River islands, refugia and genetic structuring in the endemic brown frog *Rana kukunoris* (Anura, Ranidae) of the Qinghai-Tibetan Plateau

WEIWEI ZHOU,* FANG YAN,* JINZHONG FU,† SHIFANG WU,* ROBERT W. MURPHY,*‡ JING CHE* and YAPING ZHANG*§

State Key Laboratory of Genetic Resources and Evolution, and Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, 650223, China, †Department of Integrative Biology, University of Guelph, Guelph, ON, Canada N1G 2W1, ‡Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, 100 Queen’s Park, Toronto, ON, Canada M5S 2C6, §Laboratory for Conservation and Utilization of Bio-resources, Yunnan University, Kunming, 650091, China

Abstract

Frequently, Pleistocene climatic cycling has been found to be the diver of genetic structuring in populations, even in areas that did not have continental ice sheets, such as on the Qinghai-Tibetan Plateau (QTP). Typically, species distributed on the plateau have been hypothesized to retreat to south-eastern refugia, especially during the Last Glacial Maximum (LGM). We evaluated sequence variation in the mitochondrial DNA gene *Cytb* and the nuclear DNA gene *RAG-1* in *Rana kukunoris*, a species endemic to the QTP. Two major lineages, N and S, were identified, and lineage N was further subdivided into N1 and N2. The geographical distribution and genealogical divergences supported the hypothesis of multiple refugia. However, major lineages and sublineages diverged prior to the LGM. Demographical expansion was detected only in lineage S and sublineage N2. Sublineage N1 might have survived several glacial cycles *in situ* and did not expand after the LGM because of the absence of suitable habitat; it survived in river islands. Genetic analysis and environment modelling suggested that the north-eastern edge of QTP contained a major refugium for *R. kukunoris*. From here, lineage S dispersed southwards after the LGM. Two microrefugia in northern Qilian Mountains greatly contributed to current level of intraspecific genetic diversity. These results were found to have important implications for the habitat conservation in Northwest China.

Keywords: amphibians, Last Glacial Maximum, microrefugia, Northwest China, phylogeography, Pleistocene climatic oscillations

Received 13 June 2012; revision received 30 August 2012; accepted 4 September 2012

Introduction

Pleistocene climatic oscillations are one of the most important drivers of contemporary diversity in many species and communities (Hewitt 2000, 2004). Global cyclical cooling–warming events during the Quaternary Period have resulted in periodic expansions and contractions of the ranges of species. For example, species survived in south-eastern North America during the repeated, climatically driven changes in ranges. Temperate species dispersed northwards during interglacial times (Hewitt 2004). Today, sky islands serve as interglacial refugia in arid regions for many species that are mesic restricted (or desiccation intolerant) (Bryson et al. 2010, 2011). Refugia, areas relatively unaltered through multiple glacial cycles, have played important roles in shaping the genetic structure of species and often conserved high levels of genetic diversity. In
contrast, recently occupied areas have reduced genetic diversity because of repeated bottlenecks during expansion (Hewitt 2000, 2004).

Refugia and recently occupied areas have different amounts of genetic diversity, and this observation lends to a suite of testable hypotheses. New analytical tools and more exhaustive sampling facilitate the identification of previously unidentified refugia (Stewart & Lister 2001; Bhagwat & Willis 2008). These cryptic (or micro-) refugia, defined as small areas of local distribution outside the major glacial refugia, are receiving much attention, and the roles they play in maintaining genetic diversity are hotly debated (Provan & Bennett 2008).

In contrast to Europe and North America, glaciation in temperate Asia did not involve a unified ice sheet. In contrast to other Northern Hemisphere regions, the most recent rapid uplifts of the Qinghai-Tibetan Plateau (QTP) initiated the formation of isolated montane glaciers, ice caps and valley glaciers in many high-peak regions only (Zhou et al. 2006). The QTP is largest and highest plateau on Earth. It covers more than 2.5 million km$^2$ and has an average elevation of about 4000 m above sea level (Zheng 1996). Limited fossil records of pollen suggest that during the LGM, the ice-free areas of the QTP were mainly covered with permafrost/desert-steppe and that forest islands only existed in the extreme southern areas, and especially in south-eastern and eastern parts (Tang & Shen 1996; also see Qiu et al. 2011).

Genetic studies and fossil pollen records suggest two main scenarios for the biota of the QTP: re-colonization and survival in situ. First, re-colonization of the QTP seems to have occurred from peripheral glacial refugia, which were located at the eastern edge of the plateau. Second, species survived during LGM and thus had several refugia on the QTP (Qiu et al. 2011). Although some species have been studied, especially alpine flora (Zhang et al. 2005; Meng et al. 2007; Wang et al. 2009), the influence of Quaternary glaciations on the biota of the QTP and, if present, the location of refugia and postglacial re-colonization routes still remain poorly explored. Moreover, detailed intraspecific faunal analyses on the QTP are sparse, and among vertebrates, birds have received the greatest amount of attention (Qu et al. 2010; Zhan et al. 2011).

