

Population genetics of a sentinel stream-breeding frog (*Rana boylii*)

By

RYAN A PEEK

B.S. (University of California, Davis) 2002
M.S. (University of San Francisco) 2010

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Ecology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA
DAVIS

Approved:

(Michael R. Miller), Chair

(Peter B. Moyle)

(Mark W. Schwartz)

Committee in Charge

2018

The Dissertation of Ryan A Peek is approved.

Peter B. Moyle

Mark W. Schwartz

Michael R. Miller, Committee Chair

September 2018

Population genetics of a sentinel stream-breeding frog (*Rana boylii*)

Copyright © 2018

by

Ryan A Peek

*“The good life of any river may depend on the perception of its music;
and the preservation of some music to perceive.”*

(Aldo Leopold)

*“One thing to remember is to talk to the animals. If you do,
they will talk back to you. But if you don’t talk to them, they
won’t talk back to you, then you won’t understand. And when
you don’t understand, you will fear, and when you fear, you will
destroy the animals, and if you destroy the animals, you will
destroy yourself”*

(Chief Dan George, Tseil-Waututh Nation, North Vancouver)

Acknowledgements

For all the curious people who have come before and hopefully after, I want to acknowledge you, and I hope we can do better to inspire and support those voices that may not have had the opportunities or privilege I have had. I am lucky to have had all I have had, and finishing a dissertation requires a community, and this dissertation would not have happened if it wasn't for the amazing community of family, friends, and colleagues who helped me every step of the way. In particular, thank you to my partner, wife and best-friend Leslie—you are my sun and gravity—you held me together, anchored our family, and made it possible to run this crazy academic ultra-marathon. To my dearest little tadpoles, Connor and Genevieve, you inspire me, you make me laugh every day, and you remind me the world still has hope as long as we nourish joy and curiosity. Thank you for being you, and I hope one day you forgive me for the amount of time I've spent staring at a computer. Thanks to my mom for all the support, love, and baked goods. Sibling, thank you for consistently inspiring me, listening to me, and being the best sibling one could ask for. And for all my close friends, bandmates, and officemates (you know who you are), you keep me sane, you motivate me, and you remind me every day that I really love this crazy journey. John, your shed and couch have probably single-handedly kept me anchored in ways I can't even express...also your friendship. Thank you for your time, humor, and general levity. Thanks to my Dad, who has cajoled, pestered, and annoyed me for far too long to "get a PhD", thanks for believing it was possible even when I didn't. Also, please never suggest anything like this again. And to my committee and my colleagues at the Center for Watershed Sciences, you have all been an amazing resource in providing feedback, guidance, and support throughout my graduate student career. Finally, to my cohort and fellow students in the GGE, this has been a great place to grow and mature as a scientist and researcher. Thank you all.

Ryan A Peek
September 2018
Ecology

Abstract

Rana boylii is an imperiled frog species native to CA and OR, and it is currently designated as a species of special concern (CDFW) in the state of CA. It has been petitioned as candidate for federal (USFWS) and state (CDFW) listing. As a lotic breeding amphibian, *R. boylii* is tied closely to local flow regimes in the watersheds it inhabits and is therefore particularly sensitive to alterations to the natural flow regime. Effective conservation management of this species should consider and prioritize maintenance of genetic diversity as part of any listing decision because it is closely related to the evolutionary capacity for adaptation to environmental changes. Conservation of genetic diversity in this species will require several components, including refining potential conservation units (i.e., distinct population segments) and quantifying of genetic diversity and genetic diversity trajectories across the species range. To assess these components, fine-scale and landscape-scale analyses were conducted using genomic data from over 600 samples from 89 localities across the range of the species. Six genetically-distinct groups were identified, as well as population subdivisions at local watershed scales. One major impact on *R. boylii* populations has been river regulation. River regulation has been implicated as a cause of fundamental changes to downstream aquatic ecosystems. Regulation changes the natural flow regime which may restrict population connectivity and decrease genetic diversity in some species. Since population connectivity and the maintenance of genetic diversity are fundamental drivers of long-term persistence, understanding the extent that river regulation impacts these critical attributes of genetic health is an important goal. However, the extent to which *R. boylii* populations in regulated rivers have maintained

connectivity and genetic diversity is unknown. The impacts of river regulation on *R. boylii* were investigated with genomic data to explore the potential for long-term persistence of *R. boylii* under continued regulation. *R. boylii* in regulated rivers showed striking patterns of isolation and trajectories of genetic diversity loss relative to unregulated rivers. For example, river regulation explained the greatest amount of variance in population genetic differentiation compared with other covariates including geographic distance. Importantly, patterns of connectivity and genetic diversity loss were observed regardless of regulation level but were most prominent in locations with the greatest regulation intensity. Using the same genomic data, fine-scale analyses of *R. boylii* and *R. sierrae* in a single region of the Sierra Nevada of California was conducted to evaluate the potential for hybridization between species. Hybridization between species may combine parental genotypes in ways that yield reproductively sterile or isolated lineages, and hybridization events may be short-lived and difficult to detect. Limited hybridization between the species was detected in the Feather basin, though it appears these are terminal events based on PCA, admixture, and tests of heterozygosity using species diagnostic SNPs. Finally, rangewide quantification and comparison of genomic variation across populations indicates the southern coast, southern Sierra Nevada, and Northern Sierra/Feather basin in California should have high prioritization in conservation efforts due to low genomic diversity and trajectories of diversity loss. More broadly, these results demonstrate both the critical need for regional conservation in a sentinel river species, and the utility and power of genetic methods for assessing and monitoring sensitive species across many scales.

Contents

Abstract	v
1 Flow regulation associated with decreased genetic health of a river-breeding frog species	1
Introduction	1
Methods	5
Results	11
Discussion	21
Conclusion	23
2 Hybridization between two sympatric ranid frog species in the northern Sierra Nevada, California	25
Introduction	25
Methods	28
Results	36
Discussion	46
Conclusion	49
3 Refining conservation unit boundaries of a sentinel stream-breeding frog (<i>Rana boylii</i>) using population genomics	50
Introduction	50
Methods	53
Results	57
Discussion and Conclusion	76

A Appendix: Supplemental Tables & Code	78
Colophon	80
References	81

Chapter 1

Flow regulation associated with decreased genetic health of a river-breeding frog species

Introduction

Rivers simultaneously connect and carve the landscapes through which they flow. Rivers provide corridors of connectivity for riparian and aquatic organisms such as fish, amphibians, and macroinvertebrates (Wiens 2002, Pringle 2003), while also acting as physical barriers on the landscape for many terrestrial organisms (Voelker et al. 2013, Cazé et al. 2016). Hydrologic connectivity (Pringle 2003) transfers energy, organisms and ultimately genetic variation and thus is a critical component for population persistence in dynamic systems where populations must constantly adapt to temporal and spatial changes. In Mediterranean climates, rivers have strong seasonal patterns associated with cold, wet winters and warm, dry summers. Native aquatic organisms have evolved life histories

INTRODUCTION

well adapted to these natural patterns, which are both predictable and seasonal (Yarnell et al. 2010, Tonkin et al. 2017).

River regulation, or the hydrological alteration of flow by dams and diversions, impacts the seasonal and interannual flow variability within a watershed. Regulation changes the natural flow regime and dramatically alters geomorphic and hydrologic connectivity of watersheds (Poff et al. 2007), which may restrict natural population connectivity (Schick and Lindley 2007, Shaw et al. 2016). River regulation can change flow frequency, magnitude, duration, timing, and rate of change, which can have significant impacts on aquatic organisms and ecological processes (Poff et al. 2007, Yarnell et al. 2010). River regulation, and more specifically, regulation associated with hydropower generation, has been implicated as a cause of fundamental changes to downstream aquatic ecosystems (Power et al. 1996, Bunn and Arthington 2002, Moyle et al. 2011). The hydrological regimes of over half of the world's largest rivers have been altered by large dams (Nilsson et al. 2005) and only recently has the extent of flow alteration and the associated ecosystem-level impacts been acknowledged (Pringle 2001, Dudgeon et al. 2006, Murchie et al. 2008).

Changes to abiotic processes caused by river regulation can have a substantial impact on biotic communities. The negative effects of river regulation on migration and loss of spawning habitat (Lind et al. 1996, Fuller et al. 2011, Kupferberg et al. 2012, Rolls and Bond 2017), reductions in population abundances and diversity (Zhong and Power 1996, Vorosmarty et al. 2010, Fuller et al. 2011, Werth et al. 2014, Scribner et al. 2016, Sabo et al. 2017, Guzy et al. 2018), and fragmentation (Vorosmarty et al. 2010, Werth et al. 2014, Scribner et al. 2016, Sabo et al. 2017, Guzy et al. 2018) have been well documented. However, most rivers have not been regulated for long periods (e.g., less than 100 years) compared to the time these organisms had to adapt to pre-anthropogenic

INTRODUCTION

river flow. In regulated rivers that organisms still occupy, it remains unknown whether populations can persist long-term with continued regulation. In other words, while some species may have persisted since regulation began in a system (e.g., several decades), this does not necessarily mean these populations will persist into the future under current flow regulation regimes. Thus, exploring the potential for long-term persistence of populations under different flow regimes is a crucial component for guiding conservation efforts yet remains a significant gap.

One tool that can help address this gap is the integration of genetics and hydrology to better assess the impact of river regulation on aquatic organisms (Scribner et al. 2016). Although aquatic organisms are often difficult to count and monitor by conventional methods, genetic monitoring can be a powerful tool to assess population health by revealing factors such as fragmentation and population declines. It is widely recognized that reductions in population connectivity can increase isolation and inbreeding, leading to a potential “extinction vortex” (Gilpin and Soule 1986), yet there is limited understanding of how flow alteration may impair the processes crucial for maintenance of genetic variation and thus adaptive capacity. In addition, there is a current pressing need for more effective and flexible watershed management tools, particularly in relation to monitoring aquatic populations and implementation of environmental flows (Grantham et al. 2010). Thus, population genetics could be a powerful tool to understand the influence of different flow regimes on population health and this information could facilitate improved flow management to better protect aquatic populations.

The river-breeding foothill yellow-legged frog (*Rana boylii*; FYLF) historically occurred in lower and mid-elevation streams and rivers from Southern Oregon to northern Baja California west of the Sierra-Cascade crest (Stebbins 2003). *Rana boylii* are intimately linked with river hydrology because they have evolved to spawn in synchrony with natural

INTRODUCTION

flow cues associated with seasonal spring snowmelt or rain recession periods (Kupferberg 1996, Yarnell et al. 2010, 2016, Bondi et al. 2013). However, population declines have been documented across the former range of this species, particularly in southern California and the Sierra Nevada where it has been extirpated from approximately 50 percent of its historical range (Jennings and Hayes 1994, Davidson et al. 2002).

In California, particularly in the Sierra Nevada, river regulation may be a significant environmental stressor (Lind et al. 1996, Kupferberg et al. 2012). Regulated river reaches typically alter flows by augmenting or diverting winter and spring runoff, thereby reducing or eliminating flow cues and disrupting natural flow regimes. Aseasonal flow fluctuation from river regulation can scour (detach from substrate) or desiccate *R. boylii* egg masses, and the loss of clutches may have a significant demographic impact because only one egg mass is laid per year. In many regulated rivers in the Sierra Nevada, *R. boylii* populations are now restricted to small unregulated tributaries flowing into the regulated mainstem.

Here, we investigated the impacts of river regulation on genetic diversity of *R. boylii* populations across three different flow regimes. Given that population connectivity and genetic diversity are known to be play critical roles in long-term species persistence, we explored the association between these metrics and levels of river regulation. Our goal was to assess the genetic health (through metrics of genetic diversity) of *R. boylii* under different river regulation regimes to better inform the potential for long-term persistence. These data can be used to prioritize management and conservation efforts for *R. boylii*, as well as demonstrate the potential utility of genetics for future conservation monitoring efforts in aquatic species.

METHODS

Methods

Sample collection and DNA extraction

345 *R. boylii* buccal or tissue samples were used in this study across six different rivers (Table 1.1, Appendix, S1). Field sampling was conducted following Heyer et al. (1994), under CDFW SCP Permit #0006881, with IACUC protocol #19327. Individual post-metamorphic frogs were buccal-swabbed following established protocols (Goldberg et al. 2003, Pidancier et al. 2003, Broquet et al. 2007). Each post-metamorphic individual was comprehensively swabbed underneath tongue and cheek for approximately one minute. Swabs were air dried for approximately five minutes and placed in 1.5 mL microcentrifuge tubes while in the field. Samples were stored in the laboratory at -80°C until DNA extraction. Where possible, tail clips from tadpole larvae were collected, and tadpoles greater than 15 mm total length were targeted (Wilbur and Semlitsch 1990, Parris et al. 2010). One small (<3mm) tail clip was taken per individual tadpole and dried on Whatman qualitative filter paper (grade 1) and stored at room temperature. Some older tissue samples consisted of toe clips placed in 100% ethanol for storage, and DNA extraction from these samples used Qiagen DNeasy kits following the manufacturer's protocol. Buccal swabs and tail clip DNA were extracted with a magnetic bead-based protocol (Ali et al. 2016) and quantified using Quant-iT PicoGreen dsDNA Reagent (Thermo Fisher Scientific) with an FLx800 Fluorescence Reader (BioTek Instruments).

METHODS

Table 1.1: Sampling and locality information for population genomic analysis in the Yuba, Bear, and American watersheds in the northern Sierra Nevada of California, USA. The number of individuals (n) is given for the total number sequenced per location and the number of individuals that were retained after filtering across the 8,533 baits. NHD refers to the National Hydrography Dataset by USGS (U.S. Geological Survey, National Hydrography Dataset, Digital data, accessed, August 2017).

Site Name	SiteID	Locality	River	Watershed	Regulation Type	Lat.	Lon.	Elev. (m)	NHD Stream Order	NHD Total Drainage Area (sq. km)	n initial	n retained
BEAR	20	Chicago Powerhouse	Bear	Bear	Bypass	39.17484	-120.8908	665.7900	4	136.0	6	6
BEAR_GRIN	16	Greenhorn Creek	Bear	Bear	Bypass	39.23206	-120.9019	820.5347	2	11.0	15	6
BEAR_STH2	19	Steep Hollow Creek	US	Bear	Bypass	39.19444	-120.8878	704.6578	2	45.0	7	6
BEAR_STHA	18	Hawkins Ravine	Bear	Bear	Bypass	39.18833	-120.8981	706.9961	2	4.0	3	3
BEAR_STHC	17	Steep Hollow Creek	DS	Bear	Bypass	39.20231	-120.8754	736.8513	2	52.0	13	12
MFA_AMEC	1	American Canyon	MF American	American	Hydropoeaking	38.93396	-120.9436	240.3928	2	9.0	16	6
MFA_GASC	2	Gas Canyon	MF American	American	Hydropoeaking	38.96651	-120.9325	241.8506	1	13.0	6	6
MFA_TODC	3	Todd Creek	MF American	American	Hydropoeaking	38.96385	-120.9216	367.7263	2	10.0	11	9
MFY_OREGCK	27	Oregon Creek	Middle Yuba	Yuba	Bypass	39.44188	-121.0575	620.4250	4	375.0	15	13
MFY_US_OH	28	US Our House Dam	Middle Yuba	Yuba	Bypass	39.43305	-120.9903	624.1467	4	375.0	13	12
NFA	13	Iowa Hill Mainstem	NF American	American	Unregulated	39.11115	-120.9168	386.4710	4	605.0	36	30
NFA_BUNG	8	Branch Canyon	NF American	American	Unregulated	39.63762	-120.9103	280.2874	3	27.0	15	14
NFA_EUCH	14	Euchre Bar	NF American	American	Unregulated	39.18492	-120.7620	579.5186	4	508.0	13	11
NFA_INDC	10	Indian Creek	NF American	American	Unregulated	39.05665	-120.9085	296.1071	2	24.0	12	11
NFA_POND	7	Ponderosa Bridge	NF American	American	Unregulated	38.99995	-120.9406	240.8255	5	857.0	5	5
NFA_ROBR	12	Robbers Ravine	NF American	American	Unregulated	39.10451	-120.9267	400.0005	1	4.0	30	11
NFA_SAIC	15	Sailor Canyon	NF American	American	Unregulated	39.21694	-120.4960	1005.5781	3	166.0	8	5
NFA_SHIC	9	Shirttail Creek	NF American	American	Unregulated	39.04446	-120.8994	525.7589	4	141.0	16	15
NFA_SLAR	11	Slaughter Ravine	NF American	American	Unregulated	39.09865	-120.9255	356.0791	2	6.0	8	8
NFMFA_SC	4	NFMFA-Slunk Canyon	MF American	American	Hydropoeaking	39.02237	-120.7369	521.5720	2	6.0	18	18
NFY	29	Rocky Rest Mainstem	North Yuba	Yuba	Unregulated	39.51190	-120.9774	704.6863	5	669.0	15	12
NFY_SLAKE_CGRAV	30	Slate Creek	North Yuba	Yuba	Bypass	39.68913	-120.9389	1330.9373	3	58.7	4	4
RUB_LCUTS	6	Rubicon-Long Canyon	MF American	American	Bypass	38.98887	-120.6900	415.1026	5	806.0	9	8
RUB_USPH	5	Rubicon-US Powerhouse	MF American	American	Bypass	38.99928	-120.7233	360.5759	5	816.0	11	11
SFY	26	US Canyon Creek	South Yuba	Yuba	Bypass	39.35386	-120.7342	889.7745	4	365.0	6	6
SFY_LOCA	25	Logan Creek	South Yuba	Yuba	Bypass	39.36914	-120.8526	1201.1790	1	5.0	5	4
SFY_MISC	24	Missouri Canyon	South Yuba	Yuba	Bypass	39.36096	-120.8814	1094.6312	2	5.0	8	6
SFY_ROCKCK	23	Rock Creek	South Yuba	Yuba	Bypass	39.32983	-120.9863	593.5214	4	710.0	3	3
SFY_SHADYCK	21	Shady Creek	South Yuba	Yuba	Bypass	39.35433	-121.0590	675.0255	2	15.0	14	12
SFY_SPRINGCK	22	Spring Creek	South Yuba	Yuba	Bypass	39.33233	-120.9890	595.1054	3	24.0	4	4

METHODS

Sequencing and de novo assembly

To produce a high-quality genomic resource for a frog species with a large genome size, we first interrogated a large fraction of the genome using a SbfI restriction enzyme and high-density RAD sequencing on an Illumina HiSeq (Miller et al. 2007, Baird et al. 2008). Paired-end sequence data were generated from 24 *R. boylii* individuals collected previously (Peek 2010) from coastal and Sierra Nevada populations in California, USA (Appendix, S2). RAD libraries were constructed following the protocol described in Ali et al. (2016). *De novo* loci discovery and contig extension were carried via custom PERL scripts (Miller et al. 2012) using the alignment program Novoalign and the genome assembler PRICE (Ruby et al. 2013). This pipeline resulted in a set of 77,544 RAD contigs ranging from 300 to 800 bp which served as a *de novo* partial genome reference for all subsequent downstream analyses (Appendix, S3). Using these data, we filtered data to loci with 4 or fewer SNPs, and randomly selected 10,000 loci from this subset. Using these RADSeq data, 8,533 RAD capture baits (120bp) were designed by Arbor Biosciences from the *de novo* alignment (Appendix, S4).

Rapture sequencing

We then performed Rapture on all study samples to identify putative high-quality SNPs (Appendix, S1) using RAD capture baits. Three different sequencing runs on an Illumina HiSeq were merged together, filtered, and duplicates were removed using ANGSD and Samtools (Li et al. 2009). Sampled individuals were aligned against the *de novo* partial genome reference using the BWA-MEM algorithm (Li and Durbin 2010, Li 2013) and saved to BAM format. SAMtools was used to sort, filter for proper pairs, remove PCR duplicates, and index binary alignment map (BAM), as well as merge sequences from

METHODS

multiple libraries (Li et al. 2009).

Principal component analysis

A probabilistic framework was used to discover SNPs for PCA as it does not require calling genotypes and is suitable for low-coverage sequencing data (Korneliussen et al. 2013, Fumagalli et al. 2013). All Rapture analyses were conducted using Analysis of Next Generation Sequencing Data (ANGSD) (Korneliussen et al. 2014). ANGSD analyses were conducted following methods from Prince et al. (2017), with a minimum mapping quality score (`minMapQ`) of 10, a minimum base quality score (`minQ`) of 20, and the genotype likelihood model (`GL 1`) (Li 2011). To maximize data quality, samples with less than 100,000 aligned reads were excluded (Appendix, S1) and only sites represented in at least 50% of the included samples (`minInd`) were used. Settings used in ANGSD for PCA to identify polymorphic sites included a `SNP_pval` of $1e^{-6}$, inferring major and minor alleles (`doMajorMinor 1`), estimating allele frequencies (`doMaf 2`) (Kim et al. 2011), retaining SNPs with a minor allele frequency of at least 0.05 (`minMaf`), genotype posterior probabilities calculated with a uniform prior (`doPost 2`), and the `doIBS 1` and `doCov 1` options were used to generate PCA data. Principal components (PC) summarizing population structure were derived from classic eigenvalue decomposition and were visualized using the `ggplot2` package in R (R Core Team 2017).

Genetic differentiation and diversity estimates

Mean scaled F_{ST} was used to quantify genetic differentiation between populations (Wright 1943, Rousset 1997). Genome-wide F_{ST} between population pairs was estimated by first calculating a site frequency spectrum (SFS) for each population (`doSaf`) (Nielsen et al.

METHODS

2012) with ANGSD. The two-dimensional SFS and global F_{ST} between each population pair were then estimated using realSFS (Korneliussen et al. 2014). F_{ST} was calculated between each pair of collection locations within a watershed, and the mean of all pairwise calculations within that watershed was calculated for each location. We calculated the river distances (distance along river network) between locations within watersheds using the riverdist package in R (Tyers 2017), and used the mean pairwise river distance (km) to all other locations within the watershed. These values were plotted and a generalized linear model was fitted ($F_{ST} \sim MeanRiverDistance$) in R (R Core Team 2017). To calculate Watterson's θ (θ_S) (Watterson 1975), and Tajima's θ (θ_π) (Tajima 1983), we used SFS that were estimated as described above as priors (`pest`) to calculate each statistic for each site (`doThetas`), and then averaged to obtain a single value for each statistic (Korneliussen et al. 2013).

Boosted regression tree modeling of variance in F_{ST}

We used boosted regression tree (BRT) models with the R packages gbm (Ridgeway 2015) and dismo (Hijmans et al. 2017) to assess the relative influence of river regulation as compared to other covariates. Boosted regression trees (BRT) are suitable frameworks for large and complex ecological datasets because they do not assume normality, nor linear relationships between predictor and response variables and they ignore non-informative predictor variables (Graham et al. 2008, Steel et al. 2017). BRTs use iterative boosting algorithms to combine simple decision trees to improve model performance (De'ath 2007) and provide a robust alternative to many traditional statistical methods (Phillips et al. 2006, Guisan et al. 2007). BRTs assess the relative impact of modeled variables by calculating the number of times a variable is selected for splitting a tree across all folds of the cross validation. Following Steel et al. (2017), estimates of relative influence

METHODS

for each predictor variable were used to evaluate the relative contribution a variable had in predicting the response. To evaluate the relative influence of covariates on F_{ST} , models were trained using river distance (km), elevation (m), upstream drainage area (km²), Strahler stream order, and number of samples per location. Stream segment data on elevation, length, slope, stream order, and drainage area were derived from NHD Plus attributes (U.S. Geological Survey, National Hydrography Dataset, Digital data, accessed, August 2017 at <http://nhd.usgs.gov/data.html>). In addition, $\Delta\theta$ ($\theta_\pi - \theta_S$) was included to assess the effect of genomic variation on F_{ST} across regulation types.

Model training and fitting were conducted following methods previously described in Steel et al. (2017). To reduce overfitting, the learning rate (also known as the shrinking rate) was set to 0.001. Stochastic gradient boosting was utilized to reduce prediction error (De'ath 2007) and the fraction of training data sampled to build each tree was 0.75, within the range as recommended by (Brown et al. 2012). Tree complexity was set to three to allow for second and third order interaction effects. The minimum number of observations required in the final nodes of each tree was three. A ten-fold cross-validation technique allowed us to determine the number of trees at which prediction error was minimized using the cross-validation deviance. Model performance was evaluated using the minimum estimated cross-validation deviance which maximized the estimated deviance explained.

