From promise to practice: pairing non-invasive sampling with genomics in conservation

Michael Russello, Matthew D Waterhouse, Paul D Etter, Eric A Johnson

Conservation genomics has become an increasingly popular term, yet it remains unclear whether the non-invasive sampling that is essential for many conservation-related studies is compatible with the minimum requirements for harnessing next-generation sequencing technologies. Here, we evaluated the feasibility of using genotyping-by-sequencing of non-invasively collected hair samples to simultaneously identify and genotype single nucleotide polymorphisms (SNPs) in a climate change-sensitive mammal, the American pika (*Ochotona princeps*). We identified and genotyped 3,803 high-confidence SNPs across eight sites distributed along two elevational transects using starting DNA amounts as low as 1 nanogram. Fifty-five outlier loci were detected as candidate gene regions under divergent selection, constituting potential targets for future validation. Genome-wide estimates of gene diversity significantly and positively correlated with elevation across both transects, with all low elevation sites exhibiting significant heterozygote deficit likely due to inbreeding. More broadly, our results highlight a range of issues that must be considered when pairing genomic data collection with non-invasive sampling, particularly related to field sampling protocols for minimizing exogenous DNA, data collection strategies and quality control steps for enhancing target organism yield, and analytical approaches for maximizing cost-effectiveness and information content of recovered genomic data.

Title: From promise to practice: pairing non-invasive sampling with genomics in conservation

Michael A. Russello[1*], Matthew D. Waterhouse[1], Paul D. Etter[2] and Eric A. Johnson[2]

[1] Department of Biology, University of British Columbia, Okanagan Campus, 3427 University Way, Kelowna, BC V1V 1V7, Canada

[2] SNPsaurus, Eugene, OR, USA

* Correspondence to: Michael Russello (michael.russello@ubc.ca); 250-807-8762

Running head: Pairing non-invasive sampling with genomics

Abstract

Conservation genomics has become an increasingly popular term, yet it remains unclear whether the non-invasive sampling that is essential for many conservation-related studies is compatible with the minimum requirements for harnessing next-generation sequencing technologies. Here, we evaluated the feasibility of using genotyping-by-sequencing of non-invasively collected hair samples to simultaneously identify and genotype single nucleotide polymorphisms (SNPs) in a climate change-sensitive mammal, the American pika (*Ochotona princeps*). We identified and genotyped 3,803 high-confidence SNPs across eight sites distributed along two elevational transects using starting DNA amounts as low as 1 nanogram. Fifty-five outlier loci were detected as candidate gene regions under divergent selection, constituting potential targets for future validation. Genome-wide estimates of gene diversity significantly and positively correlated with elevation across both transects, with all low elevation sites exhibiting significant heterozygote deficit likely due to inbreeding. More broadly, our results highlight a range of issues that must be considered when pairing genomic data collection with non-invasive sampling, particularly related to field sampling protocols for minimizing exogenous DNA, data collection strategies and quality control steps for enhancing target organism yield, and analytical approaches for maximizing cost-effectiveness and information content of recovered genomic data.

Introduction

There has been much discussion on the transition of conservation genetics to conservation genomics (Helyar et al. 2011; McMahon et al. 2014). Genomic analysis provides the advantage of assessing natural selection and adaptive genetic variation (Schoville et al. 2012), accurately estimating levels of genome wide diversity (Vali et al. 2008), and providing novel information for delineating conservation units (Funk et al. 2012) and informing management strategies (Hoffmann et al. 2015). Yet, it remains unclear the degree to which the conservation community as a whole has embraced genomics as a useful tool, suggesting significant gaps in methodology and analysis that must be overcome before the technology is ready for real world applications (Shafer et al. 2015). One methodological aspect that has yet to be formally considered is the sample source of DNA; many population genetic studies of elusive or endangered species must rely on non-invasively collected samples. There is an expansive literature on the use of DNA from hair, feces, feathers and other non-invasively sampled materials for investigating the ecology, behavior, and population history of wildlife species (reviewed in Beja-Pereira et al. 2009; Waits & Paetkau 2005). These studies have largely been based on single locus mitochondrial DNA (mtDNA) sequencing or multi-locus nuclear genotyping of hyper-variable loci, such as microsatellites (Taberlet et al. 1999; Waits & Paetkau 2005). Modern genotyping-by-sequencing approaches, such as those that rely on restriction-site associated DNA (RAD) tags, typically call for 1 microgram of high quality DNA for library construction (Baird et al. 2008; Etter et al. 2011). However, even highly refined DNA extraction protocols from non-invasively collected starting materials typically yield low concentrations of DNA, which may also contain PCR inhibitors (Beja-Pereira et al. 2009; Roon et al. 2003; Smith & Wang 2014; Taberlet et al. 1999; Waits & Paetkau 2005). To date, it remains unclear whether the non-

invasive sampling that is essential for many conservation-related studies is compatible with the minimum requirements for harnessing the next generation sequencing (NGS) technologies necessary for expanding conservation genetics in the genomics era.

The American pika, *Ochotona princeps*, is an example of an elusive species that has benefited from the pairing of non-invasive sampling with genetic data collection. A small lagomorph, the American pika is discontinuously distributed in mountainous areas throughout western North America from central British Columbia and Alberta, Canada, south to the Sierra Nevada in California and east to New Mexico, USA. Pikas are restricted to talus slopes in proximity to meadows that provide their food (Smith & Weston 1990). Exhibiting highly nonrandom distributions across mountaintop habitats, the average elevation of Great Basin *O. princeps* populations is currently ~783 m higher than during the late Wisconsinian (Grayson 2005). In general, lower elevational limits are constrained by an inability to tolerate high temperatures, while high altitude distribution is possible through adaptation to hypoxic environments (Beever & Smith 2011). The fragmented nature of their habitats has propelled *O. princeps* as a focal mammalian species for studies of metapopulation dynamics, island biogeography, source-sink dynamics (Beever et al. 2013; Peacock & Smith 1997a), and extinction risk in the face of climate change (Beever et al. 2010; Hafner 1993; Smith 1974; Stewart et al. 2015).

Recent genetic studies of American pika have relied on samples obtained non-invasively using hair snares, which have greatly enhanced sample sizes while minimizing sampling effort (Henry & Russello 2011). These studies revealed restricted dispersal capacity (Henry et al. 2012) and preliminary evidence for adaptive population divergence of American pika along elevation gradients at their northern range margin (Henry & Russello 2013). These latter

findings were based on amplified fragment length polymorphism (AFLP)-based genomic scans. In addition to other undesirable properties, AFLPs are anonymous, dominant markers, which precluded the identification of genes responsible for the observed adaptive divergence. Single nucleotide polymorphisms (SNPs), with their broad genomic coverage and better-understood mutation models, would overcome many of these limitations if they can be effectively genotyped within the constraints imposed by this system and others involving elusive and endangered species.

In the current study, we used nextRAD (Nextera-tagmented, reductively-amplified DNA) genotyping to collect SNP data from American pikas sampled along parallel elevational gradients to: 1) evaluate the feasibility of using DNA from non-invasively collected hair samples to simultaneously identify and genotype SNPs in an elusive species; and 2) provide preliminary insights into patterns of neutral and adaptive population divergence within this system.

