

Chromosome-scale, haplotype-resolved genomes from a Southern Beaufort Sea polar bear and a Chichagof Island brown bear

Jonathan Kirshner

University at Buffalo

Flavio Augusto da Silva Coelho

University at Buffalo

Christopher A. Osborne

State University of New York at Oswego

Johanna Merkel

University at Buffalo

Anthony P. Crupi

Alaska Department of Fish and Game

Arnar Pálsson

University of Iceland

Charlotte Lindqvist

c1243@buffalo.edu

University at Buffalo

data-descriptor

Keywords:

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11074417/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Polar and brown bears (*Ursus maritimus* and *U. arctos*) are closely related species shaped by recent divergence, recurrent admixture, and ecological differentiation, making them a useful system for studying speciation and genome evolution under environmental change. However, existing genomic resources remain uneven in quality and rely on few chromosome-scale genomes that capture limited genetic diversity, constraining analyses of structural variation and introgression. Here, we present chromosome-scale, haplotype-resolved genome assemblies and annotations for a Southern Beaufort Sea Alaskan polar bear and a Chichagof Island brown bear, a genetically distinct brown bear lineage closely tied to polar bear evolutionary history. The polar bear primary assembly spans 2.44 Gb and comprises 229 scaffolds with a scaffold N50 of 73.7 Mb, 98.5% of the total sequence length anchored to 37 pseudo-chromosomes, and a 95.7% complete BUSCO score. The brown bear assembly exhibits comparable quality, with 98% of the assembled sequence anchored to 37 pseudo-chromosomes and 95.6% BUSCO completeness. Repeat annotation identified repetitive content across 34.2% and 34.3% of the polar and brown bear assemblies, respectively, and gene model prediction identified 27,356 and 27,563 protein-coding genes, respectively. These chromosome-scale genomic resources provide a foundation for comparative, population, and conservation genomic studies in this important species complex.

Background & Summary

Polar and brown bears (*Ursus maritimus* and *U. arctos*) are sister species whose evolutionary histories reflect recent divergence, ecological specialization, and recurrent gene flow across the Northern Hemisphere¹. Polar bears are uniquely adapted to extreme Arctic marine environments, exhibiting physiological and behavioral traits associated with prolonged fasting, cold tolerance, and a lipid-rich marine diet^{2–4}. Brown bears, by contrast, occupy a wide range of terrestrial habitats and display substantial ecological, morphological, and demographic diversity^{4–6}. Despite these differences, genomic studies have demonstrated ancient and recurrent introgressive hybridization between the two lineages, resulting in a complex evolutionary history shaped by both divergence and admixture^{1,7–9}.

This evolutionary complexity makes these bear species a compelling system to study genome evolution under ecological and demographic constraints. In particular, bears provide opportunities to investigate how structural variation, chromosomal organization, and genome-wide patterns of introgression can contribute to adaptation and diversification¹⁰. However, addressing these questions requires genomic resources that accurately represent chromosome-scale structure and allelic variation. In systems characterized by recent divergence and recurrent gene flow, fragmented, incomplete, or unresolved assemblies can obscure signals of admixture, misrepresent structural variants, and limit the resolution of comparative and population genomic analyses^{11,12}.

Although considerable polar and brown bear genome-scale data are publicly available^{8,13–15}, most genome assemblies remain fragmented, lack chromosome-scale scaffolding, or do not resolve

haplotypes, and many resequencing-based analyses still rely on mapping short reads to a single linear reference genome. This framework can poorly resolve structural, regulatory, and gene-content variation, especially in repetitive or structurally complex regions, and lineage-specific reference bias may create or obscure apparent asymmetries in ancestry. Inconsistencies in sequencing technologies, scaffolding approaches, and annotation pipelines across studies further complicate direct comparisons among polar bears and distinct brown bear lineages. Together, these limitations have constrained analyses of large-scale genome structure, structural variation, and chromosomal patterns of introgression.

Furthermore, the few available chromosome-scale assemblies capture only a narrow fraction of genetic diversity within the polar-brown bear lineage. For example, brown bears from the Admiralty, Baranof, and Chichagof Islands of Southeast Alaska (the ABC brown bears) comprise a genetically distinct lineage with a well-established role in polar bear evolutionary history. These brown bear populations share a close genetic ancestry with polar bears due to ancient admixture events, making them an essential comparative lineage for investigating gene flow and divergence within this species complex^{1,7,9,16,17}. High-quality genomic resources from such lineages are critical for contextualizing polar bear genomic variation and for discerning lineage-specific from shared evolutionary processes.

