Molecular Ecology (2007) **16**, 785–796 doi: 10.1111/j.1365-294X.2006.03188.x

Defining population boundaries: use of three Bayesian approaches with microsatellite data from British natterjack toads (*Bufo calamita*)

G. ROWE*† and T. J. C. BEEBEE*
*School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK, †Biological Sciences Research Group, School of Science, University of Derby, Kedleston Road, Derby DE22 1GB, UK

Abstract

Defining boundaries between populations is often difficult in the absence of information about current levels of gene flow. Such definitions can be important, however, both for the understanding of population dynamics and for conservation planning. Recently developed Bayesian methods for analysing genetic data now provide a powerful approach to this problem. Natterjack toads *Bufo calamita* are endangered in Britain, where their distribution is restricted to four geographically discrete regions. In three of these regions the boundaries between populations are often uncertain. We therefore used Bayesian approaches with microsatellite data to try and define British natterjack population structure, and thus inform conservation management. A large sample of natterjack toads from all 38 locations in Britain where the species is native was genotyped at eight microsatellite loci. The genetic diversity of natterjack populations declined as a function of increasing latitude, echoing postglacial colonization dynamics. Comparisons of three assignment methods (STRUCTURE, BAPS and GENELAND) generated some broad similarities but also some inconsistencies in the definitions of population structure, especially in the most complex region (south Cumbria). Implications of the analyses for the future conservation of *Bufo calamita* in Britain are discussed.

Keywords: Bayesian, *Bufo calamita*, microsatellites, Population definition

Received 9 July 2006; revision received 1 September 2006; accepted 28 September 2006

Introduction

Defining the geographical boundaries of animal and plant populations is fraught with difficulty because, as with definitions of species (but even more so), there are often areas of indeterminacy at prospective borders. Natural populations generally have complex substructures with variable rates of gene flow between demes, and in many instances (particularly for widespread species) there may be no completely discrete separation of populations over large geographical distances (e.g. Brede & Beebee 2004). However for rare species, and for widespread species near their range margins, genetically separate populations may occur. It will be increasingly useful for conservation biologists to identify discrete populations because, especially if small, they may be at high risk of extinction from demographic, environmental and genetic stochasticity. Assignment methods using genetic data now offer new possibilities for defining populations of this kind (Manel *et al*. 2005), especially where boundaries between them and the permeability of intervening habitat are unclear. Here we describe the use of assignment methods to define populations near the range margin of an amphibian, the natterjack toad *Bufo calamita*.

Extensive amphibian declines are causing widespread concern around the world, especially because for many species the primary causes of decrease are unclear (Stuart *et al*. 2004). In Europe, amphibian declines have a long history and were already severe by the mid-20th century (Houlahan *et al*. 2000). In most of these cases the causes are reasonably well known, and relate primarily to habitat alteration or destruction (Beebee & Griffiths 2005). The natterjack toad *B. calamita* is a European species at the edge of its biogeographical range in Britain, where its distribution has always been restricted due to a requirement for particular habitats, mostly coastal dunes and lowland heaths

Correspondence: Prof T.J.C. Beebee, E-mail: t.j.c.beebee@sussex.ac.uk

(Beebee 1979). Natterjacks declined dramatically in Britain during the early mid 20th century, disappearing from 70 to 80% of previously occupied sites as a consequence of habitat change or loss (Beebee 1976, 1977). *Bufo calamita* now benefits from a high level of legal protection in Britain and has been the subject of a species recovery programme (Denton *et al*. 1997).

Largely because of its rarity and conservation interest, *B. calamita* is the best documented British amphibian with respect to distribution and abundance (Buckley & Beebee 2004). Survey and monitoring over the past 30 years have identified 38 currently extant, native 'populations' (which in many cases just represent well known localities for the species) as described in the National Site Register for *B. calamita* (Beebee & Buckley 2001). The last of these populations was discovered in 1993, and it is likely that the current distribution of natterjacks in Britain is now fully known. This distribution can be segregated into four distinct geographical regions, notably east/southeast England where natterjacks form a distinct clade (Rowe *et al*. 1998), Merseyside, South Cumbria and Solway (Fig. 1). The five populations in east/southeast England are clearly distinct because they are separated by at least tens of kilometres of unsuitable habitat. The 'Syderstone' samples in this region were actually taken from Sandy, a new site in eastern England where Syderstone animals were translocated in 1980 before habitat degradation caused the extinction of the Syderstone population.

By contrast, the populations identified in the Site Register for Merseyside ($n = 5$), south Cumbria ($n = 19$) and Solway ($n = 9$) are more complex, often with vague boundaries between what might really be interconnected sampling regions rather than truly separate populations. The primary purpose of this study was to use genetic information, rather than just the historic sampling sites listed in the Site Register, in an attempt to define better the number and distribution of discrete natterjack populations in these three areas of Britain. This in turn should generate a more biologically meaningful picture of natterjack distribution and population dynamics than hitherto possible, and also prove useful in strategic planning for conservation. As far as we know, this is the first attempt to assign all the populations of any species on genetic grounds at a national scale. We used three similar but distinctive methods, all with recently developed computer programs, to address this question.

Materials and methods

Sampling and genotyping

Most of the sample collection (1415 larvae from 36 sites) occurred during 1994 and 1995, followed by genotyping at eight microsatellite loci and phylogeographical analysis as

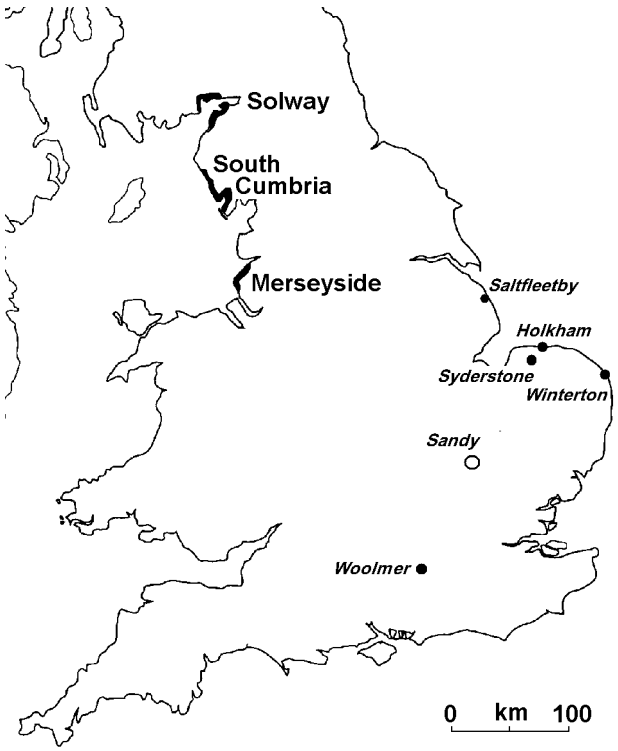

Fig. 1 Distribution of *Bufo calamita* in Britain (native sites). Open circle = Sandy, where a population derived from Syderstone was sampled.

described earlier (Rowe *et al*. 1998). Forty individuals were sampled at each site, with the following exceptions: Altcar (38), Green Road (38) and Grune (19). However, two populations from which larvae were not available in that period (Annaside and Eskmeals in South Cumbria) were sampled and genotyped in 2005. Nineteen larvae were taken at Annaside, but only eight adult toads were found at Eskmeals and in this case toe clips were used as the source of DNA. This paper therefore concerns further analysis of a largely pre-existent data set (Rowe *et al*. 1998) using analytical procedures that have only recently become available.

