UC Santa Cruz
UC Santa Cruz Previously Published Works

Title

A Beary Good Genome: Haplotype-Resolved, Chromosome-Level Assembly of the Brown Bear (Ursus arctos)

Permalink

Journal

ISSN

Authors

Armstrong, Ellie E
Perry, Blair W
Huang, Yongqing
et al.

Publication Date

DOI

Peer reviewed

A Beary Good Genome: Haplotype-Resolved, Chromosome-Level Assembly of the Brown Bear (*Ursus arctos*)

Ellie E. Armstrong [1], Blair W. Perry [1], Yongqing Huang[2], Kiran V. Garimella[2], Heiko T. Jansen[3], Charles T. Robbins[1,4], Nathan R. Tucker[5,6,*], and Joanna L. Kelley [1,*]

[1]School of Biological Sciences, Washington State University, Pullman, WA, USA

[2]Data Sciences Platform, The Broad Institute of MIT and Harvard, Cambridge, MA, USA

[3]Integrative Physiology and Neuroscience, Washington State University, Pullman, WA, USA

[4]School of the Environment, Washington State University, Pullman, WA, USA

[5]Masonic Medical Research Institute, New York, NY, USA

[6]Cardiovascular Disease Initiative, The Broad Institute of MIT and Harvard, Cambridge, MA, USA

*Corresponding author: E-mail: ntucker@mmri.edu; joanna.l.kelley@wsu.edu.

Abstract

The brown bear (*Ursus arctos*) is the second largest and most widespread extant terrestrial carnivore on Earth and has recently emerged as a medical model for human metabolic diseases. Here, we report a fully phased chromosome-level assembly of a male North American brown bear built by combining Pacific Biosciences (PacBio) HiFi data and publicly available Hi-C data. The final genome size is 2.47 Gigabases (Gb) with a scaffold and contig N50 length of 70.08 and 43.94 Megabases (Mb), respectively. Benchmarking Universal Single-Copy Ortholog (BUSCO) analysis revealed that 94.5% of single copy orthologs from Mammalia were present in the genome (the highest of any ursid genome to date). Repetitive elements accounted for 44.48% of the genome and a total of 20,480 protein coding genes were identified. Based on whole genome alignment to the polar bear, the brown bear is highly syntenic with the polar bear, and our phylogenetic analysis of 7,246 single-copy orthologs supports the currently proposed species tree for Ursidae. This highly contiguous genome assembly will support future research on both the evolutionary history of the bear family and the physiological mechanisms behind hibernation, the latter of which has broad medical implications.

Key words: long-read sequencing, Ursidae, comparative genomics, hibernation.

Introduction

Brown bears (*Ursus arctos*) are a historically wide-ranging species, formerly occupying habitat from the southern tip of North America, across most of Asia and Europe, to the northernmost tip of Africa (McLellan ei al. 2017). However, as the second largest extant terrestrial carnivore, brown bears have seen extensive reductions in their range and even total extirpations in some regions due to habitat loss, climate change, and human-wildlife conflict (Albrecht et al. 2017). As top predators, brown bears also play an important role in ecosystem function (Duffy 2003). Brown bears are interesting ecological models that show local adaptations in both diet (Bojarska and Selva 2012), morphology (Sato et al. 2011; Colangelo et al. 2012), and other life history traits (Ferguson and McLoughlin 2000).

Brown bears have emerged as a model species for population genomics and speciation due to their interesting (and not fully resolved) demographic history, which contains signals of both incomplete lineage sorting and post-speciation hybridization (Miller et al. 2012; Cahill et al. 2013, 2015; Kumar et al. 2017; Barlow et al. 2018). Additionally, brown bear hibernation has been proposed as a medical model for

Significance

Brown bears (*Ursus arctos*) are the most widespread, large terrestrial carnivore on the planet and represent an example of speciation through hybridization, as well as a medical model for sedentary lifestyle-related disease. Although a previous genome for a brown bear has been published, the reported contig N50 was low (only ~530 kb), despite being scaffolded into putative chromosomes. Genomes of this quality limit the accuracy of analyses which rely on long contiguous stretches of the genome to be assembled (such as with some demographic analyses) as well as attempts at connecting genotype to phenotype (such as in association analyses). In order to support studies on both the complex hybridization history of the brown bear and investigations into medically relevant phenotypes, we generated a fully phased, chromosome-level assembly from a male grizzly bear. The genome has a total size of 2.47 Gb and 90% of the genome is contained in 36 scaffolds, roughly corresponding to one autosome per scaffold. This high-quality genome will enable studies across a variety of disciplines, including conservation, evolution, and medicine.

several diseases, including diabetes and insulin resistance (Rigano et al. 2017) and conditions related to sedentary lifestyles (Fröbert et al. 2020). While several genome assemblies for the brown bear have been published to date (Taylor et al. 2018), these assemblies have low contiguity (i.e., contig N50 = ~530 kilobases [kb]) which limits their value when being used to study brown bear biology. In order to improve the power and breadth of future research for the brown bear, we present a fully phased, chromosome-level assembly from a male brown bear built with Pacific Biosciences (PacBio) HiFi data and scaffolded with publicly available Hi-C data. We analyze this genome for quality and completeness, report on improved annotation statistics, and compare it with other publicly available bear genomes for repetitive element composition, diversity, and demographic history.