Here, we present chromosome-scale, haplotype-resolved genome assemblies and annotations for a polar bear (*Ursus maritimus*) from the Southern Beaufort Sea subpopulation and a Chichagof Island brown bear (*U. arctos sitkensis*). Both assemblies were generated using Oxford Nanopore long-read sequencing combined with Hi-C scaffolding and processed through identical assembly, scaffolding, repeat annotation, and gene annotation pipelines. Together, these resources provide a strong genomic foundation for future studies of structural variation, admixture, demographic history, and conservation genomics across bear lineages.

Materials and Methods

Sample collection and DNA extraction

Whole blood was obtained from a female polar bear in the Southern Beaufort Sea, Alaska (Collection #20520, Lab #AK034) and preserved in EDTA following standard protocols to minimize DNA degradation. A muscle tissue sample was collected from a male brown bear from Stag Bay on Chichagof Island, Alaska (Collection #22575, Lab #ABC027). Samples were preserved at -80°C until DNA extraction was performed.

Two high-molecular-weight (HMW) genomic DNA extractions were performed from the polar bear blood sample: once from 2 mL of whole blood, and once from 1 g of frozen whole blood (wet weight). HMW DNA from the brown bear was extracted from 80 mg of muscle tissue homogenized on ice using a tissue grinder and using the QIAGEN DNeasy Blood & Tissue Kit and QIAGEN Genomic Tips 100/G according to the manufacturer's protocol. The animal tissue protocol was used instead of the blood-

specific protocol because prior optimization, including minimal/no vortexing, demonstrated improved recovery of HMW DNA and preservation of long DNA fragments.

To enrich for long DNA molecules (> 10 kb) and reduce the contribution of short fragments to total sequencing output, a short-read elimination (SRE) step was performed prior to library preparation using a PacBio Short Read Eliminator kit (PN: 102-208-300). The SRE-prepared DNA was used as input for long-read sequencing library preparation. DNA integrity was assessed by agarose gel electrophoresis to confirm the presence of HMW DNA. DNA concentration and purity were evaluated using a NanoDrop spectrophotometer prior to library preparation.

Oxford Nanopore Technologies and Hi-C sequencing

Long-read sequencing was performed using Oxford Nanopore Technologies (ONT)¹⁸ on a PromethION platform. Sequencing libraries were prepared using the SQK-LSK114 ligation sequencing kit following the manufacturer’s instructions. Sequencing was performed using R10.4.1 flow cells run on the PromethION 24 and PromethION P2 Solo, yielding approximately 101 Gb or 41.49× genome coverage for the polar bear and 111 Gb or 45.72× genome coverage for the brown bear (Table 1).

To generate chromatin conformation capture data for chromosome-scale scaffolding, haplotype phasing, and assembly validation using Hi-C-based approaches¹⁹, Hi-C libraries were prepared by Phase Genomics (Seattle, WA 98109) from the same individuals used for ONT sequencing. The Illumina-compatible libraries were sequenced as paired-end reads, generating a total of 42.7 Gb or 17.50× genome coverage and 46.3 Gb or 19.1× genome coverage for the polar bear and brown bear, respectively.

Table 1
Nanoq assessment of sequenced ONT reads.

Quality metric	Polar bear	Brown bear
Number of reads	14,271,175	23,688,534
Number of bases	101,006,005,083	111,311,684,102
N50 read length	15,076	8,589
Longest read	1,376,846	1,671,892
Shortest read	5	5
Mean read length	7,077	4,698
Median read length	3,587	3,056

Genome assembly and scaffolding

ONT sequencing data were basecalled using Dorado v.0.9.6 (<https://github.com/nanoporetech/dorado.git>) with the super-accuracy model dna_r10.4.1_e8.2_400bps_sup v.5.0.0. Basecalled BAM files were converted to FASTQ format using SAMtools v.1.16.1²⁰. ONT reads were not length- or quality-filtered prior to assembly.

Genome assemblies were generated using Hifiasm v.0.25.0^{21,22} using the *--ont* option with all other parameters retained at default values. Hi-C read pairs were provided to Hifiasm during assembly using *--h1* and *--h2* to perform haplotype phasing and improve separation of homologous chromosomes, producing primary and alternate contig assemblies. Initial assembly outputs consisted of phased contigs for each haplotype and were evaluated using standard contiguity metrics, total assembly length, and consistency with the expected genome size.

