

RESEARCH ARTICLE

Development and validation of a GT-seq panel for genetic monitoring in a threatened species using minimally invasive sampling

Molly J. Garrett¹  | Stacey A. Nerkowski¹ | Shannon Kieran¹  | Nathan R. Campbell² |
Soraia Barbosa¹  | Courtney J. Conway³  | Paul A. Hohenlohe⁴  | Lisette P. Waits¹ 

¹Department of Fish and Wildlife Sciences, College of Natural Resources, University of Idaho, Moscow, Idaho, USA

²GTseek LLC, Twin Falls, Idaho, USA

³U.S. Geological Survey, Idaho Cooperative Fish & Wildlife Research Unit, University of Idaho, Moscow, Idaho, USA

⁴Department of Biological Sciences, College of Science, University of Idaho, Moscow, Idaho, USA

Correspondence

Molly J. Garrett, Department of Fish and Wildlife Sciences, College of Natural Resources, University of Idaho, Moscow, ID, USA.

Email: mollyjgarrett@gmail.com

Funding information

National Science Foundation, Grant/Award Number: OIA-1757324; U.S. Fish and Wildlife Service; U.S. Forest Service; Idaho Department of Fish and Game

Abstract

Minimally invasive samples are often the best option for collecting genetic material from species of conservation concern, but they perform poorly in many genomic sequencing methods due to their tendency to yield low DNA quality and quantity. Genotyping-in-thousands by sequencing (GT-seq) is a powerful amplicon sequencing method that can genotype large numbers of variable-quality samples at a standardized set of single nucleotide polymorphism (SNP) loci. Here, we develop, optimize, and validate a GT-seq panel for the federally threatened northern Idaho ground squirrel (*Urocitellus brunneus*) to provide a standardized approach for future genetic monitoring and assessment of recovery goals using minimally invasive samples. The optimized panel consists of 224 neutral and 81 putatively adaptive SNPs. DNA collected from buccal swabs from 2016 to 2020 had 73% genotyping success, while samples collected from hair from 2002 to 2006 had little to no DNA remaining and did not genotype successfully. We evaluated our GT-seq panel by measuring genotype discordance rates compared to RADseq and whole-genome sequencing. GT-seq and other sequencing methods had similar population diversity and F_{ST} estimates, but GT-seq consistently called more heterozygotes than expected, resulting in negative F_{IS} values at the population level. Genetic ancestry assignment was consistent when estimated with different sequencing methods and numbers of loci. Our GT-seq panel is an effective and efficient genotyping tool that will aid in the monitoring and recovery of this threatened species, and our results provide insights for applying GT-seq for minimally invasive DNA sampling techniques in other rare animals.

KEYWORDS

conservation genomics, northern Idaho ground squirrel, single nucleotide polymorphisms, SNPs, targeted amplicon sequencing, *Urocitellus brunneus*

TAXONOMY CLASSIFICATION

Conservation genetics, Genomics, Population genetics

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Rapid advancements of genomic sequencing technologies have provided increased resolution and identification of neutral and adaptive genomic variants, which has improved our ability to monitor and manage wildlife populations and potentially mitigate the effects of environmental change and biodiversity loss (Hohenlohe et al., 2021; Walters & Schwartz, 2020). Genetic monitoring studies investigate both demographic and genetic parameters across time, often with noninvasive or minimally invasive samples (Schwartz et al., 2007). Noninvasive samples are left behind by an organism and are collected without having to catch, disturb, or sometimes even observe individuals, while minimally invasive samples are collected often through capture but via less invasive methods than blood or tissue extraction (Beja-Pereira et al., 2009; Carroll et al., 2018; Taberlet et al., 1999; Waits & Paetkau, 2005). Due to their tendency to yield low DNA quality and quantity, noninvasive and minimally invasive samples are not the best options for many next-generation sequencing approaches (Andrews et al., 2018; but see Russello et al., 2015). Restriction site-association DNA sequencing (RADseq) is one of the most common next-generation sequencing techniques used for single nucleotide polymorphism (SNP) discovery and genotyping of non-model organisms. RADseq targets a subset of the genome at a greater depth of coverage per locus than whole genome sequencing techniques, which is beneficial for genotyping large numbers of samples (Andrews et al., 2016). However, RADseq requires high-molecular-weight genomic DNA (Graham et al., 2015), which may be limiting if samples can only be collected through non-invasive or minimally invasive techniques. Therefore, targeted amplicon sequencing is one of several new techniques that provide powerful tools for SNP genotyping for low-quality samples. Amplicon sequencing generates PCR products using template-specific primers, which can have a high success rate for low-quality samples (Andrews et al., 2018).

Genotyping-in-thousands by sequencing (GT-seq; Campbell et al., 2015) is a type of amplicon sequencing that involves the development of multiplexed PCR primers. These primers target 50 to 100s of preidentified loci that are informative for a specific suite of conservation objectives. Compared to other genotyping-by-sequencing methods, GT-seq is cost effective and while originally designed on tissue samples (Campbell et al., 2015), it has worked on minimally and non-invasive samples. For example, GT-seq panels have been created and used on polar bear (*Ursus maritimus*) fecal samples (Hayward et al., 2022), western-rattlesnake (*Crotalus oreganus*) cloacal swabs and skin sheds (Schmidt et al., 2020), and Sitka black-tailed deer (*Odocoileus hemionus sitkensis*) hair and fecal samples (Burgess et al., 2022).

GT-seq panels require the previous identification of the loci of interest, generally through RAD sequencing or other approaches using high quality DNA samples. The samples used in primer and panel development should be chosen to avoid ascertain bias and to be representative of the populations in which the panel will be used, which sometimes requires that data from lower-quality samples are

also used in panel development. Additionally, an important step in panel validation is to estimate and minimize genotype discordance between GT-seq and the genotyping performed to identify the loci for the GT-seq panel. Studies that compared genotypes for the same individuals at the same loci generated with both RADseq and GT-seq report discordance rates ranging from 1.77% to 3.40% (Hayward et al., 2022; Schmidt et al., 2020; Setzke et al., 2021).

Northern Idaho ground squirrels (*Urocitellus brunneus*, hereafter NIDGS) are endemic to Idaho and federally listed as a threatened species (USFWS, 2003) with an estimated population size of 2000–3000 individuals (Wagner & Evans Mack, 2020). NIDGS occupy xeric meadows and small rocky outcroppings within coniferous forests of central Idaho. They persist within only a small fraction of their former range (Helmstetter et al., 2021) and numerous causes have been proposed to explain their rarity (Sherman & Runge, 2002; Suronen & Newingham, 2013; USFWS, 2003; Yensen & Dyni, 2020). In 2003, the United States Fish and Wildlife Service (USFWS) produced a NIDGS recovery plan that delineated actions and objectives to recover and protect the species (USFWS, 2003). One objective of the recovery plan was to more accurately define NIDGS metapopulation structure and conservation units (sensu Funk et al., 2012). The plan outlines the need to understand connectivity among metapopulations (occupied sites) and the landscape factors that impede or facilitate gene flow, as well as an improved estimate of effective population size (N_e).

NIDGS are semi-colonial and occur in distinct sites interspersed within a matrix of unoccupied areas. Many of these occupied sites have been genetically sampled since 2002, but due to the threatened status of the species, managers require less stressful, minimally invasive genetic sampling techniques including plucking hair and taking buccal swabs. Hair samples collected in 2002 and 2006 were used to evaluate genetic diversity and gene flow based on mitochondrial DNA sequencing and nuclear DNA microsatellites (Garner et al., 2005; Hoisington-Lopez et al., 2012; Zero et al., 2017). Buccal swabs collected in 2016 were sequenced with RADseq (Barbosa et al., 2021). This was the first application of genomic sequencing to NIDGS samples, which provided novel information about population structure and neutral and adaptive genetic variation across the NIDGS sites. For instance, six genetic groups were identified among 13 sampling areas using neutral loci and three distinct groups were identified using adaptive loci. Elevation appeared to be the main driver of adaptive differentiation between sites, as SNPs were associated with environmental variables such as slope, ridges, and peaks (Barbosa et al., 2021). Further, gene ontology enrichment analyses of adaptive loci found no evidence for enrichment of specific cellular components, molecular functions, or biological processes (Barbosa et al., 2021). However, this study had small sample sizes for some of the sampled sites because many of the minimally invasive samples did not produce sufficient DNA quality or quantity for RADseq. Thus, new SNP analyses that incorporate all previously sampled populations of NIDGS would greatly improve the inferences possible when assessing genetic diversity, gene flow among extant populations, and differentiation and adaptive differences among populations.

