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

"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 priviledge 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 genomically-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

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

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 (Fuller, Pope, Ashton, & Welsh, 2011; Sarah J. Kupferberg et al., 2012; Lind, Welsh, & Wilson, 1996; Rolls & Bond, 2017), reductions in population abundances and diversity (Fuller et al., 2011; Guzy, Eskew, Halstead, & Price, 2018; Lind et al., 1996; Sabo et al., 2017; Scribner et al., 2016; Vörösmarty et al., 2010; Zhong & Power, 1996), and fragmentation (Guzy et al., 2018; Sabo et al., 2017; Scribner et al., 2016; Vörösmarty et al., 2010; Werth, Schödl, & Scheidegger, 2014; Zhong & Power, 1996) 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 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 & 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, Merenlender, & Resh, 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). FYLF are intimately linked

with river hydrology because they have evolved to spawn in synchrony with natural flow cues associated with seasonal spring snowmelt or rain recession periods (Bondi, Yarnell, Lind, & Lind, 2013; S. J. Kupferberg, 1996; S. M. Yarnell et al., 2010; S. Yarnell, Peek, Epke, & Lind, 2016). 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 (Davidson, Shaffer, & Jennings, 2002; Jennings & Hayes, 1994). In California, particularly in the Sierra Nevada, river regulation may be a significant environmental stressor (Sarah J. Kupferberg et al., 2012; Lind et al., 1996). 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 FYLF 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, FYLF populations are now restricted to small unregulated tributaries flowing into the regulated mainstem.

Here, we investigate the impacts of river regulation on genetic health of FYLF 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 explore the association between these metrics and levels of river regulation. Our goal is to assess the genetic health of FYLF under different river regulation regimes to better inform the potential for long-term persistence. Addressing this question will help to inform management and conservation efforts for FYLF, as well as the potential utility of genetics for future conservation monitoring efforts in aquatic species.

Methods

Sample collection

345 FYLF buccal or tissue samples were used in this study (see Table S1). Field sampling was conducted as previously described (Heyer, Donnelly, McDiarmid, Hayek, & Foster, 1994), under CDFW SCP Permit #0006881, with IACUC protocol #19327. Individual post-metamorphic frogs were buccal-swabbed following established protocols (Broquet, Berset-Braendli, Emaresi, & Fumagalli, 2007; Goldberg, Kaplan, & Schwalbe, 2003; Pidancier, Miquel, & Miaud, 2003). 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 (Parris et al., 2010; Wilbur & Semlitsch, 1990). One small (<3mm) tail clip was taken per individual tadpole and dried on Whatman qualitative filter paper (grade 1) and stored at room temperature.

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 RAD sequencing (Baird et al., 2008; Miller et al., 2007). Paired-end sequence data were generated from 24 FYLF individuals (sampling details given in Table S2) across coastal and Sierra Nevada populations from California, USA. DNA was 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). RAD libraries were constructed using the SbfI restriction enzyme and a new RAD protocol (Ali et al., 2016). De novo loci discovery and contig extension were carried as previously described (Miller et al., 2012) using the alignment program Novoalign and the genome assembler PRICE (Ruby, Bellare, & Derisi, 2013). This pipeline resulted in a set of 77,544 RAD contigs ranging from 300 to 800 bp (Table S3) which served as a de novo partial genome reference for all subsequent downstream analyses.

Rapture sequencing

We then performed Rapture on all samples (Table S1) (2016) using 8,533 RAD capture baits (120 bp) were designed by Arbor Biosciences from the de novo alignment (Table S4). The final Rapture library was sequenced in 50% of an Illumina HiSeq 3000 lane. Rapture sequence data from each individual (Table S1) were aligned against the de novo partial genome reference using the BWA-MEM algorithm (Li, 2013; Li & Durbin, 2010) 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 multiple libraries (Li et al., 2009). BAM files from the same sample were merged before indexing using SAMtools.

