

# The role of neutral and adaptive genomic variation in population diversification and speciation in two ground squirrel species of conservation concern

Soraia Barbosa<sup>1</sup>, Kimberly Andrews<sup>1</sup>, Amanda Goldberg<sup>1</sup>, Digpal Gour<sup>1</sup>, Paul Hohenlohe<sup>1</sup>, Courtney Conway<sup>1</sup>, and Lisette Waits<sup>1</sup>

<sup>1</sup>University of Idaho

March 12, 2021

## Abstract

Understanding the neutral (demographic) and adaptive processes leading to the differentiation of species and populations is a critical component of evolutionary and conservation biology. In this context, recently diverged taxa represent a unique opportunity to study the process of genetic differentiation. Northern and southern Idaho ground squirrels (*Urocitellus brunneus* – NIDGS, and *U. endemicus* - SIDGS, respectively) are a recently diverged pair of sister species that have undergone dramatic declines in the last 50 years and are currently found in metapopulations across restricted spatial areas with distinct environmental pressures. Here we genotyped single-nucleotide polymorphisms (SNPs) from buccal swabs with restriction site-associated DNA sequencing (RADseq). With these data we evaluated neutral genetic structure at both the inter- and intra-specific level, and identified putatively adaptive SNPs using population structure outlier and genotype-environment association (GEA) analyses. At the interspecific level, we found a clear separation between NIDGS and SIDGS, and evidence for adaptive differentiation relating to differences in hibernation. At the intraspecific level, we identified 3 Evolutionarily Significant Units for NIDGS and 2 for SIDGS plus multiple Management and Adaptive Units. Elevation appears to be the main driver of adaptive differentiation in NIDGS, while neutral variation patterns match and extend that identified in previous studies using microsatellite markers. For SIDGS, neutral substructure generally reflected the effect of natural geographic barriers, while adaptive variation reflected differences in land cover and temperature. These results clearly highlight the roles of neutral and adaptive processes for understanding species and population differentiation, which can have important conservation implications in threatened species.

## Introduction

Isolation and subsequent local adaptation of populations are considered common processes that lead to speciation. Changes in habitat, barriers to dispersal, or stochastic demographic events can cause population isolation and diversification (Slatkin 1987; Steinberg *et al.* 2000). In this context, environmental variation plays an important role in facilitating or hindering connectivity, and thus in promoting the persistence of populations, their vicariance, or even extinction (Waits *et al.* 2016). On one hand, increased structural and functional connectivity facilitates the persistence of small populations that are highly susceptible to demographic stochasticity, genetic drift and density-dependent effects (Hanski 1998; Lopes & de Freitas 2012; Wittmann *et al.* 2018). On the other hand, high levels of connectivity result in more genetically homogenous populations, with less propensity for local adaptation (Kawecki & Ebert 2004). With loss of connectivity between populations, allele frequencies tend to diverge due to genetic drift, ultimately leading to neutral genetic differentiation (Rundell & Price 2009). Additionally, spatial variation in local selection pressures within a species' range, particularly when there is habitat fragmentation, can lead to changes in allele frequencies and fixation of new adaptive mutations, resulting in the emergence of adaptive differences (Orsini *et al.* 2013). This is especially important for small and isolated populations that are restricted

to increasingly unfavorable habitat, for which studies have shown that local adaptation tends to occur more rapidly (Wood *et al.* 2016; Hoffmann *et al.* 2017). Recently diverged populations represent a great opportunity to study the process of genetic differentiation (Fišer *et al.* 2018; Fletcher *et al.* 2019; Marques *et al.* 2019). The comparison of neutral and adaptive variation should provide evidence for distinguishing which processes are contributing most to differentiation, and what can be done to circumvent or sustain that diversification, depending on specific conservation goals (Orsini *et al.* 2013).

Northern and southern Idaho ground squirrels (*Urocitellus brunneus* and *U. endemicus*, respectively) are a recently diverged pair of sister species, which currently have an allopatric distribution (Figure 1) (Yensen 1991). Northern Idaho ground squirrels (hereafter NIDGS) and southern Idaho ground squirrels (hereafter SIDGS) were formerly considered two subspecies of *Spermophilus brunneus* and have been distinguished as separate species on the basis of ecological niche modelling, morphology and genetics (Gill & Yensen 1992; Yensen & Sherman 1997; Helgen *et al.* 2009; Hoisington-Lopez *et al.* 2012; USFWS 2015). Both species are rare, endemic to Idaho, and are of high conservation concern (IUCN 2000, 2018). Ecologically, both species are semi-colonial and patchily distributed, representing classic examples of metapopulation structure whereby dispersal among populations is uncommon and tends to occur in a ‘stepping stone’ manner (Yensen 1991; Yensen & Sherman 1997; USFWS 2003). NIDGS live in open meadows, grassy scabs and small rocky outcroppings at an elevation of 1100 to 2300 m within coniferous forests of central Idaho (Burak 2011; Goldberg, Conway, Mack, *et al.* 2020), and they persist within only a small fraction of their former range likely due to habitat loss and reduced population connectivity, mostly as a result of forest encroachment (Sherman & Runge 2002; Suronen & Newingham 2013; Yensen & Dyni 2020; Helmstetter *et al.* 2021). SIDGS live in sagebrush steppe and rolling hill slopes at an elevation of 630 to 1400 m in southwestern Idaho, and are currently threatened by urban and agricultural development, as well as the spreading of invasive annual plants (USFWS 2000; Lohr *et al.* 2013). Morphological differences between the two species include coat color, which tends to mimic differences in soil color between the species’ geographic ranges (Yensen 1991), pelage (longer in SIDGS), and baculum characteristics (longer with more spines in SIDGS) (Yensen 1991). Previous genetic work on NIDGS and SIDGS estimated that the divergence between NIDGS and SIDGS occurred about 32.5 (18.3-63.5) thousand years ago, during the Quaternary climate cycles, and found no subsequent gene flow between the two species (Hoisington-Lopez *et al.* 2012). Vicariant events of this magnitude have been found to be sufficient for distinct evolutionary lineages to become different species, a pattern frequently found in several North American small mammals (Hope *et al.* 2014, 2016, 2020).

Local adaptation is likely to be an important factor for ground-dwelling, hibernating, small mammals like NIDGS and SIDGS with limited dispersal abilities. Both ground squirrel species hibernate but the timing of hibernation differs, likely due to differences in elevation and climate (Yensen 1991; Goldberg & Conway 2021). This difference in emergence timing between NIDGS and SIDGS could have a genomic basis, or may simply result from a plastic response (Yensen 1991; Hut & Beersma 2011; Santos *et al.* 2017). Typically, adaptations are associated with the habitat variables that affect fitness the most, which in the case of the Idaho ground squirrels (hereafter IDGS) are likely variables associated with energy consumption, timing of food availability, soil temperature, forage quality, and risk of predation, which may vary between the active season and hibernation (Goldberg, Conway, Mack, *et al.* 2020; Goldberg, Conway, Tank, *et al.* 2020). Variation in ground squirrel hibernation emergence timing has been associated with food availability and snowpack and thus, site productivity appears to dictate differences within and possibly between species differentiation (French 1982; Goldberg & Conway 2021). These differences may be determinant for species divergence, but may also lead to intraspecific local adaptation if environmental differences are found across the species range (Kawecki & Ebert 2004; Savolainen *et al.* 2013). Previous intraspecific genetic studies on Idaho ground squirrels have found that genetic differentiation was low to moderate among NIDGS populations, with one disjunct population (Round Valley) being completely isolated (Yensen & Sherman 1997; Garner *et al.* 2005; Hoisington 2007; Hoisington-Lopez *et al.* 2012). In SIDGS, the Weiser River was documented as a barrier to dispersal between populations, but connectivity among populations on either side of the river was relatively high (Garner *et al.* 2005; Hoisington 2007). Additionally, there are reports of human-mediated translocations between localities east of the Weiser River, particularly from southern localities close to Van

Deussen to the vicinity of the Weiser River population (Yensen *et al.* 2010; Yensen & Tarifa 2012). However, translocation success in SIDGS has been very limited, especially into areas without established populations, for which the majority of the translocated individuals did not survive the first winter (Busscher 2009; Yensen *et al.* 2010). Given the currently observed low levels of gene flow among some populations within both species, population persistence might be highly dependent on locally adapted genotypes for many isolated populations (USFWS 2000; Rundell & Price 2009). Thus, to better understand the probability of population persistence under habitat and climate change, it is essential to determine the role of neutral processes in maintaining population connectivity and overall genetic diversity, and the role of adaptive processes in improving population resilience through local adaptation (Macdonald *et al.* 2018).

In this study, we aimed to develop and use genomic tools to provide novel information on neutral and adaptive genetic diversity and differentiation within and among NIDGS and SIDGS to address the following five questions: 1) how do genetic patterns of adaptive and neutral variation compare between and within species? 2) are there any populations that have elevated levels of adaptive differentiation? 3) what landscape variables are associated with loci under selection? 4) can we identify specific genes under selection and what are their putative functions? and 5) can we identify conservation units based on genomic data to inform management? To address these questions, we tested the following hypotheses: (a) geographic distance will be the main driver of differentiation for neutral loci in NIDGS and SIDGS, possibly exacerbated by geographic barriers to gene flow; (b) populations within each species will exhibit signatures of adaptation to local environmental conditions; (c) local adaptation between populations within species will be highest in NIDGS because this species occupies a more topographically diverse area (Yensen & Sherman 1997); and (d) adaptive variation will be associated with timing of hibernation, production and storage of fat, and increased metabolism at higher elevation for NIDGS (Faherty *et al.* 2018; Garcia-Elfring *et al.* 2019). Finally, we combine all of this information to identify Evolutionarily Significant Units (ESUs), Management Units (MUs) and Adaptive Units (AUs) that could warrant heightened protection due to genetic isolation or the emergence of local adaptation.