RESULTS

Results

Rapture produces high quality genomic data for *Rana boylii*

To begin investigating the impact of river regulation on *R. boylii*, we collected frog tissue and buccal samples from 30 locations in six rivers representing three different flow impairment levels associated with hydropower generation. The three flow regimes assessed were: (1) hydropeaking, where flows are pulsed on most days from late spring through fall to provide electricity during peak-use hours and for recreational whitewater rafting; (2) bypass, which diverts river flows from an upstream portion of the basin to the downstream power generation facilities; and (3) unregulated, a largely natural flow regime where no upstream controls exist to regulate flows (Figure 1.1). Flow data were obtained for each river reach using proximal USGS gaging stations (Table 1.1). We sampled a total of 345 *R. boylii* from sites in three major watersheds (Yuba, Bear, and American) in the northern Sierra Nevada of California (Figure 1.1, Table 1.1). The six study rivers share a similar Mediterranean climate, underlying geology, watershed aspect (west-slope), stream morphology (riffle-pool), and vegetative communities, but differ in the intensity of flow regulation (Steel et al. 2017). Although river regulation occurs in all three of the study watersheds, both the North Yuba and North Fork (NF) American are unregulated whereas the Middle Fork (MF) American is the only river that has a hydropeaking flow regime (Figure 1.1A).

To generate genetic data from the samples, we performed RAD Capture (a.k.a. Rapture) (Ali et al. 2016) on the samples by generating SbfI RAD libraries, capturing a subset of the RAD loci using 8,533 baits (see Methods), and sequencing the resulting library on an Illumina HiSeq. We then aligned the sequencing reads from each sample to a de novo RAD assembly (see Methods). The mean number of filtered alignments across all

RESULTS

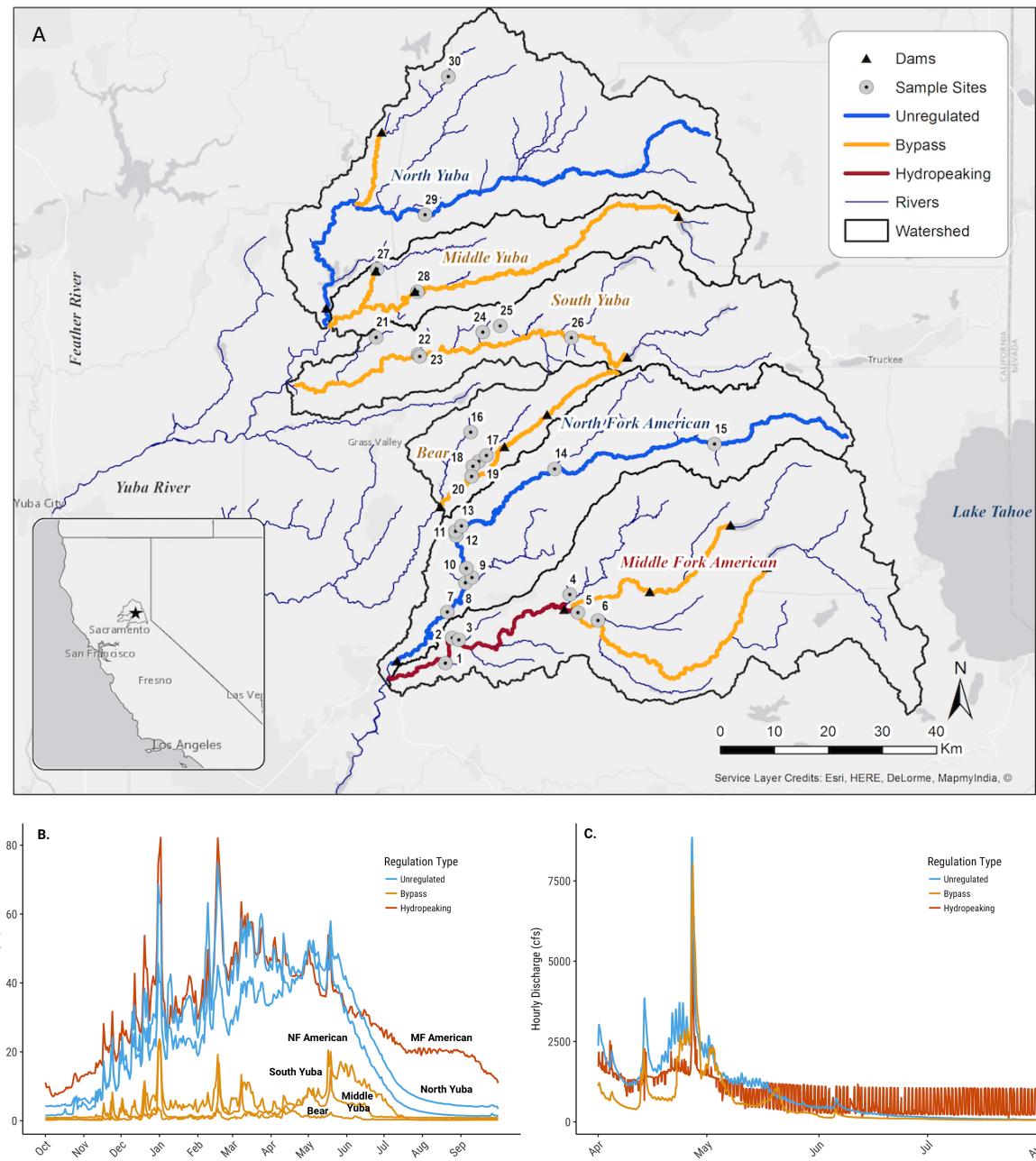


Figure 1.1: Sampling locations and flow characteristics. A) Map of sampling locations spread across six rivers. B) Comparison of annual mean daily discharge from 1981–2016 for three flow types. C) Comparison of hourly discharge in three different flow regimes in April through July 2012, Bypass (South Yuba), Hydropowering (Middle Fork American), and Unregulated (North Fork American). See Table 1.2 for USGS gaging station information.

RESULTS

Table 1.2: Metadata for USGS gaging stations with current and historic data available for each study river

Study Site	USGS Gage Number	Years of Record	Latitude	Longitude
North Yuba				
North Yuba	11413000	1931–Present	39.52500	-120.9369
Middle Yuba				
Middle Yuba	11408550	1987–Present	39.52194	-120.5825
Middle Yuba	11408700	1957–1966	39.43861	-120.8111
South Yuba				
South Yuba	11414250	1965–Present	39.31861	-120.6567
South Yuba	11417000	1942–1972	39.36056	-120.7706
NF American				
NF American	11427000	1942–Present	38.93611	-121.0228
Rubicon				
Rubicon	11433200	1959–1984	38.99250	-120.7206
Rubicon	11427765	1974–Present	39.00027	-120.7231
MF American				
MF American	11433300	1958–2011	39.00611	-120.7597
MF American	11433500	1911–1986	38.91805	-121.0142
MF American	OXB (PCWA) ¹	1997–Present	39.00600	-120.7600

¹ PCWA=Placer County Water Agency, <http://cdec.water.ca.gov/cdecstation2/?sta=OXB>

345 samples was 324,928. For downstream analysis, we selected individuals that had greater than 100,000 alignments ($n=277$), which provided sufficient data to investigate population genetic attributes at broad and fine geographic scales (see below). *R. boylii* are cryptic, and often occur in low densities within the study area. Thus, we retained a minimum of three individuals per site, and the mean number of samples per site was approximately nine (Table 1.1). With genomic data, population genetic parameters can be accurately estimated from even low sample numbers (Hotaling et al. 2018), and genomic analyses in non-model organism often use fewer loci (Narum et al. 2013). We conclude that the sequence data we obtained should be appropriate for population genetic analyses across our study area.

RESULTS

Anomalous genetic pattern in highly regulated reach of Middle Fork American watershed

To assess *R. boylii* population structure across the collection locations, we used ANGSD (Korneliussen et al. 2014) to discover 44,406 SNPs and perform principal component analysis (PCA; see Methods), which provides a dimensionless comparison of all samples. The first two principal components revealed four main groups corresponding to the Yuba, Bear, North Fork (NF) American, and Middle Fork (MF) American samples (Figure 1.2A). Unlike the Yuba watershed where all rivers clustered as one group, the two rivers within the American watershed (the NF American and MF American) were separated by both PC1 and PC2. Although the NF American watershed clustered closely with the adjacent Bear watershed, the MF American showed a surprisingly high degree of genetic differentiation from other locations (Figure 1.2A). These data suggest that there is less genetic differentiation between the NF American and the Bear watersheds, than between the NF and MF American watersheds. We conclude that measurements of overall genetic differentiation in *R. boylii* from our study area largely conform to watershed and geographic expectations, with the exception of the American watershed, which shows a surprisingly high degree of genetic differentiation between the North (unregulated) and Middle (hydropeaking) Forks.

To further investigate patterns of genetic variation within the American watershed, we performed two PCAs, one on samples from the NF American, and the other on samples from the MF American. The PCA of the NF American showed minimal differentiation among locations, with different study sites blending together and weak patterns of population structure (Figure 1.2B). In contrast, PCA of the MF American showed strong differentiation between sites (Figures 1.2C, 1.2D). The MF American PCA completely resolved all sites, with the first component (PC1) strongly differentiating the samples

RESULTS

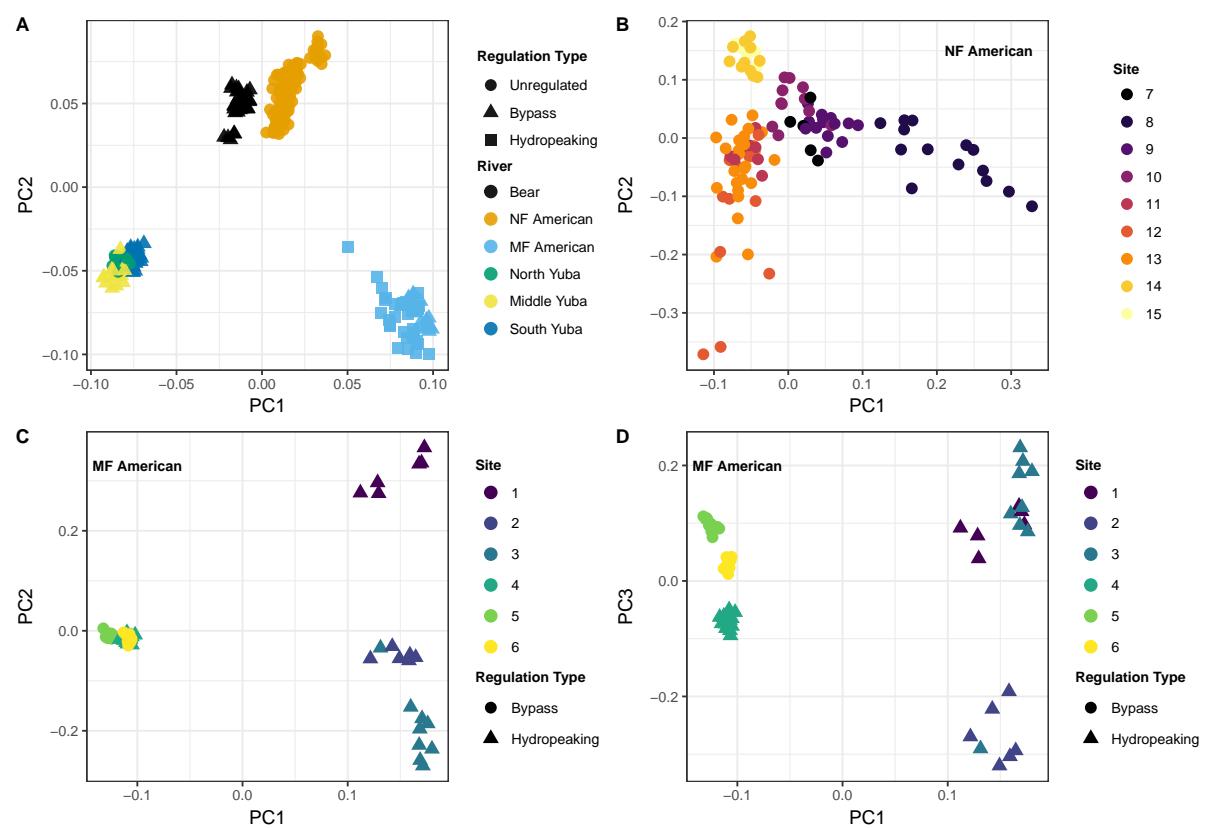


Figure 1.2: Principal component analysis of Rapture sequencing data. A) Northern Sierra Nevada ($n=277$) watersheds and regulation types; B) Unregulated NF American; C) and D) Hydropeaking MF American Reach.

RESULTS

in the hydropeaking reach from all other sites in the MF American. This pattern may be due to the differential river regulation between the two rivers; the NF American is unregulated and has weak PCA differentiation, whereas the MF American has a higher level of river regulation and all sites form distinct genetic clusters, indicative of reduced gene flow among sites within the MF American.

River regulation is the strongest predictor of genetic isolation with *R. boylii* in the Northern Sierra

To assess how patterns of genetic differentiation are associated with river regulation across our entire study area, we estimated pairwise F_{ST} (Wright 1943) between all collection locations within a river for all six rivers. We then plotted the scaled mean pairwise F_{ST} [mean F_{ST} / (1-mean F_{ST})] (Rousset 1997) for each location against the mean river distance (the average distance along the river network from each collection location to every other location within that study river). Furthermore, each location was categorized by regulation level of closest mainstem location. While there was a clear relationship between F_{ST} and river distance (as shown by the slope of regression lines in Figure 1.3A), there was a striking pattern of elevated F_{ST} by regulation type (Figure 1.3A). Even the bypass regulation type showed a distinct pattern of elevated F_{ST} compared to the unregulated type. For instance, regulated rivers with locations separated by less than 10km had F_{ST} values comparable to unregulated locations separated by mean river distances over 30 km. Hydropeaking was the most extreme pattern of the three regulation types and showed highly elevated F_{ST} values with the steepest regression coefficient. The baseline F_{ST} or global mean increased by over 0.1 between the unregulated (mean $F_{ST}=0.141$), and regulated locations (global mean for bypass $F_{ST}=0.256$, hydropeaking $F_{ST}=0.278$). This suggests a greater degree of isolation within sites in regulated river

RESULTS

reaches compared with *R. boylii* populations in unregulated reaches, as larger F_{ST} values represent reductions in heterozygosity due to population subdivision (Slatkin 1987). We conclude *R. boylii* in regulated rivers show patterns of greater population isolation and loss of heterozygosity compared to frogs in unregulated locations.

To investigate the relative influence of river regulation compared to other covariates such as river distance on genetic differentiation (i.e. F_{ST}), we used boosted regression tree (BRT) modeling. Covariates included flow regime alteration type, river distance, watershed variables derived from National hydrology data (NHD), topographic data, and allele frequency spectrum skew. We found flow regulation explained the greatest amount of variance in F_{ST} (Figure 1.3B). Thus, river regulation has a larger relative influence than mean river distance between sampling locations, which is often the most important factor influencing genetic differentiation (Wright 1943, Slatkin 1987, Rousset 1997). We conclude there is a pattern of isolation and limited connectivity between populations in regulated reaches.

RESULTS

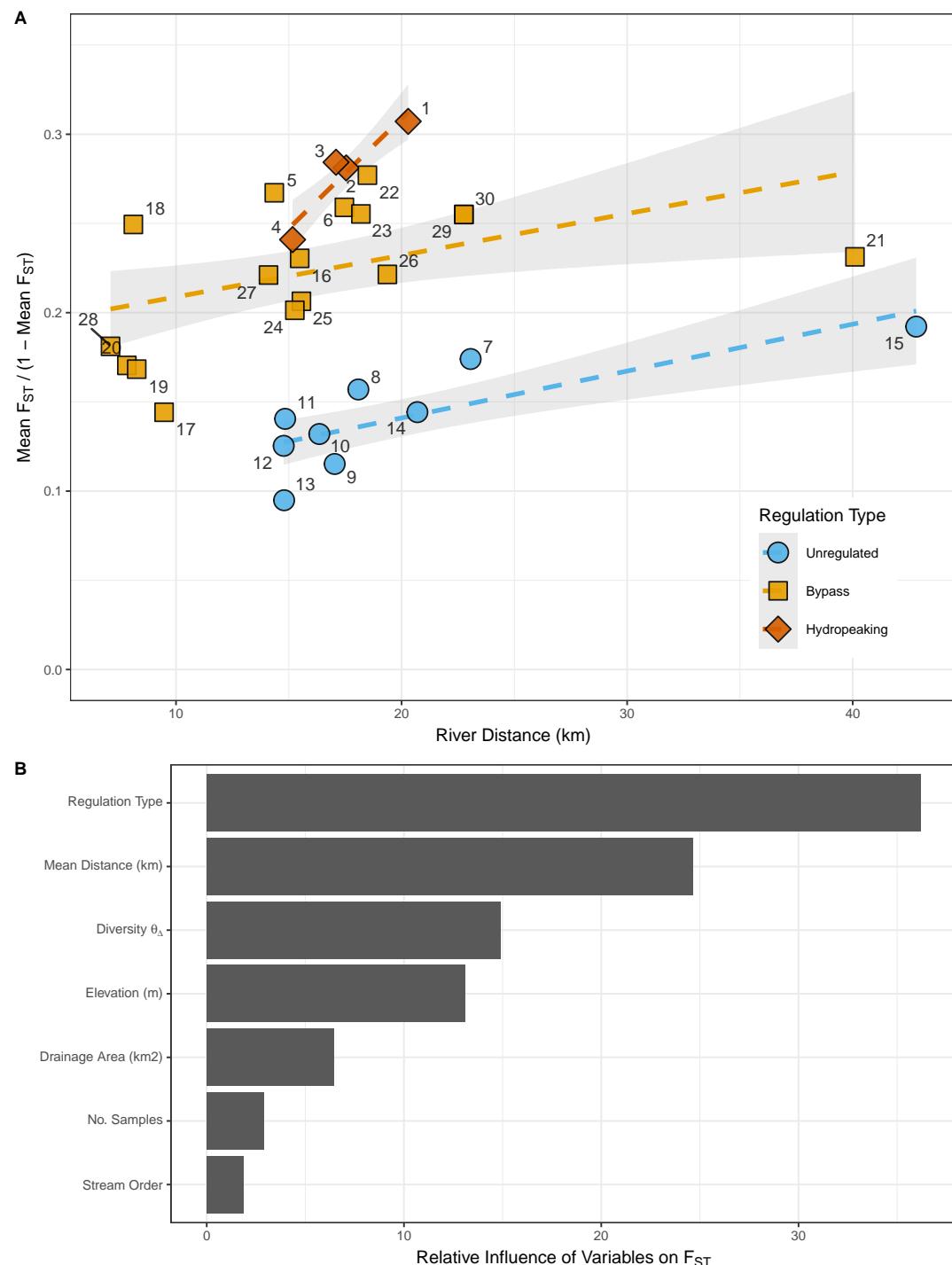


Figure 1.3: Relationship between river regulation and genetic differentiation in *R. boylii*. A) Mean pairwise F_{ST} vs. mean river distance for each location; B) Relative influence of variables on F_{ST} from boosted regression tree models.

RESULTS

River regulation strongly correlated with decreasing genetic diversity in *R. boylii*

To investigate the association between river regulation and genetic diversity trajectory (stable, increasing, or decreasing), we summarized patterns of genetic variation using two estimators of θ ($4N\mu$): Tajima's θ (θ_π) is based on the average number of pairwise differences (Tajima 1983) and Watterson's θ (θ_S) is based on the number of segregating sites (Watterson 1975). These estimators are influenced by the demographic history of a population and provide information on the trajectory of changes in genetic diversity. When genetic diversity has been stable, these estimates are generally equal; but when genetic diversity has been increasing, $\theta_\pi > \theta_S$; and when genetic diversity has been decreasing, $\theta_S > \theta_\pi$. We found zero populations sampled within regulated watersheds had evidence of increasing genetic diversity (e.g., a $\theta_\pi - \theta_S$ that was less than zero) (Figure 1.4A). The regulated locations showed a clear trajectory of genetic diversity loss (Figure 1.4A, 1.4B). Three of the four hydropeaking locations had the highest values of $\Delta\theta$ ($\theta_\pi - \theta_S$), and the global mean was significantly different from other regulation types. Although some tributary populations within unregulated watersheds also showed signs of genetic diversity loss, the mean genetic diversity trajectory at unregulated locations was largely neutral (Figure 1.4B). This indicates populations in the northern Sierra Nevada which are already limited in number are losing genetic variation, and river regulation appears to be exacerbating these patterns. We conclude there is evidence of recent genetic diversity loss across populations in the regulated river systems, regardless of regulation type.

RESULTS

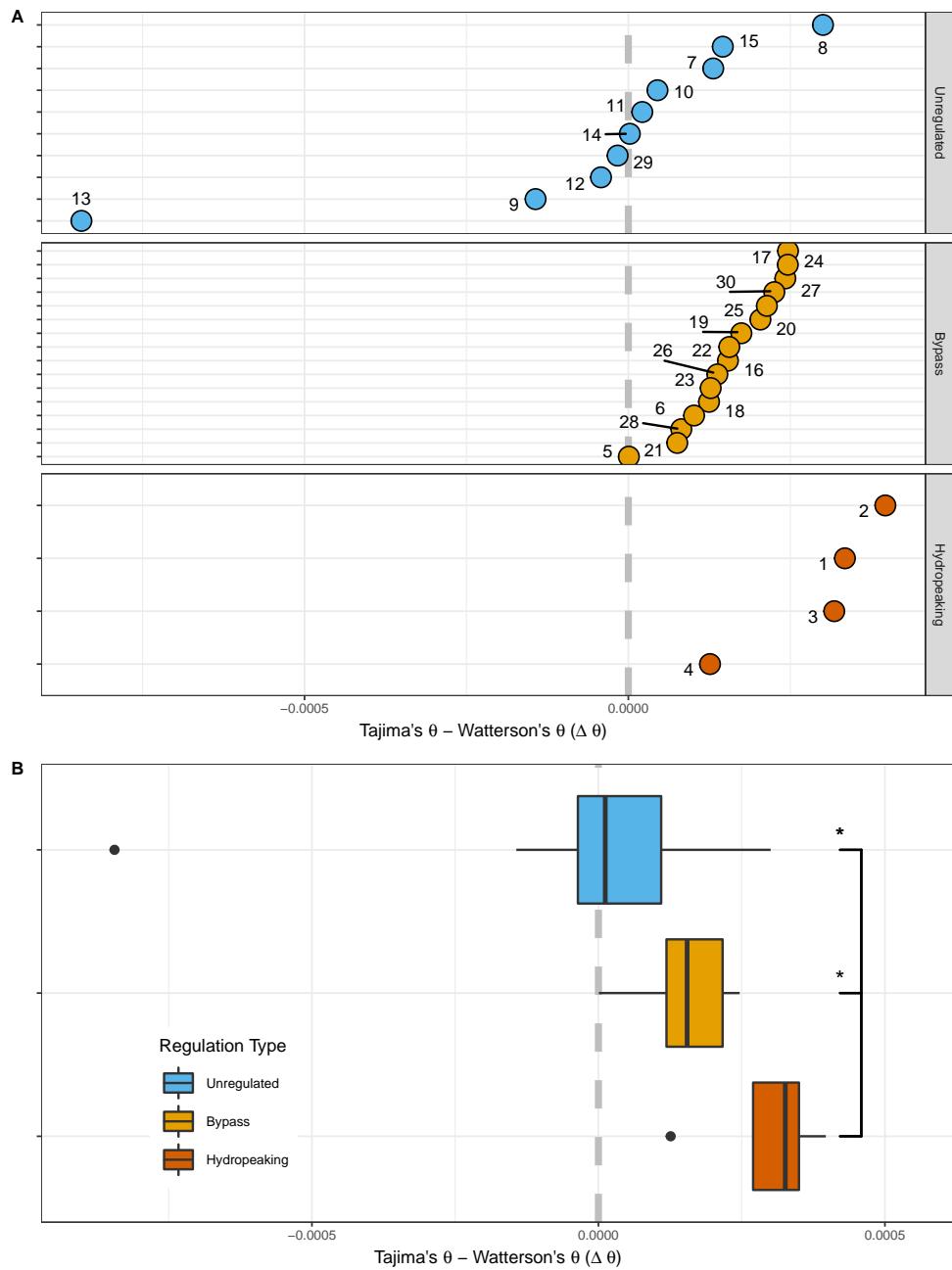


Figure 1.4: Relationship between river regulation and genetic diversity trajectory in *R. boylii*. A) Assessment of genetic diversity trajectories using $\Delta\theta$ ($\theta_\pi - \theta_S$) for each sampling location; B) Boxplots of difference between $\theta_\pi - \theta_S$ and pairwise significance between regulation groups using a pairwise Wilcoxon rank sum test with bonferroni correction ($P < 0.05$). Negative values represent trends of increasing genetic diversity, positive values represent trajectories of diversity loss, values near zero are stable.