Results and Discussion

Genome Quality and Continuity

Utilizing the trio-binning method in Hifiasm (Cheng et al. 2021), we generated a phased assembly with one haplotype phase (Hifiasm-Hap1) totaling 2.47 Gb and the other (Hifiasm-Hap2) totalling 2.46 Gb (supplementary table S1, Supplementary material online) with a contig N50 of 48.3 and 48.2 Mb, respectively. The contig L90 indicated that 55 and 54 contigs made up 90% of the total genome, respectively. Note that for long read assemblies that do not incorporate a scaffolding step, only contig statistics are reported since no scaffolds are built. After incorporation of Hi-C data, the scaffold and contig N50 for hap1 (Hifiasm-Hap1 + HiC) and hap2 (Hifiasm-Hap2 + HiC) were 70.5 and 45.6 Mb, and 70.1 and 43.9 Mb, respectively. The slight decrease in contig N50 after Hi-C data incorporation likely indicates that some misassemblies were present in the original PacBio HiFi assembly. The final composite assembly includes the autosomes and unplaced scaffolds from Hifiasm-Hap2 + HiC, the putative Y chromosome scaffolds from Hifiasm-Hap2, and the major X scaffold from Hifiasm + HiC Hap1. The composite assembly had a scaffold and contig N50 of 70.1 and 43.9 Mb and a scaffold L90 of 36 (supplementary table S1, Supplementary material online), indicating that 90% of the assembly is contained in 36 scaffolds. Given the statistics of the final assembly, it is likely that most autosomes and the X chromosome are contained in approximately one scaffold, since the diploid Ursine karyotype is 37 (Wurster-Hill and Bush 1980; Nash et al. 1998).

Approximately 1.9 Mb of putative Y chromosome scaffolds were identified in a previous version of the polar bear assembly (Bidon et al. 2015), while the most recent polar bear assembly contains approximately 1.6 Mb of putative Y scaffolds (see GCF_017311325.1 assembly report via UCSC browser). After removing a misassembly from our composite assembly, we identified a total of approximately 9.9 Mb of putative Y scaffolds based on alignment to the Y scaffolds in the polar bear assembly (supplementary table S2, Supplementary material online). However, it is likely that there are still some misassemblies within this region due to the repetitive nature of mammalian Y chromosomes (Li et al. 2013).

The final composite assembly also improved upon the two previously published assemblies for *U. arctos,* which had a scaffold/contig N50 of 36.7/0.5 Mb and 72.2/0.5 Mb, respectively (supplementary table S3, Supplementary material online). Although the assembly produced by DNAZoo has a slightly lower scaffold L90 (32; supplementary table S3, Supplementary material online), the contig N50 is improved in our assemblies by approximately 88x. Undoubtedly, Hi-C data from a male bear (current Hi-C data is from a female) and/or additional long read data from a male bear will provide further improvements and resolution to this assembly.

Benchmarking Universal Single-Copy Ortholog (BUSCO) analyses revealed that each bear haplotype phase and the composite assembly had from 96.3% to 96.5% of expected complete genes (supplementary table S4,

Supplementary material online). We observed no changes in BUSCO scores when incorporating Hi-C data (supplementary table S4, Supplementary material online), revealing that any joins or misassemblies did not impact these genic regions. The BUSCO scores from the assemblies produced here are the highest scores across any currently published bear assembly (supplementary table S5, Supplementary material online), further indicating that the final assembled genomes are of high quality.

Genomic Synteny

In order to investigate the synteny between our genome and the polar bear (the closest relative to the brown bear with an estimated divergence date of more than one million years (Bidon et al. 2014; Cronin et al. 2014; Lan et al. 2022), we performed a whole-genome alignment. A total of 38 autosomal scaffolds comprising 2.17 Gb were identified as the major autosomal scaffolds in the brown bear. These 38 scaffolds aligned to 36 autosomal scaffolds, comprising 2.30 Gb of sequence, from the polar bear. Both the polar bear and the brown bear are ursine bear species, which are known to have a stable karyotype of $2n = 74$ (Wurster-Hill and Bush 1980; Nash et al. 1998). This agrees with the alignment produced here for the polar and brown bear (fig. 1A), which showed no major chromosomal rearrangements. For both the polar and brown bear, most of all 36 autosomes and the X chromosome appear to be represented primarily by one scaffold each, and only two sequences, corresponding to NW_024424452.1 and NW_024426230.1 in the polar bear assembly, appear to contain a break in the brown bear assembly. In fig. 1A, the break in NW_024424452.1 is clearly visible (the second

scaffold from the top in maroon), but it is not visible for the much smaller scaffold (the second scaffold from the bottom) due to the plot's resolution.