Materials and Methods

Sample collection

This study was conducted in the North Cascades National Park, Washington, USA (Figure 1). Sites within this national park present the opportunity to sample American pika along steep elevational transects where climates change rapidly over short linear distances, while controlling for other environmental and historical factors. Additionally, while pika are currently abundant in the park, this area has been disproportionally affected by climate change (Karl et al. 2009). Pika populations were sampled along two elevational transects [Pyramid Peak (PP) and Thornton Lakes (TL)] between July and August 2013. Sites within transects ranged from 450 m to 1700 m,

113 representing an approximate 6°C gradient in mean annual temperature (Briggs et al. 1997) over
114 less than 6.5 km distance (Figure 1).

115 Non-invasive snares were used to obtain hair samples from 12 individuals at four sites
116 along each of the two transects (n = 96) following Henry and Russello (2011). To minimize
117 resampling the same animal, snares were set a minimum of 15 m apart and only one sample from
118 each snare was used. Subsequent genetic data were used to test the assumption that each sample
119 possessed a unique genotype (see below). All samples were collected under United States
120 Department of Interior National Park Service permit # NOCA-2014-SCI-0022 and in accordance
121 with animal care protocol (A11-0371) as approved by the University of British Columbia's
122 Animal Care & Biosafety Committee.

123

124 *DNA isolation, genomic data collection and SNP discovery*

125 Total genomic DNA was extracted using the DNA IQ Tissue and Hair Extraction Kit (Promega,
126 Madison, WI, USA) following the manufacturer's protocol. Each sample contained 60 hair
127 follicles with the majority of the hair shaft removed under a dissecting microscope to reduce
128 protein and other contamination. All DNA extractions were conducted in a separate laboratory
129 free of concentrated PCR products. Negative controls were included in each extraction to
130 monitor contamination. DNA quantifications were conducted using real-time PCR fluorescence
131 measurements of double stranded DNA (Blotta et al. 2005) and using the Quant-it kit (Life
132 Technologies, Foster City, CA).

133 Genomic DNA was converted into nextRAD genotyping-by-sequencing libraries
134 (SNPsaurus, LLC). The nextRAD method uses selective PCR primers to amplify genomic loci
135 consistently between samples. Genomic DNA (10 ng or less depending upon extraction yield)

was first fragmented with Nextera reagent (Illumina, Inc), which also ligates short adapter sequences to the ends of the fragments. Fragmented DNA was then amplified, with one of the primers matching the adapter and extending 9 arbitrary nucleotides into the genomic DNA with the selective sequence. Thus, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer will be efficiently amplified. The resulting fragments are fixed at the selective end, and have random lengths depending on the initial Nextera fragmentation. Because of this, amplified DNA from a particular locus is present at many different sizes and careful size selection of the library is not needed. For this project, an arbitrary 9-mer was chosen from those previously validated in smaller genomes, which did not appear to target repeat-masked regions in the publically available American pika genomic scaffolds (Ensembl, release 74, Ochotona_princeps.74.dna_sm.toplevel.fa) and that would approximate the results of standard RAD sequencing projects using SbfI (Baird et al. 2008).

Since these samples were collected non-invasively, it was important to assess the proportion of sequence reads in each sample that originated from the target organism relative to other environmental sources prior to conducting genotyping analysis. This was done using a custom script (SNPsaurus, LLC) that randomly sampled 1,000 high-quality reads from each sample and aligned those to the publically available American pika genomic scaffolds as well as subjected them to a BLASTN (Altschul et al. 1997) search of all sequences in the NCBI non-redundant database. Only samples that had greater than 50% sequencing reads that mapped to *Ochotona princeps* were retained for genotyping analysis.

The genotyping analysis used custom scripts (SNPsaurus, LLC) that created a reference from abundant reads present between 500 and 2000 times across the combined set of samples, mapping all of the reads to the reference allowing two mismatches. The identified variants were

159 then filtered by removing loci that had more than the expected maximum of two alleles and those
160 that were present in less than 10% of all samples.

161 Following assembly, mapping and variant detection, the data were further filtered to
162 maximize data quality. We retained only those loci that were genotyped in ≥ 50% of individuals
163 from each transect, had a minor allele frequency ≥ 0.05, and a minimum coverage of 6X for
164 homozygotes (affording 95% confidence in the genotype) while heterozygotes were required to
165 have a minimum of 2X coverage per allele for each individual. These values were chosen to
166 minimize null alleles and sequencing errors from biasing homozygote and heterozygote genotype
167 calls, respectively. We then removed loci that displayed significant deviation from Hardy-
168 Weinberg equilibrium (HWE) in more than two sites per transect as assessed using the method of
169 Guo and Thompson (1992) as implemented in GENEPOP 4.3 (Raymond & Rousset 1995; Rousset
170 2008).

171 To ensure that only non-redundant samples were included in subsequent analyses, we
172 conducted genotype matching across a random subset of 100 loci. We conducted the match
173 analysis and calculated the multi-locus probability of identity (Waits et al. 2001) for the 100
174 randomly chosen loci using GenAlEx (Peakall & Smouse 2006). Only samples with unique
175 genotypes were retained.

176

177 *Outlier locus detection and annotation*

178 Polymorphic loci were screened for statistical outliers using the Bayesian simulation method of
179 Beaumont and Balding (2004) as implemented in BAYESCAN 2.1 (Foll & Gaggiotti 2008). This
180 analysis was run independently for each transect, with all samples coded by site (PP1-PP4, TL1-
181 TL4). We used a prior odds value of 10, with 100,000 iterations and a burn-in of 50,000

182 iterations. We identified loci that were significant outliers at a q-value of 0.20. A q-value is a
183 false discovery rate (FDR) analogue of the p-value, with the former only defined in the context
184 of multiple testing, whereas the latter is defined on a single test. Consequently, a 20% threshold
185 for q-values is much more stringent than a 20% threshold for p-values in classical statistics. To
186 test for non-random association of genotypes, linkage disequilibrium was assessed between all
187 pairs of outlier loci in each population using the exact test of Guo and Thompson (1992) and
188 10,000 dememorization steps, 100 batches, and 10,000 iterations per batch as implemented in
189 GENEPOP 4.3 (Raymond & Rousset 1995; Rousset 2008). In addition, each haplotype from all
190 nextRAD-tags that contained outlier loci were subject to a BLASTN (Altschul et al. 1997) search
191 of all sequences in the NCBI non-redundant database (word size = 11; mismatch scores = 2, -3;
192 maximum e-value = 10-15). To reduce annotations to repetitive sequences in the database, we
193 required either a unique BLASTN hit or a top hit with an *e*-value that was at least an order of
194 magnitude lower than the next closest hit.
195
196 *Population genetic analyses*
197 We segregated loci into two datasets including: 1) all loci identified as an outlier ("outlier
198 dataset"); and 2) all loci not identified as an outlier ("neutral dataset"). The neutral dataset was
199 used to conduct standard population genetic analyses for quantifying the extent and distribution
200 of variation within and among sites. Within sites, proportion of polymorphic loci, observed (H_o)
201 and expected (H_e) heterozygosity, and gene diversity (N_g) were calculated using ARLEQUIN 3.5
202 (Excoffier & Lischer 2010). Global tests for heterozygote deficit were conducted using Fisher's
203 method and 10,000 dememorization steps, 100 batches, and 10,000 iterations per batch as
204 implemented in GENEPOP 4.3 (Raymond & Rousset 1995; Rousset 2008). The inbreeding

coefficient, F_{is}, was calculated for each site as implemented in GENETIX (Belkhir et al. 2004), with significance assessed using 1000 permutations. To evaluate whether site-level genetic diversity was correlated with elevation and sample size, we conducted linear regression analyses implemented in R v. 3.1 (R Development Core Team 2011).

