



Genome Resources

## Chromosome-Level Reference Genome Assembly for the American Pika (*Ochotona princeps*)

Bryson M. F. Sjodin, Kurt E. Galbreath, Hayley C. Lanier, and Michael A. Russello<sup>®</sup>

From the Department of Biology, University of British Columbia, Okanagan Campus, 3247 University Way, Kelowna, BC V1V 1V7, Canada (Sjodin and Russello); the Department of Biology, Northern Michigan University, Marquette, MI 49855, USA (Galbreath); and the Sam Noble Oklahoma Museum of Natural History and Department of Biology, University of Oklahoma, Norman, OK 73019, USA (Lanier).

Address correspondence to M. A. Russello at the address above, or e-mail: [michael.russello@ubc.ca](mailto:michael.russello@ubc.ca).

Received April 12, 2021; First decision May 14, 2021; Accepted May 20, 2021.

Corresponding Editor: Klaus-Peter Koepfli

### Abstract

The American pika (*Ochotona princeps*) is an alpine lagomorph found throughout western North America. Primarily inhabiting talus slopes at higher elevations (>2000 m), American pikas are well adapted to cold, montane environments. Warming climates on both historical and contemporary scales have contributed to population declines in American pikas, positioning them as a focal mammalian species for investigating the ecological effects of climate change. To support and expand ongoing research efforts, here, we present a highly contiguous and annotated reference genome assembly for the American pika (OchPri4.0). This assembly was produced using Dovetail de novo proximity ligation methods and annotated through the NCBI Eukaryotic Genome Annotation pipeline. The resulting assembly was chromosome-scale, with a total length of 2.23 Gb across 9350 scaffolds and a scaffold N50 of 75.8 Mb. The vast majority (>97%) of the total assembly length was found within 36 large scaffolds; 33 of these scaffolds correlated to whole autosomes, while the X chromosome was covered by 3 large scaffolds. Additionally, we identified 17 enriched gene ontology terms among American pika-specific genes putatively related to adaptation to high-elevation environments. This high-quality genome assembly will serve as a springboard for exploring the evolutionary underpinnings of behavioral, ecological, and taxonomic diversification in pikas as well as broader-scale eco-evolutionary questions pertaining to cold-adapted species in general.

**Subject Area:** Genome Resources

**Key words:** adaptation, climate change, cold tolerance, hypoxia, Lagomorpha, pikas

The American pika (*Ochotona princeps*) is a small, alpine lagomorph found along mountain ranges in western North America, with a distribution that stretches from central California in the Sierra Nevadas, to the Southern Rockies of Colorado and New

Mexico in the United States, and north to British Columbia and Alberta in the Cascades and Northern Rockies of Canada (Smith and Weston 1990; Hafner and Smith 2010). Although detected across a wide range of elevations (0–4000 m), American pikas are

primarily found at higher elevations (>2000 m), especially at more southerly points of the range (Smith and Weston 1990; Millar and Westfall 2010). American pikas are a typical rock-dwelling pika species, primarily inhabiting talus slopes, though they can be found along river banks and log piles at lower elevations and in cooler climates (Smith and Beever 2016). American pikas are habitat specialists and are well-adapted to cold, montane environments (Smith and Weston 1990); however, these adaptations can leave the species vulnerable to warmer temperatures (MacArthur and Wang 1974). In fact, warming climates across both historical and contemporary timeframes have been implicated in range contractions, population declines, and local extirpations (Beever et al. 2003, 2010, 2011; Wilkening et al. 2011; but see Smith 2020). Likewise, numerous niche models predict precipitous declines in the range of American pikas in response to climate change projections (Galbreath et al. 2009; Stewart et al. 2015; Schwalm et al. 2016). American pikas are also poor dispersers, particularly in warmer climates, further exacerbating the potential impacts of rapidly changing environments (Tapper 1973; Peacock 1997; Henry et al. 2012; Henry and Russello 2013; Robson et al. 2016; Smith and Beever 2016). This thermal sensitivity and limited dispersal ability have led many to consider the American pika as a sentinel mammalian species for the ecological effects of climate change (Beever et al. 2003; Wilkening and Ray 2016); a high-quality reference genome assembly would provide a fundamental resource for informing studies of pika ecology, evolution, and conservation moving forward.

Modern sequencing approaches, such as massively parallel sequencing, have led to an expansion of genomic resources across taxa, including non-model organisms (Muir et al. 2016). These resources include an ever-increasing number of annotated reference genomes; in fact, there are currently more than 1000 annotated reference or representative vertebrate genomes available in the NCBI RefSeq database alone (O'Leary et al. 2016). Beyond quantity, the quality of reference genomes being produced is also increasing with the introduction of new genome sequencing and assembly methods (Whibley et al. 2021). Whereas previous methods that largely relied on whole-genome shotgun sequencing often resulted in highly fragmented assemblies, modern techniques, including proximity ligation methods and single-molecule sequencing (Belton et al. 2012; Rhoads and Au 2015; Lu et al. 2016), are producing chromosome-level assemblies at substantially lower costs than previously possible (Giani et al. 2020). Reduction in reference genome fragmentation often leads to better annotation and can provide new insights, such as the discovery of novel genomic structures (Tørresen et al. 2017), enhanced characterization of coding and noncoding regions including improved gene models (Satou et al. 2008), and identification of previously unknown gene regions and biochemical pathways (Rice et al. 2017). These improvements can lead to more robust studies in ecology and evolution (Holt et al. 2018), biomedicine and agriculture (Warr et al. 2020), and conservation management (Macqueen et al. 2017).

Here, we present a chromosome-level reference genome assembly for the American pika. We assessed the quality of the new assembly relative to the previously available version and investigated synteny, orthology, and gene ontology (GO) in comparisons with the most closely related model organisms for which annotated reference genomes are available, including the European rabbit (*Oryctolagus cuniculus*) and house mouse (*Mus musculus*). In the process, we explored functional enrichment and identified genes that may be promising targets for investigating American pika adaptation to high-elevation environments.

