Conservation genomics of desert dwelling California voles (Microtus californicus) and implications for management of endangered Amargosa voles (Microtus californicus scirpensis)

Article *in* Conservation Genetics · August 2017
DOI: 10.1007/s10592-017-1010-2

6 authors, including:

Alexander Krohn
University of California, Berkeley
8 PUBLICATIONS **15** CITATIONS
SEE PROFILE

Chris J. Conroy
University of California, Berkeley
38 PUBLICATIONS **1,889** CITATIONS
SEE PROFILE

Risa Pesapane
University of California, Davis
7 PUBLICATIONS **64** CITATIONS
SEE PROFILE

Conserv Genet
DOI 10.1007/s10592-017-1010-2

RESEARCH ARTICLE

Conservation genomics of desert dwelling California voles (*Microtus californicus*) and implications for management of endangered Amargosa voles (*Microtus californicus scirpensis*)

Alexander R. Krohn[1,2] · Chris J. Conroy[2] · Risa Pesapane[3] · Ke Bi[2,4] · Janet E. Foley[3] · Erica Bree Rosenblum[1,2]

Received: 16 March 2017 / Accepted: 18 August 2017

Abstract Understanding population genetic structure and levels of genetic variation is critical for the conservation and management of imperiled populations, especially when reintroductions are planned. We used restriction-site associated DNA (RAD) sequencing to study the genetic diversity and evolutionary relationships of the endangered Amargosa vole and other closely related desert-dwelling California voles. Specifically, we sought to determine how Amargosa voles are related to other California voles, how genetic variation is partitioned among subpopulations in wild Amargosa voles, and how much genetic variation is captured within a captive insurance colony of Amargosa voles. Our multilocus nuclear dataset provides strong evidence that Amargosa voles are part of a northern clade of California voles. Amargosa voles have highly reduced genetic variation relative to other California voles, but do exhibit some sub-structure among sampled marshes. Captive Amargosa voles capture approximately half of the total genetic variation present in the wild Amargosa vole populations. We discuss the management implications of our findings in light of reintroductions planned for Amargosa voles. Our study highlights the utility of reduced representation genomic approaches, like RADseq, to resolve relationships among small populations that are difficult to study with traditional markers due to low genetic variation and few individuals left in the wild.

Keywords Captive breeding programs · Endangered species · Population structure · RADseq

Electronic supplementary material The online version of this article (doi:10.1007/s10592-017-1010-2) contains supplementary material, which is available to authorized users.

✉ Alexander R. Krohn
alex.krohn@berkeley.edu

[1] Department of Environmental, Science, Policy and Management, University of California, Berkeley, Berkeley, CA 94720-3114, USA

[2] Museum of Vertebrate Zoology, University of California, Berkeley, 3101 Valley Life Sciences Building, Berkeley, CA 94720-3160, USA

[3] Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, Davis, CA 95616, USA

[4] Computational Genomics Resource Laboratory (CGRL), California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, Berkeley, CA 97420-3102, USA

Introduction

Understanding population genetic structure and the amount of genetic diversity present in imperiled taxa is crucially important for conservation management. One key application of population genetics to conservation is when captive breeding and reintroductions are part of an endangered species recovery plan. For example, genetic data can help scientists and managers select donor populations for captive breeding and recipient populations for reintroductions (e.g., Miller et al. 2010). Moreover, genetic analyses can be used to devise a breeding strategy focused on preserving existing population genetic structure across a species range and/or increasing genetic variation through outbreeding. Risks associated with both inbreeding and outbreeding strategies (e.g., Edmands 2007; Tallmon et al. 2004) make genetic data particularly important for evaluating which reintroduction strategy is most appropriate for a species (e.g., Madsen et al. 1999; Tordoff and Redig 2001; Norton and Ashley 2004). Population genetic analyses of the relationships among imperiled taxa and sister taxa can also provide important

Published online: 21 August 2017

historical and evolutionary context for management efforts. For example, by comparing the amount of genetic variation in imperiled taxa and sister taxa, scientists and managers can assess how much genetic variation has been lost in focal taxa, and whether population sizes have been reduced, relative to close relatives that are not threatened (Docker and Heath 2003; Funk et al. 2016). Moreover, knowledge of evolutionary relationships among taxa can guide reintroductions of the closest relatives if extirpation of the imperiled group becomes unavoidable.

The Amargosa vole (*Microtus californicus scirpensis*) is an endangered subspecies of the California vole (*Microtus californicus*). Although California voles as a whole are not a species of conservation concern, isolated desert subspecies such as the Mojave River vole (*Microtus californicus mohavensis*) and the Amargosa vole are imperiled. While California voles range from southern Oregon to northern Baja California (Kellogg 1918), the Amargosa vole is geographically highly restricted and was thought to be extinct in the Twentieth Century (Kellogg 1918). The Amargosa vole was listed as endangered by the state of California in 1980 and at the U.S. federal level in 1984, after researchers confirmed the existence of a surviving population near Tecopa, CA in 1970s (USFWS 2009). Like other California voles found in the desert, the Amargosa vole is restricted to isolated mesic areas in an otherwise parched landscape. Amargosa voles in particular are tied to marshes where Olney's three-square bulrush (*Schoenoplectus americanus*), the voles' primary food source, is found. Despite conservation concern for the species, a number of outstanding questions remain about the phylogenetic context and population structure of Amargosa voles.

At the phylogenetic level, the relationships among lineages of California voles and specifically the position of Amargosa voles are currently not well resolved. Taxonomically, the California vole was initially split into 17 subspecies (Kellog 1918; Hall 1981). However, recent work has shown that California voles form distinct northern and southern mitochondrial clades, and these two clades do not correspond well to the original subspecies designations (Conroy and Neuwald 2008; Conroy et al. 2016). Moreover, previous work established that the Amargosa vole was a genetically distinct clade, but did not conclusively determine whether it belonged to the northern or southern clade of California vole (Conroy et al. 2016). Specifically, mitochondrial evidence situated Amargosa voles in the southern clade, while nuclear data suggested affinity to the northern clade. The authors hypothesized that Amargosa voles may be of hybrid origin, but inferences were based on few markers and lacked morphological data.

At the population genetic level, it is essential to understand population genetic structure and levels of genetic diversity in Amargosa voles for translocations of wild voles and for reintroduction of captive voles to the wild. When they were rediscovered, Amargosa voles were only known to occupy seven marshes around Tecopa, CA (USFWS 2009). Intensive population monitoring of Amargosa vole populations began in earnest around 2008. The Amargosa vole population size fluctuated dramatically, but by 2013 extensive trapping over the vole's peak activity times resulted in the capture of less than 200 voles (Klinger et al. 2013). In 2014, researchers noticed that the habitat in the marsh with the largest population (Marsh 1) was rapidly drying. Given that this rapidly disappearing marsh represented the majority of Amargosa voles left in the wild, managers decided to remove 20 voles from Marsh 1 to start a captive insurance colony for eventual reintroduction into the wild (Foley 2016). However, little was known about how wild Amargosa vole populations were structured genetically, how much genetic variation was left in the wild, and how much of that genetic variation was captured in the captive colony.

In this study, we used a multilocus restriction-site associated (RAD) sequencing approach to address three nested aims. Our first aim was to determine the phylogeographic structure of desert California vole populations and the affinity of Amargosa voles within this group. Our second aim was to inform the conservation of Amargosa voles by quantifying population structure and genetic diversity of wild Amargosa voles. Our third aim was to inform the management of the captive colony by assessing how much genetic diversity is captured by the captive colony. Addressing these aims together provides essential population genetic context for the conservation and recovery of imperiled Amargosa voles.