Chromosome-scale scaffolding was performed for the primary and haplotype-resolved assemblies from both species using HapHiC v.1.0.6²³ followed by manual curation using Juicebox²⁴, which leverages Hi-C contact frequencies to cluster, order, and orient contigs into chromosome-length scaffolds. Prior to this step, Hi-C reads were mapped to the contig-level assembly using BWA v.0.7.17^{25,26}. Assemblies were indexed with *bwa index*, and Hi-C read pairs were aligned using *bwa mem -5SP*, with duplicate marking handled on-the-fly using samblaster v.0.1.26²⁷. Alignments were converted to BAM and filtered using samtools view to remove non-primary and low-information mappings (*flag filter -F 3340*), generating a filtered Hi-C BAM file for scaffolding. HapHiC was run independently for each assembly using Hi-C data generated from the corresponding individual, with default settings and a specified chromosome number of 37. Hi-C interaction matrices were manually inspected in Juicebox to validate scaffold structure, identify potential misjoins, and correct assembly errors where necessary. Final contact maps were visualized using the HapHiC plot command (Fig. 1).

Following Hi-C-scaffolding, the quality of the final chromosome-scale assemblies was assessed using QUAST v.5.0.2²⁸. Assembly completeness was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) v.5.4.3^{29,30}, with the *carnivora_odb10* dataset (total 14,502 BUSCOs), for each final assembly (primary and alternate haplotypes for each individual). Together, these metrics provide an integrated assessment of assembly contiguity, structural integrity, and gene space completeness.

The polar bear primary assembly spanned 2.44 Gb, with an initial contig N50 of 38.5 Mb across 353 contigs prior to scaffolding. Scaffolding resulted in an N50 of 73.7 Mb and 229 scaffolds, with 98.47% of the total assembled sequence anchored to 37 chromosome-length scaffolds. BUSCO analysis indicated 95.7% completeness (Table 2). The phased haplotypes were similarly contiguous and complete, each with comparable genome sizes, high scaffold N50 values, and BUSCO completeness exceeding 95% (Table 3).

To identify the putative X chromosome, chromosome-scale scaffolds from the new polar bear assembly were compared with the publicly available polar bear reference genome (GCF_017311325.1) using whole-genome synteny analysis in D-GENIES³¹, with alignments generated by Minimap2³²

(**Supplementary Fig. S1A**). A single scaffold showed strong syntenic correspondence to the reference X chromosome (NW_024423319.1) (**Supplementary Fig. S1C**), including approximately 45.0 Mb of high-confidence Minimap2 alignment (MAPQ = 60). The scaffold was provisionally designated chrX; however, given that the alignment covers only part of the reference X chromosome, this assignment remains tentative pending further validation.

Comparable results were obtained for the Chichagof brown bear assemblies. The primary assembly spanned 2.44 Gb, with an initial contig N50 of 48.8 Mb and 12,675 contigs. Following scaffolding, N50 increased to 72.4 Mb and decreased the number of scaffolds to 1,529, with 37 chromosome-length scaffolds containing approximately 98.03% of the entire assembled sequence. BUSCO analysis showed 95.6% complete BUSCOs (Table 2). Scaffold N50 values exceeded 70 Mb and BUSCO completeness ranged from 92.5% to 94.3% across alternate haplotypes (Table 3), consistent with chromosome-scale assemblies and high gene-space completeness.

Table 2

Assembly quality metrics and BUSCO completeness for scaffolded primary genome assemblies. Summary statistics for the contig level assembly produced by Hifiasm, as well as Hi-C scaffolded genome assemblies produced by HapHiC, with calculated QUAST and BUSCO completeness being assessed using the *carnivora_odb10* dataset (14,502 total BUSCOs).