## Methods

### Status of American Pika Genomic Resources

The currently published reference genome for the American pika (OchPri3.0; GenBank assembly accession: GCA\_000292845.1) available on NCBI is fragmented (10 420 scaffolds, scaffold N50 = 26.9 Mb, largest scaffold = 83.7 Mb; Table 1) and not assembled to chromosomes (*O. princeps* 2N = 68; Stock 1976). Genome annotation of this assembly through the NCBI Eukaryotic Genome Annotation Pipeline (Pruitt et al. 2014; O'Leary et al. 2016) also has fewer gene alignments (17 952 gene alignments with a target sequence coverage  $\geq 50\%$ ; NCBI Annotation Release 101) to curated databases (e.g., UniProt and SwissProt) than other chromosomal assemblies, such as *O. cuniculus* (18 830 gene alignments; NCBI Annotation Release 102) and *M. musculus* (20 905 gene alignments; NCBI Annotation Release 109).

### Biological Materials, Nucleic Acid Library Preparation, DNA Sequencing, and Genome Assembly

We obtained a liver tissue sample from an adult, male pika collected in Beaverhead-Deerlodge National Forest, Montana, USA, accessioned at the University of Alaska Museum of the North (Fairbanks, AK; Accession ID: UAM:Mamm:113957). High molecular weight DNA was extracted and sent to Dovetail Genomics (Scotts Valley, CA), who performed all subsequent library construction, sequencing, assembly, and scaffolding. Briefly, 3 Chicago™ libraries were constructed following the methods described by Putnam et al. (2016) and sequenced using the Illumina HiSeq X (PE150bp) platform yielding 402 million  $2 \times 151$  bp total read pairs. These libraries were then assembled and scaffolded using the Dovetail HiRise™ software pipeline following Putnam et al. (2016) using the OchPri3.0 genome as the input assembly. Shotgun reads from the Chicago library were mapped to the input assembly and a custom statistical model was generated for evaluating the assembly. The assembly was scanned for errors, breaking misjoins where appropriate. Candidate scaffold arrangements were evaluated over multiple iterations until maximum support for the scaffold arrangements was achieved. Three Hi-C libraries were subsequently prepared according to the manufacturer's

**Table 1.** Comparison of summary statistics for 2 American pika (*O. princeps*) genome assemblies

Measure	OchPri4.0	OchPri3.0
Total length	2.231 Gb	2.230 Gb
Number of scaffolds	9350	10 420
Scaffold L50/N50	11 scaffolds; 75.8 Mb	26 scaffolds; 26.8 Mb
Scaffold L90/N90	29 scaffolds; 27.1 Mb	97 scaffolds; 4.18 Mb
Longest scaffold	153.7 Mb	83.7 Mb
Contig N50	42.1 kb	42.4 kb
Number of gaps	124 941	122 542
Percentage of genome in gaps	12.88%	12.82%
BUSCO completeness:	92.4%	92.4%
Single copy	255	259
Duplicated	10	9
Fragmented	15	12
Missing	23	23
Total	303	303

protocol. Chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DpnII, the 5' overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, cross-links were reversed, and the DNA purified from proteins. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment size and a sequencing library was generated using Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before polymerase chain reaction enrichment of the library. The library was sequenced on an Illumina HiSeq X (PE150bp) platform to generate 342 million 2 × 151 bp read pairs. Hi-C sequence reads were then assembled and scaffolded using the Dovetail HiRise™ software pipeline using the above methods with the Chicago™ assembly as an input.

### Alignment and Identification of Putative Chromosomes

The Dovetail assembly (hereafter “OchPri4.0”) produced 37 large scaffolds (>5 Mb) covering >97% of the total assembly length (Supplementary Figure S1), putatively representative of chromosomes (see Results). To evaluate this possibility, we aligned the 37 largest scaffolds to the European rabbit (*O. cuniculus*) reference genome (OryCun2.0; GenBank assembly accession: GCA\_000003625.1), which is the closest related organism with a chromosomal assembly. The European rabbit does have both a larger genome size (~2.7 Gb) and fewer chromosomes (2N = 44) than the American pika; however, Ye et al. (2011) found evidence of significant homology between Forrest's pika (*O. forresti*; 2N = 54) and European rabbit using cross-species chromosome painting. We used the *k*-mer alignment algorithm implemented by the *nucmer* program in MUMmer v4.0.0.0beta2 (Kurtz et al. 2004). This program compares 2 genomes by finding highly similar regions between 2 sequences or genomes in both the forward- and reverse-complemented orientation and can also identify structural rearrangements. For this analysis, only assembled European rabbit chromosomes were used. Only alignments that uniquely matched to both the American pika and European rabbit genomes were considered to minimize misalignments. Alignments were visualized using Dot (<https://github.com/dnanexus/dot>). Scaffolds were considered chromosomes if most of the sequence aligned to at least 1 rabbit chromosome and did not overlap with any other scaffold alignments following Marques et al. (2020). The resulting American pika putative chromosomes were evaluated for their coverage across the rabbit genome. To further validate chromosomal designations, assembled scaffold lengths were sorted by size and plotted against chromosome size estimates of American pika autosomes following Rhie et al. (2021) and Ba et al. (2020). Size estimates were obtained from a karyotype image (Wurster et al. 1971) using a custom python script (see Supplemental Materials in Rhie et al., 2021).

### Assembly Annotation

Following putative chromosome identification, the reference genome was processed and annotated through the NCBI Eukaryotic Genome Annotation Pipeline, part of the NCBI RefSeq database (Pruitt et al. 2014; O'Leary et al. 2016). This automated pipeline performs all steps of the annotation process, including repeat identification and masking via WindowMasker (Morgulis et al. 2006), transcript (including transcriptome, where available) and protein alignment, ab initio and model-based gene prediction, and mapping of curated gene products. All annotation steps were performed de novo for the

new assembly making use of previously generated American pika transcriptome data (Lemay et al. 2013). The final annotated assembly was evaluated for completeness using BUSCO v3.0.2 (Simão et al. 2015). This software aligns a curated database of single-copy orthologs to a draft genome assembly and identifies complete, duplicated, fragmented, and missing genes within the assembly.