Materials and methods

Sampling and RAD sequencing

We extracted DNA from 94 *M. californicus* tissue or blood samples using DNeasy extraction kits (Qiagen, Valencia, CA, USA). We included 66 Amargosa voles and 28 California voles that were not Amargosa voles, hereafter referred to as desert California voles, from the northern (Deep Springs, n = 7; Lark Seep, n = 5; Lake Isabella, n = 5) and southern (Victorville, n = 7; Elizabeth Lake, n = 4; Fig. 1; MVZ specimens used listed in Online Resource 1) mtDNA clades. We included four to seven Amargosa vole individuals per marsh, except for Marsh 1, where our sampling included 19 wild individuals, 10 of which were founding members of the colony. The founding members were treated as the founders of the colony, and not as wild Amargosa voles. The 66 Amargosa voles included in this study represent between 29–88% of all known Amargosa voles [based on census data from Klinger et al. (2013) and Klinger (2015)]. The seven

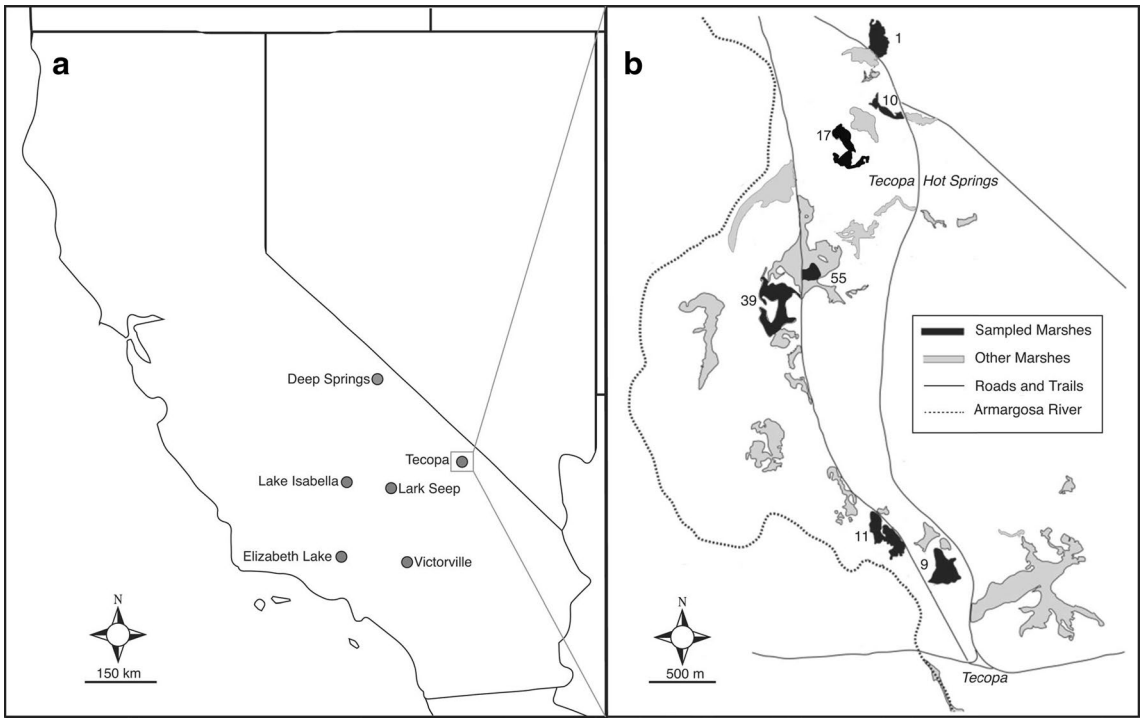

Fig. 1 Map of sampled locations of **a** desert California voles and **b** focal marshes near Tecopa, CA for the Amargosa voles. Figure modified from Neuwald (2010)

marshes from which samples were collected also represent half of all the known marshes where Amargosa voles were captured between 2014 and 2015 (n = 14). Finally, to understand changes in genetic diversity in the captive colony, we also included individuals from the F_1 (n = 2), F_2 (n = 3) and F_3 backcross generation (n = 6) from the captive colony at UC Davis.

We constructed a restriction-site associated DNA (RAD) library using the RAD protocol outlined in Ali et al. (2016) with two modifications. First, we did not do the targeted bait capture (the "Rapture" step), but instead only used the new RAD protocol with biotinylated adapters (a protocol referred to as "bestRAD"). Second, we used Pippin Prep (Sage Science, Beverly, MA, USA) instead of beads to size select fragments between 300 and 800 bp. We sequenced our library on one lane of Illumina HiSeq 4000 at the U.C. Davis Genome Center with 150 bp paired-end reads.

RAD sequence filtering and genotyping

Before beginning our analysis, we filtered our reads to include only loci with high quality data. First, we re-oriented the reverse reads, then used Stacks (Catchen et al. 2013) to demultiplex individuals from our raw reads (while allowing one mismatch in the internal barcode) and to remove PCR duplicates. Next, we followed the RADToolKit pipeline (https://github.com/CGRL-QB3-UCBerkeley/RAD). Briefly, we removed reads that did not include the restriction enzyme cut-site or had more than one error in the restriction enzyme cut-site ("filter" function), and further removed reads that were low quality based on Phred scores, or low complexity, or, had adapter and/or bacterial contamination ("cleanPE" function). Next, we used FLASH (Magoč and Salzberg 2011) to merge forward and reverse reads if they overlapped. At this point we removed any individuals that did not have enough remaining sequences to align to the reference (fewer than 250,000 reads that passed filters, n = 14), leaving 52 Amargosa voles and 28 desert California voles.

We used the Amargosa vole de novo genome assembly created at Purdue University as a reference genome. We aligned cleaned sequence reads from each individual to this reference using Novoalign (http://www.novocraft.com) and kept reads that mapped uniquely to the reference, allowing five high-quality mismatches per read when aligning. We converted the resulting alignment from SAM format to BAM format using SAMtools (Li et al. 2009). For the aligned reads we used Picard (http://picard.sourceforge.net) to add read groups and GATK (McKenna et al. 2010) to perform realignment around indels. Next, we used SAMtools and bcftools to construct a variant call format (VCF) file that contained all individual genotype calls for all sites to filter low-quality sites and create a list of high-confidence sites. We filtered sites from the VCF using a custom script ("10-SNPCleaner" available at https://github.com/

CGRL-QB3-UCBerkeley/denovoTargetCapturePopGen). We removed sites that were not present in 90% of individuals with at least 5X coverage (to minimize the effect of missing data), were not biallelic, were adjacent to indels (within 5 bp), had low mapping or base calling scores, or that had excessive heterozygosity (to purge potential paralogs). Finally, we used ANGSD (Korneliussen et al. 2014) for our data analysis on the filtered set of sites. ANGSD incorporates genotype uncertainty in the form of genotype likelihoods and posterior probability of genotype calls into the analysis and thus gives more accurate genotypes calls from lower coverage data.

Aim 1: determine phylogeographic structure for California voles and phylogeographic affinity for Amargosa voles

To quantify population structure of wild California voles from the northern, southern, and Amargosa clades, we used NGSadmix (Skotte et al. 2013) to compute the admixture proportions of individuals from one to eight possible populations (K = 1 through K = 8), representing one more than the seven total populations in our analysis. Given that uneven sampling of populations can lead to misleading inferences of admixture and population clustering (Kalinowski 2011; Puechmaille 2016; Wang 2017), we subset the Amargosa voles to one randomly chosen individual per marsh (n = 30 total, n = 7 Amargosa voles, n = 2–6 per population for other outgroup populations), as in Puechmaille (2016). To run NGSadmix, we first generated a Beagle file using ANGSD. To create this Beagle file based on our genotype likelihoods, we also removed sites that were not polymorphic (using a likelihood ratio test with a *P*-value of 1e − 6), or had a minor allele frequency below 5%, in addition to all of the above filters. We ran NGSadmix for each value of K ten times, and then used the ΔK method from Evanno et al. (2005), calculated in Clumpak (Kopelman et al. 2015), to determine which K was most likely based on its log-likelihood. If a population was admixed, we classified the population as belonging to the group with which its individuals drew more than 50% of their admixture based on the NGSadmix analyses. The NGSadmix analysis revealed that the most likely grouping of populations for both the scenario with seven Amargosa voles, and when all individuals were included, was K = 2 (see "Results"). Because the ΔK method of Evanno et al. (2005) cannot determine whether K = 2 or K = 1 is more likely, we calculated whether the two groups identified by the NGSadmix analyses were significantly differentiated by comparing the F_{ST} value of those two groups to a null distribution of F_{ST} values generated from random permutations of the individuals in each of the two groups. We calculated F_{ST} using the function of realSFS built into ANGSD. realSFS calculates F_{ST} based on the joint site frequency spectrum of the two populations of interest. In order to generate a null probability distribution of F_{ST} without making prior assumptions about the distribution of F_{ST} values, we generated F_{ST} values from the two populations indicated in the K = 2 finding (8 and 22 individuals each) using parametric bootstrapping. Thus, we calculated F_{ST} for 100 random combinations of 8 and 22 individuals from the northern clade, southern clade and wild Amargosa voles. We then compared the actual value of F_{ST} for the true groupings of individuals when K = 2 to the distribution of random F_{ST} values to determine whether the actual F_{ST} value was significantly higher than we would expect by chance. If the true value was larger than 95% of the values in the null distribution, we considered the true value significant at α = 0.05.