Levels of genetic differentiation among groups were estimated by pairwise comparisons of θ (Weir & Cockerham 1984), as calculated in GENETIX (Belkhir et al. 2004), and evaluated using 1,000 permutations. The hierarchical organization of genetic variation within and among transects was calculated using an analysis of molecular variance (AMOVA) as implemented in ARLEQUIN 3.5 (Excoffier & Lischer 2010), with significance assessed using 1,000 permutations. In addition, the model-based clustering method implemented in STRUCTURE 2.3.4 (Pritchard et al. 2000) was used to infer the number of discrete genetic units across both transects. Run length was set to 100,000 MCMC replicates after a burn-in period of 100,000 using correlated allele frequencies under a straight admixture model. We varied the number of clusters (K) from 1 to 10, with 10 replicates for each value of K. The most likely number of clusters was determined by plotting the log probability of the data (ln Pr(X|K) (Pritchard et al, 2000) across the range of K values tested and selecting the K where the value of ln Pr(X|K) plateaued as suggested in the STRUCTURE manual. We also employed the ΔK method (Evanno et al. 2005) as calculated in STRUCTURE HARVESTER (Earl 2011). Results for the identified optimal values of K were summarized using CLUMPP (Jakobsson & Rosenberg 2007) and plotted using DISTRUCT (Rosenberg 2004). In order to test for unrecognized substructure in the broader STRUCTURE analysis, we repeated the above analysis for each transect separately using neutral and outlier loci.

Results

Data quality

The mean starting DNA concentration recovered from the non-invasively collected hair samples was 0.55 ng/µl with as little as 1 ng total for some samples. The mean number of sequencing reads per sample was 1,863,634. Ten samples yielded less than 100,000 sequencing reads, likely due to the degraded quality and very low quantity of starting DNA. Nineteen additional samples had less than 50% of their sequencing reads mapping to *O. princeps*. Sixteen of these samples had high proportions of sequence reads matching with two small mammals that likely co-occur in the sampling area [*Mus musculus* (n=13) and *Spermophilus* (n=3)], with others matching *Homo sapiens* (n=2) and *Zea mays* (n=1). The above samples (n=29) with low overall sequence reads or a low proportion mapping to *O. princeps* (or both) were removed, leaving 67 samples from eight sites across two elevational transects that were subject to all downstream analyses (Figure 1; Table 2).

We identified 9,825 SNPs that met the minimum parameters for recovering genotypes. To minimize linkage, we retained the highest coverage SNP from each contig, resulting in 3,830 SNPs. Twenty-seven SNPs deviated from HWE in two or more sites per transect and were removed from the dataset. Consequently, all downstream analyses were based on genotypic data at 3,803 SNPs. All 67 of the retained samples possessed unique genotypes at a random subset of 100 loci (average probability of identity within each sampling site = 1.1×10^{-23}), suggestive of unique individuals.

Outlier loci detection and annotation

Outlier detection identified 37 loci along the TL transect and 18 loci along the PP transect, none of which were shared. There was no evidence of significant deviation from linkage equilibrium for any pairwise comparison of outlier loci across populations. Fourteen outlier loci unambiguously matched sequences from the NCBI nr database, five of which annotated to genes of known functions (Table 1). Locus 57863_76 identified from the PP transect mapped to the receptor tyrosine kinase-like orphan receptor 2 (ROR2) gene that is part of a conserved family that function in developmental processes including skeletal and neuronal development, cell movement and cell polarity (Green et al. 2008). Likewise, locus 108547_114 identified from the TL transect annotated to another gene encoding a cell surface tyrosine kinase receptor (beta-type platelet-derived growth factor receptor), but for members of the platelet-derived growth factor family (Shim et al. 2010). Locus (28594_45) was similar to the laminin alpha 3 gene in humans that codes for a protein that is essential for formation and function of the basement membrane, with additional functions in regulating cell migration and mechanical signal transduction (Hamill et al. 2010). Lastly, locus 23486_75 was annotated to the hephaestin-like 1 (HEPHL1) gene that may function as a ferroxidase and may be involved in copper transport and homeostasis, while locus 33398_46 mapped to thioredoxin-related transmembrane protein 4 (TMX4) that may act as a reductase in the calnexin folding complex (Sugiura et al. 2010).

Population genetic analyses

The proportion of polymorphic loci varied across the sampling sites, with the lower elevation sites (PP1, PP2, TL1) exhibiting substantially lower numbers (P = 0.661-0.777) than found at the mid- and high-elevation sites (P = 0.837-0.947) for both transects (Table 2). Similar trends were seen for gene diversity along both transects where the low and mid- elevation sites recorded the

lowest values relative to higher elevation sites (Table 2). Indeed, both measures of site-level genetic variation were significantly correlated with elevation (P: r^2=0.557, p=0.034; N_g: r^2=0.738 p=0.006; Figure 2). Although P significantly correlated with sample size (r^2=0.635, p=0.018), this was not the case for gene diversity (r^2=0.0813, p=0.493) or elevation (r^2=0.184, p=0.289). This general trend of increasing variation with elevation seemed to hold for observed heterozygosity along the PP transect, but were stable across TL sites (H_o: 0.336-0.368; Table 2). Yet, all low (PP1, TL1) and mid-low (PP2, TL2) sites exhibited significant, genome-wide evidence of heterozygote deficit. Interestingly, significant inbreeding was also detected at the low and mid-low elevation sites along PP, with no such evidence at the higher elevation sites (PP3, PP4; Table 2). All sites along the TL transect exhibited evidence of inbreeding (Table 2).

The AMOVA revealed that a significant amount of variation ($p < 0.0001$) was exhibited both among transect (4.14%, d.f. = 1) and among sites within transect (2.01%, d.f. = 6), with the remaining found within populations (93.05%, d.f. = 126). These patterns were congruent with those from pairwise θ estimates, with the highest values generally displayed by among transect comparisons, but where all comparisons were significant (Table 3).