DISCUSSION

Discussion

Although massive parallel sequencing (MPS) technologies have the potential to facilitate collection of high-quality genetic data in virtually any species, a number of challenges still remain for many species including low quality or non-existent reference genomes, large/complex/repetitive genomes, and high cost of processing/sequencing in studies with many samples. Amphibians are particularly challenging as many species have very large genome sizes (Nunziata et al. 2017), for example, there are only two frog reference genome assemblies available as of 2018 (Hellsten et al. 2010, Sun et al. 2015). Our results demonstrate that Rapture (Ali et al. 2016) is a suitable method to rapidly and inexpensively discover a large number of loci in a frog species with a complex genome. In this study, we used new RAD sequencing and RAD capture (Rapture) methods (Ali et al. 2016) to generate high-quality genomic data suitable for discovering and genotyping many single nucleotide polymorphisms (SNPs) in *R. boylii*. Based on this dataset, we were able to successfully characterize patterns of genetic variation within *R. boylii* as well as design a set of RAD capture baits that can be used as a genetic monitoring resource for *R. boylii* (and likely other ranid species). This highlights that the collection of genetic information, even from large numbers of samples or in complex genomes, is no longer a limitation with current genomic methods such as RAD and Rapture.

Demographic connectivity is widely recognized as a fundamental driver of long-term population persistence (Fahrig and Merriam 1985, Taylor et al. 1993). Populations must adapt over time and connectivity is a major way to transfer genetic information. For example, previous studies have shown that adaptation can occur by spreading specific alleles across large geographic distances (Miller et al. 2012, Prince et al. 2017). In many regulated river reaches in the Sierra Nevada, *R. boylii* now occur in isolated locations, breeding in tributaries rather than mainstem habitats. However, since these frogs have

DISCUSSION

the potential to move long distances (Bourque (2008) documented *R. boylii* individuals that moved over 1 km in a day), the extent to which current population connectivity has been lost due to river regulation remains unknown. Examining pairwise F_{ST} , revealed a major decrease in connectivity in populations in regulated systems, even with limited river regulation (i.e., bypass reaches). Usually isolation by distance patterns best describe variation in genetic data, yet the primary factor influencing genetic differentiation among these rivers is hydrologic alteration (Figure 1.3B). Thus, despite being able to move long distances, *R. boylii* have not been able to maintain population connectivity in regulated rivers. This demonstrates that even in species that can move relatively long distances and pass potential physical barriers (e.g., infrastructure such as dams, canals, and reservoirs likely do not represent true barriers to movement of *R. boylii*), loss of connectivity may still occur and can be revealed with genetic analysis.

Genetic diversity is also a critical component for long-term population persistence because it is closely related to the evolutionary capacity for adaptation to environmental changes (Lande and Shannon 1996, Frankham 2002, Hoffmann and Sgrò 2011, Ishiyama et al. 2015). In some cases, isolated populations can maintain genetic diversity when they are sufficiently sized (Whiteley et al. 2010), however, species of conservation concern typically have small and/or declining populations and thus may be susceptible to genetic diversity loss (Krohn et al. 2018). Throughout the Sierra Nevada, *R. boylii* have largely disappeared from regulated mainstem rivers, but the extent to which existing populations have been able to maintain genetic diversity is unclear. Strikingly, our analysis revealed that every single population within the regulated watersheds exhibits a trajectory of genetic diversity loss. Thus, genomic analysis of molecular variation provides a powerful lens to discover and assess trajectories of genetic diversity.

Our analyses, using metrics that can serve as a reasonable proxy for genetic health of a

CONCLUSION

population, does not bode well for the long-term persistence of *R. boylii* populations in regulated rivers in the Sierra Nevada. Many of these *R. boylii* populations are already losing genetic diversity and given their small size and reduced connectivity the effects of inbreeding will likely exacerbate their problems. *Rana boylii* have evolved in river systems with consistent hydrologic seasonality and predictability, despite inter-annual variation. Flow regulation has altered patterns of hydrologic seasonality and predictability in many watersheds (Kupferberg et al. 2012). Long-term population persistence may still be possible if conservation efforts utilize methods that promote or maintain genetic diversity and increase population connectivity. For example, simulations by Botero et al. (2015) demonstrated adaptation persisted in modeled populations through large environmental changes—if phenotypic strategies were appropriate before and after the change—but modeled populations declined rapidly when the current strategy was a mismatch to the current environment. Thus, *R. boylii* conservation efforts should focus on river reaches where flow management may provide opportunities to more closely mimic local natural flow regimes and improve hydrologic connectivity to prevent further genetic diversity loss.

Conclusion

Detecting evolutionary responses to within- and among-year changes in an ecological or hydrological context has previously been difficult. However, utilizing genetic data can fill these gaps and provide a highly informative process for identifying the impacts of anthropogenic and environmental change on the process of adaptation (Kahilainen et al. 2014, Botero et al. 2015). We demonstrate that an aquatic species that has adapted to local hydrology patterns shows poor genetic health (i.e., clear patterns of

CONCLUSION

decreased connectivity and trajectories of genetic diversity loss). Our results highlight the potential impact of river regulation on aquatic organisms and their potential for long term persistence. In the future, similar genetic approaches could be used in many other contexts to explore the impacts of river regulation on aquatic organisms. Taken together, our results demonstrate that genetic monitoring can be a powerful tool for assessment of population health and should be a critical component of conservation management in aquatic organisms.

Chapter 2

Hybridization between two sympatric ranid frog species in the northern Sierra Nevada, California

Introduction

Landscape changes can influence species demography and migration patterns (Li et al. 2017) which can change rates of gene flow within species. Changing migration rates and population sizes can influence population structure; thus, over time, landscape changes can cause significant changes in genetic diversity within a species. Furthermore, cross-breeding or hybridization between closely related taxa can promote gene flow (introgression) between species, which may be an important evolutionary mechanism for either homogenization (reversing initial divergence between species), speciation (from reproductive isolation of hybrid populations), or adaptation (transfer of adaptive alleles) (Mallet 2007, Abbott et al. 2013, Barrera-Guzmán et al. 2018).

INTRODUCTION

Hybridization events in vertebrates may be rare, or rarely detected, and thus identifying potential hybridization can be difficult and may be affected by sampling design, timing, and resolution of genetic markers. Therefore, occurrences of hybridization likely remain unknown, particularly in cryptic taxa. Assessing population admixture or detecting potential hybridization has previously been challenging; however, modern genetic methods provide a powerful approach to assess populations at fine geographic and evolutionary scales (Ali et al. 2016, Prince et al. 2017).

We investigate the potential for hybridization in two sympatrically occurring endemic frog species in the Sierra Nevada of California. Foothill yellow-legged frogs, *Rana boylii*, (Baird 1856) historically occurred in lower and mid-elevation (<1500 m) streams and rivers from Southern Oregon to northern Baja California west of the Sierra-Cascade crest (Stebbins 2003), whereas Sierra Nevada yellow-legged frogs, *Rana sierrae*, (Camp 1917) typically occurred from 1500 m to over 3600 m in lakes and streams (Zweifel 1955, Stebbins 2003). Population declines have been documented across the former range of both of these species; *R. sierrae* has been extirpated from over 90 percent of its historical range (Drost and Fellers 1996, Vredenburg 2004) while *R. boylii* has been extirpated from 50 percent of its historical range (Jennings and Hayes 1994, Davidson et al. 2002). Both species are of conservation concern; in 2014, the U. S. Fish and Wildlife Service (USFWS) listed *R. sierrae* as endangered under the U. S. Endangered Species Act (ESA) (USFWS 2014), and *R. boylii* is listed as a species of special concern in California and is a candidate for listing under the California and federal ESAs.

Unlike other ranid frog species with broad areas of potential intergradation (Shaffer et al. 2004), *R. boylii* and *R. sierrae* only rarely occur sympatrically. Zweifel (1955) described one historical location where these two species co-occurred, in Butte County near DeSabla. Currently the only known location where both species are found is sev-

INTRODUCTION

eral tributaries to the Feather River in the northern Sierra Nevada, California (Figure 2.1). Hybridization between these species has not previously been documented. Furthermore, breeding experiments by Zweifel (1955) between *R. sierrae* (formerly known as *R. muscosa*) and *R. boylii* yielded very low viability in fertilization and high incidences of embryological abnormalities—indicating a post-zygotic barrier between the species. However, these experiments only crossed female *R. sierrae* with male *R. boylii*, and the individuals were from very different California regions (e.g., Butte and Nevada County vs. Contra Costa County). *Rana boylii* and *R. sierrae* have very similar morphology and habitat preferences in areas where they co-occur; thus assigning individuals to species is difficult and imprecise using field identification methods. This presents a challenge for management because these sympatric species have different conservation status and management objectives. We employed modern genetic methodology to better understand *R. sierrae* and *R. boylii* where their ranges overlap. We investigated three primary questions:

1. Can hybridization be detected between two sympatrically occurring threatened and endangered (ESA) frog species in the Sierra Nevada using data generated from genome-wide single nucleotide polymorphisms (SNPs);
2. If hybrids can be detected, do genetic signatures suggest hybrid viability (i.e., can hybrids reproduce, leading to introgression between species);
3. Using coalescent modeling, are migration rates between species in sympatrically occurring populations higher than in allopatrically occurring populations in adjacent watersheds?

METHODS

Methods

Sampling and DNA Extraction

To investigate potential hybridization between *R. sierrae* and *R. boylii*, a total of 458 tadpole tail clips, buccal swabs, or tissue samples were compiled. Samples were identified to species in the field as either *R. boylii*, *R. sierrae*, or “unknown”, which were individuals which could not be visually confirmed as either species (Stebbins 2003). The samples were collected between 1992 and 2016, from three watersheds in the Sierra Nevada (the Feather, Yuba, and American) (Table 2.1, Appendix, S5). All unknown individuals were from Feather watershed localities.

The Yuba and American watersheds share a similar Mediterranean climate, underlying geology, watershed aspect (west-slope), and vegetative communities. The Feather watershed shares a similar climate but has a slightly different underlying geology and aspect than that of other watersheds in the Sierra Nevada. The Feather watershed lies in the transition zone of the northern Sierra Nevada and the Cascades/Basin and Range Province, and thus the landscape in the northern portion of the watershed is comprised largely of volcanic bedrock while the southern portion is largely granitic (Durrell 1988).

METHODS

Table 2.1: Sampling localities.

Locality	River	No. Samples	Lat.	Lon.	Elev (m)	Basin (HUC8)
MFA-AMEC	MFA	5	38.934	-120.9436	240	American
MFA-GASC	MFA	1	38.9665	-120.9325	242	American
MFA-TODC	MFA	6	38.9638	-120.9216	368	American
MFA-US-R	MFA	1	39.0075	-120.7316	360	American
NFA	NFA	12	39.1079	-120.9227	363	American
NFA-BUNC	NFA	8	39.0376	-120.9103	286	American
NFA-EUCHDS	NFA	3	39.1849	-120.762	580	American
NFA-INDC	NFA	3	39.0567	-120.9085	296	American
NFA-LyonsPk	NFA	11	39.2067	-120.3113	2529	American
NFA-POND	NFA	2	38.9999	-120.9406	241	American
NFA-ROBR	NFA	6	39.1045	-120.9267	400	American
NFA-SHIC	NFA	8	39.0417	-120.9009	341	American
NFA-SLAR	NFA	5	39.0987	-120.9255	356	American
NFMFA-SC	NFMFA	7	39.0224	-120.7369	522	American
RUB-HighlandDS	RUB	18	38.9615	-120.2422	2312	American
RUB-HighlandLk	RUB	29	38.9573	-120.2418	2383	American
RUB-LC-US	RUB	1	38.9889	-120.69	415	American
RUB-USPH	RUB	5	38.9993	-120.7233	361	American
RUB-Zitella	RUB	9	38.9595	-120.227	2335	American
RUB-ZitellaLk	RUB	2	38.9604	-120.2316	2337	American
SFA-CAMI	SFA	2	38.8115	-120.5787	725	American
FEA-BeanCk	FEA	60	39.9774	-121.091	1397	Feather
FEA-EBNFF	FEA	6	Unknown	Unknown	Unknown	Feather
FEA-GoldLk	FEA	4	39.9416	-121.136	1816	Feather
FEA-LoneRockCk	FEA	3	40.2012	-120.6453	1563	Feather
FEA-MillCk	FEA	4	39.9591	-121.1573	1891	Feather
FEA-RockLkBucksCk	FEA	1	39.9403	-121.1499	2102	Feather
FEA-RockLkSilver	FEA	8	39.9409	-121.1422	1902	Feather

METHODS

Table 2.1: Sampling localities. (*continued*)

Locality	River	No. Samples	Lat.	Lon.	Elev (m)	Basin (HUC8)
FEA-SFRockCk	FEA	27	39.8789	-121.0022	1470	Feather
FEA-SPANISH-BGulch	FEA	5	39.9546	-121.089	1283	Feather
FEA-SPANISH-RockCk	FEA	1	39.9445	-121.0221	1090	Feather
FEA-SPANISH-SilverCk	FEA	3	39.9377	-121.0849	1189	Feather
FEA-SPANISH-Wapaunsie	FEA	1	39.9523	-121.0373	1096	Feather
FEA-SpanishCk	FEA	26	39.9541	-121.0541	1192	Feather
NFF-Poe	NFF	1	39.736	-121.4702	284	Feather
FORD-Mossy-P1	FORD	1	39.381	-120.4623	2157	Yuba
FORD-Mossy-P2	FORD	3	39.3852	-120.4714	1998	Yuba
FORD-Mossy-P3	FORD	2	39.3765	-120.4603	2158	Yuba
FORD-MossyDS	FORD	19	39.3853	-120.4728	1984	Yuba
FORD-MossyPond	FORD	34	39.3781	-120.4701	2106	Yuba
FORD-NorthCkTrib	FORD	34	39.3869	-120.451	2090	Yuba
MFY-OREGCK	MFY	10	39.4419	-121.0575	620	Yuba
MFY-Remmington	MFY	1	39.4137	-120.9912	620	Yuba
MFY-US-OH	MFY	7	39.413	-120.9903	624	Yuba
NFY	NFY	12	39.5119	-120.9774	705	Yuba
NFY-SLATE-CGRav	NFY	3	39.6928	-120.9258	1457	Yuba
NFY-SLATE-Onion	NFY	3	39.6355	-121.0395	1300	Yuba
SFY	SFY	1	39.3539	-120.7342	890	Yuba
SFY-FallCk	SFY	5	39.3553	-120.7371	884	Yuba
SFY-HUMBUG	SFY	1	39.3637	-120.921	877	Yuba
SFY-LOGA	SFY	3	39.3691	-120.8526	1201	Yuba
SFY-MCKI	SFY	3	39.368	-120.8354	1084	Yuba
SFY-MISC	SFY	7	39.361	-120.8814	1095	Yuba
SFY-RockCk	SFY	3	39.3298	-120.9863	594	Yuba
SFY-Scotchman	SFY	3	39.3293	-120.777	1167	Yuba
SFY-ShadyCk	SFY	9	39.3543	-121.059	675	Yuba

METHODS

Field sampling was conducted following methods in Heyer et al. (1994) under CDFW SCP permit #0006881 and federal permit TE-40087B-0 with IACUC protocol #19327 and #04718-001. Individual post-metamorphic frogs were buccal-swabbed following established protocols (Goldberg et al. 2003, Pidancier et al. 2003, Broquet et al. 2007). Each post-metamorphic individual was comprehensively swabbed underneath tongue and inside of both cheeks for approximately 30 sec to one minute. Swabs were air dried for approximately five minutes and placed in 1.5 mL microcentrifuge tubes while in the field or placed in lysis buffer. Dried samples were stored in the laboratory at -80°C until DNA extraction. Where possible, tail clips from tadpole larvae were collected, and tadpoles greater than 15 mm total length were targeted (Wilbur and Semlitsch 1990, Parris et al. 2010). One clip was taken per individual tadpole and dried on Whatman filter paper (grade 1) and stored at room temperature or in 95% ethanol. DNA was extracted from ethanol-stored samples using Qiagen DNeasy kits following manufacturer protocol and stored at -20°C. DNA was extracted from dried buccal swabs and tail clips using an Ampure magnetic bead-based protocol (Ali et al. 2016) and stored at -20°C.

Rapture Sequencing

To produce a high-quality genomic resource for frog species with large genome sizes, we interrogated a significant fraction of the *R. boylii* genome using RAD sequencing with SbfI (Miller et al. 2007, Baird et al. 2008, Ali et al. 2016). Paired-end sequence data were generated using 24 *R. boylii* individuals collected previously (Peek 2010) from coastal and Sierra Nevada populations in California, USA (Appendix, S2). RAD libraries were constructed following the protocol described in Ali et al. (2016). De novo locus discovery and contig extension were carried out as previously described (Miller et al. 2012) using the alignment program Novoalign and the genome assembler PRICE (Ruby et al. 2013).

METHODS

This resulted in a set of 77,544 RAD contigs ranging from 300 to 800 bp which served as a de novo partial genome reference for all subsequent downstream analyses ([Appendix, S3](#)). We next removed loci with five or more SNPs, and randomly selected 10,000 loci from the remaining subset. Of these 10,000 loci, 8,533 were successfully designed into 120 bp RAD capture baits by Arbor Biosciences ([Appendix, S4](#)). Sample libraries were prepared for sequencing following RAD Capture (Rapture) methods outlined in Ali et al. ([2016](#)). These samples were then used to identify putative high-quality SNPs following sequencing.

Sampled individuals were aligned against the de novo partial genome reference using the BWA-MEM algorithm (Li and Durbin [2010](#), Li [2013](#)), and converted to BAM format and filtered for properly paired alignments using Samtools (Li et al. [2009](#)). Next, alignments from three different sequencing runs on an Illumina HiSeq were merged together and duplicates were removed using Samtools (Li et al. [2009](#)). For all downstream analysis, we selected individuals that had greater than 25,000 alignments (n=311), which provided sufficient data to investigate population genetic attributes at broad and fine geographic scales ([Appendix, S5](#)).

To generate SNP (i.e., segregating site) data, a probabilistic framework was used for all population genetic analyses as it does not require calling genotypes and is suitable for low-coverage sequencing data (Korneliussen et al. [2013](#), Fumagalli et al. [2013](#)). SNP discovery, minor allele frequencies (MAF) estimates, and genotype probabilities were conducted using ANGSD (Korneliussen et al. [2014](#)). ANGSD analyses were conducted following methods from Prince et al. ([2017](#)), with a minimum mapping quality score (`minMapQ`) of 10, a minimum base quality score (`minQ`) of 20, the genotype likelihood model (`GL 1`), specifying the Rapture bait locations using the `sites` flag, and only sites represented in at least 50% of the included samples (`minInd`) were used. Furthermore,

METHODS

genomic sites were designated as polymorphic only if MAFs were greater than 0.05 and the probability of the site not being polymorphic was less than $1e^{-6}$. Using this approach, over 44,000 polymorphic sites were identified across all *R. boylii* study samples.

PCA and Admixture

To assess population structure and coancestry, ANGSD was used to generate PCA and NG admix was used to calculate admixture. Settings used in ANGSD for PCA to identify polymorphic sites included a `SNP_pval` of $1e^{-6}$, inferring major and minor alleles (`doMajorMinor` 1), estimating genotypic likelihoods (`GL` 1), estimating allele frequencies (`doMaf` 2) (Kim et al. 2011), retaining SNPs with a minor allele frequency of at least 0.05 (`minMaf`), specifying the Rapture bait locations using the `sites` flag, estimation of genotype posterior probabilities using a uniform prior (`doPost` 2), and the `doIBS` 1 and `doCov` 1 options. Principal components (PC) summarizing population structure were derived from classic eigenvalue decomposition and were visualized using the ggplot2 package in R (R Core Team 2017). To assess admixture between *R. sierrae* and *R. boylii*, genotype likelihood data (`GL` 1) was generated in ANGSD with the same settings as above, in addition to retaining only SNPs that were shared in at least of 50% of the samples, `doPost` 2, `doGLF` 2, and limiting to higher quality alignment data (`minMapQ` 10, `minQ` 20). We then used NG admix (Skotte et al. 2013) to infer ancestry proportions in *R. sierrae* and *R. boylii* individuals. NG admix is a robust admixture method that can be applied to low-depth NGS data, and does not require called genotypes, thus reducing error associated with potential ascertainment and uncertainty in the data (Skotte et al. 2013).

METHODS

F1 vs. F2 Test with Species Diagnostic SNPs

To test whether hybrids were first generation filial (F1) hybrids or progeny from F1 hybrids from subsequent generations (e.g., F2, F3, etc.), we identified differentially fixed (i.e., species-specific) SNPs and assessed heterozygosity at these loci in hybrid individuals. F1 and F2 hybrid individuals should exhibit different degrees of heterozygosity in species-diagnostic SNPs. We called genotypes in ANGSD using a uniform prior (`doPost 2`) and the following settings: `GL 1`, `doGeno 13`, `postCutoff 0.95`, `doMaf 1`, `doMajorMinor 1`, `minInd 2`, `SNP_pval 1e-6`, `minMapQ 20`, `minQ 20`, and specifying the Rapture bait locations using the `sites` flag. The subsequent output (`*.geno.gz`) was then processed in the program R using the `dplyr` package (Wickham et al. 2018) to manipulate and filter to homozygous diagnostic SNPs. Data were filtered to include only loci where over 50 non-hybrid individuals from each species had called genotypes at a given polymorphism.

Demographic Modeling with fastsimcoal2

To quantify divergence times and migration rates between *R. sierrae* and *R. boylii*, we used coalescent simulations in fastsimcoal2 (Excoffier and Foll 2011, Excoffier et al. 2013). This maximum-likelihood modeling approach uses simulations to estimate the expected site-frequency spectra (SFS) for a demographic model of interest to calculate a composite likelihood, and then utilizes a maximization procedure to find the maximum-likelihood parameter estimates.

We calculated folded joint SFS for each species in each watershed from SNP data generated from ANGSD because the ancestral condition is unknown. For all models, we assumed the potential for bidirectional gene flow and that extant genetic clusters emerged simultaneously from a common ancestry. We tested models that allowed for popula-

METHODS

tion growth, and models with no growth. We used two conservative model scenarios to estimate divergence times and migration rates between species in each watershed. To estimate migration probabilities per generation between species within each watershed, we set the divergence time parameters between 1–1.1 billion years ago to create simplified migration-only models. To estimate divergence time between species, we used the watershed that had the lowest migration rate from the previous migration-only models, and generated divergence time estimates assuming no migration between species.

The basic steps taken to obtain final model estimates from fastsimcoal2 used a set of 25 replicate models, followed by comparison of maximum observed and expected likelihoods to select the best-fit model (Akaike 1973), then simulate new SFS using the best-fit model for parametric bootstrapping. Following Excoffier and Foll (2011), we used 1,000 randomly drawn SNPs from each SFS to generate 100,000 coalescent simulations for likelihood calculations (estimation of the expected SFS) with a maximum of 40 cycles for the conditional maximization algorithm. To select the best-fit model we selected the model replicate that minimized the difference between the maximum expected likelihood and the maximum observed likelihood. We used parametric bootstrapping to generate 95% confidence intervals for each best-fit model using 100 bootstraps for each model and selected the best model from each bootstrap based on maximum likelihoods as described above.

RESULTS

Results

Rapture produced high quality genomic data for both *R. sierrae* and *R. boylii*

Individual samples were collected across 56 different sampling localities in three different watersheds (Figure 2.1, Table 2.1). For downstream analysis, we filtered and retained 311 samples from the original sequencing data that contained a minimum of 25,000 alignments (Appendix, S5). The final merged dataset mean alignments per sample was 229,485 (Appendix, S5), and the mean number of samples per site was eight. These frog species are cryptic, and often occur in low densities, so we retained all sites in our analysis, regardless of the number of samples per locality (Table 2.1). We conclude that the sequence data we obtained should be appropriate for population genetic analyses across our study area.

