

Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*)

HAYLEY C. LANIER,* ROB MASSATTI,† QIXIN HE,† LINK E. OLSON‡ and L. LACEY KNOWLES†

*Department of Zoology and Physiology, University of Wyoming at Casper, Casper, WY 82601, USA, †Department of Ecology and Evolutionary Biology, The University of Michigan, Ann Arbor, MI 41809-1079, USA, ‡University of Alaska Museum, University of Alaska Fairbanks, Fairbanks, AK 99775, USA

Abstract

Identifying the genetic structure of a species and the factors that drive it is an important first step in modern population management, in part because populations evolving from separate ancestral sources may possess potentially different characteristics. This is especially true for climate-sensitive species such as pikas, where the delimitation of distinct genetic units and the characterization of population responses to contemporary and historical environmental pressures are of particular interest. We combined a restriction site-associated DNA sequencing (RADSeq) data set containing 4156 single nucleotide polymorphisms with ecological niche models (ENMs) of present and past habitat suitability to characterize population composition and evaluate the effects of historical range shifts, contemporary climates and landscape factors on gene flow in Collared Pikas, which are found in Alaska and adjacent regions of northwestern Canada and are the lesser-studied of North America's two pika species. The results suggest that contemporary environmental factors contribute little to current population connectivity. Instead, genetic diversity is strongly shaped by the presence of three ancestral lineages isolated during the Pleistocene (~148 and 52 kya). Based on ENMs and genetic data, populations originating from a northern refugium experienced longer-term stability, whereas both southern lineages underwent population expansion – contradicting the southern stability and northern expansion patterns seen in many other taxa. Current populations are comparable with respect to generally low diversity within populations and little-to-no recent admixture. The predominance of divergent histories structuring populations implies that if we are to understand and manage pika populations, we must specifically assess and accurately account for the forces underlying genetic similarity.

Keywords: Alaska, climate change, mammal, phylogeography, STRUCTURE

Received 30 September 2014; revision received 27 May 2015; accepted 29 May 2015

Introduction

The characterization of the genetic structure of a species and the factors that drive it are important first steps in population management because the dynamic spatio-temporal histories of populations can profoundly impact their future evolutionary potential. For example, not only do contemporary factors (i.e. geographic distance

separating populations and environmental connectivity; e.g. Storfer *et al.* 2007) contribute to patterns of genetic variation, but differences between the present and past distributions, where contemporary populations are founded from divergent ancestral source populations, may act to structure genetic diversity among extant populations (e.g. Massatti & Knowles 2014). Understanding the relative roles of these factors is critical as unrecognized historical patterns may potentially mislead analyses of population connectivity (He *et al.* 2013) and divergent histories among populations may impact the

Correspondence: Hayley Lanier, Fax: +1 307 268 2416; E-mail: hlanier@uwyo.edu

potential for similar evolutionary responses to environmental change (Palumbi *et al.* 2014). This is especially true in species with limited vagility (Franken & Hik 2004; Zgurski & Hik 2012), strong territoriality (Franken 2002) or narrow habitat requirements (Horn 2013). For climate-sensitive species such as pikas (Smith *et al.* 2004), understanding the joint influence of present environmental factors and historical population connectivity (i.e. over evolutionary timescales) on genetic composition is not only important, but absolutely pressing for population management decisions (USFWS 2010; COSEWIC 2011).

Pikas (Mammalia: Lagomorpha; Ochotonidae) are small (~150 g), alpine-associated herbivores found in mountainous regions of western North America and Asia (Smith 2008). Like their closest living relatives – rabbits and hares – pikas do not hibernate and instead spend much of their summer months collecting and caching food to survive the winter (Broadbooks 1965). Unlike rabbits, pikas are particularly sensitive to warm temperatures; the American Pika (*Ochotona princeps*) cannot survive for more than 2 h at ambient temperatures above 28 °C (82.4 °F) before succumbing to hyperthermia and death (MacArthur & Wang 1973). This sensitivity and their association with fragmented habitats such as alpine talus slopes and boulder fields are thought to have contributed to ongoing local extirpation in multiple pika species and have led to their consideration as a sentinel species for climate change (Smith *et al.* 2004). However, much of our current understanding of pika response to warming is based on either highly localized studies (e.g. Morrison & Hik 2007) or those conducted in a questionably representative portion of one species' range (e.g. Beever *et al.* 2003). Furthermore, environmental factors influencing the colonization process and thus the potential for gene flow in pikas remain poorly understood. Because genetic signals of contemporary connectivity are influenced by shared ancestry as well as ongoing gene flow, understanding how contemporary factors shape genetic structure requires a joint analysis of historical connectivity and current environmental pressures.

North America is home to two species of pika, both occurring in mountainous regions in the western half of the continent. Approximately 650 km currently separates the range of Collared Pikas (*Ochotona collaris*), found in Alaska and northwest Canada, from that of the American Pikas to the south (COSEWIC 2011). Divergence between these two sister species appears to have occurred as far back as the Pliocene (~4.6 mya; Lanier & Olson 2009). Although pikas originally evolved in Asia (Dawson 1967) and migrated into North America through the Bering Land Bridge, strong evidence suggests that *O. collaris* originated from an ancestor that

dispersed secondarily to higher latitudes from lower North American latitudes (Galbreath & Hoberg 2012). Fossil (Guthrie 1973; Harrington 2011) and molecular evidence (Lanier & Olson 2013) indicate that small-bodied pikas (similar to either *O. collaris* or *O. princeps*) were present in parts of Beringia during most of the Pleistocene; however, the location of refugial populations is not well understood (Lanier *et al.* 2015).

Studies of historical biogeography of Collared Pikas have concluded that population differentiation has been in part driven by Pleistocene population distributions (Lanier & Olson 2013). Two mitochondrial lineages are present, corresponding to inferred western (Beringian) and eastern Pleistocene refugia (Lanier & Olson 2013). Studies of mountain goats (Shafer *et al.* 2011) and alpine plants (Marr *et al.* 2008) suggest that northern British Columbia may be a hot spot of genetic diversity due to the presence of nunataks (isolated rock peaks surrounded by glaciers or ice sheets; Pielou 1991) that served as refugia. However, the limited resolution afforded by mtDNA has hindered efforts to identify the refugial locations for Collared Pikas (Lanier & Olson 2013). If nunatak refugia were important in shaping contemporary genetic diversity, we expect to observe multiple source populations (i.e. refugia) and small ancestral population sizes in genomic data. Furthermore, genomic diversity should be the highest in regions where historically separated lineages intersect (Petit *et al.* 2003), such as Wrangell-St. Elias National Park (as indicated from mtDNA; Lanier *et al.* 2015).

Ecologically, much is known about the American Pika but far less is known about the Collared Pika, particularly with regard to its response to ongoing climate change. This is due in part to the short summer field season at higher latitudes and extremely limited road access throughout the range of the Collared Pika. Most recent research on Collared Pikas has focused on a single locality in southwest Yukon Territory ('Pika Camp'; see Morrison & Hik 2007; Horn 2013). Like American Pikas, Collared Pikas specialize on alpine talus and boulder fields (MacDonald & Jones 1987). During the summer, they rarely forage >10 m into surrounding meadows (McIntire & Hik 2002). Their overwinter survival is dependent on the collection of sufficient forage (Morrison *et al.* 2009) and enough snowpack to insulate the talus they inhabit (Morrison & Hik 2007). Collared Pikas are notably poor dispersers (539 m average and 1.8 km maximum individual movements, based on genetic evidence; Zgurski & Hik 2014), which should lead to high differentiation among disjunct populations and an erosion of diversity in peripheral populations (Fig. 1). Thus, the contemporary configuration of populations relative to low-elevation barriers would suggest mountain ranges may play an important role in shaping

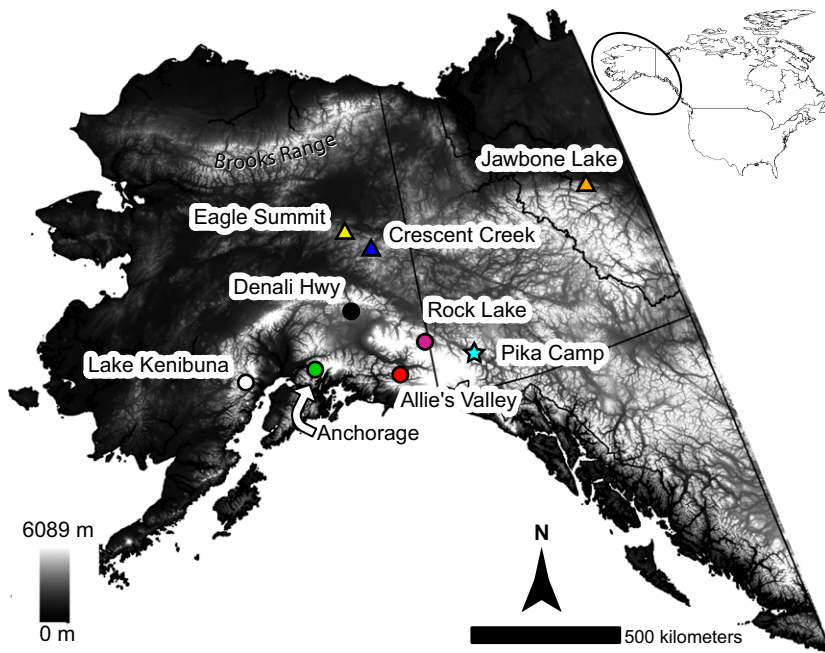


Fig. 1 Northern (triangles), southern (circles) and Pika Camp (star) populations of *Ochotona collaris* (see Table 1 for details about populations) in relation to the topography of the region.