Amphibians, as indicators of ecosystem changes, retain high-resolution signals of historical responses to environmental perturbations (Beebee 2005). Historical events, such as range expansion and contraction, and vicariance because of habitat fragmentation often leave an imprint on their contemporary genetic structure. The first scenario predicts that a frog occurring on the edges of the QTP, for example, the Hengduan Mountains, will contain higher levels of genetic diversity than populations on the plateau. In contrast, the second scenario predicts the occurrence of distinctive lineages whose origins predate the LGM. One species of brown frog, *Rana kukunoris*, occurs on both north-eastern and south-eastern edges of the QTP, and it has a restricted distribution on the plateau (Fig. 1). Endemic to the QTP region, the species occurs at elevations ranging from 1500 to 4400 m (Fei et al. 2009). To the north, *R. kukunoris* occurs along the northern Qilian Mountains, which separate the QTP from the Hoxi Corridor and Gobi Desert. Three major endorheic rivers originating from glaciers on the Qilian Mountains dominate here, and they flow northwardly and disappear in the desert. They exist as ‘river islands’ surrounded by the Gobi Desert, and they are essential for life north of the Qilian Mountains. To the south-east, *R. kukunoris* occurs in the Hengduan Mountains. To the west, it reaches to the east of Tsaidam Basin only. This complex landscape facilitates analyses of physiographic barriers to dispersal of species and the potential existence of refugia.

We employ markers from both the mitochondrial DNA (mtDNA) and nuclear DNA (nuDNA) to test hypotheses on the drivers of genetic patterning in *R. kukunoris*. Samples collected from across the entire range of this species allow us to test hypotheses on the evolutionary history and population demographics of *R. kukunoris*, and to determine whether one refugium or multiple refugia best explain the geographical structure of genetic variation. Species’ distribution modelling (SDM) (Hijmans & Graham 2006; Kidd & Ritchie 2006; Kozak et al. 2008) techniques are used to predict current and past areas with suitable habitat for *R. kukunoris*. Comparisons between SDMs for different times facilitate explorations into the drivers of genetic diversity of the species. Herein, we evaluate the possible relationship between genetic patterns, the evolutionary history of *R. kukunoris* and Pleistocene climatic shifts.

**Materials and methods**

**Population sampling, laboratory techniques and molecular data**

A total of 360 individuals from 41 localities covering almost the range of *R. kukunoris* were analysed (Fig. 1; Tables 1 and S1, Supporting information). Two individuals of *Rana chensinensis* and one individual of *R. huanniensis* formed the out-group (Zhou et al. 2012). Because mitochondrial genomic introgression was identified from *R. kukunoris* to *R. chensinensis* along peripheral regions of both species (Zhou et al. 2012), samples from the western Qinling Mountains were excluded from our study.
Tissue samples including toe tips, muscle, livers and tadpoles were obtained following Animal Use Protocols approved by the Kunming Institute of Zoology Animal Care and Ethics Committee. Genomic DNA was extracted using the standard phenol–chloroform extraction protocol (Sambrook et al. 1989). A partial sequence of the mitochondrial gene cytochrome b (Cytb) was obtained for all individuals. We sequenced 216 individuals de novo and retrieved 144 from the study by Zhou et al. (2012). A fragment of the recombinase-activating gene 1 (RAG-1) was obtained from a subset of samples (202 individuals, Table S1, Supporting information). Primer sequences for Cytb were taken from the study by Zhou et al. (2012) and for RAG-1 from the study by Hoegg et al. (2004) and Che et al. (2010).

PCR products were purified with Gel Extraction Mini Kit (Watson BioTechnologies, Shanghai, China). The purified products were used as the template DNA for cycle sequencing reactions performed using BigDye Terminator Cycle Sequencing Kit (v.2.0, Applied Biosystems) and an ABI PRISM 3730 DNA Analyzer (Applied Biosystems). The same PCR primers were used for sequencing, and in some cases, internal primers were also used (Che et al. 2010; Zhou et al. 2012).

Data analyses

Sequences were aligned using CLUSTALX 1.81 (Thompson et al. 1997) with default parameters and then optimized by eye in MEGA 4.0 (Tamura et al. 2007). Nuclear gene sequences containing more than one ambiguous site were resolved using PHASE 2.1.1 (Stephens et al. 2001; Smith et al. 2005), accepting results with a probability >90%. This method was shown to assign alleles with high accuracy and very low false positives (Garrick et al. 2010), even in cases of small sample sizes and numerous variable sites (Harrigan et al. 2008). The input files for PHASE were prepared using SEQPHASE (Flot 2010). Recombination tests for detecting the longest nonrecombining region for nuclear locus were conducted using IMGC (Woerner et al. 2007). Identical haplotypes for mtDNA and phased nuDNA alleles were collapsed using DNASP 5.10 (Librado & Rozas 2009).