RESULTS

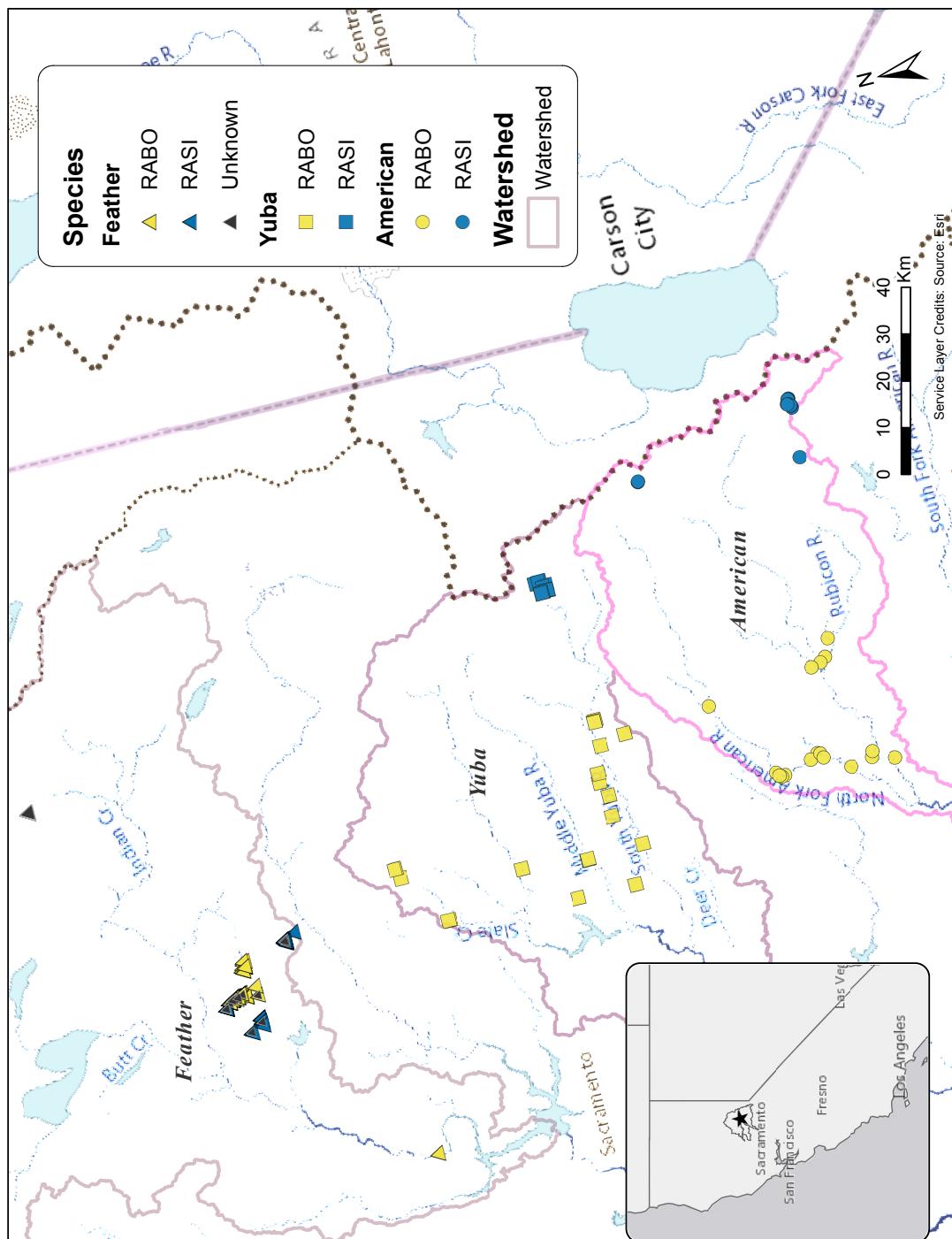


Figure 2.1: Map of sampling locations in the Feather, Yuba, and American watersheds. RABO=*R. boylii*, RASI=*R. sierrae*.

RESULTS

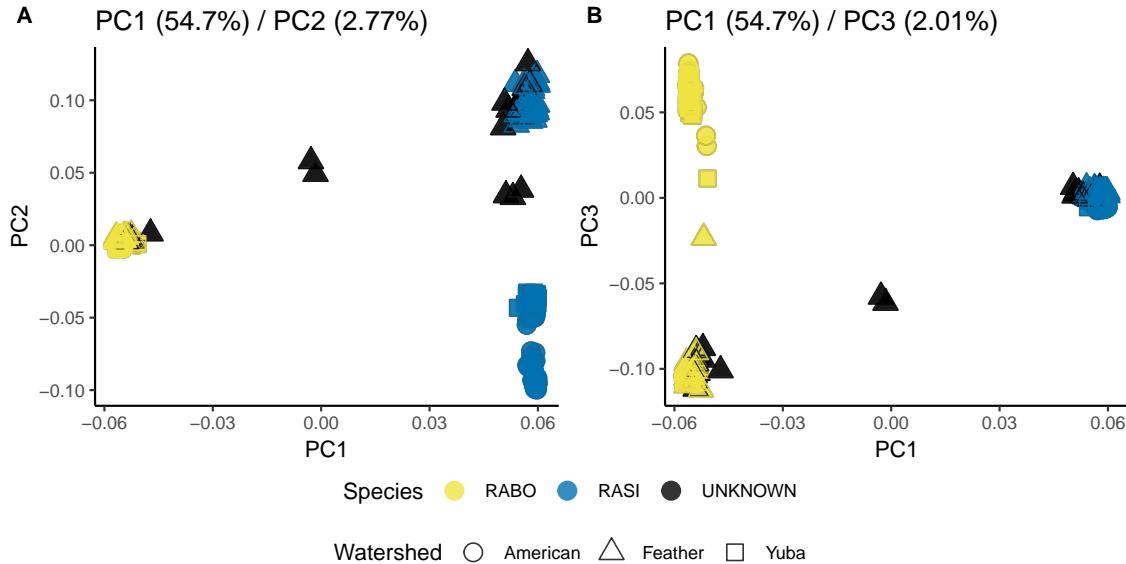


Figure 2.2: Principal component analysis of Rapture sequencing data, RABO=*R. boylii*, RASI=*R. sierrae*. A) PC1 vs. PC2; B) PC1 vs. PC3.

PCA shows strong separation between species and identifies putative hybrids

To assess within-basin population structure, principal components analysis (PCA) was used to provide a dimensionless comparison of putative SNPs across species and watersheds (Figure 2.2). Strong differentiation was observed between species (*R. sierrae* and *R. boylii*) on the PC1 axis, which accounted for approximately 55 percent of the variation. PC2 differentiated *R. sierrae* among the three watersheds (Figure 2.2A), while PC3 differentiated *R. boylii* sampling locations among the three watersheds (Figure 2.2B). Little sign of admixture between the two species appears in the PCA, however, two samples—collected in the Feather watershed and designated as “unknown” in the field—clustered halfway between the *R. sierrae* and *R. boylii* groups along PC1, suggesting these individuals were hybrids.

RESULTS

Admixture shows two unknown individuals with equal *Rana* species ancestry

To further investigate if the two unknown individuals identified in the PCA were potential hybrids of *R. sierrae* and *R. boylii*, we used NGSAdmix to assess population structure and individual ancestry from genome-wide SNPs (Skotte et al. 2013). We used k=2 to evaluate the fraction of ancestry derived from each species. Admixture showed the same two unknown samples from the Feather basin had approximately 50% ancestry from each species (*R. boylii* and *R. sierrae*), confirming their hybrid ancestry (Figure 2.3). Furthermore, ancestry in the individuals designated as “unknown” in the field also showed very low levels of mixed ancestry between the species. There were very low or nearly non-existent levels of mixed-ancestry in the American and Yuba basins as compared to the Feather. However, introgression between *R. sierrae* and *R. boylii* appears asymmetric, with a greater proportion of ancestry from *R. boylii* occurring in the *R. sierrae* samples, particularly in the Feather watershed, and predominantly in the “unknown” individuals. The putative hybrid individuals were sampled in Bean Creek, a tributary to Spanish Creek ((Figure 2.4)). Bean Creek was one of the only tributaries where both *R. sierrae* and *R. boylii* co-occur; therefore we conclude there is strong evidence for recent hybridization between *R. sierrae* and *R. boylii* in this drainage.

RESULTS

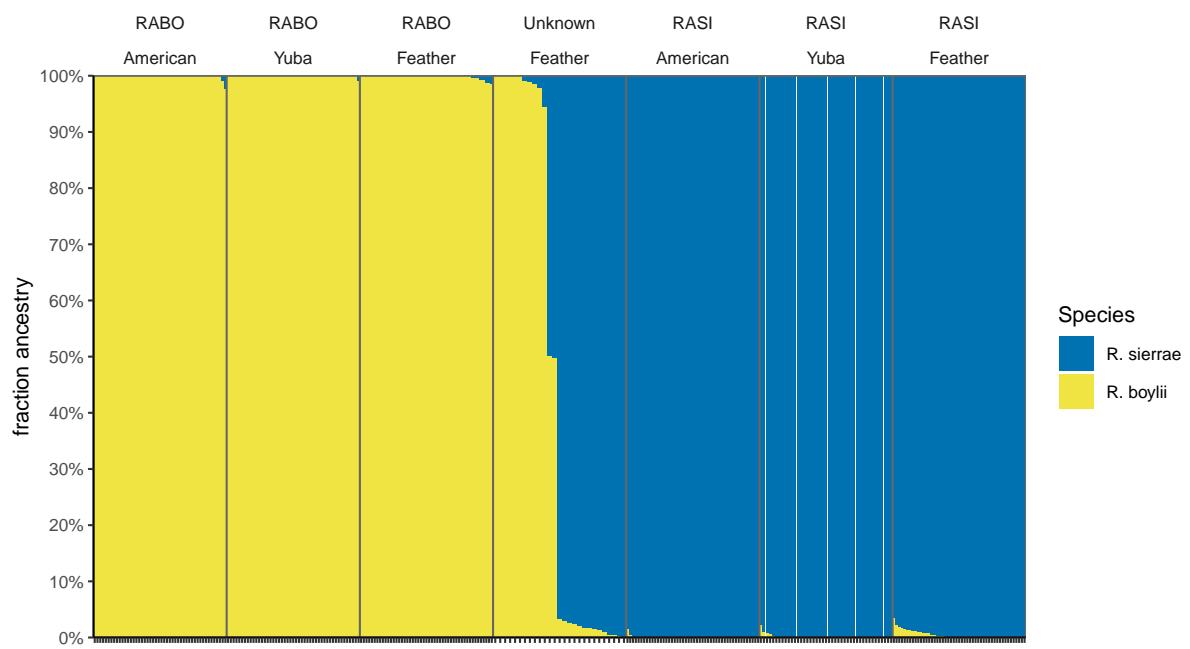


Figure 2.3: Admixture ($k=2$) of *R. sierrae* and *R. boylii* and “Unknown” *Rana* samples from the Feather, American, and Yuba watersheds.

RESULTS

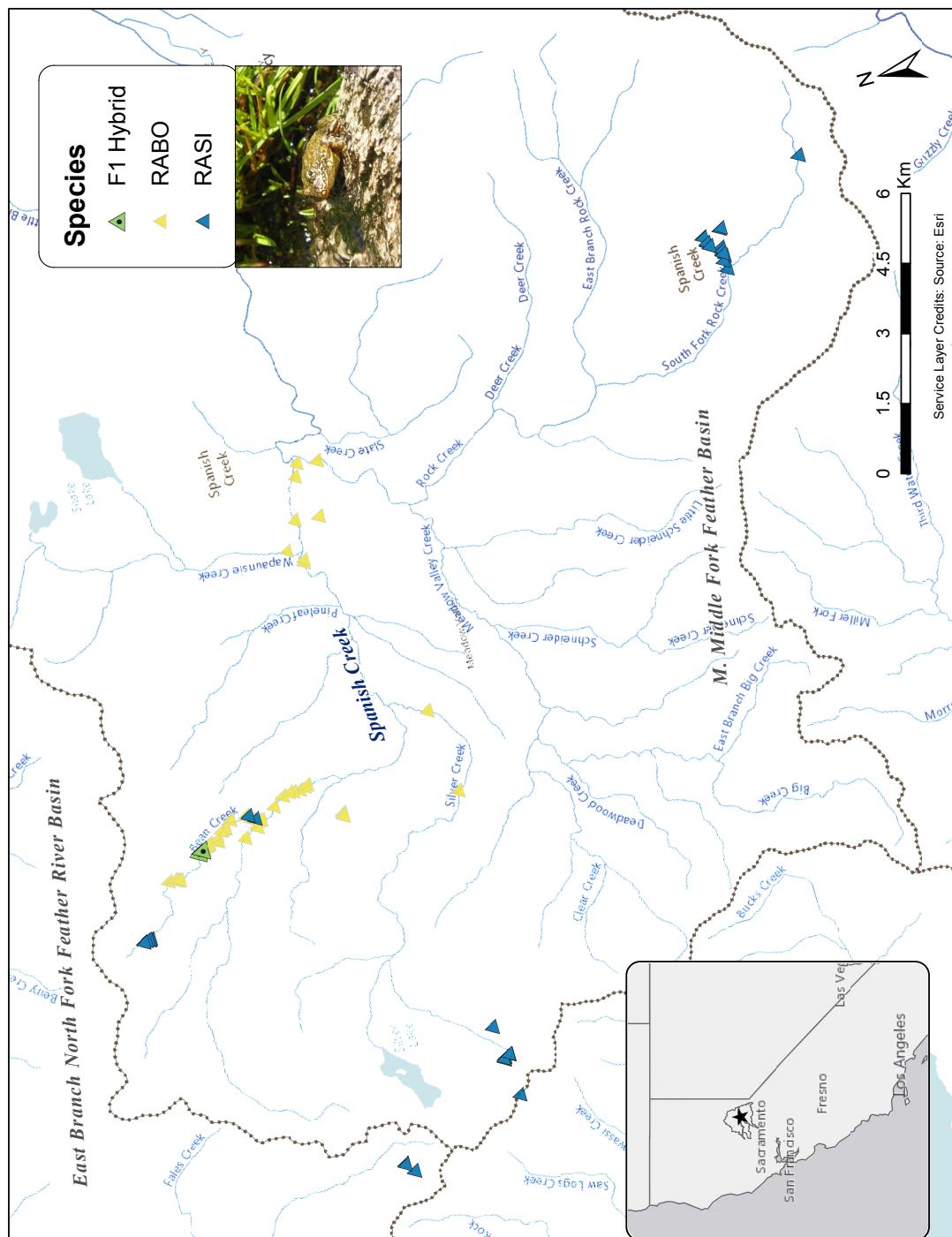


Figure 2.4: Map of sample locations in Bean Creek/Spanish Creek in the Feather watershed where hybrids were identified. RABO=*R. boylii*, RASI=*R. sierrae*.

RESULTS

F1 vs. F2 test on hybrids

To test whether the hybrids were F1 (first-generation) or F2 (progeny of two F1's), we identified species diagnostic SNPs. Our filtering process (see [Methods](#)) yielded 3,062 putative diagnostic SNPs that were homozygous for different alleles in *R. sierrae* or *R. boylii* samples and also had successfully called genotypes in the two hybrid individuals. F1 hybrids should be exclusively heterozygous at species diagnostic SNPs. In contrast, F2 hybrids should heterozygous for 50% of the species diagnostic SNPs, and homozygous at the remaining 50% with 25% allotted to each species. We observed extremely high heterozygosity and very low homozygosity (6% genotyped as *R. boylii*, 4% *R. sierrae*, and 89% were heterozygous) ([Figure 2.5](#)). This level of heterozygosity is far greater than expected for F2 individuals, and the presence of homozygous genotype calls in the hybrid individuals at species diagnostic SNPs is expected due to low coverage sequencing data; genotyping from low coverage sequencing will cause a low frequency of erroneous homozygous calls, because only one of the two alleles is sampled, causing heterozygotes to be called as homozygotes. We conclude these hybrid individuals are F1 instead of F2 individuals. Furthermore, the hybrid individuals were found to have *R. sierrae* mitochondrial DNA (Bedwell and Goldberg, *in review*), indicating the female was from a *R. sierrae* individual and the male was from *R. boylii* in both cases.

RESULTS

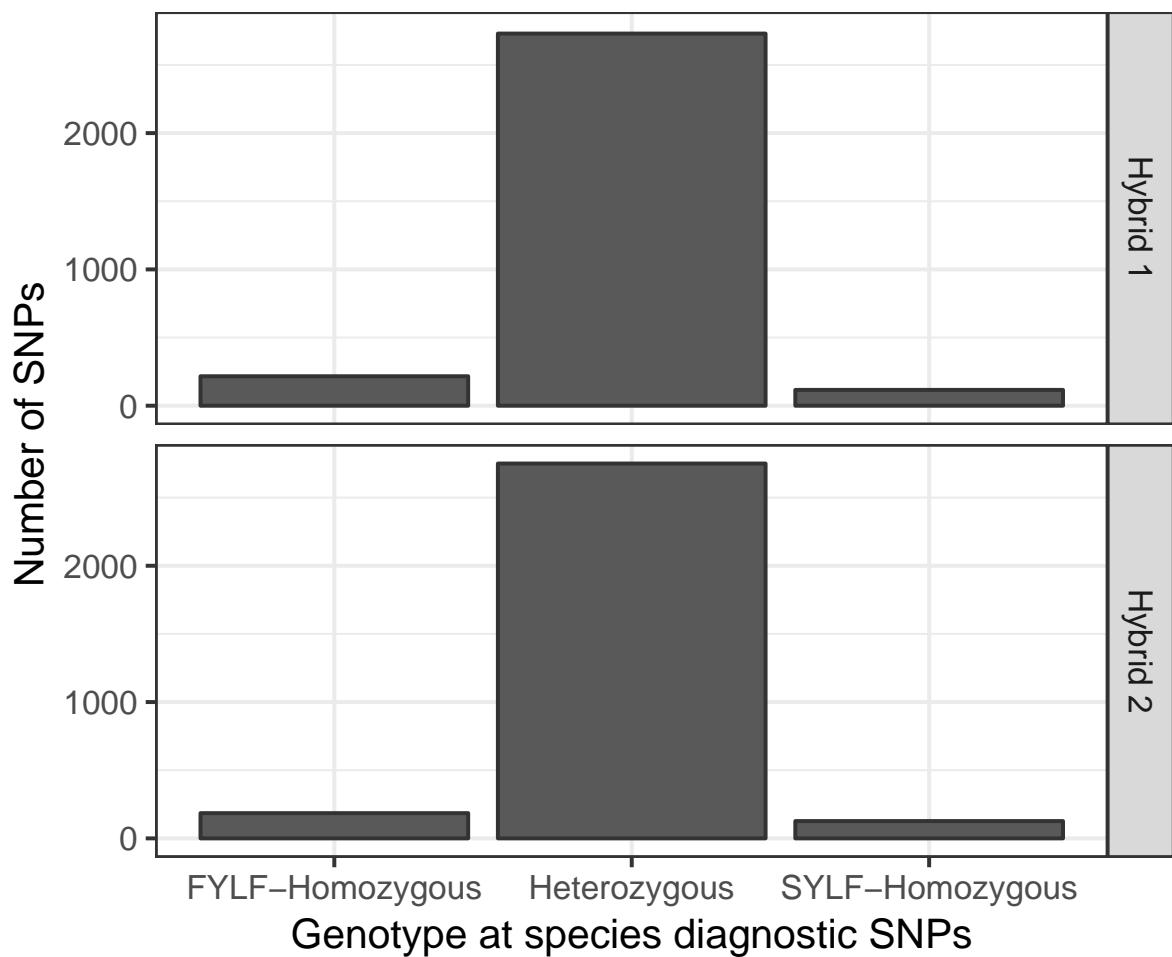


Figure 2.5: The F1 vs. F2 test using species diagnostic SNPs to assess heterozygosity in hybrid individuals. RABO=*R. boylii*, RASI=*R. sierrae*.

RESULTS

Divergence times and migration rates

To test for differential migration rates between *R. sierrae* and *R. boylii* in the Feather watershed compared to the Yuba and American, we used fastsimcoal2 (Excoffier and Foll 2011) coalescent simulations. Using all individuals except the two hybrid individuals, we found migration probability (or the per generation likelihood that any gene from one population transfers to another) from *R. sierrae* and *R. boylii* was highest in the Feather, with a mean of $5.29e^{-6}$ (95% CI $5.28e^{-6}$ – $5.30e^{-6}$) and lowest in the American watershed, $7.80e^{-7}$ ($7.78e^{-7}$ – $7.81e^{-7}$) (Figure 2.6). The migration probabilities in the Yuba watershed were lower than estimates from the Feather, but were closer in magnitude, $3.84e^{-6}$ ($3.83e^{-6}$ – $3.85e^{-6}$). We found migration rates from *R. boylii* to *R. sierrae* were extremely low across all three (Feather = $6.05e^{-7}$ [$6e^{-7}$ – $6.10e^{-7}$], Yuba = $3.76e^{-7}$ [$3.73e^{-7}$ – $3.79e^{-7}$], American = $8.29e^{-8}$ [$8.23e^{-8}$ – $8.35e^{-8}$]; mean and 95% CI) (Figure 2.6A). As observed in the admixture analysis, migration rates were asymmetric, showing F1 individuals backcrossing to individuals from *R. boylii* more often than to *R. sierrae*. We conclude migration probability rates are highest in the Feather watershed from *R. sierrae* and *R. boylii*, with very limited migration occurring from *R. boylii* to *R. sierrae*.

To estimate divergence time between species, we used the American watershed samples because migration estimates between *R. sierrae* and *R. boylii* were lowest, and we wanted to derive a conservative estimate of divergence by minimizing inaccuracy caused by migration. We ran fastsimcoal2 models with no migration and divergence times bounded between 10 kya and 4 mya. The best model based on maximum likelihood estimated the time since divergence between *R. sierrae* and *R. boylii* was 370,856 (95% CI: 370,041–371,670) generations. Typically, *R. boylii* have a generation time of 2–3 years, depending on the region (Kupferberg et al. 2009, Railsback et al. 2015), while *R. sierrae* can have a greater range of generation times, between 3–6 years because tadpoles may overwinter

RESULTS

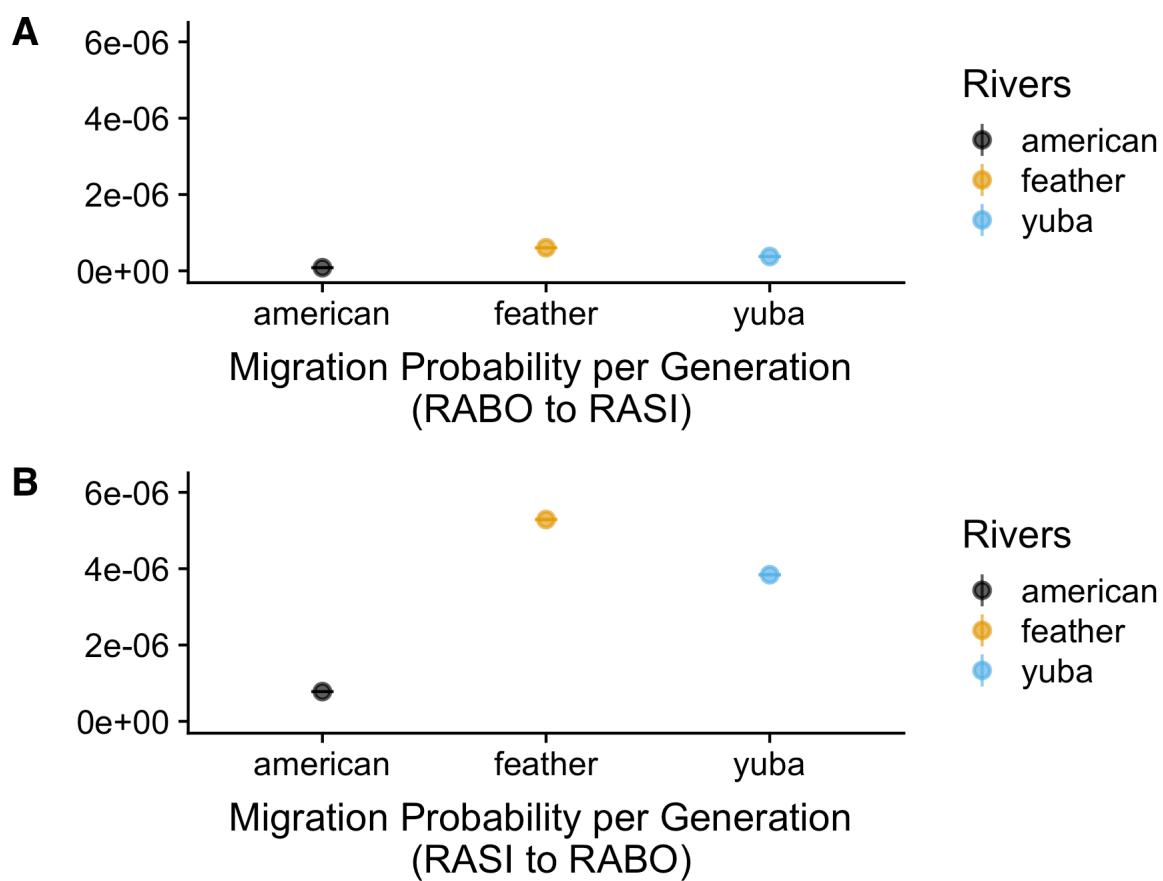


Figure 2.6: Estimates of migration probabilities from fastsimcoal2 models between the two species within the Feather, Yuba, and American watersheds. RABO=*R. boylii*, RASI=*R. sierrae*, with 95% confidence intervals from 100 bootstrapped estimates.