Table 1 Details for the sampled *O. collaris* populations, including population name and abbreviation, mountain range, the number of individuals collected per population, population coordinates (latitude, longitude) and the elevation of the collection site

Population	Mountain range	Number of individuals	GIS coordinates	Elevation (m)
Allie's Valley (AV)	Chugach Mtns	8	60.9724, -143.1411	1325
Anchorage (AN)	Chugach Mtns	6	61.4510, -148.4652	1038
Crescent Creek (CC)	Yukon-Tanana Uplands	7	64.8212, -143.7509	1542
Denali Hwy (DH)	Alaska Range	6	63.0741, -145.6358	1202
Eagle Summit (ES)	Yukon-Tanana Uplands	6	65.4833, -145.4167	1111
Jawbone Lake (JL)	Mackenzie Mtns	2	65.0000, -127.6167	868
Lake Kenibuna (LK)	Alaska Range	8	61.1564, -152.8551	1000
Pika Camp (PC)	Wrangell-St. Elias Mtns	8	61.2170, -138.2670	1798
Rock Lake (RL)	Wrangell-St. Elias Mtns	7	61.7864, -141.2095	1381

phylogeographic structure. In addition, Collared Pika populations are also known to persist on nunataks (Murray & Murray 1969) and may disperse further through glacial corridors than in unglaciated areas (COSEWIC 2011), indicating that glacial corridors may also play an important role in population connectivity. If genetic diversity is being primarily shaped by contemporary environments, gene flow should be highest between proximate populations, that is those within the same mountain ranges or connected by glaciers, and lowest between populations separated by low-elevation areas with extensive tree and shrub cover.

To simultaneously and comprehensively evaluate the roles of contemporary and historical environmental factors, we conducted a restriction site-associated DNA sequencing (RADSeq) analysis of population structure. Our objectives were to: (i) delimit genetic boundaries

within the Collared Pika, (ii) test proposed biogeographic scenarios and distinguish refugial locations, and (iii) evaluate the degree to which specific contemporary factors (habitat continuity and glacial corridors) affect gene flow. This represents the first RADSeq study in any Alaskan mammal and is thus an important opportunity to evaluate the degree to which earlier mtDNA analyses capture relevant phylogeographic patterns within a species.

Methods

RAD library preparation and sequence analysis

Samples of 58 Collared Pikas from nine populations were collected, and their DNA was extracted as described previously (Lanier & Olson 2013; Table S1,

Supporting information). Localities were chosen to broadly represent the species' range (Fig. 1; Table 1). Extracted genomic DNA was individually barcoded and processed into a reduced-complexity library using a double-digestion restriction-fragment-based procedure (for details, see Peterson *et al.* 2012). Briefly, DNA was double-digested with *EcoRI* and *MseI* restriction enzymes, followed by the ligation of Illumina adaptor sequences and unique 10-base-pair barcodes. Ligation products were pooled among samples and size-selected between 350 and 450 base pairs using a Pippin Prep (Sage Science) machine. The targeted-size ligation products were amplified by iProof™ High-Fidelity DNA Polymerase (BIO-RAD) with 12 cycles. The library was sequenced in one lane on the Illumina HiSeq2000 platform according to the manufacturer's instructions to generate paired-end 100-base-pair reads. Sequences were demultiplexed, and reads with an average Phred score of at least 30 and an unambiguous barcode and restriction cut site were retained (raw genomic data available on Dryad: doi:10.5061/dryad.jh2fp). Only the first read was maintained due to the need for unlinked single nucleotide polymorphisms (SNPs) in our analyses. Reads showed consistently high sequence quality (Fig. S1, Supporting information) and remained untrimmed, except for the barcode and restriction enzyme cut site.

Single nucleotide polymorphisms were called from RADSeq loci (i.e. contigs assembled from overlapping sets of homologous fragments), and genotypes were assigned using a maximum-likelihood statistical model (Catchen *et al.* 2011; Hohenlohe *et al.* 2012) with the STACKS v1.03 pipeline (Catchen *et al.* 2013); default settings were used except where noted below. Specifically, loci and polymorphic nucleotide sites were identified in each individual using the USTACKS program, which groups reads with a minimum coverage depth ($m = 3$) into a 'stack'; increasing the minimum depth generally helps to avoid erroneously assigning sequencing errors as SNPs (Catchen *et al.* 2013). Reads were filtered using a removal algorithm that eliminated highly repetitive stacks (i.e. stacks that exceed the expected number of reads for a single locus given the average depth of coverage, for example, when loci are members of multigene families) and a 'deleveraging algorithm' to resolve overmerged loci (i.e. nonhomologous loci misidentified as a single locus). SNPs were identified at each locus and genotypes were called using a multinomial-based likelihood model that accounts for sequencing error (Hohenlohe *et al.* 2010; Catchen *et al.* 2011; Catchen *et al.* 2013), with the upper bound of the error rate (ϵ) set to 0.1. A conservative upper bound was selected for ϵ , as these models have been developed primarily for higher-coverage data; a conservative bound was preferred over the unbounded model because the latter has

been shown to underestimate heterozygotes (Catchen *et al.* 2013). A catalogue of consensus loci among individuals was constructed with the CSTACKS program from the USTACKS output files, with loci recognized as homologous across individuals if the distance between the consensus sequences (n) was ≤ 2 . Alleles were identified in each individual against this catalogue using the SSTACKS program.

Variable RADSeq loci present in all individuals and containing only two alleles were exported from the MYSQL database after excluding two individuals that were missing more than 20% of the total loci. This data set, hereafter referred to as the 'complete data set' because it contains no missing data, was used in all analyses, except where noted. Population genetics statistics, including observed heterozygosity (H_{obs}), nucleotide diversity (π) and Wright's F -statistic (F_{IS}), were calculated using the POPULATIONS program in the STACKS pipeline (Catchen *et al.* 2013). For biallelic RADSeq loci, π is a measure of expected heterozygosity and is therefore a useful measure of genetic diversity of populations. Furthermore, F_{IS} measures the reduction in observed heterozygosity as compared to expected heterozygosity for an allele in a population, where positive values indicate nonrandom mating or cryptic population structure (Nei & Kumar 2000; Hartl & Clark 2006; Holsinger & Weir 2009). Pairwise F_{ST} values were calculated among populations in ARLEQUIN 3.5 (Excoffier & Lischer 2010), and significance ($P < 0.05$) was determined using 1000 permutations.

Characterization of population genetic structure

Population genetic structure was characterized using STRUCTURE 2.3.4 (Pritchard *et al.* 2000) with K -values ranging from 1 to 11 (two more than the total number of populations). Given the low divergence exhibited in this data set and the importance of a full site-frequency spectrum for downstream uses (described below), loci were not filtered based on minimum allele frequency. We filtered out linked SNPs from the complete data set using the `-write_single_snp` option in POPULATIONS. In addition, we analysed the data set excluding the highly divergent Pika Camp population (see Results), which could potentially obscure genetic composition of the remaining populations (see Ryan *et al.* 2007; Massatti & Knowles 2014). Ten independent runs per K were conducted, each with 100 000 burn-in and 200 000 MCMC iterations, using the 'admixture model' and the 'correlated allele frequency model' with default settings. Results based on more burn-in or MCMC iterations were not different (data not shown). STRUCTURE HARVESTER (Earl & vonHoldt 2012) and DISTRICT (Rosenberg 2004) were used to visualize results, and the most probable K was chosen based on ΔK (Evanno *et al.* 2005).

To visualize the major axes of population genetic variation, a principal components analysis (PCA) was performed on the complete data set without linked SNPs (the same as in the *STRUCTURE* analysis) using the 'adegenet' package (Jombart 2008) in R (R Core Team 2012). Major axes for genomewide SNP data were identified using the R 'dudi.pca' function. PCA is free from many of the population genetics assumptions underlying *STRUCTURE* (Gao *et al.* 2007; Jombart *et al.* 2009) and can be more useful with continuous patterns of differentiation (e.g. isolation by distance – IBD; Engelhardt & Stephens 2010) than *STRUCTURE*. The first two PCA eigenvectors and the sample locations (latitude and longitude) were also used in a Procrustes analysis with the 'vegan' package (Oksanen *et al.* 2013) in R. This analysis is especially useful for quantifying the correlation between genetic variation and geography (see Wang *et al.* 2014). Specifically, the Procrustes analysis superimposes the PCA plot of genetic variation on the geographic map, rotating the PC axes to achieve maximum similarity to the geographic distribution of sampled locations (i.e. the sum of squared differences between the two data sets are minimized); significance of the association was evaluated based on 10 000 permutations.

To examine the patterns of population relationships, a maximum-likelihood phylogeny was estimated from unlinked SNPs with a general time-reversible (GTR) + Γ substitution model using PHYLIP 3.0 (Guindon *et al.* 2010). This algorithm is able to estimate phylogenies using standard IUPAC coding for polymorphic nucleotide sites, allowing us to include all unlinked RADSeq loci within our concatenated data set; a PHYLIP file was generated by the POPULATIONS program in STACKS using the `-phylip_var` setting and a white list containing the loci names included in the complete data set. Nodal support was estimated from 1000 bootstrap replicates. The Pika Camp population was used as an outgroup, based on PCA and *STRUCTURE* results.

Patterns of population isolation

A series of Mantel and partial Mantel tests were used to test for associations between genetic distance and the geographic distance separating populations, as well as associations between genetic distance and environmental resistance, where an environmental factor mediates the degree of isolation among populations. For tests of isolation by resistance (IBR; McRae 2006), a resistance matrix based on raster maps of per-cell conductance values was generated among populations using CIRCUITSCAPE v3.5.8 (Shah 2007), where cells with high conductance values represent regions of probable migration relative to cells with low conductance values. Specifically, tests

of isolation by distance (i.e. tests of the effect of geography alone on patterns of genetic differentiation) used a map with equal per-cell conductance values among sampling populations. For tests of the impact of glaciers on patterns of genetic differentiation among populations, areas covered by glaciers were coded to have either twice or half of the conductance value as cells in nonglaciated areas to test the respective hypotheses that glaciers facilitated (i.e. high conductance) or inhibited (i.e. low conductance) pika dispersal. Two resistance matrices among populations were also generated based on the ENMs for the present and the LGM to investigate the impact of current and past environmental factors on patterns of genetic differentiation, respectively, where the cell-specific habitat suitability scores were translated into cell-specific conductance values.