We used a principal component analysis (PCA) to visualize population structure among the northern, southern and Amargosa groups of California voles, initially in all northern, southern and Amargosa voles (n = 54) using the program ngsCovar within the ngsTools package (Fumagalli et al. 2013). ngsCovar creates a covariance matrix of individual's genotypes while accounting for uncertainty in the genotype calls by incorporating genotype likelihood values from ANGSD. After plotting the Eigenvalues of the covariance matrix in R (v. 3.3.1, R Core Team 2016),we calculated 95% confidence intervals around the individuals in our PCA analysis based on their assignment in the NGSadmix analysis (see above). Specifically, we drew ellipses around clusters whose individuals that draw more than 50% of their admixture from the same group, as identified with NGSadmix using the *ellipse* package (Murdoch and Chow 2013).

To quantify which clade Amargosa voles were more closely related to, we used SpaceMix (Bradburd et al. 2016). SpaceMix creates a "geogenetic map" based on geography and genetic makeup of individuals. Using Markov chain Monte Carlo (MCMC) simulations, individuals are placed on the map according to their geographic location. Next, the program moves individuals to another geogenetic position based on the proportion of admixture that they draw from other populations or individuals, and plots those locations on a map. To use SpaceMix, we called SNPs using ANGSD. Given that we stringently filtered sites in our preprocessing steps (described above), we applied less stringent filtering when calling SNPs in ANGSD. We called SNPs that were polymorphic using a likelihood ratio test (with a *P*-value of 1e − 6), and where 70% of individuals (n = 54) had at least 8X coverage, in addition to the above filters. We removed any individuals that were missing data at 50% or more of the loci. We ran the MCMC on a model estimating both geogenetic sampling locations and admixture source locations on a flat plane for 13 million iterations, sampling the chain every 1,000 iterations and saving the model output every 100,000 iterations.

Aim 2: quantify population structure and genetic diversity in wild Amargosa voles

We quantified population structure within the Amargosa vole clade as above, using a PCA and NGSadmix. The analyses were run only for wild Amargosa voles (n = 31). Clustering in the PCA and groups deemed the most likely K in NGSadmix were concordant, made up of 15 and 16 individuals total, and both forming two groups, K = 2 (see "Results"). Again, as above, we calculated F_{ST} to quantify whether the differentiation observed between these two groups was larger than what one would expect to observe by chance. We performed parametric bootstrapping as above to generate a null distribution of F_{ST} values based on 100 random permutations of 15 and 16 individuals to determine whether the two actual groups were significantly differentiated using F_{ST}.

We quantified patterns of migration among the sampled marshes using the program EEMS (Petkova et al. 2016) to estimate effective migration surfaces for our data. Briefly, EEMS uses MCMC simulations to interpolate average migration rates across a given geographic area, based on sampled individuals with known genetic distances. Where migration is less than average, one can infer barriers that may disrupt migration. We compared levels of migration based on genetic distances to a null expectation of isolation by distance (Wright 1943). First, we called SNPs using filters similar to the SpaceMix analysis. We used ANGSD to call SNPs using SNPs present in at least 67% of individuals (n = 21) with at least 8X coverage. We saved the SNPs as a Plink file, and used this, the geographic locations of each individual and a polygon encompassing the total range of marshes around Tecopa as inputs for EEMS. We adjusted the proposal variances for the parameters in the MCMC so that they would be rejected between 10 and 40% of the time, following Petkova et al. (2016). To understand the average migration surface, we ran EEMS with 20, 40, 60, 80 and 100 demes for three runs each (as in Petkova et al. 2016 and; Rick 2015) and averaged over the results. While the number of demes is arbitrary and results are robust to the number of demes chosen (Petkova et al. 2016), we kept the number of demes similar to the true number of marshes, estimated around 50, so the program could effectively interpolate over populations that were not sampled. We plotted the average migration surface of all of the runs. Each run consisted of 1,000,000 burn-in steps, then 2,000,000 MCMC iterations sampled every 9999 steps. To quantify whether the observed migration surface could be explained by a pattern of isolation by distance (Wright 1943), we plotted the pairwise genetic difference between demes with the pairwise geographic distance between the demes, as calculated by EEMS. Using R, we ran a linear regression on the pairwise genetic and geographic distances to test whether genetic distance increased significantly with geographic distance.

We also quantified genetic diversity in the wild Amargosa voles (and in the desert California vole populations for comparison) by calculating the proportion of segregating sites (Watterson's Theta, θ_W), proportions of pairwise differences (Tajima's Estimator, π), and the degree of inbreeding. We calculated θ_W and π using the built in functionality of ANGSD, based on the site frequency spectra of the populations (Korneliussen et al. 2013). We calculated the degree of inbreeding by calculating the relatedness coefficient for each population using the program ngsRelate (Korneliussen and Moltke 2015). While incorporating the uncertainty of genotype calls using genotype likelihoods into the analysis, ngsRelate calculates the probabilities of each pair of individuals in the analysis having zero, one or two pairs of alleles that are identical by descent (IBD). Given the probabilities that two individuals have one or two alleles IBD, we calculated the coancestry, or relatedness, coefficient between those two individuals (Weir et al. 2006). Thus, for all of the pairwise combinations of individuals within a single population, we took the average coancestry coefficient as a measure of inbreeding in that population. For ngsRelate, we used the above filtering, and we only considered alleles with a minor allele frequency of at least 0.05.

We calculated effective population size using the program NeEstimator V2.01 (Do et al. 2014) with data for all wild Amargosa voles. To use NeEstimator, we called SNPs using the same procedure as we did for the SpaceMix analyses. We converted the SNP matrix output from ANGSD to the required GENEPOP input format for NeEstimator using PopGenTools 3.4.10 (https://github.com/CGRL-QB3-UCBerkeley/PopGenTools), removing sites with missing data. We calculated N_e based on LD (Hill 1981), because estimates of N_e using the LD methods of Do et al. (2014) [modified according to Waples and Do (2010)], are precise for small population sizes.

Aim 3: quantify genetic diversity in the captive colony of Amargosa voles

To assess the amount of genetic diversity captured by the captive colony, we repeated the above analyses for the dataset comprising wild and captive Amargosa voles—hereafter referred to as Colony voles—from the founding generation, the F_1 generation, the F_2 generation and the backcrossed F_3 generation (n = 21 in total). First, to confirm where the captive voles fall in the population structure of the wild Amargosa voles, we conducted PCA as above. Second, we used NGSadmix, as above, to determine the proportion of admixture that the Colony voles draw from each of the two Amargosa vole groupings. Again, we used the ΔK method of Evanno et al. (2005) to determine what the most likely number of populations was when we included the Colony voles. PCA and NGSadmix results were concordant with

K = 2 being most likely (see "Results"). We used parametric bootstrapping to generate a null distribution of F_{ST} values based on 100 random permutations of 32 and 20 individuals to determine whether the two actual groups were significantly differentiated. Third, we quantified the amount of genetic variation captured by the Colony voles by calculating θ_W, π and the relatedness coefficient as above.

Results

RAD sequencing and filtering

Our paired-end bestRAD sequencing effort yielded 408,749,750 total reads for 79 voles. After aligning our reads to the reference genome, we had 3,828,309 total sites on 9806 contigs that passed our filtering criteria. For all contigs, we had a per-individual mean coverage of 21X (median: 14.1, min: 4.18, max: 105) before starting our analyses. Different analyses used slightly different subsets of the total dataset, and we present the number of sites or SNPs included for each analysis below. Our raw bestRAD reads are archived in the Short Read Archive (Accession: SRP097751), and tutorials to recreate the analysis and plots of our results are available at https://github.com/alexkrohn/AmargosaVoleTutorials.