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 (Fumagalli et al., 2013; Korneliussen, Moltke, Albrechtsen, & Nielsen, 2013). All Rapture analyses were conducted using Analysis of Next Generation Sequencing Data (ANGSD) (Korneliussen,

Albrechtsen, & Nielsen, 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 (Table S1, S2) using 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 were 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 FST was used to quantify genetic differentiation between populations (Rousset, 1997; Wright, 1943). Genome-wide FST between population pairs was estimated by first calculating a site frequency spectrum (SFS) for each population (doSaf) (Nielsen, Korneliussen, Albrechtsen, Li, & Wang, 2012) with ANGSD. The two-dimensional SFS and global FST between each population pair were then estimated using realSFS (Korneliussen et al., 2014). FST 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 water-

shed. These values were plotted and a generalized linear model was fitted (FST ~ Mean River Distance) in R (R Core Team, 2017). To calculate Watterson's S (Watterson, 1975), and Tajima's (Tajima, 1983), we used SFS that were estimated as described above as priors (pest) to calculate each statistic for each site (doThetas), which were averaged to obtain a single value for each statistic (Korneliussen et al., 2013).

Boosted regression tree modeling of variance in FST

We used boosted regression tree (BRT) models with the R packages gbm (Ridgeway, 2015) and dismo (Hijmans, Phillips, Leathwick, & Elith, 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, Peek, Lusardi, & Yarnell, 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 (Guisan et al., 2007; Phillips, Anderson, & Schapire, 2006). 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 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 FST, 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, Δ (− S) was included to assess the effect of genomic variation on FST 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

Rapture produces high quality genomic data for FYLF

To begin investigating the impact of river regulation on FYLF, 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). Flow data were obtained for each river reach using proximal USGS gaging stations (Table S5). We sampled a total of 345 FYLF from sites in three major watersheds (Yuba, Bear, and American) in the northern Sierra Nevada of California (Figure 1A; Table 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 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

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). FYLF 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). 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, Buerkle, Davey, Miller, & Hohenlohe, 2013). We conclude that the sequence data we obtained should be appropriate for population genetic analyses across our study area.

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

To assess FYLF 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 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 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 FYLF 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 2B). In contrast, PCA of the MF American showed strong differentiation between sites (Figures 2C, 2D). The MF American PCA completely resolved all sites, with the first component (PC1) strongly differentiating the samples 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 FYLF in the Northern Sierra

To assess how patterns of genetic differentiation are associated with river regulation across our entire study area, we estimated pairwise FST (Wright, 1943) between all collection locations within a river for all six rivers. We then plotted the scaled mean pairwise FST [mean FST / (1-mean FST)] (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 (see Methods). While there was a clear relationship between FST and river distance (as shown by the slope of regression lines in Figure 3A), there was a striking pattern of elevated FST by regulation

type (Figure 3A). Even the bypass regulation type showed a distinct pattern of elevated FST. For instance, regulated rivers with locations separated by less than 10km had FST 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 FST values with the steepest regression coefficient. The baseline FST or global mean increased by over 0.1 between the unregulated (mean FST=0.141), and regulated locations (global mean for bypass FST=0.256, hydropeaking FST=0.278). This suggests a greater degree of isolation within sites in regulated river reaches compared with FYLF populations in unregulated reaches as larger FST values represent reductions in heterozygosity due to population subdivision (Slatkin, 1987). We conclude FYLF 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. FST), 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 (see below, Methods). We found flow regulation explained the greatest amount of variance in FST (Figure 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 (Rousset, 1997; Slatkin, 1987; Wright, 1943). We conclude there is a pattern of isolation and limited connectivity between populations in regulated reaches.

River regulation strongly correlated with decreasing genetic diversity in FYLF 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 (4N): Tajima's ( ) is based on the average number of pairwise differences (Tajima, 1983), and Watterson's (S) is based on the number of segregating sites (Watterson, 1975) (see methods). 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 should be equal; when genetic diversity has been increasing, > S; and when genetic diversity has been decreasing, S > . We found zero populations sampled within regulated watersheds had evidence of increasing genetic diversity (e.g., a − S that was less than zero) (Figure 4A). The regulated locations showed a clear trajectory of genetic diversity loss (Figure 4A, 4B). Three of the four hydropeaking locations had the highest values of Δ (− 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 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.