The main objective of this study was to develop a GT-seq panel for genetic monitoring of this rare federally threatened ground squirrel using minimally invasive samples. Neutral and putatively adaptive SNPs identified through the previous RADseq study (Barbosa et al., 2021) were selected for the GT-seq panel, which will be applied to assess objectives set by the USFWS species recovery plan for the NIDGS. The specific goals of this study were to: (1) develop and optimize a set of SNP loci for a GT-seq panel, (2) validate the panel by comparing against RADseq and whole genome sequencing (WGS) data, and (3) evaluate the factors that produce genotype mismatches between sequencing methods.

2 | METHODS

2.1 | Sample collection and DNA extraction

As part of larger, ongoing studies on habitat use and species delimitation, we extracted DNA from a total of 1086 samples, from hair, buccal swab, or tissue samples collected from NIDGS captured at occupied sites spanning two decades (Table S1). Sample collection

periods were 2002–2006 (Garner et al., 2005; Hoisington-Lopez et al., 2012), 2016 (Barbosa et al., 2021), and 2020. Some NIDGS sites were sampled during each collection period (Figure 1, Table S1), with the 2020 collection focused on obtaining samples from sites that had never been genetically sampled. Across all collection periods, Tomahawk live traps were baited and placed near burrows or logs at regularly spaced intervals within occupied sites. All NIDGS were trapped and handled following protocols and procedures approved by the University of Idaho Animal Care and Use Committee [IACUC #2006-35 (Garner et al., 2005; Hoisington-Lopez et al., 2012) and #2015-53 (Barbosa et al., 2021)]. During 2002–2006, we plucked approximately 50–60 hairs from the tail of each squirrel, and we stored the hairs in an envelope with silica until DNA extraction. During 2016–2020, we swabbed the inside of each captured squirrel's cheek with a sterile cotton swab (Lakewood Biomedical) to collect epithelial cells, which was repeated five times per individual. All replicate buccal swabs per individual were preserved in the same tube in Qiagen ATL buffer until DNA extraction. Separate from the above long-term dataset, 28 tissue samples were opportunistically collected since 2004, mostly through road-kill collections. Twenty-one of these tissue samples were sequenced with whole genome

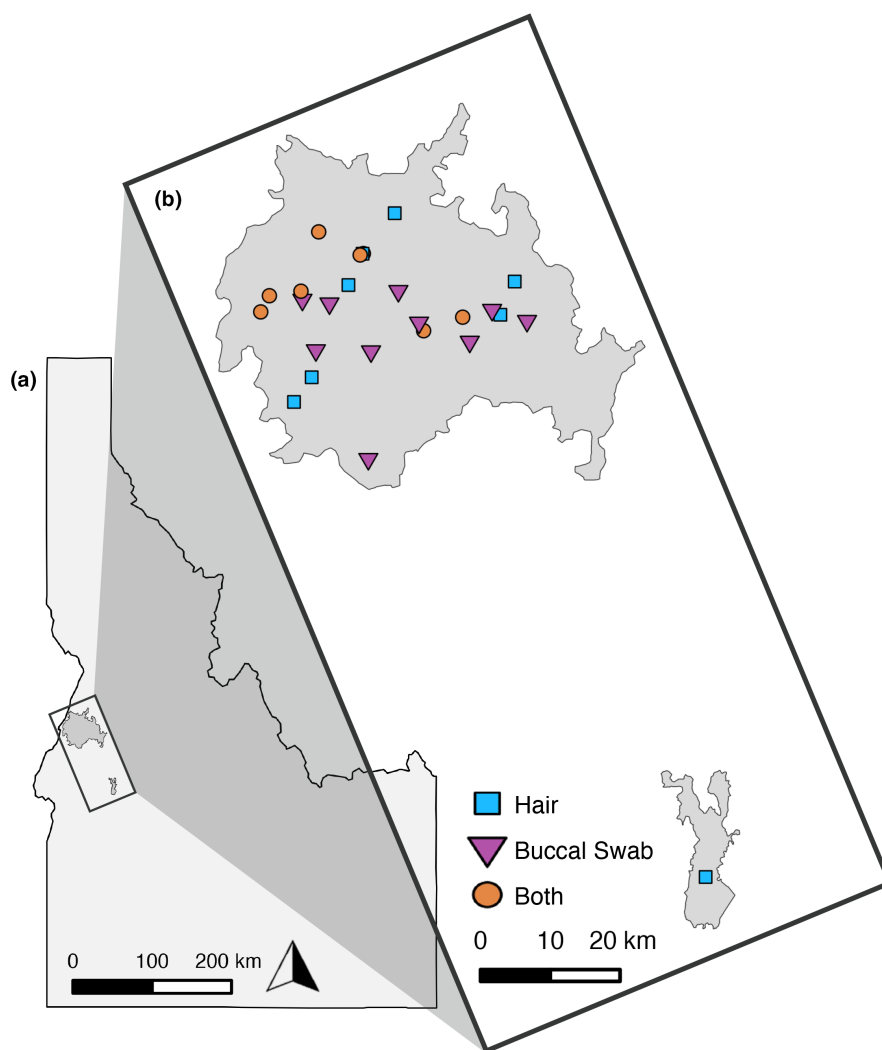


FIGURE 1 Sample distribution of hair and buccal swab samples from NIDGS in central Idaho. (a) Outline of Idaho with range of NIDGS in dark gray. (b) Detail of NIDGS range with sample types collected from each site denoted by color and shape. NIDGS range obtained from <https://ecos.fws.gov/ecp/species/2982>.

sequencing and the GT-seq panel to validate the data produced from the panel but were not included in the population genetic analyses, because they were collected outside of our study's temporal range.

We used Qiagen Blood and Tissue extraction kits (Qiagen, Inc.) to extract DNA from hair, buccal swab, and tissue samples using standard protocols (Barbosa et al., 2021; Garner et al., 2005; Hoisington-Lopez et al., 2012). The hair and buccal swab samples were extracted in the low quantity DNA facility with no PCR products or concentrated DNA. We evaluated DNA concentration of 207 hair and 454 buccal swab extracts with Qubit fluorometer quantification (Thermo Fisher Scientific). We further investigated the DNA quality and quantity of the hair extracts through gel electrophoresis ($N=30$), genomic DNA (gDNA) fragment analysis ($N=9$), and through the amplification of one microsatellite locus (GS17 145–167 bp, $N=53$).

2.2 | SNP selection and primer design

We used all 131 putatively adaptive loci that had been identified through partial redundancy analysis (pRDA) (Barbosa et al., 2021) for primer development to ensure we captured all identified adaptive loci for future analyses. Briefly, pRDA is a multivariate approach to estimate genotype–environment associations while accounting for population structure. To select neutral GT-seq candidate SNPs, we filtered from the 2663 neutral NIDGS SNPs identified by Barbosa et al. (2021) using the following parameters: <0.4 missing data, >0.1 expected heterozygosity, and 1 SNP per scaffold (3475 total scaffolds, 3444 neutral only scaffolds). This generated 510 neutral loci for panel optimization. GTseek LLC (<https://gtseek.com/>) designed the GT-seq primers for all candidate SNPs using the consensus sequence from the RADseq data from Barbosa et al. (2021) as flanking sequence around targeted SNPs. Primers were designed to be 18–24 bases in length with an annealing temperature range between 58 and 63°C before the addition of Illumina adapter sequences. Designed primer sets were then further filtered if stable hairpin structures had melting temperatures above 50°C or if primer heterodimers were predicted to form with other primers. Primers were excluded if they reported annealing temperatures to other primer sequences of 50°C or higher. As a final filtering step, primers were excluded if there were any binding sites within any of the expected amplicon sequences.