Metric	Polar bear (Hifiasm)	Brown bear (Hifiasm)	Polar bear (HapHiC)	Brown bear (HapHiC)
Total length	2,438,398,189	2,442,652,352	2,438,573,026	2,442,667,152
# Scaffolds/Contigs	353	1,675	229	1,529
Largest Scaffold/Contig	88,326,776	108,844,939	134,589,315	124,729,119
N50	38,470,189	48,840,665	73,742,978	72,400,000
N90	16,436,648	9,903,676	37,810,831	37,206,774
L50	21	19	14	14
L90	54	55	31	32
# N's per 100 kbp	-	-	0.97	0.61
GC (%)	42.50	42.38	42.50	42.38
Complete BUSCOs (C)	13,880 (95.7%)	13,861 (95.6%)	13,880 (95.7%)	13,862 (95.6%)
Complete and single-copy BUSCOs (S)	13,729 (94.7%)	13,716 (94.6%)	13,748 (94.8%)	13,727 (94.7%)
Complete and duplicated BUSCOs (D)	151 (1.0%)	145 (1.0%)	132 (0.9%)	135 (0.9%)
Fragmented BUSCOs (F)	94 (0.6%)	100 (0.7%)	95 (0.7%)	98 (0.7%)
Missing BUSCOs (M)	528 (3.6%)	541 (3.7%)	527 (3.6%)	542 (3.7%)

Table 3

Assembly statistics and BUSCO completeness for haplotype-resolved genome assemblies. Summary assembly metrics are shown for the polar bear and brown bear haplotype-resolved genome assemblies. Standard assembly statistics were calculated using QUAST. Genome completeness was assessed with BUSCO using the *carnivora_odb10* dataset (14,502 total BUSCOs).

Metric	Polar bear Haplotype 1	Polar bear Haplotype 2	Brown bear Haplotype 1	Brown bear Haplotype 2
Total length	2,457,270,318	2,424,993,920	2,326,555,968	2,391,188,180
# Scaffolds	373	179	1,636	631
Largest Scaffold	134,674,407	126,540,317	108,226,276	109,124,548
Scaffold N50	72,991,455	73,249,088	71,284,333	71,823,863
Scaffold N90	37,424,484	36,609,223	34,968,376	36,543,907
Scaffold L50	14	14	14	14
Scaffold L90	32	32	32	32
# N's per 100 kbp	1.4	1.34	1.47	1.42
GC (%)	42.63	42.46	42.40	42.38
Complete BUSCOs (C)	13,850 (95.5%)	13,775 (95.0%)	13,416 (92.5%)	13,676 (94.3%)
Complete and single-copy BUSCOs (S)	13,695 (94.4%)	13,647 (94.1%)	13,290 (91.6%)	13,550 (93.4%)
Complete and duplicated BUSCOs (D)	155 (1.1%)	128 (0.9%)	126 (0.9%)	126 (0.9%)
Fragmented BUSCOs (F)	99 (0.7%)	96 (0.7%)	101 (0.7%)	104 (0.7%)

A single putative brown bear scaffold was identified as the X chromosome using the same approach of whole-genome comparison and Minimap2 alignment against the publicly available brown bear reference genome (GCF_023065955.2) (**Supplementary Fig. S1B**). This scaffold showed clear syntenic correspondence to the reference X chromosome (NC_079873.1), including approximately 58.3 Mb of high-confidence Minimap2 alignment (MAPQ = 60) (**Supplementary Fig. S1D**). The scaffold was provisionally labeled chrX, although potential structural differences between the assembled scaffold and reference X chromosome will require further investigation.

For the brown bear, we identified a candidate Y chromosome-associated sequence using a targeted read-mapping and reassembly approach. ONT reads from the Chichagof individual were mapped to the publicly available brown bear reference genome, and reads aligning to the reference Y chromosome (NC_079874.1) were retained and reassembled with Hifiasm^{21,22}, yielding a single 21.39 Mb Y-

associated scaffold. Portions of this sequence were independently recovered in one of the original haplotype-resolved assemblies, providing additional support for its authenticity. Whole-genome comparison with the brown bear reference genome showed extensive correspondence between the candidate scaffold and the reference Y chromosome, while alignments to the reference X chromosome were comparatively sparse and discontinuous (**Supplementary Fig. S2**). Together, the targeted reassembly, independent recovery within a phased haplotype assembly, and preferential correspondence to the reference Y chromosome support assignment of this scaffold as putatively Y chromosome-associated.

Repeat Annotation

Repetitive elements were identified using RepeatModeler v.2.0.4³³ to generate *de novo*, assembly-specific repeat libraries. For each genome, a RepeatModeler database was constructed using *BuildDatabase* with the *rmblast* search engine, and repeat discovery was performed with the *-LTRStruct* option enabled to improve detection and classification of long terminal repeat (LTR) retrotransposons. Consensus repeat sequences were classified using RepeatModeler's internal classification pipeline and exported as custom repeat libraries.