Data analysis

Exact tests for conformance to Hardy–Weinberg equilibrium and analysis of isolation by distance were carried out using GENEPOP 3.4 (Raymond & Rousset 1995) using data from each of the 38 separate sampling sites. Geographical distances between sites were based on a priori autecological information (Rowe *et al*. 2000) rather than direct point-to-point measures, and our initial estimates of these distances were not subsequently modified to fit the genetic data. We assumed that toads would move along the coastlines (i.e. within mostly suitable habitat) and not traverse mountains, and also that they could cross shallow parts of strongly tidal

estuaries at low tide (notably the Duddon in south Cumbria, and the eastern Solway). Mean expected heterozygosity and pairwise F_{ST} estimations of population differentiation (200 permutations) for the 38 sites were performed using FSTAT 2.9.3 (Goudet 1995). The possibilities of recent population bottlenecks were investigated using BOTTLENECK (Cornuet & Luikart 1996), again for all 38 sites, using the two-phase model with 90% stepwise mutation. Significance was assessed using the mode-shift of allele frequencies criterion (weak evidence of a recent bottleneck) and by the more rigorous Wilcoxon test.

We made an initial comparison of genetic structures within and among the four sampling regions using analysis of molecular variance, AMOVA (Schneider *et al*. 2000), in ARLEQUIN version 2.0. Both F_{ST} estimates and AMOVA require populations to be defined in advance. We then analysed current population structure in the four regions by three methods, all based on Bayesian approaches, which permit populations to be defined by the genetic data. Firstly we used STRUCTURE version 2 (Pritchard *et al*. 2000) with a burn-in of 5×10^4 and with 10^6 iterations, and not using prior population information. In its standard mode STRUCTURE uses genotype data alone, in the absence of sample site information, to infer population clusters. True population number (K) estimates were made from a minimum of one to a maximum of 10 values for K, and always triplicated for each K-value. Separate sets of runs were carried out assuming either correlated or uncorrelated alleles. The correlated alleles model assumes that frequencies in the different populations are likely to be similar, either due to migration or shared ancestry. Admixture was excluded in the analyses of East/southeast England populations, but included in analyses of the other three regions. Means of the log probabilities were used to determine the most likely true K, and probability of the most likely true K was estimated using Bayes's rule:

$$P = \frac{e^a}{e^a + e^b + e^c}$$

Where a, b, c, etc. are the ln-probabilities for each value of K. Populations were ascribed to a particular STRUCTURE group when the highest proportion of the sampled individuals were in that group. As recommended by the authors, we adopted a hierarchical approach starting with the full set of UK sampling sites ($n = 38$) and further analysed each proposed cluster until no further subdivision was indicated in each cluster (i.e. $K = 1$). We also carried out STRUCTURE analyses using prior information, i.e. with USEPOPINFO = 1. Secondly we used BAPS 3.1 (Corander *et al*. 2003) in both individual and group analysis modes, with duplicate runs for each region. BAPS uses a stochastic optimization to infer the posterior mode of genetic structures, runs much faster than STRUCTURE or GENELAND (both of which use Markov chain Monte Carlo algorithms) and does not require multiple trials for each possible value of K. It uses both genotype and, in group mode, sample group information to infer population clusters. Maximum numbers of populations (prior information) were 10 for East/southeast England and Merseyside, and 25 for South Cumbria and Solway. Alterations to this prior over a wide range (from five to 200) did not affect the results. Although BAPS gives the most probable cluster arrangement, probabilities of the best clusters were also estimated using Bayes's rule by comparison with less probable clusters. Thirdly, we used GENELAND (Guillot *et al*. 2005) in the R-PACKAGE (Ihaka & Gentleman 1996). GENELAND uses genotype data together with geographical information (the locations in which individuals were sampled) to estimate population structure. Following the approach of Coulon *et al*. (2006), we first allowed K (the number of populations in each cluster) to vary and inferred the most probable K using five replicates with 5×10^5 Markov chain Monte Carlo iterations, the maximum rate of the Poisson process fixed at 500, the maximum number of nuclei in the Poisson-Voronoi tessellation fixed at 200, and the Dirichlet model for allele frequencies. In the five replicates for each of the four regions there was almost no variation in the median inferred K (seven for east/southeast, six for Merseyside, seven for south Cumbria and 10 for Solway), and we used these numbers in part two of the analysis. For this we included spatial information from the sampling sites, with K invariant at the value derived in the first round of analyses and other parameters as above, with 100 pixels on the X and Y axes. Delta coordinates (error for spatial coordinates) was set at 0 because the samples were all taken from one or a few small ponds at clearly specified sites, but each sample was given a slightly different coordinate, randomly assigned, within the pond cluster. For each region we carried out 20 such runs and checked the consistency of the results.

Because we sampled larvae, recent migration was investigated using BAYESASSNM, a derivative of BAYESASS (Wilson & Rannala 2003) that specifically detects first-generation progeny of migrants (Jehle *et al*. 2005). We used 3×10^6 iterations altogether for each run, including a burn-in of 10^6 iterations, and a sampling frequency of 2000. Delta values for allele frequency, migration rate and inbreeding were varied to try and ensure that accepted numbers of changes were within 40%–60% of the total, as recommended by the program authors, though this did not prove possible for delta inbreeding where accepted changes could not be reduced below 95%. Although STRUCTURE and GENELAND can also be used to identify migrants, they are not specifically designed to detect first generation progeny and we therefore did not try to use them for this purpose.

