Molecular analysis of the Spanish Creek *Rana* population (Plumas County, California) supports taxonomic placement within *Rana boylii*

Photo by Roland Knapp, Spanish Creek, April 2013

Thomas J. Poorten

Department of Environmental Science Policy and Management, University of California, Berkeley, CA 94720, poorten@berkeley.edu

Erica Bree Rosenblum

Department of Environmental Science Policy and Management, University of California, Berkeley, CA 94720

Roland A. Knapp

Sierra Nevada Aquatic Research Laboratory, University of California, 1016 Mount Morrison Road, Mammoth Lakes, CA 93546

Introduction

The main question that we sought to answer was whether a frog population in Spanish Creek, Plumas County, California is affiliated with the Sierra Nevada yellow-legged frog *Rana sierrae* or the foothill yellow-legged frog *Rana boylii*. The implementation of appropriate conservation measures for this population requires accurate taxonomic placement, but this placement is sometimes difficult. Historically, the ranges of *R. sierrae* and *R. boylii* abutted each other at mid-elevations in the southern, western, and northern Sierra Nevada, and the species were sometimes found in close proximity to each other (Zweifel 1955). Previous efforts to identify the species represented by the Spanish Creek population using morphological and genetic analyses have yielded ambiguous results (e.g., Wengert 2008). Overlap in morphological characters can result in confusion between these two species. Both species share the characteristic yellow coloration on ventral surfaces of the abdomen and the hind limbs, and in some geographic areas both species display broadly overlapping characteristics. Despite morphological similarity, previous work has found significant genetic divergence (~11% dissimilar over 2kb of mtDNA) between the two species (Macey et al. 2001). Thus molecular markers can be used to confidently diagnose species affiliation.

A previous study on the range-wide phylogeography of *R. boylii*, which used molecular markers, is relevant to our study because it included two sampling locations near our focal study site (Lind et al. 2011). This study included one individual sampled from Spanish Creek (~6km downstream from our sampling site), which was identified as *R. boylii* (Figure 1). The other sampling locality was on Bean Creek, a tributary to Spanish Creek ~2.5 km upstream from our site (Figure 1). Four samples were collected at this locality, of which three were identified as *R. boylii* and one was identified as *R. sierrae* (Lind et al. 2011). This provides evidence that *R. sierrae* can at least occasionally be found in Bean Creek. However, the identity of frogs in Spanish Creek remains unclear.

In the present study, we used a non-invasive swabbing method to collect genetic samples from seven individuals in Spanish Creek (Figure 1). We targeted our sampling within the area where extensive stream stabilization measures are proposed. We analyzed mitochondrial and nuclear gene sequences and conclude that the individuals we sampled on Spanish Creek are unambiguously affiliated with *R. boylii*. However, a larger sample size will be necessary to determine whether *R. sierrae* do in fact exist in Spanish Creek at low density.

Methods

Sample Collection

We collected seven samples from the Spanish Creek population, on property owned by the Spanish Ranch (Figure 1; 39.9576 N, -121.0817 W; approximately 11 km west of Quincy, Plumas County). Samples were taken from each of three adults and four metamorphs. The four metamorphs were all found in a single pool and may have come from the same egg clutch, which would reduce the expected genetic variability. We obtained genetic material from each individual by rubbing a rayon swab across the ventral surface for approximately 30 seconds. Swabs were air-dried, placed individually into sterile microcentrifuge tubes, and refrigerated until analysis. Frog DNA was extracted from the swabs using the Prepman Ultra Reagent according to the manufacturer's protocol.