Genome-wide repeat annotation and masking were performed using RepeatMasker v.4.1.5³⁴ with the custom libraries. Soft-masking was applied with the *-xsmall* option, preserving repeat coordinates. Both interspersed repeats and low-complexity regions were annotated, and the resulting annotations were used to quantify total repetitive content and the relative contributions of major repeat classes.

The overall repeat composition of these assemblies is consistent with expectations for mammalian genomes as well as previous brown and polar bear assemblies³⁵. The relative abundance and composition of repeat classes were highly similar between polar and brown bear assemblies, consistent with their close evolutionary relationship (Fig. 2, Table 4). Repetitive elements comprised approximately one-third of both genomes and were dominated by retroelements, particularly LINES, with smaller contributions from SINEs and LTR retrotransposons.

Table 4

Summary of repetitive elements identified in the polar bear and brown bear genome assemblies.

Repeat Class	Polar bear			Brown bear		
	# elements	Length (bp)	% sequence	# elements	Length (bp)	% sequence
Total bases masked	-	834,934,988	34.24	-	838,649,727	34.33
Total interspersed repeats	-	793,120,023	32.52	-	787,795,611	32.25
Retroelements	2,627,525	676,742,454	27.75	2,623,490	664,402,586	27.2
SINEs	803,818	115,082,281	4.72	820,018	123,875,745	5.07
LINEs	1,471,641	470,342,939	19.29	1,409,738	450,807,720	18.46
L2/CR1/Rex	162,358	25,124,977	1.03	146,478	22,224,812	0.91
RTE/Bov-B	1,534	223,205	0.01	767	135,494	0.01
L1/CIN4	1,307,749	444,994,757	18.25	1,262,493	428,447,414	17.54
LTR elements	352,066	91,317,234	3.74	393,734	89,719,121	3.67
Gypsy/DIRS1	754	198,122	0.01	100	10,639	-
Retroviral	346,938	89,835,765	3.68	389,830	88,475,702	3.62
DNA transposons	276,084	42,588,922	1.75	251,130	44,254,012	1.81
hobo-Activator	195,396	28,592,682	1.17	178,789	28,145,014	1.15
Tc1-IS630-Pogo	80,306	13,805,887	0.57	69,656	12,699,785	0.52
MULE-MuDR	225	80,564	-	-	-	-
PiggyBac	-	-	-	237	120,323	-
Unclassified	45,821	73,788,647	3.03	48,920	79,139,013	3.24
Small RNA	609,503	92,211,189	3.78	636,569	109,115,212	4.47
Satellites	345	649,048	0.03	1,371	1,572,959	0.06
Simple repeats	640,460	33,097,893	1.36	657,597	34,375,622	1.41
Low complexity	98,365	6,951,461	0.29	98,463	5,841,490	0.24

Gene annotation

Repeat-masked genome assemblies were used as input for gene prediction and annotation pipelines to reduce spurious gene model predictions arising from repetitive regions.

Protein-coding gene annotation was performed using a combined strategy integrating homology-based prediction, *ab initio* gene finding, and evidence-weighted consensus building. Annotation was conducted using an identical pipeline for both genome assemblies.

Homology-based gene prediction was carried out with GeMoMa v.1.9³⁶ using the *GeMoMaPipeline* workflow, with reference annotations from two polar bear genomes, one brown bear genome, and a dog genome; (GCF_017311325.1¹⁵, GCF_023065955.2¹⁴, GCA_000687225.1⁸, and GCA_011100685.1³⁷). GeMoMa was run with orthology inference and paralog detection enabled (*o=true, p=true*), using ReAlign scoring (*GeMoMa.Score=ReAlign*). Predicted gene models were post-filtered using stringent completeness criteria requiring a canonical start codon, terminal stop codon, and a minimum score density threshold ($\text{score/aa} \geq 2.50$). Final GeMoMa predictions were exported with extractor output enabled and without further refinement via the AnnotationFinalizer step.

In parallel, *ab initio* gene prediction was performed using BRAKER3 v.3.0.8^{38–40} on the soft-masked genome assemblies. BRAKER3 was executed in protein-supported mode (*--prot_seq*) within a Singularity container. The use of soft-masked genomes allowed BRAKER3 to account for repetitive regions during model training and gene prediction.

