Title

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

Authors and Affiliations

Ryan A. Peek*, Sean M. O'Rourke, Michael R. Miller
Center for Watershed Sciences and Department of Animal Science, University of California, Davis, CA 95616, USA
*Corresponding author. Email: rapeek@ucdavis.edu

Abstract

River regulation, or the hydrological alteration of flow by dams and diversions, 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 which river regulation impacts these critical attributes of genetic health is an important goal. Foothill yellow-legged frog (FYLF; *Rana boylii*) were historically abundant throughout many western rivers but have declined since the onset of regulation. However, the extent to which FYLF populations in regulated rivers are maintaining connectivity and genetic diversity is unknown. Here we use genetic methods to investigate the impacts of river regulation on FYLF to explore their potential for long-term persistence under continued regulation. We found FYLF 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. Although our results do not bode well for long-term persistence of FYLF populations under current flow regulation regimes, they do highlight the power of genetic monitoring for assessing population health in aquatic organisms.

Keywords

Rivers, hydropower, flow regulation, genetics, frogs, hydrologic connectivity

Significance

Hydropower is an important source of renewable energy globally, but hydropower generation modifies natural flow regimes and may alter important processes of aquatic ecosystems. Better methods for assessing the long-term impacts of river regulation on aquatic ecosystems are needed. For example, exploring the potential for long-term population persistence in aquatic species under current regulation levels is a key component for conservation management. Our study uses genetic methods to investigate the impacts of river regulation on population health of a river-breeding frog species. We found that populations in regulated rivers showed striking patterns of connectivity and genetic diversity loss relative to unregulated rivers. Our results suggest that changes to current regulation regimes may be needed to promote long-term population persistence.

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 (1, 2), while also acting as physical barriers on the landscape for many terrestrial organisms (3, 4). Hydrologic connectivity (1) 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 (5, 6).

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 (7), which may restrict natural population connectivity (8, 9). River regulation can change flow frequency, magnitude, duration, timing, and rate of change, which can have significant impacts on aquatic organisms and ecological processes (5, 7). River regulation, and more specifically, regulation associated with hydropower generation, has been implicated as a cause of fundamental changes to downstream aquatic ecosystems (10–12). The hydrological regimes of over half of the world's largest rivers have been altered by large dams (13) and only recently has the extent of flow alteration and the associated ecosystem-level impacts been acknowledged (14–16).

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 (17–21), reductions in population abundances and diversity (19–25), and fragmentation (21–26) 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 (25). 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" (27), 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 (28). 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 (29). 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 (5, 30–32). 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 (33, 34). In California, particularly in the Sierra Nevada, river regulation may be a significant environmental stressor (17, 20). 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.

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 S1). 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 (35). 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).

Table 1. Sampling and locality information for population genomic analysis of FYLF in the Yuba, Bear, and American Watersheds in the northern Sierra Nevada of California, USA. The number of individuals (n) is given for the total number sequenced per location and the number of individuals that were retained after filtering across the 8,533 baits. NHD refers to the National Hydrography Dataset by USGS (U.S. Geological Survey, National Hydrography Dataset, Digital data, accessed, August 2017).

Locality	Site ID	River	Watershed	Regulation Type	Lat.	Long.	Elev. (m)	NHD Stream Order	NHD Total Drainage Area (sq.km)	n initial	n retained
American Canyon	1	MF American	American	Hydropeaking	38.934	-120.944	240.393	2	9.0	16	6
Gas Canyon	2	MF American	American	Hydropeaking	38.967	-120.933	241.851	1	13.0	6	6
Todd Creek	3	MF American	American	Hydropeaking	38.964	-120.922	367.726	2	10.0	11	9
NFMFA-Skunk Canyon	4	MF American	American	Hydropeaking	39.022	-120.737	521.572	2	6.0	18	18
Rubicon-US Powerhouse	5	MF American	American	Bypass	38.999	-120.723	360.576	5	816.0	11	11
Rubicon-Long Canyon	6	MF American	American	Bypass	38.989	-120.690	415.103	5	806.0	9	8
Ponderosa Bridge	7	NF American	American	Unregulated	39.000	-120.941	240.826	5	857.0	5	5
Bunch Canyon	8	NF American	American	Unregulated	39.038	-120.910	286.287	3	27.0	15	14
Shirttail Creek	9	NF American	American	Unregulated	39.044	-120.899	525.759	4	141.0	16	15
Indian Creek	10	NF American	American	Unregulated	39.057	-120.909	296.107	2	24.0	12	11
Slaughter Ravine	11	NF American	American	Unregulated	39.099	-120.925	356.079	2	6.0	8	8
Robbers Ravine	12	NF American	American	Unregulated	39.105	-120.927	400.000	1	4.0	30	11
Iowa Hill Mainstem	13	NF American	American	Unregulated	39.111	-120.917	386.471	4	605.0	36	30
Euchre Bar	14	NF American	American	Unregulated	39.185	-120.762	579.519	4	508.0	13	11
Sailor Canyon	15	NF American	American	Unregulated	39.217	-120.496	1005.578	3	166.0	8	5
Greenhorn Creek	16	Bear	Bear	Bypass	39.232	-120.902	820.535	2	11.0	15	6
Steep Hollow Creek DS	17	Bear	Bear	Bypass	39.202	-120.875	736.851	2	52.0	13	12
Hawkins Ravine	18	Bear	Bear	Bypass	39.188	-120.898	706.996	2	4.0	3	3
Steep Hollow Creek US	19	Bear	Bear	Bypass	39.194	-120.888	704.658	2	45.0	7	6
Chicago Powerhouse	20	Bear	Bear	Bypass	39.175	-120.900	665.799	4	136.0	6	6
Shady Creek	21	South Yuba	Yuba	Bypass	39.354	-121.059	675.025	2	15.0	14	12
Spring Creek	22	South Yuba	Yuba	Bypass	39.332	-120.989	595.105	3	24.0	4	4
Rock Creek	23	South Yuba	Yuba	Bypass	39.330	-120.986	593.521	4	710.0	3	3
Missouri Canyon	24	South Yuba	Yuba	Bypass	39.361	-120.881	1094.631	2	5.0	8	6
Logan Creek	25	South Yuba	Yuba	Bypass	39.369	-120.853	1201.179	1	5.0	5	4
US Canyon Creek	26	South Yuba	Yuba	Bypass	39.354	-120.734	889.774	4	365.0	6	6
Oregon Creek	27	Middle Yuba	Yuba	Bypass	39.442	-121.058	620.425	4	375.0	15	13
US Our House Dam	28	Middle Yuba	Yuba	Bypass	39.413	-120.990	624.147	4	375.0	13	12
Rocky Rest Mainstem	29	North Yuba	Yuba	Unregulated	39.512	-120.977	704.686	5	669.0	15	12
Slate Creek	30	North Yuba	Yuba	Bypass	39.689	-120.939	1330.937	3	58.7	4	4