Timing of population splits in the east/southeast England region was estimated using the IM (Isolation–Migration) program (Hey & Nielsen 2004) with duplicate

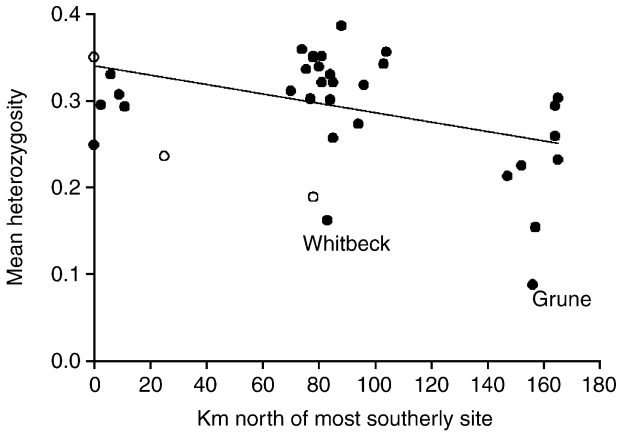

Fig. 2 Heterozygosity and latitude. Solid circles: west coast. Open circles: east coast. Distances are km north of the most southerly location in each case (Hightown on the west coast, Winterton on the east coast).

runs of 10^5 burn-ins and 10^7 iterations, following initial short runs to provide prior estimates of effective size and split time maxima. We assumed no migration between the populations for this analysis, on the basis that there must have been much inhospitable intervening habitat for at least several thousand years in this part of Britain. A generation time of four years (Rowe & Beebee 2004) and an average microsatellite mutation rate of 10^{-5} (Rowe *et al*. 2006) were used to estimate divergence times in calendar years from the t estimates generated by IM. Non-parametric tests (Spearman rank correlation and Wilcoxon rank sum) were used throughout for standard statistical analyses, using the STATISTIX software package (Tallahassee, USA).

Results

Genetic diversity

There was generally high concordance with Hardy–Weinberg equilibrium (at $P = 0.05$ after Bonferroni correction for multiple tests) for all eight loci. An exception, however, was *Bcal*μ6 in East/southeast England populations. Consistently high homozygote excesses at this locus, just in this region, implied the existence of a null allele and *Bcal*μ6 was therefore omitted from all further analyses of the East/southeast England populations. It was nevertheless retained for analyses of populations in the other regions, where this problem was not apparent.

Overall levels of genetic diversity across these microsatellite loci in British *B. calamita* populations were reported previously (Rowe *et al*. 1998). However, with a full data set for the west coast (Merseyside, south Cumbria and Solway, including the two populations sampled in 2005), genetic diversity decreased as a function of increasing latitude ($r_s = -0.351$, $n = 33$, $P = 0.046$). The amount of variation in diversity explained by latitude was weak, but the same relationship probably held on the east coast of England although too few populations occur in this region to be sure (Fig. 2). Two very small and isolated populations in Cumbria (Grune and Whitbeck) had particularly low genetic diversity. However, there was no evidence that northernmost populations have been more prone to bottlenecks than southerly ones in the Cumbria-Solway area within recent times (Table 1). 40–50% of populations in both Solway and south Cumbria showed some, generally weak (mode-shift) evidence of recent bottlenecks. Elsewhere, just one Merseyside population (Birkdale) exhibited clear signs of a recent bottleneck, whereas most of the East/southeast England populations showed weak evidence of bottlenecking. The more rigorous Wilcoxon test indicated generally lower estimates of bottlenecking throughout than did the mode-shift test, as might reasonably be expected.

Table 1 Bottleneck test results from all 38 sampling sites in the four regions

	No. populations bottlenecked according to:	
Region	Wilcoxon test	Mode-shift test
East/southeast England	0/5	4/5
Merseyside	1/5	1/5
South Cumbria	3/19	9/19
Solway	0/9	4/9

Table 2 AMOVA of all 38 British *Bufo calamita* sample sites

Source of variation	Degrees of freedom	Sums of squares	% of variation	P
Among the four sample clusters	3	1270.6	32.2	< 0.0001
Among populations within clusters	34	574.8	10.6	< 0.0001
Within populations	2846	3193.0	57.2	< 0.0001
Total	2883	5038.4		

Population structure — historical perspectives

AMOVA analysis of the four sampling clusters is reported in Table 2. The highest level of differentiation (57% of variation) was within populations, and the lowest (10%) among populations within the sampling regions. However, there was large variation among the four sampling regions and all the variances were highly significant.

Table 3 (a) shows F_{ST} estimates averaged across all loci for all population pairs in east/southeast England. Unsurprisingly, these F_{ST} estimates were relatively high (mostly

Table 3 Intersite F_{ST} estimates. All are significantly different from zero

(a) East/southeast England

Population	Saltfleetby	Holkham	Winterton	Syderstone
Holkham	0.349			
Winterton	0.255	0.143		
Syderstone	0.337	0.249	0.197	
Woolmer	0.229	0.209	0.110	0.242

(b) Merseyside

Population	Ainsdale	Birkdale	Formby	Altcar
Birkdale	0.034			
Formby	0.021	0.067		
Altcar	0.030	0.055	0.054	
Hightown	0.055	0.145	0.056	0.107

> 0.2), reflecting the large geographical and likely long temporal separations of these populations. All were significantly different from zero after Bonferroni correction for multiple comparisons. We used the IM program to estimate likely times of most recent common ancestry for two of these population pairs. Woolmer (heathland, southeast England) is about 250 km from Winterton (dunes, eastern England). The mean IM estimate of common ancestry was 8300 years before present (BP), with a standard deviation (duplicate runs) of 198 years. Winterton and Holkham (both dunes, eastern England) are about 60 km apart. In this case the mean IM estimate of common ancestry was around 5020 years BP, with a standard deviation of 311 years.

By contrast, F_{ST} estimates between the Merseyside sites were relatively low (mostly < 0.1), as shown in Table 3 (b), no doubt reflecting recent gene flow along this dune system. Nevertheless, even here all the F_{ST} estimates were significantly different from zero. Further north, mean pairwise F_{ST} estimates were significantly lower in South Cumbria, at 0.134, than in the Solway region at 0.223 (Wilcoxon Rank Sum Test, $U = 18, 153$, $P = 0.001$). This largely reflected the greater mean intersite distances in Solway (22.6 km) compared with South Cumbria (15.8 km). In some cases, intersite F_{ST} estimates were as low between South Cumbria and Solway sites as between some sites within each of these regions (e.g. Drigg × Mawbray, see Appendix). This may reflect the relatively recent segregation of populations between these two regions, which are now certainly separated by long distances of habitat impermeable to natterjacks. In South Cumbria (Fig. 3a) there was no significant isolation-by-distance (Mantel test, 1000 permutations, $P = 0.064$) whereas in Solway (Fig. 3b) there was a much stronger and significant relationship ($P = 0.002$) suggesting the occurrence of gene flow among