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, Lance, Scott, Lemmon, & Weisrock, 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 FYLF. Based on this dataset, we were able to successfully characterize patterns of genetic variation within FYLF as well as design a set of RAD capture baits that can be used as a genetic monitoring resource for FYLF (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 & Merriam, 1985; Taylor, Fahrig, Henein, & Merriam, 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, FYLF now occur in isolated locations, breeding in tributaries rather than mainstem habitats. However, since these frogs have the potential to move long distances (FYLF have been observed moving over 1 km per day (Bourque, 2008)), the extent to which current population connectivity has been lost due to river regulation remains unknown. Examining pairwise FST, 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 3B). Thus, despite being able to move long distances, FYLF 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 barriers to movement of FYLF) 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 (Frankham, 2002; Hoffmann & Sgrò, 2011; Ishiyama, Koizumi, Yuta, & Nakamura, 2015; Lande & Shannon, 1996). 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 (Frankham, 2002; Krohn et al., 2018). Throughout the Sierra Nevada, FYLF 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 serve as a reasonable proxy for genetic health, does not bode well for the long-term persistence of FYLF populations in regulated rivers in the Sierra Nevada. Many of these FYLF populations are already losing genetic diversity and given their small size and reduced connectivity the effects of inbreeding will likely exacerbate their problems. FYLF have evolved in river systems with consistent hydro-

logic seasonality and predictability, despite inter-annual variation. Flow regulation has altered patterns of hydrologic seasonality and predictability in many watersheds (Sarah J. Kupferberg et al., 2012). Long-term population persistence may still be possible if conservation efforts utilize methods that promote or maintain genetic health 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, FYLF conservation efforts should focus on river reaches where flow management may provide opportunities to more closely mimic local natural flow regimes and thus improve hydrologic connectivity.

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 (Botero et al., 2015; Kahilainen, Puurtinen, & Kotiaho, 2014). We demonstrate that an aquatic species that has adapted to local hydrology patterns shows poor genetic health (i.e., clear patterns of 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.

Data Archiving Statement: Should the manuscript be accepted, the data supporting

the results will be archived in an appropriate public repository such as Dryad, and the data DOI will be appended to the end of the article.

```
{r maxdelays, results="asis", echo=Fevalrary(knitr) kable(max_delays,
col.names = c("Airline", "Max Arrival Delay"),       caption = "Maximum
Delays by Airline",       caption.short = "Max Delays by Airline",
longtable = TRUE,       booktabs = TRUE)
```

The last two options make the table a little easier-to-read.

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).

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 (Stebbins 2003, @zweifel_ecology_1955). 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-

eral tributaries to the Feather River in the northern Sierra Nevada, California (Figure 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* species 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?

Materials and 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 S1). All unknown individuals were from Feather watershed localities.

Table 2.1: Sample Sites

Locality	River	No. Samples	Lat.	Lon.	Elev (m)	Watershed (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

Table 2.1: Sample Sites *(continued)*

Locality	River	No. Samples	Lat.	Lon.	Elev (m)	Watershed (HUC8)
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
FEA-SFRockCk	FEA	27	39.8789	-121.0022	1470	Feather
FEA-SPANISH-BGulch	FEA	5	39.9546	-121.089	1283	Feather

Table 2.1: Sample Sites *(continued)*

Locality	River	No. Samples	Lat.	Lon.	Elev (m)	Watershed (HUC8)
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

Table 2.1: Sample Sites *(continued)*

Locality	River	No. Samples	Lat.	Lon.	Elev (m)	Watershed (HUC8)
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

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). 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 (Goldberg et al. 2003). 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 (Table 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). 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 (File S3: de novo). 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 (File S4: Baits). Sample libraries were prepared for sequencing following RAD Capture (Rapture) methods outlined in Ali et al. (2016). These samples were then 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 (File S1). 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 (Table 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, 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 & Admixture

To assess population structure and coancestry, ANGSD was used to generate PCA and NGSadmix 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 2) 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 NGSadmix (Skotte et al. 2013) to infer ancestry proportions in *R. sierrae* and *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).