## Methods

### *Sampling and genotyping*

We analyzed a total of 304 Idaho ground squirrel samples. We conducted all sampling following standard biological and ethical requirements under the auspices of University of Idaho Animal Care and Use Committee Protocol #2015-53. We preserved samples consisting of buccal swabs in Qiagen ATL buffer and stored them at room temperature or 4°C until DNA extraction. For NIDGS, we obtained 232 samples from 17 populations in Adams County in central Idaho (Figure 1B), which represents approximately 60% of the known populations. For SIDGS, we collected 72 samples from 5 populations in Washington, Payette and Gem Counties (Figure 1B), which is only a small proportion of the > 100 sites identified during surveys in 2000 (CCAA 2002). Details on the location and number of samples collected per population are shown in Table 1. Samples were extracted using the Qiagen Blood and Tissue DNA extraction kit (Qiagen, Inc.), and quantified DNA using fluorometric quantitation with a Qubit double-stranded DNA High Sensitivity Assay Kit. We used genetic samples for Restriction Site Associated DNA Sequencing (RADseq) (Murphy *et al.* 2007; Baird *et al.* 2008; Andrews *et al.* 2016). We prepared libraries following Ali *et al.* (2016) using the *sbfi* restriction enzyme, but excluding the last part of that protocol relating to targeted bait capture, and instead only using the new RADseq protocol with biotinylated adapters. We built four libraries, each comprised of ~80 individually barcoded samples, and sequenced each library on one lane of Illumina HiSeq 4000 at the University of Oregon Genomics & Cell Characterization Core Facility (GC3F), with 150 base pairs (bp) paired-end reads. We conducted all bioinformatic analyses using the Institute for Bioinformatics and Evolutionary Studies (IBEST) Computational Resources Core servers at the University of Idaho. For SNP calling, we used the software *stacks* version 2.2 (Catchen *et al.* 2013) to demultiplex and remove PCR duplicates. We used the thirteen-lined ground squirrel (*Ictidomys tridecemlineatus*) genome (GCA\_000236235.1) as a reference to align our data using the software *bwa* version 0.7.17 (Li 2013). From these alignments, we used the software *SAMtools* version 1.9 (Li *et al.* 2009). We conducted our analyses at

both the interspecific and intraspecific levels in order to determine putative adaptive differences between the two sister species, but also searching for patterns of population structure and local adaptation among populations within species. To do so, we defined a first dataset consisting of all 304 IDGS, and then separated the samples by species, ‘NIDGS’ (232 samples) and ‘SIDGS’ (72 samples) datasets. We called SNPs for the ‘IDGS’ dataset using the multisample SNP caller (*mpileup*) implemented in *SAMtools*, and subsequently used VCFTools version 0.1.14 (Danecek *et al.* 2011) to exclude individuals with  $\geq 50\%$  missing data, keeping only biallelic SNPs that had  $< 50\%$  missing data, were located  $> 10,000$  base pairs apart, had  $> 2\%$  minor allele frequency and  $> 3$  reads. To produce the species specific datasets, we separated the samples by species (232 NIDGS and 72 SIDGS) and repeated the above filtering with a MAF value of 3%. We then used part of the R script available from Wright *et al.* (2019) to exclude loci with heterozygosity  $> 70\%$  and with greater than a 3:1 ratio of mean read depth for the reference allele versus the alternative allele, across individuals with these alleles, to avoid loci with significant reference bias. Finally, we used the functions from the *whoa* R package to estimate heterozygote miscalls rates and excluded loci deviating from more than one standard deviation from the mean (<http://150.185.130.98/rcran/web/packages/whoa/index.html>). For all datasets, we compared the percentage of missing data per individual (`-missing-indv`) with mean read depth (`-depth`) and heterozygosity (`-het`) obtained using VCFTools version 0.1.14 (Danecek *et al.* 2011).

### Identification of adaptive variation

To detect putative loci under selection, we used four methods that vary in whether they account for population structure and whether they allow for correlations among environmental variables. Specifically, we used a population structure outlier detection method (*pcadapt*), which does not control for population structure and does not correlate with environmental variables, and three different methods to estimate genotype-environment associations (GEAs): a Latent Factor Mixed Model (LFMM), which is a univariate regression model that summarizes environmental variation in latent factors that correct the model for population structure; a redundancy analysis (RDA), which is a multivariate approach that looks for associations of SNPs with each environmental variable separately, and a partial redundancy analysis (pRDA), which is an RDA that accounts for population structure (Forester, Landguth, *et al.* 2018; Capblancq *et al.* 2018).

For the population structure outlier detection analysis we used the function *pcadapt* from the R package *pcadapt* version 4.1.0 (Luu *et al.* 2017) to detect candidate genes under selection (‘adaptive’ loci) based on principal components analysis. To determine the number of  $K_{PC}$  to retain, we used a threshold for the singular values of the loading of each principal component (PC) above 5% of variance explained for the axes to be retained and used these  $K_{PC}$  to control for population structure in the outlier detection analysis. We selected outlier loci based on an alpha threshold of 0.01 calculated using the *qvalue* function from the *qvalue* R package (Dabney *et al.* 2010), after correcting the *p*-values for multiple comparisons with a Benjamini-Hochberg Procedure using the function *p.adjust*.

GEA analyses require that the dataset does not have any missing data, so we split samples by species, and then imputed missing genotypes based on the most frequent allele across all individuals of that species. Imputation did not consider population of origin to avoid inflating population structure, particularly given the low number of samples in some populations. We acknowledge that there will likely be a bias for increasing representation of alleles from populations with more individuals and higher minor allele frequency (MAF) (Mitt *et al.* 2017). However, this approach is more conservative in terms of the detection of local adaptation, and thus less likely to lead to false positives in terms of local adaptation. To test for associations between genetic and environmental data, we obtained bioclimatic data from across the study area using WorldClim (Fick & Hijmans 2017), as well as biomass (ESRI 2016), elevation (USGS 2012), land cover (CCRS *et al.* 2017), slope degree (ESRI 2013), soil particle size (SoilPartSize) as a proxy for burrow stability and heat retention (CONSBIO 2015a), soil temperature (SoilTemp) (Marchenko 2011) because it is likely important for hibernation (Goldberg 2018), soil orders (SoilOrder) that reflect exposure to climatic factors and biological processes (CONSBIO 2015b), land formation (LandForm) (ESRI 2015), and ecological facets (EcoFacets) which combine elevation, land formation, slope and heat load index (ESRI 2019). For categorical variables (land cover, soil order, land formation and ecological facets), we coded each category as present or absent

and analyzed them as independent variables. We used the Shapiro-Wilk normality test in the *shapiro.test* function from the *statsR* package (R Core Team 2013) to determine if each variable was normally distributed in order to determine which correlation coefficient to use. For our data, none of the variables were normally distributed, so we then used the *chart.Correlation* function from the *PerformanceAnalytics* R package (Peterson *et al.* 2014) and the Kendall correlation estimator to examine correlations among variables (see Results), and the *corr* and *corrplot* functions from the *corrplot* R package (Wei *et al.* 2017) to plot the correlation matrix. We selected a subset of variables that were not highly correlated  $<|0.7|$  for subsequent analyses. In cases where correlations were above  $|0.7|$ , we excluded the variable that was correlated with a higher number of other variables or with the highest average correlation values, even if below  $|0.7|$ .

To ensure that associations between loci and environmental variables were not confounded by population structure, for the LFMM we performed a Principal Components Analysis (PCA) on our imputed dataset using the *therda* function from the R package *vegan* (Oksanen *et al.* 2011). To determine which principal components (PCs) to retain, we then used the broken stick criterion, based on whether the observed eigenvalues were higher than the corresponding random broken stick components (Peres-Neto *et al.* 2003). We used the retained PCs as conditions when finding associations between the genetic data with the univariate uncorrelated environmental PC axes through the function *lfmm\_ridge* from the *lfmm* R package (Caye *et al.* 2019). We then evaluated the genomic inflation factor (GIF) and selected a *p*-value following François *et al.* (2016), and then identified candidate loci using a false discovery rate (FDR) of 0.1.

For the RDA, we used the function *rda* from the *vegan* R package. We then used the R function *RsquareAdj* to determine the percentage of genetic variance explained by the RDA axes, after adjusting for the number of predictors. We then used the function *anova.cca* from the *vegan* R package to check the significance of our model, as well as of each RDA axis individually, in explaining the variance in the data with an ANOVA. As a first step, we verified the variance inflation factors (VIF) of each variable using the function *vif.cca* from the *vegan* R package, and excluded the variable with the highest VIF until all variables showed values below 10, i.e. no longer representing redundant constraints (Oksanen *et al.* 2011). Then, we determined which SNPs were candidates for local adaptation by identifying those present in the tails of the RDA axes loadings that were  $>2.5$  standard deviations (two-tailed *p*-value = 0.012) from the mean.

For the pRDA, we performed the same analysis as for the RDA, except that before using the *rda* function, we performed a PCA on the genomic data to summarize neutral population structure by determining which PCs had eigenvalues greater than random, using the broken stick criterion. This analysis identified the number of PCs to condition for in the *therda* function, and then used these PCs to exclude the variance explained by neutral population structure (Rellstab *et al.* 2015; Forester, Lasky, *et al.* 2018).

For the interspecific analysis (IDGS dataset), we conducted all four analyses for detecting adaptive loci (*pcadapt*, LFMM, RDA, and pRDA), and we compared the SNPs identified as candidates in the four analyses using the *venn.diagram* function from the *VennDiagram* R package (Chen & Boutros 2011). For the intraspecific analyses (NIDGS and SIDGS datasets), we only conducted the pRDA analysis.

### *Gene ontology and species-specific adaptations*

To gain insight into the ecological and biological functions of putative adaptive loci, we identified candidate loci found within genes and the Gene Ontology (GO) terms associated with such genes (Primmer *et al.* 2013). We used the Ensembl variant effect predictor (VEP) to perform annotation of the candidate loci (McLaren *et al.* 2016) and the software SNP2GO (Szkiba *et al.* 2014) to identify cellular component, biological process, and molecular function GO terms associated with the candidate loci using an FDR of 0.05, and following the annotations of the thirteen-lined ground squirrel genome. We tested for enrichment considering all the candidates identified by the four analyses combined, as well as for each analysis individually. To evaluate the impact of the different method assumptions on GO term enrichment, we further tested for enrichment considering the identified candidate loci divided in several categories: population structure outlier approach (*pcadapt*), representing those candidates uniquely identified for the *pcadapt* analysis; genotype environment associations ('GEA'), considering the candidates uniquely identified for the LFMM, RDA and pRDA analyses;

candidates identified including population structure ('POP'), considering those loci uniquely identified for the *pcadapt* and RDA analyses; and candidates identified when excluding population structure ('noPOP'), considering those loci uniquely identified for the LFMM and pRDA analyses. For all three datasets (NIDGS, SIDGS, and the combined IDGS), we further considered the candidate SNPs resulting in non-synonymous substitutions and used the online databases Ensembl and UniProt to identify the genes and proteins involved (Bateman 2019; Yates *et al.* 2020).