### Analysis of Synteny and Gene Orthology

We identified sequence homology and gene order between the American pika genome and those of 2 model organisms, European rabbit and house mouse, using synteny analysis as implemented in Satsuma v2.0 (Grabherr et al. 2010). We chose these species based on the availability of a chromosome-level genome assembly as well as their phylogenetic relationships to American pika. This analysis can identify chromosomal rearrangements, as indicated by split alignments (Taylor et al. 2018), and can also assist in identifying chromosomes within a draft assembly (Roodgar et al. 2020). The results of the synteny analysis were visualized using Circos v0.69-8 (Krzywinski et al. 2009).

We further identified orthologous gene sets between the American pika, European rabbit, and house mouse. To begin, we assigned functional annotations to all American pika sequences using BLASTP v2.9.0 (Camacho et al. 2009) with an E-value cutoff of  $1e^{-5}$  against the protein databases NR (nonredundant proteins from NCBI) and SwissProt (Boeckmann et al. 2003). We then mapped associated GO terms for BLAST hits using Blast2GO v5.2.5 (Götz et al. 2008). We mapped GO terms to motifs and domains using InterProScan v5.50-84.0 (Cock et al. 2013; Jones et al. 2016) using default settings, and GO terms were merged with those from BLAST in Blast2GO (Götz et al. 2008).

Protein sequences for all species were downloaded from NCBI, and proteins with multiple isoforms within each species were filtered such that only the longest isoform was retained. Additionally, sequences shorter than 50 amino acids were removed. Orthologous protein sequences were detected and grouped into gene sets using a Markov Clustering algorithm (MCL) as implemented in a custom version of OrthoMCL v2.0.9 (Fischer et al. 2011; [www.github.com/apetkau/orthomclsoftware-custom](http://www.github.com/apetkau/orthomclsoftware-custom)) and automated with the orthomcl-pipeline using default settings ([www.github.com/apetkau/orthomcl-pipeline](http://www.github.com/apetkau/orthomcl-pipeline)). We found overrepresented pika-specific genes (i.e., genes that could not be clustered into any gene family and could only be found in a single species) using a hypergeometric test using BiNGO v3.0.4 (Maere et al. 2005) as implemented in Cytoscape v3.8.2 (Smoot et al. 2011). We used the entire pika GO annotations as the reference set and corrected all *P*-values using the Benjamin-Hochberg false discovery rate with a corrected significance threshold of  $\alpha = 0.05$ .

## Results

### Genome Assembly

The initial Chicago™ assembly made 1312 breaks in the input assembly (OchPri3.0) as well as 1585 joins, reducing the overall scaffold count to 10 147. Five misassemblies were identified following Hi-C proximity ligation, while an additional 802 joins were introduced, resulting in a final genome assembly of total length 2.23 Gb across 9350 scaffolds with 37 scaffolds >5 Mb (Table 1; Supplementary Figure S1). The OchPri4.0 assembly offered a substantial improvement over the input assembly (OchPri3.0), with approximately one thousand fewer scaffolds and improved contiguity (scaffold L50/N50 = 11 scaffolds/75.8 Mb; Table 1). Additionally,

nearly the entire assembly (>97%) was covered by the 37 largest scaffolds (Supplementary Figure S1), whereas the input assembly was considerably more fragmented (>95% coverage over 136 scaffolds). Each assembly had similar numbers of gaps (OchPri3.0: 12.82% of the genome; OchPri4.0: 12.88%) as well as similar BUSCO scores using the eukaryota odb9 database (Table 1).

### Chromosome Identification

Thirty-six of the 37 scaffolds >5 Mb aligned to a single rabbit chromosome, none of which overlapped in alignment (Supplementary Figure S2). Only 1 scaffold (CM25742.1; chromosome 22) had partial alignment to 2 separate rabbit chromosomes (7 and 15). Of the remaining 36 scaffolds, 33 covered putative autosomes, while 3 separate scaffolds covered the putative X chromosome (none of which overlapped). Furthermore, these scaffold alignments covered the majority of the European rabbit genome, though with significant rearrangement between the 2 genomes (Supplementary Figure S2). The 33 putative autosomes also significantly correlated with chromosome size estimates taken from the American pika karyotype (Pearson's  $r = 0.985$ ;  $df = 31$ ;  $t = 32.34$ ;  $P < 0.0001$ ; Supplementary Figure S3).

### Genome Annotation

Annotation produced 21 186 genes and 32 187 RNA transcripts, both improvements over the previous reference genome, though still fewer than for the model organisms (European rabbit and house mouse; Table 2). Repeat masking resulted in 26.22% of the genome masked, similar to the previous assembly (26.24%); this value is lower than for the other species, likely reflective of better characterization of repeats in the model organisms (Table 2).

### Syntenic Analysis and Gene Orthology

We found substantial synteny with relatively few split alignments between the American pika and European rabbit (Figure 1a), providing further support for chromosomal designations. In contrast, we detected significant genomic rearrangements between the American pika and the more distantly related house mouse (Figure 1b). We did find substantial synteny between the 3 American pika X scaffolds and the house mouse X chromosome (Figure 1b). Moreover, several

**Table 2.** Summary of annotated features for 2 American pika genomes (OchPri4.0, NCBI Annotation Release 102; OchPri3.0, NCBI Annotation Release 101), the European rabbit reference genome (OryCun2.0, NCBI Annotation Release 102), and the house mouse reference genome (GRCm39, NCBI Annotation Release 109)

Feature	OchPri4.0	OchPri3.0	OryCun2.0	GRCm39
Genes	21 186	18 976	24 120	39 728
Transcripts:	32 187	26 432	44 108	131 115
mRNA	29 688	25 678	38 445	92 486
tRNA	340	330	485	422
lncRNA	335	328	4291	23 611
rRNA	4	0	1	64
CDSs	29 701	25 678	38 445	92 499
Exons	199 586	196 561	223 017	352 138
Introns	178 616	176 085	195 269	291 708
% masked	26.22%	26.24%	37.77%	44.29% <sup>a</sup>

All genomes were annotated with the NCBI Eukaryotic Genome Annotation Pipeline (Pruitt et al. 2014; O'Leary et al. 2016).