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 as F1 vs. F2 hybrid individuals will have different degrees of heterozygosity in these 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. 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 population 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 10 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 selecting the best model from each bootstrap based on maximum likelihoods as described

above.

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 1, Table S1). For downstream analysis, we filtered and retained 311 samples from the original sequencing data that contained a minimum of 25,000 alignments (Table S5). The final merged dataset mean alignments per sample was 229,485 (Table 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 (Table S1). We conclude that the sequence data we obtained should be appropriate for population genetic analyses across our study area.

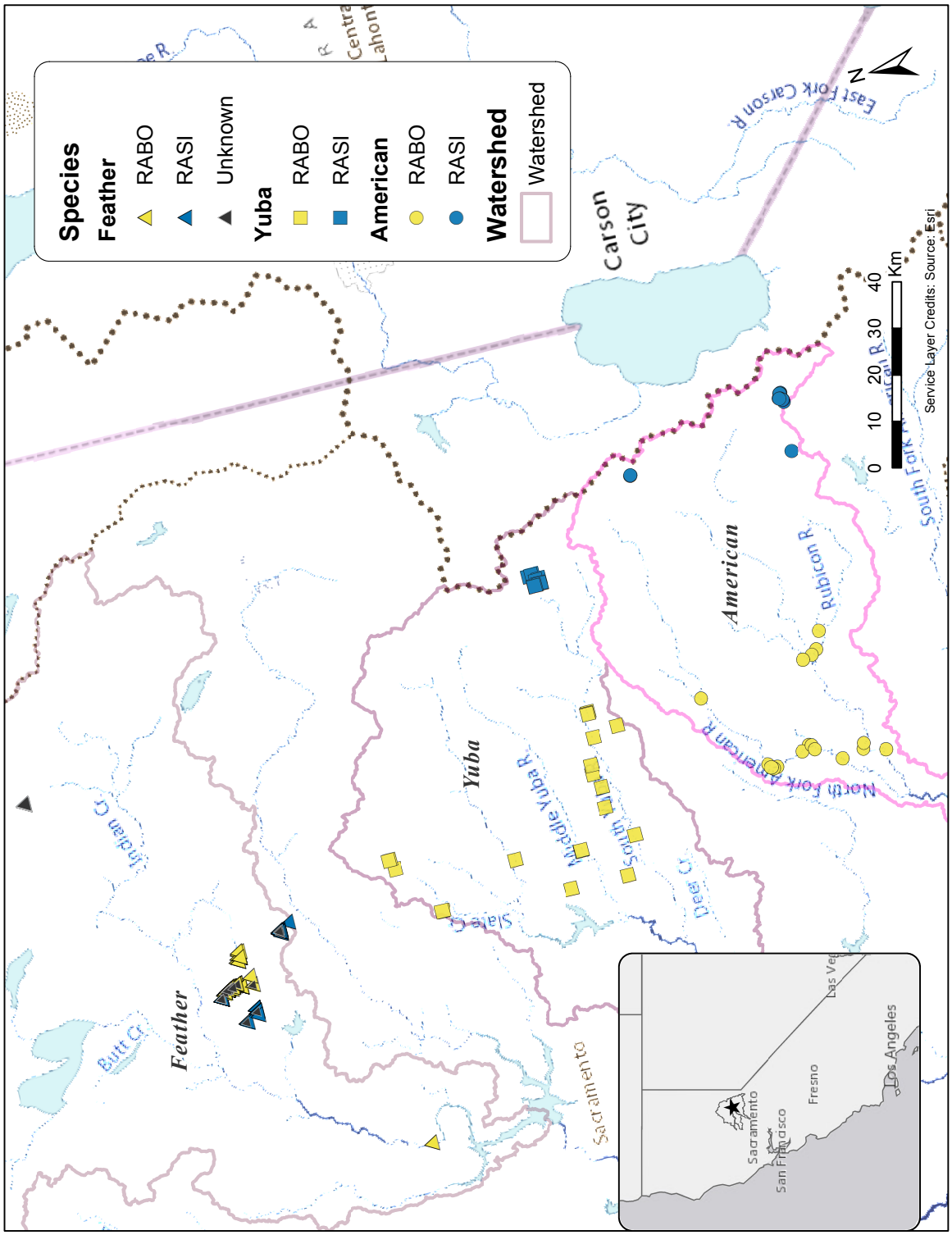

Figure 2.1: Map of sampling locations in the Feather, Yuba, and American watersheds. RABO=*R. boylii*, RASI=*R. sierrae*.

Chapter 3

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

```
#include_graphics(path = "figure/uw.png")
```

```
# Here is a reference to the UW logo: Figure \@ref(fig:uwlogo).  Note the use of th
```

INTRODUCTION

The use of modern genomic sequencing technology has greatly advanced the ability for higher resolution analyses of both geographic and ecological patterns in populations (Nunziata et al. 2017, Barbosa et al. 2018, Hendricks et al. 2018). Reduced representa-

tion sequencing methods such as restriction site-associated DNA sequencing (RADSeq) (Miller et al. 2007, Baird et al. 2008, Ali et al. 2016) provides a powerful tool 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 desired loci and allow highly efficient genotyping of thousands of individuals at once. As historical and future landscape use can influence species demography and migration patterns (Burkey 1989, Anderson and Beer 2009, Barbosa et al. 2018), these genomic tools will be invaluable for assessing critical factors for long-term persistence in sensitive populations or species.

The ecological integrity of freshwater systems and their constituent biota are rapidly declining globally (Ricciardi and Rasmussen 1999), and conservation efforts will require assessment of the adaptive capacity of populations to rapid environmental change. Given limited capacity to conserve, 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 future assessment responses to conservation actions. Thus, quantifying and linking landscape change with genetic diversity metrics may provide an important baseline to track 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. 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).