The Bayesian clustering analyses based on 3,748 neutral loci revealed strong evidence for two clusters within the dataset (ΔK_2 =249.0), corresponding to the two transects (Figure 1). When analyzing the PP transect separately, additional substructure (K = 2; Figure 3) was found using neutral (ΔK_2 = 33.1) and outlier (ΔK_2 = 123.1) loci. In both cases, the low elevation site (PP1) represents a largely distinct cluster relative to all other sites. Similarly, substructure was found along the TL transect when using outlier loci, with strong evidence for two (ΔK_2 =473.3) and three (ΔK_3 = 314.6) clusters. In the K= 2 plot, TL2 represented a distinct cluster, while in the K = 3 plot, the lower elevation sites (TL1, TL2) were each separate clusters relative to the high

296 elevation sites. Although the Evanno et al. (2005) method would favor $K = 2$, the method
297 described by Pritchard et al (2000) for inferring the optimal number of clusters would suggest K
298 = 3 given that ln Pr($X|K$) clearly plateaus at this value (Table S1). No substructure was found
299 along the TL transect based on neutral loci.
300
301 **Discussion**
302 Conservation genomics has become an increasingly popular term in the literature, yet practical
303 examples are limited (Shafer et al. 2015), including explicit consideration of the efficacy of
304 genomic data collection from non-invasively collected starting materials. Here, we demonstrated
305 the ability to identify 3,803 high confidence SNPs and recover genotypic data from low quantity
306 DNA originating from non-invasively collected American pika hair samples. These data allowed
307 us to detect outlier loci across elevational transects, identifying several candidate gene regions
308 that exhibit putative signatures of divergent selection and that can be investigated in future
309 studies for formulating mechanistic hypotheses. Moreover, the broad-scale genomic coverage
310 enabled precise estimation of population-level parameters, including standard diversity indices,
311 inbreeding, and structure within and among sampled transects.
312 We found genetic variation to be significantly correlated with elevation (Figure 2), with
313 sites at the lower fringe of American pika distribution in North Cascades National Park
314 exhibiting substantially lower levels of gene diversity. No such associations were found in a
315 previous microsatellite-based study conducted across elevationally distributed sites in British
316 Columbia, Canada (Henry et al. 2012). The detection of significant genome-wide evidence of
317 heterozygote deficit at low elevation sites in both transects further suggests inbreeding may be
318 leading to the observed patterns (Table 2), a particular concern for PP1, TL1 and TL2 given their

apparent distinctiveness from higher elevation sites (Figure 3). Due to their specific habitat requirements, patchy distribution, and life history, American pikas were long thought to regularly interbreed with close relatives based on observational studies (Smith 1993; Smith & Ivins 1983). Yet, molecular marker based studies have altered our understanding of American pika breeding behavior, revealing evidence for mate choice based on intermediate relatedness in one case (Peacock & Smith 1997b), while another found no evidence for inbreeding across elevationally distributed sites (Henry et al. 2012). This latter study conducted in Tweedsmuir South Provincial Park, British Columbia, Canada also found evidence for broad-scale and fine-scale population structure, detecting restricted gene flow among transects as well as among sites within transect and potentially driven by climatic factors (Henry et al. 2012). Although conducted at a different scale, Castillo et al. (2014) found a high degree of connectivity among geographically proximate sites in Crater Lake National Park, Oregon, USA, but restricted gene flow at a broader scale likely driven by topographic complexity and water. Here, we detected similar coarse-level patterns, detecting strong population genetic structure across transects but some evidence of connectivity between sites within transects in North Cascades National Park. It is worth noting that the three studies spanned the distribution of the Cascades lineage of American pika (Galbreath et al. 2009), conducted in the north (Henry et al. 2012), south (Castillo et al. 2014) and central (this study) portions of the range. Given that pikas are considered by some to be sentinels of climate change (Hafner 1993; Smith 1974), further investigation is warranted to infer underlying mechanisms associated with dispersal ability in pikas that may be further enhanced by comparative analyses of elevationally- and latitudinally-distributed sites.

More broadly, our results highlight a range of issues that must be considered when pairing genomic data collection with non-invasive sampling. First, sampling protocols must

endeavor to minimize non-target DNA during the collection process. In our case, the use of tape-based, non-invasive hair snares allowed us to collect genomic data from 67 individuals of American pika, but also yielded 19 samples that were almost entirely composed of DNA sequence reads from non-target organisms, primarily other small mammals that were likely using the same talus habitat. Precautions to avoid such contamination will vary according to the methods in which samples are obtained, but are of critical importance given the non-targeted nature of NGS approaches. Depending upon study questions, the use of exon-capture or other approaches for preferentially targeting the DNA of study organisms within a mixed sample may help to minimize contamination and maximize cost-effectiveness of downstream sequence data (Avila-Arcos et al. 2011; Carpenter et al. 2013; Good 2011). Exon-capture, in particular, has been effectively applied to historical DNA collected from museum specimens (Bi et al. 2013), which typically yield DNA of lower quantity and quality similar to non-invasively collected starting materials. Yet, these approaches are substantially more costly and, in the case of exon-capture, limited to expressed regions of the genome. However, for some non-invasively collected source of DNA such as feces, the use of capture approaches may be obligatory (Perry et al. 2010).

Additionally, rigorous assessments of resulting DNA sequence data must be undertaken to ensure quality control. In the current study, we used a genotyping-by-sequencing approach for reduced representation genomic data collection. We had the advantage of publicly available American pika genomic scaffolds that allowed us to initially filter our data based on SNPs assembling to these references. At present, such resources may not be available for many organisms of conservation interest. In such cases, we recommend using the closest available genome to inform reference assembly of identified SNPs (in our case, this would have been the

365 European rabbit; Lindblad-Toh et al. 2011). If no suitable reference genome is available,
366 investigators may want to consider capture approaches for genomic data collection (as discussed
367 above).

368 Analytical frameworks must also be carefully considered in relation to recovered
369 sequence coverage depth in studies using non-invasively collected samples. In our case, we used
370 explicit parameters related to coverage and amounts of allowable missing data to confidently
371 reconstruct genotypes from our sampled individuals. While there is no clear standard in the
372 literature, choice of such parameters is a balance between maximizing the number of loci and
373 minimizing null alleles when reconstructing genotypes. Yet, reconstructed genotypes may not be
374 necessary for all study questions, especially those primarily focused on estimating population–
375 level parameters rather than individual-based measures (e.g. admixture coefficients, individual
376 identification, parentage probabilities; but see Buerkle & Gompert 2013). In such cases, low
377 density genomic scans based on more individuals or sites in the genome may provide highly
378 accurate and precise population parameter estimates, even at as low as 1X coverage (Buerkle &
379 Gompert 2013; Fumagalli 2013). Analytical pipelines continue to be developed that implement
380 population genetic analysis methods that account for the statistical uncertainty of NGS data
381 (Fumagalli et al. 2014), with empirical examples now found in the literature (Cahill et al. 2013).

382 Overall, NGS data and population genomic analyses hold great promise for informing
383 conservation-related studies, substantially increasing the number of markers to allow for more
384 accurate and precise estimates of population structure and demographic parameters (Primmer
385 2009), as well as the ability to detect adaptive genetic variation for informing conservation unit
386 delimitation (Funk et al. 2012) and decision frameworks aimed at reducing the long-term impacts
387 of climate change on biodiversity (Hoffmann et al. 2015). Here, we have shown that with

careful consideration, genomic data collection is compatible with the non-invasive sampling required in practice for many conservation-related studies.

Acknowledgments

We thank Regina Rochefort, Roger Christophersen, Liesl Erb, Kelsey Robson, Shane Schoolman, and Aidan Beers for assistance with site selection and sampling. Evelyn Jensen and Andrew Veale provided feedback on the manuscript. Holly Buhler helped prepare the map figure. A special thanks to the Department of Organismic and Evolutionary Biology at Harvard University for hosting MR during his sabbatical visit, during which some of the initial development of this work occurred.