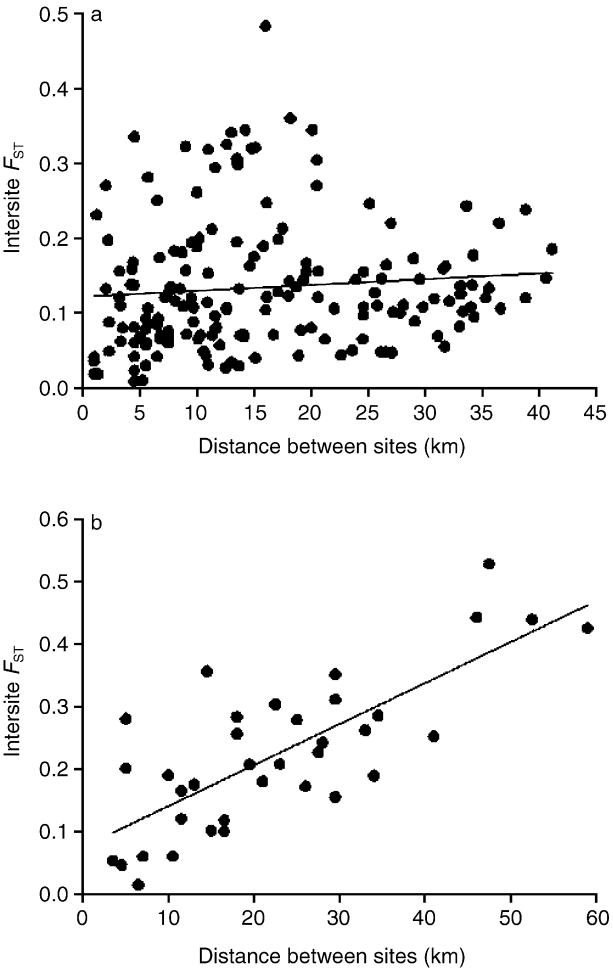

Fig. 3 Isolation by distance in South Cumbria and in Solway. a: South Cumbria; b: Solway.

these populations within historical times. A summary of all intersite pairwise F_{ST} estimates in South Cumbria and Solway is provided in the Appendix.

Population structure — today

In the first round of a hierarchical analysis allowing admixture but with uncorrelated alleles, and no prior information about samplings, the STRUCTURE program reached a plateau with the full set of British natterjack populations at an apparent K estimate of five or more. However, using the K estimator derived from the rate of change of K (Evanno *et al*. 2005) the estimate of K was two, which were East/southeast England and all the rest. A second round of analysis separated Merseyside from South Cumbria and Solway, but a third round did not distinguish the latter two regions. Nevertheless, there is no doubt that South Cumbria and Solway are truly separate regions for natterjacks today because of their large separation (> 30 km) by unsuitable habitat. Individual toads have

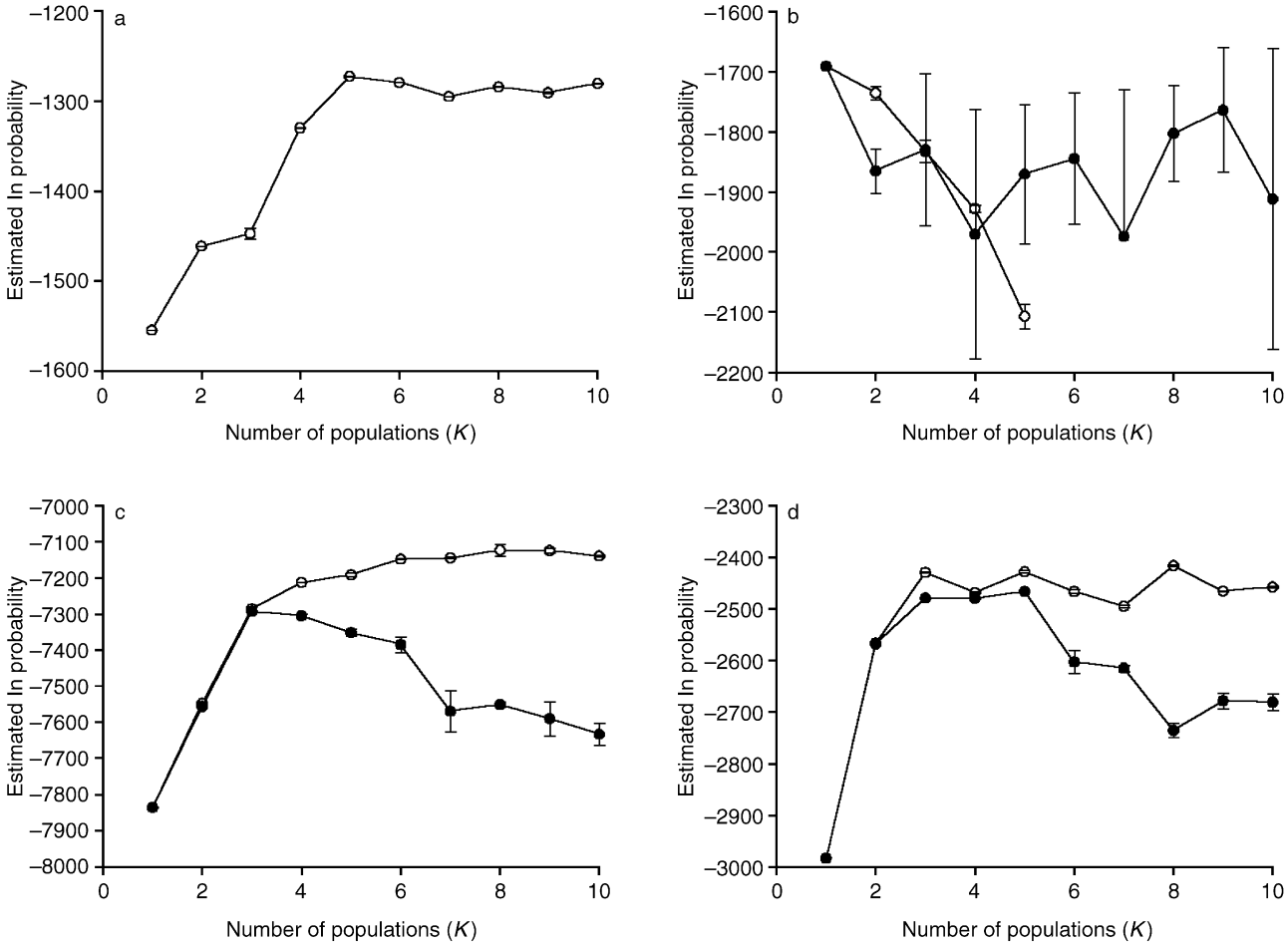

Fig. 4 STRUCTURE analyses of natterjack toad populations. a: East/southeast England; b: Merseyside; c: South Cumbria; d: Solway. Solid circles: correlated alleles; open circles: uncorrelated alleles. Bars show standard deviations.

maximum dispersal distances of no more than 3 km even in suitable habitat (Sinsch 1997). Detailed results of STRUCTURE analyses of *B. calamita* populations in the four British distribution regions are shown in Fig. 4. Standard deviations of the log probability estimates were in most cases very small, less than 1% of the mean, indicating that triplication of runs was sufficient to generate reliable inferences. In east/southeast England, STRUCTURE correctly identified the five well-separated populations with $P \approx 1$. Standard deviations were very small and there was a single likelihood peak with no tendency to increase beyond it. In Merseyside, STRUCTURE assigned all the individuals to a single population, again with $P \approx 1$. This was true whether the model used correlated or uncorrelated alleles, but standard deviations around the K estimates were much smaller with the uncorrelated than with the correlated alleles. For South Cumbria, STRUCTURE indicated either the existence of just three populations (correlated alleles model), again with $P \approx 1$, or a plateau of progressively increasing K estimates (uncorrelated alleles model). In the latter case, using K the optimum estimate was for four populations. Standard deviations around all the K estimates were low. These four populations were, however, further subdivided by continued hierarchical analysis ultimately to yield an estimate of six populations. Finally, for Solway, STRUCTURE (correlated alleles) estimated the presence of five populations with a single probability peak and an overall $P \approx 1$. With uncorrelated alleles there were three peaks, at $K = 3$, $K = 5$ and $K = 8$. Again, all standard deviations were low. With $K = 5$ (correlated alleles) or $K = 3$ (uncorrelated alleles), no further subdivisions were indicated by continued analysis. Overall results of the STRUCTURE analyses for the four regions are reported in Table 4. Using STRUCTURE with prior population information yielded exactly the same pattern of results as described above without this information.