2.3 | GT-seq library preparation and panel development

We first produced one library following the standard GT-seq library protocol in Campbell et al. (2015), but this library was unsuccessful. For all of the following libraries, we followed a low-quality DNA GT-seq protocol modified from Campbell et al. (2015) as follows: the PCR cocktail for PCR amplification of target loci included 2 μ L of input gDNA for buccal swabs and 3 μ L for hair samples, 3.5 μ L of

Qiagen Plus multiplex master mix, and 1.5 μ L of pooled primer mix (final primer concentration at each locus \approx 160 nM). Thermal cycling was conducted in 96-well plates, and cycles in PCR1 were increased from 10 to 16 cycles: [95°C – 15 min; 5 cycles (95°C – 30 s, 5% ramp down to 57°C – 30 s, 72°C – 2 min); 16 cycles (95°C – 30 s, 65°C – 20 s, and 72°C – 30 s) 4°C hold]. The product of PCR1 was diluted 10-fold, and 3 μ L of diluted PCR1 product was transferred to a new 96-well plate for PCR2. With a multichannel pipette, 1 μ L of 10 μ M well-specific i5 tag primers were added, and 1 μ L of plate-specific tagging primers and 5 μ L of Qiagen Plus multiplex master mix were added to each well. Thermal cycling conditions for PCR2 were [95°C – 15 min; 10 cycles (98°C – 10 s, 65°C – 30 s, 72°C – 30 s); 72°C – 5 min, 4°C hold]. After PCR2, only libraries with buccal swab samples were normalized using the Just-a-Plate™ 96 PCR Normalization and Purification kit (Charm Biotech LLC) according to the manufacturer's instructions. Next, 10 μ L of each sample per 96-well plate was pooled into a 1.5 mL Eppendorf tube. We then performed a two-step purification with Agencourt® AMPure® XP magnetic beads to remove smaller, unwanted fragments below 150 bp and above 300 bp. After mixing and placement on a magnetic rack, the supernatant was discarded, and the immobilized beads were washed twice with 70% ethanol. The purified libraries were eluted with 30 μ L of low TE.

Two rounds of optimization determined which primers would be included in the panel. The first optimization used samples that were sequenced with RADseq in Barbosa et al. (2021) while the second round included samples from the RADseq study and new samples that had not been genotyped before and had high DNA concentrations. After the first and second optimizations, primers were removed based on output plots from *GTseq_SeqTest.pl* and *GTseq_Primer-Interaction-Test.pl* (<https://github.com/GTseq>). Output plots from these scripts reported primers responsible for primer hetero-dimer artifacts, proper amplicon targets, and primer sets that produced no sequences. Primers were omitted if they had no amplification or were otherwise negatively affecting genotyping success or target capture rates (e.g., if they were over-abundant or had off target amplification).

A third round of optimization determined the concentration of each primer. Using the candidate primer set defined after the second optimization, two formulations were created with varying concentrations of each primer. The first was a standard primer mix with equal concentrations of each primer (following Campbell et al., 2015), with 160 nM per primer in the pooled primer mix. The second was a varied primer concentration mix, where 80 nM was used for primers that performed in the top 25% of the second optimization, 320 nM was used for primers that performed in the bottom 25% of the second optimization, and 160 nM for all other primers. Both primer mixes were applied to the same high-quality samples.

GT-seq libraries were sequenced using a single-read 118 bp cycle on Illumina NovaSeq SP at the University of Oregon Genomics and Cell Characterization Core Facility (G3CF) and a paired-end 75 bp cycle on Illumina NextSeq at Seqmatic LLC (San Francisco, CA). We sequenced 13 libraries: five of hair extracts only, six of buccal swab extracts only, one mixed of hair and buccal swab extracts, and one of tissue extracts only. The GT-seq pipeline on GitHub (<https://github.com>)

com/GTseq) was used to process raw sequence data, and only the forward sequencing reads from the paired-end data were genotyped to match the single-read data. Briefly, the *GTseq_Genotyper_v2.pl* perl script uses in silico probe sequences to count the occurrence of each allele within individual fastq files. Genotypes for each locus are created by the ratio of allele 1 to allele 2, where allele ratios >10 are homozygous for allele 1, ratios <0.1 are homozygous for allele 2, and ratios between 0.2 and 5.0 are heterozygous. Loci with <10 reads are not genotyped (Campbell et al., 2015). This genotyping method will be referred to as the allele ratio (AR) pipeline.

2.4 | Tissue whole genome sequencing and genotyping

We sequenced 21 tissue samples using whole-genome resequencing protocols. Illumina library preparation was undertaken by the University of Idaho Genomics and Bioinformatics Resources Core (GBRC; Moscow, Idaho) using the Illumina Nextera library preparation kit following manufacturer's protocols. Samples were PE150-sequenced on a single S4 lane of a Novaseq 4000X by the University of Oregon G3CF. Samples averaged 114M reads/sample before filtering.

For this project, samples were genotyped exclusively at the loci used for GT-seq sequencing. First, samples were aligned to the same genome used when genotyping the RADseq dataset, downloaded from NCBI (SpeTri2.0, accession #GCF_000236235.1, accessed 7/09/2023). Sequences were aligned using *bwa mem* (Li & Durbin, 2010). Alignments were sorted and filtered to remove duplicates and exclude improperly paired reads. Genotypes were called in *angsd* 0.981 (Korneliusson et al., 2014) using the following settings: -minDepth 126 (avg 6x), -minInds 16 (76%), -minMaf 0.05, -postCutoff 0.95 -snp-pval 1e-6, -minQ 20, -minMapQ 10. A sites file with the positions of the GT-seq loci was passed to *angsd* along with the list of input bamfiles.

2.5 | Genotyping discordance

To validate the GT-seq panel, we compared two sets of samples: buccal swabs that were genotyped by both RADseq and GT-seq and tissue samples genotyped by both WGS and GT-seq.

We calculated genotyping discordance by extracting filtered (>50% missing data) GT-seq loci from the RADseq data and comparing genotypes at those loci for 53 individuals that were sequenced with both RADseq and GT-seq. In Barbosa et al. (2021), RADseq SNPs were genotyped using the multinomial maximum-likelihood framework STACKS (Catchen et al., 2013), and VCFTOOLS (Danecek et al., 2011) was used to exclude individuals with ≥50% missing data and include biallelic SNPs with <50% missing data located >10,000bp apart with >2% minor allele frequency and >3 reads. This genotyping pipeline will be referred to as the maximum-likelihood (ML) pipeline. First, we calculated genotyping discordance between GT-seq data derived from the AR pipeline and RADseq data derived

from the ML pipeline. To remove discrepancies in discordance calculations due to differences in genotyping calling methods, we also used the AR pipeline to genotype the RADseq data and then calculated discordance between GT-seq and RADseq data derived from the AR pipeline (https://github.com/mjgarrett/NIDGS_scripts).

Twenty-one NIDGS tissue samples were genotyped at an average of 7x coverage with WGS and also by the GT-seq panel. The same set of loci used in the RADseq comparison was extracted from the WGS data and used to calculate discordance for the genotypes produced from the tissue samples between WGS and GT-seq.

2.6 | Population genetic analyses

We compared the results from genetic diversity and population structure analyses produced by GT-seq and RADseq for the 53 individuals sequenced with both methods. We selected only neutral loci to measure and compare estimates of these demographic processes across methods. We created three datasets: all 2663 (filtered) neutral SNPs identified and used in the RADseq study (hereafter referred to as RAD_2663), the 207 neutral SNPs used and genotyped in the GT-seq panel and pipeline (hereafter referred to as GT_207), and the same 207 GT-seq loci extracted from the RADseq sequencing data (hereafter referred to as RAD_207).

We performed analyses for sites with ≥8 individuals for all three datasets. From R package HIERFSTAT (Goudet, 2005), H_O , H_E , and F_{IS} were estimated with function *basic.stats*, and pairwise F_{ST} was estimated with functions *genet.dist* and *boot.ppfst* with 999 bootstraps.