Gene predictions from GeMoMa and BRAKER3 were integrated into a final consensus annotation using EvidenceModeler v.2.1.0 (EVM)⁴¹. Predictions were combined using an explicit weighting scheme that prioritized homology-based evidence (GeMoMa weight = 3, and GAF = 2) over *ab initio* predictions (AUGUSTUS, and GeneMark.hmm3 each weighted = 1). EVM was run with a segment size of 1 Mb and an overlap of 100 kb to accommodate chromosome-scale scaffolds. The resulting EVM consensus gene models represent the final protein-coding annotations used for downstream analyses.

For the polar bear assembly, the EVM consensus annotation contained 27,356 genes and BUSCO analysis recovered 89.2% complete BUSCOs. For the brown bear assembly, the EVM consensus annotation comprised 27,563 genes, with a BUSCO completeness of 88.4% (Table 5).

Table 5

BUSCO assessment of protein-based gene annotations of the polar bear and brown bear genome assemblies.

Metric	Polar bear			Brown bear		
	GeMoMa	BRAKER3	EVM	GeMoMa	BRAKER3	EVM
Total genes annotated	33,713	20,134	27,356	33,886	28,626	27,563
Complete BUSCOs (C)	13,097 (90.3%)	12,207 (84.2%)	12,943 (89.2%)	13,101 (90.3%)	6,993 (48.2%)	12,822 (88.4%)
Complete and single-copy BUSCOs (S)	12,975 (89.5%)	9,078 (62.6%)	12,830 (88.5%)	12,981 (89.5%)	5,904 (40.7%)	12,712 (87.7%)
Complete and duplicated BUSCOs (D)	122 (0.8%)	3,129 (21.6%)	113 (0.8%)	120 (0.8%)	1,089 (7.5%)	110 (0.8%)
Fragmented BUSCOs (F)	338 (2.3%)	117 (0.8%)	359 (2.5%)	334 (2.3%)	583 (4.0%)	408 (2.8%)
Missing BUSCOs (M)	1,067 (7.4%)	2,178 (15.0%)	1,200 (8.3%)	1,067 (7.4%)	6,926 (47.8%)	1,272 (8.8%)
Total BUSCO groups searched	14,502 (100%)	14,502 (100.0%)	14,502 (100%)	14,502 (100%)	14,502 (100.0%)	14,502 (100%)

Telomeric repeat content assessment

Terminal telomeric repeat content was assessed to evaluate assembly contiguity and chromosome-end recovery. Mammalian telomeres consist of tandem (TTAGGG)_n repeat arrays, and authentic chromosome termini are expected to show elevated telomeric repeat density within terminal windows^{42,43}. However, telomeric repeat detection alone does not establish complete, telomere-to-telomere assembly. Telomeric motifs (TTAGGG) were quantified in the scaffolded polar and brown bear assemblies using the Telomere Identification Toolkit v. v0.2.7⁴⁴ (tidk; <https://github.com/tolkit/telomeric-identifier.git>). We ran “tidk explore” (minimum repeat length = 5, maximum = 12), followed by “tidk find” with a 20 kb sliding window (--window 20000). Forward and reverse repeat counts were summed to generate total telomeric signal per window.

Terminal telomeric signal was assessed by identifying the maximum signal within the left (5') and right (3') terminal regions of each scaffold. These values were visualized in left-versus-right scatterplots using a custom Python script (pandas/matplotlib), and a conservative threshold of signal ≥ 250 repeats/20kb window was used to identify strong terminal telomeric enrichment⁴⁵. In the primary polar bear assembly, 34 scaffolds contained telomeric signal at one or both termini, including nine scaffolds with signal at both ends and 25 scaffolds at only one end. In the primary brown bear assembly, 34 scaffolds contained

telomeric signal at one or both termini, including nine scaffolds with signal at both ends and 25 scaffolds at one end only. (Fig. 3)

Technical Validation

Raw sequencing read quality was assessed using NanoQ v.0.10.0⁴⁶, which was used to summarize read length distributions, total yield, and per-read quality metrics for Oxford Nanopore sequencing data prior to genome assembly.