Aim 1: determine phylogeographic structure for California voles and phylogeographic affinity for Amargosa voles

California voles exhibit significant population structure across their range. For all of our results we refer to individuals that group together in the PCA as "clusters", and groups identified by NGSadmix as "groups." The PCA, based on 20,000 variable SNPs, shows that genetic variation partitions into three clusters: the northern cluster includes Deep Springs, Lake Isabella and Lark Seep; the southern cluster includes Victorville and Elizabeth Lake; and the Amargosa voles form the final cluster, distinct from both northern and southern clusters, but closer to the northern cluster (Fig. 2). The three clusters of California voles from the PCA fall into two major groups (K = 2 using Evanno et al. 2005) according to the NGSadmix analysis, based on

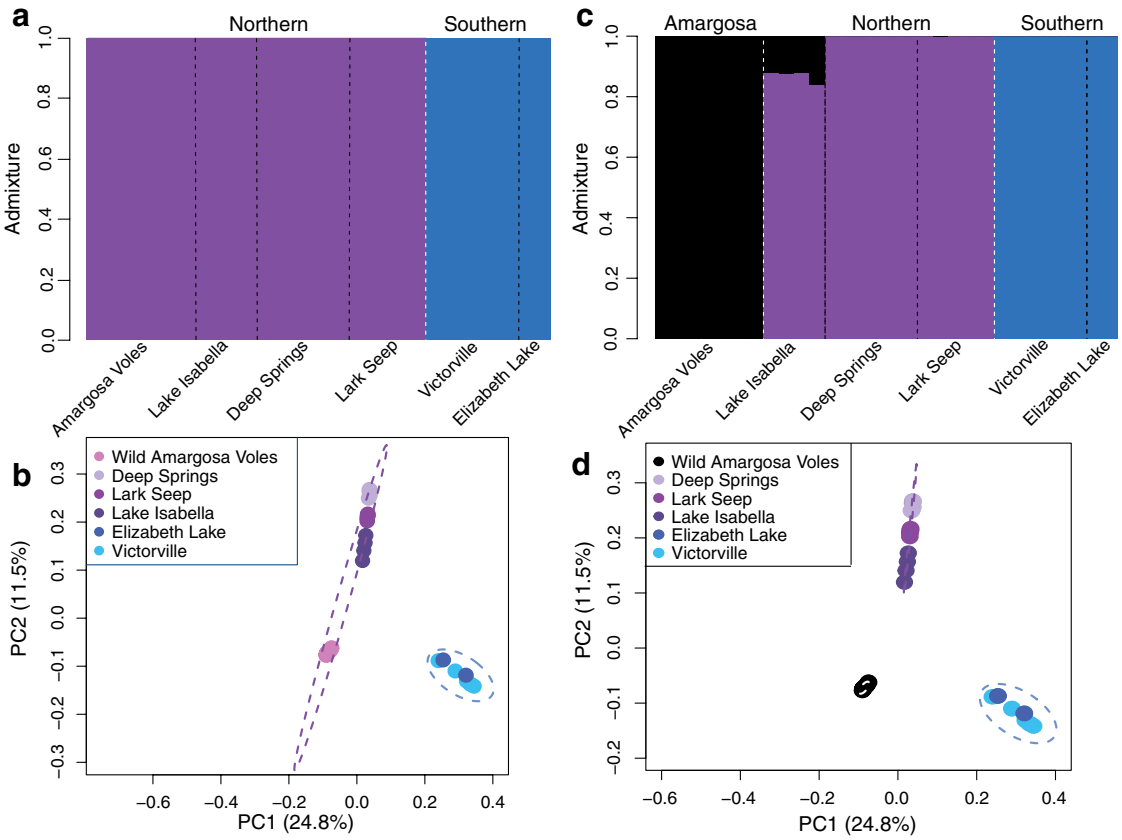

Fig. 2 Population structure analyses for California voles including Amargosa voles (native to the area around Tecopa, CA). Plots show NGSadmix and PCA for K = 2 (**a** and **b**, respectively), the most likely K, and for K = 3 (**c** and **d**, respectively), for comparison. *Dashed black lines* separate geographic sampling locations for the NGSadmix plots, while *dashed white lines* separate the groupings of each K. *Ellipses* on PCA plots are 95% confidence intervals around the mean of the two groupings identified in the corresponding NGSadmix plot

68,521 sites in 30 individuals (Fig. 2a, b). For comparison, we also present K = 3, which confirms that the Amargosa voles are genetically distinct from—but most closely related to—other northern clade voles (Fig. 2c, d). The results of the other NGSadmix runs (K = 4 through K = 8) are presented in Online Resource 2. The F_{ST} value between the northern and southern groups is 0.64. Based on our parametric bootstrapping analysis, an F_{ST} value of 0.64 is significantly higher than expected by chance ($P < 0.01$). The groupings recovered in both the PCA and NGSadmix analyses are concordant with the northern and southern clade designations of Conroy and Neuwald (2008) based on a smaller molecular dataset. According to the NGSAdmix analysis, the Amargosa voles share more admixture with the populations from the northern clade than with the populations from the southern clade.

SpaceMix analyses of 3801 SNPs also reveal that the Amargosa voles share more admixture with the northern clade than with the southern clade (Fig. 3). Plots of the joint marginal distributions, acceptance rates, and sample covariance each indicate that MCMC mixed well, achieved an acceptance rate near the desired 44%, and that the model adequately described the data. In the analysis, individuals from the southern clade draw admixture from a southern region of geogenetic space. Individuals from the northern clade, however, draw admixture from a common location away from the southern source of admixture, and geographically between Tecopa and Deep Springs, CA. Individuals from the Amargosa clade shift northwestwardly away from their sampled geographic location towards the convergence of the other northern clade individuals in geogenetic space, indicating that the Amargosa voles draw more admixture from the northern clade than the southern clade (Fig. 3), and corroborating the NGSadmix findings.

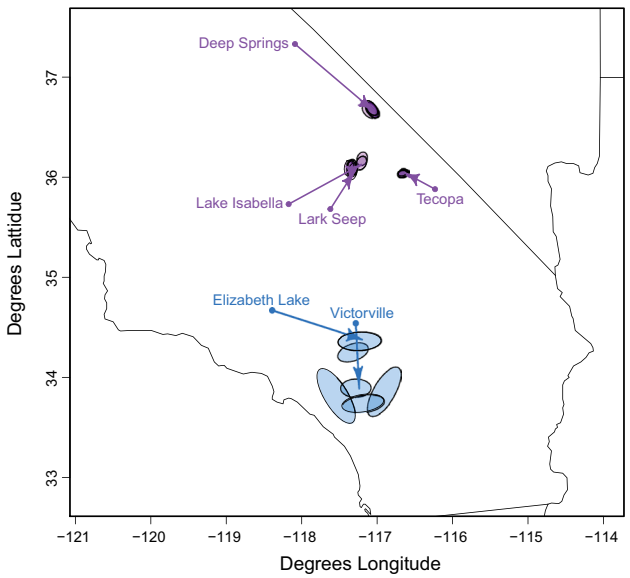

Fig. 3 SpaceMix plot of desert California voles (from Deep Springs, Lake Isabella, Lark Seep, Elizabeth Lake and Victorville, CA) and Amargosa voles (from Tecopa, CA). *Dots* represent actual geographic sampling location for each population. Each *ellipse* represents the 95% confidence interval of the "geogenetic" location of each sampled individual: the location of the individual if it could be pulled away from its geographic location proportional to—and in the direction of—its source of admixture. *Arrows point* from the geographic sampling location to the approximate mean geogenetic location of all the individuals from one sampling locality. *Purple* represents the northern clade and *blue* represents the southern clade, based on the K = 2 (see "Results" and Fig. 2)

Aim 2: quantify population structure and genetic diversity in wild Amargosa voles

The wild subpopulations of Amargosa voles are moderately structured genetically. The PCA, based on 20,000 variable SNPs, shows that the existing genetic variation forms two clusters (Fig. 4). This corresponds to a roughly north–south genetic break between Marshes 17 and 55. Population assignment analyses with NGSadmix, based on 7,774 SNPs, show a similar grouping, but indicate that Marshes 10 and 17, which are very close geographically, show admixture between the two groups (Fig. 4). The most likely K is 2 and the F_{ST} value between the two groups is 0.09. Our parametric bootstrapping analysis identified an F_{ST} value of 0.09 as significantly higher than what one would expect by chance ($P < 0.01$). The results of the other NGSadmix runs (K = 3 through K = 7) are presented in Online Resource 3.