Population structure was evaluated using two approaches: Bayesian clustering analysis as implemented in program STRUCTURE (Pritchard et al., 2000) and maximum likelihood clustering analysis as implemented in program ADMIXTURE (Alexander et al., 2009). For STRUCTURE, we applied an admixture model with a burn-in period of 100,000 and a run length of 500,000 MCMC replicates to determine the number of genetic clusters (K) from 1 to 10 over 10 iterations each. The optimal K for each dataset was identified according to the Evanno method (Evanno et al., 2005) and the rate of change in the likelihood values (Pritchard et al., 2000). For ADMIXTURE, we used the *admixture-wrapper.py* python script (<https://github.com/dportik/admixture-wrapper>) to find the optimal K from 1 to 10 which was run over 50 repetitions and determined from its cross-validation procedure, which fold number was set to 10. Both STRUCTURE and ADMIXTURE outputs were visualized with the following R script (https://github.com/sakura81/PYRA_Genomics).

3 | RESULTS

3.1 | GT-seq panel development

From the 2663 neutral and 131 putatively adaptive NIDGS SNPs discovered in the Barbosa et al. (2021) RADseq study, primers were designed for 510 neutral and all 131 putatively adaptive SNPs (Table 1).

TABLE 1 NIDGS GT-seq panel optimization with the selection of SNPs from the Barbosa et al. (2021) RADseq study, primer development, optimization, and the composition of the final NIDGS GT-seq panel.

	Neutral	Adaptive	Total
RADseq NIDGS SNPs	2663	131	2794
Primer development	510	131	641
First optimization			
Candidate SNPs	510	131	641
SNPs removed	55	13	68
Second optimization			
Candidate SNPs	455	118	573
SNPs removed	231	37	268
Final GT-seq NIDGS panel	224	81	305

After the first optimization, 55 neutral and 13 adaptive SNPs were removed due to primer artifacts, off-target amplicons, and over-amplification of the target sequence which can indicate multiple genomic copies of the target amplicon. The second optimization began with a panel of 573 SNPs, consisting of 455 neutral and 118 adaptive SNPs. After the second optimization, 231 neutral and 37 adaptive SNPs produced primer artifacts, off-target amplification, and over amplification of the target amplicon. In the third optimization, the standard mix with equal concentrations for each primer had a higher genotyping success than the variable primer mix, 82.6% and 74.2% respectively. Therefore, the optimized GT-seq panel includes primers for 305 SNP loci (224 neutral and 81 adaptive SNPs).

3.2 | Sample quality and GT-seq genotyping success

Qubit concentration for hair samples and buccal swabs ranged from >0.05–1.62 to 0.78–36.7 ng/μL, respectively. For the hair extracts, gel electrophoresis revealed poor DNA quality, and gDNA fragment analysis reported DNA concentrations from 0.0132 to 0.873 ng/μL. Forty-seven out of 53 hair extracts amplified one microsatellite locus.

Sequencing yielded 112,211 mean reads per individual (range=1556–4,377,287) for hair samples, 201,904 reads per individual (range=10–18,205,105) for buccal swabs, and 3,392,926 reads per individual (range=120,622–12,277,944) for tissue samples. Hair samples yielded 1552 mean on-target reads per individual (range=3–107,011), buccal swabs yielded 116,184 mean on-target reads per individual (range=1–11,219,616), and tissue samples had 1,330,799 mean on-target reads per individual (range=10,104–4,992,869).

After we filtered out loci that had no data from 50% or more of our samples, 280 SNPs remained in the GT-seq dataset. The percentage of loci genotyped from hair samples was lower than buccal swabs and tissue samples (Figure 2). The average percentage of loci genotyped differed among the three sample types: 5.55 for

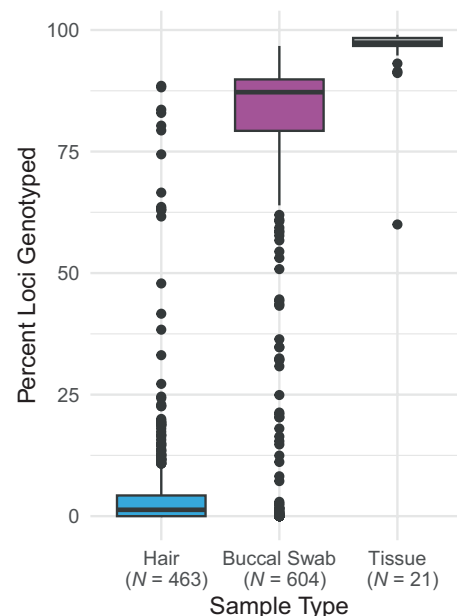


FIGURE 2 Boxplot of percent loci genotyped by sample type.

hair samples (min=0, max=88.52, median=1.31), 73.16 for swab samples (min=0, max=96.72, median=87.21), and 95.48 for tissue samples (min=60, max=99.02, median=97.38).

3.3 | Genotyping discordance

When we compared the GT-seq data derived from the AR pipeline to the RADseq data derived from the ML pipeline, we used 280 SNPs and 53 shared individuals. We found a discordance rate of 7% (Table 2). Of the discordant SNPs, 77% have a homozygous genotype from RADseq and a heterozygous genotype from GT-seq. For the GT-seq data and the RADseq data both genotyped with the AR pipeline, 274 SNPs were used to compare the same 53 individuals, as six SNPs were removed in the re-calling of the RADseq data with the AR pipeline. For the GT-seq and RADseq data derived from the AR pipeline, the discordance rate is 4%. Of the discordant SNPs, 66% have a homozygous genotype from RADseq and a heterozygous genotype from GT-seq. The same 280 SNPs from the first comparison were used to evaluate concordance between WGS and GT-seq tissue data, which has a discordance rate of 2%. Of the discordant SNPs, 40% have a homozygous genotype from WGS and a heterozygous genotype from GT-seq. More information about concordance and types of discordance is in Table S2.

3.4 | Population diversity estimates

Overall, population diversity metrics are comparable across sequencing methods (Figure 3). RAD_2663 has the lowest H_O and H_E estimates, and RAD_207 has the highest H_E estimates. For both RADseq datasets, $H_O < H_E$, while GT_207 suggested more heterozygotes

TABLE 2 Discordance between RADseq, WGS, and GT-seq for 53 NIDGS buccal swab samples and 21 tissue samples.

Sequencing methods	Sample type	RADseq pipeline	SNPs	Mean number of loci per individual					
				Both methods	Discordant loci	RAD/WGS hom, GT het	RAD/WGS het, GT hom	RAD/WGS hom, GT hom	Discordance rate
RADseq vs. GT-seq	Swab	ML	280	198.04	12.46	9.85 (77%)	1.87 (17%)	0.75 (6%)	7%
RADseq vs. GT-seq	Swab	AR	274	97.37	3.19	2.10 (66%)	0.9 (22%)	0.19 (7%)	4%
WGS vs. GT-seq	Tissue	—	280	213.55	4.7	1.85 (40%)	2.75 (53%)	0.10 (2%)	2%

Note: Discordance is a mismatch in diploid genotypes between datasets for the same individual. RADseq pipeline: the maximum likelihood (ML) or allele ratio (AR) genotyping pipeline was used to genotype RADseq data; Both methods: when RADseq and GT-seq or WGS and GT-seq call a locus; RAD/WGS hom, GT het: of the discordant loci, RADseq or WGS called a homozygote while GT-seq called a heterozygote; RAD/WGS het, GT hom: of the discordant loci, RADseq or WGS called a heterozygote while GT-seq called a homozygote; RAD/WGS hom, GT hom: of the discordant loci, RADseq or WGS and GT-seq called a homozygote. Discordance rate calculated by dividing mean discordant loci per individual over loci genotyped by both methods. Percentages in previous categories are of discordant SNPs, while Discordance rate is of all SNPs.

than what would be expected under Hardy-Weinberg conditions. Thus, F_{IS} values for GT_207 are negative, and F_{IS} values are lower in RAD_2663 than RAD_207. Pairwise F_{ST} values are lowest in RAD_2263 and highest in GT_207 (Table 3).

Patterns of proportion of assignment to a given K between programs STRUCTURE (Figure 4) and ADMIXTURE (Figures S5–S7) were consistent across K values and sequencing methods. However, among the STRUCTURE results, the best supported K s are different between sequencing methods, although there was not strong statistical support for a single “best” K for any of the datasets. For the RAD_2663 dataset, $K=6$ is best supported (Figure S1), while $K=2$ (with subpeak at $K=5$) and $K=4$ are best supported in RAD_207 (Figure S2) and GT_207 (Figure S3), respectively. Nevertheless, the genetic grouping patterns at each K among all three datasets are broadly consistent.