### Population structure

Population structure analyses were performed separately for each species. To obtain a set of "neutral markers" for each species for these analyses, we first removed from each dataset the loci identified as candidates for adaptation by the pRDA. Next, we identified loci that deviated from Hardy-Weinberg Equilibrium (HWE) proportions for all samples combined using the function `-hardy` from VCFtools, and we removed loci with  $p < 0.05$  from the dataset.

We first estimated isolation-by-distance (IBD) using the `mantel` function from the *vegan* R package (Oksanen *et al.* 2011) for both the neutral and adaptive datasets. We used the function `dist.genpop` from the *adegenet* R package (Jombart & Ahmed 2011) to calculate Nei's genetic distances (Nei 1972, 1978) between all pairs of populations, and plotted those genetic distances against pairwise Euclidean geographic distances (average geographic location of each population). We performed this for both NIDGS and SIDGS datasets and for neutral and adaptive loci separately.

We performed PCA to obtain a visual inspection of population structure without underlying models of evolution. We used the `snpgdsPCA` function from the *SNPRelate* R package (Zheng *et al.* 2012) to perform the PCA for the complete dataset (all loci), the neutral dataset, and the adaptive dataset of each species. We used the `ggplot` function from the *ggplot2* R package to construct the plots (Kassambara 2018). For the neutral data, we also evaluated population structure for each species using an approach with an underlying Bayesian model to estimate the number of populations ( $K$ ), using the program *STRUCTURE* (Pritchard *et al.* 2000) and the *ParallelStructure* R package (Besnier & Glover 2013). We applied an admixture model run for  $1 \times 10^6$  generations, and five replicates per  $K$ , with  $K = 1$  to  $K = 13$  in NIDGS and to  $K = 5$  in SIDGS, which correspond to the total number of populations sampled for each species. We then determined the optimal number of populations ( $K$ ) according to the Evanno method and the rate of change in the likelihood values (Evanno *et al.* 2005).

We estimated interspecific population differentiation using the IDGS total dataset, and intraspecific population differentiation for NIDGS and SIDGS 'neutral' and 'adaptive' datasets, using the functions `genet.dist` and `boot.pfst` from the R package *hierfstat* to calculate genetic divergence ( $F_{ST}$ ) with 999 bootstraps between all geographic sites (Goudet 2005). We then built correlation matrices of pairwise  $F_{ST}$  between all populations using the function `corrplot` from the *corrplot* R package (Wei *et al.* 2017). Using the same datasets, we determined genetic diversity by estimating observed and expected heterozygosity ( $H_O$  and  $H_E$ , respectively) per population. To estimate  $H_O$  and  $H_E$ , we used the function `basic.stats` from the *hierfstat* R package (Goudet 2005), averaging locus heterozygosity per population. We then performed a paired Wilcoxon test to identify significant differences between  $H_O$  and  $H_E$  for both neutral and adaptive loci among populations of each species, between  $H_E$  of both neutral and adaptive loci among populations of each species, and  $H_E$  between neutral and adaptive datasets for each species, using the function `t.test` from the *stats* R package.

### Defining conservation units

We applied the framework of Funk *et al.* (2012) to identify evolutionary significant units (ESUs), management units (MUs) and adaptive units (AU). ESUs were delineated based on the population structure detected using all loci. MUs were delineated based on neutral loci in HWE to detect demographically independent units. AUs were delineated using loci identified as adaptive using the methods outlined above. Assignment to different groups within each conservation unit type was based on visual inspection of the PCAs, accounting for all significant PC axes. For ESU definition we also considered mitochondrial DNA data from other studies (Hoisington 2007; Hoisington-Lopez *et al.* 2012), which tends to reflect biogeographic patterns at the levels

of intermediate to distant past (Zink & Barrowclough 2008), and for MUs we also considered the proportion of assignment to the  $K$  inferred populations, supported by levels of population differentiation,  $F_{ST}$  (Palsbøll *et al.* 2006).

## Results

### *Sampling and SNP calling*

We extracted DNA from a total of 233 NIDGS and 71 SIDGS buccal swab samples for which DNA yields varied from 0.1 to 51.0 ng/ $\mu$ l. Of 304 samples for which we conducted RADseq, for the IDGS dataset we kept 80 individuals and 4,227 SNPs after filtering the sequence data, with an average mean read depth of coverage of 9.3 reads per individual (min. 3.6 and max. 20.6) (Figure S1A, Supporting information). For the NIDGS dataset, we kept a total of 80 individuals and 3,575 SNPs, resulting in an average mean read depth of 7.5 (min. 0.9 and max. 18.8) (Figure S1B, Supporting information). For the SIDGS dataset, we kept 41 individuals and 2,348 SNPs, resulting in an average mean read depth of 4.9 (min. 1.0 and max. 16.1) (Figure S1C, Supporting information). Heterozygosity tended to decrease with increasing missing data and decreasing mean read depth, but no single population showed a particular bias towards low read depths across all individuals (Figures S1B and S1C, Supporting information).

### *Interspecific adaptive variation*

For the *pcadapt* analysis, we considered the first two PCs that explained a total of  $\sim 16\%$  of the genetic variance (Figure S2A, Supporting information). The first PC explained  $\sim 10\%$  of variance and clearly separated NIDGS from SIDGS, and some variation within SIDGS, while the remaining axes mostly reflected intraspecific variation within NIDGS (Figures S2B and S2C, Supporting information). Using a value of  $\alpha = 0.01$ , we detected 48 outlier loci using an adjustment of the  $p$ -values following the Benjamini-Hochberg procedure.

For the GEAs, most environmental variables were not normally distributed, and  $p$ -values for the Shapiro-Wilk normality were all lower than 0.001 for all variables. Based on the results of the Kendall correlation analysis, we kept 18 variables for the GEA analyses, given that these showed low correlation with one another (Figure S3A, Supporting information). The PCA performed with these variables shows that only PC1 explained more variance than the random broken stick criterion, and thus was kept as a summary of the environmental data to compare in the LFMM univariate analysis (Figure S4A, Supporting information). The PCA performed for the genomic data showed that only PC1 explained more variance than the random broken stick criterion, which suggests that the number of populations was two (Figure S3B, Supporting information). We ran the LFMM using  $K = 2$ , which resulted in a distribution of unadjusted  $p$ -values with a GIF of 2.42, which was then adjusted to 1.4 (Figure S5, Supporting information). Using a false discovery rate (FDR) of 10%, the LFMM analysis identified 52 candidate loci.

We then performed the RDA and pRDA for the IDGS dataset. For both analyses, from the 18 environmental variables used, we excluded three (BIO1, BIO2 and LF\_VA) due to high VIF. We obtained an adjusted  $r^2$  of 0.08 and 0.06 for the RDA and pRDA, respectively, across 15 axes, which correspond to the proportion of genetic variance explained by the environmental predictors kept in each analysis. The ANOVA showed that both analyses were significant at  $p < 0.001$ , and that the first four axes of both analyses were significant at  $p < 0.001$  (Figure S6, Supporting information). Imputation of the missing data appeared to not have created biases in population structure given the similarities between the *pcadapt* and the RDA PCA loadings (Figures 2A and 2B, and S2B and S2C, Supporting information). The first RDA axis explained 4.8% of the variance, while RDA2, RDA3 and RDA4 explained 2.8%, 2.5% and 2.2%, respectively. The remaining axes combined (RDA5-15) explained the remaining 13.5% of the total variance (Figure 2A and 2B). In this analysis, there was a clear separation of NIDGS and SIDGS individuals in RDA1, while RDA2-4 mostly separated populations within species. The environmental variables that loaded the most strongly with RDA1, and thus the separation between NIDGS and SIDGS, were ‘Steep Slopes’ (LF\_SS, associated with 20 SNPs), ‘Grassland/Herbaceous’ (LC\_GH, associated with 32 SNPs) and ‘Soil Temperature’ (soilPartSize, associated with 20 SNPs) (Table 2 and Figure 2A). When accounting for population structure with the pRDA analysis, we no longer find a clear distinction between NIDGS and SIDGS, and instead variation appears to mimic

that of the RDA without RDA1 (Figure 2). The first pRDA axis explained 2.9% of the variance (similarly to RDA2), while pRDA2, pRDA3 and pRDA4 explained 2.6%, 2.3% and 2.1%, respectively. The remaining axes combined (RDA5-15) explained the remaining 12.7% of the total variance (Figure 2C and 2D). We also found a total of 316 SNPs associated with 17 environmental variables, with 38.6% (122 SNPs) also associated with the land formation variable ‘Ridges and Peaks’ (LF\_RP) (Table 2).

Considering all analyses combined, we identified a total of 490 candidate loci, none of which were found by all four analyses (Figure 4). Comparing the population structure outlier approach (*pcadapt*) to the GEAs (LFMM, RDA and pRDA), there are 29 loci found by all GEAs vs. 18 found only by the population structure outlier approach (Figure 4). Methods that do not account (i.e. correct) for population structure (*pcadapt* and RDA) identified 21 loci in common, while methods that account for population structure (LFMM and pRDA) identified one locus in common. The largest overlap between analyses was 168 loci found by both multivariate GEAs (RDA and pRDA).

#### *Intraspecific adaptive variation*

To test for adaptive differences among populations of NIDGS and SIDGS, we performed partial redundancy analyses (pRDA) on each dataset separately. The environmental variables that we used differed among datasets, resulting in 19 variables for NIDGS, while only four variables were kept for SIDGS due to a large number of correlations (Figures S3B and S3C, respectively, Supporting information).