Estimation of demographic history

To better understand the historical demographic trends for pika populations, we estimated divergence time, gene flow and population size changes among different populations using the site-frequency spectrum (SFS). Populations were grouped into three clusters based on the PCA, Procrustes and population tree analyses to capture the main characteristics in the population histories: northern group (Jawbone Lake, Eagle Summit, Crescent Creek), southern group (Denali Hwy, Allie's Valley, Anchorage, Lake Kenibuna, Rock Lake) and Pika Camp. The data set used for this analysis was generated by exporting all RADSeq loci from the MySQL database containing two alleles (39 797 RADSeq loci). Furthermore, we selected only one SNP per RADSeq locus, and it had to be present in at least 80% of the individuals in each cluster (i.e. northern, southern, or Pika Camp); the final SFS contained 10 892 variable SNPs. The sample sizes of gene copies for the northern, southern and Pika Camp clusters are 24, 56 and 8, respectively. Because we did not include invariable sites in the SFS, we fixed the parameter for ancestral pika population size (N_{anc} ; Fig. 2) to be 114 000 to enable the estimation of other parameters in FASTSIMCOAL2 (see Excoffier *et al.* 2013; see Table S2, Supporting information for simulation details). The population tree recovered a shared common ancestor between the northern and southern groups relative to Pika Camp. Population sizes were allowed to vary, and gene flow was permitted between the three population lineages during recent expansions (Fig. 2). Fifty independent parameter estimations were performed in FASTSIMCOAL2 to achieve the maximum composite likelihood of the joint SFS, in which parameterized simulations were sampled from priors and the parameter estimation was optimized through 40 cycles of a conditional maximization algorithm (ECM). Divergence time

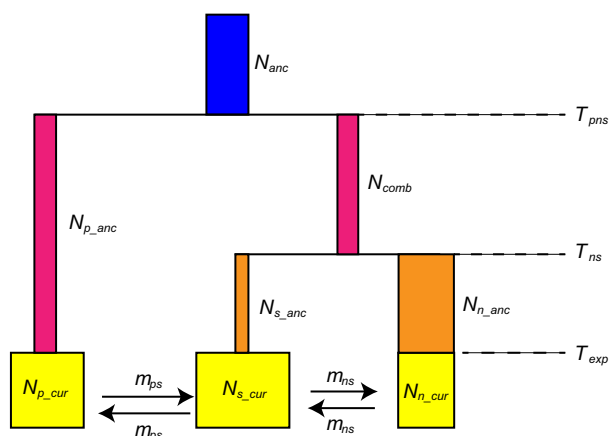


Fig. 2 Hypothesized demographic history of pika populations used in *FASTSIMCOAL2* analyses. Pika ancestors diverged (T_{pns} generations ago) into ancestral populations of Pika Camp (N_{p_anc}) and the other populations (N_{ns}). Later, the divergence into southern (N_{s_anc}) and northern refugia (N_{n_anc}) occurred, and populations experienced recent expansions and exchanged migrants. The estimates of these parameters are listed in Table 4.

estimates depend on the genomic mutation rate estimate, which is also associated with some uncertainty. The point estimates were selected from the run with the highest maximum composite likelihood. Confidence intervals of parameter estimates were obtained by 100 parametric bootstrapping runs from simulated SFS of the point estimates (see doi:10.5061/dryad.jh2fp in Dryad for specific settings).

Environmental niche modelling

Environmental niche models (ENMs) of *O. collaris* were generated from bioclimatic variables for the present and the Last Glacial Maximum (LGM; 21 kya) with *MAXENT* v3.3.3e (Phillips *et al.* 2006) using the following parameters: regularization multiplier = 1; max number of background points = 10 000; replicates = 50; and replicated run type = cross-validate. Georeferenced distribution points (161 distinct localities) from vetted occurrence data used in the modelling were representative of the entire range of *O. collaris* (Dryad doi:10.5061/dryad.jh2fp).

We used 19 bioclimatically informative variables to model present-day distributions (*WORLDCIM* v1.4; Hijmans *et al.* 2005) and LGM distributions (*PMIP2-CCSM*; Braconnot *et al.* 2007). To avoid overfitting of the distribution models, the geographic extent of the environmental layers was reduced to an area approximately 20% larger than the known distribution of the species (Anderson & Raza 2010). To guard against the inherent difficulties involved in extrapolating distribu-

tions into novel climates (reviewed in Alvarado-Serrano & Knowles 2013), an iterative approach was used to generate ENMs for the LGM. Specifically, multivariate environmental similarity surfaces (MESS maps) were used to identify bioclimatic variables that result in areas of low reliability due to predicted values that are outside of the range of present-day environmental values (Elith *et al.* 2010). *MAXENT* was rerun excluding these out-of-range variables, and this process of analysis with MESS maps was repeated until no LGM variables were out of range compared to present-day bioclimatic variables. Because MESS maps do not indicate changes in the correlations among the environmental variables used for LGM reconstructions (Elith *et al.* 2010), we checked our ENM for the LGM using only the most informative variable (Bio10) to ensure that we were not reporting errant distributional patterns. In addition, a present-day ENM was generated using the subset of variables that were not out of range during the LGM and compared to a ENM constructed using all climatic variables with >5% importance (determined by jackknifing) to assess their similarity.

Because underlying genetic similarity may be driven by habitat similarity, we examined environmental gradients in the 19 present-day bioclimatic variables. These variables were reduced by PCA to 19 orthogonal axes using the 'Principal Components' tool in the Spatial Analyst package of *ARCGIS* 10.3 (ESRI), to examine gradients that might influence patterns of population connectedness and/or genetic differentiation. The PC values were extracted for each *O. collaris* sampling point used in *MAXENT* with the 'Extract Multi Values to Points' tool in the Spatial Analyst package. Informative PC axes were graphed using R (R Core Team 2012) and contrasted with the relative configuration of sampled populations in environmental and genetic space.

Results

Sequence data quality and processing

More than 169 million reads were produced from one lane of Illumina sequencing for 58 *O. collaris* individuals (average of $3\,035\,160 \pm 1\,271\,412$ per individual; Table 2 and Fig. S2, Supporting information). After excluding two individuals containing prohibitively high missing data (>20% of the loci; raw, filtered and analysed reads shown in Table S1, Supporting information), we identified 513 443 RADSeq loci containing 89 167 SNPs. Subsequently, we identified 4156 RADSeq loci containing at least one biallelic SNP and that were present across all individuals (in contrast to other RADSeq data sets; see Wagner *et al.* 2013; Massatti & Knowles

Table 2 Population summary statistics calculated for the 4156 RADSeq loci. Included are the average number of individuals genotyped at each locus (N), the proportion of SNPs unique to each population (% private), the percentage of SNPs (% polymorphic) in each population, the average nucleotide diversity (π), the average observed heterozygosity per locus (H_{obs}) and the Wright's inbreeding coefficient (F_{IS}). The total number of DNA sites (polymorphic + invariable) in the RADSeq loci is 375 480

	N	% private	% polymorphic	π	H_{obs}	F_{IS}
Allie's Valley	8	8.8	0.26	0.00073	0.00072	0.00004
Anchorage	6	6.1	0.15	0.00048	0.00043	0.00013
Crescent Creek	7	11.9	0.29	0.00087	0.00090	-0.00006
Denali Highway	6	7.1	0.21	0.00073	0.00067	0.00012
Eagle Summit	6	4.3	0.15	0.00059	0.00064	-0.00012
Jawbone Lake	2	3.4	0.10	0.00055	0.00055	0.00000
Lake Kenibuna	8	5.6	0.10	0.00032	0.00037	-0.00010
Pika Camp	6	12.6	0.21	0.00061	0.00064	-0.00005
Rock Lake	7	8.0	0.33	0.00104	0.00105	0.00004

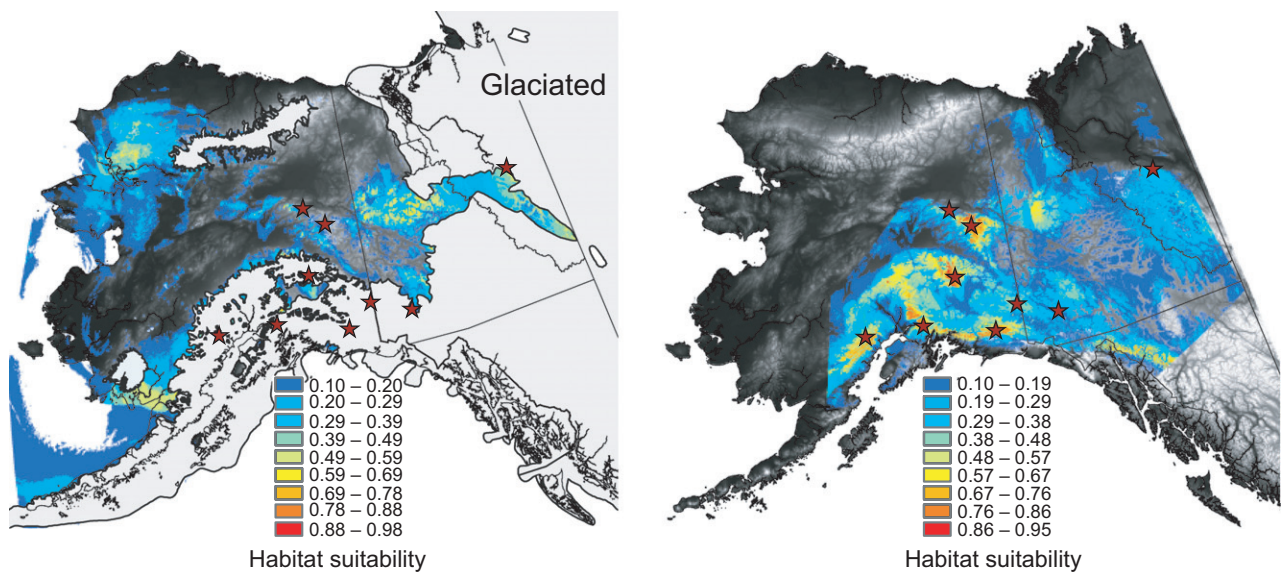


Fig. 3 Habitat suitability predicted by the ENMs for the LGM (on the left) and the present (on the right) relative to sampled populations (stars). The glacial extent at the height of the Wisconsin glaciation is shown in light blue, and the colour scale represents habitat suitability. Note that the glacial reconstruction is based on independent geologic information from glacial moraines.