4 | DISCUSSION

Here, we developed a GT-seq panel of 305 loci for genetic monitoring of NIDGS than can be used on DNA extracted from minimally invasive buccal swab samples. We validated the GT-seq panel by comparing genotypes and downstream genetic diversity and structure analyses across sequencing methods, and we evaluated factors that produced discordance between methods.

4.1 | GT-seq panel optimization

In spite of efforts to limit the potential for off-target sequence capture and primer artifacts when designing primers for GT-seq, 106 of the attempted 411 loci were excluded during testing of the panel. Typical conversion rates from designed primers to a functioning assay in GT-seq panels average about 73% (Burgess et al., 2022; Chang et al., 2022; Hayward et al., 2022; McKinney et al., 2022; Schmidt et al., 2020) and for our study the rate was 74%. Of the filtered primer sets, 24 had no sequences reported for the target amplicon, 37 produced an abundance of off-target sequences, three

produced an overabundance of the target sequence, and the remaining 58 were omitted to prevent primer artifacts from forming in PCR or due to poor results in the plots produced by the *GTseq_SeqTest.pl* and *GTseq_Primer-Interaction-Test.pl* scripts. In this case, more loci than average were omitted for producing an abundance of off-target sequences with properly paired primers. This means that some primer sets were targeting multiple genomic locations and were not efficiently capturing data from the target loci. This may be due to target locus sequences being repeated or present in transposable elements in the NIDGS genome. In Barbosa et al. (2021), RADseq data were aligned to the thirteen-lined ground squirrel (*Ictidomys tridecemlineatus*) genome, as no NIDGS genome is currently available. This could be a possible reason many primer sets amplify in multiple regions, and this represents a general challenge for working with rare or threatened species where a reference genome is only available for a closely related species.

4.2 | Performance of hair samples

The NIDGS hair samples performed poorly on GT-seq. Despite producing similar raw read quantities to buccal swabs, hair samples yielded almost no on-target reads. The hair samples were originally used in two studies that amplified eight nuclear DNA microsatellites, which ranged in 106–246bp in size, to evaluate diversity and divergence in the NIDGS (Garner et al., 2005; Hoisington-Lopez et al., 2012). When evaluated for their use in RADseq, it was determined the hair samples did not have high enough DNA quantities required by the technique. Despite the hair samples working in previous studies with microsatellite amplification, little to no DNA remained in the hair samples that we used for this study. These extracted DNA samples had been stored in a -80°C freezer after DNA extraction but had been through multiple freeze-thaw cycles which could have degraded the DNA. However, GT-seq and other SNP panels have been successfully developed to be used with DNA extracted from hair samples in other species including Sitka black-tailed deer, coyotes (*Canis latrans*), and tigers (*Panthera tigris*) (Burgess et al., 2022; Eriksson et al., 2020; Natesh et al., 2019).

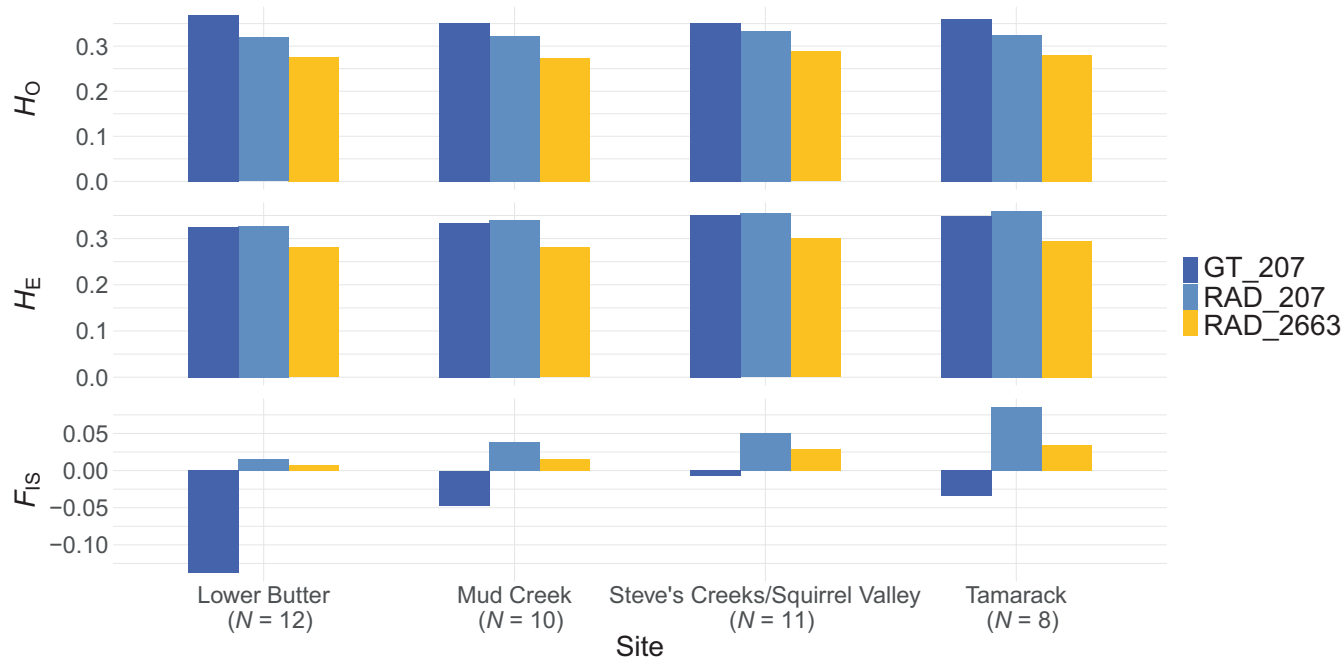


FIGURE 3 Population genetic diversity estimates of NIDGS by sequencing method for sites with more than eight individuals. RAD_2336 metrics estimated with 2336 neutral SNPs, RAD_207 and GT_207 with 207 neutral SNPs. F_{IS} , inbreeding coefficient; H_E , expected heterozygosity; H_O , observed heterozygosity.

4.3 | Factors contributing to discordance

Two main factors likely explain genotype discordance: sequencing technique and genotyping method. This is the first study to design a GT-seq panel from SNPs identified from RADseq performed on minimally invasive samples, rather than high-quality samples like tissue or blood. RADseq with low coverage (e.g., <5× read depth) has a high probability of only sampling one chromosome of a diploid individual at a locus, which may cause heterozygotes to be called homozygotes in downstream analyses (Andrews et al., 2016; Nielsen et al., 2011). In comparing the ML pipeline derived and AR pipeline derived RADseq data to the AR pipeline derived GT-seq data, 79% and 65% of discordant SNPs are from RADseq calling a homozygote and GT-seq calling a heterozygote. If these homozygote-heterozygote mismatch SNPs are removed, our discordance rate falls to 1.4% and less than 1%, in line with or below other studies (Hayward et al., 2022; Schmidt et al., 2020; Setzke et al., 2021). Moreover, in all genotyping pipeline comparisons, there are no heterozygote-heterozygote discordant SNPs, which supports the view that discordance is more often due to allelic drop-out, particularly in lower-coverage RADseq, than to detection of a false allele. In our data, GT-seq provided higher coverage than RADseq overall, with GT-seq having a mean coverage of 540.78× per locus, per individual and RADseq having a mean coverage of 9.65× per locus, per individual for the same 280 loci and 53 individuals. This is typical of how these two approaches are generally used because of the large difference in numbers of loci targeted and the trade-off with total sequencing effort. A similar bias toward RADseq homozygous calls

has also been found in other RADseq and SNP panel comparisons (Hess et al., 2015; Schmidt et al., 2020).

We found that genotype discordance is higher when calls are made from different genotyping pipelines. There is variability in the genotyping pipelines used in previous studies that compared RADseq and GT-seq data. Schmidt et al. (2020) and Setzke et al. (2021) both used RADseq data called with the ML pipeline and GT-seq data called with the AR pipeline. Bootsma et al. (2020) used GTscore (<https://github.com/gjmckinney/GTscore>) for GT-seq genotyping and only retained GT-seq genotypes if they had 60× coverage in their RADseq data, which was genotyped with the ML pipeline. Finally, Hayward et al. (2022) called their GT-seq data multiple ways, first with the AR pipeline and then with the same ML approach as their double digest RADseq (ddRADseq) data. However, these studies have a RADseq and GT-seq genotyping discordance rate between 1.1% and 3.4%. Our discordance rates for the tissue samples and the AR pipeline derived RADseq and GT-seq data fall within this range. Thus, discordance is partly due to genotyping method, which is expected given how the choice and parameterization of genotyping methods is known to affect population genomic data (Mastretta-Yanes et al., 2015; Paris et al., 2017; Rivera-Colón & Catchen, 2022).