Repetitive Content

Across all bears, a total of 40.21–47.54% of each genome was made up of repetitive elements (fig. 1B). Across most repetitive element classes, a majority of the bears had comparable numbers, but differed most substantially in the total amount of "Small RNA", "Unclassified", and "Other" (comprising satellites, simple repeats, and low complexity regions) (fig. 1B). Consistent with previous results, we found that long interspersed nuclear elements (LINEs) represented the largest percentage of repetitive elements in the Ursidae family (Srivastava et al. 2019; Zhu et al. 2020), however, one study reported fewer total repeats in the American black bear (*Ursus americanus*) and a greater number of repeats overall in the giant panda (*Ailuropoda melanoleuca*) (Srivastava et al. 2019). Interestingly, while most Ursid species contain a relatively low percent of small RNA repetitive elements, we find expansions of this repetitive element class in the Andean bear (*Tremarctos ornatus*) and the Japanese black bear (*Ursus thibetanus japonicus*) (Fig. 1B). As these genomes have some of the lowest quality scores (see supplementary tables S3 and S4, Supplementary material online) improved assemblies will be needed to investigate whether this is an artifact of misassembly or reflects the actual repetitive element content.

To investigate putative telomeric repeats, we examined the repeat content along the ends of each of the longest 36 autosomal scaffolds in both our assembly and the assembly of the polar bear (GCF_017311325.1). Using non-
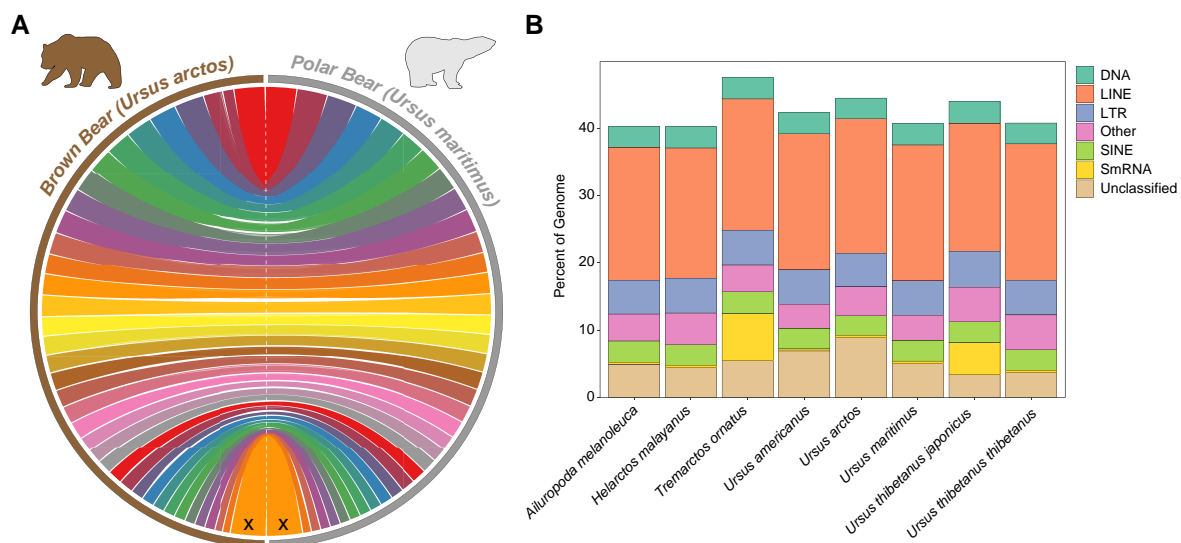
FIG. 1.—(A) Whole genome alignment between the brown bear and the polar bear containing all predicted autosomal scaffolds and the X chromosome. (B) Repeat content across Ursidae.

overlapping sliding 10 kb windows, we found that along the last 1 Mb of most of the scaffolds from both the brown bear assembled here and the polar bear, a clear increase in repetitive content can be observed (supplementary fig. S1, Supplementary material online). This increase is more pronounced in the brown bear (supplementary fig. S1A, Supplementary material online), which is unsurprising given the contig N50 in the brown bear is ~43.9 Mb, whereas the contig N50 in the polar bear is ~1.4 Mb (supplementary tables S1 and S3, Supplementary material online). This indicates that the brown bear has a much more contiguous assembly and at least part of the telomeric regions are included in the assembly, despite fewer large contig joins (see e.g. scaffold L90 in supplementary tables S1 and S3, Supplementary material online).

Gene Content

A total of 29,516 genes and pseudogenes were predicted in the genome assembly by the National Center for Biotechnology Information's annotation pipeline (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Ursus_arctos/102/). Of these, a total of 20,480 were protein coding, 5,160 were non-coding, 3,630 were pseudogenes, and 219 were immunoglobulin gene segments. Compared to the previously annotated genome for *U. arctos* (Taylor et al. 2018), this assembly adds 632 protein coding genes, reduces the number of non-coding and pseudogene sequences by 1,901 and 41, respectively, and increases the number of immunoglobulin gene segments by 100. This evidence, along with evidence provided by BUSCO (see section: *Quality and continuity*), indicates that the more complete and contiguous assembly has resulted in an updated assembly and annotation of gene regions.