For NIDGS, PCA of the genomic data showed that none of the PCs have eigenvalues greater than random, suggesting  $K = 1$  (Figure S7A, Supporting information). However, this result does not mean there is no genetic structure in the data, but rather that it might not be particularly strong. Thus, because we have evidence from previous studies that NIDGS are divided in at least two groups (Garner *et al.* 2005; Hoisington-Lopez *et al.* 2012; Zero *et al.* 2017), we decided to condition the pRDA using PC1 (i.e. assuming  $K = 2$ ). From the 19 environmental variables used, we excluded four (BIO\_1, BIO\_12, BIO\_15 and LF\_VA) due to high VIF. We then obtained an adjusted  $r^2 = 0.06$  across 15 axes. The ANOVA showed that the pRDA was significant at  $p < 0.001$ , and that the first three axes were significant at  $p < 0.001$ . The first pRDA axis explained 3.3% of the variance, while pRDA2 and pRDA3 explained 3.0% and 2.7%, respectively. The remaining axes combined (pRDA4-15) explained 14.0% of the remaining variance. From this analysis, three populations were the most distinct in association to particular environmental variables: Lower Butter (LB) mostly associated with Local Ridge in Plain (LF\_LRP) and slope, Rocky Top (RT) mostly associated with Ridges and Peaks (LF\_RP), and Tamarack (TA) mostly associated with soil particle size (soilPartSize) (Figures 3A and 3B).

For SIDGS, the PCA on the genomic data showed that PC1 had eigenvalues greater than random, suggesting  $K = 2$  (Figure S7B, Supporting information). From the four environmental variables used, all were kept for the final pRDA, which had an adjusted  $r^2=0.01$ . The ANOVA showed that the pRDA was significant at  $p < 0.001$ , and that only the first axis was significant at  $p < 0.001$ , explaining 3.7% of the variance. The remaining three axes combined (pRDA2-4) explained 6.7% of the remaining variance. From this analysis, Paddock (PA) was the most distinct population, correlating mostly with grassland areas (LC\_GH), as opposed to the other populations which are more associated with shrub/scrubland, and Annual Mean Temperature and Isothermality (Figure 3C).

#### *Gene ontology*

For the IDGS dataset, we obtained a total of 490 candidate SNPs, of which only four were found in coding genes and produced non-synonymous substitutions (Table 3). Of these four, two were assigned to uncharacterized proteins and both were found to be fixed for the reference allele in NIDGS and variable in SIDGS. Considering the other two candidate SNPs, one was found on the B3GNT8 gene, which codes for the hexosyltransferase protein. This SNP was identified as a candidate by the RDA, and its strongest association was with land cover type “Grassland/Herbaceous” (0.384). The second missense candidate SNP was found on the NPR1 gene, which codes for the guanylate cyclase protein. It was identified as a candidate by both the *pcadapt* and RDA analyses, and in the latter it was most strongly associated with “soil particle size”

(0.397). For details regarding the assignment of the remaining candidate loci to additional variant classes, refer to Table 3. Considering the results for each analysis separately, the *pcadapt*, LFMM, RDA and pRDA identified 48, 52, 31 and 316 candidate loci, respectively, with 18, 13, 89 and 139 candidate SNPs uniquely identified, respectively (Figure 4). Considering the different categories considered, we considered 18, 472, 337 and 362 candidate loci for the population structure outlier approach (*pcadapt*), genotype environment associations (‘GEA’), and including (‘POP’) and excluding (‘noPOP’) population structure, respectively, of which 18, 26, 21 and 1 were unique to analysis type (Figure 4). Gene ontology enrichment analysis of candidate SNPs found no evidence for enrichment of particular cellular components, biological processes or molecular functions for any of the individual or combined analyses considered.

For the NIDGS and SIDGS datasets, we found 150 and 18 candidate SNPs, respectively. For NIDGS, two candidate SNPs were found in coding regions and resulted in non-synonymous substitutions, although both were assigned to uncharacterized proteins. For SIDGS, no SNPs were found to result in non-synonymous substitutions (Table 3). For details regarding the assignment of the remaining candidate loci to additional categories, refer to Table 3. For these two datasets, we also found no significant GO term enrichment for all identified candidate SNPs.

#### *Population structure and conservation units*

After excluding the candidate adaptive loci and loci deviating from HWE, we kept 2,663 and 1,711 SNPs for the NIDGS and SIDGS, respectively. With these sets of neutral loci, we determined that there is a significant pattern of IBD for NIDGS ( $r = 0.136$ ,  $p = 0.039$ ) but not for SIDGS ( $r = 0.556$ ,  $p = 0.090$ ). IBD was not significant for the NIDGS adaptive dataset ( $r = 0.018$ ,  $p = 0.396$ ), or for the SIDGS adaptive dataset ( $r = -0.378$ ,  $p = 0.950$ ).

PCA results for the total and neutral datasets in both species were identical (Figure S8, Supporting information). From the NIDGS PCA analysis, there were four axes for the total and neutral datasets and three axes for the adaptive dataset explaining a significant proportion of the variance (Figures 5A, 5B and S8, Supporting information). Both the neutral and adaptive datasets identified similar populations as the most distinct, particularly Lower Butter (LB), Rocky Top (RT) and Tamarack (TA) (Figure S8, Supporting information). Mud Creek (MC) was only distinctive at the neutral level. Considering potential groups, at the first two PCs, Lower Butter (LB) was clearly separated from an eastern group (MC, TA, PV, LV) and a western group (all other populations). Based on these data, we defined three ESUs for NIDGS: ESU1 represented by the western group, ESU2 represented by Lower Butter, and ESU3 represented by the eastern group (Figure 6A). From the SIDGS PCA analysis, we considered as significant one axis for the total the neutral and the adaptive datasets, the first two identified Olds Ferry (OF) as the most distinctive populations, while the adaptive dataset showed a clear separation of Paddock (PA) (Figure 5C and 5D, and S8, Supporting information). Based on these data we defined two ESUs for SIDGS: ESU1 represented by Olds Ferry, and ESU2 represented by all other populations east of the Weiser River (Figure 6B).

The PCA neutral results for both species agree with the STRUCTURE analyses (Figure 6). For NIDGS, STRUCTURE suggested separation of NIDGS into two main groups corresponding to a western group (ESU1 and ESU2) and also an eastern group (ESU3) (Figure 6A,  $K = 2$ ). Here, Lower Butter was included in the western group, although there was a high proportion of admixture with the eastern group as well. Lower Butter (ESU2) was differentiated from all other populations at all  $K$  values starting from  $K = 3$  (Figure 6A). The STRUCTURE analysis indicated a second peak in the marginal likelihood for  $K = 6$  using the Evanno *et al.* (2005) method, suggesting the presence of fine-scale population structure within NIDGS (Figures 6A, and S9A, S9B and S10A, Supporting information). At  $K = 6$ , the eight populations of ESU1 split into three geographically distinct populations or MUs, and the four populations of ESU3 split into three MUs, with Lost Valley showing evidence of admixture (Figure 6A). In terms of adaptive variation, within ESU1 we identified two AUs from eight populations, corresponding to the same structure found for MUs, ESU2 (Lower Butter) was also a distinct AU, and for ESU3, only Tamarack was adaptively differentiated (Figure 6A). These results are also reflected in the pairwise  $F_{ST}$  estimates, where there is no significant differentiation among 7 of the 8 populations of ESU1, even at the adaptive level (Figures S11A and S11B, Supporting information).

Rocky Top (ESU1), Lower Butter (ESU2), and Mud Creek (ESU3) were the NIDGS populations with the highest neutral average pairwise  $F_{ST}$ , with 0.09, 0.08 and 0.08, respectively, considering an overall average of 0.06, while for the highest adaptive average pairwise  $F_{ST}$  was found for Rocky Top (ESU1) and Lower Butter (ESU2), with 0.26 and 0.20, respectively, considering an average of 0.10. The SIDGS STRUCTURE results also mimicked those from the neutral PCA where individuals were found to be divided in two groups, with a separation of Olds Ferry from all other SIDGS localities at  $K = 2$  (Figure 6B). This resulted in the definition of two MUs, according to our criteria. In terms of adaptive differentiation, Paddock was identified as a separate AU.

In terms of population differentiation, average  $F_{ST}$  between NIDGS and SIDGS was 0.143, compared to 0.05 and 0.100 within NIDGS and SIDGS for the same dataset, respectively. At the intraspecific level and considering the NIDGS and SIDGS datasets separately, average neutral pairwise  $F_{ST}$  was highest for Olds Ferry (0.13), considerably higher than the average value of 0.04 (Figure S11C, Supporting information). However, as seen in the SIDGS adaptive PCA, Paddock (PA) was the most distinct population and also had the highest average pairwise  $F_{ST}$  (0.24), considerably higher than the average value of 0.15 (Figure S11D, Supporting information).

Considering the levels of genetic diversity, for the IDGS dataset, no NIDGS population showed significantly decreased  $H_O$  compared  $H_E$ , but Lower Butter did show significantly higher  $H_O$  than  $H_E$  (Table 4). Contrarily, for the same dataset, two out of four SIDGS populations showed significantly lower  $H_O$  than  $H_E$ . SIDGS populations tended to have significantly lower  $H_E$  than NIDGS populations (Table S1, Supporting information). At the intraspecific level, for NIDGS we observed an average of neutral  $H_O$  of 0.237, which was significantly lower than the average  $H_E$  of 0.280 (Table 4). Values of heterozygosity for the adaptive dataset were lower than the neutral dataset, and  $H_O$  (0.199) was lower than  $H_E$  (0.190). Of the 10 populations for which we estimated  $H_E$ , four had significantly lower neutral  $H_O$  than  $H_E$ , while only two showed lower  $H_O$  than  $H_E$  for the adaptive dataset (Table 4). Considering the comparisons among populations, Steve's Creek/Squirrel Valley (SS,  $H_E = 0.305$ ), Tamarack (TA,  $H_E = 0.296$ ) and Mud Creek (MC,  $H_E = 0.286$ ) showed significantly higher neutral  $H_E$  (Table S1, Supporting information) than all other NIDGS populations. For the adaptive dataset, Rocky Top (RT,  $H_E = 0.294$ ), Tamarack (TA,  $H_E = 0.270$ ) and Lower Butter (LB,  $H_E = 0.263$ ) showed the highest  $H_E$  values, also significantly higher than most other NIDGS populations (Table S1, Supporting information). Comparing the  $H_E$  estimated for the neutral and adaptive datasets, Mud Creek (MC), Rocky Top (RT), and Steve's Creek/Squirrel Valley (SS) had significantly lower adaptive than neutral  $H_E$  (Table S1, Supporting information). For the SIDGS, we observed an average of neutral  $H_O$  of 0.231, which was significantly lower than the average  $H_E$  of 0.276. Values of heterozygosity for the adaptive dataset were similar to the neutral dataset, again with  $H_O$  (0.232) lower than  $H_E$  (0.275). Of the four populations for which we estimated  $H_E$ , three presented significantly lower neutral  $H_O$  than  $H_E$ , while none showed lower  $H_O$  than  $H_E$  for the adaptive dataset (Table 4). Considering the comparison among populations, Olds Ferry (OF,  $H_E = 0.287$ ) and Paddock (PA,  $H_E = 0.283$ ) showed significantly higher neutral  $H_E$  than the other three SIDGS populations (Table S1, Supporting information). For the adaptive dataset, Paddock (PA,  $H_E = 0.410$ ) showed significantly higher  $H_E$  compared to all other SIDGS populations, as well as when compared to its neutral  $H_E$  estimate (Table S1, Supporting information).