Sequence Data Collection

We performed PCR to amplify three molecular markers – gene fragments of Cytochrome B (CytB), NADH Dehydrogenase 2 (ND2), and Tyrosinase (Tyr). The first two genes are mitochondria-derived and the third is nuclear. No published nuclear data are available for *R. sierrae*, so we sequenced the Tyrosinase gene fragment from a known *R. sierrae* individual sampled previously in Yosemite National Park (Table 1). The primers were custom designed to anneal to mitochondrial and nuclear regions that are conserved between *R. sierrae* and *R. boylii* based on published sequence data in Genbank (Table 1). We performed Sanger sequencing at the DNA Sequencing Facility (UC Berkeley) using forward and reverse primers. Cases where PCR amplification produced multiple amplicons, as observed on agarose gel electrophoresis, were excluded from sequencing runs. The amplification discrepancies likely resulted from variable template concentration among samples, which is an apparent disadvantage of using swab-extracted DNA. Further optimization of PCR conditions will be required to obtain increased PCR specificity for such samples (e.g. optimize template dilutions, primer annealing temperature, $MgCl_2$ concentration).

Data Analysis

For each locus, we aligned the sequence reads to reference sequences using the multiple sequence alignment tool ClustalX (ver. 2.1; Larkin et al. 2007). The sequence data from the four metamorphs did not contain any polymorphic sites and, therefore, we used data from only one of the metamorphs and all three adults for subsequent analyses. As a first step to estimate genetic divergence between the samples and reference sequences, we calculated the percentage of nucleotide positions

that differed for each pair of sequences using the *ape* package in the R statistical software (Paradis et al. 2004). Next we performed a phylogenetic analysis of the aligned sequences using a maximum likelihood approach. This approach differs from the sequence divergence analysis in that a particular model of sequence evolution is applied while seeking to reconstruct ancestral relationships among the sequences. This procedure attempts to account for the vagaries of the DNA mutation process among sites in the genome. We determined the best fitting model of sequence evolution using the jModelTest software (ver. 2.1.3; Posada 2008). The selected models were HKY+G for CytB and ND2; and K80+I for Tyrosinase. We estimated the maximum likelihood phylogenetic tree for each locus using PAUP* (ver. 4.0b10, Swofford 2003). We estimated nodal support using 1,000 bootstrap replicates with the same maximum likelihood model parameters. The bootstrap values, which were calculated as the percentage of bootstrap replicates that contain a given node, provide a statistical measure for the certainty of the inferred phylogenetic interrelationships.

Results

Comparison of the pairwise sequence divergence between the sequences from Spanish Creek and the reference sequences indicate that for all three loci (two mitochondrial and one nuclear) the Spanish Creek sequences are more similar to *R. boylii* than to *R. sierrae* (Table 2). The range of divergence values among the three loci may result from variable mutation rate among loci and/or from variation in the source locations of the reference samples. In either case, all three within-locus comparisons show that the Spanish Creek samples are less diverged from the *R. boylii* reference sequences than from the *R. sierrae* reference sequence. The cytochrome B locus, in particular, provides strong support for placement within *R. boylii*. For this locus (770 bp), three out of four Spanish Creek samples are identical to the *R. boylii* reference sequence, and the fourth sample contains only a single polymorphic site.

The phylogenetic analysis provided further support for the taxonomic placement of the Spanish Creek samples within *R. boylii*. For all three loci, the Spanish Creek samples were connected to the known *R. boylii* samples by very short branch lengths (Figure 2-4). The bootstrap values provide strong support for the placement of the samples into the *R. boylii* clade. In some cases, relationships among clades were weakly supported. This phylogenetic uncertainty likely results from insufficient data to sort out the order of short internal branches. Regardless of the uncertainty in internal branches, we emphasize that the Spanish Creek samples were consistently joined to *R. boylii* with strong support.

Discussion

The results presented here strongly support taxonomic placement of the Spanish Creek samples within *R. boylii*. However, further work is needed to determine if only *R. boylii* occur in the area. To determine if *R. sierrae* is absent will require more intensive sampling over multiple seasons.

Given our results, we cannot completely rule out alternative, yet unlikely, interpretations regarding the ancestry of the samples. For example, one could hypothesize that sex-biased gene flow occurred in the population's genealogical history. With uniparental inheritance of the mitochondrial genome, there is a possibility that an *R. boylii*-type mitochondrial genome resides alongside a nuclear genome with mixed ancestry of *R. boylii* and *R. sierrae*. Although the sequence data from the nuclear Tyrosinase gene does not support this scenario, additional data from multiple nuclear loci are necessary to rule it out. Further work can be done to test the hypothesis that the population contains mixed ancestry of *R. boylii* and *R. sierrae*. However, prior evidence suggests that admixture is unlikely. For example, breeding trials between *R. boylii* and *R. sierrae* adults were not successful (Zweifel 1955, page 262). Additionally, the large amount of sequence divergence observed in the mitochondrial sequences suggests long-term reproductive isolation between these two species.