Amargosa voles show little migration between the two clusters identified in the PCA and NGSadmix analysis. EEMS analysis show that while there is some migration within each of the two clusters from the PCA and NGSadmix analysis, there appears to be a barrier to migration between the two clusters (Fig. 5). There is more migration than average among the group comprised of Marshes 9, 11, 39 and 55, and average to less than average migration among the marshes in the other group (Fig. 5). The region of decreased migration coincides with a small developed area of Tecopa Hot Springs, although suitable marshes do exist in between Marshes 55 and 17 (Fig. 1). Importantly, genetic differentiation does not increase positively with geographic distance in the wild Amargosa voles ($P = 0.726$; Fig. 6), indicating that genetic differences that gave rise to the patterns of migration observed in the EEMS analysis are different than a null expectation of isolation by distance.

Amargosa voles show half as much genetic variation as other desert-dwelling California vole populations (Table 1). The Amargosa vole samples show similar, low levels of genetic diversity across all marsh localities and most individual marshes had approximately half of the total genetic diversity present in all of the wild Amargosa voles, supporting our findings that genetic variation is partitioned into two clusters (θ_W; Table 2). As a group, two Amargosa voles are,

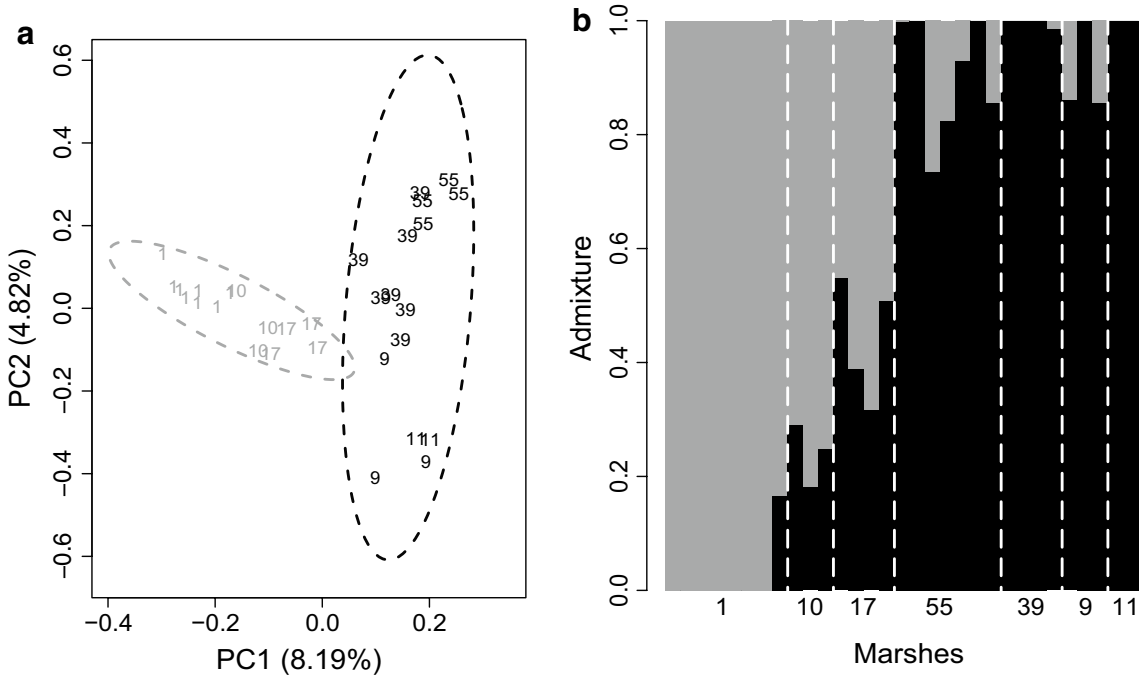

Fig. 4 Population structure analysis for wild Amargosa voles. **a** PCA plot of genetic variation showing individuals (labeled by the marsh where they were captured) into two slightly overlapping clusters. **b** NGSadmix analyses suggest there are two groups with some admixture between the two groups. *Ellipses* on the PCA are 95% confidence intervals around the mean of the NGSadmix groups

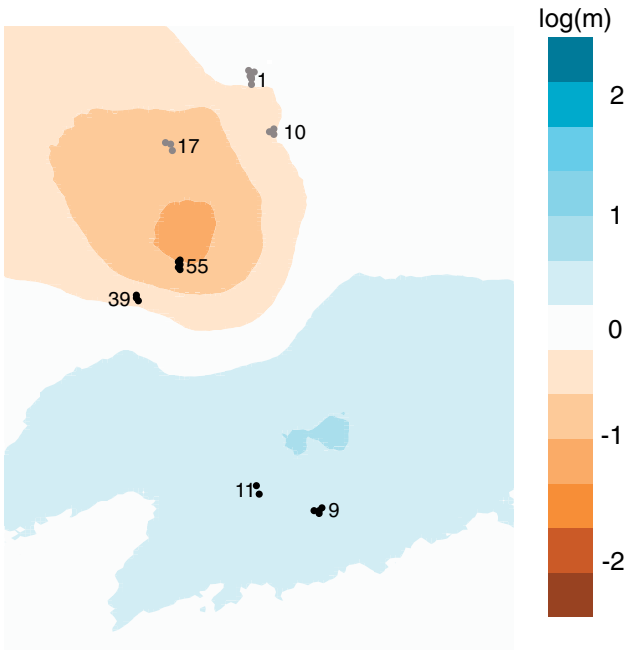

Fig. 5 Estimated effective migration surface of wild Amargosa voles. *Black* or *grey dots* represent geographic locations of sampled individuals, and numbers are names of the marshes. *Colors* indicate, on a $\log_{10}$ scale, whether interpolated migration between the sampled individuals is above or below the average level of 0. There is less than average amount of migration between Marshes 55 and 17, and more than average migration between Marshes 9 and 11. Interpolations of migration levels does not necessarily indicate suitable habitat

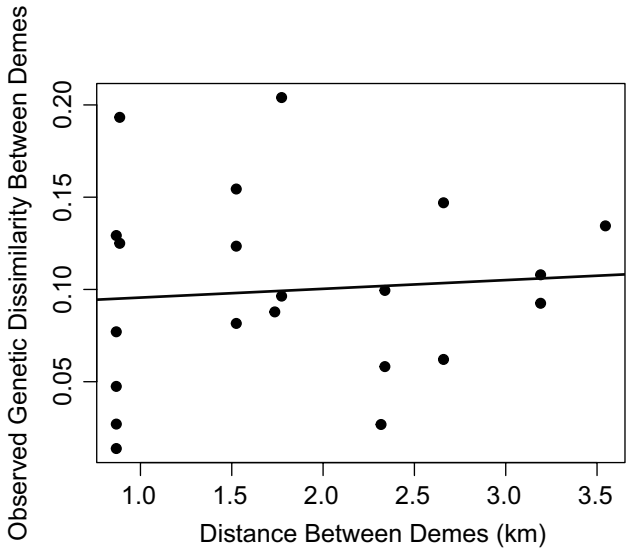

Fig. 6 Isolation by distance plot for wild Amargosa voles. The relationship between observed genetic distance and geographic distance for each pairwise combination of demes is not significant (P = 0.72)

on average, more related to each other than two voles from any one of the other measured desert California vole populations (Table 1). Looking at each marsh separately, relatedness coefficients are higher and more varied than among the larger populations. Accordingly, Amargosa voles have

Table 1 Percent of segregating sites (Watterson's Theta, θ_W), percent of pairwise differences (Tajima's estimator, π), the relatedness coefficient and sample sizes (n) for desert California voles and Amargosa voles showing reduced genetic diversity in Amargosa voles relative to other California Voles

Population	θ_W	π	Relatedness	n
Wild Amargosa voles	0.631	0.409	0.024	31
Deep Springs	1.009	1.131	0.003	6
Lark Seep	1.224	1.357	0	5
Lake Isabella	1.541	1.606	0.004	4
Elizabeth Lake	1.433	*1.469*	*0*	2
Victorville	1.329	1.383	0.003	6
Northern Clade Voles	1.456	1.559	0	15
Southern Clade Voles	1.487	1.428	0	8

