Genetic analysis of Kerry natterjack toad (*Bufo calamita*) populations

Report prepared for the National Parks and Wildlife Service June 2008

Shoshanna May and Trevor J.C. Beebee

School of Life Sciences University of Sussex Falmer Brighton BN1 9QG UK

Summary

- (1) Six natterjack toad populations (four around Castlemaine Harbour and two on the north coast of the Dingle peninsula) were genotyped at nine polymorphic microsatellite loci.
- (2) All populations had similar and moderate levels of genetic diversity, comparable with those on the coast of north-west England.
- (3) Even around Castlemaine Harbour the toad populations were substantially differentiated, implying little migration between sites within historical times.
- (4) Phylogenetics and estimates of divergence times supported the hypothesis that populations on the north coast of Dingle separated from those around Castlemaine Harbour many thousands of years ago, and are not recent introductions.

Background

Natterjack toads *Bufo calamita* are part of Ireland's Lusitanian biota and were first recorded in the country almost 200 years ago (Mackay, 1836). Recent genetic studies (Beebee & Rowe, 2000; Rowe *et al.*, 2006) confirmed that natterjacks are native to Ireland and probably colonised Kerry in the period immediately after the Younger Dryas cooling, circa 9-10,000 years before the present. However, until the mid 20th century the toads were only recorded from the Dingle Bay (Castlemaine Harbour) area and a few outlying sites further south around the Waterville peninsula. A surprising discovery in the late 1960s was a large natterjack population on the north side of the Dingle peninsula, between Castlegregory and Fermoyle (Gresson & O'Dubhda, 1971). The question then arose as to why such a large population had not been noticed earlier, and therefore whether it might be the result of a relatively recent translocation from the well-known populations further south. As long ago as the early 1800s people were evidently moving toads around in Kerry. Ward (1864) states "Mr Townsend adds, that on one occasion he removed a few dozen of them to a coastguard station, north-east of Dingle, that is to say some miles west of Inch Point".

Molecular genetic methods offer the possibility of resolving this question, and thus determining whether the north Dingle natterjack populations are longstanding or relatively new. Such an approach should also clarify the genetic diversity of individual natterjack populations in Ireland and the extent to which they differ from each other, all of which information is relevant to conservation management. We therefore set out to investigate the genetic structure of six Irish natterjack toad populations using polymorphic microsatellite markers, and to test the following alternative hypotheses: (1) Castlemaine Harbour and north Dingle populations are equally old, and have been separated since the initial colonisation of Ireland. Intervening habitat (both along the rocky coast and across the mountainous peninsula) appears impermeable to natural natterjack movements.

(2) North Dingle is a relatively recent (within the past 200 years) translocation, presumably by humans.

Methods

(1) *Sampling*. About 40 natterjack toad larvae were caught by net at each of six sites in county Kerry (see Figure 1) during a visit in early June 2007. Four sites were in the Castlemaine Harbour area (Inch, Roscullen, Glenbeigh and Yganavan), and two were north Dingle populations (Castlegregory and Tullaree). Entire larvae (two sites) or c. 5 mm tail fin clips from larvae (four sites) were immediately preserved in 100% ethanol for return to the laboratory. Fin-clipped larvae were released alive at their sites of origin.

(2) *Microsatellite genotyping*. DNA was extracted from all the larval samples using Chelex resin (Walsh *et al.*, 1991). Eighteen polymorphic microsatellite loci have previously been characterised in *Bufo calamita* (*Bcalµ*1-12, Rowe *et al.*, 1997; 2001; Buca1-6, Rogell *et al.*, 2005). All of these were tested with Kerry *B. calamita* samples, but only nine proved polymorphic in the Irish toads. All of these nine (*Bcalµ*1, *Bcalµ*2 *Bcalµ*3 *Bcalµ*4 *Bcalµ*5 *Bcalµ*8 *Bcalµ*11, Buca1 and Buca2) were used in the full population study. PCR conditions for amplification, and subsequent genotype analyses, were as described elsewhere (Rowe *et al.*, 1997; 2001; Rogell *et al.*, 2005), except for the following minor alterations: for *Bcalµ*11 the PCR annealing temperature in some cases was reduced to 53 °C; and for Buca1 and Buca2, we used 30 PCR cycles and 90 second elongation times.