In summary, our results support previous work indicating that *R. boylii* occur in Spanish Creek, but more work is needed to determine whether *R. sierrae* also occur in the area and to describe the exact distribution of both species.

Literature Cited

Bossuyt F, Brown RM, Hillis DM, Cannatella DC, Milinkovitch MC (2006) Phylogeny and biogeography of a cosmopolitan frog radiation: Late cretaceous diversification resulted in continent-scale endemism in the family Ranidae. Systematic Biology 55: 579-594

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

Lind AJ, Spinks PQ, Fellers GM, Shaffer HB (2011) Rangewide phylogeography and landscape genetics of the Western U.S. endemic frog *Rana boylii* (Ranidae): implications for the conservation of frogs and rivers. Conservation Genetics 12: 269-284.

Macey JR, Strasburg JL, Brisson JA, Vredenburg VT, Jennings M, Larson A (2001) Molecular phylogenetics of western North American frogs of the *Rana boylii* species group. Molecular Phylogenetics and Evolution 19: 131143.

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289-290.

Posada D (2008) JModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25: 1253–1256.

Schoville SD, Tunstall TS, Vredenburg VT, Backlin AR, Gallegos E, Wood DA, Fisher RN (2011) Conservation genetics of evolutionary lineages of the endangered mountain yellow-legged frog, *Rana muscosa* (Amphibia: Ranidae), in southern California. Biological Conservation 144:2031–2040.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Wengert G (2008) Habitat use, home range, and movements of mountain yellow-legged frogs (*Rana muscosa*) in Bean and Spanish Creeks on the Plumas National Forest. MGW Biological and Klamath Wildlife Resources, unpublished report.

Zweifel RG (1955) Ecology, distribution, and systematics of frogs of the *Rana boylei* group. University of California Publications in Zoology 54:207–292.

Table 1. Information on the reference sequences used in this study.

Locus	Species	Genbank identifier	Specimen Identifier	Source location	Source coordinates	Reference
CytB (mtDNA)	*R. boylii*	HM804099.1	HBS27566	Missouri Canyon, tributary to Bear River, Nevada Co., CA	39.217 N, -120.917 W	Lind et al. 2011
	R. sierrae	HM804086.1	KM8	Bean Creek, near crossing of USFS Road 25N17, Plumas National Forest, Plumas Co., CA	39.958 N, -121.068 W	Lind et al. 2011
ND2 (mtDNA)	R. boylii	AF314019	MVZ 148941	Along Pope Creek approx. 4.8 km NW of Lake Berryessa, Napa Co., CA	38.625 N, -122.323 W	Macey et al. 2001
	R. sierrae	JF727240.1	H2	Kuna Basin, Yosemite National Park, CA	37.813 N -119.224 W	Schoville et al. 2011
tyrosinase exon 1 (nucDNA)	*R. boylii*	DQ347184.1	isolate 1045, MVZ 148929	along Butts Creek, 0.6 km NW Napa County line, CA	38.709 N, -122.470 W	Bossuyt et al. 2006
	R. sierrae	N/A	RKS7365	Pond 2.1 km ENE of Mt. Conness, Yosemite National Park, CA	37.970 N, -119.345 W	present study

Table 2. Pairwise sequence divergence (%) between Spanish Creek samples and reference sequences for three loci (mean ± 1 standard deviation).