<sup>a</sup>Repeat masking was performed with RepeatMasker; all other genomes were masked with WindowMasker.

other house mouse chromosomes (e.g., chromosome 10; Figure 1b) seem to be covered by a small number of American pika chromosomes and could be indicative of Robertsonian translocations, as breakpoints between each American pika chromosome-alignment appear in centromeric locations.

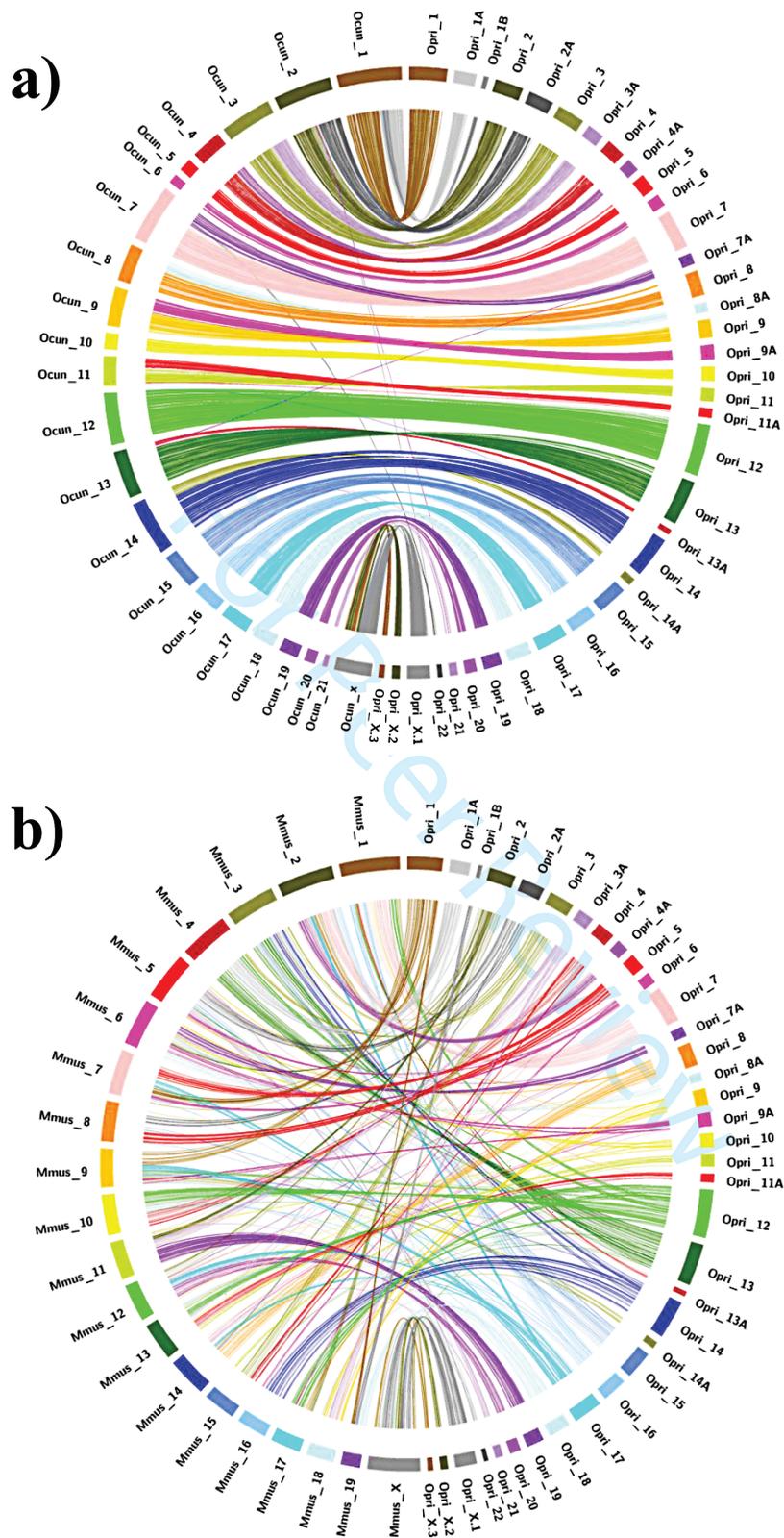
We found that 94.1% of filtered American pika genes clustered into 14 813 gene families, with 75 genes across 33 families and 1112 unclustered genes unique to the species (Supplementary Table S1). Of these unclustered genes, 1069 could be annotated with GO terms. Among these annotated American pika-specific genes, we detected significant over-representation for 152 GO terms following correction for multiple testing (Supplementary Table S2). We found 4 enriched GO terms (cytochrome-c oxidase activity [GO:0004129]; heme-copper terminal oxidase activity [GO:0015002]; oxidoreductase activity [GO:0016675 and GO:0016676]; Table 3) that have been previously linked to hypoxia response (Qiu et al. 2012; Li et al. 2018). Additionally, 10 enriched GO terms related to mitochondrial development and activity could be indicative of increased metabolic activity in American pikas (Table 3), 2 of which (mitochondrion [GO:0005739]; mitochondrial inner membrane [GO:0005743]) were consistent with upregulated genes in high-elevation Himalayan pikas (*O. roylei*; Solari et al. 2018). We also found 1 GO term associated with DNA double-strand break processing (GO:0000729) and 2 terms for the regulation of DNA repair (GO:0045739; GO:0006282), which could have implications for cellular response to UV damage (Table 3; Yang et al. 2015).

## Discussion

### Improvements to the American Pika Genome

Improving reference genome assemblies often can lead to novel discoveries. For example, improvements to the Atlantic cod (*Gadus morhua*) reference genome uncovered a high density of tandem repeats, with many associated with gaps in the previous assembly (Tørresen et al. 2017). A more contiguous American alligator (*Alligator mississippiensis*) genome led to the discovery of significantly enriched estrogen-responsive genes critical for sex determination (Rice et al. 2017). Moreover, an improved reference genome for rainbow trout (*Oncorhynchus mykiss*) identified novel structural variants between 2 separate lineages and featured a scaffolded sex-determination gene (*sdY*) in the Y chromosome sequence (Gao et al. 2021). Likewise, improvements to the desert poplar (*Populus euphratica*) assembly identified previously undescribed gene family expansions and unique structural variants likely involved with adaptation to xeric environments (Zhang et al. 2020), while an improved large yellow croaker (*Larimichthys crocea*) reference genome led to the discovery of numerous adaptations to diverse environmental conditions (Mu et al. 2018). With greater numbers of high-quality reference genomes being produced, novel discoveries will continue to provide unprecedented insights into species evolution.