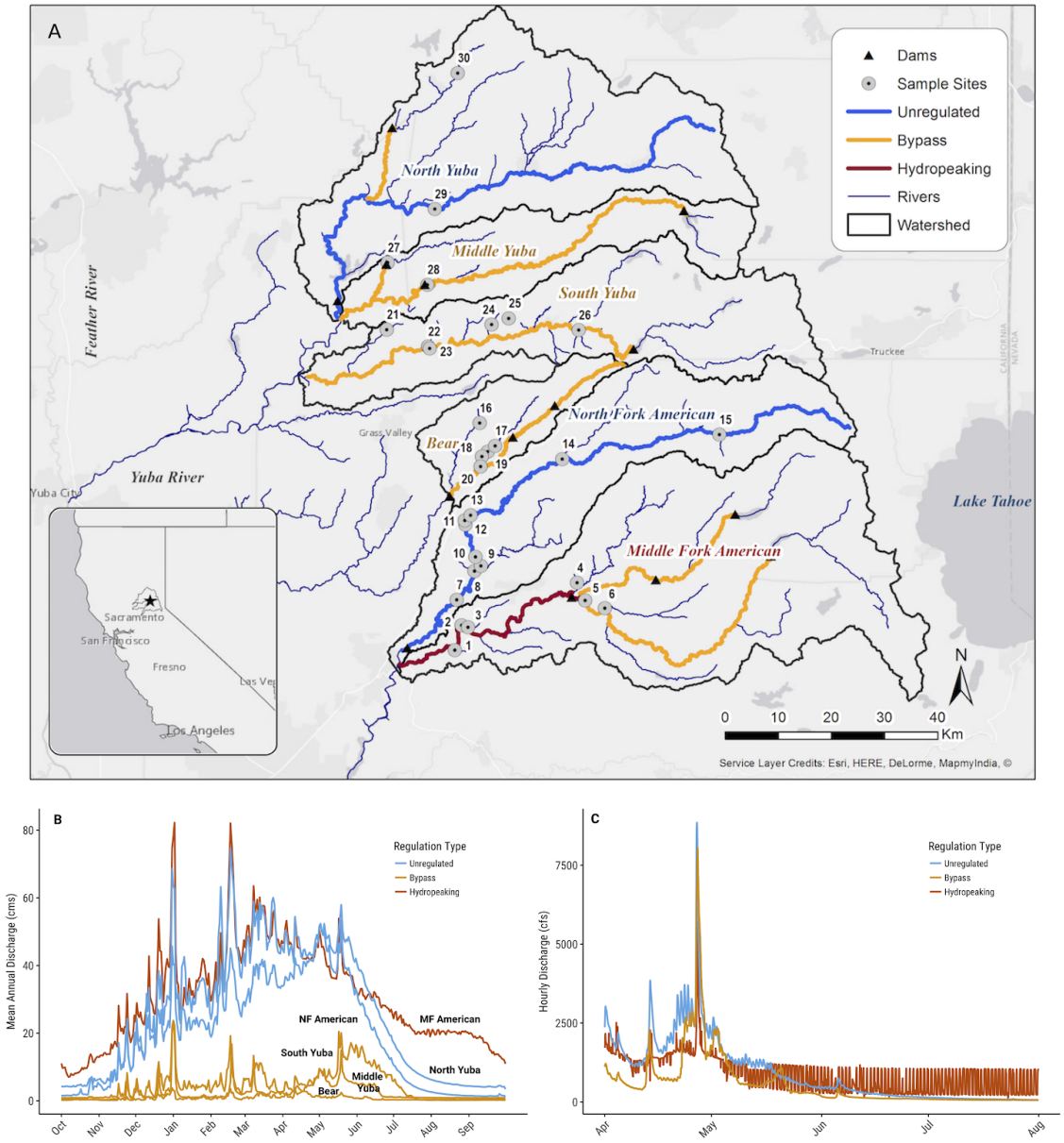

Figure 1. Sampling locations and flow characteristics. A) Map of sampling locations spread across six rivers. B) Comparison of annual mean daily discharge from 1981–2016 for three flow types. C) Comparison of hourly discharge in three different flow regimes in April through July 2012, Bypass (South Yuba), Hydropeaking (Middle Fork American), and Unregulated (North Fork American). See Table S1 for USGS gaging station information.