Locus	Spanish Creek vs. *Rana boylii*	Spanish Creek vs. *Rana sierrae*
CytB (mtDNA)	0.03 ± 0.06	14.28 ± 0.55
ND2 (mtDNA)	0.75 ± 0.12	16.44 ± 0.04
Tyr (nucDNA)	0.92 ± 0.06	1.62 ± 0.11

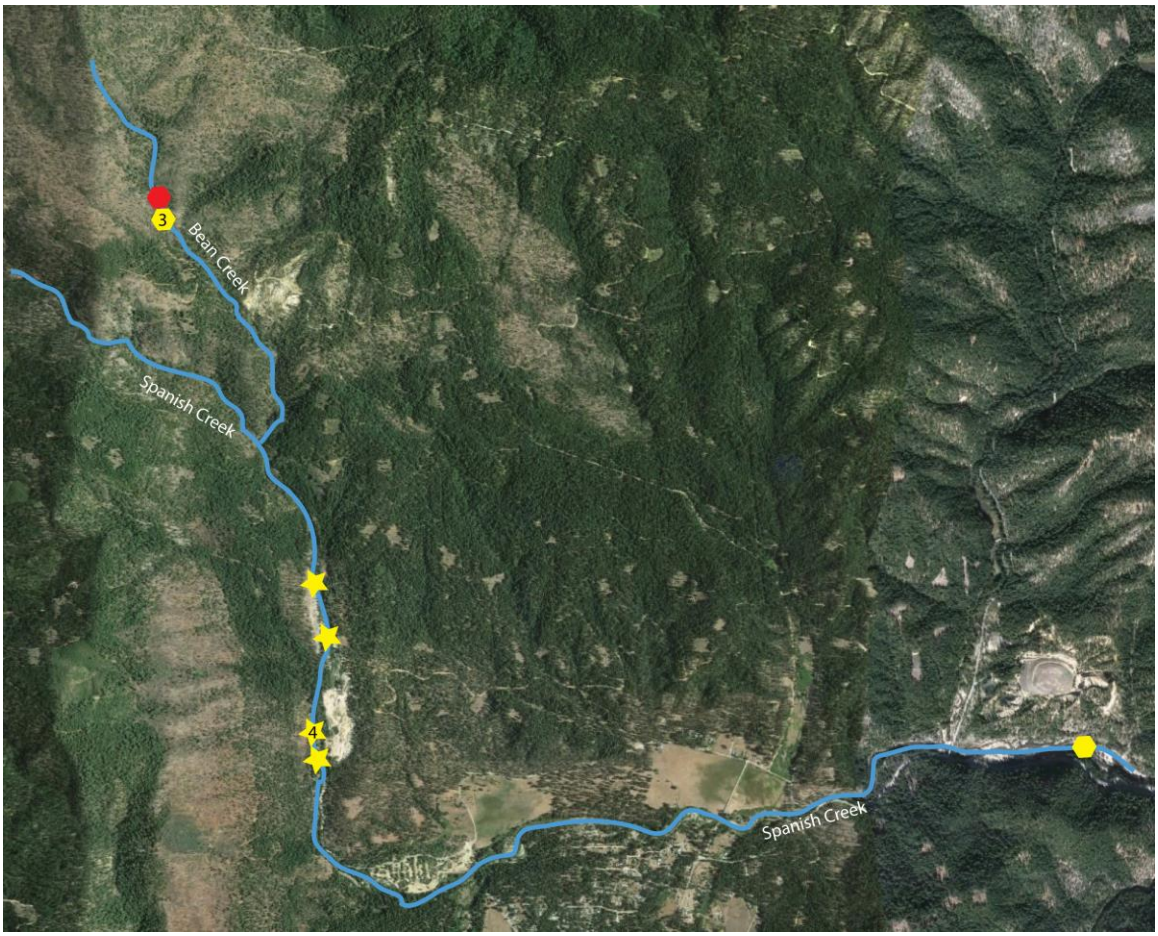

Figure 1. Map of the Spanish Creek study area showing the location of skin swab collections made during the current study (stars) and collections made during previous studies (polygons). Yellow symbols indicate samples identified using molecular methods as *Rana boylii* and the red symbol indicates *Rana sierrae*. Numbers inside of symbols indicate the number of samples collected. The lack of a number indicates that a single sample was collected from the site.