BAPS analysis in the absence of sampling group information (i.e. in individual mode, with no prior information on the group in which the individual was sampled) generated complex groups with very mixed constitutions (data not shown). These analyses proposed 12 populations in east/southeast England, 17 in Merseyside, 32 in south Cumbria

Table 4 Current population structure assessed using three assignment methods

Region	Sample sites	STRUCTURE populations: uncorrelated alleles	BAPS populations	GENELAND populations
East/southeast	Saltfleetby	5 populations:	5 populations:	4 populations:
	Holkham	Saltfleetby	Saltfleetby	Saltfleetby
	Syderstone	Holkham	Holkham	Holkham +
	Winterton	Syderstone	Syderstone	Winterton
	Woolmer	Winterton	Winterton	Syderstone
	(n = 5)	Woolmer	Woolmer	Woolmer
Merseyside	Birkdale	1 population	3 populations:	2 populations:
	Ainsdale		(a) Birkdale	(a) Birkdale
	Formby		(b) Ainsdale +	(b) Ainsdale +
	Altcar		Formby +	Formby + Altcar
	Hightown		Altcar	+ Hightown
	(n = 5)		(c) Hightown	
South Cumbria	North Walney	6 populations:	11 populations:	5 populations(?):
	Sandscale	(a) North Walney +	(a) North Walney	(a) North Walney
	Askam	Eskmeals	(b) Whitbeck	(b) Whitbeck
	Dunnerholme	(b) Whitbeck	(c) Sandscale +	(c) Sandscale +
	Soutergate	(c) Sandscale +	Askam +	Askam +
	Sandside	Askam +	Eskmeals	Dunnerholme +
	Subberthwaite	Dunnerholme +	(d) Dunnerholme	Soutergate +
	Foxfield	Soutergate +	+ Soutergate	Sandside +
	Lady Hall	Sandside +	+ Sandside	Foxfield +
	Green Road	Foxfield	(e) Subberthwaite	Subberthwaite
	Millom	(d) Subberthwaite +	(f) Foxfield	(d) Lady Hall +
	Haverigg	Millom +	(g) Lady Hall +	Green Road +
	Summer Hill	Haverigg	Green Road	Millom +
	Whitbeck	(e) Lady Hall +	(h) Millom	Haverigg
	Annaside	Green Road	(i) Haverigg	(e) Summer Hill +
	Eskmeals	(f) Summer Hill +	(j) Summer Hill +	Annaside +
	Drigg	Annaside +	Annaside	Eskmeals +
	Sellafield	Drigg +	(k) Drigg +	Drigg +
	Braystones	Sellafield +	Sellafield +	Sellafield +
	(n = 19)	Braystones	Braystones	Braystones
Solway	Mawbray	For K = 3	6 populations:	4 populations:
	Silloth	(see Fig. 4):	(a) Mawbray +	(a) Mawbray +
	Grune	(a) Mawbray +	Silloth	Silloth + Grune +
	Anthorn	Silloth + Grune	(b) Grune	Anthorn
	ICI	+ Anthorn	(c) Anthorn	(b) ICI + Priestside east
	Priestside east	(b) ICI + Priestside	(d) ICI +	(c) Priestside west + Caerlaverock
	Priestside west	west + Priestside	Priestside west	(d) Southerness
	Caerlaverock	east +	+ Priestside east	
	Southerness	Caerlaverock	(e) Caerlaverock	
	(n = 9)	(c) Southerness	(f) Southerness	

For STRUCTURE and BAPS, population allocations were as described in Methods. For geneland, population allocations were based on the highest posterior probabilities where 2/3 of the top scorers yielded identical results; or (South Cumbria only) on clusters of populations which occurred at the highest frequencies among the 20 different outcomes.

and 19 in Solway. However, in group mode (i.e. where information was provided about which group each sample was taken from) BAPS correctly ascribed the five populations in east/southeast England with $P \approx 1$. BAPS in this mode split Merseyside into three populations, again with $P \approx 1$, while in south Cumbria it indicated the existence of 11 populations (P = 0.91) and in Solway BAPS indicated six populations with $P \approx 1$. Duplicate runs of BAPS in group mode yielded identical results for all the regions, and these are summarized in Table 4.

GENELAND analyses were assessed for consistency by determining how many of the 20 runs for each sampling cluster produced identical results. We also examined more closely the three runs which generated the highest mean posterior probabilities of population membership (Coulon *et al*. 2006). GENELAND sometimes generates 'ghost'

populations that do not correspond to any sampling sites (Coulon *et al*. 2006), so the K estimates produced in the first round of the analyses do not necessarily correspond to the 'real' groups identified in the second round. For east/southeast England, 15 of the 20 runs generated different groupings and no grouping was indicated more than twice. Increasing the number of iterations to 10^6 did not improve this outcome. However, the differences between groupings were often rather small. Two of the three with the highest posterior probabilities gave the same pattern of four populations. A similar type of data set was obtained from the Merseyside cluster, except that in this case 10 out of the 20 runs were identical. Even so, the three runs with the highest mean posterior probabilities were not in this group of 10 though two of these three were identical to each other. For south Cumbria, all 20 runs yielded different results. Finally, for Solway 10 out of the 20 runs were identical and these included two out of the three with the highest mean posterior probabilities. The results of all the GENELAND analyses are also incorporated in Table 4.

BAYESASSNM analysis of the Merseyside, South Cumbria and Solway population clusters did not detect any evidence of first generation progeny from migrants among any of the sampling sites (results not shown). This result was consistent over a wide range of delta values for allele frequency, migration rates and inbreeding.

Discussion

Genetic diversity

The overall pattern of genetic diversity in British natterjacks can be explained at least partly on the basis of historical events. In general, species with southern refugia during the Pleistocene glaciations are expected to show progressively lower genetic diversity as the range edge is approached because, during postglacial colonization, each new founder event would have represented only a proportion of the parent population (Ibrahim *et al*. 1996). This pattern is seen for natterjacks across their entire range in Europe, with diversity declining from Iberia northwards and eastwards to Britain and the Baltic (Beebee & Rowe 2000; Rowe *et al*. 2006). British natterjacks therefore have substantially lower genetic diversity than those in neighbouring mainland Europe. The range edge effect may still be detectable at a fine scale within Britain, on both east and west coasts, where genetic diversity declines northwards (Fig. 2). The relationship between diversity and latitude is probably weak because of other factors, especially differences in population sizes and structures, that also have large effects on genetic diversity. There was no evidence to suggest that low diversity near the northern range edge (Solway) reflected a higher risk of population bottlenecking than was true in south Cumbria, and isolation-by-distance analysis confirmed that gene flow between sites has been at least as effective in Solway as in south Cumbria. The pattern we now see probably therefore still retains a small echo of events in the postglacial colonization period some 10 000 years ago.