As with many amphibians in California (Davidson 2004, Peek 2010, Thomson et al. 2016), there have been significant population declines 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). *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 relative genomic diversity 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 that is currently a candidate for listing. 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 the potential for long-term persistence across this species' range.

MATERIALS AND 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.

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 (**Table S1**). 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 (**Supplemental File S2**). 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 (**Supplemental File S3**). 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 ge-

netic 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 (Korneliussen et al. 2014, Fumagalli et al. 2014). 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 package in R (R Core Team 2017). To assess admixture in *R. boylii*, genotype likelihood data (-GL 2 and -doGLF 2) was generated in ANGSD with the same settings as above. 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 site 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 organism 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 a maxIter of 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 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, $\theta_S > \theta_\pi$. 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 (see Methods). 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 1, Table 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 1), there was higher resolution sam-

pling 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 1).

Appendix A

The First Appendix

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readibility and/or setup.

In the main Rmd file

```
# This chunk ensures that the huskydown package is
# installed and loaded. This huskydown package includes
# the template files for the thesis.
if(!require(devtools))
  install.packages("devtools", repos = "http://cran.rstudio.com")
if(!require(huskydown))
  devtools::install_github("benmarwick/huskydown")
library(huskydown)
#if(!require(huskydown))
#  devtools::install_github("danovando/gauchodown")
#library(gauchodown)
library(knitr)
```

In Chapter ??:

Appendix B

The Second Appendix, for Fun

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 huskydown and bookdown.