To generate genetic data from the samples, we performed RAD Capture (a.k.a. Rapture) (36) 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 (37), and genomic analyses in non-model organism often use fewer loci (38). 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 (39) 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.

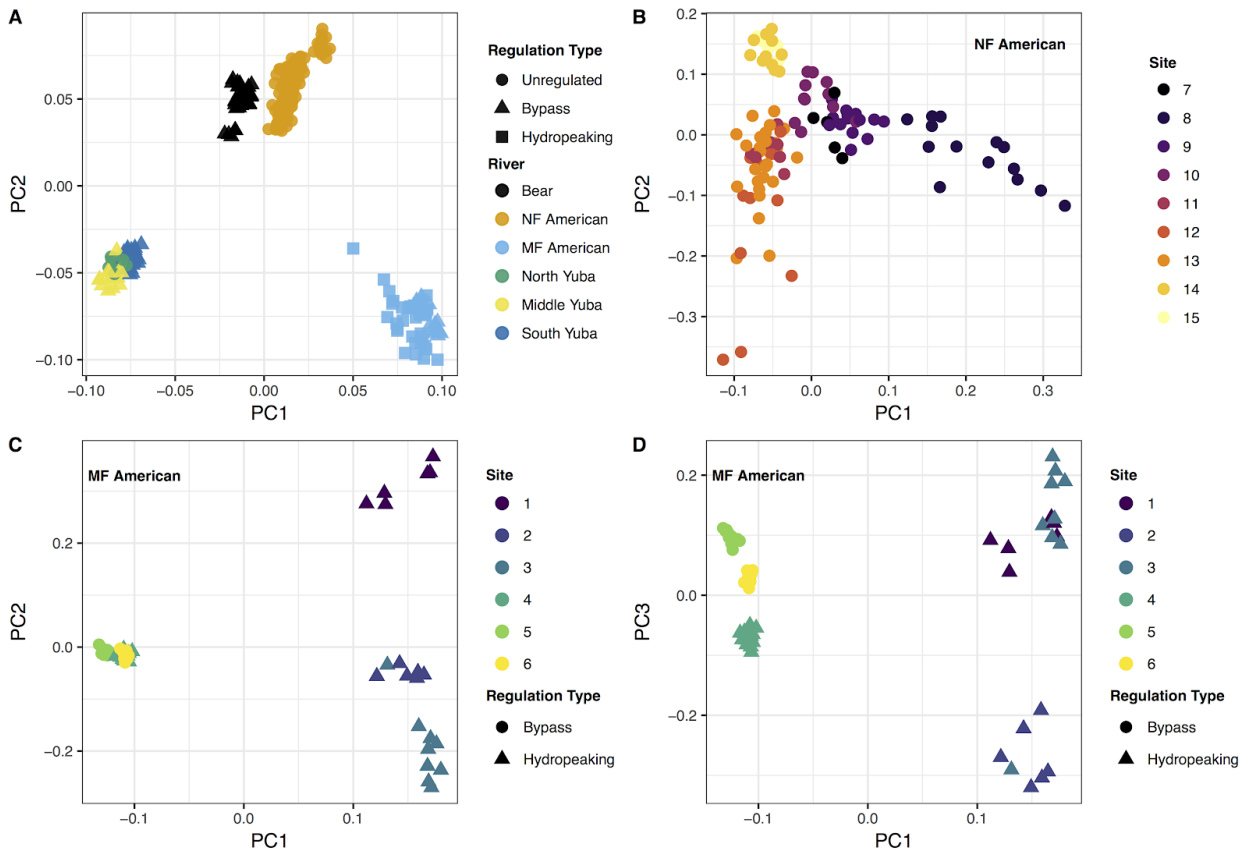

Figure 2. Principal component analysis of Rapture sequencing data. A) Northern Sierra Nevada (n=277) watersheds and regulation types; B) Unregulated NF American; C) and D) Hydropeaking MF American Reach.

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 F_{ST} (40) between all collection locations within a river for all six rivers. We then plotted the scaled mean pairwise F_{ST} [mean F_{ST} / (1-mean F_{ST})] (41) 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 F_{ST} and river distance (as shown by the slope of regression lines in Figure 3A), there was a striking pattern of elevated F_{ST} by regulation type (Figure 3A). Even the bypass regulation type showed a distinct pattern of elevated F_{ST}. For instance, regulated rivers with locations separated by less than 10km had F_{ST} values comparable to unregulated locations separated by mean river distances over 30 km. Hydropeaking was the most extreme pattern of the three regulation types and showed highly elevated F_{ST} values with the steepest regression coefficient. The baseline F_{ST} or global mean increased by over 0.1 between the unregulated (mean F_{ST}=0.141), and regulated locations (global mean for bypass F_{ST}=0.256, hydropeaking F_{ST}=0.278). This suggests a greater degree of isolation within sites in regulated river reaches compared with FYLF populations in unregulated reaches as larger F_{ST} values represent reductions in heterozygosity due to population subdivision (42). We conclude FYLF in regulated rivers show patterns of greater population isolation and loss of heterozygosity compared to frogs in unregulated locations.