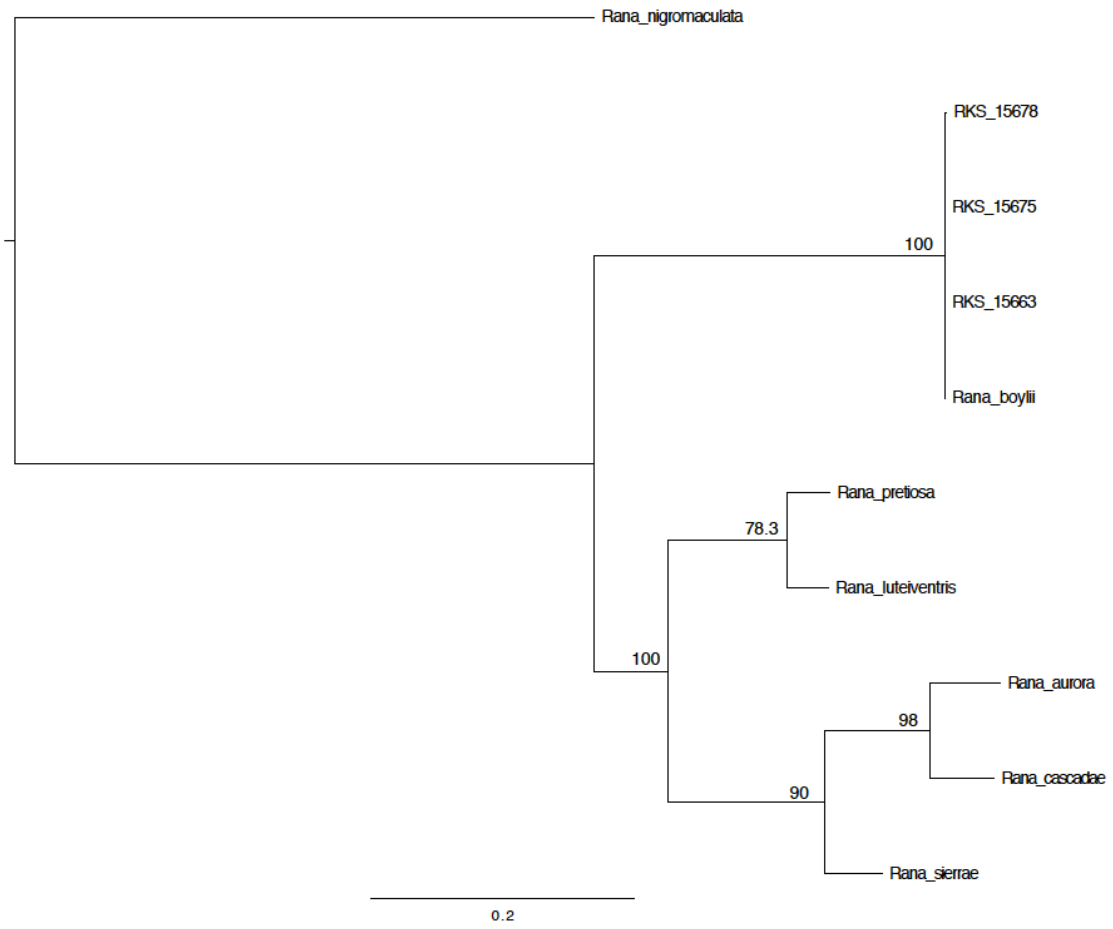

Figure 2. Maximum likelihood phylogenetic tree for Cytochrome B, 770 bp (mitochondrial). The nodes labeled as "RKS_XXXX" are the Spanish Creek samples. Internal nodes without bootstrap values had less than 50% support from bootstrap replicates. The outgroup is *Rana nigromaculata*. The reference sequences were downloaded from Genbank. *Rana nigromaculata*: AY803878.1; *Rana boylii*: HM804099.1; *Rana pretiosa*: EU708875.1; *Rana luteiventris*: AY016660.1; *Rana aurora*: EU552226.1; *Rana cascadae*: EU708878.1; *Rana sierrae*: HM804086.1.