4.4 | Genetic diversity and population structure

Our GT-seq dataset produced similar patterns of genetic diversity and population structure to the RADseq dataset with one exception: population-level F_{IS} values were consistently negative for the

TABLE 3 Pairwise F_{ST} (Weir & Cockerham, 1984) for NIDGS of sites with more than eight individuals across three datasets (RAD_2663, RAD_207, GT_207).

Dataset	LB vs. MC	LB vs. SS	LB vs. TA	MC vs. SS	MC vs. TA	SS vs. TA
RAD_2663	0.089 (0.081, 0.097)	0.080 (0.073, 0.088)	0.073 (0.066, 0.081)	0.094 (0.086, 0.101)	0.067 (0.060, 0.074)	0.0811 (0.074, 0.089)
RAD_207	0.093 (0.070, 0.120)	0.086 (0.062, 0.116)	0.074 (0.049, 0.099)	0.114 (0.089, 0.140)	0.071 (0.049, 0.140)	0.097 (0.069, 0.123)
GT_207	0.110 (0.080, 0.143)	0.094 (0.068, 0.123)	0.095 (0.070, 0.123)	0.129 (0.103, 0.157)	0.089 (0.064, 0.118)	0.110 (0.082, 0.142)

Note: Estimates were calculated with 999 bootstraps with confidence intervals for 0.025 and 0.975 given in parentheses.

GT-seq dataset. This finding agrees with other GT-seq validation studies. Bootsma et al. (2020) and Schmidt et al. (2020) also report higher H_O than H_E values for GT-seq data, as well as negative F_{IS} values. Positive F_{IS} values are often expected, resulting from both biological and methodological factors. Hidden population structure or excess relatedness in the set of sampled individuals can elevate observed heterozygosity relative to expectations based on treating the samples as representing a single population. Allelic drop-out as a result of low or uneven coverage can also lead to positive F_{IS} values, when true heterozygotes are genotyped as homozygotes but there is no bias in which allele is observed. In contrast, negative F_{IS} values are harder to explain. One possible explanation is amplification of duplicate sequences or paralogs, causing sequence divergence at different parts of the genome to resemble heterozygosity. Our GT-seq panel design and optimization aimed to remove any primer pairs that bind in multiple genomic locations, but it is possible that some remain. A second possibility is mis-genotyping true homozygotes as heterozygotes at single loci. As described above, we did not observe any heterozygote–heterozygote discordance, suggesting that detection of a false allele that is not present in the population is unlikely. However, detection of a false allele that is present in other individuals in the population could result from factors like index hopping during library preparation, and this could elevate the frequency of observed heterozygotes. To test the possible effect of genotyping method, we used another genotyping pipeline GTscore (<https://github.com/gjmckinney/GTscore>) to genotype the 53 individuals sequenced with GT-seq. GTscore called all the same genotypes as the original AR pipeline and thus resulted in the same higher H_O than H_E values for GT-seq data and negative F_{IS} values. For details and results on GTscore, see Supporting Information. The negative F_{IS} values observed both here and in other GT-seq studies warrant further investigation.

Despite any concerns about heterozygosity and individual genotypes, estimates of genetic population structure in NIDGS are broadly consistent when estimated with different sequencing methods and quantities of loci. The population structuring across K values between ADMIXTURE and STRUCTURE was also consistent (Figure 4 and Figures S5–S7). Nonetheless, population assignment appears cleaner (i.e., fewer individuals with admixed ancestry) in the GT_207 dataset compared to RAD_207. This is likely because although these datasets include the same 207 SNPs, the RAD_207 dataset has higher levels of missing data than GT_207 because of overall sequencing coverage (Table S2), resulting in greater uncertainty in genetic cluster assignment of some individuals in RAD_207. The clustering from the GT_207 dataset appears more similar to the RAD_2663 results (Figure 4), illustrating that the larger number of loci in a typical RADseq study can compensate for higher levels of missing data. The loci in the GT-seq panel were chosen in part to allow delineation of neutral population structure, and our results show that they do successfully capture this signal compared to the full RADseq data. As a result, the GT-seq panel promises to be effective for population clustering and ancestry assignment in future work.

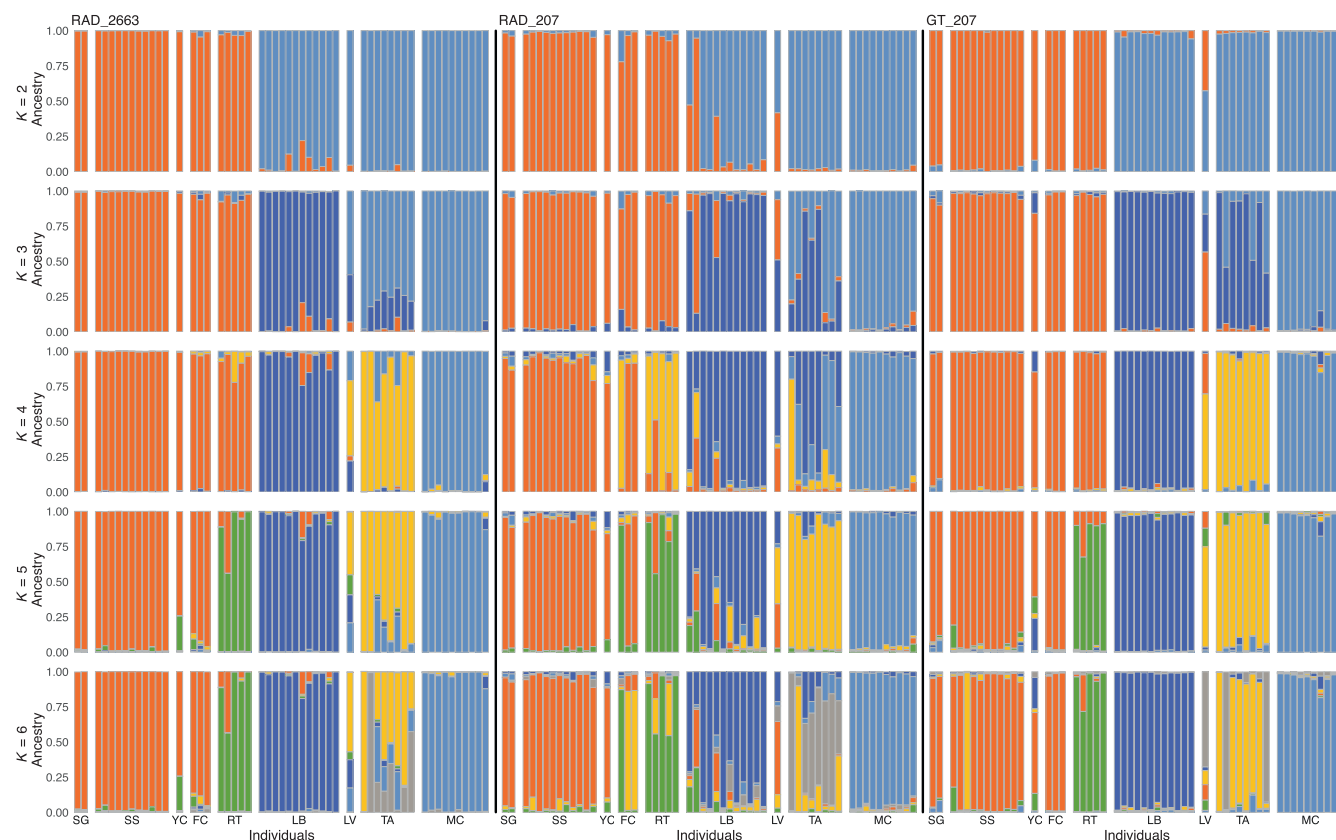


FIGURE 4 STRUCTURE bar plots displaying inferred clustering and individual ancestry estimates of NIDGS ($n=53$) for all three neutral SNP datasets (RAD_2663, RAD_207, and GT_207) for K_s 2–6. Each color represents a distinct genetic cluster, each vertical bar represents the proportion of ancestry of a single individual to the different genetic cluster. Individuals are grouped into populations which are ordered geographically from west to east: FC, Fawn Creek; LB, Lower Butter; LV, Lost Valley; MC, Mud Creek; RT, Rockytop; SG, Summit Gulch; SS, Steve's Creek/Squirrel Valley; TA, Tamarack; YC, YCC.