References

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, and Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:3389-3402.

Avila-Arcos MC, Cappellini E, Alberto Romero-Navarro J, Wales N, Victor Moreno-Mayar J, Rasmussen M, Fordyce SL, Montiel R, Vielle-Calzada J-P, Willerslev E, and Gilbert MTP. 2011. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Scientific Reports* 1.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, and Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376.

Beaumont MA, and Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* 13:969-980.

Beever EA, Dobrowski SZ, Long J, Mynsberge AR, and Piekielek NB. 2013. Understanding relationships among abundance, extirpation, and climate at ecoregional scales. *Ecology* 94:1563-1571.

Beever EA, Ray C, Mote PW, and Wilkening JL. 2010. Testing alternative models of climate-mediated extirpations. *Ecological Applications* 20:164-178.

Beever EA, and Smith AT. 2011. *Ochotona princeps. IUCN Red List of Threatened Species*.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, and Luikart G. 2009. Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources* 9:1279-1301.

421 Belkhir K, Borsa P, Chikhi L, Raufaste N, and Bonhomme F. 2004. GENETIX 4.0.5.2.,
422 Windows TM software for population genetics. Laboratoire Génome, Populations,
423 Interactions, CNRS UMR 5171. Montpellier, France: Université de Montpellier II.
424 Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, and Moritz C. 2013. Unlocking the
425 vault: next-generation museum population genomics. *Molecular Ecology* 22:6018-6032.
426 Blotta I, Prestinaci F, Mirante S, and Cantafora A. 2005. Quantitative assay of total dsDNA with
427 PicoGreen reagent and real-time fluorescent detection. *Annali dell'Istituto superiore di*
428 *sanita* 41:119-123.
429 Buerkle CA, and Gompert Z. 2013. Population genomics based on low coverage sequencing:
430 how low should we go? *Molecular Ecology* 22:3028-3035.
431 Cahill JA, Green RE, Fulton TL, Stiller M, Jay F, Ovsyanikov N, Salamzade R, St John J,
432 Stirling I, Slatkin M, and Shapiro B. 2013. Genomic evidence for island population
433 conversion resolves conflicting theories of polar bear evolution. *Plos Genetics* 9.
434 Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, Rasmussen
435 M, Gravel S, Guillen S, Nekhrizov G, Leshtakov K, Dimitrova D, Theodossiev N,
436 Pettener D, Luiselli D, Sandoval K, Moreno-Estrada A, Li Y, Wang J, Gilbert MTP,
437 Willerslev E, Greenleaf WJ, and Bustamante CD. 2013. Pulling out the 1%: Whole-
438 Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries.
439 *American Journal of Human Genetics* 93:852-864.
440 Castillo JA, Epps CW, Davis AR, and Cushman SA. 2014. Landscape effects on gene flow for a
441 climate-sensitive montane species, the American pika. *Molecular Ecology* 23:843-856.
442 Earl DA. 2011. Structure harvester v0.6.5. *Available at*
443 http://taylor0.biology.ucla.edu/struct_harvest/.
444 Etter PD, Bassham S, Hohenlohe PA, Johnson EA, and Cresko WA. 2011. SNP discovery and
445 genotyping for evolutionary genetics using RAD sequencing. *Methods in molecular*
446 *biology (Clifton, NJ)* 772:157-178.
447 Evanno G, Regnaut S, and Goudet J. 2005. Detecting the number of clusters of individuals using
448 the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611-2620.
449 Excoffier L, and Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform
450 population genetics analyses under Linux and Windows. *Molecular Ecology Resources*
451 10:564-567.
452 Foll M, and Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for
453 both dominant and codominant markers: a bayesian perspective. *Genetics* 180:977-993.
454 Fumagalli M. 2013. Assessing the effect of sequencing depth and sample size in population
455 genetics inferences. *PLoS ONE* 8.
456 Fumagalli M, Vieira FG, Linderoth T, and Nielsen R. 2014. ngsTools: methods for population
457 genetics analyses from next-generation sequencing data. *Bioinformatics* 30:1486-1487.
458 Funk CW, McKay JK, Hohenlohe PA, and Allendorf FW. 2012. Harnessing genomics for
459 delineating conservation units. *Trends in Ecology and Evolution* 27:489-496.
460 Galbreath KE, Hafner DJ, and Zamudio KR. 2009. When Cold Is Better: Climate-Driven
461 Elevation Shifts Yield Complex Patterns of Diversification and Demography in an Alpine
462 Specialist (American Pika, *Ochotona Princeps*). *Evolution* 63:2848-2863.
463 Good JM. 2011. Reduced representation methods for subgenomic enrichment and next-
464 generation sequencing. *Methods in Molecular Biology* 772:85-103.
465 Grayson DK. 2005. A brief history of Great Basin pikas. *Journal of Biogeography* 32:2103-
466 2111.

Green JL, Kuntz SG, and Sternberg PW. 2008. Ror receptor tyrosine kinases: orphans no more. *Trends in Cell Biology* 18:536-544.

Guo SW, and Thompson EA. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361-372.

Hafner DJ. 1993. North-American Pika (Ochotona-Princeps) as a Late Quaternary Biogeographic Indicator Species. *Quaternary Research* 39:373-380.

Hamill KJ, Paller AS, and Jones JCR. 2010. Adhesion and Migration, the Diverse Functions of the Laminin alpha 3 Subunit. *Dermatologic Clinics* 28:79-+.

Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogden R, Limborg MT, Cariani A, Maes GE, Diopere E, Carvalho GR, and Nielsen EE. 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* 11:123-136.

Henry P, and Russello MA. 2011. Obtaining high-quality DNA from elusive small mammals using low-tech hair snares. *European Journal of Wildlife Research* 57:429-435.

Henry P, and Russello MA. 2013. Adaptive divergence along environmental gradients in a climate-change-sensitive mammal. *Ecology and Evolution* 3:3906-3917.

Henry P, Sim Z, and Russello MA. 2012. Genetic evidence for highly restricted dispersal along continuous altitudinal gradients in the climate-change sensitive American pika. *PLoS ONE* 7:e39077.

Hoffmann A, Griffin P, Dillon S, Catullo R, Rane R, Byrne M, Jordan R, Oakeshott J, Weeks A, Joseph L, Lockhart P, Borevitz J, and Sgrò C. 2015. A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses* 2:1 (24 pages).

Jakobsson M, and Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801-1806.

Karl TR, Melillo JM, Peterson TC, Anderson DM, Boesch DF, Burkett V, Carter LM, Cohen SJ, Grimm NB, Hatfield JL, Hayhoe K, Janetos A, Kaye JA, Lawrimore J, McCarthy J, McGuire AD, Miles E, Mills E, Overpeck JT, Patz J, Pulwarthy R, Santer B, Savonis MJ, Schwartz HG, Shea E, Stone J, Udall BH, Walsh J, Wehner MF, Wilbanks TJ, and Wuebbles D. 2009. *Global Climate Change Impacts in the United States: A State of Knowledge Report from the U.S. Global Change Research Program*. New York: Cambridge University Press.

Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, Kheradpour P, Ernst J, Jordan G, Mauceli E, Ward LD, Lowe CB, Holloway AK, Clamp M, Gnerre S, Alfoeldi J, Beal K, Chang J, Clawson H, Cuff J, Di Palma F, Fitzgerald S, Flicek P, Guttman M, Hubisz MJ, Jaffe DB, Jungreis I, Kent WJ, Kostka D, Lara M, Martins AL, Massingham T, Moltke I, Raney BJ, Rasmussen MD, Robinson J, Stark A, Vilella AJ, Wen J, Xie X, Zody MC, Worley KC, Kovar CL, Muzny DM, Gibbs RA, Warren WC, Mardis ER, Weinstock GM, Wilson RK, Birney E, Margulies EH, Herrero J, Green ED, Haussler D, Siepel A, Goldman N, Pollard KS, Pedersen JS, Lander ES, Kellis M, Broad I, Baylor Coll M, and Washington U. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478:476-482.

McMahon BJ, Teeling EC, and Hoglund J. 2014. How and why should we implement genomics into conservation? *Evolutionary Applications* 7:999-1007.

512 Peacock MM, and Smith AT. 1997a. The effect of habitat fragmentation on dispersal patterns,
513 mating behavior, and genetic variation in a pika (Ochotona princeps) metapopulation.
514 *Oecologia* 112:524-533.
515 Peacock MM, and Smith AT. 1997b. Nonrandom mating in pikas *Ochotona princeps*: evidence
516 for inbreeding between individuals of intermediate relatedness. *Molecular Ecology*
517 6:801-811.
518 Peakall R, and Smouse PE. 2006. GENALEX 6: genetic analysis in Excel. Population genetic
519 software for teaching and research. *Molecular Ecology Notes* 6:288-295.
520 Perry GH, Marioni JC, Melsted P, and Gilad Y. 2010. Genomic-scale capture and sequencing of
521 endogenous DNA from feces. *Molecular Ecology* 19:5332-5344.
522 Primmer CR. 2009. From Conservation Genetics to Conservation Genomics. *Year in Ecology*
523 *and Conservation Biology 2009* 1162:357-368.
524 Pritchard JK, Stephens M, and Donnelly P. 2000. Inference of population structure using
525 multilocus genotype data. *Genetics* 155:945-959.
526 R Development Core Team. 2011. R: A language and environment for statistical computing. R
527 Foundation for Statistical Computing, Vienna, Austria.
528 Raymond M, and Rousset F. 1995. GENEPOP (version-1.2) – population genetics software for
529 exact tests and ecumenicism. *Journal of Heredity* 86:248-249.
530 Roon DA, Waits LP, and Kendall KC. 2003. A quantitative evaluation of two methods for
531 preserving hair samples. *Molecular Ecology Notes* 3:163-166.
532 Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure.
533 *Molecular Ecology Notes* 4:137-138.
534 Rousset F. 2008. GENEPOP ' 007: a complete re-implementation of the GENEPOP software for
535 Windows and Linux. *Molecular Ecology Resources* 8:103-106.
536 Schoville SD, Bonin A, Francois O, Lobreaux S, Melodelima C, and Manel S. 2012. Adaptive
537 genetic variation on the landscape: methods and cases. *Annual Review of Ecology,*
538 *Evolution, and Systematics* 43:23-43.
539 Shafer ABA, Wolf JBW, Alves PC, Bergström L, Bruford MW, Brännström I, Colling G, Dalén
540 L, Meester LD, Ekblom R, Fawcett KD, Fior S, Hajibabaei M, Hill JA, Hoezel AR,
541 Höglund J, Jensen EL, Krause J, Kristensen TN, Krützen M, McKay JK, Norman AJ,
542 Ogden R, Österling EM, Ouborg NJ, Piccolo J, Popović D, Primmer CR, Reed FA,
543 Roumet M, Salmona J, Schenekar T, Schwartz MK, Segelbacher G, Senn H, Thaulow J,
544 Valtonen M, Veale A, Vergeer P, Vijay N, Vilà C, Weissensteiner M, Wennerström L,
545 Wheat CW, and Zieliński P. 2015. Genomics and the challenging translation into
546 conservation practice. *Trends in Ecology & Evolution* 30:78-87.
547 Shim AH-R, Liu H, Focia PJ, Chen X, Lin PC, and He X. 2010. Structures of a platelet-derived
548 growth factor/propeptide complex and a platelet-derived growth factor/receptor complex.
549 *Proceedings of the National Academy of Sciences of the United States of America*
550 107:11307-11312.
551 Smith AT. 1974. Distribution and Dispersal of Pikas - Influences of Behavior and Climate.
552 *Ecology* 55:1368-1376.
553 Smith AT. 1993. *The natural history of inbreeding and outbreeding in small mammals*.
554 Smith AT, and Ivins BL. 1983. Colonization in a Pika Population - Dispersal Vs Philopatry.
555 *Behavioral Ecology and Sociobiology* 13:37-47.
556 Smith AT, and Weston ML. 1990. *Ochotona princeps*. *Mammalian Species* 352:1-8.

557 Smith O, and Wang J. 2014. When can noninvasive samples provide sufficient information in
558 conservation genetics studies? *Molecular Ecology Resources* 14:1011-1023.
559 Stewart JAE, Perrine JD, Nichols LB, Thorne JH, Millar CI, Goehring KE, Massing CP, and
560 Wright DH. 2015. Revisiting the past to foretell the future: summer temperature and
561 habitat area predict pika extirpations in California. *Journal of Biogeography* 42:880-890.
562 Sugiura Y, Araki K, Iemura S-i, Natsume T, Hoseki J, and Nagata K. 2010. Novel Thioredoxin-
563 related Transmembrane Protein TMX4 Has Reductase Activity. *Journal of Biological*
564 *Chemistry* 285:7135-7142.
565 Taberlet P, Waits LP, and Luikart G. 1999. Noninvasive genetic sampling: look before you leap.
566 *Trends in Ecology & Evolution* 14:323-327.
567 Vali U, Einarsson A, Waits L, and Ellegren H. 2008. To what extent do microsatellite markers
568 reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*
569 17:3808-3817.
570 Waits LP, Luikart G, and Taberlet P. 2001. Estimating the probability of identity among
571 genotypes in natural populations: cautions and guidelines. *Molecular Ecology* 10:249-
572 256.
573 Waits LP, and Paetkau D. 2005. Noninvasive genetic sampling tools for wildlife biologists: A
574 review of applications and recommendations for accurate data collection. *Journal of*
575 *Wildlife Management* 69:1419-1433.
576 Weir BS, and Cockerham CC. 1984. Estimating F-statistics for the analysis of population
577 structure. *Evolution* 38:1358-1370.

578

579 **Supporting Information Captions**

580 Table S1. Summarized STRUCTURE output including ΔK for the Thornton Lake elevational

581 transect based on 37 outlier loci.

582

583 **Data Accessibility**

584 All SNP genotypic data are deposited in Dryad doi: ####.

585

586 **Figure Captions**

587 Figure 1. Sites in North Cascades National Park, Washington, USA where America pika hair
588 samples were non-invasively collected. Topographic lines represent 100 m elevation. Inset
589 shows a STRUCTURE bar plot depicting the model-based clustering results for all sites within the
590 Pyramid Peak (PP) and Thornton Lake (TL) elevational transects based on 3,748 neutral single
591 nucleotide polymorphisms.