Phylogenetic analyses of the Cytb data were conducted using Bayesian inference (BI), maximum-likelihood (ML), and maximum parsimony (MP). BI analyses were performed using MRBAYES 3.1.2 (Ronquist & Huelsenbeck 2003). Nucleotide substitution models were selected based on codon position using the Akaike information criterion in MRMODELTEST 2.3 (Nylander 2004). The best-fit models for the first, second and third codon positions were K80, F81 and GTR+I, respectively. Each analysis used four Metropolis-coupled Markov chain Monte Carlo chains, with default heating values, and run for 10 million generations while sampling trees every 1000 generations. The first 50% sampled trees were discarded as burn-in. Sampled trees were analysed using AWTY (Nylander et al. 2008) to confirm satisfactory convergence of topological split frequencies. ML analyses were conducted using RAxML 7.0.4 (Stamatakis et al. 2008). These analyses implemented the GTR+I+G model for each data partition. Nodal support values were estimated from 100 nonparametric bootstrap replicates. MP analyses were implemented using PAUP* 4.0b10a (Swofford 2002). Heuristics searches with TBR were executed for 1000 random addition replicates with all characters treated as unordered and equally weighted. To assess nodal reliabilities, bootstrap analysis was conducted using 1000 replicates.

We used NETWORK 4.5 (Bandelt et al. 1999) to build a median-joining network (MJN) for both Cytb and RAG-1. After generating the initial MJN, the MP option was
chosen to remove excessive links and median vectors (Polzin & Daneshmand 2003). For mtDNA, we analysed Cytb data directly. For the nuclear genes, we used the longest nonrecombining region generated from IMGC.

Genetic variation including nucleotide diversity (\( \pi \)), haplotype diversity (\( H \)) and the number of segregating sites (\( S \)) was calculated for the mtDNA data using ARLEQUIN 3.5 (Excoffier & Lischer 2010). Isolation-by-distance analysis was carried out using a Mantel test performed with ALLELES IN SPACE (AIS) (Miller 2005). Genetic landscape shapes for both mtDNA and nuDNA data sets were also generated with AIS. The LANDSCAPE SHAPE INTERPOLATION module of AIS was used to estimate spatial patterns of genetic diversity between sampled midpoints. Residual genetic distances were used because significant correlations between genetic and geographical distances were detected using the Mantel test.

Population structure was explored using spatial analysis of molecular variation (SAMOVA) (Dupanloup et al. 2002) with 100 simulated annealing processes. Data from mtDNA and nuDNA were analysed separately. The method required the a priori definition of the number of groups (\( K \)) and generated \( F_{SC} \), \( F_{ST} \) and \( F_{CT} \) using an AMOVA approach (Excoffier et al. 1992). By exploring the behaviour of the indices \( F_{CT} \) and \( F_{SC} \) for different values of \( K \), it was possible to identify the optimum number of population groups for a set of sample populations (Dupanloup et al. 2002). We explored values of \( K \) (number of groups) ranging from 2 to 10 and selected a \( K \) value for which \( F_{CT} \) reached a plateau.

We assessed historical changes in effective population size using multiple approaches. Because population subdivision could have masked the effect of expansion, we performed these analyses for each lineage separately. First, Tajima’s \( D \) (Tajima 1989) and Fu’s \( F_s \) statistics (Fu 1997) were calculated using ARLEQUIN 3.5 to seek evidence of demographical expansions. For each test, we assessed significance by generating null distributions from 10 000 coalescent simulations. Second, pairwise mismatch distributions were implemented in ARLEQUIN 3.5 to test for signals of demographical expansion. This analysis compared the observed frequencies of pairwise differences of haplotypes with those expected under a single sudden expansion model, which would have generated a smooth, unimodal mismatch distribution (Rogers & Harpending 1992). An expected distribution under a model of sudden demographical expansion was generated using 10 000 permutations, and statistical significance was tested sum of squared deviation (SSD) and raggedness index (Rag). Finally, coalescent-based extended Bayesian skyline plots (EBSP; Heled & Drummond (2008)) were generated by BEAST 1.6.1 (Drummond & Rambaut 2007) for both nuDNA and mtDNA data to explore demographical history. EBSP provided a temporal reference to demographical events such as bottlenecks and expansions. The HKY+I+G model was assigned to the nuclear marker, and GTR+I+G model was assigned to the mtDNA. A divergence rate of 3.6% per million years for the closely related species Rana arvalis (Babik et al. 2004) was treated as reference rate of mtDNA. Relative rates of the nuDNA locus were inferred based on rates of the mtDNA. We enforced a strict molecular clock to simplify the coalescent model, and we ran four replicate analyses, each 20 million years.

Table 1 Sample locality number, sample sizes (\( n \)) and the Cytb haplotypes detected in each locality of Rana kukunoris. Greater details are provided in Table S1.