Phylogenetics

Using a total of 7,246 single-copy orthologs from the mammalia_odb10 BUSCO dataset, we generated a consensus tree for Ursidae (fig. 2A). Previous studies have found a substantial amount of gene tree-species tree discordance due to possible incomplete lineage sorting and/or post-speciation hybridization (Cahill et al. 2013, 2015; Kumar et al. 2017; Barlow et al. 2018; Wang et al. 2022a). While most of these studies have focused on the potential hybridization history of the brown bear and polar bear, recent work has suggested that incomplete lineage sorting and post-speciation gene flow may be more prominent throughout the clade than previously expected (Kumar et al. 2017). Our results reiterate the basic species-tree topology and quartet scores revealed a high amount of gene tree-species tree discordance across the Ursid topology (fig. 2A). Reiterating the results from Kumar et al. 2017, we found strong signals of discordance both among American black bears, brown bears and polar bears (45–

54% of genes supported the main tree topology in this lineage), but also in the Asiatic bear lineages (40–50%). However, while previous studies have relied on consensus sequence generation, *de novo* assemblies and reference-free alignments like those performed here avoid mapping and consensus generation-related errors when performing evolutionary inference (Gopalakrishnan et al. 2017; Armstrong et al. 2020; Westbury et al. 2021; Prasad et al. 2022). Utilization of such methods may be essential for understanding evolution in young lineages like the bears, with high degrees of incomplete lineage sorting and hybridization.

Heterozygosity

To investigate the relative diversity across bear species, we estimated observed heterozygosity for all species of bear for which whole-genome sequence data and assemblies are available. We also included various subspecies and/or populations (see supplementary table S6, Supplementary material online for details). We found that even within bear species, heterozygosity values vary widely depending on population. Similar to previous results, we found that the endangered Apennine brown bears from Spain had the lowest heterozygosity values ([Endo et al. 2021]; fig. 2B). The other brown bear tested from Europe had higher values than those from the isolated Hokkaido brown bear (ssp *lasiotus*) population. Polar bears, the Japanese black bear, and the sun bear also show remarkably low heterozygosity values, similar to those from the endangered Apennine brown bears. The Tibetan black bear (*Ursus thibetanus thibetanus*) showed the highest heterozygosity across any of the other bear species, especially compared to its island counterpart, the Japanese black bear. However, there is limited information on the connectivity and history of the mainland Asiatic black bear across its range, so additional individuals should be sequenced to establish if this estimate is representative of the larger population/species.

Demographic History

We investigated the demographic history across the bear clade using pairwise sequentially markovian coalescent (PSMC). Our analyses are mostly consistent with previous investigations of demographic history in bears, including an increase in effective population size (N_e) approximately 120 kya for most mainland continental species (fig. 2C). Interestingly, although previous analyses have shown this N_e increase to be apparent in the Alaskan populations of the North American brown bear (Miller et al. 2012; Endo et al. 2021), we do not observe this in bears sampled from the lower 48 states. Moreover, inconsistencies with previous estimates of total N_e across species appear to be attributable to differences in the selected mutation rate.
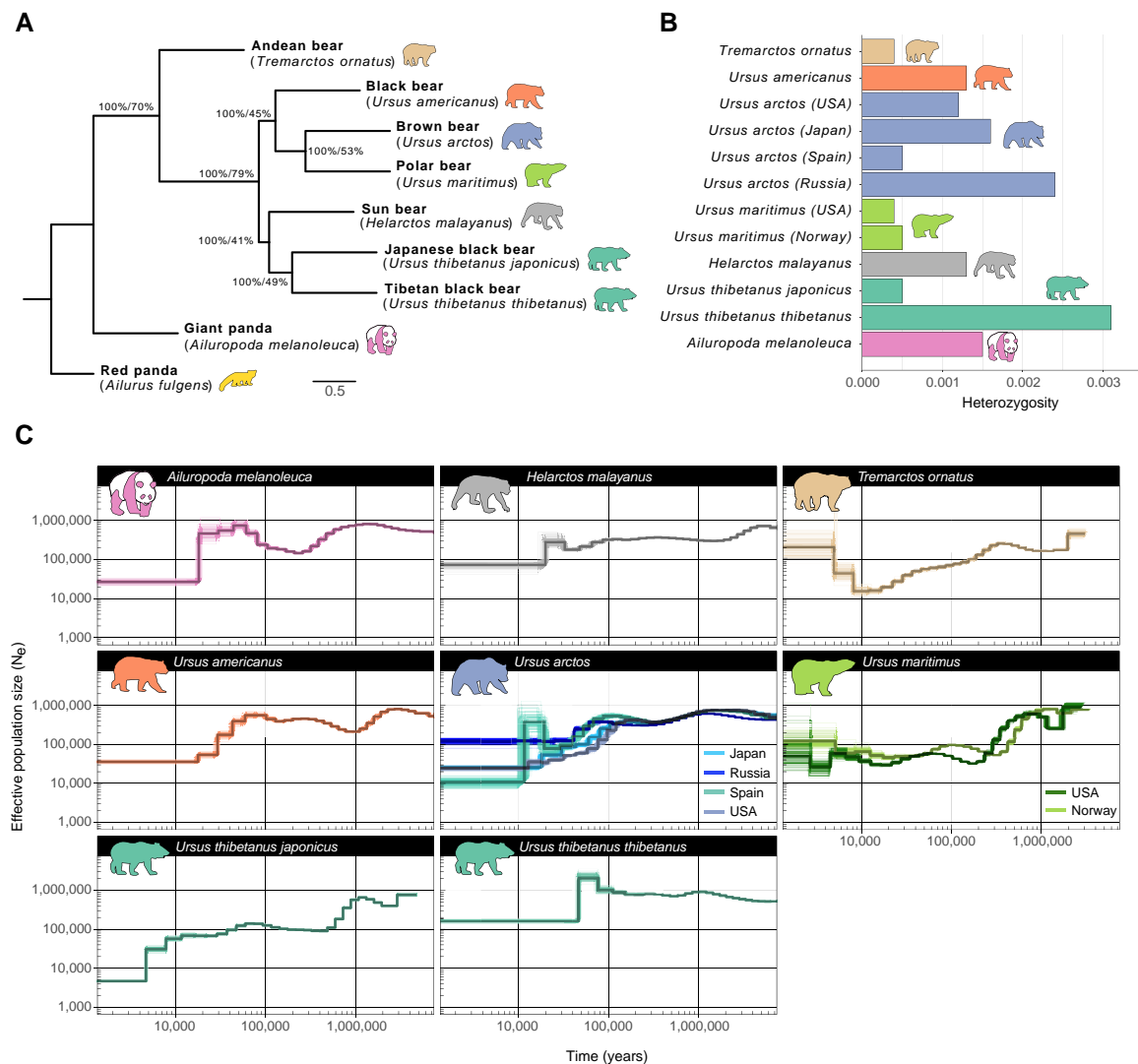
Fig. 2.—(*A*) Consensus phylogenetic tree generated from 7,246 single-copy orthologs. (*B*) Observed heterozygosity across various bear species and subspecies as calculated by angsd. (*C*) Effective population size estimates over time for focal bear species and populations by PSMC. Thick bold lines indicate estimates from the full dataset, while thinner lines indicate 100 individual bootstrap replicates.