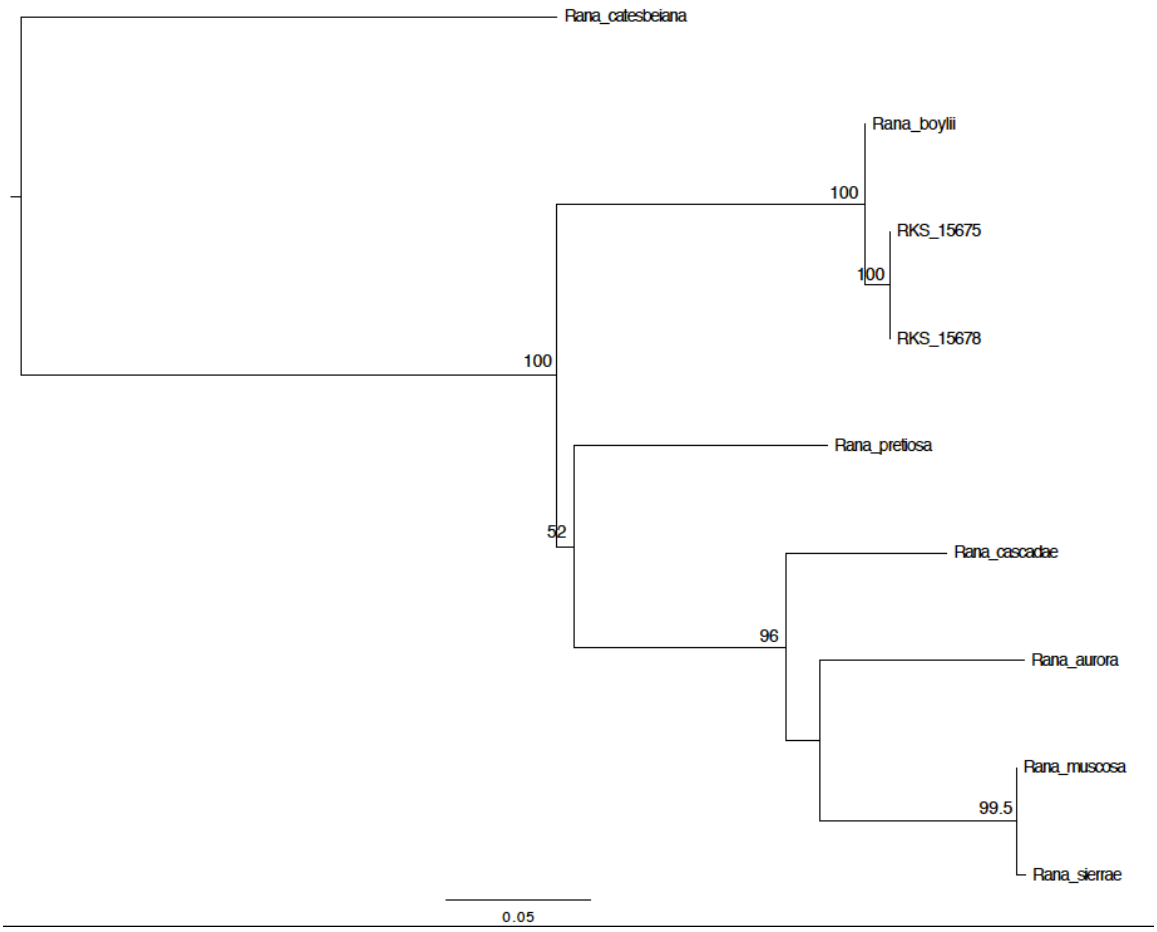

Figure 3. Maximum likelihood phylogenetic tree for NADH Dehydrogenase 2, 796 bp (mitochondrial). The nodes labeled as "RKS_XXXX" are the Spanish Creek samples. Internal nodes without bootstrap values had less than 50% support from bootstrap replicates. The outgroup is *Rana catebeiana*. The reference sequences were downloaded from Genbank. *Rana catesbeiana*: F314016.1; *Rana boylii*: AF314019.1; *Rana pretiosa*: AF314020.1; *Rana cascadae*: AF314022.1; *Rana aurora*: AF314021.1; *Rana muscosa*: AF314025.1; *Rana sierrae*: JF727240.1.