<table>
<thead>
<tr>
<th>Locality no.</th>
<th>( n )</th>
<th>Haplotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>H1, H2, H3, H4, H5, H6</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>H7, H8, H9, H10, H11</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>H1, H5, H12, H13, H14, H15, H16</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>H17, H19, H20, H21</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>H22</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>H5</td>
</tr>
<tr>
<td>7</td>
<td>27</td>
<td>H1, H5, H16, H23, H24, H25</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>H1, H26, H27</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>H1, H5, H23, H28, H29, H30, H31, H32</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>H5, H33</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>H1</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>H5, H34, H35</td>
</tr>
<tr>
<td>13</td>
<td>20</td>
<td>H5, H36, H37</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>H5</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>H1, H38</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td>H5, H39, H40</td>
</tr>
<tr>
<td>17</td>
<td>10</td>
<td>H1, H15, H41, H42</td>
</tr>
<tr>
<td>18</td>
<td>3</td>
<td>H5</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>H1, H43</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>H1, H5, H44</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>H1</td>
</tr>
<tr>
<td>22</td>
<td>6</td>
<td>H1, H5, H45</td>
</tr>
<tr>
<td>23</td>
<td>6</td>
<td>H5</td>
</tr>
<tr>
<td>24</td>
<td>6</td>
<td>H5</td>
</tr>
<tr>
<td>25</td>
<td>6</td>
<td>H5, H47</td>
</tr>
<tr>
<td>26</td>
<td>5</td>
<td>H1, H5</td>
</tr>
<tr>
<td>27</td>
<td>6</td>
<td>H5</td>
</tr>
<tr>
<td>28</td>
<td>2</td>
<td>H48</td>
</tr>
<tr>
<td>29</td>
<td>6</td>
<td>H5, H49, H50</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>H5</td>
</tr>
<tr>
<td>31</td>
<td>5</td>
<td>H1, H51</td>
</tr>
<tr>
<td>32</td>
<td>3</td>
<td>H1</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>H52</td>
</tr>
<tr>
<td>34</td>
<td>4</td>
<td>H1, H5, H53</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>H5</td>
</tr>
<tr>
<td>36</td>
<td>1</td>
<td>H5</td>
</tr>
<tr>
<td>37</td>
<td>3</td>
<td>H1, H53</td>
</tr>
<tr>
<td>38</td>
<td>5</td>
<td>H5</td>
</tr>
<tr>
<td>39</td>
<td>25</td>
<td>H1, H54, H55</td>
</tr>
<tr>
<td>40</td>
<td>10</td>
<td>H56, H57</td>
</tr>
<tr>
<td>41</td>
<td>35</td>
<td>H1, H38, H56, H58, H59, H60</td>
</tr>
</tbody>
</table>
generations long. Primarily, convergence was assessed by comparing replicate plots, to confirm that each replicate produced the same demographical patterns.

Estimated divergence time

We employed a likelihood ratio test (LRT) to assess whether or not our Cytb data followed a constant rate of molecular evolution (Huelsenbeck & Crandall 1997). Significance of the LRT assumed that the expectation of twice the absolute value of the difference in support (−Ln) under the clock vs. the nonclock model was $\sum_{i=1}^{n-2} X_i$, where $n$ equalled the number of sequences in the data set (Felsenstein 1981). The test was based on unconstrained and clock-enforced matrilineal genealogies estimated using PHYLIP 3.69 (Felsenstein 2004).

Time to most recent common ancestor was estimated using a Bayesian approach implemented in BEAST. The divergence rate of Cytb was fixed at 3.6% per million years (Babik et al. 2004). All analyses used the GTR+I+G model of nucleotide substitution. Four tree prior models (coalescent: constant size, exponential growth, logistic growth and expansion growth) were implemented, and the best model was detected using a Bayes factor test. Final analyses were undertaken with two independent runs of 20 million generations while sampling every 1000 trees. Burn-in and convergence of the chains were determined with TRACER 1.5 (Rambaut & Drummond 2007). The measures of effective sample sizes (ESS) were used to determine the Bayesian statistical significance of each parameter. The sampled trees were also analysed using AWTY to confirm satisfactory convergence of topological split frequencies.

Species distribution modelling

Data used in species distribution models (SDMs) were built based on geographical information system (GIS), including layers of environmental variables. Accessible public databases (Elith et al. 2006; Hijmans & Graham 2006) provided the fine-scaled climatic data required for this effort.

We used the maximum entropy model implemented in MAXENT 3.3.1 (Phillips et al. 2006; Phillips & Dudik 2008) to generate SDMs based on current climatic data. Locality 35 was not used because of the absence of GPS data. MAXENT outperformed similar methods of niche reconstruction (Elith et al. 2006; Phillips & Dudik 2008) when using presence-only data (Elith et al. 2006; Hijmans & Graham 2006). The probability of a species’ occurrence considered presence data with a uniform probability distribution (maximum entropy). We obtained 19 environmental variables from the WorldClim database with resolutions of 2.5 m (Hijmans et al. 2005) as environment layers (Table S2, Supporting information). We used the default parameters of MAXENT and configured the machine-learning algorithm to use 75% of species records for training and 25% for testing the model.