## Discussion

We examined the genetic basis of adaptive divergence between two endemic and vulnerable sister species of Idaho ground squirrels and assessed the diversity and differentiation at both neutral and adaptive loci among populations of each species. As in previous studies (Garner *et al.* 2005; Hoisington 2007; Hoisington-Lopez *et al.* 2012), we observed a clear genetic separation between northern (NIDGS) and southern (SIDGS) Idaho ground squirrels. Within species, we identified different clustering of populations when considering neutral and adaptive variation separately, which suggests that local adaptation and demographic processes are both important in shaping genetic variation and both should be considered in management and conservation planning of these species.

### *Interspecific neutral and adaptive divergence*

Previous studies have found substantial differences in the ecological niches occupied by these two rare and endemic ground squirrels (Hoisington-Lopez *et al.* 2012; Zero *et al.* 2017), as well as distinct morphology and life history traits (Yensen 1991; Hoisington 2007), supporting the hypothesis that these two taxa have levels of divergence consistent with species level differences (Gill & Yensen 1992; Garner *et al.* 2005). Using the largest molecular dataset generated to date, we observed a clear genetic separation between northern (NIDGS) and southern (SIDGS) Idaho ground squirrels (Figures 2 and S2B, Supporting information). When we controlled for population structure to remove the variation associated with the interspecific differences, the resulting population structure was similar to that of the intraspecific pRDA analyses, indicating that this analysis was providing information regarding within-species variation (Figure 2). Therefore, we considered the RDA to produce more meaningful results than the pRDA results for interspecific comparisons of adaptive variation. This divergence was also clearly shown by  $F_{ST}$  values, which were significantly higher between species, than within species. As in line with previous microsatellite work (Garner *et al.* 2005; Hoisington-Lopez *et al.* 2012), SIDGS populations were generally less diverse than NIDGS and were more frequently found to have significantly lower  $H_O$  than  $H_E$ , which suggests effects of inbreeding (Rousset & Raymond 1995).

### *Intraspecific neutral and adaptive variation*

The persistence of populations that are increasingly isolated depends largely upon the level of connectivity between them (Hodgson *et al.* 2011). Thus, the high differentiation observed among some populations within NIDGS might result from loss of suitable habitat and dispersal corridors between populations, currently leading to the isolation (and possible extinction) of populations (Yensen 1999; Sherman & Runge 2002; Barrett 2005). Such patterns are frequent in other small mammals with historically reduced ranges across altitudinal gradients (Waterhouse *et al.* 2018; Bi *et al.* 2019). For NIDGS, the neutral dataset identified two main groups (termed here ‘western’ and ‘eastern’, Figure 6A,  $K = 2$ ) which corroborates previous studies that used microsatellite loci (Garner *et al.* 2005; Hoisington-Lopez *et al.* 2012; Zero *et al.* 2017). Our results also identified more fine-scale structure within the eastern part of the range than within the western part of the range, indicating higher connectivity across the populations sampled from the western region. This result is similar to that of previous microsatellite analyses for the eastern part of the range (although this study did not analyze the exact same populations; Garner *et al.* 2005), but contrasts somewhat with our hypothesis (a) of IBD and the results of an allozyme study, which found significant IBD within the western region (Gavin *et al.* 1999) and with unpublished results from Hoisington (2007) which found evidence for additional substructure and restricted gene flow within both the eastern and western groups. Very similar patterns of differentiation have also been found using mitochondrial DNA (Hoisington 2007). For SIDGS, the neutral dataset identified the highest differentiation between populations on different sides of the Weiser River, which corroborates previous studies (Garner *et al.* 2005; Hoisington 2007; Zero *et al.* 2017). Interestingly, although Olds Ferry was the most distinct population at the neutral level, it was not substantially differentiated from other populations on the eastern side of the Weiser River at the adaptive level, suggesting that its distinction is mostly due to demographic (i.e. neutral) processes (Garner *et al.* 2005; Hoisington-Lopez *et al.* 2012; Zero *et al.* 2017). Previous studies sampled a larger number of sites (10-11) and detected additional subgroups both east and west of the Weiser River (Hoisington 2007; Zero *et al.* 2017). Earlier work on SIDGS, which included a larger number of populations and larger spatial area, found human disturbance (impervious surfaces and agriculture) and small-scale topographic complexity to restrict gene flow while higher at site heat load index, growing season precipitation and frost free period facilitated gene flow (Zero *et al.* 2017). Translocations among populations east of the Weiser River were performed to supplement small and isolated populations, and might have led to an increased level of genetic homogeneity at the neutral level (Yensen *et al.* 2010; Weeks *et al.* 2011; Yensen & Tarifa 2012; Landguth & Balkenhol 2012). Successful translocations would result in low neutral and adaptive differentiation, but previous studies have found low rates of translocation success (Panek 2005; Busscher 2009; Yensen *et al.* 2010; Smith *et al.* 2019). These translocations were performed mainly into areas near the Weiser River population and might explain the low genetic differentiation observed between this population and all others east of the Weiser River (Figure S11C, Supporting information). Similarly, adaptive differentiation was lowest for WR,

suggesting some benefits of translocations (increased genetic diversity and masking of deleterious alleles) outweighing the potential negative effects such as outbreeding depression (Weeks *et al.* 2011). Still, most of the SIDGS populations showed lower than expected neutral heterozygosity (Table 4), which suggests effects of inbreeding leading to reduced genetic diversity. This agrees with previous studies that have found lower genetic diversity in SIDGS than NIDGS, and point to the need of increased protection of this species (Garner *et al.* 2005).

#### *Adaptive associations with the environment*

As we hypothesized, some populations within each species exhibited signatures of adaptation to local environmental conditions. At the interspecific level, SIDGS were generally more associated with grassland and higher soil particle size, which is positively correlated with soil temperature, while associations of NIDGS were varied and appeared to vary mostly on a population basis (Figures 2 and S3A, Supporting information). These differences are within our expectations of adaptive differentiation between the two species, given that NIDGS and SIDGS occur in high and low elevations, respectively (Hoisington-Lopez *et al.* 2012). These genomic signatures of adaptive differences between NIDGS and SIDGS might also reflect associations with timing of snow melt, or site productivity, which are related to difference in hibernation (Hoisington-Lopez *et al.* 2012; Zero *et al.* 2017; Goldberg, Conway, Mack, *et al.* 2020).

Intraspecific differences can quickly arise in populations that become isolated and/or that are distributed across highly variable environmental gradients (Doebeli & Dieckmann 2003; Smith *et al.* 2019), and there is an important role of adaptive variants and their specific adaptations in the maintenance of genetic diversity and the long-term persistence of threatened species (Rubidge *et al.* 2012). Small populations can naturally become locally adapted in the process of isolation, resulting in a gradual increase in adaptive differentiation from other populations, compared to neutral differentiation (Doebeli & Dieckmann 2003; Holderegger *et al.* 2006; Wood *et al.* 2016). In agreement with our initial hypothesis (c), NIDGS showed a higher number of distinct populations both the neutral and adaptive level, with Rocky Top and Lower Butter as the most distinct populations and Tamarack to a lesser extent (Figures S11A and S11B, Supporting information). None showed particularly low genetic diversity or significantly lower  $H_O$  than  $H_E$  (indicative of inbreeding), indicating that these populations are unique but also demographically stable, despite local bottlenecks (Assis *et al.* 2013). Rocky Top and Lower Butter were mainly associated with environmental variables relating to elevation (Figure 3), and they are the highest elevation populations sampled in this study (around 1700 and 1600 m above sea level, respectively). This suggests that these populations might have particular adaptations to elevation compared to other NIDGS populations, a pattern also observed in other North American ground squirrels (Eastman *et al.* 2012). Tamarack is located on a west facing slope which remains colder until later during the day, similarly to populations at higher elevation sites. Tamarack was also found to be associated with increased soil particle size (Figure 3), which is positively correlated with soil temperature (Figure S3B, Supporting information). Combined, Tamarack NIDGS may reflect an adaptation to the colder temperatures despite living at ~1,275m (relatively lower elevation for NIDGS total range). The highest adaptive differentiation in SIDGS was observed for Paddock (PA), with patterns of local adaptation particularly associated with grassland land cover rather than shrub/scrubland, as well as increase Annual Mean Temperature and Isothermality. Greater densities of SIDGS have been associated with increased cover of perennial grasses and diversity of native perennial plants (Lohr *et al.* 2013). Additionally, burrow density appears to be influenced by the presence of native forbs typical of shrub/scrubland, which was an environmental variable associated with adaptive variation of the remaining SIDGS populations (Figure 6C). Forbs have a large influence in protein intake for fat storage before hibernation, and thus it is a strong predictor of overwintering survival of all age classes, even when present at low densities (Barrett 2005). The results of the GEA could indicate a particular adaptation of Paddock to areas with lower forb density, but further sampling of other SIDGS populations, as well as diet studies would be needed to confirm this hypothesis.

We did not find particular Gene Ontology terms associated with putative functions of genes in our analysis, which could be related to a lack of annotations of the thirteen-lined ground squirrel genome, or also the

fact that no particular Gene Ontology term stands out as causative of the differentiation of the two species (Szkiba *et al.* 2014). Still, we found two non-synonymous substitutions in known genes identified by the RDA, one of which was also identified by *pcadapt*, the NPR1. This gene has been found to be highly expressed in brown adipose tissue of thirteen-lined ground squirrels during hibernation, relating to heat production during periodic arousals from hypothermic torpor (Hampton *et al.* 2013). Adaptations relating to adipose tissue are common in high altitude adapted rodents (Gossmann *et al.* 2019), and these associations suggest that there are adaptive differences between NIDGS and SIDGS relating to their hibernation patterns. In fact, one study on Columbian ground squirrel (*Urocitellus columbianus*) found that emergence timing was heritable and hence, squirrels' emergence date may also reflect local adaptations (Lane *et al.* 2011). In accordance to our hypothesis (d), altitude differences might have resulted in adaptive differences between NIDGS and SIDGS in production and storage of fat, and increased metabolism and oxygen transport at higher elevations, as seen in other mammals (Faherty *et al.* 2018; Waterhouse *et al.* 2018; Garcia-Elfring *et al.* 2019).