Values in italics should be interpreted with caution as π and the relatedness coefficient are not robust to small sample sizes, even with thousands of loci. Note that the Northern Clade Voles comprise Deep Springs, Lark Seep and Lake Isabella populations, while the Southern Clade Voles comprise Elizabeth Lake and Victorville populations

Table 2 Percent of segregating sites (Watterson's Theta, θ_W), percent of pairwise differences (Tajima's estimator, π), the relatedness coefficient and sample sizes (n) for wild individuals and captive colony generations of Amargosa voles

	θ_W	π	Relatedness	n
Marsh 1	0.339	0.355	0.137	8
Marsh 9	0.348	0.394	0.079	4
Marsh 10	0.378	*0.396*	*0.087*	3
Marsh 11	0.369	*0.384*	*0.265*	2
Marsh 17	0.311	0.355	0.122	4
Marsh 39	0.378	0.415	0.05	6
Marsh 55	0.339	0.378	0.186	4
Founders	0.316	0.259	0.007	10
F1	0.257	*0.261*	*0*	2
F2	0.268	*0.268*	*0.055*	3
'F3'	0.246	0.231	0.034	6
Wild Amargosa voles	0.631	0.409	0.024	31
Captive Amargosa voles	0.305	0.257	0.012	21

Values in italics should be interpreted with caution as π and the relatedness coefficient are not robust to small sample sizes, even with thousands of loci. The F_3 generation is in quotation marks because it is a backcrossing of F_1 and F_2 individuals

extremely small effective population sizes. Based on 4605 SNPs, wild Amargosa voles have an effective population size of 32.3 individuals (95% confidence interval = 32.0–32.5).

Aim 3: quantify genetic diversity in the captive colony of Amargosa voles

The Colony voles contain a subset of variation from the northern group of wild Amargosa voles according to both the PCA and NGSadmix analyses. Compared to the PCA and NGSadmix plots of just the wild Amargosa voles (Fig. 4), the addition of the Colony voles does not change the major clusters (Fig. 7). The most likely grouping is still K = 2. The results from the other NGSadmix runs (K = 3 through K = 9) are presented in Online Resource 4. Unsurprisingly, the Captive voles appear in the same cluster as Marsh 1, the marsh from which the colony is derived (Fig. 7). The F_{ST} value between the two groups present in the PCA and NGSAdmix analyses is 0.14, higher than the values between the wild Amargosa vole groups. The 0.14 F_{ST} value is also significantly greater than what one would expect by chance ($P < 0.01$). However, the results with the addition of the Colony voles are qualitatively the same:

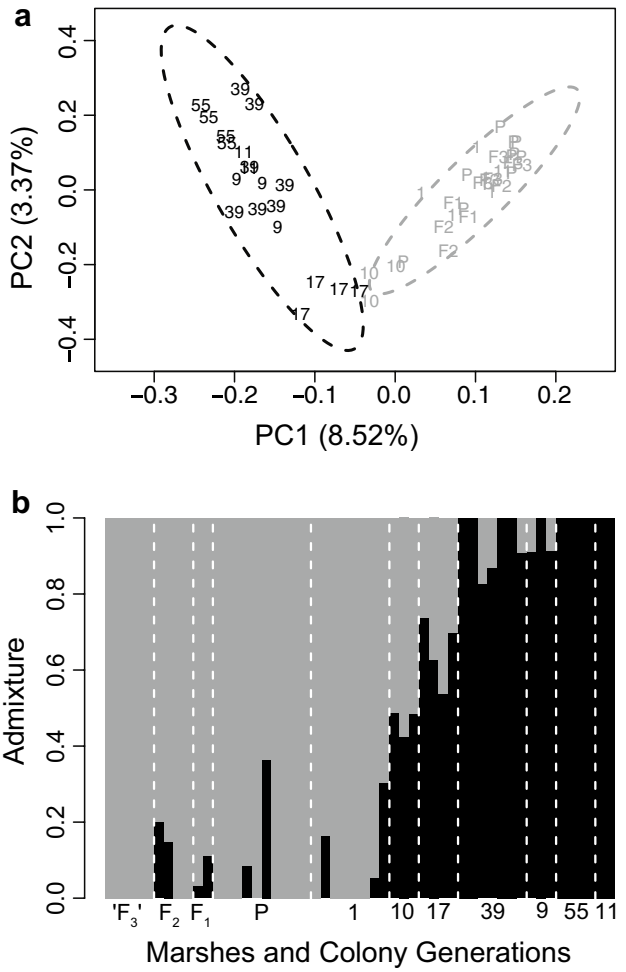

Fig. 7 Population structure analysis of wild and captive Amargosa voles. **a** PCA plot and **b** NGSadmix analysis show two clusters of voles, with the captive voles entirely aligned with samples from their Marsh 1 source population. *Dashed white lines* separate marshes or colony generations. *P* represents the founding members of the Colony voles, from Marsh 1. Subsequent colony generations are noted as F_1, F_2 and 'F_3', given that F_3 individuals are backcrosses between F_1 and F_2 individuals

the Colony voles only contain the genetic variation of the group that includes Marsh 1.

The Colony voles have less genetic diversity than the combined sample of wild Amargosa voles. According to both θ_W and π, the Colony voles contain approximately half of the total genetic variation present in the wild Amargosa voles (Table 2). The Colony voles capture approximately as much variation as any one of the wild marshes. When compared against the average relatedness of two randomly drawn wild Amargosa voles from any marsh, the Colony voles show similar degrees of inbreeding (Table 2).

Discussion

Using genome-wide markers, we evaluated the phylogeography and conservation genomics of desert dwelling California vole populations, with a focus on the endangered Amargosa vole populations. Our goal was to quantify genetic variation and detect fine scale patterns of population structure that could not be resolved using individual markers (e.g., microsatellite loci, Neuwald 2010). Because Amargosa voles are so rare, tissue samples are difficult to obtain. Therefore we used low-coverage sequencing methods (Korneliussen et al. 2014); with many loci, fewer individuals are needed to robustly resolve population structure, especially when analyses are individual-based rather than relying on a priori population boundaries (Prunier et al. 2013; Barley et al. 2015; Kardos et al. 2015; Gottscho et al. 2017). Our results suggest that Amargosa voles are genetically most similar to the northern clade of California voles. Within the geographically restricted Amargosa voles, we found evidence of moderate population structure. Amargosa voles exhibit reduced genetic diversity across their small range and have a small effective population size. We also show that the captive colony of Amargosa voles has a similar genetic signature to the wild population from which the founding members originated, and thus captures only some of the total genetic variation present in wild Amargosa voles. Below we detail these findings and their implications for management.

Our analyses suggest there are two clades of California vole, one northern clade and one southern clade, with the Amargosa voles genetically distinct but most closely allied with the northern clade. We found several lines of evidence that the Amargosa voles are more closely related to the northern, rather than southern clade. First, our NGSadmix analysis suggest that the Amargosa voles are genetically most similar to the northern clade California voles. For both $K = 2$ (the most likely K, according to Evanno et al. 2005) or $K = 3$, NGSadmix analyses reveal that the Amargosa voles and northern California voles share more admixture with each other than either group does with the southern clade (Fig. 2). Second, Amargosa voles draw admixture from the same geogenetic space as other voles previously shown to be in the northern clade (Fig. 3; Conroy and Neuwald 2008; Conroy et al. 2016). Although the Amargosa voles are more closely allied with the northern clade, there is evidence that they are genetically distinct. PCA plots comparing genetic variation among desert California vole populations and Amargosa voles show Amargosa voles as a distinct cluster separate from the rest of the northern voles (Fig. 2). Moreover, NGSadmix analyses with $K = 3$ show Amargosa voles as a distinct genetic group.