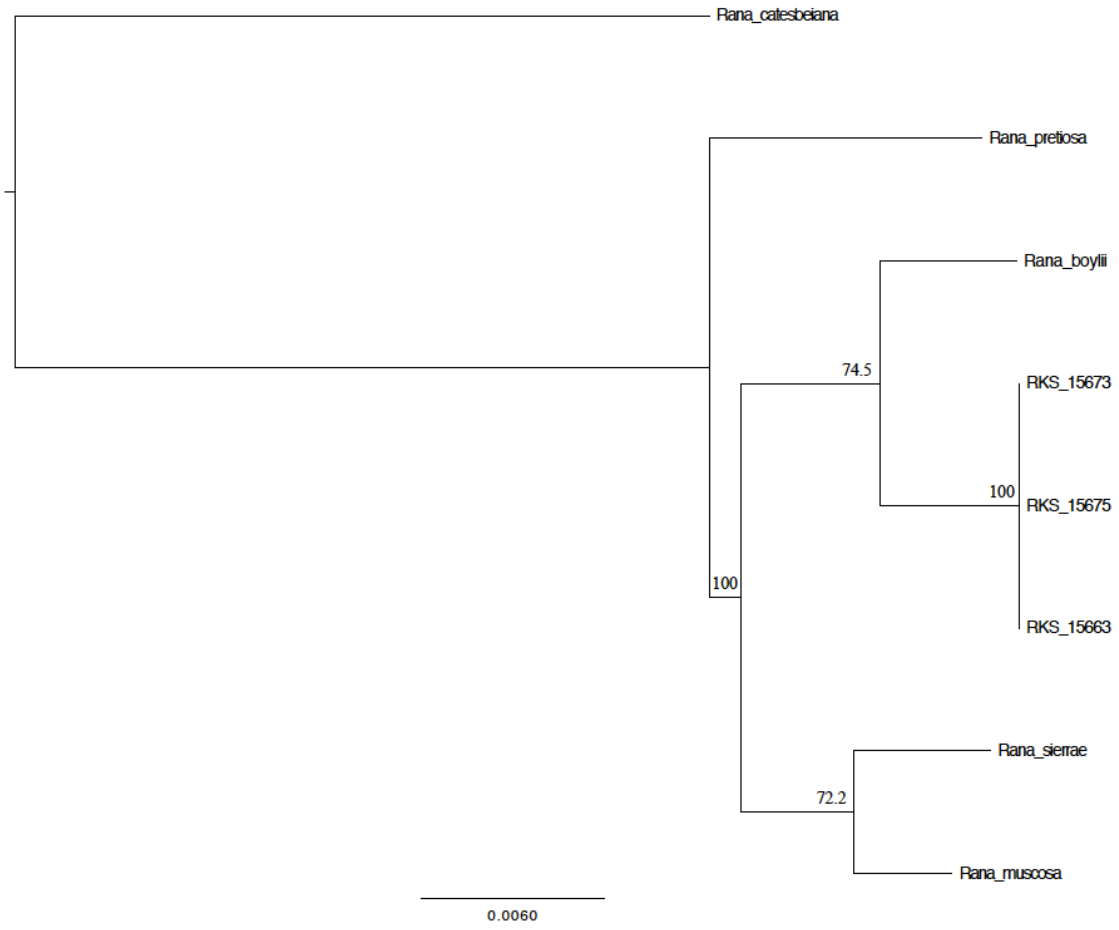

Figure 4. Maximum likelihood phylogenetic tree for Tyrosinase, 483 bp (nuclear). The nodes labeled as "RKS_XXXX" are the Spanish Creek samples. Internal nodes without bootstrap values had less than 50% support from bootstrap replicates. The outgroup is *Rana catesbeiana*. The reference sequences were downloaded from Genbank. *Rana catesbeiana*: DQ360044.1; *Rana pretiosa*: FJ845513.1; *Rana boylii*: DQ347184.1; *Rana muscosa*: DQ282945.1.