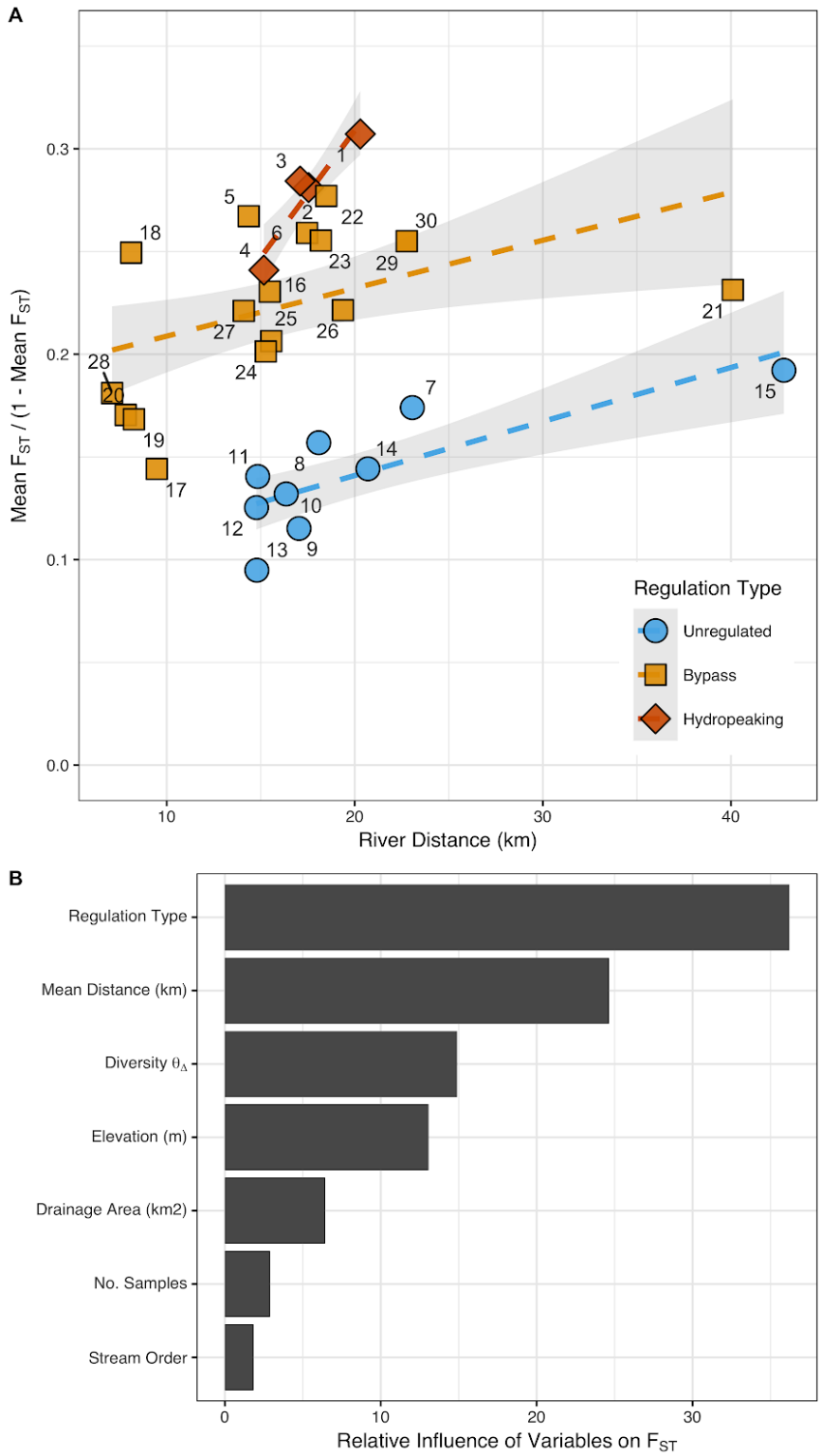

Figure 3. Relationship between river regulation and genetic differentiation in FYLF. A) Mean pairwise F_{ST} vs. mean river distance for each location; B) Relative influence of variables on F_{ST} from boosted regression tree models.

To investigate the relative influence of river regulation compared to other covariates such as river distance on genetic differentiation (i.e. F_{ST}), we used boosted regression tree (BRT) modeling. Covariates included flow regime alteration type, river distance, watershed variables derived from National hydrology data (NHD), topographic data, and allele frequency spectrum skew (see below, Methods). We found flow regulation explained the greatest amount of variance in F_{ST} (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 (40–42). 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 θ ($4N\boldsymbol{\mu}$): Tajima's θ (θ_π) is based on the average number of pairwise differences (43), and Watterson's θ (θ_S) is based on the number of segregating sites (44) (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, $\theta_\pi > \theta_S$; and when genetic diversity has been decreasing, $\theta_S > \theta_\pi$. We found zero populations sampled within regulated watersheds had evidence of increasing genetic diversity (e.g., a $\theta_\pi - \theta_S$ that was less than zero) (Figure 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 $\Delta\theta$ ($\theta_\pi - \theta_S$), and the global mean was significantly different from other regulation types. Although some tributary populations within unregulated watersheds also showed signs of genetic diversity loss, the mean genetic diversity trajectory at unregulated locations was largely neutral (Figure 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.