Very low genetic diversity (expected heterozygosity < 0.2) was apparent at four sites (Saltfleetby, Whitbeck, Grune and Anthorn). At Saltfleetby, this low diversity is certainly associated with low fitness (Rowe & Beebee 2003, 2005) and is probably consequent on recent events (habitat loss or alteration) that have reduced the size and increased the isolation of this population. We do not know whether low fitness is an issue at the other three sites.

Population structure

Assignment tests offer the prospect of a more rational definition of populations than was previously possible. However, the three attempts with STRUCTURE, BAPS and GENELAND on British natterjack populations yielded slightly different results. STRUCTURE permits analyses based on three types of model, notably: (i) discrete populations with no admixture; (ii) populations of mixed ancestry (admixture with uncorrelated alleles); and (iii) mixed ancestry in which some loci are in linkage disequilibrium (admixture with correlated alleles). For the natterjack data, model (i) was clearly the most appropriate for east/southeast England, and model (ii) was probably the most appropriate for the other regions. Although it is possible with STRUCTURE to invoke prior information concerning the sampled populations, this facility made no difference to our analyses and is not strictly comparable with the group mode of BAPS. However, BAPS performed poorly in individual mode (the most comparable analysis with STRUCTURE) because it generated large numbers of groups that bore no resemblance to the actual toad distributions. This was even true in east/southeast England, where there could be no doubt about the true population structure. BAPS in group mode, taking account of sampling site information in addition to genotype data, yielded much more credible results. Eskmeals was an outlier in both STRUCTURE and BAPS analyses, since it cannot possibly form part of a population including either North Walney or Sandscale and Askam. It is probable that this anomaly was due to the very small sample size (just eight individuals) from Eskmeals, which in terms of genetic distance from its nearest neighbours was more closely affiliated to Drigg than to Annaside, although all the log Bayes factor values for affiliation to other groups were substantially lower than zero (data not shown). It was interesting to note that in general BAPS indicated more genetically distinct populations than did STRUCTURE, whereas in tests with simulated data the converse was true, with STRUCTURE tending to overestimate *K* (Waples &

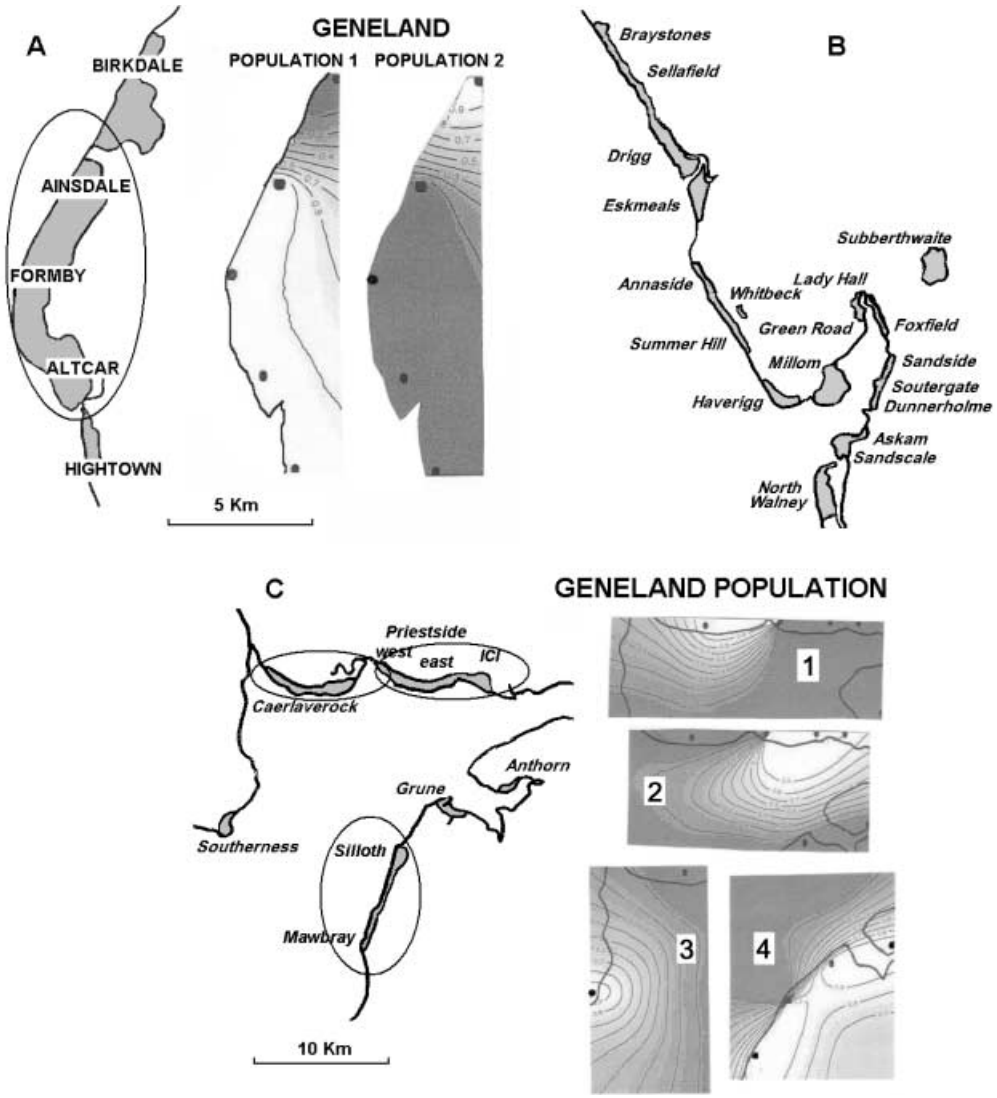

Fig. 5 Current population structures. A: Merseyside; B: South Cumbria; C: Solway. Shaded areas correspond to *Bufo calamita* habitat. Enclosures (circles, ellipsoids) show population clusters common to all three assignment programs. GENELAND population probabilities are shown for Merseyside ($n = 2$) and Solway ($n = 4$). Probability of population membership increases as shading intensity decreases, and solid circles show the sampling sites.