Our results corroborate past findings on the relationships among California voles, although with increased resolution as we have included many thousands more sites across the genome. Gill (1984) first proposed that California voles may be incipient species along a north–south break based on increased male sterility and decreased female fertility in north–south pairings and allozyme differences. Other authors also confirmed the same north–south break (Conroy and Neuwald 2008; Conroy et al. 2016). Looking specifically at the placement of Amargosa voles within California voles, our results align well with previous findings based on fewer nuclear markers. Using 12 microsatellites and 3 nuclear introns, Conroy et al. (2016) found that the Amargosa voles shared nuclear alleles with northern clade voles. In fact, the admixture source location inferred by our SpaceMix analysis is similar to the location that Conroy et al. (2016) proposed as the ancestral location from which the northern clade was derived. Conroy et al. (2016) proposed the Southern Sierra Nevadas as the ancestral location, and our SpaceMix analysis shows the northern clade groups converging near the Owens River Valley and Southern Sierra region (Fig. 3). Our results do not support the finding that Amargosa voles are part of the southern clade of California voles, as Conroy et al. (2016) found with their mitochondrial locus and a few microsatellite alleles. The discrepancy between mtDNA and nuclear DNA could be due to cytonuclear discordance, a phenomenon common in rodents (Yang and Kenagy 2009; Taylor and Hoffman 2012; Pagès et al. 2013) and other animals (Di Candia and Routman 2007; Lindell et al. 2008; Singhal and Moritz 2012). The discrepancy between our results and the microsatellite findings of Conroy et al. (2016) is likely due to the fact that we sampled many thousands more loci that show a pattern of overall connectivity to the northern, rather than southern clades of California voles. Thus, Amargosa voles are genetically distinct but most closely related to voles from the northern clade, but further phylogenetic studies including outgroups will help resolve whether Amargosa voles are nested within or sister to the northern clade of California voles. Nevertheless, our data support a paleo-hydrological connection between Tecopa and other northern clade populations as suggested by Conroy et al. (2016).

We also found that the Amargosa voles have substantially less genetic diversity than other desert California vole populations (Table 1). Levels of genetic variation in non-Amargosa populations of desert California voles are similar to other rodent species (Jezkova et al. 2015; Teng et al. 2016), whereas the Amargosa vole populations typically exhibit less than half the genetic variation of other desert California vole populations (Table 1). Unsurprisingly for a population with such low genetic diversity, the inferred effective population size of Amargosa voles (32.0–32.5 voles) is very small, indicating that their protected status is well warranted. Given the low census population size (Klinger et al. 2013) and low inferred effective population size, the observed lack of genetic diversity could be due to a recent, or ongoing, population decline. However, population sizes of California vole fluctuate tremendously over relatively short time periods (Krebs 1966), with corresponding decreases in genetic variation with large population crashes (Bowen 1982). Therefore, it is possible that past population size decreases have also contributed to low genetic diversity, or that population sizes were always quite low. Demographic simulations will be necessary to determine whether ongoing and/or past bottlenecks are responsible for the low genetic diversity present in Amargosa voles.

Despite the low genetic variation in wild Amargosa voles, we found some substructure with genetic variation partitioned into two groups (Figs. 4, 7). F_{ST} between the northern and southern marshes is relatively low (0.009), but significantly different than zero, suggesting moderate population structure (Hartl and Clark 1997). The fact that such a low F_{ST} value is statistically significant may be due to low genetic variation within each of the groups. Given that F_{ST} is low, the barrier to gene flow between the two groups (between Marshes 55 and 17 in Fig. 5) is likely permeable. The only previous study of Amargosa vole's population genetics used only three microsatellite loci, but also found structure despite low genetic variation (Neuwald 2010). Given that we sampled different marshes than Neuwald (2010), it is unsurprising that we found different boundaries for our genetic clusters, although we did both find north–south breaks. Given the low amount of genetic variation (Table 1), low effective population size, differences in fixed alleles among populations (Table 2), and known fluctuations in California vole census sizes (Krebs 1966; Bowen 1982), it is likely that drift is playing an important role in the remaining Amargosa vole populations.

We also evaluated genetic variation in the Amargosa voles in a captive colony and found that Colony voles cluster with individuals from Marsh 1 from which they were derived, and only contain a subset of the genetic diversity present in wild Amargosa voles (Figs. 4, 7; Table 2). The genetic signature of Colony voles is unsurprising as the 20 founding members of the captive colony all originated from Marsh 1 (Foley 2016). The captive colony has been managed well to avoid further inbreeding by pairing founders captured at distant trap locations and selecting subsequent pairs of breeders guided by pedigree. Given that even two individuals from the F_1 generation of Colony voles are, on average, less related to each other than two individuals from the marsh that they were taken from (Table 1), the founding individuals were well chosen to prevent inbreeding in the colony. Additionally, the breeding procedures at the captive colony have maintained relatedness values lower than those present in the wild marshes, indicating that workers are successfully avoiding inbreeding with mate pairings. Moving forward, scientists and managers now can decide whether to interbreed voles from the two Amargosa vole subgroups so that the colony can preserve more of the genetic variation present in the wild. In other systems, previous work has shown that interpopulation hybridization may create outbreeding depression that may be equally as harmful as inbreeding depression (Edmands 2007). However, given the genetic composition of the current captive colony, should a catastrophic event wipe out the few remaining wild Amargosa voles, the genetic diversity that is not already present in the Colony voles would be lost. Moreover, genetic divergence between the northern and southern marshes is only moderate. One solution would be to bring individuals from Marshes 9, 11, 39 or 55 into captivity and either interbreed them with the existing colony animals or maintain them separately.

In conclusion, we used RADseq to sequence thousands of loci across the genomes of 79 desert-dwelling California voles to resolve the phylogeographic position of Amargosa voles and inform the conservation of Amargosa voles. Our results have management implications for reintroductions of Amargosa voles. Erring on the cautious side, voles from the colony can be released into suitable marshes within the area of increased migration around Marsh 1, 17 and 10 (Fig. 5). By bringing voles from the area that contains Marshes 9, 11, 39 and 55 into captivity without interbreeding them with the current colony, an insurance colony could be created for that sub-group (Fig. 5) while avoiding potential effects of outbreeding. However, genetic diversity is low in Amargosa voles overall and F_{ST} is relatively low between the two sub-groups, so an alternative would be to use voles from the current colony for wider-spread reintroductions or allow interbreeding in the colony between the two subgroups. If Amargosa voles become extirpated and managers consider reintroductions of sister populations, voles from the most proximate northern clade populations are the closest living relatives to Amargosa voles. Our study highlights the importance of proactively understanding genetic structure in wild and captive populations before reintroductions are attempted (La Haye et al. 2017). Additionally, our study highlights the use of genome-wide markers to resolve relationships among

populations that are difficult to sequence reliably with traditional markers due to low genetic variation. Future work should focus on increasing sample sizes and monitoring changes in genetic diversity in the Amargosa voles over time both in the colony and the wild to further aid in the recovery of these imperiled populations.

Acknowledgements Funding was provided by the Bureau of Land Management (SPO Number 201223906). We thank Andrew DeWoody, and the DeWoody lab at Purdue University for access to a draft Amargosa vole genome. We thank the Museum of Vertebrate Zoology (UC Berkeley) for access to tissues of California voles, and Jim Patton for capturing and giving insight on *Microtus* generally. We thank two anonymous reviewers and Rosemary Gillespie for their helpful feedback on this manuscript. This work was approved by the UC Davis IACUC (Permit Numbers 19741 and 18179).

References

Ali OA, O'Rourke SM, Amish SJ et al (2016) Rad capture (rapture): flexible and efficient sequence-based genotyping. Genetics 202:389–400. doi: 10.1534/genetics.115.183665

Barley AJ, Monnahan PJ, Thomson RC et al (2015) Sun skink landscape genomics: assessing the roles of micro-evolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. Mol Ecol 24:1696–1712. doi: 10.1111/mec.13151

Bowen BS (1982) Temporal dynamics of microgeographic structure of genetic variation in *Microtus californicus*. J Mammal 63:625–638

Bradburd GS, Ralph PL, Coop GM (2016) A spatial framework for understanding population structure and admixture. PLoS Genet 12:1–38. doi: 10.1371/journal.pgen.1005703

Catchen J, Hohenlohe PA, Bassham S et al (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. doi: 10.1111/mec.12354

Conroy CJ, Neuwald JL (2008) Phylogeographic study of the California vole, *Microtus californicus*. J Mammal 89:755–767. doi: 10.1644/07-MAMM-A-189R1.1

Conroy CJ, Patton JL, Lim MCW et al (2016) Following the rivers: historical reconstruction of California voles *Microtus californicus* (Rodentia: Cricetidae) in the deserts of eastern California. Biol J Linn Soc 119:80–98. doi: 10.1111/bij.12808