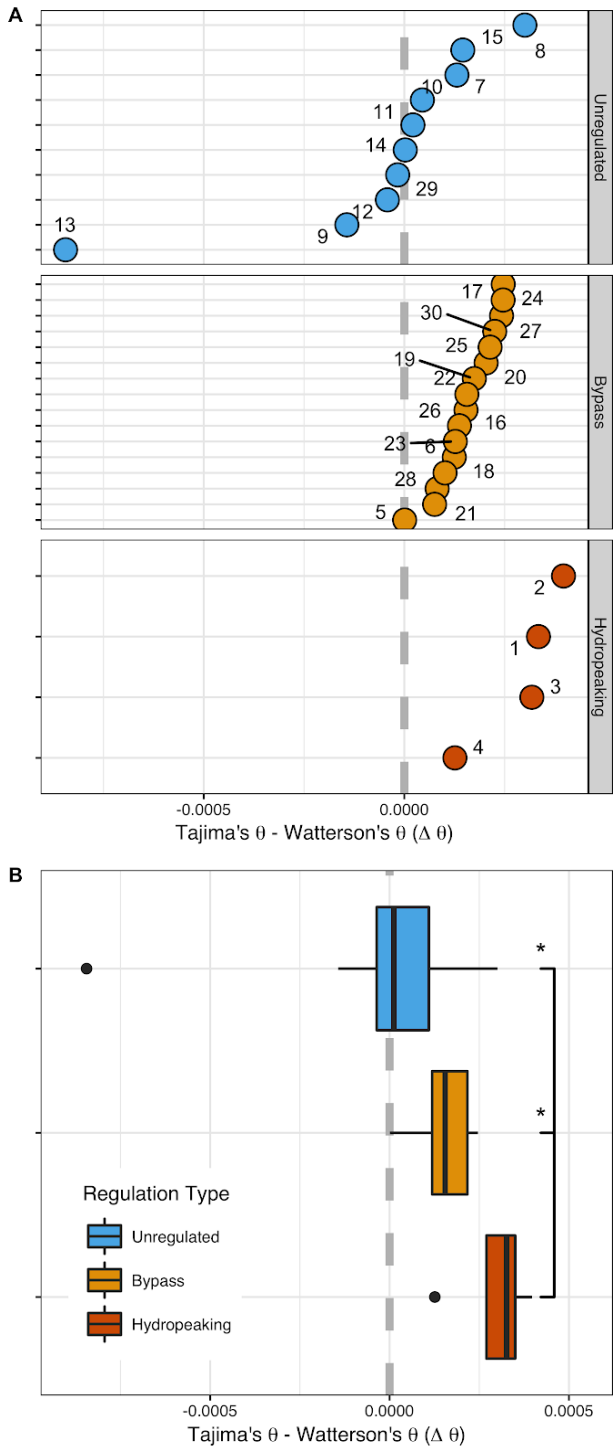

Figure 4. Relationship between river regulation and genetic diversity trajectory in FYLF. A) Assessment of genetic diversity trajectories using Δθ (Tajima's θ_π minus Watterson's θ_S) for each sampling location; B) Boxplots of difference between θ_π - θ_S and pairwise significance between regulation groups using a pairwise Wilcoxon rank sum test with bonferroni correction ($P < 0.05$). Negative values represent trends of increasing genetic diversity, positive values represent trajectories of diversity loss, values near zero are stable.

Discussion

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 (45)for example, there are only two frog reference genome assemblies available as of 2018 (46, 47). Our results demonstrate that Rapture (36) 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 (36) 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 (48, 49)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 (50, 51). 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 (52)), the extent to which current population connectivity has been lost due to river regulation remains unknown. Examining pairwise F_{ST}, revealed a major decrease in connectivity in populations in regulated systems, even with limited river regulation (i.e., bypass reaches). Usually isolation by distance patterns best describe variation in genetic data, yet the primary factor influencing genetic differentiation among these rivers is hydrologic alteration (Figure 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 (53–56). In some cases, isolated populations can maintain genetic diversity when they are sufficiently sized (57), however, species of conservation concern typically have small and/or declining populations and thus may be susceptible to genetic diversity loss (53, 58). 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 hydrologic seasonality and predictability, despite inter-annual variation. Flow regulation has altered patterns of hydrologic seasonality and predictability in many watersheds (17). 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. (59) 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 (59, 60). 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.

Methods

Sampling and Study Sites

345 FYLF buccal or tissue samples were used in this study (see Table S2). Field sampling was conducted as previously described (61), under CDFW SCP Permit #0006881, with IACUC protocol #19327. Individual post-metamorphic frogs were buccal-swabbed following established protocols (62–64). 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 (65, 66). 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 (67, 68). Paired-end sequence data were generated from 24 FYLF individuals (sampling details given in Table S3) across coastal and Sierra Nevada populations from California, USA. DNA was extracted with a magnetic bead–based protocol (36) 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 (36). De novo loci discovery and contig extension were carried as previously described (50) using the alignment program Novoalign and the genome assembler PRICE (69). This pipeline resulted in a set of 77,544 RAD contigs ranging from 300 to 800 bp (Table S4) 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) (36) using 8,533 RAD capture baits (120 bp) were designed by Arbor Biosciences from the de novo alignment (Table S5). The final Rapture library was sequenced in 50% of an Illumina HiSeq 3000 lane. Rapture sequence data from each individual (Table S2) were aligned against the de novo partial genome reference using the BWA-MEM algorithm (70, 71) 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 (72). 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 (73, 74). All Rapture analyses were conducted using Analysis of Next Generation Sequencing Data (ANGSD) (39). 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) (75). To maximize data quality, samples with less than 100,000 aligned reads were excluded (Table S2, S3) 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) (76), 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 (77).