Gaggiott 2006). GENELAND, which takes account not only of different sampling sites but also their geographical relationships to each other, generated population structures that were often similar to those of STRUCTURE and BAPS. GENELAND incorporates specific geographical information concerning the relative situations of the sampling sites. However, interpretation of the output was complicated by inconsistencies between runs, perhaps because the Markov chains failed to converge with our data even after large numbers of iterations. Longer runs were impractical because each took up to eight hours with some of our data sets. It was surprising that even for east/southeast England there was no simple consensus for five separate populations (only two out of the 20 runs yielded this result). Nevertheless, the run outputs for the other regions generated credible outcomes usually closer to STRUCTURE than to BAPS. The commonest groups identified by GENELAND in south Cumbria, the most complex region, were arguably the most credible of all three analyses and GENELAND alone placed Eskmeals (sometimes) among its geographical neighbours. However, the large variation among runs made the output of GENELAND difficult to evaluate objectively with our data set. Overviews of the results of all three methods are given in Table 4, and in Fig. 5 for the three complex regions. GENELAND maps of probabilistic population membership are provided for Merseyside and Solway, but not for south Cumbria where the GENELAND results were complex and more variable. In general, this study highlights the importance of using multiple methods to assess the most likely population

structures in wild populations, and the difficulties in coming to definitive conclusions.

There seems to be very little movement of individual toads between any of the natterjack sample populations, as judged by the BAYESASSNM analysis, since no larvae that had a parent from a different sampling site were detected out of the 1434 sampled. This is perhaps not surprising considering the intersite distances involved. With crested newts (*Triturus cristatus*) only very few such hybrid offspring were detected even in ponds mostly within 1–2 km of each other (Jehle *et al*. 2005). All the evidence (F_{ST} estimates, isolation by distance, assignment analyses) suggests that gene flow must occur in the Merseyside, south Cumbria and Solway regions, but sufficiently rarely that it was not detected directly in a single sampling exercise.

F-statistics indicated that there has, over historical time, been significant gene flow within the Merseyside, South Cumbria and Solway toad population groups. By contrast, F_{ST} and IM analyses suggested that in east/southeast England the populations have always been more isolated. This is probably because heathlands, a major habitat for natterjacks in this region, were generally less well interconnected than the primarily coastal habitats of northwest England and southern Scotland. The common ancestry date for Woolmer (southern England) and Winterton (East Anglia) of around 8300 years BP coincides with the time at which Britain became an island (Vincent 1990), and forest cover became complete following recovery from the last (Younger Dryas) cold period. This would have effectively separated many natterjack populations since forest is an impermeable barrier to this species at northern latitudes (Beebee 1979). Holkham and Winterton are both coastal sites in East Anglia, but even here there is little evidence of recent connectivity. The last ancestry estimate of around 5000 years BP coincides with the onset of a prolonged cooling phase coincident with the disappearance of pond tortoises (*Emys orbicularis*) in East Anglia (Beebee & Griffiths 2000) and this may have rendered previously permeable habitats unsuitable.

Conservation implications

In general it will be important to maintain habitat continuity within the genetically discrete toad population clusters in Merseyside, south Cumbria and Solway. The consensus genetic groupings in these three regions, using the three analytical procedures, are shown in Fig. 5. These groups should be treated as single management units, though it is quite likely that some other populations (i.e. those grouped by some methods but not others) are not genetically discrete. In practice this means sustaining the quality and extents of current terrestrial and aquatic habitats, and preventing the development of obstacles to toad movement between breeding ponds. Where genetically discrete populations are in close proximity, and were probably continuous in recent times, contact should be restored by improving habitat permeability. Creation or restoration of ponds, especially near site border areas, should facilitate gene flow in these situations. Small and/or isolated populations with low genetic diversity should be given high priority for increasing their effective population size. A minimum target should be to attain effective population sizes of > 50 (meaning census sizes of > 200, and thus average spawn string counts of > 100) at all the genetically discrete sites (Rowe & Beebee 2004). For many populations this should be achievable by extending the amounts of aquatic and/or terrestrial habitats as required. However, where low genetic diversity is associated with low fitness (e.g. low rates of tadpole survival), genetic rescue should be considered (Tallmon *et al*. 2004). Such a situation certainly exists at for the Saltfleetby population (Rowe & Beebee 2003, 2005). Genetic rescue would involve the introduction of toads from a nearby, outbred population.

Acknowledgements

Thanks to Inga Zeisset for running the BAYESASSNM analyses, Gilles Guillot for advice concerning GENELAND, two anonymous referees and Jukka Corander for comments on an earlier manuscript, and to English Nature for financial support.

References

Beebee TJC (1976) The natterjack toad (*Bufo calamita*) in the British Isles: a study of past and present status. *British Journal of Herpetology*, **5**, 515–521.

Beebee TJC (1977) Environmental change as a cause of natterjack toad (*Bufo calamita*) declines in Britain. *Biological Conservation*, **11**, 87–102.

Beebee TJC (1979) Review of scientific information pertaining to the natterjack toad *Bufo calamita* throughout its geographical range. *Biological Conservation*, **16**, 107–134.

Beebee TJC, Buckley J (2001) *Natterjack toad (*Bufo calamita*) site register for the UK 1970–99 inclusive*. Unpublished report, University of Sussex and The Herpetological Conservation Trust.

Beebee TJC, Griffiths RA (2000) *Amphibians and Reptiles*. Harper Collins, London.

Beebee TJC, Griffiths RA (2005) The amphibian decline crisis: a watershed for conservation biology? *Biological Conservation*, **125**, 271–285.

Beebee TJC, Rowe G (2000) Microsatellite analysis of natterjack toad *Bufo calamita* Laurenti populations: consequences of dispersal from a Pleistocene refugium. *Biological Journal of the Linnean Society*, **69**, 367–381.

Brede EG, Beebee TJC (2004) Contrasting population structures in two sympatric anurans: implications for species conservation. *Heredity*, **92**, 110–117.

Buckley J, Beebee TJC (2004) Monitoring the conservation status of an endangered amphibian: the natterjack toad (*Bufo calamita*) in Britain. *Animal Conservation*, **7**, 221–228.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics*, **163**, 367–374.

Cornuet J-M, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Coulon A, Guillot G, Cosson JF *et al.* (2006) Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology*, **15**, 1669–1679.

Denton JS, Hitchings SP, Beebee TJC, Gent A (1997) A recovery program for the natterjack toad (*Bufo calamita*) in Britain. *Conservation Biology*, **11**, 1329–1338.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Goudet J (1995) FSTAT version 1.2: a computer program to calculate *F*-statistics. *Journal of Heredity*, **86**, 485–486.

Guillot G, Mortimer F, Estoup A (2005) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

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.

Houlahan JE, Findlay CS, Schmidt BR, Meyer AH, Kuzmin SL (2000) Quantitative evidence for global amphibian population declines. *Nature*, **404**, 752–755.

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

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.

Jehle R, Wilson GA, Arntzen JW, Burke T (2005) Contemporary gene flow and the spatio-temporal genetic structure of subdivided newt populations (*Triturus cristatus, T. marmoratus*). *Journal of Evolutionary Biology*, **18**, 619–628.