2014); <4% of RADSeq loci (156) contained more than one SNP.

Ecological niche modelling and climatic characterizations of population localities

Both present-day and LGM ENMs were based on seven bioclimatic variables; this represents a reduced set from the original 19 variables to remove variables that were out of range during the LGM. The specific variables used in the modelling were temperature seasonality (Bio4), annual temperature range (Bio7), mean temperature of the warmest quarter (Bio10), precipitation of the driest month (Bio14), precipitation seasonality (Bio15), precipitation of the wettest quarter (Bio16) and

precipitation of the driest quarter (Bio17) (Fig. S3, Supporting information).

Based on predicted geographic areas of past habitat suitability (i.e. the ENM for the LGM), a potential refugium in northwestern Alaska, in addition to southwestern and eastern refugia, is suggested (Fig. 3). A comparison of ENMs from the LGM and present suggests that some populations have experienced significant distributional shifts over time, especially when LGM glaciers are overlaid on the ENMs (note that the reconstruction of glacial distributions is based on independent geologic data – e.g. glacial moraines). For example, more southern and coastal populations were presumably displaced from their current distributions by glaciers, although the geographic locality of potential

refugial populations remains unclear. Populations in the interior (Crescent Creek, Eagle Summit and Jawbone Lake; Fig. 1) also likely experienced distributional shifts; however, such shifts appear less pronounced, as would be predicted given that this region remained relatively unglaciated. Suitable habitat is projected both along the margins of the glaciers and in more ice-free regions in the interior (Figs 3 and S4, Supporting information).

The sampled Collared Pika populations are geographically and environmentally representative of the entire species – they span the spectrum characterized by the PCA of environmental characteristics as well as the latitudinal and longitudinal extent for all pika populations throughout their range (Fig. S5, Supporting information). They are geographically organized along PC1, which is strongly correlated with temperature seasonality (Bio4), with populations farther from the coast occupying areas of greater temperature seasonality. In contrast to PC1, a population's distance from the coast has less influence on its location along PC2, which is strongly positively correlated with annual precipitation (Bio12) and moderately positively correlated with precipitation of the wettest quarter (Bio16). Variation along this axis is less consistent compared to the coastal-to-interior gradient because some population localities near the coast have less annual precipitation than others (Fig. S5, Supporting information).

Genetic diversity within and among populations

Populations are quite comparable with respect to generally low within-population genetic diversity as measured by π (0.00032 to 0.00104, calculated across all DNA sites

in every RADSeq locus comprising the complete data set) as well as other common measures (e.g. observed heterozygosities; Table 2). Although differentiation among populations was significant, with F_{ST} values ranging from 0.23 to 0.84 (Table S3, Supporting information), no significant pattern of isolation by distance was detected ($r = 0.4433$, P -value = 0.116; Fig. 4). However, examination of the isolation-by-distance pattern (i.e. the association of geographic distance and F_{ST} between population pairs; Fig. 4) shows much higher F_{ST} values between Pika Camp and other populations. If Pika Camp is excluded, a significant pattern of isolation by distance is observed among the remaining populations ($r = 0.7278$; P -value = <0.001; Table 3). Likewise, distances from the isolation-by-resistance analyses are also significant if the Pika Camp population is excluded (Table 3). However, any significant signal of environmental factors mediating population isolation, as captured by resistance matrices informed by the current environment or a model of glacial-facilitated dispersal, becomes insignificant after controlling for the geographic distance separating populations (Table 3). Log transformations of the geographic or genetic distances had no impact on the significance of the results.

Population structure

STRUCTURE analyses of 4000 unlinked SNPs from the complete data set identified $K = 2$ (the K with the highest ΔK when considering different values of K , hereafter referred to as the most probable K) as the most probable number of genetic groups (Fig. 5). This arrangement reflects the pronounced differentiation of the Pika Camp

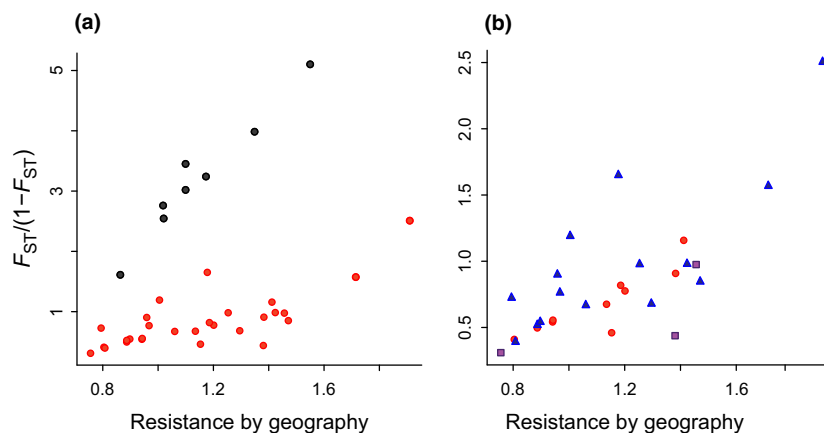


Fig. 4 Association between pairwise genetic differentiation among populations and the geographic distance between populations (i.e. isolation by distance, IBD) rescaled according to habitats suitability for (a) all the populations, with comparisons to Pika Camp population shown in black, and (b) excluding the Pika Camp population. For analyses without Pika Camp, the squares denote comparisons among the just the three most northern populations (i.e. Jawbone Lake, Eagle Summit and Crescent Creek), comparisons between these northern populations and the other populations (triangles), and comparisons among the more southern populations (circles). Note a pattern of IBD is not significant (see a; $r = 0.4433$; P -value = 0.116) unless the Pika camp population is excluded (see b; $r = 0.7278$; P -value = <0.001).

Table 3 Results of isolation-by-resistance as calculated using Mantel and partial Mantel tests (with geography and the glacier or ENM as covariates) between the pairwise F_{ST} values with resistance matrices (i.e. rescaled geographic distances according to the suitability of habitats) separating populations (see also Fig. 7). Four resistance matrices are tested: the first is calculated from geography only, the second from glacier influence, the third from current habitat suitability score and the last from the past suitability. Correlation coefficients (r) and the P -values from 1000 permutation tests are shown, with significant results shown in bold. In partial Mantel tests, geography is treated as a covariate with other testing predictors. Note that these results are based on comparisons without Pika Camp

	Mantel test		Partial Mantel test	
	r	P	r	P
Geography	0.7278	<0.001		
Glacier (easier)	0.7103	0.001	0.2433	0.24
Glacier (difficult)	0.6002	0.013	-0.1561	0.684
ENM – current	0.7492	0.002	0.2735	0.199
ENM – 21K	0.272	0.281	0.3942	0.143

population from the other populations, which is also evident in the PCA (Fig. 6). Many of the other sampled populations are genetically distinct, with differences in allele frequencies that separate the populations in both the PCA and STRUCTURE analyses. Specifically, at $K = 4$ (the second-most probable number of genetic clusters when all populations are analysed together) or $K = 3$ (the most probable number of genetic clusters when the divergent Pika Camp population is excluded; Fig. S6, Supporting information), the different genetic clusters are composed of populations with little-to-no admixture, with the notable exceptions of the Rock Lake, Denali Hwy and Allie's Valley populations (Fig. 5).

In general, the co-ancestry of populations is not strictly predicted by their co-occurrence on individual mountain ranges (Fig. 5). Populations within a mountain range may be more similar genetically to populations from different mountain ranges than to other populations within the same range. For example, the Rock Lake population is genetically more similar in composition to the Allie's Valley population than it is to the Pika Camp population (Fig. 5). Likewise, the Lake Kenibuna and Anchorage populations are quite similar genetically, even though they occur in different mountain ranges (Figs 1 and 5).

Despite the lack of clear partitioning of genetic variation by mountain ranges, there is evidence of regional structuring. This is apparent in the three groupings recovered in the PCA: (i) Pika Camp; (ii) a northern grouping of Crescent Creek, Eagle Summit and Jawbone Lake; and (iii) the rest of the coastal and interior popu-

lations (Lake Kenibuna, Anchorage, Denali Hwy, Rock Lake and Allie's Valley). Patterns of population relatedness suggest a similar regional structuring (Fig. S7, Supporting information). The northern clade composed of Eagle Summit, Crescent Creek and Jawbone Lake is sister to the remaining southern populations, although Rock Lake is not a part of either regional group identified from the concatenated SNPs (Fig. S7, Supporting information). Rock Lake's sister relationship to the northern and southern clades may be attributable to the inclusion of two individuals with a purportedly mixed ancestry between Pika Camp and the coastal/interior lineage (Fig. 6).

The Procrustes analysis identified a significant similarity score between the populations in PC space and their actual geographic locations ($t_0 = 0.62$; $P < 1.0 \times 10^{-5}$). Examination of the rotated genetic coordinates graphically shows that the sampled populations are distributed over a narrower longitudinal extent than expected based on the geographic location of populations (Fig. 7). For example, both the Jawbone Lake population to the east and the Lake Kenibuna population to the west are much more closely related to more centrally located populations than expected. In contrast, the Pika Camp population is more divergent genetically than would be predicted by geography alone (i.e. the population occupies a more distant area of PC space relative to the other populations). Other populations display a fairly tight association between the rotated genetic coordinates and their geographic location, including the Eagle Summit, Crescent Creek and Denali Hwy populations (i.e. lines connecting the individuals to their geographic location are fairly short). The superimposition of genetic coordinates on geography also shows that the more coastal populations – specifically, Lake Kenibuna and Anchorage to the west and Allie's Valley and Rock Lake populations to the east – are fairly tightly clustered, and positioned in a space corresponding to a more central longitudinal location.