Genetic Differentiation Estimates

Mean scaled F_{ST} was used to quantify genetic differentiation between populations (40, 41). Genome-wide F_{ST} between population pairs was estimated by first calculating a site frequency spectrum (SFS) for each population (doSaf) (78) with ANGSD. The two-dimensional SFS and global F_{ST} between each population pair were then estimated using realSFS (39). F_{ST} was calculated between each pair of collection locations within a watershed, and the mean of all pairwise calculations within that watershed was

calculated for each location. We calculated the river distances (distance along river network) between locations within watersheds using the riverdist package in R (79), and used the mean pairwise river distance (km) to all other locations within the watershed. These values were plotted and a generalized linear model was fitted (F_{ST} ~ Mean River Distance) in R (77). To calculate Watterson's θ_S (44), and Tajima's θ_π (43), 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 (74).

Boosted regression tree modeling of variance in F_{ST}

We used boosted regression tree (BRT) models with the R packages *gbm* (80) and *dismo* (81) 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 (35, 82). BRTs use iterative boosting algorithms to combine simple decision trees to improve model performance (83) and provide a robust alternative to many traditional statistical methods (84, 85). 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.To evaluate the relative influence of covariates on F_{ST}, models were trained using river distance (km), elevation (m), upstream drainage area (km^2), Strahler stream order, and number of samples per location. Stream segment data on elevation, length, slope, stream order, and drainage area were derived from NHD Plus attributes (U.S. Geological Survey, National Hydrography Dataset, Digital data, accessed, August 2017 at http://nhd.usgs.gov/data.html). In addition, $\Delta\theta$ (θ_π- θ_S) was included to assess the effect of genomic variation on F_{ST} across regulation types.

Model training and fitting were conducted following methods previously described in (35). 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 (83), and due to our relatively small sample size, the fraction of training data sampled to build each tree was 0.75, within the range as recommended by (86). 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, as well as to evaluate model performance. Finally, the

models were evaluated for interactions by comparing the estimated deviance explained for models with first, second, and third-order interactions.

Acknowledgements

This research builds on work from Amy Lind, Sarah Kupferberg, and Sarah Yarnell. Many thanks to all of them for insight, and to many who helped collect/prepare samples: Corey Luna, Kristen Hein Strohm, Rick Wachs, and Sarah Mussulman. This research could not have been conducted without access to specimens from Brad Shaffer at the Krebb Museum, and field samples from Caren Goldberg and Mallory Bedwell.

References

1. Pringle C (2003) What is hydrologic connectivity and why is it ecologically important? *Hydrol Process* 17(13):2685–2689.

2. Wiens JA (2002) Riverine landscapes: taking landscape ecology into the water. *Freshw Biol* 47(4):501–515.

3. Voelker G, et al. (2013) River barriers and cryptic biodiversity in an evolutionary museum. *Ecol Evol* 3(3):536–545.

4. Cazé ALR, et al. (2016) Could refuge theory and rivers acting as barriers explain the genetic variability distribution in the Atlantic Forest? *Mol Phylogenet Evol* 101:242–251.

5. Yarnell SM, Viers JH, Mount JF (2010) Ecology and Management of the Spring Snowmelt Recession. *Bioscience* 60(2):114–127.

6. Tonkin JD, Bogan MT, Bonada N, Rios-Touma B, Lytle DA (2017) Seasonality and predictability shape temporal species diversity. *Ecology* 98(5):1201–1216.

7. Poff NL, Olden JD, Merritt DM, Pepin DM (2007) Homogenization of regional river dynamics by dams and global biodiversity implications. *Proc Natl Acad Sci U S A* 104(14):5732–5737.

8. Shaw EA, Lange E, Shucksmith JD, Lerner DN (2016) Importance of partial barriers and temporal variation in flow when modelling connectivity in fragmented river systems. *Ecol Eng* 91(Supplement C):515–528.

9. Schick RS, Lindley ST (2007) Directed connectivity among fish populations in a riverine network. *J Appl Ecol* 44(6):1116–1126.

10. Power ME, Dietrich WE, Finlay JC (1996) Dams and Downstream Aquatic Biodiversity: Potential Food Web Consequences of Hydrologic and Geomorphic

Change. *Environ Manage* 20(6):887–895.

11. Bunn SE, Arthington AH (2002) Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. *Environ Manage* 30(4):492–507.

12. Moyle PB, Katz JVE, Quiñones RM (2011) Rapid decline of California's native inland fishes: A status assessment. *Biol Conserv* 144(10):2414–2423.

13. Nilsson C, Reidy CA, Dynesius M, Revenga C (2005) Fragmentation and flow regulation of the world's large river systems. *Science* 308(5720):405–408.

14. Pringle CM (2001) Hydrologic Connectivity and the Management of Biological Reserves: A Global Perspective. *Ecol Appl* 11(4):981–998.