Here, we predict a nearly doubled N_e compared to previous results (Cahill et al. 2013; Liu et al. 2014; Kumar et al. 2017; Zhu et al. 2020; Endo et al. 2021; Lan et al. 2022), which is consistent with expectations for having a mutation rate that is approximately halved (Nadachowska-Brzyska et al. 2016). This mutation rate is likely more accurate since it was calculated directly through trio sequencing (Wang et al. 2022).

Materials and Methods

Sample Acquisition and Library Preparation

To build a phased genome assembly using a trio-binning strategy, we collected blood samples from a bear trio (Adak, offspring; Oakley, mother; John, father). Protocols were followed according to (Joyce-Zuniga et al. 2016). All procedures were approved by the Washington State University Institutional Animal Care and Use Committee under protocol number ASAF 6546.

For short read whole genome sequencing of parental DNA, genomic DNA was isolated from frozen blood using the Gentra Puregene kit (Qiagen). Polymerase chain reaction-free whole genome sequencing libraries prepared at the Genomics Platform of the Broad Institute and were paired-end sequenced (2×150 bp) on a HiseqX to an estimated depth of 30×.

For long read whole genome sequencing, high molecular weight DNA was extracted from 3 mL of fresh blood. Before DNA extraction, red blood cells were lysed using the RBC lysis buffer from the Gentra Puregene kit

(Qiagen), and white blood cells were pelleted and washed. DNA from these cells was isolated using the Monarch HMW DNA Extraction Kit for Tissue (New England Biosciences, T3060). For PacBio library preparation, ≥3 ug of high molecular weight genomic DNA was sheared to ~15 kb using the Megaruptor 3 (Diagenode B06010003), with DNA repair and ligation of PacBio adapters accomplished with the PacBio SMRTbell Express Template Prep Kit 2.0 (100-938-900). Incomplete ligation products were removed with the SMRTbell Enzyme Clean Up Kit 2.0 (PacBio 101-938-500). Libraries were then size-selected for 15 kb ±- 20% using the PippinHT with 0.75% agarose cassettes (Sage Science). Following Qubit dsDNA High Sensitivity assay quantification (Thermo Q32854), libraries were diluted to 60 pM per SMRT cell, hybridized with PacBio V5 sequencing primer, and bound with SMRT seq polymerase using Sequel II Binding Kit 2.2 (PacBio 101-908-100). CCS sequencing was performed on the Sequel IIe using 8M SMRT Cells (101-389-001) and the Sequel II Sequencing 2.0 Kit (101-820-200), PacBio's adaptive loading feature was used with a 2 h pre-extension time and 30 h movie time per SMRT cell (3 cells in total). Initial quality filtering, base calling, adapter marking, and Circular Consensus Sequence (CCS) error correction was done automatically on the Sequel IIe. Sequencing yielded an estimated depth of coverage of 32X.