This document was typeset using the XeTeX typesetting system, and the University of Washington Thesis class class created by Jim Fox. Under the hood, the University of Washington Thesis LaTeX template is used 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 all the data files, have been organised into an R package, xxx, which is available at `https://github.com/xxx/xxx`. A hard copy of the thesis can be found in the University of Washington library.

This version of the thesis was generated on 2018-09-10 17:50:58. The repository is currently at this commit:

The computational environment that was used to generate this version is as follows:

```
Session info -------------------------------------------------------------

 setting  value
 version  R version 3.5.1 (2018-07-02)
 system   x86_64, darwin15.6.0
 ui       X11
 language (EN)
 collate  en_US.UTF-8
 tz       America/Los_Angeles
 date     2018-09-10

Packages -----------------------------------------------------------------

 package     * version    date       source
 assertthat    0.2.0      2017-04-11 CRAN (R 3.5.0)
 backports     1.1.2      2017-12-13 CRAN (R 3.5.0)
 base        * 3.5.1      2018-07-05 local
 bindr         0.1.1      2018-03-13 CRAN (R 3.5.0)
 bindrcpp      0.2.2      2018-03-29 CRAN (R 3.5.0)
 bookdown    * 0.7        2018-02-18 CRAN (R 3.5.0)
 broom         0.5.0      2018-07-17 CRAN (R 3.5.0)
 cellranger    1.1.0      2016-07-27 CRAN (R 3.5.0)
 cli           1.0.0      2017-11-05 CRAN (R 3.5.0)
 colorspace    1.3-2      2016-12-14 CRAN (R 3.5.0)
 compiler      3.5.1      2018-07-05 local
 crayon        1.3.4      2017-09-16 CRAN (R 3.5.0)
 datasets    * 3.5.1      2018-07-05 local
```

```
devtools    * 1.13.6     2018-06-27 CRAN (R 3.5.0)
digest        0.6.16     2018-08-22 CRAN (R 3.5.1)
dplyr       * 0.7.6      2018-06-29 CRAN (R 3.5.0)
evaluate      0.11       2018-07-17 CRAN (R 3.5.0)
forcats     * 0.3.0      2018-02-19 CRAN (R 3.5.0)
ggplot2     * 3.0.0.9000 2018-09-04 Github (tidyverse/ggplot2@6e545dc)
git2r         0.23.0     2018-07-17 CRAN (R 3.5.0)
glue          1.3.0      2018-07-17 CRAN (R 3.5.0)
graphics    * 3.5.1      2018-07-05 local
grDevices   * 3.5.1      2018-07-05 local
grid          3.5.1      2018-07-05 local
gtable        0.2.0      2016-02-26 CRAN (R 3.5.0)
haven         1.1.2      2018-06-27 CRAN (R 3.5.0)
hms           0.4.2      2018-03-10 CRAN (R 3.5.0)
htmltools     0.3.6      2017-04-28 CRAN (R 3.5.0)
httr          1.3.1      2017-08-20 CRAN (R 3.5.0)
huskydown   * 0.0.5      2018-09-04 Github (benmarwick/huskydown@3ef00c9)
jsonlite      1.5        2017-06-01 CRAN (R 3.5.0)
kableExtra  * 0.9.0      2018-05-21 CRAN (R 3.5.0)
knitr       * 1.20       2018-02-20 CRAN (R 3.5.0)
lattice       0.20-35    2017-03-25 CRAN (R 3.5.1)
lazyeval      0.2.1      2017-10-29 CRAN (R 3.5.0)
lubridate     1.7.4      2018-04-11 CRAN (R 3.5.0)
magrittr      1.5        2014-11-22 CRAN (R 3.5.0)
memoise       1.1.0      2017-04-21 CRAN (R 3.5.0)
methods     * 3.5.1      2018-07-05 local
```

```
modelr        0.1.2      2018-05-11 CRAN (R 3.5.0)
munsell       0.5.0      2018-06-12 CRAN (R 3.5.0)
nlme          3.1-137    2018-04-07 CRAN (R 3.5.1)
pillar        1.3.0      2018-07-14 CRAN (R 3.5.0)
pkgconfig     2.0.2      2018-08-16 CRAN (R 3.5.0)
plyr          1.8.4      2016-06-08 CRAN (R 3.5.0)
purrr       * 0.2.5      2018-05-29 CRAN (R 3.5.0)
R6            2.2.2      2017-06-17 CRAN (R 3.5.0)
Rcpp          0.12.18    2018-07-23 CRAN (R 3.5.1)
readr       * 1.1.1      2017-05-16 CRAN (R 3.5.0)
readxl        1.1.0      2018-04-20 CRAN (R 3.5.0)
rlang         0.2.2      2018-08-16 CRAN (R 3.5.0)
rmarkdown     1.10       2018-06-11 cran (@1.10)
rprojroot     1.3-2      2018-01-03 CRAN (R 3.5.0)
rstudioapi    0.7        2017-09-07 CRAN (R 3.5.0)
rvest         0.3.2      2016-06-17 CRAN (R 3.5.0)
scales        1.0.0.9000 2018-08-29 Github (hadley/scales@0f7a186)
stats       * 3.5.1      2018-07-05 local
stringi       1.2.4      2018-07-20 CRAN (R 3.5.0)
stringr     * 1.3.1      2018-05-10 CRAN (R 3.5.0)
tibble      * 1.4.2      2018-01-22 CRAN (R 3.5.0)
tidyr       * 0.8.1      2018-05-18 CRAN (R 3.5.0)
tidyselect    0.2.4      2018-02-26 CRAN (R 3.5.0)
tidyverse   * 1.2.1      2017-11-14 CRAN (R 3.5.0)
tools         3.5.1      2018-07-05 local
utils       * 3.5.1      2018-07-05 local
```

```
viridisLite   0.3.0      2018-02-01 CRAN (R 3.5.0)
withr         2.1.2      2018-08-29 Github (jimhester/withr@8b9cee2)
xfun          0.3        2018-07-06 CRAN (R 3.5.0)
xml2          1.2.0      2018-01-24 CRAN (R 3.5.0)
yaml          2.2.0      2018-07-25 CRAN (R 3.5.0)
```