Estimates of historical demography

Analyses based on the site-frequency spectrum in FASTSIMCOAL2 recovered a deep division between Pika Camp and populations from the other two lineages, occurring around 148 kya (95% CI: 126–323 kya; Table 4). This divergence estimate was 2–5 times earlier than the split between northern and southern populations, estimated at 52 kya (39–74 kya), with all three divergences predating the LGM (Table 4; Fig. 2). Demographic results also suggest that northern populations have stayed fairly constant in size or decreased slightly over time, with relatively broad overlaps between effective population sizes estimated at the LGM and in the recent. In con-

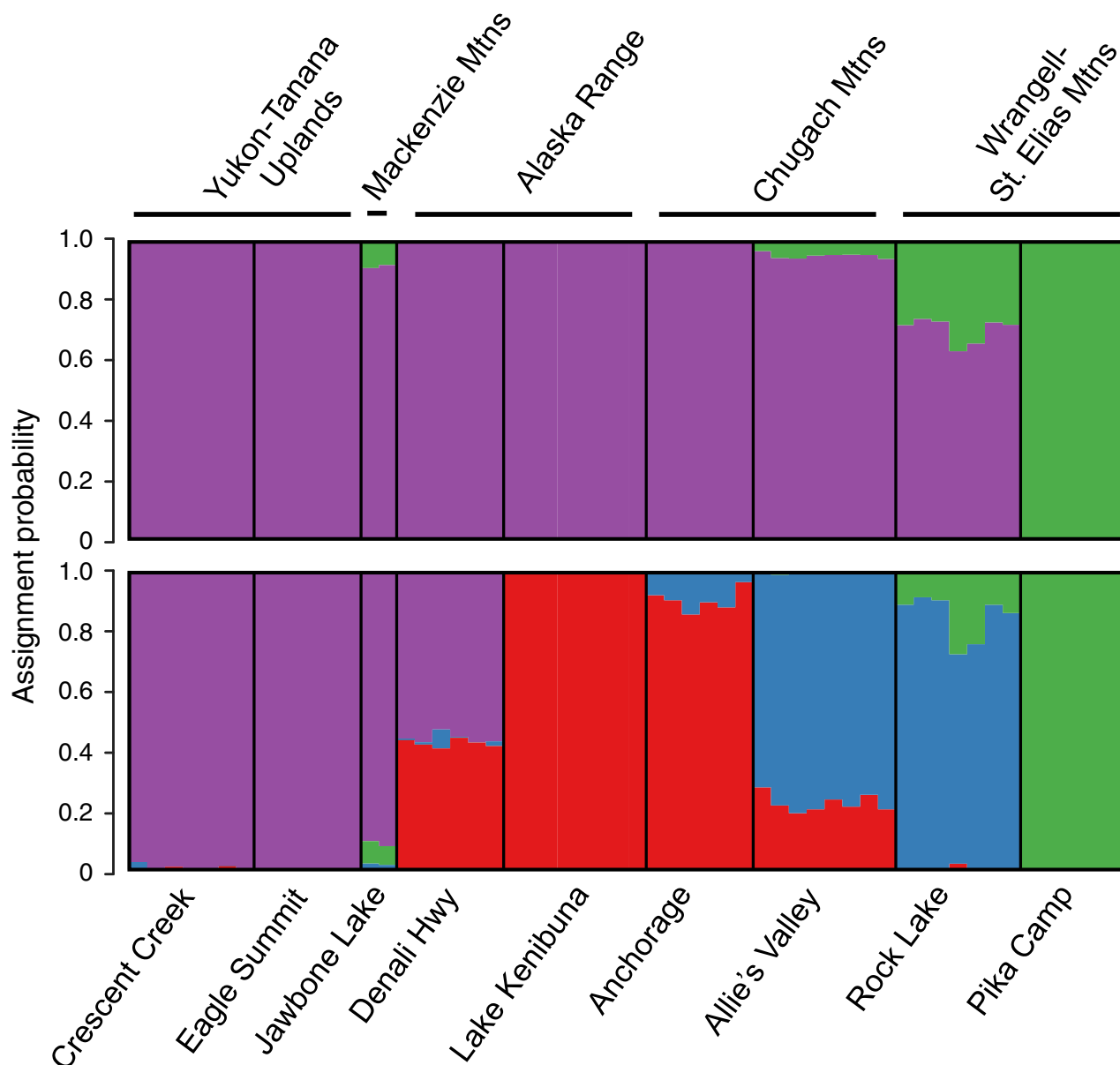


Fig. 5 Plots of posterior probabilities for individuals assigned to K genetic clusters from STRUCTURE analyses for $K = 2$ and $K = 4$ (each separate block corresponds to one analysis). Each of the K genetic clusters within a given analysis is shown as a different colour delimited by a thin black line, with the corresponding population names listed along the bottom of the plot. The regional membership of populations (i.e. mountain ranges) is listed along the topmost analysis of $K = 2$.

trast, southern populations and Pika Camp appear to have experienced a dramatic (10–40 \times) population expansion starting around 36 kya (22–46 kya). Gene flow estimates were low; the inferred migration rate per generation between the southern populations and Pika Camp (2.55E-06) was nearly twice that between the northern and southern populations (1.44E-06). Estimates of current population sizes were larger in southern populations than in the northern populations, although

with a broad 95% CI on Pika Camp due to the small sample size.

Discussion

Our results indicate that contemporary factors alone (i.e. current habitat continuity and glacial corridors) are not sufficient to explain connectivity among populations of Collared Pikas across their range. Instead, the results

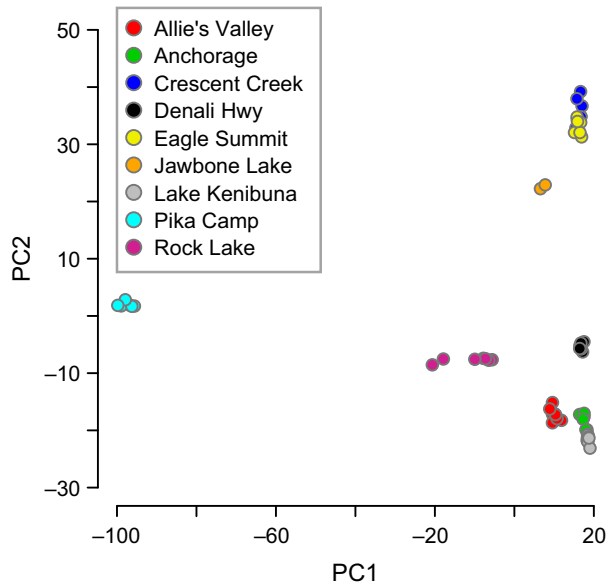


Fig. 6 Distribution of individuals along PC1 and PC2 axes of genetic variation based on the analysis of 4000 unlinked SNPs; individuals are colour-coded according to their population identities. PC1 and PC2 explained 15.4% and 5.6% of the variation among individuals, respectively.

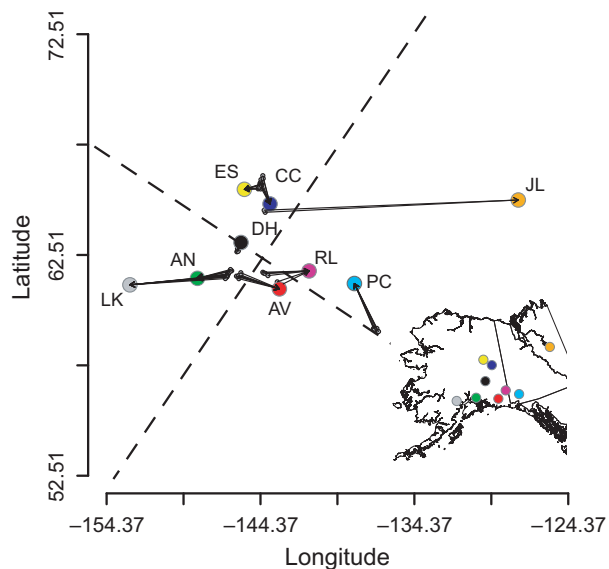


Fig. 7 Procrustes-transformed PCA plot of genetic variation with each individual mapped in PC space (the small black circles) relative to the geographic location of populations (shown by the larger, coloured circles). Dashed lines show the orientation of PC1 (oriented NW-SE) and PC2 (oriented NE-SW) for the genetic data (explaining 15.4% and 5.6% of the genetic variation, respectively) relative to the geographic longitude and latitudinal axes (the x and y axes). The length of the line connecting individuals in PC space to their geographic location represents the extent of the deviation from the expected pattern of genetic variation based on geography. Population abbreviations are shown in Table 1.

Table 4 Inferred parameters and their confidence intervals (95%) under the demographic model illustrated in Fig. 2. Note that estimates of time are shown in units of generations, with a generation time of 2 years (Table S2, Supporting information)

Parameters	Point estimates	Lower bound	Upper bound
N_{s_cur}	298 059	227 406	462 018
N_{n_cur}	167 623	117 761	278 937
N_{p_cur}	736 419	528 636	1 105 579
N_{s_anc}	28 425	36 884	62 470
N_{n_anc}	285 409	139 884	508 318
N_{p_anc}	18 896	16 476	37 899
N_{comb}	45 186	41 898	112 117
M_{ns}	1.44E-06	6.278E-07	1.92E-06
M_{ps}	2.55E-06	1.89E-06	4.07E-06
T_{exp}	18 276	11 178	23 114
T_{ns}	26 674	19 581	37 285
T_{pns}	72 822	62 950	161 261

provide strong support for the predominance of three divergent lineages, likely separated in different Pleistocene refugia, with population expansion among lineages predating the Last Glacial Maximum. These three separate lineages (one northern and two southern) correspond to the major genetic boundaries observed within contemporary populations. The populations in our study are similar with respect to generally low genomic variation within populations and little-to-no contemporary gene flow between populations. These results support a growing biogeographic paradigm of Pleistocene substructure in eastern Beringia (Shafer *et al.* 2010; Lanier *et al.* 2015), with implications for population management as well as future studies of adaptation and climate change.