(3) Data analysis. Genotype data files were inter-converted for the various analytical software programs using CREATE (Coombs et al., 2008) to minimise errors. MICRO-CHECKER (van Oosterhout et al., 2004) was employed to test for scoring errors. All loci x population combinations were tested for compliance with Hardy Weinberg equilibrium, and for linkage equilibrium, using GENEPOP 3.4 (Raymond & Rousset, 1995). To quantify genetic diversities, observed (Ho) and expected (He) heterozygosities, and allelic richness, were estimated using FSTAT 2.9.3 (Goudet, 1995). Possible population bottlenecks were evaluated with BOTTLENECK (Cornuet & Luikart, 1996), using a twophase model with 90% stepwise mutation. To assess population structure we used FSTAT to estimate global F_{ST} statistics and F_{STS} among the population pairs, and GENEPOP (ISOLDE) to investigate isolation by distance effects. STRUCTURE (Pritchard et al., 2000 was also employed to investigate how clearly the sampled populations were discrete entities. For this analysis we used a burn-in of 5 x 10^4 followed by 10^6 iterations for each run, triplicated for each value of K (estimated number of true populations) ranging from 1-6. We used the uncorrelated alleles model without admixture. Relationships among the sampled populations were estimated using the UPGMA (based on chord distances, Cavalli-Sforza & Edwards, 1967) method in the PHYLIP 3.66 program package (Felsenstein, 1993), with 1000 allele frequency bootstraps in each analysis. Chord distances have proved the most reliable with microsatellite data (Takazaki & Nei, 2008).

Figure 1: Distribution of *Bufo calamita* in Co. Kerry, 2007. Shaded areas = habitat with natterjacks present. Sampling sites are circled.



Finally, coalescent times between populations and average historical effective population sizes were estimated using IM (Hey & Nielsen, 2004), assuming an average microsatellite stepwise mutation rate of 10^{-5} /locus/generation (Rowe *et al*, 2006). These analyses were duplicated with 10^{7} steps, burn-ins of 10^{5} steps, five Markov chains and ten chain swaps for each run. Values of scalars for θ_1 (= $4Ne\mu$ of population 1) maximum and minimum times of population splitting, and for *t* (time of coalescence) were determined empirically during short preliminary runs. We set inter-site migration rates (m1 and m2, migrations in both directions) to zero for north Dingle x Castlemaine Harbour studies on the basis that intervening mountain habitat on the Dingle peninsula is impermeable to *B. calamita*.

Results

Genetic diversity

Initial tests with MICR-CHECKER indicated no scoring problems, and just four examples (among 9 loci x 6 populations, = 54 analyses) of significant homozygote excess above expectations. These were not associated with any single population or locus. No cases were found of significant linkage disequilibrium between the nine loci, and only a single instance (after Bonferroni correction of probability values, 54 comparisons) of significant deviation from Hardy-Weinberg equilibrium was detected. We therefore proceeded to assess genetic diversities using the full set of nine polymorphic loci. Results are summarised in Table 1.

| Population | Ν | Mean Ho | Mean He | Mean allelic richness |
|---------------|----|---------|---------|-----------------------|
| Yganavan | 39 | 0.423 | 0.442 | 2.64 |
| Glenbeigh | 36 | 0.432 | 0.428 | 2.89 |
| Roscullen | 40 | 0.389 | 0.389 | 2.22 |
| Inch | 40 | 0.295 | 0.308 | 2.42 |
| Castlegregory | 40 | 0.340 | 0.353 | 1.86 |
| Tullaree | 40 | 0.289 | 0.286 | 1.89 |

Table 1. Genetic diversities of Irish natterjack toad populations.

N = sample size. Estimates of genetic diversity are averages across nine loci

All the sampled populations were broadly similar with respect to all three estimators of genetic diversity, albeit with some indication of slightly lower allelic richness in the northern populations (Tullaree and Castlegregory) relative to those around Dingle Bay.

Bottleneck tests indicated (using Wilcoxon tests of significance, and mode-shift assessment of allele frequency distributions) that three of the six populations have undergone size reductions within recent times. These were Glenbeigh, Roscullen, and Castlegregory. Inch, Yganavan and Tullaree gave no strong indication of recent (within about the last 200 year) bottlenecks, although Tullaree did show a mode shift in that direction.