Genome Assembly

The haplotype-resolved assemblies were built using Hifiasm (Cheng et al. 2021), and yak (https://github.com/lh3/yak/releases/tag/v0.1) following the documentation (see https://hifiasm.readthedocs.io/en/latest/trio-assembly.html#trio-binning-assembly). Briefly, yak is used for collecting parent-specific *k*-mer distributions with parental short reads. These *k*-mer distributions are then used for binning the (CCS) long reads of the offspring into paternal-specific and maternal-specific reads. Hifiasm is then used with the appropriately partitioned reads for constructing the haplotype-specific assemblies (paternal and maternal). The pipeline (https://github.com/broadinstitute/long-read-pipelines/blob/3.0.39/wdl/tasks/Hifiasm.wdl) performing this Hifiasm step was written in WDL.

After the trio-phased assembly was built using Hifiasm (Cheng et al. 2021), we subsequently used publicly available Hi-C data for the brown bear (courtesy of DNAZoo: DNAZoo.org) to further scaffold the assembly. To incorporate these data, we used Juicer (Durand et al. 2016) according to the standard DNA Genome Assembly Cookbook instructions (https://aidenlab.org/assembly/manual_180322.pdf). We used both haplotypes generated by Hifiasm in the previous step as input (separately). We then used the 3D-DNA pipeline (Dudchenko et al. 2017) to generate a draft assembly for both haplotypes generated from Hifiasm.

To determine putative sex chromosomes in each haplotype we used BLAST (Altschul et al. 1990) to identify which scaffolds/contigs in our genomes best aligned to the polar bear Y scaffolds (ASM1731132v1). The X scaffold was identified using whole genome alignments (see *Whole genome alignment* below). We found evidence that the male bear haplotype (Hifiasm-Hap2 + HiC) contained a misassembly of the Y and X chromosomes. To correct this, we removed the two scaffolds containing BLAST hits from the Y chromosome and reincorporated the raw components of these scaffolds from the Hifiasm-Hap2 assembly into this genome. Lastly, to make a mappable genome that contained both sex chromosomes, we identified the major X chromosome scaffold from Hifiasm-Hap1 + HiC and incorporated it into the assembly. For our purposes, we refer to this as the "composite" assembly.

Quality and Continuity Assessment

We assessed the continuity and quality of each genome first using the Assemblathon2 scripts (Bradnam et al. 2013) followed by Benchmarking Universal Single-Copy Orthologs (BUSCOv5.3.0; (Simão et al. 2015) analysis. We analyzed all available bear assemblies using the *mammalia_odb10* datasets, with flags "–augustus" and "-m genome". For each species/subspecies, we selected the assembly with the best statistics from the Assemblathon2 and BUSCOv5.3.0 results to be included in the phylogenetics, repeat content analysis, and demographic history analyses. To see a complete description of which genomes were used, please refer to supplementary table S6, Supplementary material online.

Repeat Content

To assess the relative repeat content across the bear genomes, we used a combination of homology-based repeat finding, as well as *de novo* repeat finding. Briefly, we first used RepeatMaskerv4.0.9 (Smit et al. 1996) to mask repeats based on known repeat databases using flags "-species Ursidae", "-a", and "-gccalc" (Smit et al. 1996; Jurka et al. 2005). We then used the partially masked genome generated in the previous step as input to RepeatModeler v1.0.11 *BuildDatabase*, and subsequently performed *de novo* repeat finding using RepeatModelerv1.0.11 (Smit and Hubley 2008). Last, a masked file with both known and de novo repeats was produced by running RepeatMasker v4.0.9 with the flags "-a", and "-gccalc," and a final library produced from the previous step as input with the initial masked file. Total repeat content was calculated by adding the values from the initial and de novo steps. Repeat content was visualized and plotted in R (R Core Team 2013) using ggplot2v.3.3.6 (Wickham 2011).

To assess the possible presence of telomeric repeats across the putative autosomal scaffolds, we examined the ends of the 36 largest scaffolds (excluding the X scaffold)

in both our assembly and the assembly from the most recent polar bear (GCF_017311325.1). To accomplish this, we divided the genome into non-overlapping 10 kb windows using BEDTools (Quinlan and Hall 2010) and calculated the density of repeats (as defined by the % of the window occupied by repeats) in each 10 kb window. We then plotted the density of repeats starting at the ends of each scaffold for 1Mb total.

Whole Genome Alignment

The brown bear genome assembly (composite) was aligned to the polar bear genome assembly (ASM1731132v1) to investigate assembly completeness, as well as genomic synteny. For this analysis, all scaffolds from both genomes were aligned. Genomes were aligned following scripts from https://github.com/mcfrith/last-genome-alignments using LASTv921 (Kiełbasa et al. 2011). Genome alignment was visualized using the CIRCA software (http://omgenomics.com/circa) by plotting only the major scaffold(s) aligning to the putative 36 autosomes in the polar bear and the major X chromosome scaffold (see GCF_017311325.1 assembly report via UCSC browser). The major alignment was determined as the scaffold belonging to the query assembly (brown bear) that comprised a majority of the alignments to the putative polar bear chromosomes.