15. Dudgeon D, et al. (2006) Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol Rev Camb Philos Soc* 81(2):163–182.

16. Murchie KJ, et al. (2008) Fish response to modified flow regimes in regulated rivers: research methods, effects and opportunities. *River Res Appl* 24(2):197–217.

17. Kupferberg SJ, et al. (2012) Effects of flow regimes altered by dams on survival, population declines, and range-wide losses of California river-breeding frogs. *Conserv Biol* 26(3):513–524.

18. Rolls RJ, Bond NR (2017) Environmental and Ecological Effects of Flow Alteration in Surface Water Ecosystems. *Water for the Environment* (Elsevier), pp 65–82.

19. Fuller TE, Pope KL, Ashton DT, Welsh HH (2011) Linking the Distribution of an Invasive Amphibian (Rana catesbeiana) to Habitat Conditions in a Managed River System in Northern California. *Restor Ecol* 19(201):204–213.

20. Lind AJ, Welsh HH Jr, Wilson RA (1996) The effects of a dam on breeding habitat and egg survival of the foothill yellow-legged frog (Rana boylii) in Northwestern Calfifornia. *Herpetol Rev* 27(2):62–66.

21. Zhong Y, Power G (1996) Environmental impacts of hydroelectric projects on fish resources in China. *Regul Rivers: Res Mgmt* 12(1):81–98.

22. Vörösmarty CJ, et al. (2010) Global threats to human water security and river biodiversity. *Nature* 467(7315):555–561.

23. Guzy JC, Eskew EA, Halstead BJ, Price SJ (2018) Influence of damming on anuran species richness in riparian areas: A test of the serial discontinuity concept. *Ecol Evol*. doi:10.1002/ece3.3750.

24. Sabo JL, et al. (2017) Designing river flows to improve food security futures in the

Lower Mekong Basin. *Science* 358(6368). doi:10.1126/science.aao1053.

25. Scribner KT, et al. (2016) Applications of Genetic Data to Improve Management and Conservation of River Fishes and Their Habitats. *Fisheries* 41(4):174–188.

26. Werth S, Schödl M, Scheidegger C (2014) Dams and canyons disrupt gene flow among populations of a threatened riparian plant. *Freshw Biol* 59(12):2502–2515.

27. Gilpin M, Soule M (1986) Minimum Viable Populations: Processes of Species Extinction, Conservation Biology. *Sunderland, Massachusetts: Sin Assoc*:19–34.

28. Grantham TE, Merenlender AM, Resh VH (2010) Climatic influences and anthropogenic stressors: an integrated framework for streamflow management in Mediterranean-climate California, USA. *Freshw Biol* 55(s1):188–204.

29. Stebbins RC (2003) *A Field Guide to Western Reptiles and Amphibians* (Houghton Mifflin Harcourt, Boston). 3rd Ed.

30. Kupferberg SJ (1996) Hydrologic and geomorphic factors affecting conservation of a river-breeding frog (Rana boylii). *Ecol Appl* 6(4):1332–1344.

31. Bondi CA, Yarnell SM, Lind AJ, Lind A (2013) Transferability of habitat suitability criteria for a stream breeding frog (*Rana boylii*) in the Sierra Nevada, California. *Conservation and Biology* 8(1):88–103.

32. Yarnell S, Peek R, Epke G, Lind A (2016) Management of the Spring Snowmelt Recession in Regulated Systems. *J Am Water Resour Assoc* 52(3):723–736.

33. Jennings MR, Hayes MP (1994) *Amphibian and reptile species of special concern in California* (California Department of Fish and Game, Rancho Cordova).

34. Davidson C, Shaffer HB, Jennings MR (2002) Spatial Tests of the Pesticide Drift, Habitat Destruction, UV-B, and Climate-Change Hypotheses for California Amphibian Declines. *Conserv Biol* 16(6):1588–1601.

35. Steel AE, Peek RA, Lusardi RA, Yarnell SM (2017) Associating metrics of hydrologic variability with benthic macroinvertebrate communities in regulated and unregulated snowmelt-dominated rivers. *Freshw Biol*. doi:10.1111/fwb.12994.

36. Ali OA, et al. (2016) RAD Capture (Rapture): Flexible and Efficient Sequence-Based Genotyping. *Genetics* 202(2):389–400.

37. Hotaling S, et al. (2018) Demographic modelling reveals a history of divergence with gene flow for a glacially tied stonefly in a changing post-Pleistocene landscape. *J Biogeogr* 45(2):304–317.

38. Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013)

Genotyping-by-sequencing in ecological and conservation genomics. *Mol Ecol* 22(11, Sp. Iss. SI):2841–2847.

39. Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15:356.

40. Wright S (1943) Isolation by Distance. *Genetics* 28(2):114–138.

41. Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145(4):1219–1228.

42. Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236(4803):787–792.

43. Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105(2):437–460.

44. Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 7(2):256–276.

45. Nunziata SO, Lance SL, Scott DE, Lemmon EM, Weisrock DW (2017) Genomic data detect corresponding signatures of population size change on an ecological time scale in two salamander species. *Mol Ecol* 26(4):1060–1074.

46. Sun Y-B, et al. (2015) Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes. *Proc Natl Acad Sci U S A* 112(11):E1257–62.

47. Hellsten U, et al. (2010) The genome of the Western clawed frog Xenopus tropicalis. *Science* 328(5978):633–636.

48. Fahrig L, Merriam G (1985) Habitat Patch Connectivity and Population Survival. *Ecology* 66(6):1762–1768.

49. Taylor PD, Fahrig L, Henein K, Merriam G (1993) Connectivity Is a Vital Element of Landscape Structure. *Oikos* 68(3):571–573.