Pleistocene fragmentation and genetic structure

Genomic data strongly suggest that Collared Pika populations originated from three divergent ancestral sources, representing northern, eastern and coastal/central (hereafter southern) refugia, which were sundered well before the LGM. These data, as well as the projected habitat at the LGM, suggest multiple potential allopatric refugia as well as several important patterns. First and foremost, the genetic distinctiveness of the southeast Pika Camp population most likely represents an isolated ancestral source, which shared ancestry with other Collared Pika populations during the last interglacial (~148 kya [126–323 kya]; Table 4). Second, the population tree, PCA and STRUCTURE analyses all suggest that the remaining populations were colonized from northern and southern lineages that diverged well before the LGM (~52 kya [39–75 kya]; interstadial marine isotope stage 3). Third, the southern and Pika Camp

lineages have undergone population expansions, whereas the northern lineage has remained relatively stable over the same period (Table 4). This work provides additional evidence that the pattern of southern stability and northern expansion described for many taxa (Hewitt 2004) may be insufficient to describe localized range expansions at higher latitudes.

Several lines of evidence indicate that colonization from a single ancestral source in the recent past is unlikely. For example, the lack of a significant pattern of isolation-by-distance (Fig. 4), unless the Pika Camp population is excluded, and the absence of a latitudinal pattern of genetic diversity (levels of diversity are fairly comparable across populations; Table 2) would be expected under a single refugium hypothesis but are not supported by the data. Furthermore, the most likely number of genetic clusters from the STRUCTURE analyses does not correspond to the number of sampled populations (Fig. 5), suggesting that population-specific demographic events, such as strong bottlenecks associated with population colonization, do not dominate or override the signal associated with regional differentiation. If the signal of shared ancestry was completely overridden by demographic factors (e.g. genetic drift) associated with colonization, $K = 9$ would be the most probable STRUCTURE result. Additionally, demographic simulations suggest smaller, isolated source populations with subsequent population expansions, not the recent separation and population contraction that would be expected if a single large population was fragmented during post-Pleistocene warming.

These results also highlight the important roles of habitat stability and turnover during the Pleistocene. Based on the ENMs (Fig. 3), habitat in the northern periphery of the Collared Pika's range (including Eagle Summit, Crescent Creek and Jawbone Lake populations) appears to have remained relatively constant, whereas southern and Pika Camp populations underwent broad expansion and commensurate population growth. While the prediction of northern stability differs from previous ENM reconstructions (Lanier & Olson 2013), it is consistent with genetic data based on the level of differentiation relative to the other populations (Fig. 7). Given this support and the fact that the ENMs presented herein exclude out-of-range climate variables – a preferred approach for extrapolating distributions into novel climates (Alvarado-Serrano & Knowles 2013) – these reconstructions are likely a more accurate representation of historical distributions. Furthermore, the ancestral and current population estimates (Table 4) provide additional support for long-term stability of northern populations, with demographic estimates that remain constant (or decline slightly) from the LGM to the present. High philopatry in Collared Pikas (e.g.

Franken 2002), as well as inferred boundaries in the present created by unsuitable habitat (e.g. Fig. 3), indicates that the pattern of shared ancestry among distant northern populations is not the result of ongoing gene flow, but instead is a signature of shared ancestry. Any past connectivity among these populations is likely also being eroded due to recent and ongoing tree and shrub encroachment into arctic and subarctic regions (Sturm *et al.* 2001).

Geographic locality of refugial source populations

Collectively, these results suggest relative, but not exact, locations for three refugial populations. Northern populations of Crescent Creek, Eagle Summit and Jawbone Lake probably originated from a geographically proximate source as opposed to a distant northern refuge in the Brooks Range. Despite apparently suitable habitat in the western Brooks Range during the LGM (Fig. 3; Lanier & Olson 2013 – Fig. 6), the lack of present-day populations in this region (despite focused survey efforts by LEO and colleagues) and the relationship between geography and genetic differentiation (Fig. 7) suggest that this was an unlikely ancestral source for any of the populations in our study. A proximate refugium is also indicated by the fact that Eagle Summit and Crescent Creek populations are only slightly more divergent than expected based on geography alone, and Jawbone Lake is much less divergent than would be expected based on geographic distance (Fig. 7). This suggests that the refugial source for the northern populations may have been closer to Eagle Summit and Crescent Creek than the current Jawbone Lake site. Pleistocene fossil pikas have been reported from northeast and southwest of Eagle Summit/Crescent Creek sites (Guthrie 1973; Harrington 2011), areas with higher predicted habitat suitability at the LGM (Fig. 3). Estimated population sizes for the northern lineage have remained fairly constant (and generally high) since ~52 kya, as would be expected if the lineage underwent little Pleistocene movement. Further tests of population movements (e.g. He *et al.* 2013) may help distinguish probable refugial locations and colonization routes for these regions.

While an eastern refugium is clearly indicated by our data, its exact location is difficult to determine from our data (see also Lanier & Olson 2013). Pika Camp individuals are genetically distinct from geographically proximate populations (Fig. 5), but with only one location represented in our data set, there are no comparative Procrustes deformations for this lineage. Reconstructions of glacial margins indicate that this site (Fig. 3) was probably ice free during the LGM, although much of the surrounding habitat was not; thus, populations

may have persisted in situ and expanded to fill available habitat as it became available. An alternative possibility is suggested by the sub-Laurentide origin of Collared Pikas, which diverged from American Pikas during the Pliocene (Lanier & Olson 2009; Galbreath & Hoberg 2012). Simulations indicate that the Pika Camp lineage diverged from the other two lineages around 148 kya (marine isotope stage 6), predating both the Wisconsin glaciation and the last interglacial. Based on the timing, a population of Collared Pikas could have been isolated within or south of the ice sheet and undergone northward migration during an interstadial. However, estimates suggest population expansion around 36 kya (22–46 kya) (Table 4), well before the most recent opening of the ice-free corridor between the Laurentide and Cordilleran ice sheets at 15 kya (Dixon 2013). In either model, eastern populations may have occupied nunatak refugia or ice sheets margins, as has been suggested for Mountain Goats (Nagorsen & Keddie 2000; Shafer *et al.* 2011) and alpine plants (Marr *et al.* 2008), but multiple nunatak microrefugia are not suggested by our data. As Collared Pika populations from the Coast Mountains belong to the same mtDNA haplogroup as those in Pika Camp, additional fresh tissues from these sites could provide further resolution on refugial locations for the lineage.

The joint consideration of the geographic distribution of populations and the patterns of genetic variation among populations provides an opportunity to hypothesize about the potential location of a southern refugium from which the remaining coastal/central populations were colonized. During the LGM, ice sheets covered most of the current population locations, leaving diffusely distributed suitable habitat. Because the ENMs do not identify a particular circumscribed area with high suitability values, the geographic location of a southern refugium is difficult to predict from these results (Fig. 3). However, the Procrustes shifts suggest that populations from this lineage likely had a central location (Fig. 7). Lake Kenibuna, Anchorage and Allie's Valley are genetically closer to one another than expected based on geography alone (Fig. 7). Both the Lake Kenibuna and Anchorage populations exhibit genetic patterns consistent with an ancestral population at the Denali Highway longitude, as do the more eastern Allie's Valley and Rock Lake populations. Results also indicate that long-term survival in nunataks is unlikely, as localized nunatak refugia would result in a stronger correlation of geography and populations in PC space. Although we cannot rule out the possibility of two central refugia, where only one contributed to Rock Lake and Allie's Valley (see Fig. 5), it is also possible that population differentiation is indicative of environmental differences (i.e. Rock Lake and Allie's

Valley are differentiated from the other coastal populations along PC2; Fig. S5, Supporting information).

Relevance of this work to eastern Beringian phylogeography

As the first genomic study to evaluate phylogeographic patterns in an Alaskan mammal, our results are particularly revealing with regard to the degree of nuance that such data sets can provide to existing regional studies. Previous research employing mitochondrial DNA (Lanier & Olson 2013) identified two haplogroups in Collared Pikas: an Eastern haplogroup occurring in populations from Pika Camp, Rock Lake and the surrounding populations; and a Beringian haplogroup, which encompasses populations from the rest of the range (from Lake Kenibuna to Allie's Valley, north through Eagle Summit and Jawbone Lake). The geographic location of this suture zone differed slightly based on marker type, with some Allie's Valley individuals carrying mtDNA haplotypes from the Eastern lineage (Lanier & Olson 2013) on a mostly Beringian genomic background (Fig. 5). Given that dispersal is not sex-biased in Collared Pikas (Zgurski & Hik 2012), this may indicate a moving contact zone among lineages (Toews & Brelsford 2012). This further underscores the importance of combining genomic data sets with analytical approaches that can harness their power in revealing phylogeographic patterns across species.

Our results also add to a growing body of regional phylogeographic work that reveals regional complexity and shared geographic barriers within the Beringian refugium. For example, the deep split between Pika Camp and the other lineages mirrors the Beringian–Eastern (B–E) split in mtDNA evident across multiple species of codistributed mammals, including Singing Voles (Weksler *et al.* 2010), Nearctic Brown Lemmings (Fedorov *et al.* 2003) and Hoary Marmots (Lanier *et al.* 2015). It has been suggested that B–E divisions in these species could be the result of isolation into two refugia (Fedorov *et al.* 2003; Weksler *et al.* 2010), one located in central Beringia, and one located either along the northern margin between ice sheets or south of the ice sheets (Lanier *et al.* 2015), possibly corresponding to the northern British Columbia hot spot identified in previous work (Marr *et al.* 2008; Shafer *et al.* 2011). The third divergence between northern and southern Collared Pika lineages revealed in genomic data, in addition to the deeper divergence between Pika Camp and the other populations, provides further evidence of regional substructure. These results are similar to the roughly sympatric Arctic Ground Squirrel, which includes four distinct lineages (Galbreath *et al.* 2011), three of which appear to

have originated in central or eastern Beringia. Although previous studies have suggested that the coastal ice shield and Yukon River Delta may underlie these breaks (see review in Shafer *et al.* 2010), neither of these putative barriers appears to correspond with observed diversity in Collared Pikas, suggesting that ecological barriers may be driving the observed patterns. Likewise, the estimated timing of these divergences between lineages (148 and 52 kya) and the inferred population expansion indicates that the glacial contraction/postglacial expansion paradigm (Hewitt 2004) may be overly simplistic for Beringian taxa. Instead, regional diversity appears to be shaped by older historical fragmentation, similar to Arctic Ground Squirrels (Galbreath *et al.* 2011) and Singing Voles (Weksler *et al.* 2010), patterns that would have been impossible to discern from mtDNA alone.