Phylogenetics

We built a phylogenetic tree for all the members of Ursidae that have a genome assembly using the single-copy BUSCOs (see *Quality and continuity assessment*). We first extracted all single-copy BUSCOs generated with the mammalia_odb10 dataset, since this dataset resulted in higher numbers of complete, single-copy BUSCO's across all *de novo* bear assemblies. Only genes which had a representative sequence from each species/subspecies were included. Each gene was then aligned using MAFFTv.7.490 (Katoh and Standley 2013) with the flags "–ep 0", "–genafpair", and "–maxiterate 1000". Alignments were then trimmed using Gblocks v.091b (Castresana 2000) with flag "-t D". Resulting files were then used as input into IQ-TREE 2 v. 2.1.3 (Minh et al. 2020) with flags "-bb 1000", "-nt AUTO", and "-m GTR+I+G". Lastly, we concatenated the maximum likelihood trees and built a species tree using ASTRAL-III v5.7.8 (Mirarab et al. 2014; Zhang et al. 2018) with flags "-gene-only" and "-t 2" to annotate the tree. The resulting tree was then plotted in FigTree v1.4.3 (Rambaut 2007) and manually rooted on *Ailurus fulgens* (red panda).

Demographic History

We used the PSMC method to investigate demographic history across Ursidae, (Li and Durbin 2011). We analyzed data representing species, subspecies, and distinct populations

of bears for which whole-genome sequencing data was available (see supplementary table S6, Supplementary material online). Each genome was indexed using BWA *index* with flags "-a bwtsw", and short-read data subsequently mapped using BWA-MEM. SAM files were converted to BAM format and sorted and an index generated. Subsequently, variant sites were called according to the suggested commands (see https://github.com/lh3/psmc). We used a minimum depth of 10 and a maximum depth of 100 for all samples except for the polar bear from Alaska, which was run with a minimum depth of 5 and a maximum depth of 50 due to it having a lower average sequencing depth (see supplementary table S6, Supplementary material online).

We next generated PSMC curves with 100 bootstraps using the suggested parameters linked above, with a mutation rate of 0.9225×10-9 per bp per year (Wang et al. 2022b) and a generation time of 10 years. Although there are a number of different generation times used for bears, we selected a generation time of 10 because we believe this to be a conservative estimate of generation time based on previous field studies (McLellan et al. 2017). We do note however, that small shifts in generation time are unlikely to impact the results of PSMC and only doubling this time will considerably impact results (Nadachowska-Brzyska et al. 2016). PSMC results were imported into R using psmcr v. 0.1–4 (see github.com/emmanuelparadis/psmcr) and plotted using ggplot2 v.3.3.6 (Wickham 2011).

Heterozygosity

We estimated heterozygosity for each unique subspecies of bear (for individuals see *Demographic history*). Using the previously generated bam files as input in the program angsd v.0.931 (Korneliussen et al. 2014), we set the reference and the ancestral sequence as the genome assembly for each respective species, along with the flags "-GL 1", "-dosaf 1", "-fold 1", "-minQ 20", and "-minmapq30". We generated folded spectra using the reference as the ancestral sequence, since the ancestral sequence is unknown. Subsequently, we ran the command realSFS within angsd and subsequently calculated the heterozygosity in R (R Core Team2013).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Data Availability

All data associated with this project has been deposited under NCBI BioProject Accession PRJNA807323. Intermediate assemblies and phased haplotypes available at https://doi.org/10.7273/000003791.

Literature Cited

Albrecht J, et al. 2017. Humans and climate change drove the Holocene decline of the brown bear. Sci Rep. 7:10399.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.

Armstrong EE, et al. 2020. Long live the king: chromosome-level assembly of the lion (*Panthera leo*) using linked-read, Hi-C, and long-read data. BMC Biol. 18:3.

Barlow A, et al. 2018. Partial genomic survival of cave bears in living brown bears. Nat Ecol Evol. 2:1563–1570.

Bidon T, et al. 2014. Brown and polar Bear Y chromosomes reveal extensive male-biased gene flow within brother lineages. Mol Biol Evol. 31:1353–1363.

Bidon T, Schreck N, Hailer F, Nilsson MA, Janke A. 2015. Genome-wide search identifies 1.9 Mb from the polar bear Y chromosome for evolutionary analyses. Genome Biol Evol. 7:2010–2022.

Bojarska K, Selva N. 2012. Spatial patterns in brown bear *Ursus arctos* diet: the role of geographical and environmental factors. Mamm Rev. 42:120–143.

Bradnam KR, et al. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. Gigascience 2:10.

Cahill JA, et al. 2013. Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. PLoS Genet. 9:e1003345.

Cahill JA, et al. 2015. Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. Mol Ecol. 24:1205–1217.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17:540–552.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 18:170–175.

Colangelo P, et al. 2012. Cranial distinctiveness in the Apennine brown bear: genetic drift effect or ecophenotypic adaptation? Biol J Linn Soc Lond. 107:15–26.