Population structure

| Table 2. Global F_{ST} estimates | | | |
|---|------------------------|--|--|
| Locus | Global F _{ST} | | |
| Bcalµ l | 0.145 | | |
| Bcalµ2 | 0.329 | | |
| Bcalµ3 | 0.137 | | |
| Bcalµ4 | 0.129 | | |
| Bcalµ5 | 0.054 | | |
| Bcalµ8 | 0.209 | | |
| Bcalµ11 | 0.059 | | |
| Buca1 | 0.265 | | |
| Buca2 | 0.169 | | |

Global estimates of F_{ST} across all six populations indicated considerable variation among loci, as shown in Table 2. Loci with estimates > 1 standard deviation away from the mean

of 0.166 ($Bcal\mu 2$, $Bcal\mu 5$ and $Bcal\mu 11$) were excluded from subsequent F_{ST} analyses because of possible bias from directional or balancing selection. Pairwise estimates of F_{ST} among all six populations, based on the six remaining loci, are shown in Table 3. All were significantly different from zero at the 5% level, as estimated by permutation (300 iteration) tests after Bonferroni correction.

| | Yganavan | Glenbeigh | Roscullen | Inch | Castlegregory |
|---------------|----------|-----------|-----------|-------|---------------|
| Glenbeigh | 0.205 | | | | |
| Roscullen | 0.093 | 0.095 | | | |
| Inch | 0.109 | 0.238 | 0.128 | | |
| Castlegregory | 0.197 | 0.166 | 0.192 | 0.232 | |
| Tullaree | 0.159 | 0.219 | 0.144 | 0.300 | 0.217 |

Table 3. Pairwise F_{ST} estimates

The mean F_{ST} estimate across all 15 population pairs was 0.180, and across the six pairs within Castlemaine Harbour was 0.145. Both of these estimates are relatively high. Isolation by distance, investigated using Mantel tests with 10,000 permutations, was insignificant whether direct or coastline distances between the sites were employed. Even around Castlemaine Harbour, therefore, the sampled populations were isolated and must experience very little inter-site migration.

Analysis using the STRUCTURE program gave similar results with the selected six or with all nine loci. K (number of true population cluster) estimates increased asymptotically from 1-6, but using the Evanno *et al* (2005) procedure with the full nine loci there was an inference of two main clusters:

Group 1: 92% of Roscullen samples, 74% of Glenbeigh samples, 77% of Yganavan samples.

Group 2: 67% of Tullaree samples, 97% of Castlegregory samples, 60% of Inch samples.

Inch was therefore the population most difficult to ascribe to particular group, apparently split between Dingle Bay (40%) and north of the Dingle peninsula (60%).

Population divergence

To obtain a clearer picture of the genetic relationships of the Irish natterjack populations, we carried out phylogenetic analyses using PHYLIP with the UPGMA method and Cavalli-Sforza chord distances. This approach has generated the most robust phylogenetic trees when used previously with British natterjack toad populations (Rowe *et al.*, 1998). With the Irish data alone, we used both the full data set and the data set with the six selected loci, and both gave similar tree topologies (as did analyses with neighbourjoining and maximum likelihood methods). Results using the six selected loci are shown in Figure 2A. All three major groupings had bootstrap values of >50% and can therefore be considered robust. Tullaree and Castlegregory clustered together, as did Glenbeigh and Roscullen, with Yganavan closely related to the latter group. Consistent with the STRUCTURE results described above, Inch appeared as an outgroup intermediate between the Dingle Bay and North Dingle population clusters. We therefore carried out a second analysis using a Cumbrian population (Sandscale) genotyped in 1996 as an outgroup. In this case we could only use the six polymorphic loci with genotypes

Figure 2.

A: Phylogeographical relationships of Irish natterjack populations.



available for all populations, notably $Bcal\mu$ 1-5, and $Bcal\mu$ 8. Results are shown in Figure 2B. This analysis clearly separated the north Dingle populations from those around Castlemaine Harbour, with high bootstrap values.



B: Phylogeography including a Cumbrian outgroup population.