Assembly and annotation completeness were evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO) v.5.4.3 with the *carnivora_odb10* dataset, which contains 14,502 conserved orthologous genes expected to be present as single-copy in carnivoran genomes. BUSCO analyses were conducted independently on the final chromosome-scale genome assemblies and on the corresponding predicted protein-coding gene sets to assess levels of completeness, duplication, fragmentation, and missing genes. BUSCO results were summarized as percentages of complete single-copy, complete duplicated, fragmented, and missing orthologs for each assembly and annotation.

Standard assembly statistics were calculated using QUAST v.5.0.2²⁸ to assess overall assembly contiguity and structural quality. Reported metrics included total assembly length, number of contigs and scaffolds, contig and scaffold N50 values, maximum contig and scaffold lengths, GC content, and the number and total length of assembly gaps represented by ambiguous bases (Ns). These statistics were calculated separately for each primary and alternate haplotype assembly, enabling direct comparison of assembly quality across haplotypes and between species.

We also performed whole-genome alignments between our assemblies and the publicly available polar bear and brown bear reference genomes using D-GENIES. These comparisons provided visual support for chromosome-scale scaffold correspondence, including the provisional identification of sex-chromosome scaffolds. (**Supplementary Fig. S1**)

Telomeric repeat content was also compared with publicly available polar bear and brown bear reference genomes using the same *tidk*-based approach. This provided a consistent framework for evaluating relative chromosome-end recovery and visualizing differences in terminal telomeric signal among the new assemblies and existing genomic resources. (**Supplementary Fig. S2**)

When compared with current public RefSeq assemblies, the Southern Beaufort Sea polar bear assembly exhibited significant improvement over the current NCBI reference genome, with a contig N50 approximately 186-fold higher and approximately 88-fold fewer contigs, indicating substantially longer continuous sequence blocks prior to Hi-C scaffolding. The Chichagof Island brown bear assembly is best interpreted as a high-quality, lineage-specific complement to *UrsArc2.0*, with a higher contig N50 and comparable chromosome-scale scaffolding, but with a higher number of contigs and scaffolds and slightly lower reported BUSCO completeness. (**Supplementary Table S1**).

Data Records

Raw Oxford Nanopore long-read sequencing data for the polar bear and brown bear have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession XXXXX, with individual run accessions SRRXXXXXX–SRRXXXXXX. Hi-C sequencing data generated for chromosome-scale scaffolding are available under the same BioProject with run accessions SRRXXXXXX–SRRXXXXXX.

Chromosome-scale genome assemblies generated in this study have been deposited in NCBI GenBank. The *Ursus maritimus* assembly is available under accession GCA/GCF_XXXX.X, and the *Ursus arctos* assembly is available under accession GCA/GCF_XXXX.X. These assemblies include scaffolded primary assemblies as well as associated haplotype-resolved contigs where applicable.

Gene annotations produced using GeMoMa and EvidenceModeler (EVM) have been deposited alongside their respective genome assemblies under the same accession numbers and include corresponding GFF3 files and predicted protein and transcript FASTA files.

Datasets are provided in standard formats, including FASTA for genome assemblies, GFF3 for gene annotations, and FASTA for predicted protein and transcript sequences. These resources collectively provide a comprehensive genomic framework for comparative and population genomic analyses of Arctic bear lineages.

Declarations

Data Availability

The datasets described in Data Records are available through NCBI under BioProject PRJXXXXXX. Raw sequencing reads are deposited in the Sequence Read Archive under accessions SRRXXXXXX–SRRXXXXXX, and genome assemblies are available in GenBank under accessions GCA/GCF_XXXXXXX.X (polar bear) and GCA/GCF_XXXXXXX.X (brown bear). Gene annotation files are available alongside the corresponding assemblies. Custom scripts are available on GitHub (<https://github.com/jonnyPizza>).

Primary and phased haplotype assemblies, as well as annotations are available through Dryad at: https://datadryad.org/share/LINK_NOT_FOR_PUBLICATION/Ih2-bjXh35D-mS5Gt85siPIUWbSnuZs_2T2koi5z0Q.

Code Availability

All analyses were performed using publicly available bioinformatics software. Specific program versions, parameters, and implementation details are provided in the Methods. Custom scripts used to generate **Fig. 3** are available on GitHub (<https://github.com/jonnyPizza>).

Competing interests

The authors declare no competing interests.