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.

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 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.

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.

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.

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

phylogenetics and evolution 101:242–251.

Davidson, C. 2004. DECLINING DOWNWIND: AMPHIBIAN POPULATION DECLINES IN CALIFORNIA AND HISTORICAL PESTICIDE USE. Ecological applications: a publication of the Ecological Society of America 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.

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.

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.

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.

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. Measuring and monitoring biological diversity. Standard methods for amphibians. Smithsonian Institution Press, Washington DC.

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.

Jennings, M. R., and M. P. Hayes. 1994. Amphibian and reptile species of special concern in California. Final Report. California Department of Fish; Game Inland Fisheries Division, Rancho Cordova.

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.

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.

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.

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. 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 Calfifornia. Herpetological review 27:62–66.

Mallet, J. 2007. Hybrid speciation. Nature 446:279–283.

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.

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 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.

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. PhD Thesis, University of San Francisco.

Pidancier, N., C. Miquel, and C. Miaud. 2003. Buccal swabs as a non-destructive tissue sampling method for DNA analysis in amphibians. The 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.

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.

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.

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.

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.

Shaffer, H. B., G. M. Fellers, S. R. Voss, J. C. Oliver, and G. B. Pauly. 2004. Species 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.

Stebbins, R. C. 2003. A Field Guide to Western Reptiles and Amphibians. 3rd editions. Houghton Mifflin Harcourt, Boston.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.

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.

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.

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.

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.

Wright, S. 1943. Isolation by Distance. Genetics 28:114–38.

Yarnell, S. M., J. H. Viers, and J. F. Mount. 2010. Ecology and Management of the Spring Snowmelt Recession. Bioscience 60:114–127.

Zweifel, R. G. 1955. Ecology, distribution, and systematics of frogs of the Rana boylei group. University of California Publications in Zoology 54:207–291.