We then used the IM program to estimate historical divergence times of population pairs, first utilising the largest populations from around Dingle Bay (Yganavan) and north of the Dingle peninsula (Castlegregory). For this study we used the full nine loci, and results were averaged across two program runs of 10^7 iterations. A similar comparison of Inch and Castlegregory was also carried out using IM, this time with just eight loci (*Bcalµ*4 was fixed for the same allele in both populations, and therefore uninformative) because Inch is the closest population to Castelgregory and thus the one from which any introductions may most likely have originated. Results are shown in Table 4.

| Demelotion | T () - f - l' (0007 | $\mathbf{Eff}_{1} = \mathbf{A}^{1}_{1} = \mathbf{A}^{$ |
|-----------------|-----------------------------------|--|
| estimated using | g IM (average of two runs). | |
| Table 4. Popula | ation parameters (Castlegregory x | Yganavan; Castelegregory x Inch) |

| Population | T (years) of divergence (90% confidence limits) Castlegregory x: | Effective population size (90% confidence limits) |
|---------------|--|---|
| Castlegregory | | 849 (295-3,448) |
| Yganavan | 6,520 (1,720-15,520) | 1,688 (562-6,565) |
| Inch | 5,040 (1,080 - 10,480) | 1,010 (505-3,870) |

Evidently Castelegregory and Yganavan or Inch last shared a common ancestor more than at least 1,000 years ago, and probably much longer, a result not compatible with.recent introduction of natterjacks to Castelgregory from the Castlemaine Harbour area. All three of the comparator populations had large effective (historical average since time of split) sizes, although Inch is probably much smaller today.

For comparison we also estimated divergence times of two populations around Castlemaine Harbour (Roscullen and Glenbeigh) with the relatively low inter-site F_{ST}

estimate of 0.095, inferring some gene flow between them. We made two sets of estimates (Table 5), either not permitting or permitting gene flow.

| non (averages of the fails). | | | | |
|------------------------------|---------------------------------|---|-----------------|--|
| Conditions | T (years) of divergence (90% | Effective population size (90% confidence limits) | | |
| | confidence limits) | Roscullen | Glenbeigh | |
| No gene flow | 3,100 (520-9,920) | 719 (143-2,708) | 775 (155-3,255) | |
| Allowing gene flow | 5,585 (1,080-16,320) | 286 (143-2,423) | 620 (155-2,945) | |

Table 5. Population parameters (Roscullen x Glenbeigh) estimated without or with gene flow (averages of two runs).

The more realistic divergence time estimate, allowing gene flow, was similar to that of Castlegregory from the Castlemaine Harbour populations suggesting a common ancestral population existed in the Kerry region probably more than 5,000 years ago. IM also estimates migration rates, which were about the same in both directions and inferred that since separation one toad has moved between these sites approximately once every 500 years. In practice this integrates movements around the bay, which previously maintained several intervening populations (and still has one, at Dooaghs).

Discussion

With respect to genetic diversities and the extent of genetic differentiation between populations, Irish natterjacks are strikingly similar to those in south Cumbria (Rowe *et al.*, 2007). Mean expected heterozygosities in both areas were between 0.35-0.40, mean intersite F_{STS} were around 0.15, and in neither area was there significant isolation by distance. It is to the Cumbrian natterjacks that the Irish populations are most closely related, and together with other populations in north-west England they probably shared a common ancestry in a Lusitanian refuge at the end of the Younger Dryas cold period some 10,000 years ago (Rowe *et al.*, 2006).

These results imply that the Irish natterjack populations are not, and have not recently been, in genetic equilibrium even around Castlemaine harbour where historical accounts implied they were continuously distributed. The south and east coasts of the harbour probably had much more good toad habitat than they do today, before the construction of extensive sea walls and improved inland drainage. However, there have always been substantial barriers to toad movement around the harbour bay, such as the rivers Caragh, Laune and Maine. Much of the northern coast between Roscullen and Inch is drier and less suitable than the south coast as natterjack habitat, and Inch toads may (as the genetic data suggest) have been isolated from others around the harbour for a long time. The situation in Ireland and Cumbria contrasts with the continuous dune habitat along the Merseyside coast, where F_{ST} estimates between natterjack breeding areas are much lower (averaging around 0.06) implying regular gene flow along that coast.

The effective population sizes of natterjack toad populations average around 10% of census sizes (Rowe & Beebee, 2004). The IM analyses suggest that, averaged over the time since they diverged several thousand years ago, the six sampled populations in Kerry must have all had census sizes of several thousands. It is important to note that these estimates are historical averages and do not apply to the current situation. Numbers

of natterjacks on Inch, in particular, must be much lower today than the historical average for the site.