Here, we produced a substantially improved reference genome assembly for the American pika. Our assembly offers considerable gains in contiguity, with >97% of the genome accounted for in 36 large scaffolds (Table 1; Supplementary Figure S1); the previous assembly required 178 scaffolds to reach this same threshold. We found that 33 of the new large scaffolds putatively represented autosomal chromosomes, while the X chromosome was covered by 3 separate scaffolds (Supplementary Figure S2). These chromosome-scale scaffolds strongly correlated with physical chromosome measurements, providing further evidence for chromosomal designations;



**Figure 1.** Synteny analysis and structural differences between American pika (Opri) and (a) European rabbit (Ocun) and (b) house mouse (Mmus) chromosomes. Only alignments > 1kb in length are shown for clarity.

thus, this assembly represents the first chromosome-level reference genome assembly for the American pika. We also found substantial improvements in annotation products, with 2210 genes, 5755 transcripts, 4023 coding domain sequences, and 2531 introns and

3025 additional annotations. Of the genes and transcripts annotated, 17 116 were identical between OchPri3.0 and OchPri4.0, with minor and major changes made to 13 437 and 1537 transcripts, respectively (see the NCBI *O. princeps* Annotation Release 102 for a

**Table 3.** Enriched GO terms of American pika (*O. princeps*) putatively linked to high-elevation adaptation

GO ID	Type	Description
GO:0004129	MF	Cytochrome-c oxidase activity
GO:0015002	MF	Heme-copper terminal oxidase activity
GO:0016675	MF	Oxidoreductase activity, acting on heme group of donors
GO:0016676	MF	Oxidoreductase activity, acting on heme group of donors, oxygen as acceptor
GO:0015078	MF	Hydrogen ion transmembrane transporter activity
GO:0005740	CC	Mitochondrial envelope
GO:0005743	CC	Mitochondrial inner membrane
GO:0031966	CC	Mitochondrial membrane
GO:0044455	CC	Mitochondrial membrane part
GO:0005742	CC	Mitochondrial outer membrane translocase complex
GO:0044429	CC	Mitochondrial part
GO:0005746	CC	Mitochondrial respiratory chain
GO:0005739	CC	Mitochondrion
GO:0070469	CC	Respiratory chain
GO:0000729	BP	DNA double-strand break processing
GO:0045739	BP	Positive regulation of DNA repair
GO:0006282	BP	Regulation of DNA repair

Enriched GO terms were detected from species-specific American pika genes using a hypergeometric test with false discovery rate correction for multiple testing. MF, molecular function; CC, cellular component; BP, biological process.

description of changes). There were 2988 newly annotated features in OchPri4.0, whereas 509 annotated features from OchPri3.0 were deprecated. Furthermore, we detected large-scale synteny between the American pika and European rabbit genomes, a finding that had been previously undescribed.

### Putative Signatures of Adaptation in the American Pika Genome

High-elevation habitats often have extreme environmental conditions such as low atmospheric oxygen (i.e., hypoxia) and low ambient temperatures (Sun et al. 2018). Hypoxic conditions can lead to reduced oxygen supply to tissues, limiting both aerobic metabolism and hindering thermogenesis in endotherms, exacerbating the effects of low ambient temperatures (Chevion and Brumfield 2012). Species must adapt to these unique challenges in order to survive life at high elevations, often at a genomic level (Chevion and Brumfield 2012; Qu et al. 2020). We found significant GO enrichment for genes putatively associated with adaptation to high-elevation environments in the American pika (Table 3). These included genes related to mitochondria development and structure (Solari et al. 2018), DNA repair (Yang et al. 2015), and response to hypoxic stress (Qiu et al. 2012; Li et al. 2018; Solari et al. 2018). American pikas have a high basal metabolic rate, which results in a very high mean body temperature (~40 °C; MacArthur and Wang, 1974). This elevated metabolic rate likely evolved to support higher levels of thermogenesis in response to cold climates. Previous studies have also found evidence for putative adaptation by the American pika to hypoxic and cold conditions (Lemay et al. 2013; Rankin et al. 2017; Waterhouse et al. 2018; Wang et al. 2020) as well as some evidence for adaptation to increased UV radiation (Wang et al. 2020). Further investigation will be necessary, however, to demonstrate links between genomic enrichment and phenotypic change in this species.

### Limitations and Future Directions

The assembly produced here can be improved further. While we were able to produce and identify chromosome-level scaffolds, most

of the scaffolds (9314) remain unlocalized to a particular chromosome. While these unplaced scaffolds do represent a minority of the genome (<3% of the total assembly length) and are generally quite short (mean length 7150 bp; range 1000 bp to 7.26 Mb), placement within chromosome scaffolds could help with gap filling and reduce the overall proportion of the genome found in gaps, which remains high even in the current assembly (Table 1). Many of these gaps are the result of long, repetitive sequences that cannot be spanned using short-read sequencing (English et al. 2012; Tørresen et al. 2017). One approach for gap filling and reducing the overall number of scaffolds is to incorporate sequencing from long-read technologies such as those offered from Oxford Nanopore (ONT) and Pacific Biosciences (Menlo Park, CA). These long-read sequencing platforms provide substantially longer read lengths than short-read platforms (typically <500 bp reads), producing read lengths in the tens of thousands to hundreds of thousands and even millions of base pairs (longest read from ONT was >2 Mb; Payne et al. 2019; Whibley et al. 2021). Long reads may span the entire length of a repeat region, which can be used to increase contiguity and resolve uncertainties in the assembly, often resulting in much higher-quality genome assemblies than those produced by short-read sequencing alone (Whibley et al. 2021).