### *Conservation unit recommendations*

Both species have undergone accentuated population declines in the last 50 years (Yensen 1999; Gavin *et al.* 1999; Evans-Mack 2003), suggesting high susceptibility to environmental and landscape changes. As a result, NIDGS were listed as threatened under the federal Endangered Species Act (ESA) in 2000 (Clark 2000), and as Critically Endangered by the International Union for Conservation of Nature (Hafner 1998). SIDGS are listed as Vulnerable by the IUCN (IUCN 2018) and were a former candidate for listing under the ESA, although they were ultimately not listed. We made use of a minimally invasive sampling method, in order to minimize the impact of sampling on populations (Carroll *et al.* 2018). Buccal swabs, as used in this study, have been specifically used in high throughput sequencing (HTS) genotyping using targeted sequencing (Chang *et al.* 2007; McMichael *et al.* 2009), and have been shown to allow for non-targeted HTS genotyping, such as RAD sequencing, in amphibians (Peek *et al.* 2019). To our knowledge, this is the first study genotyping buccal swabs from a mammal with a non-targeted HTS approach, adding to the growing literature aiming to use less invasive sampling strategies to examine genomic variation in species of conservation concern (Andrews *et al.* 2018; Carroll *et al.* 2018).

The decline of these two species has resulted in metapopulations approaching a state of nonequilibrium, where many populations became small and isolated, and thus especially prone to both diversification and extinction (Harrison & Taylor 1997; Hanski & Gaggiotti 2004; Pironon *et al.* 2017). Metapopulations can retain genetic variation more readily than simply isolated subpopulations from a once panmictic population, and can be more resilient to extinction due to their intrinsic colonization-extinction-recolonization dynamics (Nee & May 1992; Levin 1995; Gavrillets *et al.* 2000). However, this resilience is dependent on how many individuals and how much genetic diversity each subpopulation maintains (Fahrig & Merriam 1994). NIDGS populations appear to have a more homogenous metapopulation in the western portion of their range based on low and mostly non-significant levels of pairwise  $F_{ST}$ , but form a more fragmented patch network in the eastern portion of their range, based on the neutral and adaptive loci (Figures 5 and 6, and S9, Supporting information). Overall we observed similar patterns of genetic structure as seen in previous studies for both NIDGS and SIDGS (Garner *et al.* 2005; Hoisington 2007).

Following the recommendations of Funk *et al.* (2012), we defined three evolutionarily significant units (ESUs) for NIDGS (Figure 6A). Within the ESU1, we identified two management and adaptive units (MUs and AUs, respectively), corresponding to the separation of Rocky Top from all other populations. Lower Butter was identified as a separate conservation unit, ESU2, which was distinctive at both the neutral and adaptive levels from all other populations. Within the ESU3, of the four populations sampled, three were considered separate MUs, while Lost Valley which was represented by a single individual showed mixed ancestry (Figure 6A). We also identified Tamarack as a distinct AU. Interestingly, ESU1 and ESU3 are also supported by previous work using mitochondrial data (Hoisington 2007; Hoisington-Lopez *et al.* 2012). The presence of long branches between haplogroups and absence of star-like pattern in the mitochondrial network suggests that these haplogroups diverged in the intermediate to distant past, allowing for enough

time for these populations to become demographically and adaptively differentiated (Zink & Barrowclough 2008; Hoisington-Lopez *et al.* 2012). However, the previous mitochondrial work found a third haplogroup for populations not sampled in this study, and also did not include Lower Butter. So additional mitochondrial and genomic work is necessary to validate and potentially add to the conservation unit assignments found in this study for NIDGS. For SIDGS, we identified two ESUs due to a clear separation of the populations east of the Weiser River with those from Olds Ferry (Figures 2 and 6, and S2, Supporting information). This separation is also evident at the neutral level (Table 4 and Figures 5 and 6), as Olds Ferry also represents a distinct MU from the remaining populations. Previous genetic studies of SIDGS included more sampling locations than our study, and they also detected strong genetic differentiation for populations east and west of the Weiser River using mtDNA sequence data and microsatellite loci (Garner *et al.* 2005; Hoisington 2007; Hoisington-Lopez *et al.* 2012). Paddock was clearly distinct as an AU within SIDGS (Figure 5, Figure S10B, Supporting information), but it did not show significant neutral divergence to be considered a distinct MU. There is evidence for the vulnerability of SIDGS populations due to the construction of roads using gravity models, and that might explain the significant reduction in genetic diversity in most populations east of the Weiser River (Table 4). These populations appear to be the most vulnerable, potentially as a result of low dispersal propensity and population declines associated with loss of habitat and consequently, connectivity (Barrett 2005; Panek 2005). This isolation could have resulted in further uniqueness in terms both neutral and adaptive diversity, and could explain the identification of Paddock as a separate AU for SIDGS.

Our results are important for developing management strategies in response to land use and habitat changes (Henry & Russello 2013). If translocations are warranted, they will likely be most effective if performed between populations within the same ESU and ideally also within the same AU, to allow for increased connectivity without compromising potential effects of local adaptation. Protecting AUs is essential to preserve adaptive diversity of populations which already reflects local adaptations that might be unique. However, further sampling is required to validate the results of those populations with lower number of samples, given that due to their low sample size it is less likely that adaptive variation would be detected (Lotterhos & Whitlock 2015). The inclusion of additional populations would also be ideal, not only for confirming the patterns found in this study, but also for identifying other conservation units that might warrant special protection.

## Conclusions

We developed new genomic tools to examine two rare and recently diverged species of ground squirrels, identifying multiple sets of new loci with around 3000 SNPs each using minimally invasive sampling methods. Our results corroborate previous studies regarding the clear differentiation between NIDGS and SIDGS and provide further details regarding the differentiation of these two sister species at the adaptive level. We additionally analyzed the demographic and adaptive variation of each species independently and determined that local adaptation played a more prominent role in differentiation among NIDGS populations, while geographic barriers appear to be the largest determinant of genetic differentiation in SIDGS. The different demographic and adaptive groups within both species may require dedicated management strategies in order to protect both neutral (MUs) and adaptive (AUs) diversity. In this context, we are the first to identify adaptive loci that distinguish not only the two species, but also distinguish among populations within each species, and the associations of these adaptive loci with environmental variables. The differences we detected will potentially help inform management plans aiming to protect the evolutionary and adaptive capacity of populations. Our results suggest clear metapopulation structure in both species with strong isolation-by-distance and isolation-by-barriers, as well as evidence of recent local adaptation within some of these small and isolated metapopulations.

## Acknowledgments

Data collection and analyses performed by the IBEST Genomics Resources Core at the University of Idaho were supported in part by NIH COBRE grant P30GM103324 and NSF EPSCOR (OIA-1757324). Research was supported by the College of Natural Resources at the University of Idaho, the U.S. Fish and Wildlife Service, and the U.S. Geological Survey. Greg Burak helped secure funding. Idaho Department of Fish and

Game (Bill Bosworth and Diane Evans Mack) provided samples for northern and southern Idaho ground squirrels. Payette National Forest provided funding that helped support collection of samples for northern Idaho ground squirrels. Laboratory support was provided by Jennifer Adams. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## References

- Ali OA, O'Rourke SM, Amish SJ *et al.* (2016) RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics* , **202** , 389–400.
- Andrews KR, De Barba M, Russello MA, Waits LP (2018) Advances in Using Non-invasive, Archival, and Environmental Samples for Population Genomic Studies. In: *Population Genomics* , pp. 1–37. Springer, Cham.
- Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* , **17** , 81–92.
- Assis J, Castilho Coelho N, Alberto F *et al.* (2013) High and distinct range-edge genetic diversity despite local bottlenecks (SJ Goldstien, Ed.). *PLoS ONE* , **8** , e68646.
- Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* , **3** , e3376.
- Barrett JS (2005) Population Viability of the Southern Idaho Ground Squirrel (*Spermophilus brunneus endemicus*): Effects of an Altered Landscape. Boise State University.
- Bateman A (2019) UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research* , **47** , D506–D515.
- Besnier F, Glover KA (2013) ParallelStructure: A R Package to Distribute Parallel Runs of the Population Genetics Program STRUCTURE on Multi-Core Computers (M Anisimova, Ed.). *PLoS ONE* , **8** , e70651.
- Bi K, Linderoth T, Singhal S *et al.* (2019) Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLoS Genetics* , **15** , e1008119.
- Burak GS (2011) *Status review for northern Idaho ground squirrel (Spermophilus brunneus brunneus)* . Boise, ID.
- Busscher KL (2009) Improving Success of Translocating Southern Idaho Ground Squirrels (*Spermophilus endemicus*) .
- Capblancq T, Luu K, Blum MGB, Bazin E (2018) Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources* , **18** , 1223–1233.
- Carroll EL, Bruford MW, DeWoody JA *et al.* (2018) Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications* , **11** , 1094–1119.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* , **22** , 3124–3140.
- Caye K, Jumentier B, Lepeule J, François O (2019) LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies (J Kelley, Ed.). *Molecular Biology and Evolution* , **36** , 852–860.
- CCAA (2002) *Southern Idaho Ground Squirrel Candidate Conservation Agreement with Assurances for the Soulen Livestock Company, Inc.*
- CCRS, CCMEQ, NRCAN *et al.* (2017) 2010 Land Cover of North America at 30 meters.
- Chang ML, Terrill RL, Bautista MM *et al.* (2007) Large-scale SNP genotyping with canine buccal swab DNA. *Journal of Heredity* , **98** , 428–437.