DISCUSSION

as many as three years (Knapp et al. 2003, 2016). We may assume the ancestral condition was derived from *R. boylii* (Macey et al. 2001, Vredenburg et al. 2007, Yuan et al. 2016), therefore we suggest a generation time between two and three years, which means *R. sierrae* likely diverged from *R. boylii* 741 kya to 1.1 mya. This time period corresponds to the early Pleistocene during an era of episodic glaciation (“the Great Ice Age”—where distributions likely contracted and lineages became isolated—followed by subsequent interglaciation, where distributions expanded (Birkeland 1964, Gillespie and Zehfuss 2004).

Discussion

We identified strong divergence between *R. sierrae* and *R. boylii* across all three watersheds, evidence of two F1 hybrids, and low levels of asymmetric introgression primarily in the Feather basin. Hybridization between *R. sierrae* and *R. boylii* has not been previously documented based on field observations and breeding experiments (Zweifel 1955).

It is unlikely that there is currently the potential for major introgression between *R. sierrae* and *R. boylii*, particularly as hybridization initially may not be adaptive and is often selected against (Abbott et al. 2013, Streicher et al. 2014). Although hybridization may be common between some amphibian species (Malone and Fontenot 2008) and can even occur between highly divergent taxa—up to 21 million years divergent (Prager and Wilson 1975)—our data show there is strong pattern of divergence between *R. sierrae* and *R. boylii* with limited hybridization and introgression between the species. Furthermore, there are currently few localities where *R. sierrae* and *R. boylii* occur sympatrically, and populations of either species are typically sparse in the Sierra (Kupferberg et al. 2012, Catenazzi and Kupferberg 2013). Additionally, these two species may be strongly influ-

DISCUSSION

enced by elevation due to life history differences (*R. sierrae* are typically found in higher elevations and are capable of overwintering as tadpoles while *R. boylii* are not). Previous work suggests that elevation strongly influences genetic structure in frogs (Monsen and Blouin 2004)—further reinforced by the patterns of strong divergence between species within watersheds that we observe in our data.

There remains the potential for low-levels of naturally occurring hybridization and introgression between *R. sierrae* and *R. boylii*, but currently both species appear to have clear genotypic divergence even in an area of sympatry. Thus, our data suggest this is unlikely to be a major concern for conservation management. While there is a potential for misclassification of individuals in intermediate locations, genetic testing, and/or monitoring could be a useful tool for clarifying species and population boundaries as well as population size estimates. Successful tests of not only hybridization, but timing of divergence events (as well as better understanding bottlenecks and population expansion) based on landscape history can be informative in understanding what events may have driven divergence. The landscape of the Sierra Nevada during the Pleistocene epoch was one of repeated glaciation (Moore and Moring 2013). Rivers flowing into the present-day Central Valley were being alternately eroded by west-flowing streams during interglaciation or covered in glaciers. It is therefore likely that adaptation to colder climates (e.g., freezing lakes and streams) may have provided an advantage to individuals or populations occurring in localities where the effects of glaciation were most prominent. *Rana sierrae* are uniquely adapted to persist in short-growing periods common in the high Sierras—tadpoles may overwinter multiple years before metamorphosing—thus *R. sierrae* may have diverged from *R. boylii* because of their ability to persist in colder climates, common during periods of glaciation during the Pleistocene.

While timing of *R. sierrae* and *R. boylii* divergence in the Sierra Nevada may correspond

DISCUSSION

with the onset of the Pleistocene glaciation, more recent anthropogenic changes may be a stronger driver of current population connectivity and structure. With increased global temperatures and more variable winter periods in the Sierra Nevada, long-term persistence of high elevation species such as *R. sierrae* may reside in the ability for the species to adapt to significant change. Future hybridization events could provide paths for introgression of selectively favored alleles between *R. sierrae* and *R. boylii*, and our data show this could be possible, given the presence of gene flow between species; even very limited introgression could provide adaptive alleles for subsequent positive selection.

In rare species with small population sizes, hybridization outcomes that fail to produce successful offspring (sterile F1 hybrids) may have a greater cost on the species with low numbers of effective breeders, affecting both locally adapted populations and negatively impacting the probability of population persistence in a given region (Pagano et al. 2003). For *R. sierrae*, current patterns of hybridization do not appear likely to affect population persistence; however under future scenarios (e.g., warming climate, range contraction, population crashes) the loss of even several breeding individuals (via reproduction with *R. boylii*) may have a significant impact in declining populations. Given the pattern of asymmetric admixture and migration observed in the Feather watershed, it is likely there are more *R. boylii* present in the region the hybrids were observed than *R. sierrae*. This difference could potentially lead to greater competition for *R. sierrae* females and reducing male *R. sierrae* reproductive success through the loss of mating opportunities. This may lead to a reduction in the fitness of *R. sierrae* females because the female deposits one egg clutch per year.

CONCLUSION

Conclusion

Assessing the impacts of current landscape and watershed change on the genetic variation of organisms, particularly sensitive and endangered species, may be a crucial tool for monitoring and more robust restoration, translocation, and conservation efforts. Future conservation of these species will require several key components, including establishing higher resolution population boundaries across the species' ranges, particularly in the northern Sierra Nevada, delineation of distinct population segments that can be utilized in conservation management, and quantification of relative genomic health of these groups. Identification of hybridization is a key step towards better delineating management units and further understanding what conservation steps may be taken.

Acknowledgements

Many thanks to all who helped collect/provide/prepare samples: Corey Luna, Rick Wachs, and Sarah Musselman. Thanks to Cathy Brown, Colin Dillingham, and all the USFS field crew members who helped collect these samples.

Chapter 3

Refining conservation unit boundaries of a sentinel stream-breeding frog (*Rana boylii*) using population genomics

Introduction

Modern genomic sequencing technology has permitted much higher resolution analyses of both geographic and ecological patterns in populations (Nunziata et al. 2017, Hendricks et al. 2018, Barbosa et al. 2018). Reduced representation sequencing methods such as restriction site-associated DNA sequencing (RADSeq) (Miller et al. 2007, Baird et al. 2008, Ali et al. 2016) provide a powerful means to address ecological genomics questions at scales that were previously impossible using traditional field methods. Furthermore, new methods such as RAD Capture (Rapture) (Ali et al. 2016) adapt RADSeq to target

INTRODUCTION

desired loci and allow highly efficient genotyping of thousands of individuals at once. Because historical and future landscape changes can influence species demography and migration patterns (Burkey 1989, Anderson and Beer 2009, Barbosa et al. 2018), these genomic tools are increasingly valuable for assessing critical factors for long-term persistence of sensitive populations or species.

Freshwater biota are rapidly declining globally (Ricciardi and Rasmussen 1999), and conservation efforts require assessment of the adaptive capacity of populations to rapid environmental change. Given limited resources available for conservation, it is important to define and establish clear geographic boundaries for conservation units such as distinct population segments across a species' range. Delineation of distinct population segments can be used for prioritizing objectives in conservation management. Furthermore, quantification and comparison of relative genetic diversity within and among populations can provide additional information as a benchmark for assessment of conservation actions. Thus, quantifying and linking landscape change with genetic diversity metrics allows tracking of how sensitive populations respond to future environmental change (through reduced adaptive potential) as well as evaluating whether restoration efforts are effective (i.e., increasing genetic connectivity, diversity, effective breeder/population size).

Amphibians are particularly sensitive to changes in the ecosystem due to their physiology and ontogeny (Davidson et al. 2002, Beebee and Griffiths 2005); thus the ability to utilize environmental variables as life history cues can be especially important. In highly dynamic riverine environments, organisms must constantly adapt to temporal and spatial changes. One such sentinel stream-breeding species is the foothill yellow-legged frog (*Rana boylii*), a native to California and Oregon which historically occurred in lower elevation (0-1500m) streams and rivers from Southern Oregon to northern Baja California west of the Sierra-Cascade crest (Stebbins 2003). As a lotic breeding amphibian, *R.*

INTRODUCTION

boylii is tied closely to the local hydrology in watersheds it inhabits, and therefore it is particularly sensitive to alterations to flow regimes (Kupferberg 1996, Lind et al. 1996, Kupferberg et al. 2012).

Similar to many amphibian species in California, there have been significant population declines across the former range of this species (Davidson 2004, Peek 2010, Thomson et al. 2016). Particularly in southern California and the Sierra Nevada, *R. boylii* has been extirpated from approximately 50 percent of its historical range (Jennings and Hayes 1994, Davidson et al. 2002). *Rana boylii*, currently designated as a species of special concern (CDFW) in the state of CA, has been petitioned as candidate for listing under the federal (USFWS) Endangered Species Act (USFWS 2014) as well as the state (CDFW) Endangered Species Act.

Effective conservation management of this species will need to consider and prioritize maintenance of genetic diversity as part of any listing decision because it is closely related to the evolutionary capacity for adaptation to environmental changes (Lande and Shannon 1996). Thus, utilizing genetic data provides a potentially informative process for identifying the impacts of anthropogenic and environmental change on the process of adaptation. Establishing high-resolution genetic boundaries for populations across the species range as well as quantification of genomic metrics (i.e., genomic diversity, population connectivity) would help managers prioritize conservation actions.

A recent study by McCartney-Melstad et al. (2018) identified five major clades in *R. boylii* with strong geographically structured genetic subdivision across its range in California and Oregon. Here we provide an additional population genomic analysis across the range of this declining sentinel stream species. We provide additional geographic and genetic resolution to McCartney-Melstad et al. (2018), as well as quantify genetic diversity metrics across subpopulations and clades as both a reference and assessment of

METHODS

the potential for long-term persistence across the species' range.

Methods

Sampling and DNA extraction

A total of 1103 individual tadpole tail clips, buccal swabs, or tissue samples were compiled, collected between 1992 and 2016 across the range of *R. boylii*. Field sampling was conducted following methods in Heyer et al. (1994) under CDFW SCP Permit #0006881, with IACUC protocol #19327. Individual post-metamorphic frogs were buccal-swabbed following established protocols (Goldberg et al. 2003, Pidancier et al. 2003, Broquet et al. 2007). Each post-metamorphic individual was comprehensively swabbed underneath tongue and cheek for approximately one minute. Swabs were air dried for approximately five minutes and placed in 1.5 mL microcentrifuge tubes while in the field. Samples were stored in the laboratory at -80°C until DNA extraction. Where possible, tail clips from tadpole larvae were collected, and tadpoles greater than 15 mm total length were targeted (Wilbur and Semlitsch 1990, Parris et al. 2010). One clip was taken per individual tadpole and dried on Whatman filter paper (grade 1) and stored at room temperature. Some older tissue samples consisted of toe clips placed in 100% ethanol for storage, and DNA extraction from these samples used Qiagen DNeasy kits following the manufacturer's protocol. Buccal swabs and tail clip DNA were extracted using an Ampure magnetic bead-based protocol (Ali et al. 2016). DNA samples were stored at -20°C.

METHODS

Generating high-quality sequencing data

To produce a high-quality genomic resource for frog species with large genome sizes, we interrogated a significant fraction of the *R. boylii* genome using a SbfI restriction enzyme and high-density RAD sequencing on an Illumina HiSeq (Miller et al. 2007, Baird et al. 2008). Paired-end sequence data were generated using 24 *R. boylii* individuals (Appendix, S2). RAD libraries were constructed following the protocol described in Ali et al. (2016). De novo loci discovery and contig extension were carried out via custom PERL scripts (Miller et al. 2012), the alignment program Novoalign and the genome assembler PRICE (Ruby et al. 2013). This pipeline resulted in a set of 77,544 RAD contigs ranging from 300 to 800 bp which served as a de novo partial genome reference for all subsequent downstream analyses (Appendix, S3). Using these data, we filtered data to loci with 4 or fewer SNPs, and randomly selected 10,000 loci from this subset. Using these RADSeq data, 8,533 RAD capture baits (120bp) were designed by Arbor Biosciences from the de novo alignment (Appendix, S4). The number of polymorphic loci identified across all *R. boylii* study samples was 44,406. RAPTURE was then used to identify putative high-quality SNPs.

Three different sequencing runs on an Illumina HiSeq were merged together, filtered, and duplicates were removed using ANGSD and Samtools (Li et al. 2009). Sampled individuals were aligned against the de novo partial genome reference using the BWA-MEM algorithm (Li and Durbin 2010, Li 2013) and saved to BAM format. To generate SNP (segregating site) data, a probabilistic framework was used for all population genetic analyses as it does not require calling genotypes and is suitable for low-coverage sequencing data (Korneliussen et al. 2013, Fumagalli et al. 2013). Estimates of per site minor allele frequencies (MAF), genotype probabilities and SNP discovery were conducted using ANGSD and NGStools (Fumagalli et al. 2014, Korneliussen et al. 2014).

METHODS

Genomic sites were designated as polymorphic only if MAFs were greater than 0.05 and the probability of the site not being polymorphic was less than 10^{-12} . ANGSD analyses were conducted following methods from Prince et al. (2017), with a minimum mapping quality score (`minMapQ`) of 10, a minimum base quality score (`minQ`) of 20, the genotype likelihood model (`GL 1`), and only sites represented in at least 50% of the included samples (`minInd`) were used (Li and Durbin 2011).

Quantifying genetic structure

To characterize and quantify genetic population structure within and among watersheds, we conducted principal component analysis (PCA) using data subsampled to different alignment thresholds (e.g., all individuals with a minimum of 100,000 alignments) to determine the amount of data needed for population analyses. For downstream analysis, we selected individuals that had greater than 100,000 alignments. To assess population structure and coancestry, ANGSD was used to generate PCA and NGSadmix was used to estimate admixture. Settings used in ANGSD for PCA to identify polymorphic sites included a `SNP_pval` of $1e^{-6}$, inferring major and minor alleles (`doMajorMinor 1`), estimating allele frequencies (`doMaf 2`) (Kim et al. 2011), retaining SNPs with a minor allele frequency of at least 0.05 (`minMaf`), estimation of genotype posterior probabilities using a uniform prior (`doPost 2`), specifying the RAPTURE bait locations using the `sites` flag, calculating the PCA matrix with the `doIBS 1` and `doCov 1` options, and limiting the analysis to higher quality alignment data (`minMapQ 10, minQ 20`). Principal components (PC) summarizing population structure were derived from classic eigenvalue decomposition and visualized using the `ggplot2` and `cowplot` packages in R (Wickham 2016, R Core Team 2017, Wilke 2018). To assess admixture in *R. boylii*, genotype likelihood data (`GL 2` and `doGLF 2`) was generated in ANGSD with the same settings as above.

METHODS

We then used NGSadmix (Skotte et al. 2013) to infer ancestry proportions in *R. boylii* individuals. NGSadmix is a robust admixture method that can be applied to low-depth NGS data, and does not require called genotypes, thus reducing error associated with potential ascertainment and uncertainty in the data (Skotte et al. 2013).

Genetic differentiation and diversity estimates

Rana boylii are cryptic, and often occur in low densities within the study area. Thus, we retained a minimum of three individuals per locality for estimates of genetic diversity and F_{ST} . With genomic data, population genetic parameters can be accurately estimated from even low sample numbers (Hotaling et al. 2018), and genomic analyses in non-model organisms often use fewer loci (Narum et al. 2013). To quantify genetic variation and differentiation, pairwise population differentiation (F_{ST}) was calculated and scaled [mean F_{ST} / (1-mean F_{ST})] to examine the relationship between genetic differentiation and geographic distance between populations (Wright 1943, Weir and Cockerham 1984, Rousset 1997). F_{ST} was estimated by first calculating a folded site frequency spectrum (SFS) for each population from site allele frequencies (SAF) in ANGSD (`doSaf 1, fold 1, minMapQ 10, minQ 20, GL 2`) and specifying the Rapture bait locations using the `sites` flag (Nielsen et al. 2012). The two-dimensional SFS between each population pair were then estimated from folded `SAF.idx` files using `maxIter 100` with `realSFS` (Korneliussen et al. 2014). F_{ST} statistics were then calculated from two-dimensional SFS (2DSFS) for each possible pairwise combination of unique collection locations using an estimator preferable for small sample sizes implemented in ANGSD (`whichFST 1`). These values were plotted in R.

We summarized patterns of genetic variation using two estimators of θ ($4N\mu$): Tajima's θ (θ_π) is based on the average number of pairwise differences (Tajima 1983) and Watterson's

RESULTS

θ (θ_S) is based on the number of segregating sites (Watterson 1975). These estimators are influenced by the demographic history of a population and provide information on the trajectory of changes in genetic diversity. When genetic diversity has been stable, these estimates are generally equal; but when genetic diversity has been increasing, $\theta_\pi > \theta_S$; and when genetic diversity has been decreasing, $\theta_S > \theta_\pi$. To calculate θ statistics from Rapture data, we used folded SFS in ANGSD with `GL 2`, `doThetas 1`, `doSaf 1 fold 1`, and `pest`. Outputs were used to calculate each statistic for each site using `thetaStat` with `make_bed` and then `do_stat`. These data were averaged over the sites to obtain a single “genome-wide” value for each statistic for each locality (Korneliussen et al. 2013).

Results

A total of 1,103 individual samples were sequenced using Rapture. For principal components analysis (PCA) and admixture, we selected samples that had greater than 100,000 alignments and had 1 or more individuals per sampling locality. For localities with greater than ten individuals, we randomly sampled a maximum of 10 samples, yielding 480 total samples from 89 distinct localities across the range of the species (Figure 3.1, Table 3.1). These localities overlap many of the localities used in McCartney-Melstad et al. (2018), with a few notable differences. There were more individuals available for analyses at most of the localities (Table 3.1), there was higher resolution sampling in certain areas (i.e., the northern coast of California, the Feather watershed), and a two additional localities fall outside of the clades delineated by McCartney-Melstad et al. (2018) (i.e., Locality 1 in the SF American basin in El Dorado County, and Locality 4 in the Honey-Eagle Lakes basin in Lassen County; Figure 3.1).

RESULTS

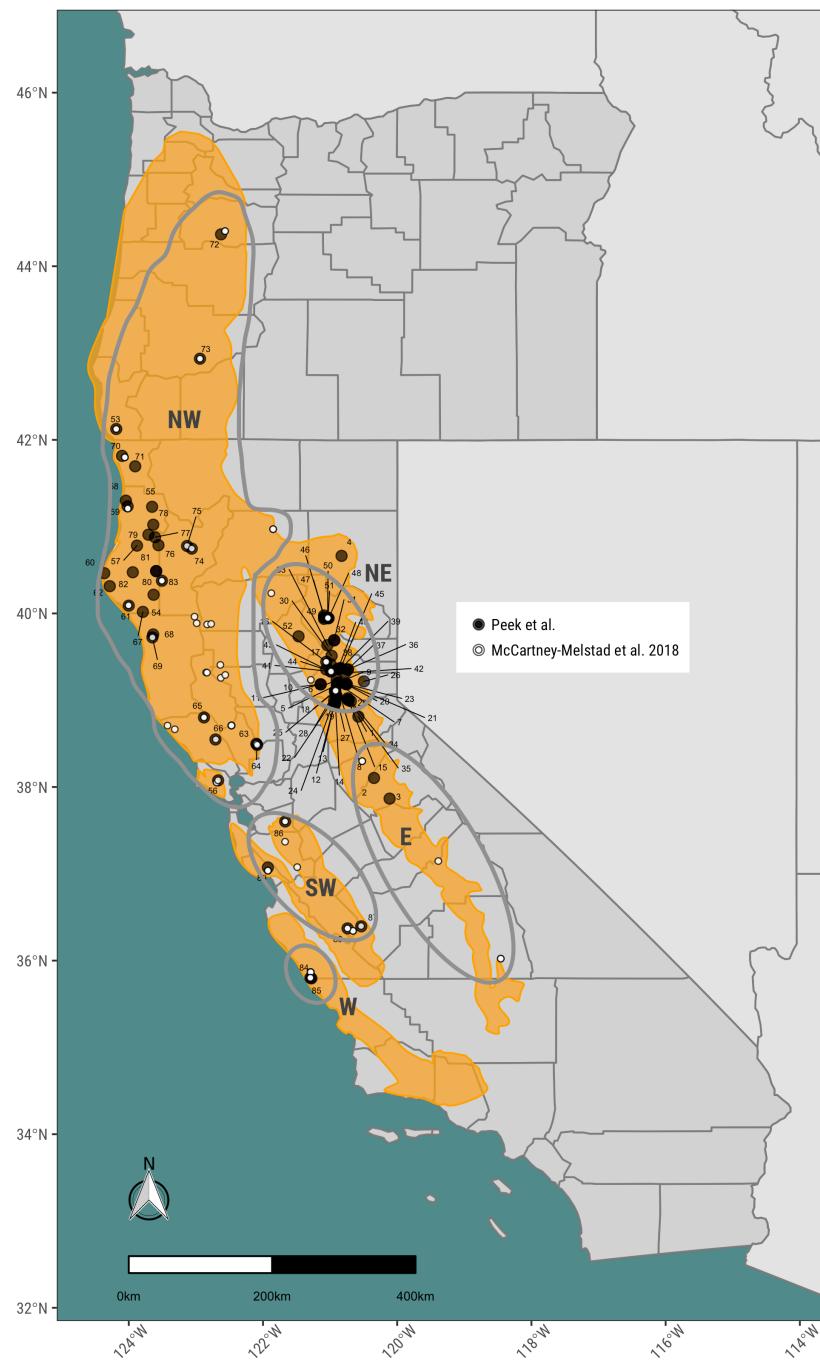


Figure 3.1: Map comparing sampling localities between the *R. boylii* study of McCartney-Melstad et al. (2018) and the current study with a simplified historical range (yellow) of *R. boylii* (adapted from USFS 2011 and Thomson et al. (2016)). Clades identified by McCartney-Melstad et al. (2018) are circled.

RESULTS

Table 3.1: Sampling Localities and clade groupings.