Assuming niche conservatism over time (Peterson et al. 1999; Holt 2003; Wiens & Graham 2005), we projected this model to LGM climatic layers to predict the distribution of R. kukunoris during the LGM. Both the community climate system model (CCSM) (Collins et al. 2006) and model for interdisciplinary research on climate (MIROC) (Hasumi & Emori 2004) were used to generate predictions for distributions during the LGM. We obtained these data layers from the WorldClim database at a spatial resolution of 2.5 m. Maximum training sensitivity plus specificity was applied as threshold rule for all the analyses because sensitivity–specificity combined approaches were suggested as good thresholds (Liu et al. 2005).

Results

Sequence information

A 798-bp segment of Cytb was obtained from 360 individuals of R. kukunoris collected from 41 populations including 144 sequences from the study by Zhou et al. (2012). The sequences contained 60 unique haplotypes for the in-group (Tables 1 and S1, Supporting information). Each of the three individuals in out-group had unique haplotypes. In total, 70 sites were variable, 31 of which were potentially parsimony informative.

One nuclear gene fragment, RAG-1, was obtained from a subset of our samples (Table S1, Supporting information). The data included 202 sequences. After trimming the ends, a 575-bp fragment was resolved, of which four sites exhibited variation and three were potentially parsimony informative. All newly obtained sequences were deposited in GenBank (Table S1, Supporting information).

Matrilineal genealogy

The BI, ML and MP analyses obtained similar topologies (Figs 1 and S1, Supporting information) and recovered distinctive lineages N and S. Lineage N was comprised of sublineages N1 and N2 from localities 2 and 4, respectively (Figs 1 and S1, Supporting information). Haplotypes from all other localities grouped in lineage S with an unresolved polytomy. Lineage S and sublineage N2 were highly supported by ML and MP, but not BI (Fig. S1, Supporting information). The network analyses for Cytb (Fig. 2) showed similar patterns to the gene tree (Figs 1 and S1, Supporting information) and further indicated three clusters for lineage S. Star-like
clusters I and III were connected by cluster II. Haplotypes H5 and H1 were common and widespread. Whereas clusters I and III were widely distributed south of the Qilian Mountains, samples of cluster II occurred in the north-eastern edge of the QTP and in a few localities only.

Owing to a limited number of potentially parsimony-informative sites, the network of RAG-1 did not produce a well-resolved tree. However, four private alleles were recovered in populations from the northern Qilian Mountains, and each locality had two alleles (Fig. 3): A1 and A2 in locality 4, and A5 and A7 in locality 2. Further, most samples from locality 41 had private allele A6. This locality occurred in the border zone between the northern and southern lineages along the eastern edge of the northern Qilian Mountains. The widespread allele in the centre of the network (A3) mainly occurred south of the Qilian Mountains.

Genetic structure and demographical history

For Cytb, the overall nucleotide diversity (\( \pi \)) was 0.00414, haplotype diversity (\( H \)) was 0.835, and the number of segregating sites (\( S \)) was 70. The Mantel test detected a significant correlation between geographical distance and genetic distance (\( r = 0.177, P < 0.001 \)). Genetic landscape shapes for R. kukunoris based on mtDNA and nuDNA (Fig. 4a, b) produced a clear surface plot, and the analyses resolved higher genetic diversity in northern parts of the QTP than in other places.

The results of SAMOVA are illustrated in Fig. 5. Table S3 (Supporting information) presented more detailed data. For the mtDNA data, \( F_{CT} \) equalled 0.7517 when \( K = 3 \), and it stabilized for higher values of \( K \). Two populations from north of the Qilian Mountains (localities 2 and 4) were suggested as two separated groups, and other samples were included in the third group. This pattern was consistent with the matrilineal genealogy (Fig. 1) and network analyses (Fig. 2). Compared with the mtDNA analyses, SAMOVA based on nuDNA data suggested division of four groups. The population from the border zone between north and south of Qilian Mountains (locality 41) was separated as the extra group.

For sublineage N1, Tajima’s \( D \) and Fu’s \( F_s \) were negative but not significant (Table 2). Mismatch distribution analyses rejected the hypothesis of sudden expansion (\( P_{SSD} = 0.0003 \) and \( P_{RAG} = 0.9988 \)), and EBSP failed to detect population expansion (Fig 6a). For sublineage N2, significant negative values of Tajima’s \( D \) and Fu’s \( F_s \) indicated population expansion (Table 2). The mismatch distribution analyses did not reject the
sudden expansion model ($P_{\text{SSD}} = 0.4822$ and $P_{\text{RAG}} = 0.565$). EBSP obtained similar results and suggested that expansion happened at about 10,000 years ago (Fig. 6b). For lineage S, all three analyses detected demographical expansion and EBSP suggested that the expansion occurred about 12,000 years ago (Table 2; Fig. 6c).