592

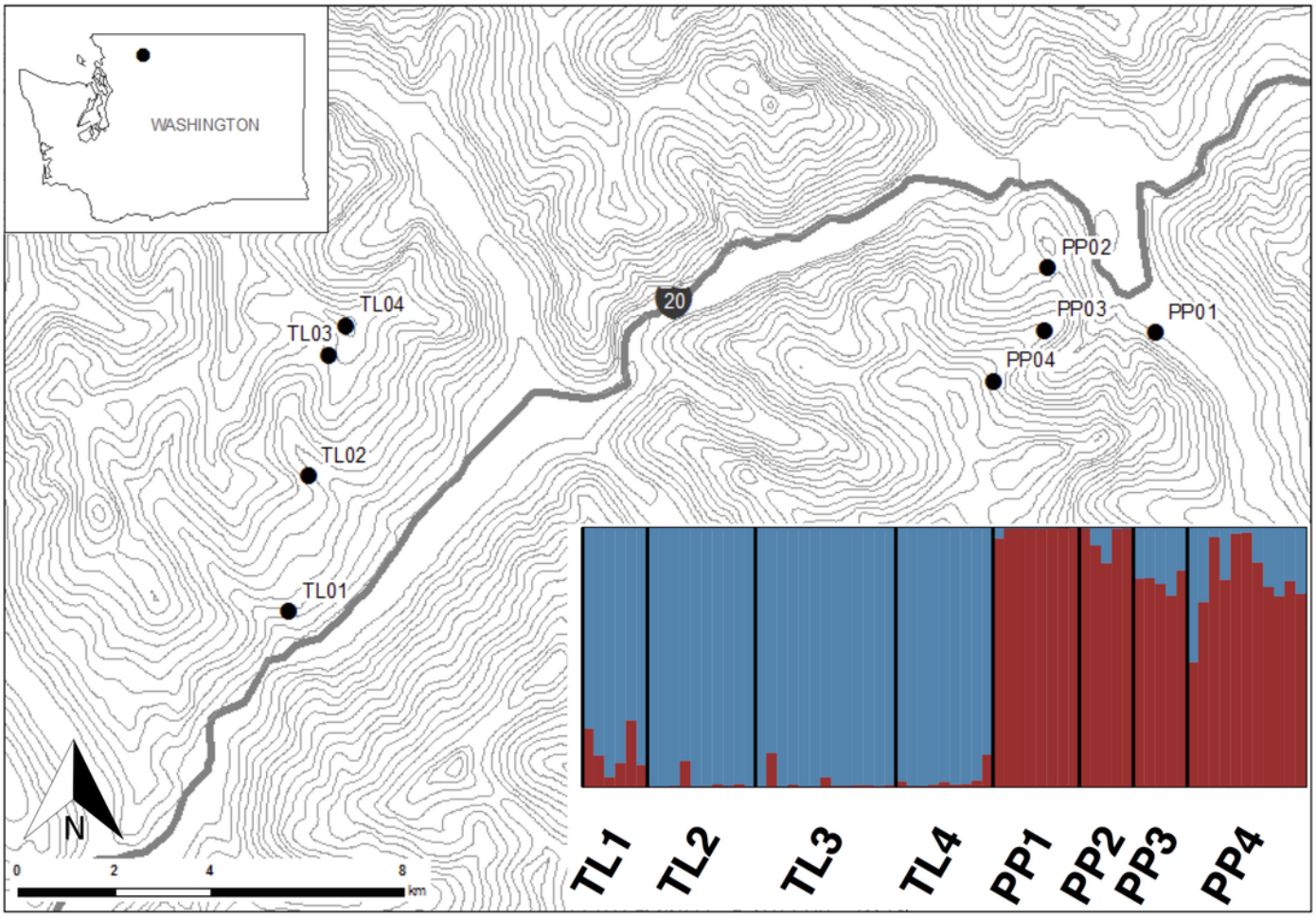
593 Figure 2. Elevational patterns of genetic diversity within American pika samples in the North
594 Cascade National Park. Solid line shows the correlation between proportions of polymorphic loci
595 (P; circles) with elevation (r^2=0.557 p=0.034). Dashed line shows the correlation between gene
596 diversity (N_g; squares) with elevation (r^2=0.738 p=0.006).

597

598 Figure 3. STRUCTURE bar plots depicting the model-based clustering results for Thornton Lake
599 (TL) and Pyramid Peak (PP) sites based on outlier loci (above) and neutral loci (below).
600 Analyses for the TL transect revealed evidence for both K = 2 (ΔK = 473.3) and K = 3 (ΔK =
601 314.6; plot shown) based on 37 outlier loci, and K = 1 (K = 2 plot shown for display purposes)
602 based on 3748 neutral loci. Analyses for the PP transect revealed evidence for K = 2 (ΔK =
603 123.1) based on 18 outlier loci, and K = 2 (ΔK = 33.1) based on 3,748 neutral loci.

1

Figure 1: Sites in North Cascades National Park, Washington, USA where America pika hair samples were non-invasively collected.



2

Figure 2: Elevational patterns of genetic diversity within American pika samples in the North Cascade National Park.

Solid line shows the correlation between proportions of polymorphic loci (*P*; circles) with elevation (r^2=0.557 p=0.034). Dashed line shows the correlation between gene diversity (N_g; squares) with elevation (r^2=0.738 p=0.006).