- Chen H, Boutros PC (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* , **12** , 35.
- Clark JR (2000) Endangered and threatened wildlife and plants; determination of threatened status for the northern Idaho ground squirrel. *Federal Register* , **65** , 17779–17786.
- CONSBIO (2015a) STATSGO Derived Soil Particle Size in 11 Western States in USA (Data Basin Dataset).1:250,000.
- CONSBIO (2015b) STATSGO Derived Soil Order in 11 Western States in USA (Data Basin Dataset).1:250,000.
- Dabney A, Storey JD, Warnes GR (2010) qvalue: Q-value estimation for false discovery rate control. R package version 1.0.
- Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics* , **27** , 2156–2158.
- Doebeli M, Dieckmann U (2003) Speciation along environmental gradients. *Nature* , **421** , 259–264.
- Eastman LM, Morelli TL, Rowe KC, Conroy CJ, Moritz C (2012) Size increase in high elevation ground squirrels over the last century. *Global Change Biology* , **18** , 1499–1508.
- ESRI (2013) Terrais: Slope Map.
- ESRI (2015) Land Formation.
- ESRI (2016) “World Biomass” [basemap].Scale not given. “World Biomass 2000 (1km)”. Aug 25, 2016.
- ESRI (2019) World Ecological Facets Landform Classes.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* , **14** , 2611–2620.
- Evans-Mack D (2003) *Northern Idaho ground squirrel population monitoring 2003 annual report* . McCall, ID.
- Faherty SL, Villanueva-Cañas JL, Blanco MB, Albà MM, Yoder AD (2018) Transcriptomics in the wild: hibernation physiology in free-ranging dwarf lemurs. *Molecular Ecology* , **27** , 709–722.
- Fahrig L, Merriam G (1994) Conservation of fragmented populations. *Conservation Biology* , **8** , 50–59.
- Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* , **37** , 4302–4315.
- Fišer C, Robinson CT, Malard F (2018) Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology* , **27** , 613–635.
- Fletcher NK, Acevedo P, Herman JS *et al.* (2019) Glacial cycles drive rapid divergence of cryptic field vole species. *Ecology and Evolution* , 1–13.
- Forester BR, Landguth EL, Hand BK, Balkenhol N (2018) Landscape Genomics for Wildlife Research. In: *Population Genomics* , pp. 1–40. Springer, Cham.
- Forester BR, Lasky JR, Wagner HH, Urban DL (2018) Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology* , **27** , 2215–2233.
- François O, Martins H, Caye K, Schoville SD (2016) Controlling false discoveries in genome scans for selection. *Molecular Ecology* , **25** , 454–469.
- French AR (1982) Intraspecific differences in the pattern of hibernation in the ground squirrel *Spermophilus beldingi* . *Journal of Comparative Physiology* , **148** , 83–91.

- Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution* , **27** , 489–496.
- Garcia-Elfring A, Barrett RDH, Millien V (2019) Genomic signatures of selection along a climatic gradient in the northern range margin of the white-footed mouse (*Peromyscus leucopus* ). *Journal of Heredity* , **110** , 684–695.
- Garner A, Rachlow JL, Waits LP (2005) Genetic diversity and population divergence in fragmented habitats: conservation of Idaho ground squirrels. *Conservation Genetics* , **6** , 759–774.
- Gavin TA, Sherman PW, Yensen E, May B (1999) Population genetic structure of the northern Idaho ground squirrel (*Spermophilus brunneus brunneus* ). *Journal of Mammalogy* , **80** , 156–168.
- Gavrilets S, Acton R, Gravner J (2000) Dynamics of speciation and diversification in a metapopulation. *Evolution* , **54** , 1493–1501.
- Gill AE, Yensen E (1992) Biochemical differentiation in the Idaho ground squirrel *Spermophilus brunneus* (Rodentia: Sciuridae). *Great Basin Natura* , **52** , 155–159.
- Goldberg AR (2018) Diet, disease, and hibernation behavior of northern Idaho ground squirrels. University of Idaho.
- Goldberg AR, Conway CJ (2021) Hibernation behavior of a federally-threatened ground squirrel: climate change and habitat selection implications. *Journal of Mammalogy* , in press.
- Goldberg AR, Conway CJ, Mack DE, Burak G (2020) Winter Versus Summer Habitat Selection in a Threatened Ground Squirrel. *Journal of Wildlife Management* , **84** , 1548–1559.
- Goldberg AR, Conway CJ, Tank DC *et al.* (2020) Diet of a rare herbivore based on DNA metabarcoding of feces: Selection, seasonality, and survival. *Ecology and Evolution* , **10** , 7627–7643.
- Gossmann TI, Shanmugasundram A, Börno S *et al.* (2019) Ice-Age Climate Adaptations Trap the Alpine Marmot in a State of Low Genetic Diversity. *Current Biology* , **29** , 1712-1720.e7.
- Goudet J (2005) hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes* , **5** , 184–186.
- Hafner DJ (1998) *North American Rodents: Status Survey and Conservation Action Plan* . IUCN.
- Hampton M, Melvin RG, Andrews MT (2013) Transcriptomic analysis of brown adipose tissue across the physiological extremes of natural hibernation (K Dahlman-Wright, Ed.). *PLoS ONE* , **8** , e85157.
- Hanski I (1998) Metapopulation dynamics. *Nature* , **396** , 41–49.
- Hanski I, Gaggiotti OE (2004) *Ecology, Genetics, and Evolution of Metapopulations* . Elsevier Academic Press, Burlington, MA, USA.
- Harrison S, Taylor AD (1997) Empirical evidence for metapopulation dynamics. In: *Metapopulation Biology: Ecology, Genetics, and Evolution* (eds Hanski I, Gilpin M), pp. 27–42. Academic Press, San Diego, CA, USA.
- Helgen KM, Cole FR, Helgen LE, Wilson DE (2009) Generic revision in the Holarctic ground squirrel genus *Spermophilus*. *Journal of Mammalogy* , **90** , 270–305.
- Helmstetter NA, Conway CJ, Stevens BS, Goldberg AR (2021) Balancing transferability and complexity of species distribution models for rare species conservation (Y Fourcade, Ed.). *Diversity and Distributions* , **27** , 95–108.
- Henry P, Russello MA (2013) Adaptive divergence along environmental gradients in a climate-change-sensitive mammal. *Ecology and Evolution* , **3** , 3906–3917.

- Hodgson JA, Moilanen A, Wintle BA, Thomas CD (2011) Habitat area, quality and connectivity: striking the balance for efficient conservation. *Journal of Applied Ecology* , **48** , 148–152.
- Hoffmann AA, Sgrò CM, Kristensen TN (2017) Revisiting adaptive potential, population size, and conservation. *Trends in Ecology and Evolution* , **32** , 506–517.
- Hoisington-Lopez JL, Waits LP, Sullivan J (2012) Species limits and integrated taxonomy of the Idaho ground squirrel (*Urocitellus brunneus*): genetic and ecological differentiation. *Journal of Mammalogy* , **93** , 589–604.
- Hoisington J (2007) Conservation genetics, landscape genetics and systematics of the two subspecies of the endemic Idaho ground squirrel (*Spermophilus brunneus*). PhD Thesis, University of Idaho.
- Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology* , **21** , 797–807.
- Hope AG, Malaney JL, Bell KC *et al.* (2016) Revision of widespread red squirrels (genus: *Tamiasciurus*) highlights the complexity of speciation within North American forests. *Molecular Phylogenetics and Evolution* , **100** , 170–182.
- Hope AG, Panter N, Cook JA, Talbot SL, Nagorsen DW (2014) Multilocus phylogeography and systematic revision of North American water shrews (genus: *Sorex*). *Journal of Mammalogy* , **95** , 722–738.
- Hope AG, Stephens RB, Mueller SD, Tkach V V, Demboski JR (2020) Speciation of North American pygmy shrews (Eulipotyphla: Soricidae) supports spatial but not temporal congruence of diversification among boreal species. *Biological Journal of the Linnean Society* , **129** , 41–60.
- Hut RA, Beersma DGM (2011) Evolution of time-keeping mechanisms: early emergence and adaptation to photoperiod. *Philosophical Transactions of the Royal Society B: Biological Sciences* , **366** , 2141–2154.
- IUCN SSMSG (2000) *Urocitellus brunneus*. *The IUCN Red List of Threatened Species* .
- IUCN SSMSG (2018) *Urocitellus endemicus*. *The IUCN Red List of Threatened Species* .
- Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* , **27** , 3070–3071.
- Kassambara A (2018) ggpubr: “ggplot2” based publication ready plots. R package version 0.1.7.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* , **7** , 1225–1241.
- Landguth EL, Balkenhol N (2012) Relative sensitivity of neutral versus adaptive genetic data for assessing population differentiation. *Conservation Genetics* , **13** , 1421–1426.
- Lane JE, Kruuk LEB, Charmantier A *et al.* (2011) A quantitative genetic analysis of hibernation emergence date in a wild population of Columbian ground squirrels. *Journal of Evolutionary Biology* , **24** , 1949–1959.
- Levin DA (1995) Metapopulations: an arena for local speciation. *Journal of Evolutionary Biology* , **8** , 635–644.
- Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* , 1303.3997.
- Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , **25** , 2078–2079.
- Lohr K, Yensen E, Munger JC, Novak SJ (2013) Relationship between habitat characteristics and densities of southern Idaho ground squirrels. *The Journal of Wildlife Management* , **77** , 983–993.

- Lopes CM, de Freitas TRO (2012) Human impact in naturally patched small populations: genetic structure and conservation of the burrowing rodent, tuco-tuco (*Ctenomys lami*). *The Journal of Heredity* , **103** , 672–681.
- Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology* , **24** , 1031–1046.
- Luu K, Bazin E, Blum MGB (2017) pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources* , **17** , 67–77.
- Macdonald SL, Llewelyn J, Phillips BL (2018) Using connectivity to identify climatic drivers of local adaptation. *Ecology Letters* , **21** , 207–216.
- Marchenko S (2011) GIPL1.3 simulated mean annual ground temperature (MAGT) in Celsius averaged for particular decade for the entire Alaskan permafrost domain. NAD83, Alaska Albers projection.
- Marques DA, Meier JI, Seehausen O (2019) A combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution* .
- McLaren W, Gil L, Hunt SE *et al.* (2016) The Ensembl Variant Effect Predictor. *Genome Biology* , **17** , 122.
- McMichael GL, Gibson CS, O’Callaghan ME *et al.* (2009) DNA from buccal swabs suitable for high-throughput SNP multiplex analysis. *Journal of Biomolecular Techniques* , **20** , 232–235.
- Mitt M, Kals M, Pärn K *et al.* (2017) Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *European Journal of Human Genetics* , **25** , 869–876.
- Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W (2007) Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Research* , **17** , 413–421.
- Nee S, May RM (1992) Dynamics of metapopulations: habitat destruction and competitive coexistence. *Journal of Animal Ecology* , **61** , 37–40.
- Nei M (1972) Genetic distance between populations. *The American Naturalist* , **106** , 283–292.
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* , **89** , 583–590.
- Oksanen J, Blanchet FG, Kindt R *et al.* (2011) vegan: Community Ecology Package. , R package version 2.0–5.
- Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* , **22** , 5983–5999.
- Palsbøll PJ, Bé Rubé M, Allendorf FW (2006) Identification of management units using population genetic data. *Trends in Ecology and Evolution* , **22** , 11–16.
- Panek KC (2005) Dispersal, Translocation, and Population Connectivity in Fragmented Populations of Southern Idaho Ground Squirrels. Boise State University.
- Peek RA, Bedwell M, O’Rourke SM *et al.* (2019) Hybridization between two parapatric ranid frog species in the northern Sierra Nevada, California, USA. *Molecular Ecology* , mec.15236.
- Peres-Neto PR, Jackson DA, Somers KM (2003) Giving meaningful interpretation to ordination axes: Assessing loading significance in principal component analysis. *Ecology* , **84** , 2347–2363.
- Peterson GB, Carl P, Boudt K *et al.* (2014) PerformanceAnalytics: econometric tools for performance and risk analysis. *R package version, 1(3541)* , 107.