Taken together, the genetic analyses infer that the first hypothesis (an old origin for the north Dingle nattertjack populations) is most probably correct. Firstly, Castlegregory and Tullaree toads had levels of genetic diversity broadly comparable with those around Castlemaine Harbour. Recent introduction would very likely have involved a small number of individuals, and thus only produced a sub-sample of the diversity further south. Secondly, phylogenetic analysis using a Cumbrian outgroup clearly separated north Dingle and Castlemaine Harbour population clusters, implying a distant common ancestry. Thirdly, the time estimates for common ancestry estimated by IM support an ancient divergence (several thousand years, on average) for populations north of the Dingle peninsula and those around Castlemaine Harbour. Although the 90% confidence limits on this time are very wide (largely due to the limited genetic diversity in the Irish populations), they tend to exclude a recent establishment of the north Dingle populations. The first hypothesis, notably a common origin for all the Irish populations soon after the Younger Dryas cooling, is therefore supported by the genetic data.

The results of this study emphasise the need for conservation of all the surviving Irish natterjack toad populations, and the importance of improving habitat between the populations around Castlemaine Harbour to restore inter-site migration wherever possible. Long-term population isolation increases the risk of local extinction by a range of random factors, including inbreeding depression if toad numbers should fall. In 2007, however, the genetic diversities of all the sampled populations were reassuringly high.

Acknowledgements

Thanks to Jessica Scriven for preliminary studies with some of the microsatellite loci; to Aurelie Aubry, Tim O'Donoghue, Pascal Dower, Frank McMahon and Ferdia Marnell for help with sampling; and to the National Parks and Wildlife Service (Ireland) for funding this project.

References

- Beebee, T.J.C. & 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.
- Cavalli-Sforza, L.L. & Edwards, A.W.E. (1967) Phylogenetic analysis: models and estimation procedures. *Evolution* **21**, 550-570.
- Coombs, J.A., Letcher, B.H. & Nislow, K.H. (2008) CREATE: a software to create input files from diploid genetic data for 52 software programs. *Molecular Ecology Resources* **8**, 578-580.
- 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.
- 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.
- Felsenstein, J. (1993) PHYLIP (Phylogeny inference package) version 3.5c. University of Washington, Seattle, WA.
- Goudet, J. (1995) FSTAT (vers. 1.2): a computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.
- Gresson, R.A.R. & O'Dubhda, S. (1971) Natterjack toads *Bufo calamita* Laur, at Castlegregory and Fermoyle, Co. Kerry. *Irish Naturalists' Journal* **17**, 9-11.
- Hey, J. & Nielsen, R. (2004) Multilocus methods for estimating population sizes, migration rates and divergence times, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis. Genetics* **167**, 747-760.
- Mackay, J.T. (1836) The natterjack toad (*Bufo rubeta*) occurs wild in Ireland. *Magazine* of Natural History **9**, 316-317.
- Pritchard, J.K., Stephens, M. & Donelly, 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.
- Rogell, B., Gyllenstrand, N. & Hoglund, J. (2005) Six polymorphic microsatellite loci in the natterjack toad, *Bufo calamita*. *Molecular Ecology Notes* **5**, 639-640.
- Rowe, G. & Beebee, T.J.C. (2004) Reconciling genetic and demographic estimates of effective population size in the anuran amphibian *Bufo calamita*. *Conservation Genetics* 5, 287-298
- Rowe, G., Beebee, T.J.C. & Burke, T. (1997) PCR primers for polymorphic microsatellite loci in the anuran amphibian *Bufo calamita*. *Molecular Ecology* 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 introduced populations. *Molecular Ecology* 7, 751-760.
- Rowe, G., Beebee, T.J.C. & Burke, T. (2001) A further four polymorphic microsatellite loci in the natterjack toad *Bufo calamita*. *Conservation Genetics* **1**, 371-372.
- Rowe, G., Harris, D.J. & Beebee, T.J.C. (2006) Lusitania revisited: A phyleogeographic analysis of the natterjack toad *Bufo calamita* across its entire biogeographical range. *Molecular Phylogenetics & Evolution* **39**, 335-346.

- Takazaki, N. & Nei, M. (2008) Empirical tests of the reliability of phylogenetic trees constructed with microsatellite data. *Genetics* **178**, 385-392.
- Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P.F. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4, 535-538.
- Walsh, P.S., Metzger, D.A. & Higuchi, R. (1991) Chelex R 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**, 506-513.
- Ward, Hon. Mrs (1864) The natterjack toad in Ireland. Intellectual Observer 5, 227-233.