**Divergence time estimate**

The null hypothesis of clock-like sequence evolution for Cytb was not rejected by the LRT ($\chi^2 = 32.91$, d.f. = 61, $P > 0.05$). The Bayes factor test suggested expansion as being the best tree prior model, and it was significantly better than other models (Table 3). *Rana chensinensis* and *R. kukunoris* (Node 1 in Table 4) diverged approximately 1.3 Ma (95% CI: 0.86–1.71 Ma). Divergence between lineages N and S (Node 2 in Table 4) happened about 0.43 Ma (95% CI: 0.26–0.6 Ma). Sublineages N1 and N2 split (Node 3 in Table 4) at about 0.22 Ma (95% CI: 0.11–0.3 Ma). The radiation of lineage S (Node 4 in Table 4) happened at about 0.27 Ma (95% CI: 0.17–0.35 Ma).

**Species distribution modelling**

We built a SDM (Fig. 7) using 40 sites for *R. kukunoris*. The SDM had excellent predictive power for occurrences under current conditions. The AUC for the receiver operating characteristic curve of the test data (0.873) was close to that of the training data (0.925) and substantially better than that of a random model (0.5), which implied that the results greatly differed from random.

Areas that retained suitable habitat for *R. kukunoris* in the LGM based on CCSM and MIROC differed slightly, but both approaches suggested range fluctuation after the LGM. Range shift was more serious in the south-eastern edge of the QTP compared with the north-eastern edge. During the LGM, large areas in the northern parts of the QTP retained suitable habitat for...
R. kukunoris, whereas suitable habitat in the south was rare. Suitable habitat was also detected north of the Qilian Mountains during the LGM.

Discussion

Lineage divergence

Taken together, the matrilineal genealogy (Fig. 1) and network analyses (Fig. 2) reveal a clear north-south split (N and S) in *R. kukunoris*. The Qilian Mountains generally separate the N and S lineages, and no locality has mtDNA haplotypes from both lineages. A second split occurs between northern populations (N1 and N2). The analysis of mitochondrial variation by SAMOVA supports this three-group division (Fig. 5). Our nuDNA data conform to this pattern in that four private alleles (A1

**Table 2** Statistics of neutrality test based on *Cytb* data for each lineage and sublineages

<table>
<thead>
<tr>
<th></th>
<th>Lineage S</th>
<th>Lineage N1</th>
<th>Lineage N2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tajima’s <em>D</em></td>
<td>−2.26773</td>
<td>−1.37919</td>
<td>−2.08319</td>
</tr>
<tr>
<td><em>P</em>-value</td>
<td>≤ 0.001</td>
<td>0.071</td>
<td>0.004</td>
</tr>
<tr>
<td><em>Fs</em></td>
<td>−27.12550</td>
<td>−1.82660</td>
<td>−2.75693</td>
</tr>
<tr>
<td><em>P</em>-value</td>
<td>≤ 0.001</td>
<td>0.060</td>
<td>0.005</td>
</tr>
</tbody>
</table>

*R. kukunoris*, whereas suitable habitat in the south was rare. Suitable habitat was also detected north of the Qilian Mountains during the LGM.

© 2012 Blackwell Publishing Ltd
Table 3 Bayes factor of each tree prior model. 2lnBF ($H_1$ vs. $H_0$) > 10 is treated as decisive support for hypothesis

<table>
<thead>
<tr>
<th>Tree prior model</th>
<th>Ln (Bayes factor)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>$-2133.839$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$-2134.786$</td>
</tr>
<tr>
<td>Expansion</td>
<td>$-2121.020$</td>
</tr>
<tr>
<td>Logistic</td>
<td>$-2137.925$</td>
</tr>
</tbody>
</table>

Table 4 Results of divergence time estimate. Time of divergence estimates among monophyletic lineages in millions of years (Ma) at 1.8%/Ma substitution rates calculated from Bayesian coalescent phylogenetic estimation of time to most recent common ancestor (implemented in BEAST). Showing mean divergence time (Ma) and 95% highest posterior density (HPD) range

<table>
<thead>
<tr>
<th>Node</th>
<th>Mean divergence time (Ma)</th>
<th>95% HPD range (Ma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.30</td>
<td>0.86–1.71</td>
</tr>
<tr>
<td>2</td>
<td>0.43</td>
<td>0.26–0.60</td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td>0.11–0.30</td>
</tr>
<tr>
<td>4</td>
<td>0.27</td>
<td>0.17–0.35</td>
</tr>
</tbody>
</table>

and A2 in locality 2 and A5 and A7 in locality 4) occur in populations north of the Qilian Mountains (Fig. 3).