SiteID	Locality	River	Clade	Lat.	Lon.	HUC6	County	n
								samples
1	SFA-CAMI	SFA	Unknown	38.81151	-120.5787	180201	El Dorado	6
2	STAN-RoseCk	STAN	East	38.10420	-120.3460	180400	Tuolumne	2
3	TUO-Clavey	TUO	East	37.86643	-120.1134	180400	Tuolumne	4
4	ANTV-PineCk	ANTV	Unknown	40.66316	-120.8288	180800	Lassen	1
5	BEAR	BEAR	North-East	39.17484	-120.8998	180201	Nevada	6
6	BEAR-GRHN	BEAR	North-East	39.23206	-120.9019	180201	Nevada	6
7	BEAR-STH2	BEAR	North-East	39.19444	-120.8878	180201	Nevada	6
8	BEAR-STHA	BEAR	North-East	39.18833	-120.8981	180201	Nevada	3
9	BEAR-STHC	BEAR	North-East	39.20231	-120.8754	180201	Nevada	10
10	DEER-ClearCk	DEER	North-East	39.18287	-121.1410	180201	Nevada	3
11	DEER-CLEC	DEER	North-East	39.18287	-121.1410	180201	Nevada	2
12	MFA-AMEC	MFA	North-East	38.93396	-120.9436	180201	El Dorado	6
13	MFA-GASC	MFA	North-East	38.96651	-120.9325	180201	Placer	6
14	MFA-TODC	MFA	North-East	38.96385	-120.9216	180201	Placer	9
15	MFA-US-R	MFA	North-East	39.00754	-120.7316	180201	Placer	1
16	MFY-OREGCK	MFY	North-East	39.44188	-121.0575	180201	Nevada	10
17	MFY-US-OH	MFY	North-East	39.41305	-120.9903	180201	Nevada	10
18	NFA	NFA	North-East	39.10722	-120.9239	180201	Placer	10
19	NFA-BUNC	NFA	North-East	39.03762	-120.9103	180201	Placer	10
20	NFA-EUCHDS	NFA	North-East	39.18492	-120.7620	180201	Placer	5
21	NFA-EUCHUS	NFA	North-East	39.18444	-120.7505	180201	Placer	4
22	NFA-INDC	NFA	North-East	39.05665	-120.9085	180201	Placer	10
23	NFA-NFNFA	NFA	North-East	39.18839	-120.7592	180201	Placer	2
24	NFA-POND	NFA	North-East	38.99995	-120.9406	180201	Placer	5
25	NFA-ROBR	NFA	North-East	39.10451	-120.9267	180201	Placer	10
26	NFA-SAIC	NFA	North-East	39.21694	-120.4960	180201	Placer	5
27	NFA-SHIC	NFA	North-East	39.04073	-120.9014	180201	Placer	10
28	NFA-SLAR	NFA	North-East	39.09865	-120.9255	180201	Placer	8
29	NFMFA-SC	NFMFA	North-East	39.02237	-120.7369	180201	Placer	10
30	NFY	NFY	North-East	39.51190	-120.9774	180201	Sierra	10

RESULTS

Table 3.1: Sampling Localities and clade groupings. (*continued*)

SiteID	Locality	River	Clade	Lat.	Lon.	HUC6	County	n
								samples
31	NFY-SLATE	NFY	North-East	39.69219	-120.9399	180201	Plumas	1
32	NFY-SLATE-CGRev	NFY	North-East	39.68913	-120.9389	180201	Plumas	3
33	NFY-SLATE-Onion	NFY	North-East	39.63414	-121.0400	180201	Plumas	2
34	RUB-LC-US	RUB	North-East	38.98887	-120.6900	180201	Placer	8
35	RUB-USPH	RUB	North-East	38.99928	-120.7233	180201	Placer	10
36	SFY	SFY	North-East	39.35386	-120.7342	180201	Nevada	1
37	SFY-FallCk	SFY	North-East	39.35532	-120.7370	180201	Nevada	5
38	SFY-LOGA	SFY	North-East	39.36914	-120.8526	180201	Nevada	4
39	SFY-MCKI	SFY	North-East	39.36805	-120.8354	180201	Nevada	2
40	SFY-MISC	SFY	North-East	39.36096	-120.8814	180201	Nevada	6
41	SFY-RockCk	SFY	North-East	39.32983	-120.9863	180201	Nevada	3
42	SFY-Scotchman	SFY	North-East	39.31653	-120.7733	180201	Nevada	2
43	SFY-ShadyCk	SFY	North-East	39.35433	-121.0590	180201	Nevada	10
44	SFY-SpringCk	SFY	North-East	39.33233	-120.9890	180201	Nevada	3
45	SFY-THIMC	SFY	North-East	39.36465	-120.8462	180201	Nevada	1
46	FEA-BeanCk	FEA	North-East	39.97650	-121.0901	180201	Plumas	10
47	FEA-SPANISH-BGulch	FEA	North-East	39.95460	-121.0887	180201	Plumas	6
48	FEA-SPANISH-RockCk	FEA	North-East	39.94449	-121.0221	180201	Plumas	1
49	FEA-SPANISH-SilverCk	FEA	North-East	39.93733	-121.0904	180201	Plumas	4
50	FEA-SPANISH-Wapaunsie	FEA	North-East	39.95226	-121.0373	180201	Plumas	1
51	FEA-SpanishCk	FEA	North-East	39.94637	-121.0325	180201	Plumas	10
52	NFF-Poe	NFF	North-East	39.73598	-121.4702	180201	Butte	4
53	CHECO-NookCk	CHECO	North-West	42.12500	-124.1870	171003	Curry	1
54	EEL	EEL	North-West	40.21518	-123.6303	180101	Humboldt	10

RESULTS

Table 3.1: Sampling Localities and clade groupings. (*continued*)

SiteID	Locality	River	Clade	Lat.	Lon.	HUC6	County	n
								samples
55	KLAM-Aiken	KLAM	North-West	41.22842	-123.6524	180102	Humboldt	2
56	LAGUN-HalleckCk	LAGUN	North-West	38.07650	-122.6680	180500	Marin	1
57	MAD-MapleCk	MAD	North-West	40.78113	-123.8778	180101	Humboldt	8
58	MAD-RedwoodCk	MAD	North-West	41.29934	-124.0430	180101	Humboldt	9
59	MAD-RedwoodCk2	MAD	North-West	41.23420	-124.0171	180101	Humboldt	9
60	MAT-BearRiver	MAT	North-West	40.46383	-124.3652	180101	Humboldt	9
61	MAT-HeadwaterMattole	MAT	North-West	40.08992	-124.0003	180101	Humboldt	1
62	MAT-LowerMattole	MAT	North-West	40.31408	-124.2830	180101	Humboldt	5
63	PUT-COLD	PUT	North-West	38.49829	-122.0988	180201	Solano	9
64	PUT-WildhorseCk	PUT	North-West	38.48616	-122.0866	180201	Solano	10
65	RUSS-HUNTCK	RUSS	North-West	38.80166	-122.8793	180101	Sonoma	1
66	RUSS-MWSprngCk	RUSS	North-West	38.54757	-122.7086	180101	Sonoma	10
67	SFEEL	SFEEL	North-West	40.01786	-123.7896	180101	Humboldt	1
68	SFEEL-Cedar	SFEEL	North-West	39.75784	-123.6364	180101	Mendocino	10
69	SFEEL-Elder	SFEEL	North-West	39.72383	-123.6483	180101	Mendocino	2
70	SMITH	SMITH	North-West	41.81553	-124.1003	180101	Del Norte	1
71	SMITH-HurdygurdyCk	SMITH	North-West	41.69500	-123.9043	180101	Del Norte	10
72	SSANTIAM	SSANTIAM	North-West	44.36750	-122.6250	170900	Linn	7
73	SUMPQUA	SUMPQUA	North-West	42.93500	-122.9390	171003	Douglas	2
74	TRIN-ConnerCk	TRIN	North-West	40.74656	-123.0607	180102	Trinity	1
75	TRIN-NFTrinity	TRIN	North-West	40.77713	-123.1318	180102	Trinity	1
76	TRIN-SFTrinity	TRIN	North-West	40.78619	-123.5575	180102	Humboldt	6
77	TRIN-SFTrinity-SandyBar	TRIN	North-West	40.87797	-123.6088	180102	Humboldt	9
78	TRIN-TishTang	TRIN	North-West	41.02208	-123.6351	180102	Humboldt	9

RESULTS

Table 3.1: Sampling Localities and clade groupings. (*continued*)

SiteID	Locality	River	Clade	Lat.	Lon.	HUC6	County	n
								samples
79	TRIN-WillowCk	TRIN	North-West	40.90705	-123.7073	180102	Humboldt	1
80	VANDZ-Dinsmore	VANDZ	North-West	40.48759	-123.5918	180101	Humboldt	9
81	VANDZ-Mill	VANDZ	North-West	40.48759	-123.5918	180101	Humboldt	1
82	VANDZ-RootCreek	VANDZ	North-West	40.47422	-123.9400	180101	Humboldt	4
83	VANDZ-Shanty	VANDZ	North-West	40.37670	-123.5061	180101	Trinity	3
84	SANCARP-DutraCk	SANCARP	South-West	35.80217	-121.2932	180600	Monterey	10
85	SANCARP-SanCarpoforoCk	SANCARP	South-West	35.79533	-121.2787	180600	San Luis Obispo	2
86	ALA-ArroyoMocho	ALA	West	37.60299	-121.6691	180500	Alameda	5
87	DRY-Cantua_AL	DRY	West	36.39510	-120.5359	180300	Fresno	1
88	PAJ-ClearCk	PAJ	West	36.37111	-120.7400	180600	San Benito	3
89	SOQUEL-EBSC	SOQUEL	West	37.07300	-121.9270	180600	Santa Cruz	10

PCA and Admixture shows strong separation between California ecoregions

To compare and assess population structure patterns across the range of *R. boylii*, PCA were used to provide a dimensionless comparison of genetic variation. The patterns of genetic differentiation largely conformed to the clades described by McCartney-Melstad et al. (2018) (North-West [NW], South-West [SW], West [W], North-East[NE], and East [E]), but some notable differences were observed. We used regional names for clades, thus the North Coast=NW, the Central Coast=SW, the South Coast=W, the

RESULTS

Northern Sierra=NE, the Southern Sierra=E (Figure 3.2). The main difference evident in the PCA was the presence of a distinct group previously part of the Northern Sierra, comprised of samples from the Feather basin. Additional subdivision was evident in the Central/Southern coastal region (two groups which match SW and W clades), and the Sierra Nevada (two groups, matching the NE and E clades), yielding seven total distinct structured groups. Interestingly, a small cluster of individuals from the North Fork (NF) Feather River consistently clustered along intermediate axes between the North Coast and Northern Sierra/Southern Sierra groups. These individuals were consistently intergrades between the larger Feather group and the Northern Coastal group, regardless of the PC axes.

A single sampling locality from Lassen National Forest near Eagle Lake (Locality 4, Table 3.1) clustered with the Northern Sierra group, suggesting the geographic boundary for this genetic group should be extended (Figure 3.2). Furthermore, individuals from the South Fork (SF) American basin clustered with the Southern Sierra (East) group, which would extend the boundary delineated by McCartney-Melstad et al. (2018).

To assess how strongly the Feather and Southern/Central coastal samples may have affected PCA structure, we ran several additional separate post-hoc PCA analyses where all Feather samples were excluded, and all Southern/Central coastal samples were excluded. Patterns remained consistent with those observed when all samples were used (Figure 3.2), regardless of which samples were excluded or retained. We conclude there is significant congruence with the McCartney-Melstad et al. (2018) genetic clades, but an additional group appears distinct from the original five delineated, the Feather watershed, and boundaries should be expanded to include the SF American basin in the Southern Sierra, and the Honey-Eagle Lakes basin in the Northern Sierra.

RESULTS

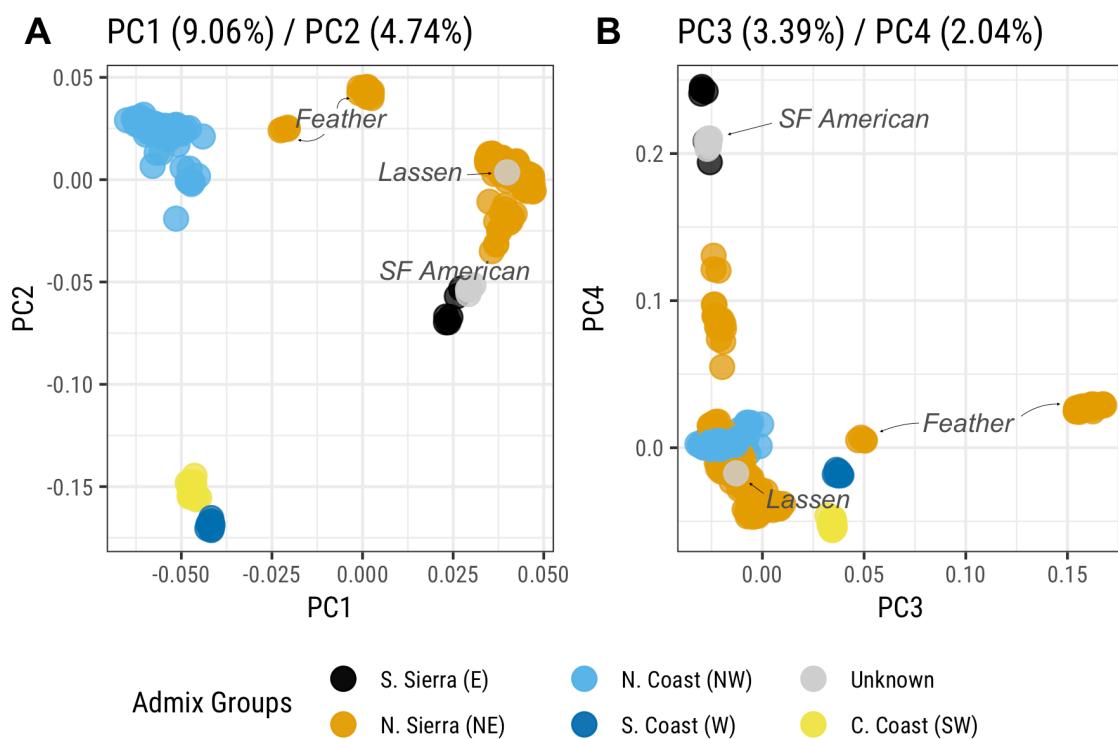


Figure 3.2: Principal component analysis of Rapture data colored by genetic groups (in parentheses) from McCartney-Melstad et al. (2018). A) PC1 vs. PC2; B) PC3 vs. PC4. Note, samples from South Fork American River and Lassen County (gray) were not previously studied and thus had no assigned clade. *Rana boylii* in the Feather basin were assigned as part of the Northern Sierra (NE) clade by McCartney-Melstad et al. (2018).

RESULTS

To further evaluate population structure across the range of *R. boylii*, we used NGSAdmix estimate the proportion of coancestry among individuals (admixture) from genome-wide SNPs. We used a range of k-values to evaluate the number of potential groups based on genetic ancestry and found strong patterns of divergence which largely matched the patterns observed in the PCA. While the northern and central coastal (SW and W) groups showed strong divergence from all other groups, they did not separate into distinct groups until k=9 (Figure 3.3). At k=3 through k=8, the N. Sierra and N. Coast groups continued to show patterns of subdivision, while other groups like the N. Sierra-Feather remained strongly distinct with little or no admixture observed. At k=9, the Central Coast and South Coast split into distinct groups, with additional patterns of substructure and admixture observed in the N. Sierrae (NE) and the N. Coast (NW) groups (Figure 3). While the Central/Southern Coastal populations are clearly distinct from all other clades, our NGSadmix analysis of our data shows weaker support for delineating Central (SW) and Southern (W) groups as distinct.

Sampling localities were then updated with genetic groups identified from admixture and PCA analyses to delineate geographic boundaries for these clades, building on McCartney-Melstad et al. (2018) (Figure 3.4). In summary, we identified *R. boylii* from the Feather Basin as a unique genetic group, the Southern Sierra (East) clade should therefore extend to include the South Fork American basin in El Dorado County (Locality 1), and the Northern Sierra (North-East) clade should be expanded to include Honey-Eagle Lakes basin in Lassen County (Locality 4, see Figure 3.1 and Figure 3.4).

RESULTS

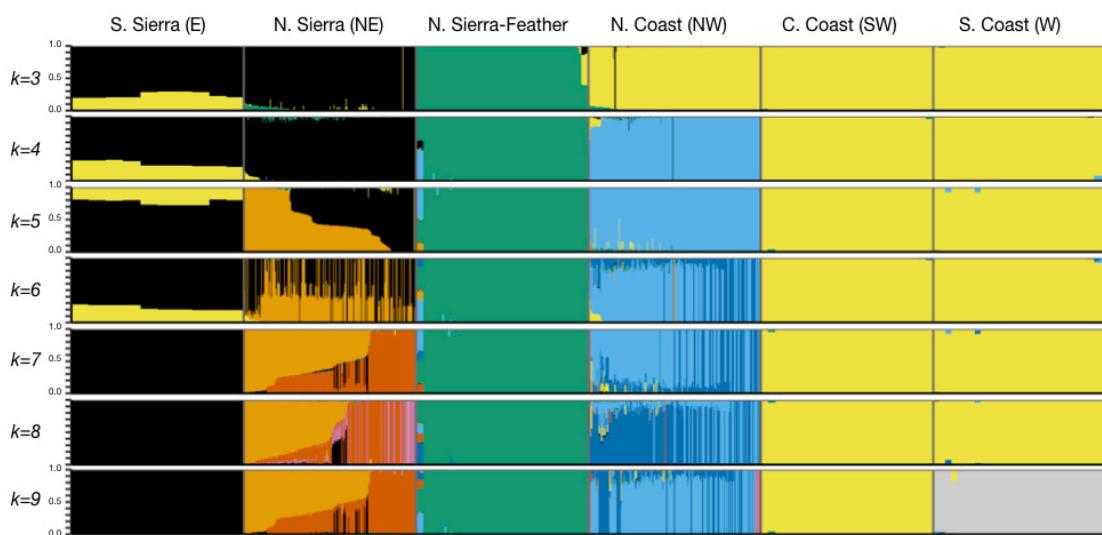


Figure 3.3: NGSAdmixture for $k=3$ through $k=9$ for *R. boylii* across major geographic groupings (adapted from McCartney-Melstad et al. (2018)). The two Central/Southern Coastal groups did not differentiate until a $k=9$.

RESULTS

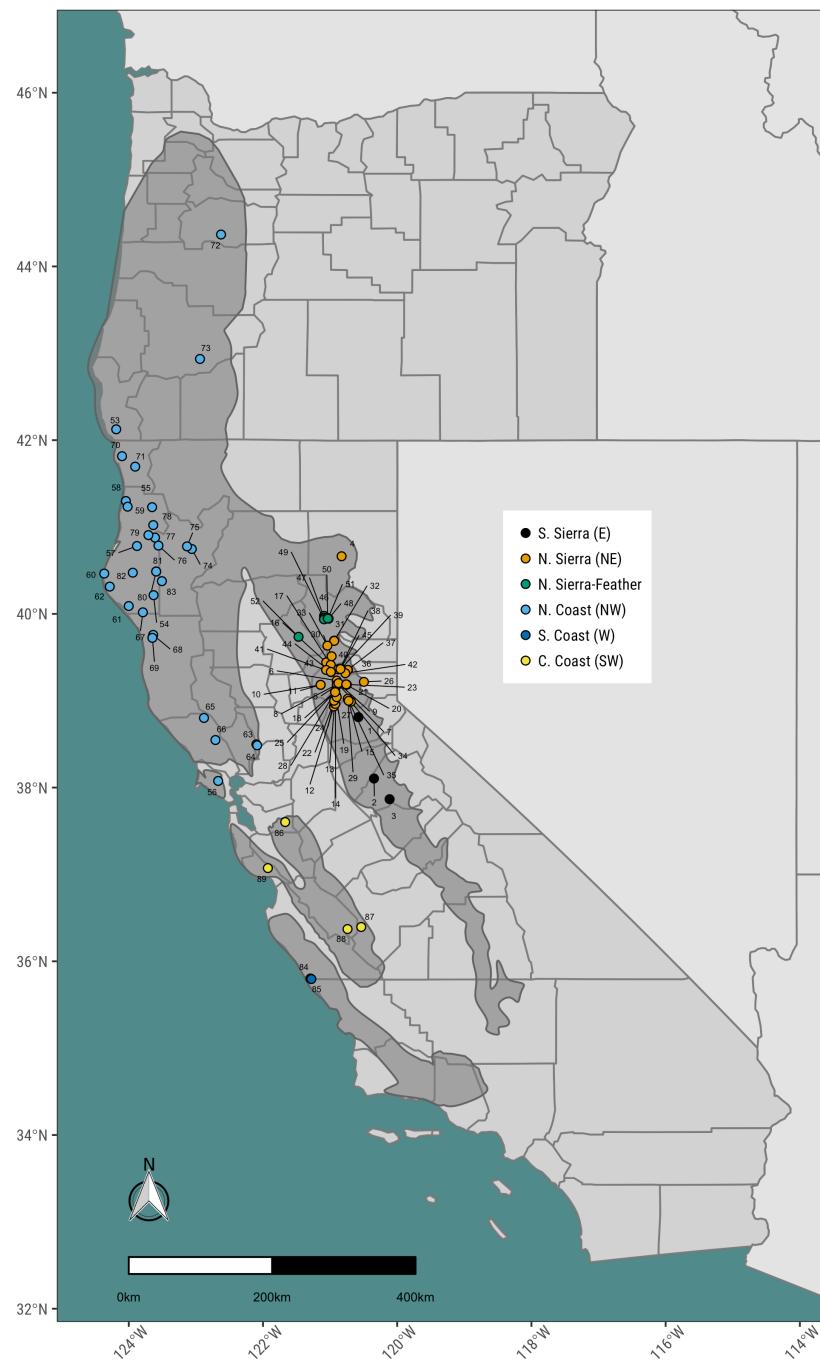


Figure 3.4: Map of localities colored by genetic groupings. The Northern Sierra-Feather group adds an additional genetically distinct group, the Feather Basin to previous clades; Locality 1 extends the boundary for the Southern Sierra (E) genetic clade up to the SF American basin in El Dorado County, and Locality 4 extends the Northern Sierra (NE) clade into Lassen County.

RESULTS

F_{ST} shows strong divergence across regions

To assess how distinct *R. boylii* populations were across the range of the species using F_{ST} and genetic diversity metrics, we selected all samples that had greater than 100,000 alignments from localities with at least three individuals, yielding a total of 630 individuals from 63 localities for analysis (Table 3.2). Pairwise F_{ST} was calculated between each population pair (total combinations = 1,953), values ranged from 0 to 0.646. Extremely high F_{ST} values were observed between Central/South Coastal (SW/W) localities and all other localities from other regions, indicating the coastal regions are strongly divergent from *R. boylii* in other parts of the species' range (Figure 3.5A). Comparison of pairwise F_{ST} values against geographic distance showed higher F_{ST} values between clades than within. Furthermore, the largest geographic distances between locality pairs (i.e., between Southern Coastal localities and Northern Coastal localities in Oregon) correlated with higher F_{ST} values. However, the highest F_{ST} values were not associated with the greatest geographic distances, as the Sierra (NE/E) and Southern Coast (SW/W) comparisons had the greatest pairwise F_{ST} values (> 0.6) of any pairwise comparison (Figure 3.5).