- Pironon S, Papuga G, Villellas J *et al.* (2017) Geographic variation in genetic and demographic performance: new insights from an old biogeographical paradigm. *Biological Reviews* , **92** , 1877–1909.
- Primmer CR, Papakostas S, Leder EH, Davis MJ, Ragan MA (2013) Annotated genes and nonannotated genomes: cross-species use of Gene Ontology in ecology and evolution research. *Molecular Ecology* , **22** , 3216–3241.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* , **155** , 945–959.
- R Core Team (2013) R: A language and environment for statistical computing. , 201.
- Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* , **24** , 4348–4370.
- Rousset F, Raymond M (1995) Testing heterozygote excess and deficiency. *Genetics* , **140** , 1413–1419.
- Rubidge EM, Patton JL, Lim M *et al.* (2012) Climate-induced range contraction drives genetic erosion in an alpine mammal. *Nature Climate Change* , **2** , 285–288.
- Rundell RJ, Price TD (2009) Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution* , **24** , 394–399.
- Santos MJ, Smith AB, Thorne JH, Moritz C (2017) The relative influence of change in habitat and climate on elevation range limits in small mammals in Yosemite National Park, California, U.S.A. *Climate Change Responses* , **4** , 7.
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics* , **14** , 807–820.
- Sherman PW, Runge MC (2002) Demography of a population collapse: the northern Idaho ground squirrel (*Spermophilus brunneus brunneus* ). *Ecology* , **83** , 2816–2831.
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* , **236** , 787–792.
- Smith AB, Beever EA, Kessler AE *et al.* (2019) Alternatives to genetic affinity as a context for within-species response to climate. *Nature Climate Change* , **9** , 787–794.
- Steinberg EK, Patton JL, Lacey EA (2000) Genetic structure and the geography of speciation in subterranean rodents: opportunities and constraints for evolutionary diversification. In: *Life underground, the biology of subterranean rodents* (eds Lacey E, Patton JL, Cameron GN), pp. 301–331. University of Chicago Press, Chicago, IL, USA.
- Suronen EF, Newingham BA (2013) Restoring habitat for the northern Idaho ground squirrel (*Urocitellus brunneus brunneus* ): Effects of prescribed burning on dwindling habitat. *Forest Ecology and Management* , **304** , 224–232.
- Szkiba D, Kapun M, von Haeseler A, Gallach M (2014) SNP2GO: functional analysis of genome-wide association studies. *Genetics* , **197** , 285–289.
- USFWS USF and WS (2000) *Endangered and threatened wildlife and plants; Determination of threatened status for the northern Idaho ground squirrel* .
- USFWS (2003) *Recovery plan for the Northern Idaho ground squirrel ( Spermophilus brunneus brunneus )* . Portland, OR.
- USFWS (2015) *Federal Register, Vol. 80, No. 120. Rules and Regulations.*
- USGS (2012) National Elevation Dataset, 1 arc-second.

Waits LP, Cushman SA, Spear SF (2016) Applications of landscape genetics to connectivity research in terrestrial animals. In: *Landscape Genetics: Concepts, Methods, Applications* (eds Balkenhol N, Cushman SA, Storfer AT, Waits LP), pp. 199–219. Wiley-Blackwell, Oxford, UK.

Waterhouse MD, Erb LP, Beever EA, Russello MA (2018) Adaptive population divergence and directional gene flow across steep elevational gradients in a climate-sensitive mammal. *Molecular Ecology* , **27** , 2512–2528.

Weeks AR, Sgrò CM, Young AG *et al.* (2011) Assessing the benefits and risks of translocations in changing environments: a genetic perspective. *Evolutionary Applications* , **4** , 709–725.

Wei T, Simko V, Levy M (2017) corrplot: Visualization of a correlation matrix.

Wittmann MJ, Stuis H, Metzler D (2018) Genetic Allee effects and their interaction with ecological Allee effects. *Journal of Animal Ecology* , **87** , 11–23.

Wood JLA, Yates MC, Fraser DJ (2016) Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. *Evolutionary Applications* , **9** , 640–657.

Wright BR, Grueber CE, Belov K, Hogg CJ, McLennan EA (2019) Too much of a good thing? Finding the most informative genetic dataset to answer conservation questions. *Molecular Ecology Resources* , 0–2.

Yates AD, Achuthan P, Akanni W *et al.* (2020) Ensembl 2020. *Nucleic Acids Research* , **48** , D682–D688.

Yensen E (1991) Taxonomy and distribution of the Idaho ground squirrel, *Spermophilus brunneus* . *Journal of Mammalogy* , **72** , 583–600.

Yensen E (1999) *Population survey of the southern Idaho ground squirrel*, *Spermophilus brunneus* endemicus. Caldwell, ID.

Yensen E, Dyni EJ (2020) Why is the northern Idaho ground squirrel rare? *Northwest Science* , **94** , 1–23.

Yensen E, Sherman PW (1997) *Spermophilus brunneus* . *Mammalian Species* , 1–5.

Yensen E, Tarifa T (2012) Can southern Idaho ground squirrels be translocated successfully? *Annual Report, Zoo Boise Conservation Fund* , 1–37.

Yensen E, Tarifa T, Clayton Z (2010) Can southern Idaho ground squirrels be translocated successfully? *Annual Report, Zoo Boise Conservation Fund* , 1–36.

Zero VH, Barocas A, Jochimsen DM *et al.* (2017) Complementary network-based approaches for exploring genetic structure and functional connectivity in two vulnerable, endemic ground squirrels. *Frontiers in Genetics* , **8** , 1–13.

Zheng X, Levine D, Shen J *et al.* (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* , **28** , 3326–3328.

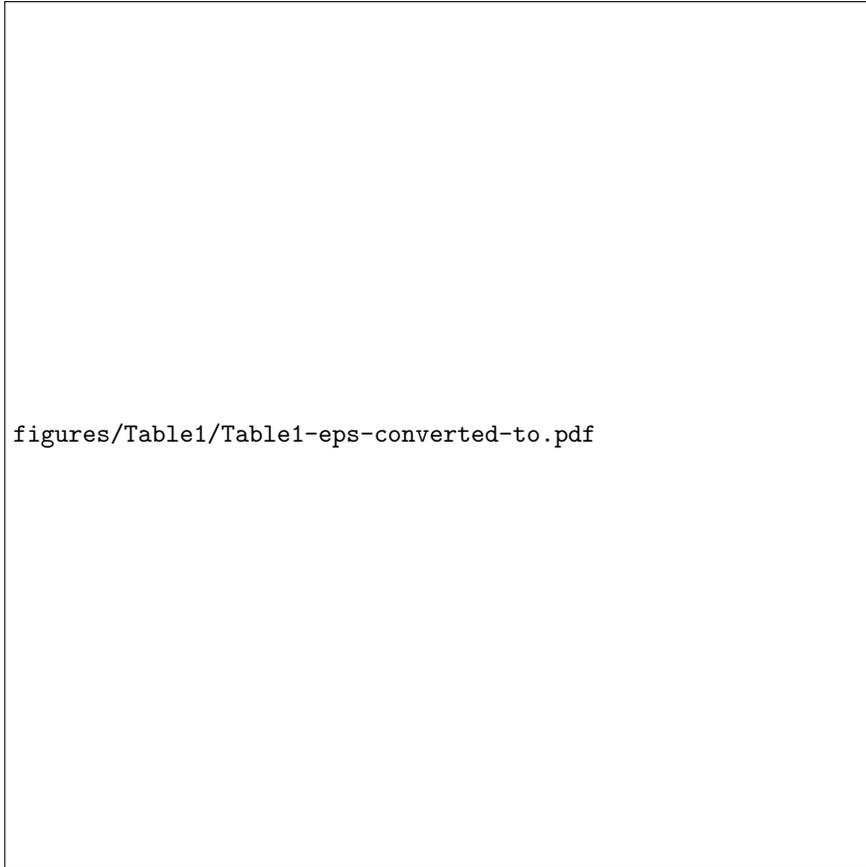
Zink R, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* , 1–15.

### Data Accessibility

Mapped bam files for the total 304 ground squirrels are available on NCBI's public Sequence Read Archive (XXX). Additional meta-data for each individual ground squirrel, and respective scores for all environmental variables tested can be found in DRYAD, as well as the final three VCF files used for all analyses with respective R code employed.

### Author Contributions

CC and LPW designed the study; ARG and LPW participated in sample collection; SB, KRA and DSG prepared the DNA for RAD-sequencing, SB processed the RAD-seq data and performed all analyses. SB wrote the manuscript and all authors contributed to editing and revising the manuscript.



figures/Table2/Table2-eps-converted-to.pdf

figures/Table3/Table3-eps-converted-to.pdf

figures/Table4/Table4-eps-converted-to.pdf

figures/Figure1/Figure1-eps-converted-to.pdf

figures/Figure2/Figure2-eps-converted-to.pdf

figures/Figure3/Figure3-eps-converted-to.pdf

figures/Figure4/Figure4-eps-converted-to.pdf

figures/Figure5/Figure5-eps-converted-to.pdf

figures/Figure6/Figure6-eps-converted-to.pdf