50. Miller MR, et al. (2012) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Mol Ecol* 21(2):237–249.

51. Prince DJ, et al. (2017) The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Sci Adv* 3(8):e1603198.

52. Bourque RM (2008) Spatial ecology of an inland population of the Foothill yellow-legged frog (Rana boylii) in Tehama County, California. MSc (Humboldt State

University). Available at: http://scholarworks.calstate.edu/handle/2148/437.

53. Frankham R (2002) *Introduction to Conservation Genetics* (Cambridge University Press, Cambridge, UK ; New York).

54. Ishiyama N, Koizumi I, Yuta T, Nakamura F (2015) Differential effects of spatial network structure and scale on population size and genetic diversity of the ninespine stickleback in a remnant wetland system. *Freshw Biol* 60(4):733–744.

55. Lande R, Shannon S (1996) THE ROLE OF GENETIC VARIATION IN ADAPTATION AND POPULATION PERSISTENCE IN A CHANGING ENVIRONMENT. *Evolution* 50(1):434–437.

56. Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. *Nature* 470(7335):479–485.

57. Whiteley AR, et al. (2010) Genetic variation and effective population size in isolated populations of coastal cutthroat trout. *Conserv Genet* 11(5):1929–1943.

58. Krohn AR, et al. (2018) Conservation genomics of desert dwelling California voles (Microtus californicus) and implications for management of endangered Amargosa voles (Microtus californicus scirpensis). *Conserv Genet* 19(2):383–395.

59. Botero CA, Weissing FJ, Wright J, Rubenstein DR (2015) Evolutionary tipping points in the capacity to adapt to environmental change. *Proc Natl Acad Sci U S A* 112(1):184–189.

60. Kahilainen A, Puurtinen M, Kotiaho JS (2014) Conservation implications of species–genetic diversity correlations. *Global Ecology and Conservation* 2:315–323.

61. Heyer WR, Donnelly MA, McDiarmid RW, Hayek L-AC, Foster MS (1994) *Measuring and monitoring biological diversity. Standard methods for amphibians* (Smithsonian Institution Press, Washington DC).

62. Broquet T, Berset-Braendli L, Emaresi G, Fumagalli L (2007) Buccal swabs allow efficient and reliable microsatellite genotyping in amphibians. *Conserv Genet* 8(2):509–511.

63. Pidancier N, Miquel C, Miaud C (2003) Buccal swabs as a non-destructive tissue sampling method for DNA analysis in amphibians. *Herpetol J* 13(4):175–178.

64. Goldberg CS, Kaplan ME, Schwalbe CR (2003) From the frog's mouth: buccal swabs for collection of DNA from amphibians. *Herpetol Rev* 34(3):220–221.

65. Parris KM, et al. (2010) Assessing ethical trade-offs in ecological field studies. *J*

Appl Ecol 47(1):227–234.

66. Wilbur HM, Semlitsch RD (1990) Ecological Consequences of Tail Injury in Rana Tadpoles. *Copeia* 1990(1):18–24.

67. Miller MR, et al. (2007) RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biol* 8(6):R105.

68. Baird NA, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3(10):e3376.

69. Ruby JG, Bellare P, Derisi JL (2013) PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. *G3* 3(5):865–880.

70. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.

71. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bioGN]*. Available at: http://arxiv.org/abs/1303.3997.

72. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

73. Fumagalli M, et al. (2013) Quantifying population genetic differentiation from next-generation sequencing data. *Genetics* 195(3):979–992.

74. Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R (2013) Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* 14:289.

75. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21):2987–2993.

76. Kim SY, et al. (2011) Estimation of allele frequency and association mapping using next-generation sequencing data. *BMC Bioinformatics* 12:231.

77. R Core Team (2017) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria) Available at: https://www.R-project.org/.

78. Nielsen R, Korneliussen T, Albrechtsen A, Li Y, Wang J (2012) SNP calling, genotype calling, and sample allele frequency estimation from New-Generation Sequencing data. *PLoS One* 7(7):e37558.

79. Tyers M (2017) *riverdist: River Network Distance Computation and Applications*

Available at: https://CRAN.R-project.org/package=riverdist.

80. Ridgeway G (2015) *gbm: Generalized Boosted Regression Models* (http://CRAN.R-project.org/package-gbm).

81. Hijmans RJ, Phillips S, Leathwick J, Elith J (2017) *dismo: Species Distribution Modeling* (R package) Available at: https://CRAN.R-project.org/package=dismo.

82. Graham CH, et al. (2008) The influence of spatial errors in species occurrence data used in distribution models. *J Appl Ecol* 45(1):239–247.

83. De'ath G (2007) Boosted trees for ecological modeling and prediction. *Ecology* 88(1):243–251.

84. Guisan A, et al. (2007) What Matters for Predicting the Occurrences of Trees: Techniques, Data, or Species' Characteristics? *Ecol Monogr* 77(4):615–630.

85. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Modell* 190(3):231–259.

86. Brown LR, et al. (2012) Predicting biological condition in southern California streams. *Landsc Urban Plan* 108(1):17–27.

87. Brown LR, et al. (2012) Predicting biological condition in southern California streams. *Landsc Urban Plan* 108(1):17–27.