Differences in environment and landscape probably drive the split between lineages N and S. An important climatic transition zone occurs in the region of the Wushao Mountains, which are located north-east of the Qilian Mountains (Fig. 1). North of Qilian Mountains contains the Hoxi Corridor, which is dominated by an arid climate unsuitable for amphibians. Geographically, the Wushao Mountains also mark the boundary between the endorheic rivers (Shule, Hei and Shiyang rivers) and rivers that drain into seas, such as the Yellow River.

Sublineages N1 and N2 occur exclusively at the foot of the northern Qilian Mountains. Presently, *R. kukunoris* only occurs near some creeks. Three major northwards flowing endorheic rivers originate from glaciers on the Qilian Mountains (Fig. S2, Supporting information). Our samples come from two of these rivers. Locality 2 is from the Hei River Basin and locality 4 from the Shiyang River Basin. These rivers do not connect with each other, and the southern edge of the Gobi Desert isolates them. These ‘river islands’ provide suitable, limited habitat for the frog; intervening deserts restricted dispersal between the river islands because of desertification and aridity.

The *SAMOVA* and MJN analyses for the nuDNA data suggest that locality 41 is an independent group (Fig. 3 and 5). Locality 41 consists of the small, endorheic Shagou River, which flows independently from the other river systems (Fig. S2, Supporting information). Most samples from locality 41 have allele A6. The discordance between mtDNA and nuDNA may indicate gene flow between locality 41 and populations south of Qilian Mountains. Some individuals in lineages S and N share allele A3 (Fig. 3), and this occurrence indicates either gene flow or the retention of ancestral polymorphisms. Most populations have A3 and private nuDNA alleles only occur in localities 2 (A5 and A7) and 4 (A1 and A2) (Fig. 3; Table S1, Supporting information).

Additional informative nuclear data from *R. kukunoris* and sister species, such as *R. chensinensis*, are required to distinguish between the possibilities.

Estimated dates for the diversification of *R. kukunoris* vary depending on the assumed substitution rate. Application of the fastest estimated rate for Cytb (Babik et al. 2004) suggests a Pleistocene diversification around 0.43 Ma, long before the LGM (0.023–0.018 Ma). Using secondary calibration points from the study by Bossuyt et al. (2006) and Zhou et al. (2012) estimated a much older time (3.5 Ma). In terms of lineages and sublineages within *R. kukunoris*, both calibrations yield divergence times that predate the LGM. No evidence suggests that *R. kukunoris* re-treated to a single refugium at time of maximum glaciation.

**Glacial refugia and demographic expansion**

Our study supports the hypothesis of multiple refugia and suggests they occurred on the northern and north-eastern edge of the QTP. These refugia are associated with maternal lineages and patterns of nuDNA variation.

Lineage S occurs on the eastern escarpment of the QTP south of the Qilian Mountains, and it does not have discernible sublineages. The absence of sublineages probably reflects a post-LGM expansion around 12 000 years ago. The star-like networks, Tajima’s *D*, Fu’s *Fs*, mismatch distribution and EBSP support this scenario. Ecological modelling identifies the north-eastern edge of the QTP as a refugium during the LGM (Fig. 7). SDMs based on current climatic data suggest that the Hengduan Mountains did not offer suitable habitat for *R. kukunoris* during the LGM (Fig. 7). Thus, quick expansion after the LGM from a single refugium best explains the present genetic pattern in lineage S.

Lineages N1 (locality 2) and N2 (locality 4) in the northern Qilian Mountains are unexpected. Their origin dates to 0.22 Ma and the minimum date of 0.11 Ma long predates the LGM (0.021 Ma). Locality 2 has private alleles A1 and A2 in *RAG-1*, and locality 4 possesses A5 and A7. The environmental modelling
(Fig. 7) supports this multiple refugia hypothesis by revealing suitable habitat along the northern Qilian Mountains during the LGM. Population expansion is detected in locality 4 (sublineage N2) by all three analyses. But in locality 2 (sublineage N1), this is excluded by mismatch distribution analyses and the other two analyses (Fig. 6). Results of SDMs suggest that locality 2 lies outside of the suitable habitat both now and during the LGM after applying the threshold (Fig. 6b, d, f). Further, locality 41 may be a microrefugium because the SAMOVA based on nuDNA indicates it is a distinct group. However, owing to the limited information from RAG-1, we do not further speculate on this possibility. Additional mtDNA and nuDNA data are required to test the hypothesis.

Suitable habitats in refugia may permit species to persist for long periods of time and even to speciate. Unique genotypes and high levels of diversity often occur in these locations (Hewitt 2000). Our analyses conform to this tendency because higher levels of genetic diversity occur in the north-eastern edges of the QTP (Fig. 4). The tendency for genetic differences to increase northwardly also occurs in the landscape genetic analysis (Fig. 4), supporting our hypothesis of multiple refugia for *R. kukunoris* in the north-eastern edges of the QTP and northern Qilian Mountains.