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, **20**, 136–142.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

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

Rowe G, Beebee TJC (2003) Population on the verge of a mutational meltdown? Fitness costs of genetic load for an amphibian in the wild. *Evolution*, **57**, 177–181.

Rowe G, Beebee TJC (2004) Reconciling genetic and demographic estimators of effective population size in the anuran amphibian *Bufo calamita*. *Conservation Genetics*, **5**, 287–298.

Rowe G, Beebee TJC (2005) Intraspecific competition disadvantages inbred natterjack toad (*Bufo calamita*) genotypes over outbred ones in a shared pond environment. *Journal of Animal Ecology*, **74**, 71–76.

Rowe G, Beebee TJC, Burke T (1998) Phylogeography of the natterjack toad *Bufo calamita* in Britain: genetic differentiation of native and translocated populations. *Molecular Ecology*, **7**, 751–760.

Rowe G, Beebee TJC, Burke T (2000) A microsatellite analysis of natterjack toad, *Bufo calamita*, metapopulations. *Oikos*, **88**, 641–651.

Rowe G, Harris JD, Beebee TJC (2006) Lusitania revisited: a phyleogeographic analysis of the natterjack toad *Bufo calamita* across its entire biogeographical range. *Molecular Phylogenetics and Evolution*, **39**, 335–346.

Schneider S, Kueffer JM, Roessli D, Excoffier L (2000) ARLEQUIN version 1.1: a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva.

Sinsch U (1997) Postmetamorphic dispersal and recruitment of first breeders in a *Bufo calamita* metapopulation. *Oecologia*, **112**, 42–47.

Stuart SN, Chanson JS, Cox NA *et al.* (2004) Status and trends of amphibian declines and extinctions worldwide. *Science*, **306**, 1783–1786.

Tallmon DA, Luikart G, Beaumont MA (2004) Comparative evaluation of a new effective population size estimator based on approximate Bayesian computation. *Genetics*, **167**, 977–988.

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

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Both the authors have a longstanding interest in conservation genetics, and in particular, the use of hypervariable molecular markers to address issues of phylogeography and population viability. The emphasis of our research has been on amphibians, especially European species of anurans.

Appendix

Pairwise F_{ST} estimates for South Cumbria and Solway

	Population number																										
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	0.14	0.18	0.16	0.19	0.15	0.21	0.34	0.20	0.19	0.21	0.16	0.17	0.16	0.25	0.16	0.24	0.24	0.19	0.27	0.26	0.43	0.35	0.21	0.27	0.25	0.25	0.21
2		0.04	0.01*	0.03	0.04	0.03	0.11	0.11	0.07	0.07	0.07	0.04	0.25	0.12	0.05*	0.09	0.10	0.11	0.09	0.11	0.21	0.16	0.11	0.12	0.10	0.16	0.22
3			0.08	0.04	0.08	0.06	0.18	0.10	0.07	0.09	0.07	0.07	0.32	0.16	0.05*	0.11	0.10	0.13	0.16	0.17	0.27	0.24	0.20	0.22	0.19	0.22	0.25
4				0.02	0.02*	0.03	0.14	0.12	0.07	0.09	0.06	0.05	0.29	0.12	0.04*	0.11	0.11	0.12	0.13	0.16	0.22	0.18	0.14	0.14	0.12	0.17	0.21
5					0.04	0.06	0.14	0.11	0.08	0.11	0.09	0.08	0.33	0.13	0.05*	0.13	0.12	0.14	0.13	0.16	0.22	0.17	0.13	0.15	0.21	0.16	0.23
6						0.08	0.16	0.17	0.12	0.16	0.13	0.11	0.30	0.14	0.10*	0.16	0.16	0.14	0.20	0.23	0.26	0.20	0.13	0.15	0.14	0.18	0.14
7							0.12	0.11	0.09	0.04	0.07	0.08	0.34	0.16	0.07*	0.13	0.12	0.15	0.18	0.19	0.26	0.25	0.19	0.19	0.14	0.20	0.26
8								0.23	0.20	0.17	0.26	0.18	0.48	0.27	0.22	0.17	0.18	0.22	0.17	0.19	0.19	0.25	0.23	0.19	0.16	0.23	0.41
9									0.02*	0.06	0.11	0.07	0.32	0.14	0.11	0.10	0.08	0.12	0.17	0.17	0.34	0.28	0.21	0.28	0.23	0.23	0.33
10										0.06	0.08	0.03	0.31	0.12	0.07*	0.05	0.06	0.11	0.14	0.14	0.29	0.25	0.22	0.26	0.23	0.25	0.32
11											0.06	0.07	0.32	0.20	0.08*	0.11	0.10	0.15	0.18	0.16	0.33	0.32	0.22	0.27	0.21	0.26	0.32
12												0.07	0.28	0.20	0.07*	0.14	0.14	0.15	0.21	0.31	0.37	0.30	0.21	0.25	0.21	0.27	0.25
13													0.27	0.09	0.08	0.03	0.04	0.07	0.08	0.09	0.22	0.18	0.19	0.20	0.20	0.25	0.30
14														0.34	0.32	0.34	0.36	0.30	0.38	0.37	0.64	0.46	0.32	0.41	0.39	0.44	0.39
15															0.25	0.13	0.13	0.10	0.20	0.25	0.43	0.16	0.22	0.25	0.23	0.26	0.33
16																0.13	0.12	0.19	0.21	0.20	0.41	0.37	0.26	0.30	0.23	0.25	0.31
17																	0.01*	0.08	0.10	0.10	0.22	0.20	0.24	0.24	0.23	0.27	0.38
18																		0.05	0.12	0.11	0.22	0.21	0.23	0.24	0.22	0.24	0.35
19																			0.16	0.17	0.28	0.19	0.17	0.18	0.18	0.21	0.26
20																				0.01*	0.17	0.10	0.17	0.16	0.19	0.25	0.43
21																					0.20	0.19	0.21	0.21	0.23	0.29	0.44
22																						0.28	0.36	0.26	0.30	0.35	0.53
23																							0.18	0.12	0.18	0.24	0.44
24																								0.05	0.06	0.10	0.26
25																									0.05	0.12	0.31
26																										0.06	0.28
27																											0.28

Population list:
1–19: South Cumbria; 20–28: Solway. 1-North Walney; 2-Sandscale; 3-Askam; 4-Dunnerholme; 5-Soutergate; 6-Sandside; 7-Subberthwaite; 8-Foxfield; 9-Ladyhall; 10-Green Road; 11-Millon; 12-Haverigg; 13-Summer Hill; 14-Whitbeck; 15-Annaside; 16-Eskmeals; 17-Drigg; 18-Sellafield; 19-Braystones; 20-Mawbray; 21-Silloth; 22-Grune; 23-Anthorn; 24-ICI; 25-Priestside East; 26-Priestside West; 27-Caerlaverock; 28-Southerness. *Estimates that were NOT significantly different from zero at $P = 0.05$, after Bonferroni correction.