We were able to identify 3 large scaffolds covering the American pika X chromosome, though we were unable to resolve the order and orientation of these scaffolds into a single molecule. This problem of incompletely assembled sex chromosomes is not unique to American pikas; the mammalian X and Y chromosomes contain many highly repetitive regions that often cannot be assembled using just short-read sequencing (Tomaszkiewicz et al. 2017). The mammalian Y chromosome can be particularly challenging to assemble, as this chromosome is primarily composed of repetitive sequences. One method to identify Y chromosome scaffolds is to compare candidate scaffolds from a draft assembly to annotated Y chromosome assemblies of a related species using BLAST, though these rely heavily on Y-linked genes, which may not be present on all Y scaffolds (Ba et al. 2020). Long-read sequencing can again be used to help resolve these difficult to assemble sex chromosomes and has been recently employed to produce the first

gapless, telomere-to-telomere assembly of the human X chromosome (Miga et al. 2020). Future incorporation of long-read sequencing into the American pika reference assembly could help resolve both the X and Y chromosomes into single, representative scaffolds, and, in general, improve contiguity and completeness across this assembly.

## Conclusions

Here, we generated the first chromosome-level reference genome assembly for the American pika. This assembly offers significant improvements over the previously published genome in terms of contiguity and annotation quality and exhibited large-scale synteny with the European rabbit genome. Moreover, we detected significant GO enrichment that may provide targets for future studies related to the genetic basis of adaptation to high-elevation environments, including investigations into gene family expansions and contractions, and identification of genes undergoing positive selection. These types of analyses have been fruitful in detecting putative signatures of adaptation in similar, high-elevation systems (Qiu et al. 2012; Shao et al. 2015; Li et al. 2018; Sun et al. 2018). The inclusion of long-read sequencing could provide further improvements to this assembly and should be explored in the future.

Moving forward, this reference genome will serve as a vital resource for research on the American pika and other mammals that occupy similar environments. Given that ecological disruption due to ongoing climate warming disproportionately affects species associated with high elevations and high latitudes (e.g., Cardillo et al. 2006; Moritz et al. 2008), enhancing capacity for genome-enabled investigations in these systems is a priority (Colella et al. 2020; Theodoridis et al. 2021). Such genomic perspectives hold great promise for addressing challenging questions regarding the relative roles of climate, landscape complexity, and interspecific interactions in the assembly of biotic communities (e.g., for North American pikas, see Lanier and Olson 2009; Galbreath et al. 2020). Furthermore, across Asia and North America, most pika species are closely tied to sensitive high-elevation environments, yet they exhibit diverse ecologies and behaviors (Smith et al. 2018). Thus, the high-quality genome assembly presented here will serve as a springboard for exploring the evolutionary underpinnings of behavioral, ecological, and taxonomic diversification in pikas as well as broader-scale eco-evolutionary questions pertaining to cold-adapted species in general.

## Supplementary Material

Supplementary material can be found at *Journal of Heredity* online.

## Funding

Dovetail Genomics Research Impact Award (to M.A.R. and H.C.L.); and Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN-2019-04621) (to M.A.R.) and Canada Graduate Scholarship (to B.M.F.S).

## Acknowledgments

We are grateful to Link Olson and the University of Alaska Museum of the North for providing the tissue sample used in this study. Computational resources were made available by Compute Canada through the Resources for Research Groups program.

## Data Availability

The reference genome assembly generated in this study is available in the NCBI BioProject repository under the accession PRNJNA649356.

## References

- Ba H, Cai Z, Gao H, Qin T, Liu W, Xie L, Zhang Y, Jing B, Wang D, Li C. 2020. Chromosome-level genome assembly of Tarim red deer, *Cervus elaphus yarkandensis*. *Sci Data*. 7(1):187.
- Beever EA, Brussard PF, Berger J. 2003. Patterns of apparent extirpation among isolated populations of pikas (*Ochotona princeps*) in the Great Basin. *J Mammal*. 84:37–54.
- Beever EA, Ray C, Mote PW, Wilkening JL. 2010. Testing alternative models of climate-mediated extirpations. *Ecol Appl*. 20:164–178.
- Beever EA, Ray C, Wilkening JL, Brussard PF, Mote PW. 2011. Contemporary climate change alters the pace and drivers of extinction. *Glob Change Biol*. 17:2054–2070.
- Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*. 58:268–276.
- Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, et al. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*. 31:365–370.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*. 10:421.
- Cardillo M, Mace GM, Gittleman JL, Purvis A. 2006. Latent extinction risk and the future battlegrounds of mammal conservation. *Proc Natl Acad Sci U S A*. 103:4157–4161.
- Cheviron ZA, Brumfield RT. 2012. Genomic insights into adaptation to high-altitude environments. *Heredity (Edinb)*. 108:354–361.
- Cock PJA, Grünig BA, Paszkiewicz K, Pritchard L. 2013. Galaxy tools and workflows for sequence analysis with applications in molecular plant pathology. *PeerJ*. 1:e167.
- Colella JP, Talbot SL, Brochmann C, Taylor EB, Hoberg EP, Cook JA. 2020. Conservation genomics in a changing Arctic. *Trends Ecol Evol*. 35:149–162.
- English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, et al. 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One*. 7:e47768.
- Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, Shanmugam D, Roos DS, Stoeckert CJ. 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. *Curr Protoc Bioinformatics*. Chapter 6, Unit 6.12:1–19.
- Galbreath KE, Hafner 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, Toman HM, Li C, Hoberg EP. 2020. When parasites persist: tapeworms survive host extinction and reveal waves of dispersal across Beringia. *Proc Biol Sci*. 287:20201825.
- Gao G, Magadan S, Waldbieser GC, Youngblood RC, Wheeler PA, Scheffler BE, Thorgaard GH, Palti Y. 2021. A long reads-based de-novo assembly of the genome of the Arlee homozygous line reveals chromosomal rearrangements in rainbow trout. *G3*. 11:jkab052.
- Giani AM, Gallo GR, Gianfranceschi L, Formenti G. 2020. Long walk to genomics: history and current approaches to genome sequencing and assembly. *Comput Struct Biotechnol J*. 18:9–19.
- Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 36:3420–3435.

- Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, Di Palma F, Lindblad-Toh K. 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics*. 26:1145–1151.
- Hafner DJ, Smith AT. 2010. Revision of the subspecies of the American pika, *Ochotona princeps* (Lagomorpha: Ochotonidae). *J Mammal*. 91:401–417.
- Henry P, Russello MA. 2013. Adaptive divergence along environmental gradients in a climate-change-sensitive mammal. *Ecol Evol*. 3:3906–3917.
- 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.
- Holt C, Campbell M, Keays DA, Edelman N, Kapusta A, Maclary E, Domyan ET, Suh A, Warren WC, Yandell M, et al. 2018. Improved genome assembly and annotation for the Rock Pigeon (*Columba livia*). *G3 (Bethesda)*. 8:1391–1398.
- Jones HP, Holmes ND, Butchart SH, Tershy BR, Kappes PJ, Corkery I, Aguirre-Muñoz A, Armstrong DP, Bonnaud E, Burbidge AA, et al. 2016. Invasive mammal eradication on islands results in substantial conservation gains. *Proc Natl Acad Sci U S A*. 113:4033–4038.
- Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res*. 19:1639–1645.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol*. 5:R12.
- Lanier HC, Olson LE. 2009. Inferring divergence times within pikas (*Ochotona* spp.) using mtDNA and relaxed molecular dating techniques. *Mol Phylogenet Evol*. 53:1–12.
- 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.
- Li JT, Gao YD, Xie L, Deng C, Shi P, Guan ML, Huang S, Ren JL, Wu DD, Ding L, et al. 2018. Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. *Proc Natl Acad Sci U S A*. 115:8406–8411.
- Lu H, Giordano F, Ning Z. 2016. Oxford Nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics*. 14:265–279.
- MacArthur RA, Wang LC. 1974. Behavioral thermoregulation in the pika *Ochotona princeps*: a field study using radiotelemetry. *Can J Zool*. 52:353–358.
- Macqueen DJ, Primmer CR, Houston RD, Nowak BF, Bernatchez L, Bergseth S, Davidson WS, Gallardo-Escárate C, Goldammer T, Guiguen Y, et al.; FAASG Consortium. 2017. Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genomics*. 18:484.
- Maere S, Heymans K, Kuiper M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. 21:3448–3449.
- Marques JP, Seixas FA, Farello L, Callahan CM, Good JM, Montgomery WL, Reid N, Alves PC, Boursot P, Melo-Ferreira J. 2020. An annotated draft genome of the Mountain Hare (*Lepus timidus*). *Genome Biol Evol*. 12:3656–3662.
- Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, Brooks S, Howe E, Porubsky D, Logsdon GA, et al. 2020. Telomere-to-telomere assembly of a complete human X chromosome. *Nature*. 585:79–84.
- Millar CI, Westfall RD. 2010. Distribution and climatic relationships of the American pika (*Ochotona princeps*) in the Sierra Nevada and Western Great Basin, U.S.A.; periglacial landforms as refugia in warming climates. *Arct Antarct Alp Res*. 42:76–88.
- Morgulis A, Gertz EM, Schäffer AA, Agarwala R. 2006. WindowMasker: window-based masker for sequenced genomes. *Bioinformatics (Oxford, England)*. 22:134–141.
- Moritz C, Patton JL, Conroy CJ, Parra JL, White GC, Beissinger SR. 2008. Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science*. 322:261–264.
- Mu Y, Huo J, Guan Y, Fan D, Xiao X, Wei J, Li Q, Mu P, Ao J, Chen X. 2018. An improved genome assembly for *Larimichthys crocea* reveals hepcidin gene expansion with diversified regulation and function. *Commun Biol*. 1:195.
- Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, Zhang J, Weinstock GM, Isaacs F, Rozowsky J, et al. 2016. Erratum to: The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol*. 17:78.
- O’Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 44:D733–D745.
- Payne A, Holmes N, Rakyar V, Loose M. 2019. BulkVis: a graphical viewer for Oxford Nanopore bulk FAST5 files. *Bioinformatics*. 35:2193–2198.
- 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 Ecol*. 8:340–350.
- Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell CM, Hart J, Landrum MJ, McGarvey KM, et al. 2014. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res*. 42:D756–D763.
- Putnam NH, O’Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res*. 26:342–350.
- Qiu Q, Zhang G, Ma T, Qian W, Wang J, Ye Z, Cao C, Hu Q, Kim J, Larkin DM, et al. 2012. The yak genome and adaptation to life at high altitude. *Nat Genet*. 44:946–949.
- Qu Y, Chen C, Xiong Y, She H, Zhang YE, Cheng Y, DuBay S, Li D, Ericson PGP, Hao Y, et al. 2020. Rapid phenotypic evolution with shallow genomic differentiation during early stages of high elevation adaptation in Eurasian tree sparrows. *Natl Sci Rev*. 7:113–127.
- Rankin AM, Galbreath KE, Teeter KC. 2017. Signatures of adaptive molecular evolution in American pikas (*Ochotona princeps*). *J Mammal*. 98:1156–1167.
- Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 592:737–746.
- Rhoads A, Au KF. 2015. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics*. 13:278–289.
- Rice ES, Kohno S, John JS, Pham S, Howard J, Lareau LF, O’Connell BL, Hickey G, Armstrong J, Deran A, et al. 2017. Improved genome assembly of American alligator genome reveals conserved architecture of estrogen signaling. *Genome Res*. 27:686–696.
- Robson KM, Lamb CT, Russello MA. 2016. Low genetic diversity, restricted dispersal, and elevation-specific patterns of population decline in American pikas in an atypical environment. *J Mammal*. 97:464–472.
- Roodgar M, Babveyh A, Nguyen LH, Zhou W, Sinha R, Lee H, Hanks JB, Avula M, Jiang L, Jian R, et al. 2020. Chromosome-level de novo assembly of the pig-tailed macaque genome using linked-read sequencing and HiC proximity scaffolding. *Gigascience*. 9:giaa069.
- Satou Y, Mineta K, Ogasawara M, Sasakura Y, Shoguchi E, Ueno K, Yamada L, Matsumoto J, Wasserscheid J, Dewar K, et al. 2008. Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon populations. *Genome Biol*. 9:R152.
- Schwalm D, Epps CW, Rodhouse TJ, Monahan WB, Castillo JA, Ray C, Jeffress MR. 2016. Habitat availability and gene flow influence diverging local population trajectories under scenarios of climate change: a place-based approach. *Glob Chang Biol*. 22:1572–1584.
- Shao Y, Li J-X, Ge R-L, Zhong L, Irwin DM, Murphy RW, Zhang Y-P. 2015. Genetic adaptations of the plateau Zokor in high-elevation burrows. *Sci Rep*. 5:17262.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31:3210–3212.
- Smith AT. 2020. Conservation status of American pikas (*Ochotona princeps*). *J Mammal*. 101:1466–1488.