4.5 | Future applications

Our GT-seq panel will be applied as a monitoring tool in future and ongoing sampling to address recovery objectives set by USFWS (2003), such as defining metapopulations and conservation units based on neutral and adaptive loci (Funk et al., 2012; Turbek et al., 2023) and estimating N_e . The inclusion of both neutral and putatively adaptive loci is important for genetic monitoring practices and informing conservation action. Because of the high quality of DNA required by RADseq, Barbosa et al. (2021) provided a snapshot in time rather than a study of temporal change in genetic parameters for NIDGS. Our GT-seq panel will be used to gauge the genetic health of NIDGS by monitoring changes in allele frequencies over time and in response to any future management actions, disturbances, or stochastic events (Flanagan et al., 2018; Schwartz et al., 2007; Shafer et al., 2015). As levels of gene flow were found to be variable among NIDGS sites (Barbosa et al., 2021), it will be important to assess how demographic processes change through time. Moreover, population persistence may be dependent on locally adapted genotypes for many of the isolated populations (Rundell & Price, 2009). Additionally, due to different evolutionary histories at each site, the differences in standing genetic variation and adaptive potential are important considerations for breeding or reintroduction programs

(Flanagan et al., 2018; Hansen et al., 2012). This panel can help illuminate these factors as well as carry the genetic monitoring of NIDGS into the future as climate change and anthropogenic pressures may increase risks for this threatened species.

AUTHOR CONTRIBUTIONS

Molly J. Garrett: Conceptualization (equal); data curation (lead); formal analysis (lead); investigation (lead); methodology (equal); project administration (equal); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (equal). **Stacey A. Nerkowski:** Investigation (supporting); methodology (supporting); project administration (supporting); supervision (supporting); writing – review and editing (equal). **Shannon Kieran:** Formal analysis (supporting); investigation (supporting); writing – review and editing (equal). **Nathan R. Campbell:** Methodology (equal); software (lead); writing – review and editing (equal). **Soraia Barbosa:** Formal analysis (supporting); investigation (supporting); writing – review and editing (equal). **Courtney J. Conway:** Conceptualization (equal); funding acquisition (equal); project administration (supporting); supervision (equal); writing – review and editing (equal). **Paul A. Hohenlohe:** Conceptualization (equal); funding acquisition (equal); investigation (supporting); methodology (supporting); project administration (supporting); supervision (equal); writing – review and editing

(equal). **Lisette P. Waits:** Conceptualization (equal); funding acquisition (equal); investigation (supporting); methodology (supporting); project administration (supporting); supervision (equal); writing – review and editing (equal).

ACKNOWLEDGEMENTS

This publication was made possible by the NSF Idaho EPSCoR Program and by the National Science Foundation under award number OIA-1757324. U.S. Forest Service (Payette National Forest), U.S. Fish and Wildlife Service, and Idaho Department of Fish and Game for funding that allowed for capture and sampling of northern Idaho ground squirrels. Austin Allison, Alice Morris, and many seasonal field technicians for their help obtaining samples. The OX Ranch for providing field lodging for the field crew collecting samples. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This study was performed under the auspices of University of Idaho IACUC protocols #2006-35, 2015-53, 2019-28, and 2022-10.

OPEN RESEARCH BADGES



This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://github.com/mjgarrett/NIDGS_scripts.

DATA AVAILABILITY STATEMENT

SNP genotyping pipelines, concordance scripts, and the GT-seq SNP dataset in .csv format are available on GITHUB (https://github.com/mjgarrett/NIDGS_scripts). VCF files from the RADseq data are available on DRYAD (<https://doi.org/10.5061/dryad.sj3tx965c>).

ORCID

Molly J. Garrett <https://orcid.org/0000-0002-5757-7762>

Shannon Kieran <https://orcid.org/0000-0002-2855-6487>

Soraia Barbosa <https://orcid.org/0000-0001-7806-4256>

Courtney J. Conway <https://orcid.org/0000-0003-0492-2953>

Paul A. Hohenlohe <https://orcid.org/0000-0002-7616-0161>

Lisette P. Waits <https://orcid.org/0000-0002-1323-0812>

REFERENCES

- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Andrews, K., Good, J. M., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2), 81–92. <https://doi.org/10.1038/nrg.2015.28>
- Andrews, K. R., De Barba, M., Russello, M. A., & Waits, L. P. (2018). Advances in using non-invasive, archival, and environmental samples for population genomic studies. In P. A. Hohenlohe & O. P. Rajora (Eds.), *Population genomics: Wildlife. Population genomics* (pp. 63–99). Springer. https://doi.org/10.1007/13836_2018_45
- Barbosa, S., Andrews, K. R., Goldberg, A. R., Gour, D. S., Hohenlohe, P. A., Conway, C. J., & Waits, L. P. (2021). The role of neutral and adaptive genomic variation in population diversification and speciation in two ground squirrel species of conservation concern. *Molecular Ecology*, 30, 1–31.
- Beja-Pereira, A., Oliveira, R., Alves, P. C., Schwartz, M. K., & Luikart, G. (2009). Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources*, 9(5), 1279–1301. <https://doi.org/10.1111/j.1755-0998.2009.02699.x>
- Bootsma, M. L., Gruenthal, K. M., McKinney, G. J., Simmons, L., Miller, L., Sass, G. G., & Larson, W. A. (2020). A GT-seq panel for walleye (*Sander vitreus*) provides important insights for efficient development and implementation of amplicon panels in non-model organisms. *Molecular Ecology Resources*, 20(6), 1706–1722. <https://doi.org/10.1111/1755-0998.13226>
- Burgess, B. T., Irvine, R. L., & Russello, M. A. (2022). A genotyping-in-thousands by sequencing panel to inform invasive deer management using noninvasive fecal and hair samples. *Ecology and Evolution*, 12(6), e8993. <https://doi.org/10.1002/ece3.8993>
- Campbell, N. R., Harmon, S. A., & Narum, S. R. (2015). Genotyping-in-thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources*, 15(4), 855–867. <https://doi.org/10.1111/1755-0998.12357>
- Carroll, E. L., Bruford, M. W., DeWoody, J. A., Leroy, G., Strand, A., Waits, L., & Wang, J. (2018). Genetic and genomic monitoring with minimally invasive sampling methods. *Evolutionary Applications*, 11(7), 1094–1119. <https://doi.org/10.1111/eva.12600>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Chang, S. L., Ward, H. G. M., Elliott, L. D., & Russello, M. A. (2022). Genotyping-in-Thousands by sequencing panel development and application for high-resolution monitoring of introgressive hybridization within sockeye salmon. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-07309-x>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Eriksson, C. E., Ruprecht, J., & Levi, T. (2020). More affordable and effective noninvasive single nucleotide polymorphism genotyping using high-throughput amplicon sequencing. *Molecular Ecology Resources*, 20(6), 1505–1516. <https://doi.org/10.1111/1755-0998.13208>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Flanagan, S. P., Forester, B. R., Latch, E. K., Aitken, S. N., & Hoban, S. (2018). Guidelines for planning genomic assessment and monitoring of locally adaptive variation to inform species conservation. *Evolutionary Applications*, 11(7), 1035–1052. <https://doi.org/10.1111/eva.12569>
- Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, 27(9), 489–496. <https://doi.org/10.1016/j.tree.2012.05.012>
- Garner, A., Rachlow, J. L., & Waits, L. P. (2005). Genetic diversity and population divergence in fragmented habitats: Conservation of Idaho ground squirrels. *Conservation Genetics*, 6(5), 759–774. <https://doi.org/10.1007/s10592-005-9035-3>
- Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5(1), 184–186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>

- Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., Manzon, R. G., Martino, J. A., Pierson, T., Rogers, S. M., Wilson, J. Y., & Somers, C. M. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, 15(6), 1304–1315. <https://doi.org/10.1111/1755-0998.12404>
- Hansen, M. M., Olivieri, I., Waller, D. M., Nielsen, E. E., & The GeM Working Group. (2012). Monitoring adaptive genetic responses to environmental change. *Molecular Ecology*, 21(6), 1311–1329. <https://doi.org/10.1111/j.1365-294X.2011.05463.x>
- Hayward, K. M., Clemente-Carvalho, R. B. G., Jensen, E. L., de Groot, P. V. C., Branigan, M., Dyck, M., Tschritter, C., Sun, Z., & Loughheed, S. C. (2022). Genotyping-in-thousands by sequencing (GT-seq) of noninvasive faecal and degraded samples: A new panel to enable ongoing monitoring of Canadian polar bear populations. *Molecular Ecology Resources*, 22(5), 1906–1918. <https://doi.org/10.1111/1755-0998.13583>
- Helmstetter, N. A., Conway, C. J., Stevens, B. S., & Goldberg, A. R. (2021). Balancing transferability and complexity of species distribution models for rare species conservation. *Diversity and Distributions*, 27(1), 95–108. <https://doi.org/10.1111/ddi.13174>
- Hess, J. E., Campbell, N. R., Docker, M. F., Baker, C., Jackson, A., Lampman, R., McIlraith, B., Moser, M. L., Statler, D. P., Young, W. P., Wildbill, A. J., & Narum, S. R. (2015). Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. *Molecular Ecology Resources*, 15(1), 187–202. <https://doi.org/10.1111/1755-0998.12283>
- Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, 30(1), 62–82. <https://doi.org/10.1111/mec.15720>
- Hoisington-Lopez, J. L., Waits, L. P., & Sullivan, J. (2012). Species limits and integrated taxonomy of the Idaho ground squirrel (*Urocitellus brunneus*): Genetic and ecological differentiation. *Journal of Mammalogy*, 93(2), 589–604. <https://doi.org/10.1644/11-mamm-a-021.1>
- Korneliusson, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, 15(1), 356. <https://doi.org/10.1186/s12859-014-0356-4>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5), 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), 28–41. <https://doi.org/10.1111/1755-0998.12291>
- McKinney, G. J., Barry, P. D., Pascal, C., Seeb, J. E., Seeb, L. W., & McPhee, M. V. (2022). A new genotyping-in-thousands-by-sequencing single nucleotide polymorphism panel for mixed-stock analysis of chum salmon from coastal Western Alaska. *North American Journal of Fisheries Management*, 42(5), 1134–1143. Portico. <https://doi.org/10.1002/nafm.10805>
- Natesh, M., Taylor, R. W., Truelove, N. K., Hadly, E. A., Palumbi, S. R., Petrov, D. A., & Ramakrishnan, U. (2019). Empowering conservation practice with efficient and economical genotyping from poor quality samples. *Methods in Ecology and Evolution*, 10(6), 853–859. <https://doi.org/10.1111/2041-210X.13173>
- Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, 12(6), 443–451. <https://doi.org/10.1038/nrg2986>
- Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for STACKS. *Methods in Ecology and Evolution*, 8(10), 1360–1373. <https://doi.org/10.1111/2041-210X.12775>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Rivera-Colón, A. G., & Catchen, J. (2022). Population genomics analysis with RAD, reprised: Stacks 2. In C. Verde & D. Giordano (Eds.), *Marine genomics. Methods in molecular biology* (Vol. 2498) (pp. 99–149). Humana. https://doi.org/10.1007/978-1-0716-2313-8_7
- Rundell, R. J., & Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, 24(7), 394–399. <https://doi.org/10.1016/j.tree.2009.02.007>
- Russello, M. A., Waterhouse, M. D., Etter, P. D., & Johnson, E. A. (2015). From promise to practice: Pairing non-invasive sampling with genomics in conservation. *PeerJ*, 3, e1106. <https://doi.org/10.7717/peerj.1106>
- Schmidt, D. A., Campbell, N. R., Govindarajulu, P., Larsen, K. W., & Russello, M. A. (2020). Genotyping-in-thousands by sequencing (GT-seq) panel development and application to minimally invasive DNA samples to support studies in molecular ecology. *Molecular Ecology Resources*, 20(1), 114–124. <https://doi.org/10.1111/1755-0998.13090>
- Schwartz, M. K., Luikart, G., & Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, 22(1), 25–33. <https://doi.org/10.1016/j.tree.2006.08.009>
- Setzke, C., Wong, C., & Russello, M. A. (2021). Genotyping-in-thousands by sequencing of archival fish scales reveals maintenance of genetic variation following a severe demographic contraction in kokanee salmon. *Scientific Reports*, 11(1), 22798. <https://doi.org/10.1038/s41598-021-01958-0>
- Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fawcett, K. D., Fior, S., Hajibabaei, M., Hill, J. A., Hoezel, A. R., Höglund, J., Jensen, E. L., Krause, J., Kristensen, T. N., ... Zieñiński, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution*, 30(2), 78–87. <https://doi.org/10.1016/j.tree.2014.11.009>
- Sherman, P. W., & Runge, M. C. (2002). Demography of a population collapse: The northern Idaho ground squirrel (*Spermophilus brunneus brunneus*). *Ecology*, 83(10), 2816–2831.
- Suronen, E. F., & Newingham, B. A. (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. <https://doi.org/10.1016/J.FORECO.2013.04.026>
- Taberlet, P., Waits, L. P., & Luikart, G. (1999). Noninvasive genetic sampling: Look before you leap. *Trends in Ecology & Evolution*, 14(8), 323–327. [https://doi.org/10.1016/S0169-5347\(99\)01637-7](https://doi.org/10.1016/S0169-5347(99)01637-7)
- Turbek, S. P., Funk, W. C., & Ruegg, K. C. (2023). Where to draw the line? Expanding the delineation of conservation units to highly mobile taxa. *Journal of Heredity*, 114(4), 300–311. <https://doi.org/10.1093/jhered/esad011>
- USFWS. (2003). Recovery plan for the Northern Idaho Ground Squirrel (*Spermophilus brunneus brunneus*). (p. 68).
- Wagner, B., & Evans Mack, D. (2020). Long-term population monitoring of northern Idaho ground squirrel: 2020 implementation and population estimates.
- Waits, L. P., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: A review of applications and recommendations for accurate data collection. *The Journal of Wildlife Management*, 69(4), 1419–1433. [https://doi.org/10.2193/0022-541X\(2005\)69\[1419:NGSTFW\]2.0.CO;2](https://doi.org/10.2193/0022-541X(2005)69[1419:NGSTFW]2.0.CO;2)
- Walters, A. D., & Schwartz, M. K. (2020). Population genomics for the management of wild vertebrate populations. In P. A. Hohenlohe & O. P. Rajora (Eds.), *Populations genomics: Wildlife. Population genomics* (pp. 419–436). Springer. https://doi.org/10.1007/13836_2020_72
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370. <https://doi.org/10.2307/2408641>
- Yensen, E., & Dyni, E. J. (2020). Why is the northern Idaho ground squirrel rare? *Northwest Science*, 94(1), 1–23. <https://doi.org/10.3955/046.094.0101>

Zero, V. H., Barocas, A., Jochimsen, D. M., Pelletier, A., Giroux-Bougard, X., Trumbo, D. R., Castillo, J. A., Evans Mack, D., Linnell, M. A., Pigg, R. M., Hoisington-Lopez, J., Spear, S. F., Murphy, M. A., & Waits, L. P. (2017). Complementary network-based approaches for exploring genetic structure and functional connectivity in two vulnerable, endemic ground squirrels. *Frontiers in Genetics*, 8. <https://doi.org/10.3389/fgene.2017.00081>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Garrett, M. J., Nerkowski, S. A., Kieran, S., Campbell, N. R., Barbosa, S., Conway, C. J., Hohenlohe, P. A., & Waits, L. P. (2024). Development and validation of a GT-seq panel for genetic monitoring in a threatened species using minimally invasive sampling. *Ecology and Evolution*, 14, e11321. <https://doi.org/10.1002/ece3.11321>