite analysis of *Bufo calamita* Laurenti populations: consequences of dispersal from a Pleistocene refugium. Biol. J. Linn. Soc. 69, 367–381.

Buonaccorsi, V.P., McDowell, J.R., Graves, J.E., 2001. Reconciling patterns of inter-ocean molecular variance from four classes of molecular markers in blue marlin (*Makaira nigricans*). Mol. Ecol. 10, 1179–1196.

Cavalli-Sforza, L.L., Menozzi, P., Piazza, A., 2000. The History and Geography of Human Genes. Princeton University Press, New Jersey.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1660.

Corbet, G.B., 1961. Origin of the British insular races of small mammals and of the 'Lusitanian' fauna. Nature 191, 1037–1040.

Corbet, G.B., 1962. The Lusitanian element in the British fauna. Sci. Prog. 50, 177–191.

Devoy, R.J., 1985. The problem of Late Quaternary land bridges between Britain and Ireland. Quat. Sci. Rev. 4, 43–58.

Felsenstein, J., 1993. PHYLIP (Phylogeny inference package) Version 3.5c. Dept. of Genetics, University of Washington, Seattle.

Fitzsimmons, N.N., Moritz, C., Limpus, C.J., Pope, L., Prince, R., 1997. Geographic structure of mitochondrial and nuclear gene polymorphisms in Australian green turtle populations and male-biased gene flow. Genetics 147, 1843–1854.

Forbes, E., 1846. On the connexion between the distribution of the existing fauna and flora of the British Isles, and the geological changes which have affected their area, especially during the epoch of the northern drift. Mem. Geol. Survey Great Br. 1, 336–432.

Goudet, J., 1995. FSTAT (Vers. 1.2): a computer program to calculate F-statistics. J. Hered. 86, 485–486.

Hanfling, B., Hellemans, B., Volckaert, F.A.M., Carvalho, R., 2002. Late glacial history of the cold-adapted freshwater fish *Cottus gobio*, revealed by microsatellites. Mol. Ecol. 11, 1717–1729.

Hanotte, O., Bradley, D.G., Ochieng, J.W., Verjee, Y., Hill, E.W., Rege, J.E., 2000. African pastoralism: genetic imprints of origins and migrations. Science 296, 336–339.

Hewitt, G., 2000. The genetic legacy of the ice ages. Nature 405, 907–913.

Hewitt, G.M., 2001. Speciation, hybrid zones and phylogeography-or seeing genes in space and time. Mol. Ecol. 10, 537–549.

Hey, J., Nielsen, R., 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167, 747–760.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Ibrahim, K.M., Nichols, R.A., Hewitt, G.M., 1996. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity 77, 282–291.

Liu, W., Lathrop, A., Fu, J.Z., Yang, D.T., Murphy, R.W., 2000. Phylogeny of East Asian bufonids inferred from mitochondrial DNA sequences (Anura: amphibia). Mol. Phylogenet. Evol. 14, 423–435.

Lu, G., Basley, D.J., Bernatchez, L., 2001. Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridisation between lineages of lake whitefish (*Coregonus clupeaformis*): relevance for speciation. Mol. Ecol. 10, 965–985.

Macey, J.R., Schulte, J.A., Larson, A., Fang, Z.L., Wang, Y.Z., Tuniyev, B.S., Papenfuss, T.J., 1998. Phylogenetic relationships of toads in the *Bufo bufo* species group from the eastern escarpment of the Tibetan plateau: a case of vicariance and dispersal. Mol. Phylogenet. Evol. 9, 80–87.

Monsen, K.J., Blouin, M.S., 2003. Genetic structure in a montane ranid frog: restricted gene flow and nuclear-mitochondrial discordance. Mol. Ecol. 12, 3275–3286.

O'Rourke, F.J., 1970. The Fauna of Ireland. Mercier Press, Cork.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Praeger, R.L., 1939. The relations of the flora and fauna of Ireland to those of other countries. Proc. Linn. Soc. 151, 192–213.

Preece, R.C., Coxon, P., Robinson, J.E., 1986. New biostratigraphic evidence of the post-glacial colonisation of Ireland and for Mesolithic forest disturbance. J. Biogeog. 13, 487–509.

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

Renssen, H., Vandenberghe, J., 2003. Investigation of the relationship between Permafrost distribution in NW Europe and extensive sea ice cover in the North Atlantic Ocean during the cold phases of the last glaciation. Quat. Sci. Rev. 22, 209–223.

Rowe, G., Beebee, T.J.C., 2004. Reconciling genetic and demographic estimates of effective population size in the anuran amphibian *Bufo calamita*. Conserv. Genet. 5, 287–298.

Rowe, G., Beebee, T.J.C., Burke, T., 1997. PCR primers for polymorphic microsatellite loci in the anuran amphibian *Bufo calamita*. Mol. Ecol. 6, 401–402.

Rowe, G., Beebee, T.J.C., Burke, T., 1998. Phylogeography of the natterjack toad *Bufo calamita* in Britain: genetic differentiation of native and translocated populations. Mol. Ecol. 7, 751–760.

Rowe, K.C., Heske, E.J., Brown, P.W., Paige, K.N., 2004. Surviving the ice: northern refugia and postglacial colonisation. Proc. Natl. Acad. Sci. USA 101, 10355–10359.

Schlotterer, C., 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma 109, 365–371.

Schneider, S., Roseli, D., Excoffier, L., 1997. ARLEQUIN ver 11: a software for population genetic analysis. Genetics Biometry Laboratory, University of Geneva.

Sinsch, U., Seidel, D., 1995. Dynamics of local and temporal breeding assemblages in a *Bufo calamita* metapopulation. Aust. J. Ecol. 20, 351–361.

Swofford, D.L., 2002. PAUP: Phylogenetic Analysis Using Parsimony (and Other Methods). Sinauer, Sunderland.

Taberlet, P., Fumagalli, L., Wust-Saucy, G., Cossons, J-F., 1998. Comparative phylogeography and postglacial colonisation routes in Europe. Mol. Ecol. 7, 453–464.

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

Takezaki, N., Rzhetsky, A., Nei, M., 1995. Phylogenetic test of the molecular clock and linearised trees. Mol. Biol. Evol. 12, 823–833.

Vincent, P., 1990. The Biogeography of the British Isles. Routledge, London.

Vigouroux, Y., Matsuoka, Y., Doebley, J., 2003. Directional evolution for microsatellite size in maize. Mol. Biol. Evol. 20, 1480–1483.

Walsh, P.G., Metzger, D.A., Higuchi, R., 1991. Chelex R100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10, 506–513.

Webb, D.A., 1983. The flora of Ireland in its European context. J. Life Sci. Roy. Dublin Soc. 4, 143–160.

Zeisset, I., Beebee, T.J.C., 2001. Determination of biogeographical range: an application of molecular phylogeography to the European pool frog *Rana lessonae*. Proc. Roy. Soc. Lond. B 268, 933–938.