Di Candia MR, Routman EJ (2007) Cytonuclear discordance across a leopard frog contact zone. Mol Phylogenet Evol 45:564–575. doi: 10.1016/j.ympev.2007.06.014

Do C, Waples RS, Peel D et al (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol Ecol Resour 14:209–214. doi: 10.1111/1755-0998.12157

Docker MF, Heath DD (2003) Genetic comparison between sympatric anadromous steelhead and freshwater resident rainbow trout in British Columbia, Canada. Conserv Genet 4:227–231. doi: 10.1023/A:1023355114612

Edmands S (2007) Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. Mol Ecol 16:463–475. doi: 10.1111/j.1365-294X.2006.03148.x

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Foley J (2016) Annual performance report: captive breeding of a model species and reconstructing habitat ahead of translocation of the critically endangered Amargosa vole. Presented to BLM and USFWS

Fumagalli M, Vieira FG, Korneliussen TS et al (2013) Quantifying population genetic differentiation from next-generation sequencing data. Genetics 195:979–992. doi: 10.1534/genetics.113.154740

Funk WC, Lovich RE, Hohenlohe PA et al (2016) Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). Mol Ecol 25:2176–2194. doi: 10.1111/mec.13605

Gill AE (1984) Partial reproductive isolation of subspecies of the California vole, Microtus californicus. Genetica 52–53:105–117

Gottscho AD, Wood DA, Vandergast AG et al (2017) Lineage diversification of fringe-toed lizards (Phrynosomatidae: *Uma notata* complex) in the Colorado Desert: delimiting species in the presence of gene flow. Mol Phylogenet Evol 106:103–117. doi: 10.1016/j.ympev.2016.09.008

Hall ER (1981) The mammals of North America. John Wiley & Sons, New York

Hartl DL, Clark AG (1997) Principles of population genetics, 3rd edn. Sinauer Associates, Sunderland, 542 pp

Hill W (1981) Estimation of effective population size from data on linkage disequilibrium. Genet Res Camb 38:209–216. doi: 10.1017/S0016672300020553

Jezkova T, Riddle BR, Card DC et al (2015) Genetic consequences of postglacial range expansion in two codistributed rodents (genus *Dipodomys*) depend on ecology and genetic locus. Mol Ecol 24:83–97. doi: 10.1111/mec.13012

Kalinowski ST (2011) The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity 106:625–632. doi: 10.1038/hdy.2010.95

Kardos M, Luikart G, Allendorf FW (2015) Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity 115:63–72. doi: 10.1038/hdy.2015.17

Kellog R (1918) A revision of the *Microtus californicus* group of meadow mice. Univ Calif Publ Zool 21:1–42

Klinger R (2015) Estimation of the total population size of the Amargosa vole. Report to United States Geological Survey, Bishop

Klinger R, Cleaver M, Anderson S, Maier P, Clark J (2013) Short-term population dynamics, demography, and habitat selection by the Amargosa vole. Report to the BLM. United States Geological Survey, Western Ecological Research Station, Bishop

Kopelman NM, Mayzel J, Jakobsson M et al (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191. doi: 10.1111/1755-0998.12387

Korneliussen TS, Moltke I (2015) NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics 31:4009–4011. doi: 10.1093/bioinformatics/btv509

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 Bioinform 14:289. doi: 10.1186/1471-2105-14-289

Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: analysis of next generation sequencing data. BMC Bioinform 15:356. doi: 10.1186/s12859-014-0356-4

Krebs CJ (1966) Demographic changes in fluctuating populations of *Microtus californicus*. Ecol Monogr 36:239–273. doi: 10.2307/1942418

La Haye MJJ, Reiners TE, Raedts R et al (2017) Genetic monitoring to evaluate reintroduction attempts of a highly endangered rodent. Conserv Genet. doi: 10.1007/s10592-017-0940-z

Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Lindell J, Méndez-De La Cruz FR, Murphy RW (2008) Deep biogeographical history and cytonuclear discordance in the black-tailed brush lizard (*Urosaurus nigricaudus*) of Baja California. Biol J Linn Soc 94:89–104. doi: 10.1111/j.1095-8312.2008.00976.x

Madsen T, Shine R, Olsson M, Wittzell H (1999) Conservation biology: restoration of an inbred adder population. Nature 402:34–35. doi: 10.1038/46941

Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963. doi: 10.1093/bioinformatics/btr507°

McKenna A, Hanna M, Banks E et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. doi: 10.1101/gr.107524.110

Miller W, Wright SJ, Zhang Y et al (2010) Optimization methods for selecting founder individuals for captive breeding or reintroduction of endangered species. Pac Symp Biocomput 53:43–53. doi: 9789814295291_0006

Murdoch D, Chow ED (2013) ellipse: Functions for drawing ellipses and ellipse-like confidence regions. R package version 0.3-8. https://CRAN.R-project.org/package=ellipse

Neuwald JL (2010) Population isolation exacerbates conservation genetic concerns in the endangered Amargosa vole, *Microtus californicus scirpensis*. Biol Conserv 143:2028–2038. doi: 10.1016/j.biocon.2010.05.007

Norton JE, Ashley MV (2004) Genetic variability and population differentiation in captive Baird's tapirs (*Tapirus bairdii*). Zoo Biol 23:521–531. doi: 10.1002/zoo.20031

Pagès M, Bazin E, Galan M et al (2013) Cytonuclear discordance among Southeast Asian black rats (*Rattus rattus* complex). Mol Ecol 22:1019–1034. doi: 10.1111/mec.12149

Petkova D, Novembre J, Stephens M (2016) Visualizing spatial population structure with estimated effective migration surfaces. Nat Genet 48:94–100. doi: 10.1101/011809

Prunier JG, Kaufmann B, Fenet S et al (2013) Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: towards a better assessment of functional connectivity using an individual-based sampling scheme. Mol Ecol 22:5516–5530. doi: 10.1111/mec.12499

Puechmaille SJ (2016) The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Mol Ecol Resour 16:608–627. doi: 10.1111/1755-0998.12512

R Core Team (2016). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rick J (2015) Genetic diversity, structure, and hybridization in a harvested gray wolf (*Canis lupus*) population in Minnesota. MS Thesis, University of Minnesota

Singhal S, Moritz C (2012) Testing hypotheses for genealogical discordance in a rainforest lizard. Mol Ecol 21:5059–5072. doi: 10.1111/j.1365-294X.2012.05747.x

Skotte L, Korneliussen TS, Albrechtsen A (2013) Estimating individual admixture proportions from next generation sequencing data. Genetics 195:693–702. doi: 10.1534/genetics.113.154138

Tallmon DA, Luikart G, Waples RS (2004) The alluring simplicity and complex reality of genetic rescue. Trends Ecol Evol 19:489–496. doi: 10.1016/j.tree.2004.07.003

Taylor ZS, Hoffman SMG (2012) Microsatellite genetic structure and cytonuclear discordance in naturally fragmented populations of deer mice (*Peromyscus maniculatus*). J Hered 103:71–79. doi: 10.1093/jhered/esr100

Teng H, Zhang Y, Shi C et al (2016) Whole-genome sequencing reveals genetic variation in the Asian House Rat. G3 6:1969–1977. doi: 10.1534/g3.116.029504

Tordoff HB, Redig PT (2001) Role of genetic background in the success of reintroduced peregrine falcons. Conserv Biol 15:528–532. doi: 10.1046/j.1523-1739.2001.015002528.x

United States Fish and Wildlife Service (2009) Amargosa vole *Microtus californicus scirpensis* 5-year review: summary and evaluation. United States Fish and Wildlife Service, Ventura, p. 29

Wang J (2017) The computer program structure for assigning individuals to populations: easy to use but easier to misuse. Mol Ecol. doi: 10.1111/1755-0998.12650

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evol Appl 3:244–262. doi: 10.1111/j.1752-4571.2009.00104.x

Weir BS, Anderson AD, Hepler AB (2006) Genetic relatedness analysis: modern data and new challenges. Nat Rev Genet 7:771–780. doi: 10.1038/nrg1960

Wright S (1943) Isolation by distance. Genetics 28:114–138

Yang D-S, Kenagy GJ (2009) Nuclear and mitochondrial DNA reveal contrasting evolutionary processes in populations of deer mice (*Peromyscus maniculatus*). Mol Ecol 18:5115–5125. doi: 10.1111/j.1365-294X.2009.04399.x