Table 3.2: Sampling localities with at least 3 individuals, used in population analyses of genetic diversity

SiteID	Locality	River	Clade	Lat.	Lon.	HUC6	County	n samples
1	SFA-CAMI	SFA	Unknown	38.81151	-120.5787	180201	El Dorado	6
2	STAN-RoseCk	STAN	East	38.10420	-120.3460	180400	Tuolumne	2
3	TUO-Clavey	TUO	East	37.86643	-120.1134	180400	Tuolumne	4
4	ANTV-PineCk	ANTV	Unknown	40.66316	-120.8288	180800	Lassen	1
5	BEAR	BEAR	North-East	39.17484	-120.8998	180201	Nevada	6
6	BEAR-GRHN	BEAR	North-East	39.23206	-120.9019	180201	Nevada	6
7	BEAR-STH2	BEAR	North-East	39.19444	-120.8878	180201	Nevada	6
8	BEAR-STHA	BEAR	North-East	39.18833	-120.8981	180201	Nevada	3

RESULTS

Table 3.2: Sampling localities with at least 3 individuals, used in population analyses of genetic diversity (*continued*)

SiteID	Locality	River	Clade	Lat.	Lon.	HUC6	County	n samples
9	BEAR-STHC	BEAR	North-East	39.20231	-120.8754	180201	Nevada	10
10	DEER-ClearCk	DEER	North-East	39.18287	-121.1410	180201	Nevada	3
11	DEER-CLEC	DEER	North-East	39.18287	-121.1410	180201	Nevada	2
12	MFA-AMEC	MFA	North-East	38.93396	-120.9436	180201	El Dorado	6
13	MFA-GASC	MFA	North-East	38.96651	-120.9325	180201	Placer	6
14	MFA-TODC	MFA	North-East	38.96385	-120.9216	180201	Placer	9
15	MFA-US-R	MFA	North-East	39.00754	-120.7316	180201	Placer	1
16	MFY-OREGCK	MFY	North-East	39.44188	-121.0575	180201	Nevada	10
17	MFY-US-OH	MFY	North-East	39.41305	-120.9903	180201	Nevada	10
18	NFA	NFA	North-East	39.10722	-120.9239	180201	Placer	10
19	NFA-BUNC	NFA	North-East	39.03762	-120.9103	180201	Placer	10
20	NFA-EUCHDS	NFA	North-East	39.18492	-120.7620	180201	Placer	5
21	NFA-EUCHUS	NFA	North-East	39.18444	-120.7505	180201	Placer	4
22	NFA-INDC	NFA	North-East	39.05665	-120.9085	180201	Placer	10
23	NFA-NFNFA	NFA	North-East	39.18839	-120.7592	180201	Placer	2
24	NFA-POND	NFA	North-East	38.99995	-120.9406	180201	Placer	5
25	NFA-ROBR	NFA	North-East	39.10451	-120.9267	180201	Placer	10
26	NFA-SAIC	NFA	North-East	39.21694	-120.4960	180201	Placer	5
27	NFA-SHIC	NFA	North-East	39.04073	-120.9014	180201	Placer	10
28	NFA-SLAR	NFA	North-East	39.09865	-120.9255	180201	Placer	8
29	NFMFA-SC	NFMFA	North-East	39.02237	-120.7369	180201	Placer	10
30	NFY	NFY	North-East	39.51190	-120.9774	180201	Sierra	10
31	NFY-SLATE	NFY	North-East	39.69219	-120.9399	180201	Plumas	1
32	NFY-SLATE-	NFY	North-East	39.68913	-120.9389	180201	Plumas	3
	CGRav							
33	NFY-SLATE-	NFY	North-East	39.63414	-121.0400	180201	Plumas	2
	Onion							
34	RUB-LC-US	RUB	North-East	38.98887	-120.6900	180201	Placer	8
35	RUB-USPH	RUB	North-East	38.99928	-120.7233	180201	Placer	10
36	SFY	SFY	North-East	39.35386	-120.7342	180201	Nevada	1

RESULTS

Table 3.2: Sampling localities with at least 3 individuals, used in population analyses of genetic diversity (*continued*)

SiteID	Locality	River	Clade	Lat.	Lon.	HUC6	County	n samples
37	SFY-FallCk	SFY	North-East	39.35532	-120.7370	180201	Nevada	5
38	SFY-LOGA	SFY	North-East	39.36914	-120.8526	180201	Nevada	4
39	SFY-MCKI	SFY	North-East	39.36805	-120.8354	180201	Nevada	2
40	SFY-MISC	SFY	North-East	39.36096	-120.8814	180201	Nevada	6
41	SFY-RockCk	SFY	North-East	39.32983	-120.9863	180201	Nevada	3
42	SFY-Scotchman	SFY	North-East	39.31653	-120.7733	180201	Nevada	2
43	SFY-ShadyCk	SFY	North-East	39.35433	-121.0590	180201	Nevada	10
44	SFY-SpringCk	SFY	North-East	39.33233	-120.9890	180201	Nevada	3
45	SFY-THIMC	SFY	North-East	39.36465	-120.8462	180201	Nevada	1
46	FEA-BeanCk	FEA	North-East	39.97650	-121.0901	180201	Plumas	10
47	FEA-SPANISH-BGulch	FEA	North-East	39.95460	-121.0887	180201	Plumas	6
48	FEA-SPANISH-RockCk	FEA	North-East	39.94449	-121.0221	180201	Plumas	1
49	FEA-SPANISH-SilverCk	FEA	North-East	39.93733	-121.0904	180201	Plumas	4
50	FEA-SPANISH-Wapaunsie	FEA	North-East	39.95226	-121.0373	180201	Plumas	1
51	FEA-SpanishCk	FEA	North-East	39.94637	-121.0325	180201	Plumas	10
52	NFF-Poe	NFF	North-East	39.73598	-121.4702	180201	Butte	4
53	CHETCO-NookCk	CHETCO	North-West	42.12500	-124.1870	171003	Curry	1
54	EEL	EEL	North-West	40.21518	-123.6303	180101	Humboldt	10
55	KLAM-Aiken	KLAM	North-West	41.22842	-123.6524	180102	Humboldt	2
56	LAGUN-HalleckCk	LAGUN	North-West	38.07650	-122.6680	180500	Marin	1
57	MAD-MapleCk	MAD	North-West	40.78113	-123.8778	180101	Humboldt	8
58	MAD-RedwoodCk	MAD	North-West	41.29934	-124.0430	180101	Humboldt	9
59	MAD-RedwoodCk2	MAD	North-West	41.23420	-124.0171	180101	Humboldt	9
60	MAT-BearRiver	MAT	North-West	40.46383	-124.3652	180101	Humboldt	9

RESULTS

Table 3.2: Sampling localities with at least 3 individuals, used in population analyses of genetic diversity (*continued*)

SiteID	Locality	River	Clade	Lat.	Lon.	HUC6	County	n samples
61	MAT- HeadwaterMattole	MAT	North-West	40.08992	-124.0003	180101	Humboldt	1
62	MAT- LowerMattole	MAT	North-West	40.31408	-124.2830	180101	Humboldt	5
63	PUT-COLD	PUT	North-West	38.49829	-122.0988	180201	Solano	9
64	PUT-WildhorseCk	PUT	North-West	38.48616	-122.0866	180201	Solano	10
65	RUSS-HUNTCK	RUSS	North-West	38.80166	-122.8793	180101	Sonoma	1
66	RUSS- MWSprngCk	RUSS	North-West	38.54757	-122.7086	180101	Sonoma	10
67	SFEEL	SFEEL	North-West	40.01786	-123.7896	180101	Humboldt	1
68	SFEEL-Cedar	SFEEL	North-West	39.75784	-123.6364	180101	Mendocino	10
69	SFEEL-Elder	SFEEL	North-West	39.72383	-123.6483	180101	Mendocino	2
70	SMITH	SMITH	North-West	41.81553	-124.1003	180101	Del Norte	1
71	SMITH- HurdygurdyCk	SMITH	North-West	41.69500	-123.9043	180101	Del Norte	10
72	SSANTIAM	SSANTIAM	North-West	44.36750	-122.6250	170900	Linn	7
73	SUMPQUA	SUMPQUA	North-West	42.93500	-122.9390	171003	Douglas	2
74	TRIN-ConnerCk	TRIN	North-West	40.74656	-123.0607	180102	Trinity	1
75	TRIN-NFTrinity	TRIN	North-West	40.77713	-123.1318	180102	Trinity	1
76	TRIN-SFTrinity	TRIN	North-West	40.78619	-123.5575	180102	Humboldt	6
77	TRIN-SFTrinity- SandyBar	TRIN	North-West	40.87797	-123.6088	180102	Humboldt	9
78	TRIN-TishTang	TRIN	North-West	41.02208	-123.6351	180102	Humboldt	9
79	TRIN-WillowCk	TRIN	North-West	40.90705	-123.7073	180102	Humboldt	1
80	VANDZ-Dinsmore	VANDZ	North-West	40.48759	-123.5918	180101	Humboldt	9
81	VANDZ-Mill	VANDZ	North-West	40.48759	-123.5918	180101	Humboldt	1
82	VANDZ- RootCreek	VANDZ	North-West	40.47422	-123.9400	180101	Humboldt	4
83	VANDZ-Shanty	VANDZ	North-West	40.37670	-123.5061	180101	Trinity	3

RESULTS

Table 3.2: Sampling localities with at least 3 individuals, used in population analyses of genetic diversity (*continued*)

SiteID	Locality	River	Clade	Lat.	Lon.	HUC6	County	n samples
84	SANCARP-DutraCk	SANCARP	South-West	35.80217	-121.2932	180600	Monterey	10
85	SANCARP-SanCarpoforoCk	SANCARP	South-West	35.79533	-121.2787	180600	San Luis Obispo	2
86	ALA-ArroyoMocho	ALA	West	37.60299	-121.6691	180500	Alameda	5
87	DRY-Cantua_AL	DRY	West	36.39510	-120.5359	180300	Fresno	1
88	PAJ-ClearCk	PAJ	West	36.37111	-120.7400	180600	San Benito	3
89	SOQUEL-EBSC	SOQUEL	West	37.07300	-121.9270	180600	Santa Cruz	10

RESULTS

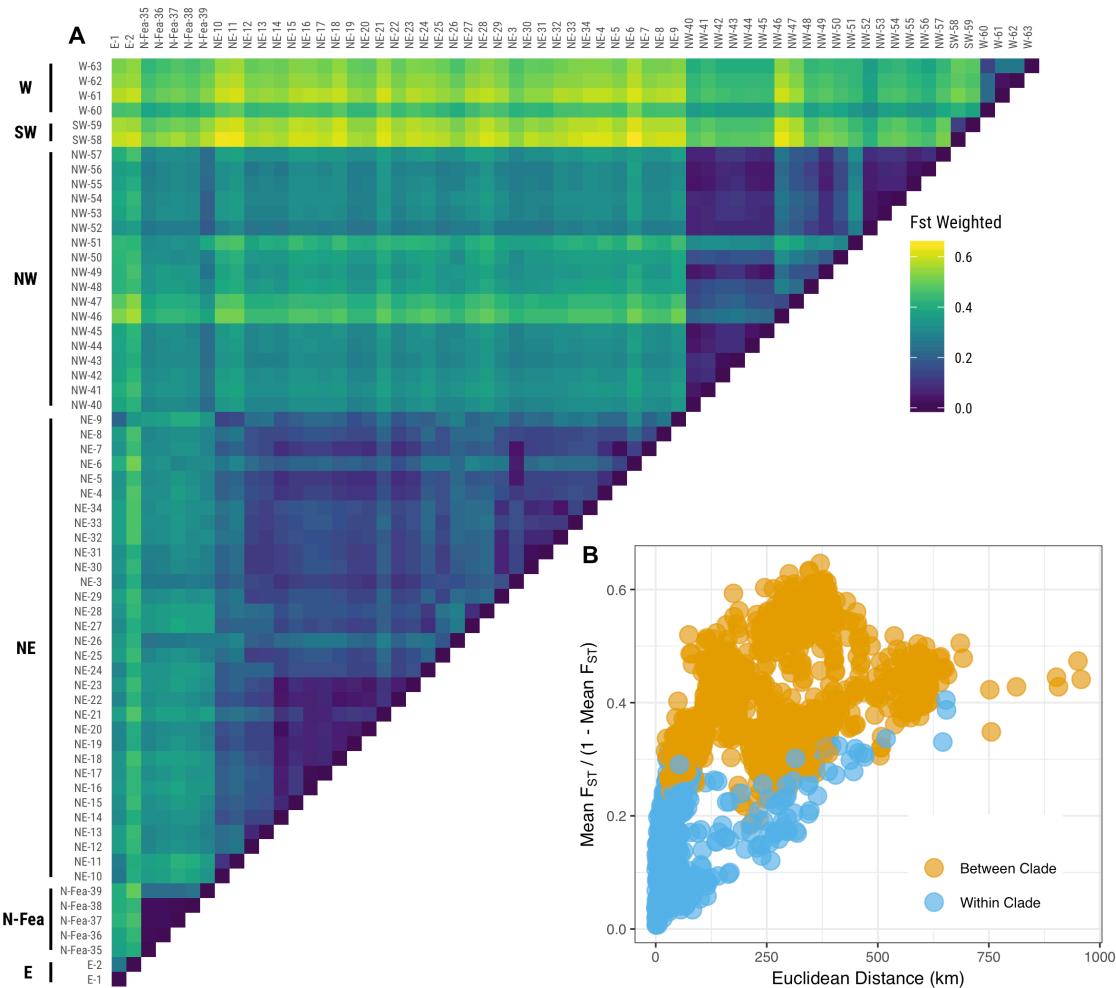


Figure 3.5: Plots of F_{ST} for all localities sampled, A) comparison of pairwise F_{ST} for all localities grouped by *R. boylii* clades; and B) pairwise F_{ST} comparisons between and within clades vs. geographic distance between localities.

RESULTS

Genetic variation is highest in the North-West and lowest in the East/ South-West clades

Genetic diversity θ estimates for each region showed *R. boylii* from the North Coast (NW) group had the greatest range in genetic variation, with both the most diverse and least diverse θ estimates (Figure 3.6). Populations from the Southern Sierra (E) and Central Coast (SW) had the lowest mean Tajima's θ_π (0.0064) and Watterson's θ_W (0.0059 = E, 0.0060 = SW), but both these groups had the lowest number of sampling localities. The $\Delta\theta$ was positive across all groups except the North Coast, which exhibited the lowest value, while the Central Coast (SW) and Southern Sierra (E) estimates were highest. Both coastal groups (SW and W) had wide ranges of $\Delta\theta$ estimates in different locations. With the exception of the North Coast, which contains the greatest genetic diversity and a trajectory of increasing genetic diversity, overall genetic diversity data for *R. boylii* show a trajectory of genetic diversity loss, with many locations exhibiting positive $\Delta\theta$ (Figure 3.6D).

RESULTS

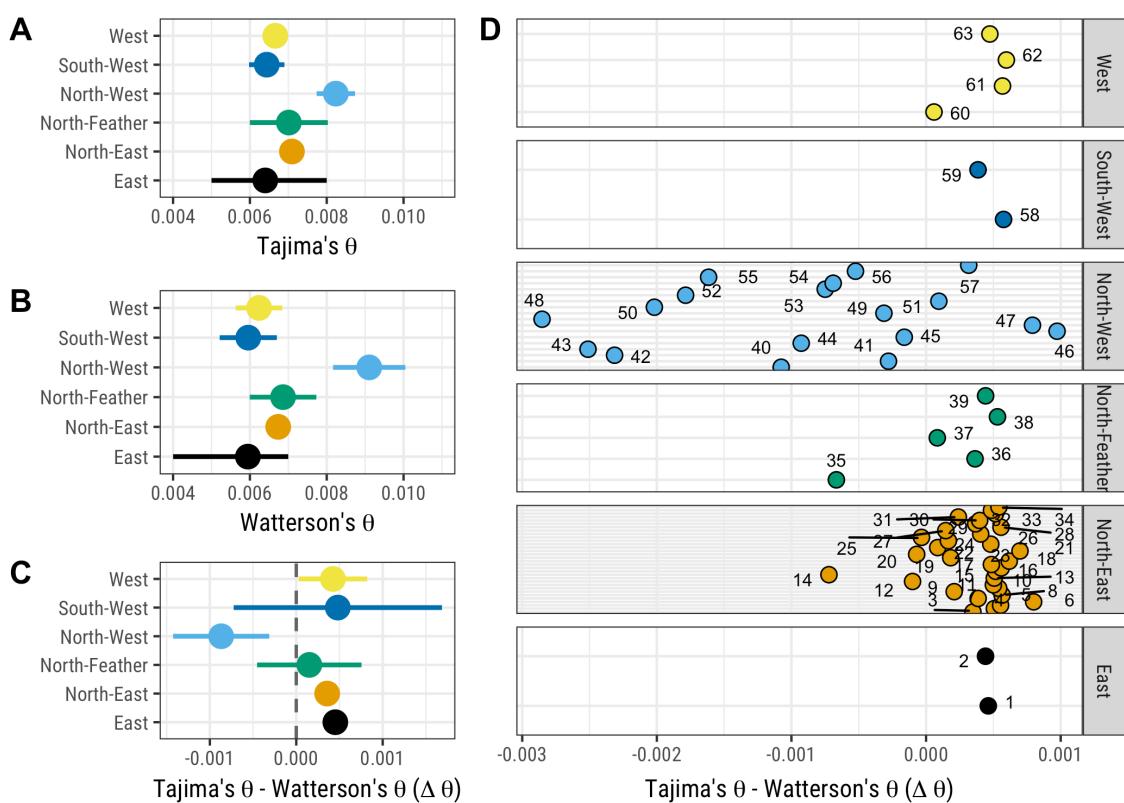


Figure 3.6: Tajima's θ_π and Watterson's θ_S diversity estimates for *R. boylii* by clade groups (adapted from McCartney-Melstad et al. (2018)) with 95% CI. A) Tajima's θ_π by clade; B) Watterson's θ_S by clade; C) $\theta_\pi - \theta_S$ aggregated by clade and D) for each locality within clades.

DISCUSSION AND CONCLUSION

Discussion and Conclusion

As a species with a broad geographic range, *Rana boylii* utilizes a wide array of different landscapes throughout California and Oregon. However, our study indicates there are strong patterns of geographic structure and substantial differences in patterns of genetic diversity across the species' range. The more comprehensive sampling provided by this data set complements a recent genomic study of *R. boylii* (McCartney-Melstad et al. 2018) and largely supports the same clades identified using a different RADSeq method. Moreover, our PCA and admixture results are congruent with the five major clades identified in McCartney-Melstad et al. (2018), with the exception of an additional novel group from the Feather watershed. This congruence indicates these clades are important units which should be considered carefully in planning conservation actions (i.e., such as translocation, recolonization, and habitat restoration) for the species.

We also show that genetic data can provide resolution to prioritize populations and regions at multiple scales. Evaluating the efficacy of potential restoration actions can be difficult, particularly with rare and cryptic species. Genetics provides a tool that can more effectively track changes in population connectivity or genetic trajectories (i.e., low diversity, high F_{ST}) which correspond to the potential for long-term population persistence. Populations in the Northern Coastal clade contained the greatest amount of genetic variation, yet pairwise comparisons with populations from Central and Southern Coast localities showed extremely high F_{ST} values, indicating these groups are genetically very distinct. In addition, the Central/Southern Coast groups have relatively low genetic diversity and trajectories of genetic diversity loss.

Recent sampling has largely occurred in the North Coast, Northern Sierra, and Northern Sierra-Feather localities. The limited number of sampling localities from the south-

DISCUSSION AND CONCLUSION

ern extent of the range of *R. boylii* correlates with population declines (Davidson 2004, Adams et al. 2017). Thus, conservation actions should prioritize protecting sites in the Central/Southern Coast and Southern Sierra clades, as these represent both the most genetically divergent and distinct *R. boylii*, and simultaneously the most at-risk populations. Extremely high F_{ST} values between populations from the same species as well as strong phylogeographic structuring suggest limited gene flow through biogeographic barriers (Spinks et al. 2010, Voelker et al. 2013, Cazé et al. 2016). This suggests conservation management should consider these clades as distinct, and future actions that include translocation or captive breeding should utilize genetic data to assess potential impacts such as outbreeding depression in locally adapted populations.

R. boylii is an important sentinel stream species—not only as link between both aquatic and riparian ecosystems—but also as a sensitive gage of how altered hydrology affects riverine environments. The impacts of regulated hydrology on *R. boylii* populations have been well documented (Kupferberg 1996, Lind et al. 1996, Yarnell et al. 2012, Kupferberg et al. 2012, Catenazzi and Kupferberg 2013). For populations in regulated streams, genetic monitoring may be a crucial component for assessing flow management actions on the probability of long-term persistence (via maintenance of genetic diversity), as well as tracking population responses (i.e., using metrics like F_{ST} and genetic diversity θ) to restoration actions.

Acknowledgements

Many thanks to all who helped collect/prepare samples: Corey Luna, Rick Wachs, and Sarah Musselman. This research could not have been conducted without access to specimens from Brad Shaffer at the Krebb Museum at UCLA, and field samples from Caren Goldberg and Mallory Bedwell at Washington State University.

Appendix A

Appendix: Supplemental Tables & Code

This includes repository links to supplementary tables and code for analyses and figure generation.

Supplemental Tables

- Table S1: [Chapter 1 Rapture Samples](#) can be found at:
 - github.com/peek_phd_supplemental_materials
- Table S2: [RADSeq samples](#) used for designing baits can be found at:
 - github.com/peek_phd_supplemental_materials
- Table S3: [de Novo](#) alignment for *R. boylii*. Zipped file is 4.16 MB and can be found at:
 - github.com/peek_phd_supplemental_materials

APPENDIX A. APPENDIX: SUPPLEMENTAL TABLES & CODE

- Table S4: RAD Capture baits can be found at:
 - github.com/peek_phd_supplemental_materials
- Table S5: Chapter 2 Hybrid Samples can be found at:
 - github.com/peek_phd_supplemental_materials

Code

The pipeline for analyses and associated code can be found at:

- **Miller Lab:** <https://github.com/miller-genetics-lab>
- <https://ryanpeek.github.io/radseq/>

Colophon

This document is set in EB Garamond, Source Code Pro and Lato. The body text is set at 11pt with *lmr*.

It was written in R Markdown and *LATEX*, and rendered into PDF using `ucd-aggiedown` and `bookdown`, largely based on the `huskydown` template.

This document was typeset using the XeTeX typesetting system, and the University of California Thesis class (2012). The template meets UC Davis Graduate Studies dissertation requirements as of Fall 2018 to ensure that documents conform precisely to submission standards. Other elements of the document formatting source code have been taken from the Latex, Knitr, and RMarkdown templates for UC Berkeley's graduate thesis, and Dissertate: a LaTeX dissertation template to support the production and typesetting of a PhD dissertation at Harvard, Princeton, and NYU.

The source files for this thesis, along with data files are available at https://github.com/ryanpeek/peek_phd_supplemental_materials.

This version of the thesis was generated on 2018-09-17 16:58:49.

References

- Abbott, R., D. Albach, S. Ansell, J. W. Arntzen, S. J. E. Baird, N. Bierne, J. Boughman, A. Brelsford, C. A. Buerkle, R. Buggs, R. K. Butlin, U. Dieckmann, F. Eroukhmanoff, A. Grill, S. H. Cahan, J. S. Hermansen, G. Hewitt, A. G. Hudson, C. Jiggins, J. Jones, B. Keller, T. Marczewski, J. Mallet, P. Martinez-Rodriguez, M. Möst, S. Mullen, R. Nichols, A. W. Nolte, C. Parisod, K. Pfennig, A. M. Rice, M. G. Ritchie, B. Seifert, C. M. Smadja, R. Stelkens, J. M. Szymura, R. Väinölä, J. B. W. Wolf, and D. Zinner. 2013. Hybridization and speciation. *Journal of evolutionary biology* 26:229–246.
- Adams, A. J., A. P. Pessier, and C. J. Briggs. 2017. Rapid extirpation of a North American frog coincides with an increase in fungal pathogen prevalence: Historical analysis and implications for reintroduction. *Ecology and evolution* 7:10216–10232.
- Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 in B. N. Petrov and F. Csaki, editors. *Proceedings of the Second International Symposium on Information Theory*.
- Ali, O. A., S. M. O'Rourke, S. J. Amish, M. H. Meek, G. Luikart, C. Jeffres, and M. R. Miller. 2016. RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics* 202:389–400.
- Anderson, J. J., and W. N. Beer. 2009. Oceanic, riverine, and genetic influences on

References

- spring chinook salmon migration timing. Ecological applications: a publication of the Ecological Society of America 19:1989–2003.
- Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS one 3:e3376.
- Baird, S. F. 1856. Descriptions of new genera and species of North American Frogs. Pages 59–62 *in* Proceedings of the Academy of Natural Sciences of Philadelphia. Philadelphia, Academy of Natural Sciences of Philadelphia.
- Barbosa, S., F. Mestre, T. A. White, J. Paupério, P. C. Alves, and J. B. Searle. 2018. Integrative approaches to guide conservation decisions: Using genomics to define conservation units and functional corridors. Molecular ecology.
- Barrera-Guzmán, A. O., A. Aleixo, M. D. Shawkey, and J. T. Weir. 2018. Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. Proceedings of the National Academy of Sciences of the United States of America 115:E218–E225.
- Beebee, T. J. C., and R. A. Griffiths. 2005. The amphibian decline crisis: A watershed for conservation biology? Biological conservation 125:271–285.
- Birkeland, P. W. 1964. Pleistocene Glaciation of the Northern Sierra Nevada, North of Lake Tahoe, California. The Journal of geology 72:810–825.
- Bondi, C. A., S. M. Yarnell, A. J. Lind, and A. Lind. 2013. Transferability of habitat suitability criteria for a stream breeding frog (*Rana Boylii*) in the Sierra Nevada, California. Conservation and Biology 8:88–103.
- Botero, C. A., F. J. Weissing, J. Wright, and D. R. Rubenstein. 2015. Evolutionary

References

- tipping points in the capacity to adapt to environmental change. *Proceedings of the National Academy of Sciences of the United States of America* 112:184–189.
- Bourque, R. M. 2008, December. Spatial ecology of an inland population of the Foothill yellow-legged frog (*Rana boylii*) in Tehama County, California. PhD Thesis, Humboldt State University.
- Broquet, T., L. Berset-Braendli, G. Emaresi, and L. Fumagalli. 2007. Buccal swabs allow efficient and reliable microsatellite genotyping in amphibians. *Conservation Genetics* 8:509–511.
- Brown, L. R., J. T. May, A. C. Rehn, P. R. Ode, I. R. Waite, and J. G. Kennen. 2012. Predicting biological condition in southern California streams. *Landscape and urban planning* 108:17–27.
- Bunn, S. E., and A. H. Arthington. 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. *Environmental management* 30:492–507.
- Burkey, T. V. 1989. Extinction in Nature Reserves - the Effect of Fragmentation and the Importance of Migration between Reserve Fragments. *Oikos* 55:75–81.
- Camp, C. L. 1917. Notes on the systematic status of the toads and frogs of California. *University of California publications in zoology* 17:115–125.
- Catenazzi, A., and S. J. Kupferberg. 2013. The importance of thermal conditions to recruitment success in stream-breeding frog populations distributed across a productivity gradient. *Biological Conservation* 168:40–48.
- Cazé, A. L. R., G. Mäder, T. S. Nunes, L. P. Queiroz, G. de Oliveira, J. A. F. Diniz-Filho, S. L. Bonatto, and L. B. Freitas. 2016. Could refuge theory and rivers acting as barriers explain the genetic variability distribution in the Atlantic Forest? *Molecular*

References

- phylogenetics and evolution 101:242–251.
- Davidson, C. 2004. Declining downwind: Amphibian population declines in California and historical pesticide use. *Ecological Applications* 14:1892–1902.
- Davidson, C., H. B. Shaffer, M. R. Jennings - Conservation Biology, and 2002. 2002. Spatial tests of the pesticide drift, habitat destruction, UV-B, and climate-change hypotheses for California amphibian declines. *Conservation biology: the journal of the Society for Conservation Biology* 16:1588–1601.
- De'ath, G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88:243–251.
- Drost, C. A., and G. M. Fellers. 1996. Collapse of a Regional Frog Fauna in the Yosemite Area of the California Sierra Nevada, USA. *Conservation biology: the journal of the Society for Conservation Biology* 10:414–425.
- Dudgeon, D., A. H. Arthington, M. O. Gessner, Z.-I. Kawabata, D. J. Knowler, C. Lévêque, R. J. Naiman, A.-H. Prieur-Richard, D. Soto, M. L. J. Stiassny, and C. A. Sullivan. 2006. Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological reviews of the Cambridge Philosophical Society* 81:163–182.
- Durrell, C. 1988. *Geologic History of the Feather River Country, California*. University of California Press.
- Excoffier, L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll. 2013. Robust demographic inference from genomic and SNP data. *PLoS genetics* 9:e1003905.
- Excoffier, L., and M. Foll. 2011. Fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27:1332–1334.