Cronin MA, Rincon G, Meredith RW. 2014. Molecular phylogeny and SNP variation of polar bears (*Ursus maritimus*), brown bears (U. arctos), and black bears (U. americanus) derived from genome 2014. J At Mol Phys. 105:312–323.

Dudchenko O, et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356: 92–95.

Duffy JE. 2003. Biodiversity loss, trophic skew and ecosystem functioning. Ecol Lett. 6:680–687.

Durand NC, et al. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3:95–98.

Endo Y, Osada N, Mano T, Masuda R. 2021. Demographic history of the brown bear (*Ursus arctos*) on Hokkaido Island, Japan, based on whole-genomic sequence analysis. Genome Biol Evol. 13:evab195.

Ferguson SH, McLoughlin PD. 2000. Effect of energy availability, seasonality, and geographic range on brown bear life history. Ecography 23:193–200.

Fröbert O, Frøbert AM, Kindberg J, Arnemo JM, Overgaard MT. 2020. The brown bear as a translational model for sedentary lifestyle-related diseases. J Intern Med. 287:263–270.

Gopalakrishnan S, et al. 2017. The wolf reference genome sequence (*Canis lupus* lupus) and its implications for Canis spp. population genomics. BMC Genomics 18:495.

Joyce-Zuniga NM, et al. 2016. Positive reinforcement training for blood collection in grizzly bears (*Ursus arctos* horribilis) results in undetectable elevations in serum cortisol levels: a preliminary investigation. J Appl Anim Welf Sci. 19:210–215.

Jurka J, et al. 2005. Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 110:462–467.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.

Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. Genome Res. 21: 487–493.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics 15:356.

Kumar V, et al. 2017. The evolutionary history of bears is characterized by gene flow across species. Sci Rep. 7:46487.

Lan T, et al. 2022. Insights into bear evolution from a Pleistocene polar bear genome. Proc Natl Acad Sci U S A. 119:e2200016119.

Li G, et al. 2013. Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. Genome Res. 23:1486–1495.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. Nature 475:493–496.

Liu S, et al. 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell 157:785–794.

McLellan BN, Proctor MF, Huber D, Michel S. 2017. *Ursus arctos*. The IUCN red list of threatened species 2016: e. T41688A45034772.

Miller W, et al. 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proc Natl Acad Sci U S A. 109:E2382–E2390.

Minh BQ, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 37: 1530–1534.

Mirarab S, et al. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541–i548.

Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. Mol Ecol. 25: 1058–1072.

Nash WG, Wienberg J, Ferguson-Smith MA, Menninger JC, O'Brien SJ. 1998. Comparative genomics: tracking chromosome evolution in

the family ursidae using reciprocal chromosome painting. Cytogenet Cell Genet. 83:182–192.

Prasad A, Lorenzen ED, Westbury MV. 2022. Evaluating the role of reference-genome phylogenetic distance on evolutionary inference. Mol Ecol Resour. 22:45–55.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842.

R Core Team. 2013. R: A language and environment for statistical computing. https://cran.microsoft.com/snapshot/2014-09-08/web/packages/dplR/vignettes/xdate-dplR.pdf.

Rambaut A. 2007. FigTree, a graphical viewer of phylogenetic trees.

Rigano KS, et al. 2017. Life in the fat lane: seasonal regulation of insulin sensitivity, food intake, and adipose biology in brown bears. J Comp Physiol B. 187:649–676.

Sato Y, Nakamura H, Ishifune Y, Ohtaishi N. 2011. The white-colored brown bears of the Southern Kurils. Ursus 22:84–90.

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.

Smit AFA, Hubley R. 2008. RepeatModeler Open-1.0. Available fom http://www.repeatmasker.org.

Smit AFA, Hubley R, Green P. 1996. RepeatMasker.

Srivastava A, et al. 2019. Genome assembly and gene expression in the American black bear provides new insights into the renal response to hibernation. DNA Res. 26:37–44.

Taylor GA, et al. 2018. The Genome of the North American brown bear or grizzly: *Ursus arctos* ssp. horribilis. Genes 9: 598.

Wang MS, et al. 2022a. A polar bear paleogenome reveals extensive ancient gene flow from polar bears into brown bears. Nat. Ecol. Evol.: 6:936-944.

Wang RJ, et al. 2022b. Hibernation shows no apparent effect on germline mutation rates in grizzly bears. bioRxiv. 2022.03.15.481369.

Westbury MV, et al. 2021. Ecological specialization and evolutionary reticulation in extant hyaenidae. Mol Biol Evol. 38:3884–3897.

Wickham H. 2011. Ggplot2. Wiley Interdiscip Rev Comput Stat. 3: 180–185.

Wurster-Hill DH, Bush M. 1980. The interrelationship of chromosome banding patterns in the giant panda (Ailuropoda melanoleuca) hybrid bear (Ursus middendorfi× Thalarctos maritimus), and other …. Cytogenet. Genome Res 27:147–154.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153.

Zhu C, et al. 2020. Draft genome assembly for the tibetan black bear (Ursus thibetanus thibetanus). Front. Genet 11:231.

Associate editor: Bonnie Fraser