Our discovery of northward refugia contrasts with the trend for southwards re-treating during the LGM. Lineage S appears to have dispersed southwards after LGM. The genetic pattern for *R. kukunoris* strongly contrasts with the prevailing trend for the region in which the south-eastern edge of the plateau holds a higher level genetic diversity, which is based mostly on analyses of alpine plants including herbs (*Pedicularis longiflora*) (Yang et al. 2008) and conifers (*Juniperus przewalskii*) (Zhang et al. 2005). Several recent studies support the occurrence of multiple refugia in the region, such as for the blood pheasant (*Ithaginis cruentus*; Zhan et al. 2011), the fossorial rodent of plateau zokor (*Eospalax baileyi*; Tang et al. 2010) and the pine *Picea crassifolia* (Meng et al. 2007). Present study from *R. kukunoris* is consistent with this scenario that multiple refugia are largely responsible for the present spatial distribution of this plateau endemic frog. However, importantly, our study indicates that the separated microrefugia lay in the northern edge of QTP (the north of Qilian Mountains), not on the plateau.

**Implications for conservation**

The maintenance of genetic diversity prevents loss of the evolutionary potential of species. The distinct
genetic differentiation within *R. kukunoris* has important implications, not only for the conservation of this species, but also for all biota in Northwest China.

Genetic criteria can be used to design at evolutionarily significant units (ESUs) and management units (MUs) (Moritz 1994). Within this framework, our analyses suggest that ranges of the three maternal lineages (N1, N2 and S) qualify as MUs. All of these require protection because of their genetic uniqueness, which involves exclusive mtDNA haplotypes and private nuclear alleles (Fig. 3).

The northern Qilian Mountains merit special attention. Desertification is the major ecological problem in Northwest China. Recently, both the snowline and timberline have risen in these mountains, and this decreases the supply of water for the associated rivers. Further, livestock overgrazing and excessive farming are causing habitat reduction (Qin 2002). Presently, *R. kukunoris* only occurs in and near these endorheic river basins. Protection measurements for *R. kukunoris* should focus on habitat conservation, and our genetic analyses can help guide conservation efforts. Other species in this region require evaluation for their conservation needs, and planned sustainable development can reduce anthropogenic pressures on the habitat of these ‘river islands’.

Conclusions

Our study documents genetic structure within *R. kukunoris*. The arid environment and landscape features, such as river systems north of the Qilian Mountains, drive divergence between the lineages and sublineages. Genetic and environmental data identify the north-eastern edge of the QTP as major refugium for *R. kukunoris*. Two microrefugia in the northern Qilian Mountains have also had a great influence on the genetic structure of this frog. Populations associated with lineage S and sublineage N2 appear to have experienced sudden demographical expansions after the LGM. Our results suggest that populations north of the Qilian Mountains require monitoring to assure perpetuity of their genetic diversity via habitat conservation in Northwest China.

Acknowledgements

We thank Yongbiao Xu, Shaoying Liu and Yang Liu for assistance in the field and for providing tissue samples in their care. Technical support in the laboratory of Yaping Zhang is gratefully acknowledged. This work was supported by grants from the National Basic Research Program of China (973 Program, 2007CB411600), the National Natural Science Foundation of China (30700065), programmes of Chinese Academy of Sciences, the Bureau of Science and Technology of Yunnan Province, and a Natural Sciences and Engineering Research Council (Canada) Discovery Grant 3148.

References


© 2012 Blackwell Publishing Ltd
RIVER ISLANDS, REFUGIA AND BROWN FROGS


Heled J, Drummond A (2008) Bayesian inference of population size history from multiple loci. BMC Evolutionary Biology, 8, 289.


Qu YX, Cheng X, Comes HP (2011) Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world’s most diverse temperate flora. Molecular Phylogenetics and Evolution, 59, 225–244.


© 2012 Blackwell Publishing Ltd
This work is part of the PhD dissertation of W.Z. whose research uses genetic markers to address the evolution of Chinese amphibians. F.Y. is a PhD student pursuing diversity and conservation genetics of amphibians and reptiles. J.F. is a molecular phylogeneticist mainly interested in speciation, hybridization, and biogeography of amphibians and reptiles. S.W. is a technician. J.C. is a herpetologist who mainly focuses on the evolution of Asian amphibians. R.W.M. and Y.Z. are broadly interested in genetics and evolution.

Data accessibility

DNA sequences have been deposited in GenBank under Accession numbers JX486130–JX486547. Details regarding individual samples are available in Table S1, and the bioclimatic data used in species’ distribution modeling for Rana kukunoris are listed in Table S2 (Supporting information).

Supporting information

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

Fig. S1 Maximum likelihood matrilineal genealogy based on the 63 haplotypes from Cytb. Vertical bars show the lineage/sublineage assignment.

Fig. S2 River system in the northern Qilian Mountains.

Table S1 Detailed information for specimens included in this study.

Table S2 Information of bioclimatic data used in SDMs construction.

Table S3 Results of AMOVA based on population groups suggested by SAMOVA.