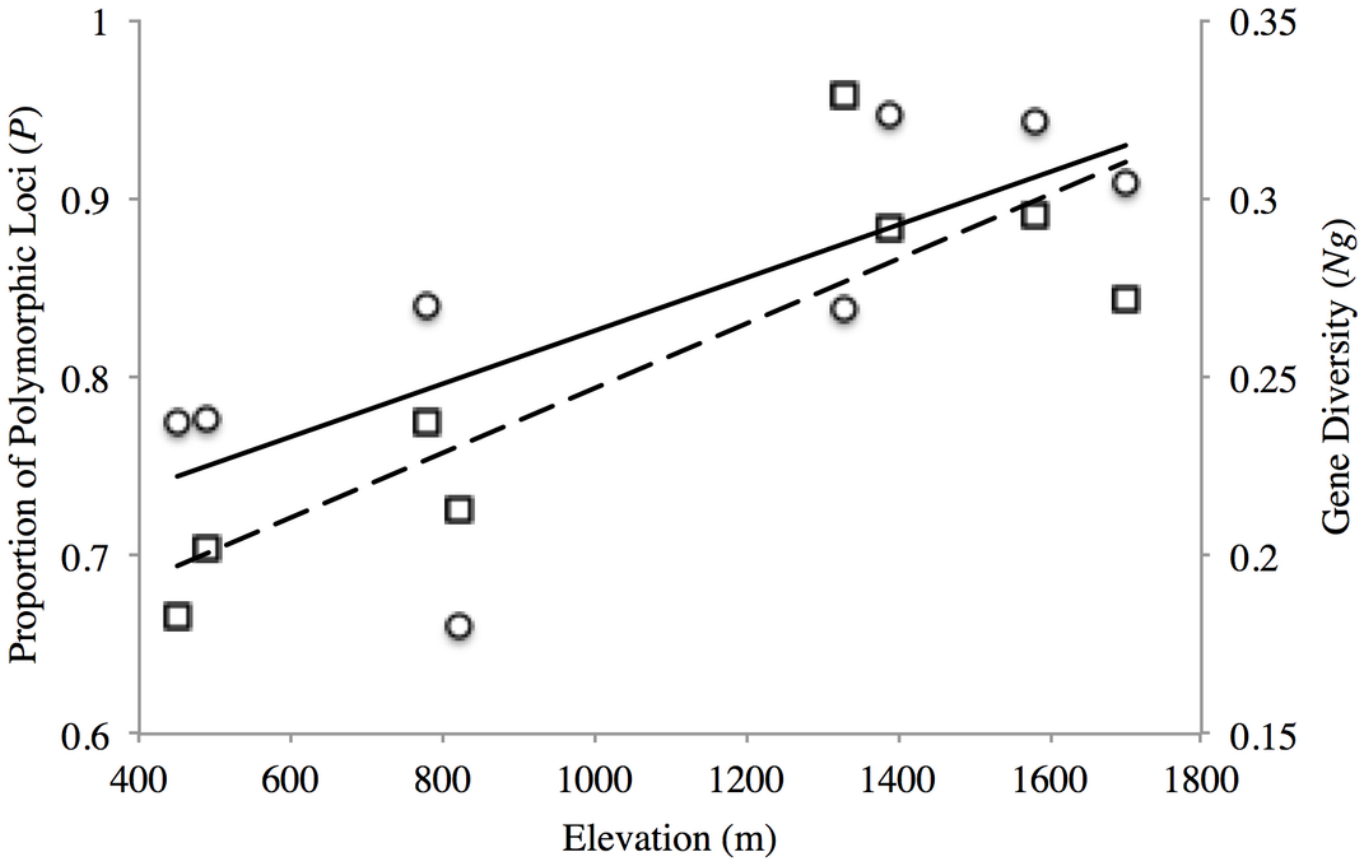


Table 1(on next page)

Summary of outlier loci detected for American pika sampling sites along the Pyramid Peak (PP) and Thornton Lake (TL) elevational transects in North Cascades National Park.

2 Table 1. Summary of outlier loci detected for American pika sampling sites along the Pyramid

3 Peak (PP) and Thornton Lake (TL) elevational transects in North Cascades National Park.

Locus	SNP	Transect	F_{ST}*	Top Blast hit (Accession)	Abbreviated description
21404_70	C / T	PP	0.315	AC234826	*Ochotona princeps* clone VMRC40-45K5
33398_46	C / T	PP	0.261	XM004593395	*Ochotona princeps* thioredoxin-related transmembrane protein 4
57863_76	A / G	PP	0.255	NG008089.1	*Homo sapiens* receptor tyrosine kinase-like orphan receptor 2
23902_347	C / T	TL	0.248	AC237024	*Ochotona princeps* clone VMRC40-172G4
46878_140	C / G	TL	0.234	AC234901	*Oryctolagus cuniculus*, clone 0087B06
72966_67	A / G	PP	0.229	AL358859	Human DNA sequence from clone RP11-545G13 on chromosome 1
59691_160	C / T	TL	0.223	AC234021	*Ochotona princeps* clone VMRC40-93N24
23486_75	A / C	TL	0.209	XM004585173	*Ochotona princeps* hephaestin-like 1
110148_49	A / G	TL	0.196	AC165118	*Oryctolagus cuniculus*, clone 16788057J9
94981_43	A / T	PP	0.194	AC236101	*Ochotona princeps* clone VMRC40-347J6
43241_27	C / T	TL	0.171	AC233835	*Ochotona princeps* clone VMRC40-526O13
108547_114	A / T	TL	0.165	XM004587540	*Ochotona princeps* platelet-derived growth factor receptor, beta
87086_98	C / G	TL	0.155	XM004593191	*Ochotona princeps* putative uncharacterized protein FLJ46204-like
28594_45	C / T	TL	0.153	NG007853	*Homo sapiens* laminin, alpha 3

4 *F_{ST} values significantly higher than under neutral expectations; averaged over populations.

Table 2(on next page)

Genetic variation within American pika samples sites along the Pyramid Peak (PP) and Thornton Lake (TL) elevational transects in North Cascades National Park.

Table 2. Genetic variation within American pika samples sites along the Pyramid Peak (PP) and Thornton Lake (TL) elevational transects in North Cascades National Park.

Site	Elevation	*n*	*P*	H_o	H_e	N_g	F_{is}
PP1	450	8	0.774	0.282*	0.372	0.183	0.260*
PP2	820	5	0.661	0.314*	0.425	0.213	0.295*
PP3	1330	5	0.837	0.403	0.403	0.329	0.001
PP4	1580	11	0.943	0.383	0.359	0.295	-0.071
TL1	490	6	0.777	0.368*	0.400	0.202	0.088*
TL2	780	10	0.839	0.339*	0.362	0.237	0.067*
TL3	1390	13	0.947	0.336	0.349	0.292	0.039*
TL4	1700	9	0.908	0.356	0.364	0.272	0.023*

Elevation in meters; sample size (*n*); proportion of polymorphic loci (*P*); observed heterozygosity (H_o); unbiased expected heterozygosity (H_e); gene diversity (N_g); inbreeding coefficient (F_{is}); * $p<0.05$.

Table 3(on next page)

Pairwise θ estimates for American pika sites within and among the Pyramid Peak (PP) and Thornton Lake (TL) elevational transects in North Cascades National Park.

Table 3. Pairwise θ estimates for American pika sites within and among the Pyramid Peak (PP) and Thornton Lake (TL) elevational transects in North Cascades National Park.

Site	PP1	PP2	PP3	PP4	TL1	TL2	TL3	TL4
PP1	-	0.066	0.056	0.054	0.105	0.109	0.101	0.098
PP2		-	0.040	0.049	0.096	0.101	0.087	0.088
PP3			-	0.020	0.072	0.072	0.062	0.056
PP4				-	0.065	0.068	0.059	0.053
TL1					-	0.051	0.040	0.042
TL2						-	0.027	0.030
TL3							-	0.015
TL4								-

Results based on 3,748 neutral single nucleotide polymorphisms. All pairwise θ estimates were significant ($p < 0.05$). Pairwise values for among transect comparisons shaded in gray.

3

Figure 3: Structure bar plots depicting the model-based clustering results for Thornton Lake (TL) and Pyramid Peak (PP) sites based on outlier loci (above) and neutral loci (below).

Analyses for the TL transect revealed evidence for both $K = 2$ ($\Delta K = 473.3$) and $K = 3$ ($\Delta K = 314.6$; plot shown) based on 37 outlier loci, and $K = 1$ ($K = 2$ plot shown for display purposes) based on 3748 neutral loci. Analyses for the PP transect revealed evidence for $K = 2$ ($\Delta K = 123.1$) based on 18 outlier loci, and $K = 2$ ($\Delta K = 33.1$) based on 3,748 neutral loci.