Genomic variation and gene flow

Across the range of Collared Pikas, genomic data reveal significant differentiation among populations, but generally low variation within populations (Table 2; Fig. 6). This is likely the result of habitat specialization and limited dispersal among alpine-associated populations, which will decrease genomic similarities (e.g. Ferchaud *et al.* 2010). Both North American pika species are notably poor dispersers (most individuals disperse <1 km and only a few exceed 2 km; Peacock 1997; Zgurski & Hik 2014). Despite this, geographic proximity does not predict whether populations were colonized from the same or different refugial source. For example, the most geographically isolated population, Jawbone Lake, shares common ancestry with two other northern populations, as well as a small amount of ancestry with Pika Camp to the south (Figs 5 and 6), whereas Pika Camp is genetically distinct from all other populations. Marginal populations (including Jawbone Lake) also contained the lowest levels of polymorphism, but genetic variability was low in all populations. Genomic diversity was slightly, but not greatly, higher in populations exhibiting shared ancestry among historically separate lineages (e.g. Rock Lake) (Table 2).

Although gene flow is generally low in pikas (Smith & Ivins 1983; Zgurski & Hik 2012), STRUCTURE results do suggest that some populations (in particular Rock Lake, Allie's Valley and the Denali Highway) may have experienced colonization from multiple ancestral sources (Fig. 5). Simulations also recovered very low levels of historical migration between the southern populations and Pika Camp (Table 4); however, recent or ongoing migration is not indicated for most populations. As Collared Pikas have generally low population sizes (~1.7 animals/ha; COSEWIC 2011) and live in naturally fragmented habitats, it is likely that drift quickly fixes

migrant alleles within a population, so multiple lines of genomic evidence provide important support of past connectivity among sites.

Importance of environmental factors

Although there appears to be some a correspondence between regional environmental differences and population relatedness, this relationship may be more indicative of geographic proximity and common ancestry than environmental gradients. Although the northern populations in the interior cluster together in both environmental and genetic space, the environmental variables included in our models were not good predictors of population relationships overall. For example, the populations originating from the southern refugium (i.e. Denali Highway, Lake Kenibuna, Anchorage and Allie's Valley) share common ancestry but are characterized by fairly different climates (Fig. S5, Supporting information). Likewise, Rock Lake, which shares alleles with both Allie's Valley and Pika Camp, is geographically, but not climatically, intermediate (Figs 1 and S5, Supporting information). Thus, the environmental and genetic similarities among northern populations may be driven by shared history across a similar habitat rather than contemporary environmental gradients.

The predominant ancestral signal in the IBD and IBR analyses (Fig. 4; Table 3) also underscores the insignificant role of environmental variables at this scale. Under any IBD model, we would expect geographically marginal populations, such as Jawbone Lake, to be much more dissimilar to the other populations, and populations that are geographically proximate, such as Pika Camp and Rock Lake, to be much more similar, yet neither holds true (Figs 6 and 7). After excluding Pika Camp, the resulting significant pattern of IBD generally corresponds to a comparison of populations between northern and southern area (i.e. colonization from two ancestral sources). Further support comes from the isolation-by-resistance analyses, which are nonsignificant with respect to contemporary environmental parameters, environments during the LGM, and glacially mediated dispersal (as suggested by Lanier & Olson 2013; Table 3) when geography is included as a covariate.

This is not to say that dispersal distance and environment do not matter; rather, they do not appear to impact rangewide patterns; further and finer-scale analyses are needed to identify and quantify the importance of these variables locally. For example, fine-scale analyses in Collared Pikas have identified significant IBD patterns at the local (<2 km) scale (Zgurski & Hik 2014) and restricted dispersal along altitudinal transects (Henry *et al.* 2012). Landscape genetics analyses of closely related American Pikas (*O. princeps*) revealed that

dispersal is limited by landscape features such as topographic relief (e.g. cliffs), water and, to a lesser extent, west-facing slopes (Castillo *et al.* 2014). Whether these features represent important barriers to dispersal in Collared Pikas at the local scale remains unknown.

Relevance to adaptation and climate change

One of the most pressing and important aspects this work suggests is the incorporation of population history into studies of the effects of ongoing climate change. As temperature-sensitive alpine specialists, pikas may be particularly sensitive to climate change (Smith *et al.* 2004). These threats may be amplified in Collared Pikas, which have been shown to harbour less genetic variation and possibly less adaptive potential than their more southerly congeners (Lanier & Olson 2013). From a management perspective, identifying discrete historical genetic breaks that may lead to significant genetic differentiation among populations (i.e. due to different refugial origins), and possibly differential adaptive potential, is important. Recent species status updates in Canada classified Collared Pikas as a species of 'special concern' but noted that no designatable units had been identified within the species (COSEWIC 2011). Our results indicate three different, historically derived genetic lineages that should be considered in future management strategies and ecological studies. In particular, targeting future studies at populations within each of these distinct lineages will provide greater resolution of environmental interactions as well as a stronger broad-scale understanding of how the species as a whole is responding to changing climates.

Multiple documented extirpations of American Pikas in the Great Basin underscore the concern that climate change may force one or both North American species to extinction (Beever *et al.* 2003; Smith *et al.* 2004). However, the clearly divergent and disparate histories exhibited by the extant populations sampled here suggest that we cannot simply assume populations represent interchangeable units that will respond in the same fashion to ongoing change (Palumbi *et al.* 2014). In the case of Collared Pikas, nearly all longitudinal research (e.g. Franken & Hik 2004; Morrison & Hik 2007; Zgurski & Hik 2014) has been carried out at Pika Camp (Fig. 1), which is the most divergent and perhaps least representative site with respect to the allelic variation observed across the species' range. While we may be able to use studies of local adaptation, behaviour and response to climate change in pikas to inform our expectations with regard to climate change impacts (Erb *et al.* 2011; Henry *et al.* 2012; Horn 2013; Lemay *et al.* 2013), strong support for divergent histories and divergent alleles among populations means that we cannot assume that all pop-

ulations will respond in a similar manner. Divergent histories (Galbreath *et al.* 2009) may explain why American Pikas appear to be responding idiosyncratically to environmental conditions and/or warming across their range (e.g. Simpson 2009; Millar & Westfall 2010; Rodhouse *et al.* 2010; Jeffress *et al.* 2013). If we are to fully understand the potential for populations to adapt locally to warming temperatures, we must specifically assess and accurately account for the role of historical associations shaping variation and shared alleles among populations, as well as current population connectivity.

Acknowledgements

This work was funded by NSF (DEB 1118815 to LLK, DDIG DEB 0808619 to LEO and HCL) and the Alaska Department of Fish and Game (State Wildlife Grant T-1-6 to LEO). In addition, we thank B. Jacobsen and A. Gunderson, for access to samples at the University of Alaska Museum, and D. Hik and J. Zgurski for providing samples of DNA from Pika Camp. Finally, we are extremely grateful to the 18 collectors (not including the authors) without whose arduous efforts to responsibly collect and prepare voucher specimens this study would not have been possible.

References

- Alvarado-Serrano DF, Knowles LL (2013) Ecological niche models in phylogeographic studies: applications, advances and precautions. *Molecular Ecology Resources*, **14**, 233–248.
- Anderson RP, Raza A (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography*, **37**, 1378–1393.
- Beever EA, Brussard PE, Berger J (2003) Patterns of apparent extirpation among isolated populations of pikas (*Ochotona princeps*) in the Great Basin. *Journal of Mammalogy*, **84**, 37–54.
- Braconnot P, Otto-Bliesner B, Harrison S *et al.* (2007) Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum-Part 1: experiments and large-scale features. *Climate of the Past*, **3**, 261–277.
- Broadbooks HE (1965) Ecology and distribution of the pikas of Washington and Alaska. *American Midland Naturalist*, **73**, 299–335.
- Castillo JA, Epps CW, Davis AR, Cushman SA (2014) Landscape effects on gene flow for a climate-sensitive montane species, the American pika. *Molecular Ecology*, **23**, 843–856.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *Genes, Genomes, Genetics*, **1**, 171–182.
- Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.
- COSEWIC (2011) *COSEWIC assessment and status report on the Collared Pika Ochotona collaris in Canada*. Committee on the Status of Endangered Wildlife in Canada, Ottawa, Ontario.
- Dawson MR (1967) Lagomorph history and the stratigraphic record. *University of Kansas Department of Geology Special Publications*, **2**, 287–316.