References

- Fahrig, L., and G. Merriam. 1985. Habitat Patch Connectivity and Population Survival. *Ecology* 66:1762–1768.
- Frankham, R. 2002. Introduction to Conservation Genetics. Cambridge University Press, Cambridge, UK ; New York.
- Fuller, T. E., K. L. Pope, D. T. Ashton, and H. H. Welsh. 2011. Linking the Distribution of an Invasive Amphibian (*Rana catesbeiana*) to Habitat Conditions in a Managed River System in Northern California. *Restoration Ecology* 19:204–213.
- Fumagalli, M., F. G. Vieira, T. S. Korneliussen, T. Linderoth, E. Huerta-Sánchez, A. Albrechtsen, and R. Nielsen. 2013. Quantifying population genetic differentiation from next-generation sequencing data. *Genetics* 195:979–992.
- Fumagalli, M., F. G. Vieira, T. Linderoth, and R. Nielsen. 2014. ngsTools: Methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* 30:1486–1487.
- Gillespie, A. R., and P. H. Zehfuss. 2004. Glaciations of the Sierra Nevada, California, USA. Pages 51–62 *in* J. Ehlers and P. L. Gibbard, editors. *Developments in Quaternary Sciences*. Elsevier.
- Gilpin, M., and M. Soule. 1986. Minimum Viable Populations: Processes of Species Extinction, *Conservation Biology*. Sunderland, Massachusetts: Sin. Assoc.:19–34.
- Goldberg, C. S., M. E. Kaplan, and C. R. Schwalbe. 2003. From the frog's mouth: Buccal swabs for collection of DNA from amphibians. *Herpetological review* 34:220–221.
- Graham, C. H., J. Elith, R. J. Hijmans, A. Guisan, A. Townsend Peterson, B. A. Loiselle, and The Nceas Predicting Species Distributions Working Group. 2008. The influence of spatial errors in species occurrence data used in distribution models. *The Journal*

References

- of applied ecology 45:239–247.
- Grantham, T. E., A. M. Merenlender, and V. H. Resh. 2010. Climatic influences and anthropogenic stressors: An integrated framework for streamflow management in Mediterranean-climate California, USA. Freshwater biology 55:188–204.
- Guisan, A., N. E. Zimmermann, J. Elith, C. H. Graham, S. Phillips, and A. T. Peterson. 2007. What Matters for Predicting the Occurrences of Trees: Techniques, Data, or Species' Characteristics? Ecological monographs 77:615–630.
- Guzy, J. C., E. A. Eskew, B. J. Halstead, and S. J. Price. 2018. Influence of damming on anuran species richness in riparian areas: A test of the serial discontinuity concept. Ecology and evolution.
- Hellsten, U., R. M. Harland, M. J. Gilchrist, D. Hendrix, J. Jurka, V. Kapitonov, I. Ovcharenko, N. H. Putnam, S. Shu, L. Taher, I. L. Blitz, B. Blumberg, D. S. Dichmann, I. Dubchak, E. Amaya, J. C. Detter, R. Fletcher, D. S. Gerhard, D. Goodstein, T. Graves, I. V. Grigoriev, J. Grimwood, T. Kawashima, E. Lindquist, S. M. Lucas, P. E. Mead, T. Mitros, H. Ogino, Y. Ohta, A. V. Poliakov, N. Pollet, J. Robert, A. Salamov, A. K. Sater, J. Schmutz, A. Terry, P. D. Vize, W. C. Warren, D. Wells, A. Wills, R. K. Wilson, L. B. Zimmerman, A. M. Zorn, R. Grainger, T. Grammer, M. K. Khokha, P. M. Richardson, and D. S. Rokhsar. 2010. The genome of the Western clawed frog *Xenopus tropicalis*. Science 328:633–636.
- Hendricks, S., E. C. Anderson, T. Antao, L. Bernatchez, B. R. Forester, B. Garner, B. K. Hand, P. A. Hohenlohe, M. Kardos, B. Koop, A. Sethuraman, R. S. Waples, and G. Luikart. 2018. Recent advances in conservation and population genomics data analysis. Evolutionary applications.
- Heyer, W. R., M. A. Donnelly, R. W. McDiarmid, L.-A. C. Hayek, and M. S. Foster. 1994.

References

- Measuring and monitoring biological diversity. Standard methods for amphibians. Smithsonian Institution Press, Washington DC.
- Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith. 2017. Dismo: Species Distribution Modeling. R package.
- Hoffmann, A. A., and C. M. Sgrò. 2011. Climate change and evolutionary adaptation. *Nature* 470:479–485.
- Hotaling, S., C. C. Muhlfeld, J. J. Giersch, O. A. Ali, S. Jordan, M. R. Miller, G. Luikart, and D. W. Weisrock. 2018. Demographic modelling reveals a history of divergence with gene flow for a glacially tied stonefly in a changing post-Pleistocene landscape. *Journal of biogeography* 45:304–317.
- Ishiyama, N., I. Koizumi, T. Yuta, and F. Nakamura. 2015. Differential effects of spatial network structure and scale on population size and genetic diversity of the ninespine stickleback in a remnant wetland system. *Freshwater biology* 60:733–744.
- Jennings, M. R., and M. P. Hayes. 1994. Amphibian and reptile species of special concern in California. Final Report. California Department of Fish and Game Inland Fisheries Division, Rancho Cordova.
- Kahilainen, A., M. Puurtinen, and J. S. Kotiaho. 2014. Conservation implications of speciesGenetic diversity correlations. *Global Ecology and Conservation* 2:315–323.
- Kim, S. Y., K. E. Lohmueller, A. Albrechtsen, Y. Li, T. Korneliussen, G. Tian, N. Grarup, T. Jiang, G. Andersen, D. Witte, T. Jorgensen, T. Hansen, O. Pedersen, J. Wang, and R. Nielsen. 2011. Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC bioinformatics* 12:231.
- Knapp, R. A., G. M. Fellers, P. M. Kleeman, D. A. W. Miller, V. T. Vredenburg, E. B.

References

- Rosenblum, and C. J. Briggs. 2016. Large-scale recovery of an endangered amphibian despite ongoing exposure to multiple stressors. *Proceedings of the National Academy of Sciences of the United States of America* 113:11889–11894.
- Knapp, R., K. Matthews, H. Preisler, and R. Jellison. 2003. Developing probabilistic models to predict amphibian site occupancy in a patchy landscape. *Ecological applications: a publication of the Ecological Society of America* 13:1069–1082.
- Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC bioinformatics* 15:356.
- Korneliussen, T. S., I. Moltke, A. Albrechtsen, and R. Nielsen. 2013. Calculation of Tajima’s D and other neutrality test statistics from low depth next-generation sequencing data. *BMC bioinformatics* 14:289.
- Krohn, A. R., C. J. Conroy, R. Pesapane, K. Bi, J. E. Foley, and E. B. Rosenblum. 2018. Conservation genomics of desert dwelling California voles (*Microtus californicus*) and implications for management of endangered Amargosa voles (*Microtus californicus scirpensis*). *Conservation genetics* 19:383–395.
- Kupferberg, S. J. 1996. Hydrologic and geomorphic factors affecting conservation of a river-breeding frog (*Rana boylii*). *Ecological applications: a publication of the Ecological Society of America* 6:1332–1344.
- Kupferberg, S. J., W. J. Palen, A. J. Lind, S. Bobzien, A. Catenazzi, J. Drennan, and M. E. Power. 2012. Effects of flow regimes altered by dams on survival, population declines, and range-wide losses of California river-breeding frogs. *Conservation biology: the journal of the Society for Conservation Biology* 26:513–524.
- Kupferberg, S., A. Lind, J. Mount, and S. Yarnell. 2009. Pulsed flow effects on the

References

- Foothill Yellow-Legged Frog (*Rana boylii*): Integration of empirical, experimental, and hydrodynamic modeling approaches. California Energy Commission, PIER.
- Lande, R., and S. Shannon. 1996. The Role of Genetic Variation in Adaptation and Population Persistence in a Changing Environment. *Evolution; International Journal of Organic Evolution* 50:434–437.
- Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
- Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595.
- Li, H., and R. Durbin. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li, Y., X.-X. Zhang, R.-L. Mao, J. Yang, C.-Y. Miao, Z. Li, and Y.-X. Qiu. 2017. Ten Years of Landscape Genomics: Challenges and Opportunities. *Frontiers in plant science* 8:2136.
- Lind, A. J., H. H. Welsh Jr, and R. A. Wilson. 1996. The effects of a dam on breeding habitat and egg survival of the foothill yellow-legged frog (*Rana boylii*) in Northwestern California. *Herpetological review* 27:62–66.

References

- Macey, J. R., J. L. Strasburg, J. A. Brisson, V. T. Vredenburg, M. Jennings, and A. Larson. 2001. Molecular phylogenetics of western North American frogs of the *Rana boylii* species group. *Molecular phylogenetics and evolution* 19:131–143.
- Mallet, J. 2007. Hybrid speciation. *Nature* 446:279–283.
- Malone, J. H., and B. E. Fontenot. 2008. Patterns of reproductive isolation in toads. *PloS one* 3:e3900.
- McCartney-Melstad, E., M. Gidiş, and H. B. Shaffer. 2018. Population genomic data reveal extreme geographic subdivision and novel conservation actions for the declining foothill yellow-legged frog. *Heredity*.
- Miller, M. R., J. P. Brunelli, P. A. Wheeler, S. Liu, C. E. Rexroad 3rd, Y. Palti, C. Q. Doe, and G. H. Thorgaard. 2012. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular ecology* 21:237–249.
- Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* 17:240–248.
- Monsen, K. J., and M. S. Blouin. 2004. Extreme isolation by distance in a montane frog *Rana cascadae*. *Conservation genetics* 5:827–835.
- Moore, J. G., and B. C. Moring. 2013. Rangewide glaciation in the Sierra Nevada, California. *Geosphere* 9:1804–1818.
- Moyle, P. B., J. V. E. Katz, and R. M. Quiñones. 2011. Rapid decline of California's native inland fishes: A status assessment. *Biological conservation* 144:2414–2423.
- Murchie, K. J., K. P. E. Hair, C. E. Pullen, T. D. Redpath, H. R. Stephens, and S. J. Cooke. 2008. Fish response to modified flow regimes in regulated rivers: Research

References

- methods, effects and opportunities. *River research and applications* 24:197–217.
- Narum, S. R., C. A. Buerkle, J. W. Davey, M. R. Miller, and P. A. Hohenlohe. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* 22:2841–2847.
- Nielsen, R., T. Korneliussen, A. Albrechtsen, Y. Li, and J. Wang. 2012. SNP calling, genotype calling, and sample allele frequency estimation from New-Generation Sequencing data. *PloS one* 7:e37558.
- Nilsson, C., C. A. Reidy, M. Dynesius, and C. Revenga. 2005. Fragmentation and flow regulation of the world’s large river systems. *Science* 308:405–408.
- Nunziata, S. O., S. L. Lance, D. E. Scott, E. M. Lemmon, and D. W. Weisrock. 2017. Genomic data detect corresponding signatures of population size change on an ecological time scale in two salamander species. *Molecular ecology* 26:1060–1074.
- Pagano, A., A. Dubois, D. Lesbarrères, and T. Lodé. 2003. Frog alien species: A way for genetic invasion? *Comptes rendus biologies* 326 Suppl 1:S85–92.
- Parris, K. M., S. C. McCall, M. A. McCarthy, B. A. Minteer, K. Steele, S. Bekessy, and F. Medvecky. 2010. Assessing ethical trade-offs in ecological field studies. *The Journal of applied ecology* 47:227–234.
- Peek, R. 2010. Landscape Genetics of Foothill Yellow-Legged Frogs (*Rana boylii*) in regulated and unregulated rivers: Assessing connectivity and genetic fragmentation. Master’s Thesis, University of San Francisco.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological modelling* 190:231–259.
- Pidancier, N., C. Miquel, and C. MIAUD. 2003. Buccal swabs as a non-destructive tissue

References

- sampling method for DNA analysis in amphibians. *Herpetological Journal* 13:175–178.
- Poff, N. L., J. D. Olden, D. M. Merritt, and D. M. Pepin. 2007. Homogenization of regional river dynamics by dams and global biodiversity implications. *Proceedings of the National Academy of Sciences of the United States of America* 104:5732–5737.
- Power, M. E., W. E. Dietrich, and J. C. Finlay. 1996. Dams and Downstream Aquatic Biodiversity: Potential Food Web Consequences of Hydrologic and Geomorphic Change. *Environmental management* 20:887–895.
- Prager, E. M., and A. C. Wilson. 1975. Slow evolutionary loss of the potential for interspecific hybridization in birds: A manifestation of slow regulatory evolution. *Proceedings of the National Academy of Sciences of the United States of America* 72:200–204.
- Prince, D. J., S. M. O'Rourke, T. Q. Thompson, O. A. Ali, H. S. Lyman, I. K. Saglam, T. J. Hotaling, A. P. Spidle, and M. R. Miller. 2017. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science advances* 3:e1603198.
- Pringle, C. 2003. What is hydrologic connectivity and why is it ecologically important? *Hydrological processes* 17:2685–2689.
- Pringle, C. M. 2001. Hydrologic Connectivity and the Management of Biological Reserves: A Global Perspective. *Ecological applications: a publication of the Ecological Society of America* 11:981–998.
- Railsback, S. F., B. C. Harvey, S. K. Kupferberg, M. M. Lang, S. McBain, and H. H. Welsh Jr. 2015. Modeling potential river management conflicts between frogs and salmonids. *Canadian journal of fisheries and aquatic sciences. Journal canadien des sciences halieutiques et aquatiques*.

References

- R Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ricciardi, A., and J. B. Rasmussen. 1999. Extinction rates of North American freshwater fauna. *Conservation biology: the journal of the Society for Conservation Biology* 13:1220–1222.
- Ridgeway, G. 2015. Gbm: Generalized Boosted Regression Models. <http://CRAN.R-project.org/package-gbm>.
- Rolls, R. J., and N. R. Bond. 2017. Environmental and Ecological Effects of Flow Alteration in Surface Water Ecosystems. Pages 65–82 in *Water for the Environment*. Elsevier.
- Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1219–1228.
- Ruby, J. G., P. Bellare, and J. L. Derisi. 2013. PRICE: Software for the targeted assembly of components of (Meta) genomic sequence data. *G3* 3:865–880.
- Sabo, J. L., A. Ruhi, G. W. Holtgrieve, V. Elliott, M. E. Arias, P. B. Ngor, T. A. Räsänen, and S. Nam. 2017. Designing river flows to improve food security futures in the Lower Mekong Basin. *Science* 358.
- Schick, R. S., and S. T. Lindley. 2007. Directed connectivity among fish populations in a riverine network. *The Journal of applied ecology* 44:1116–1126.
- Scribner, K. T., W. H. Lowe, E. Landguth, G. Luikart, D. M. Infante, G. E. Whelan, and C. C. Muhlfeld. 2016. Applications of Genetic Data to Improve Management and Conservation of River Fishes and Their Habitats. *Fisheries* 41:174–188.
- Shaffer, H. B., G. M. Fellers, S. R. Voss, J. C. Oliver, and G. B. Pauly. 2004. Species

References

- boundaries, phylogeography and conservation genetics of the red-legged frog (*Rana aurora/draytonii*) complex. *Molecular ecology* 13:2667–2677.
- Shaw, E. A., E. Lange, J. D. Shucksmith, and D. N. Lerner. 2016. Importance of partial barriers and temporal variation in flow when modelling connectivity in fragmented river systems. *Ecological engineering* 91:515–528.
- Skotte, L., T. S. Korneliussen, and A. Albrechtsen. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195:693–702.
- Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. *Science* 236:787–792.
- Spinks, P. Q., R. C. Thomson, and H. B. Shaffer. 2010. Nuclear gene phylogeography reveals the historical legacy of an ancient inland sea on lineages of the western pond turtle, *Emys marmorata* in California. *Molecular ecology* 19:542–556.
- Stebbins, R. C. 2003. A Field Guide to Western Reptiles and Amphibians. 3rd editions. Houghton Mifflin Harcourt, Boston.
- Steel, A. E., R. A. Peek, R. A. Lusardi, and S. M. Yarnell. 2017. Associating metrics of hydrologic variability with benthic macroinvertebrate communities in regulated and unregulated snowmelt-dominated rivers. *Freshwater biology*.
- Streicher, J. W., T. J. Devitt, C. S. Goldberg, J. H. Malone, H. Blackmon, and M. K. Fujita. 2014. Diversification and asymmetrical gene flow across time and space: Lineage sorting and hybridization in polytypic barking frogs. *Molecular ecology* 23:3273–3291.
- Sun, Y.-B., Z.-J. Xiong, X.-Y. Xiang, S.-P. Liu, W.-W. Zhou, X.-L. Tu, L. Zhong, L. Wang, D.-D. Wu, B.-L. Zhang, C.-L. Zhu, M.-M. Yang, H.-M. Chen, F. Li, L. Zhou, S.-H. Feng, C. Huang, G.-J. Zhang, D. Irwin, D. M. Hillis, R. W. Murphy, H.-M. Yang,

References

- J. Che, J. Wang, and Y.-P. Zhang. 2015. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proceedings of the National Academy of Sciences of the United States of America* 112:E1257–62.
- Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–60.
- Taylor, P. D., L. Fahrig, K. Henein, and G. Merriam. 1993. Connectivity Is a Vital Element of Landscape Structure. *Oikos* 68:571–573.
- Thomson, B., A. Wright, and B. Shaffer. 2016. California Amphibian and Reptile Species of Special Concern. University of California Press, Berkeley, CA., <http://arssc.ucdavis.edu>.
- Tonkin, J. D., M. T. Bogan, N. Bonada, B. Rios-Touma, and D. A. Lytle. 2017. Seasonality and predictability shape temporal species diversity. *Ecology* 98:1201–1216.
- Tyers, M. 2017. Riverdist: River Network Distance Computation and Applications.
- USFWS. 2014. Endangered and Threatened Wildlife and Plants; Endangered Species Status for Sierra Nevada Yellow-Legged Frog and Northern Distinct Population Segment of the Mountain Yellow-Legged Frog, and Threatened Species Status for Yosemite Toad. *Federal register* 79:24255–24310.
- Voelker, G., B. D. Marks, C. Kahindo, U. A'genonga, F. Bapeamoni, L. E. Duffie, J. W. Huntley, E. Mulotwa, S. A. Rosenbaum, and J. E. Light. 2013. River barriers and cryptic biodiversity in an evolutionary museum. *Ecology and evolution* 3:536–545.
- Vorosmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann, and P. M. Davies. 2010. Global threats to human water security and river biodiversity. *Nature* 467:555–61.

References

- Vredenburg, V. 2004. Reversing introduced species effects: Experimental removal of introduced fish leads to rapid recovery of a declining frog. *Proceedings of the National Academy of Sciences* 101:7646–7650.
- Vredenburg, V., R. Bingham, R. Knapp, J. Morgan, C. Moritz, and D. Wake. 2007. Concordant molecular and phenotypic data delineate new taxonomy and conservation priorities for the endangered mountain yellow-legged frog. *Journal of Zoology* 271:361–374.
- Watterson, G. A. 1975. Number of Segregating Sites in Genetic Models without Recombination. *Theoretical Population Biology* 7:256–276.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-Statistics for the Analysis of Population-Structure. *Evolution; international journal of organic evolution* 38:1358–1370.
- Werth, S., M. Schödl, and C. Scheidegger. 2014. Dams and canyons disrupt gene flow among populations of a threatened riparian plant. *Freshwater biology* 59:2502–2515.
- Whiteley, A. R., K. Hastings, J. K. Wenburg, C. A. Frissell, J. C. Martin, and F. W. Allendorf. 2010. Genetic variation and effective population size in isolated populations of coastal cutthroat trout. *Conservation genetics* 11:1929–1943.
- Wickham, H. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H., R. François, L. Henry, and K. Müller. 2018. *Dplyr: A Grammar of Data Manipulation*.
- Wiens, J. A. 2002. Riverine landscapes: Taking landscape ecology into the water. *Freshwater biology* 47:501–515.

- Wilbur, H. M., and R. D. Semlitsch. 1990. Ecological Consequences of Tail Injury in *Rana* Tadpoles. *Copeia* 1990:18–24.
- Wilke, C. O. 2018. Cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'.
- Wright, S. 1943. Isolation by Distance. *Genetics* 28:114–38.
- Yarnell, S. M., A. J. Lind, and J. F. Mount. 2012. Dynamic flow modelling of riverine amphibian habitat with application to regulated flow management. *River Research and Applications* 28:177–191.
- Yarnell, S. M., J. H. Viers, and J. F. Mount. 2010. Ecology and Management of the Spring Snowmelt Recession. *Bioscience* 60:114–127.
- Yarnell, S., R. Peek, G. Epke, and A. Lind. 2016. Management of the Spring Snowmelt Recession in Regulated Systems. *JAWRA Journal of the American Water Resources Association* 52:723–736.
- Yuan, Z.-Y., W.-W. Zhou, X. Chen, N. A. Poyarkov Jr, H.-M. Chen, N.-H. Jang-Liaw, W.-H. Chou, N. J. Matzke, K. Iizuka, M.-S. Min, S. L. Kuzmin, Y.-P. Zhang, D. C. Cannatella, D. M. Hillis, and J. Che. 2016. Spatiotemporal Diversification of the True Frogs (Genus *Rana*): A Historical Framework for a Widely Studied Group of Model Organisms. *Systematic biology* 65:824–842.
- Zhong, Y., and G. Power. 1996. Environmental impacts of hydroelectric projects on fish resources in China. *Regulated Rivers: Research & Management* 12:81–98.
- Zweifel, R. G. 1955. Ecology, distribution, and systematics of frogs of the *Rana boylei* group. *University of California Publications in Zoology* 54:207–291.