- Smith A, Beever, E. 2016. IUCN Red List of Threatened Species: *Ochotona princeps*. IUCN Red List of Threatened Species. Available from: URL <https://www.iucnredlist.org/species/41267/45184315>
- Smith AT, Johnston CH, Alves PC, Hackländer K. 2018. *Lagomorphs: pikas, rabbits, and hares of the world*. Baltimore (MD): Johns Hopkins University Press.
- Smith AT, Weston ML. 1990. *Ochotona princeps*. *Mamm Species*. 352:1–8.
- Smoot ME, Ono K, Ruschinski J, Wang PL, Ideker T. 2011. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 27:431–432.
- Solari KA, Ramakrishnan U, Hadly EA. 2018. Gene expression is implicated in the ability of pikas to occupy Himalayan elevational gradient. *PLoS One*. 13:e0207936.
- Stewart JAE, Perrine JD, Nichols LB, Thorne JH, Millar CI, Goehring KE, Massing CP, Wright DH. 2015. Revisiting the past to foretell the future: summer temperature and habitat area predict pika extirpations in California. *J Biogeogr*. 42:880–890.
- Stock AD. 1976. Chromosome banding pattern relationships of hares, rabbits, and pikas (order Lagomorpha). A phyletic interpretation. *Cytogenet Cell Genet*. 17:78–88.
- Sun YB, Fu TT, Jin JQ, Murphy RW, Hillis DM, Zhang YP, Che J. 2018. Species groups distributed across elevational gradients reveal convergent and continuous genetic adaptation to high elevations. *Proc Natl Acad Sci U S A*. 115:E10634–E10641.
- Tapper SC. 1973. *The spatial organisation of pikas (Ochotona) and its effect on population recruitment* [PhD thesis]. Canada: University of Alberta (Edmonton, AB, Canada). Available from: URL <https://search.proquest.com/docview/250072010/citation/EBDFBA58DE1B4007PQ/1>
- Taylor GA, Kirk H, Coombe L, Jackman SD, Chu J, Tse K, Cheng D, Chuah E, Pandoh P, Carlsen R, et al. 2018. The genome of the North American brown bear or grizzly: *Ursus arctos* ssp. *horribilis*. *Genes*. 9:598.
- Theodoridis S, Rahbek C, Nogues-Bravo D. 2021. Exposure of mammal genetic diversity to mid-21st century global change. *Ecography*. 44:817–831.
- Tomaszkiewicz M, Medvedev P, Makova KD. 2017. Y and W chromosome assemblies: approaches and discoveries. *Trends Genet*. 33:266–282.
- Tørresen OK, Star B, Jentoft S, Reinart WB, Grove H, Miller JR, Walenz BP, Knight J, Ekholm JM, Peluso P, et al. 2017. An improved genome assembly uncovers prolific tandem repeats in Atlantic cod. *BMC Genomics*. 18:95.
- Wang X, Liang D, Jin W, Tang M, Liu S, Zhang P. 2020. Out of Tibet: genomic perspectives on the evolutionary history of extant pikas. *Mol Biol Evol*. 37:1577–1592.
- Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson HA, Flicek P, et al. 2020. An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience*. 9:giaa051.
- Waterhouse MD, Erb LP, Beever EA, Russello MA. 2018. Adaptive population divergence and directional gene flow across steep elevational gradients in a climate-sensitive mammal. *Mol Ecol*. 27:2512–2528.
- Whibley A, Kelley JL, Narum SR. 2021. The changing face of genome assemblies: guidance on achieving high-quality reference genomes. *Mol Ecol Resour*. 21:641–652.
- Wilkening JL, Ray C. 2016. Characterizing predictors of survival in the American pika (*Ochotona princeps*). *J Mammal*. 97:1366–1375.
- Wilkening JL, Ray C, Beever EA, Brussard PF. 2011. Modeling contemporary range retraction in Great Basin pikas (*Ochotona princeps*) using data on microclimate and microhabitat. *Quat Int*. 235:77–88.
- Wurster DH, Snapper JR, Benirschke K. 1971. Unusually large sex chromosomes: new methods of measuring and description of karyotypes of six rodents (*Myomorpha* and *Hystricomorpha*) and one lagomorph (*Ochotonidae*). *Cytogenet Genome Res*. 10:153–176.
- Yang Y, Wang L, Han J, Tang X, Ma M, Wang K, Zhang X, Ren Q, Chen Q, Qiu Q. 2015. Comparative transcriptomic analysis revealed adaptation mechanism of *Phrynocephalus erythrurus*, the highest altitude lizard living in the Qinghai-Tibet Plateau. *BMC Evol Biol*. 15:101.
- Ye J, Nie W, Wang J, Su W, Jing M, Graphodatsky AS, Yang F. 2011. Genome-wide comparative chromosome map between human and the Forrest's pika (*Ochotona forresti*) established by cross-species chromosome painting: further support for the Glires hypothesis. *Cytogenet Genome Res*. 132:41–46.
- Zhang Z, Chen Y, Zhang J, Ma X, Li Y, Li M, Wang D, Kang M, Wu H, Yang Y, et al. 2020. Improved genome assembly provides new insights into genome evolution in a desert poplar (*Populus euphratica*). *Mol Ecol Resour*. 20:781–794.