- Dixon EJ (2013) Late Pleistocene colonization of North America from Northeast Asia: new insights from large-scale paleogeographic reconstructions. *Quaternary International*, **285**, 57–67.
- Earl DA, vonHoldt BM (2012) Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
- Elith J, Kearney M, Phillips S (2010) The art of modeling range-shifted species. *Methods in Ecology and Evolution*, **1**, 330–342.
- Engelhardt BE, Stephens M (2010) Analysis of population structure: a unifying framework and novel methods based on sparse factor analysis. *PLoS Genetics*, **6**, e1001117.
- Erb LP, Ray C, Guralnick R (2011) On the generality of a climate-mediated shift in the distribution of the American pika (*Ochotona princeps*). *Ecology*, **92**, 1730–1735.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software Structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics*, **9**, e1003905.
- Fedorov VB, Goropashnaya AV, Jaarola M, Cook JA (2003) Phylogeography of lemmings (*Lemmus*): no evidence for postglacial colonization of Arctic from the Beringian refugium. *Molecular Ecology*, **12**, 725–731.
- Ferchaud A, Lyet A, Cheylan M *et al.* (2010) High genetic differentiation among French populations of the Orsini's viper (*Vipera ursinii ursinii*) based on mitochondrial and microsatellite data: implications for conservation management. *Journal of Heredity*, **102**, 67–78.
- Franken RJ (2002) *Demography and metapopulation dynamics of collared pikas (Ochotona collaris) in the southwest Yukon*. Master's Thesis, University of Alberta, Edmonton, Alberta.
- Franken RJ, Hik DS (2004) Influence of habitat quality, patch size and connectivity on colonization and extinction dynamics of collared pikas *Ochotona collaris*. *Journal of Animal Ecology*, **73**, 889–896.
- Galbreath KE, Hoberg EP (2012) Return to Beringia: parasites reveal cryptic biogeographic history of North American pikas. *Proceedings of the Royal Society B, Biological Sciences*, **279**, 371–378.
- Galbreath KE, Hafer DJ, Zamudio KR (2009) When cold is better: climate-driven elevation shifts yield complex patterns of diversification and demography in an alpine specialist (American Pika, *Ochotona princeps*). *Evolution*, **63**, 2848–2863.
- Galbreath KE, Cook JA, Eddingsaas AA, DeChaine EG (2011) Diversity and demography in Beringia: multilocus tests of paleodistribution models reveal the complex history of Arctic ground squirrels. *Evolution*, **65**, 1879–1896.
- Gao H, Williamson S, Bustamante CD (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics*, **176**, 1635–1651.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the PhyML 3.0. *Systematic Biology*, **59**, 307–321.
- Guthrie RD (1973) Mummified pika (*Ochotona*) carcass and dung pellets from Pleistocene deposits in interior Alaska. *Journal of Mammalogy*, **54**, 970–971.
- Harrington CR (2011) Pleistocene vertebrates of the Yukon Territory. *Quaternary Science Reviews*, **30**, 2341–2354.
- Hartl DL, Clark AG (2006) *Principles of Population Genetics*, 4th edn. Sinauer Associates, Sunderland, Massachusetts.
- He Q, Edwards D, Knowles LL (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, **67**, 3386–3402.
- Henry P, Sim Z, Russello MA (2012) Genetic evidence for restricted dispersal along continuous altitudinal gradients in a climate change-sensitive mammal: the American Pika. *PLoS ONE*, **7**, e39077.
- Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B*, **359**, 183–195.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.
- Hohenlohe PA, Catchen J, Cresko WA (2012) Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. *Methods in Molecular Biology*, **888**, 235–260.
- Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature Reviews Genetics*, **10**, 639–650.
- Horn HL (2013) *The role of habitat quality and climate in the dynamics of occupancy and survival of a population of collared pikas (Ochotona collaris) in the Ruby Range, Yukon Territory*. Master's Thesis, University of Alberta, Edmonton, Alberta.
- Jeffress MR, Rodhouse TJ, Ray C, Wolff S, Epps CW (2013) The idiosyncrasies of place: geographic variation in the climate-distribution relationships of the American pika. *Ecological Applications*, **23**, 864–878.
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.
- Jombart T, Pontier D, Dufour AB (2009) Genetic markers in the playground of multivariate analysis. *Heredity*, **102**, 330–341.
- Lanier HL, Olson LE (2009) Inferring divergence times within pikas (*Ochotona* spp.) using mtDNA and relaxed molecular dating techniques. *Molecular Phylogenetics and Evolution*, **53**, 1–12.
- Lanier HL, Olson LE (2013) Deep barriers, shallow divergences: reduced phylogeographical structure in the collared pika (Mammalia: Lagomorpha: *Ochotona collaris*). *Journal of Biogeography*, **40**, 466–478.
- Lanier HC, Gunderson AM, Weksler M, Fedorov VB, Olson LE (2015) Comparative phylogeography highlights the double-edged sword of climate change faced by arctic- and alpine-adapted species. *PLoS ONE*, **10**, e0118396.
- Lemay MA, Henry P, Lamb CT, Robson KM, Russello MA (2013) Novel genomic resources for a climate change sensitive mammal: characterization of the American pika transcriptome. *BMC Genomics*, **14**, 311.
- MacArthur RA, Wang LCH (1973) Physiology of thermoregulation in the pika, *Ochotona princeps*. *Canadian Journal of Zoology*, **51**, 11–16.
- MacDonald SO, Jones C (1987) *Ochotona collaris*. *Mammalian Species*, **281**, 1–4.

- Marr KL, Allen GA, Hebda RJ (2008) Refugia in the Cordilleran ice sheet of western North America: chloroplast DNA diversity in the Arctic-alpine plant *Oxyria digyna*. *Journal of Biogeography*, **35**, 1323–1334.
- Massatti R, Knowles LL (2014) Microhabitat differences impact phylogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution*, **10**, 2833–2846.
- McIntire EJB, Hik DS (2002) Grazing history versus current grazing: leaf demography and compensatory growth of three alpine plants in response to a native herbivore (*Ochotona collaris*). *Journal of Ecology*, **90**, 348–359.
- McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.
- Millar CI, Westfall RD (2010) Distribution and climatic relationships of the American Pika (*Ochotona princeps*) in the Sierra Nevada and Western Great Basin, USA; periglacial landforms as refugia in warming climates. *Arctic, Antarctic and Alpine Research*, **42**, 76–78.
- Morrison SF, Hik DS (2007) Demographic analysis of a declining pika *Ochotona collaris* population: linking survival to broadscale climate patterns via spring snowmelt patterns. *Journal of Animal Ecology*, **76**, 899–907.
- Morrison SF, Pelchat G, Donahue A, Hik DS (2009) Influence of food hoarding behavior on the over-winter survival of pikas in strongly seasonal environments. *Oecologia*, **159**, 107–116.
- Murray BM, Murray DF (1969) Notes on mammals in alpine areas of the northern St. Elias Mountains, Yukon Territory, and Alaska. *Canadian Field-Naturalist*, **83**, 331–338.
- Nagorsen DW, Keddie G (2000) Late Pleistocene mountain goats (*Oreamnos americanus*) from Vancouver Island: biogeographic implications. *Journal of Mammalogy*, **81**, 666–675.
- Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York City, New York.
- Oksanen J, Blanchet FG, Kindt R *et al.* (2013) vegan: Community Ecology Package. R package version 2.0-7. Available at <http://CRAN.R-project.org/package=vegan>.
- Palumbi SR, Barshis DJ, Traylor-Knowles N, Bay RA (2014) Mechanisms of reef coral resistance to future climate change. *Science*, **344**, 895–898.
- Peacock MM (1997) Determining natal dispersal patterns in a population of North American pikas (*Ochotona princeps*) using direct mark-resight and indirect genetic methods. *Behavioral Ecology*, **8**, 340–350.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.
- Petit RJ, Aguinagalde I, de Beaulieu J-L *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modeling*, **19**, 231–259.
- Pielou EC (1991) *After the Ice Age: the return of life to glaciated North America*. University of Chicago Press, Chicago, Illinois.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- R Core Team (2012) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rodhouse TJ, Beever EA, Garrett LK *et al.* (2010) Distribution of American pikas in a low-elevation lava landscape: conservation implications from the range periphery. *Journal of Mammalogy*, **91**, 1287–1299.
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.
- Ryan PG, Bloomer P, Moloney CL, Grant TJ, Delpont W (2007) Ecological speciation in South Atlantic island finches. *Science*, **315**, 1420–1423.
- Shafer ABA, Cullingham CI, Cote SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, **19**, 4589–4621.
- Shafer ABA, Côté SD, Coltman DW (2011) Hot spots of genetic diversity descended from multiple Pleistocene refugia in an alpine ungulate. *Evolution*, **65**, 125–138.
- Shah VB (2007) *An interactive system for combinatorial scientific computing with an emphasis on programmer productivity*. PhD Thesis, University of California, Santa Barbara, California.
- Simpson WG (2009) American pikas inhabit low-elevation sites outside the species' previously described bioclimatic envelope. *Western North American Naturalist*, **69**, 243–250.
- Smith AT (2008) The world of pikas. In: *Lagomorph biology: evolution, ecology and conservation* (eds Alves PC, Ferrand N, Hackländer K), pp. 89–102. Springer, Berlin.
- Smith AT, Ivins BL (1983) Colonization in a pika population: dispersal vs. philopatry. *Behavioral Ecology and Sociobiology*, **13**, 37–47.
- Smith AT, Li W, Hik DS (2004) Pikas as harbingers of global warming. *Species*, **41**, 4–5.
- Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity*, **98**, 128–142.
- Sturm M, Racine C, Tape K (2001) Climate change: increasing shrub abundance in the Arctic. *Nature*, **411**, 546–547.
- Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.
- USFWS (2010) 12-month finding on a petition to list the American pika as threatened or endangered. *Federal Register*, **75**, 6437–6471.
- Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.
- Wang C, Zöllner S, Rosenberg NA (2014) A quantitative comparison of the similarity between genes and geography in worldwide human populations. *PLoS Genetics*, **8**, e1002886.
- Weksler M, Lanier HC, Olson LE (2010) Eastern Beringian biogeography: historical and spatial genetic structure of singing voles (*Microtus miurus*) in Alaska. *Journal of Biogeography*, **37**, 1414–1431.
- Zgurski JM, Hik DS (2012) Polygynandry and even-sexed dispersal in a population of collared pikas, *Ochotona collaris*. *Animal Behavior*, **83**, 1075–1082.

Zgurski JM, Hik DS (2014) Gene flow and the restoration of genetic diversity in a fluctuating collared pika (*Ochotona collaris*) population. *Conservation Genetics*, **15**, 37–48.

H.C.L. and L.L.K. designed the study; H.C.L. and L.E.O. collected most of the samples; H.C.L. and Q.H. conducted Illumina Library Prep; R.M. and Q.H. analysed the data; all authors contributed to the manuscript.

Data accessibility

Raw genomic data, trimmed data input files, climate data and vetted occurrence data: Dryad doi:10.5061/dryad.jh2fp.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Summary of genomic data collected in each individual.

Table S2 Parameters used in FASTSIMCOAL2 analyses.

Table S3 Population pairwise F_{ST} values.

Table S4 PC1 and PC2 loading scores for environmental variability among pika populations.

Fig. S1 Mean quality scores of across reads averaged across individuals.

Fig. S2 Resulting number of retained and discarded reads from Illumina sequencing.

Fig. S3 Response curve showing relationship between bioclimatic variables and predicted pika presence.

Fig. S4 Habitat suitability maps without population locations.

Fig. S5 Dispersion of sampled populations in environmental space.

Fig. S6 Posterior probabilities for individuals assigned in STRKUCTURE when Pika Camp is excluded from the analysis.

Fig. S7 Maximum-likelihood estimate of population tree.