Acknowledgments

We thank Todd Atwood from the U.S. Geological Survey Alaska Science Center for providing the polar bear tissue sample. We thank the Trevor Krabbenhoft lab at the University at Buffalo for access to and assistance with ONT sequencing. This work was supported with funding from the National Science Foundation (EF award #2221988 and DEB award #1556565). J.K. salary was paid by grant 2511413-051 of the Icelandic Research Fund, to A.P.

Author contributions

C.L. conceived and supervised the research; J.K. collected data, assembled the genomes and performed downstream analyses, and drafted the manuscript; F.A.S.C. collected data and performed preliminary data analyses; C.A.O. and J.M. contributed to genome assemblies and validation; A.P.C. collected the brown bear sample. A.P. provided resources and materials. J.K. and C.L. revised the final manuscript with contributions from all authors.

References

1. Lan, T. *et al.* Insights into bear evolution from a Pleistocene polar bear genome. *Proc. Natl. Acad. Sci.* 119, e2200016119 (2022).
2. Stirling, I. & Archibald, W. R. Aspects of Predation of Seals by Polar Bears. *J. Fish. Res. Board Can.* 34, 1126–1129 (1977).
3. Welch, A. J. *et al.* Polar Bears Exhibit Genome-Wide Signatures of Bioenergetic Adaptation to Life in the Arctic Environment. *Genome Biol. Evol.* 6, 433–450 (2014).
4. Luna-Arangur , C. & V zquez-Dom nguez, E. Bears into the Niche-Space: Phylogeography and Phyloclimatic Model of the Family Ursidae. *Diversity* 16, 223 (2024).
5. Hilderbrand, G. V. *et al.* Body size and lean mass of brown bears across and within four diverse ecosystems. *J. Zool.* 305, 53–62 (2018).
6. Haroldson, M. A. *et al.* Brown Bear (*Ursus arctos*; North America). in *Bears of the World: Ecology, Conservation and Management* (eds Melletti, M. & Penteriani, V.) 162–195 (Cambridge University Press, Cambridge, 2020). doi:10.1017/9781108692571.014.
7. Miller, W. *et al.* Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc. Natl. Acad. Sci.* 109, E2382–E2390 (2012).
8. Liu, S. *et al.* Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* 157, 785–794 (2014).
9. Lepp l , K., da Silva Coelho, F. A., Richter, M., Albert, V. A. & Lindqvist, C. Five-leaf Generalizations of the D-statistic Reveal the Directionality of Admixture. *Mol. Biol. Evol.* 41, msae198 (2024).
10. Zhang, L., Reifov , R., Halenkov , Z. & Gompert, Z. How Important Are Structural Variants for Speciation? *Genes* 12, 1084 (2021).

11. Ahmad, B. *et al.* Mango pangenome reveals dramatic impacts of reference bias on population genomic analyses. *Hortic. Res.* 12, uhaf166 (2025).
12. Bredemeyer, K. R. *et al.* The value of hybrid genomes: Building two highly contiguous reference genome assemblies to advance *Canis* genomic studies. *J. Hered.* 115, 480–486 (2024).
13. Taylor, G. A. *et al.* The Genome of the North American Brown Bear or Grizzly: *Ursus arctos* ssp. *horribilis*. *Genes* 9, 598 (2018).
14. Armstrong, E. E. *et al.* A Beary Good Genome: Haplotype-Resolved, Chromosome-Level Assembly of the Brown Bear (*Ursus arctos*). *Genome Biol. Evol.* 14, evac125 (2022).
15. Laidre, K. L. *et al.* Glacial ice supports a distinct and undocumented polar bear subpopulation persisting in late 21st-century sea-ice conditions. *Science* 376, 1333–1338 (2022).
16. Cahill, J. A. *et al.* Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Mol. Ecol.* 24, 1205–1217 (2015).
17. da Silva Coelho, F. A. *et al.* Ancient bears provide insights into Pleistocene ice age refugia in Southeast Alaska. *Mol. Ecol.* 32, 3641–3656 (2023).
18. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* 39, 1348–1365 (2021).
19. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long range interactions reveals folding principles of the human genome. *Science* 326, 289–293 (2009).
20. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009).
21. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18, 170–175 (2021).
22. Cheng, H. *et al.* Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* 10.1038/s41587-022-01261-x (2022) doi:10.1038/s41587-022-01261-x.
23. Zeng, X. *et al.* Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nat. Plants* 10, 1184–1200 (2024).
24. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* 3, 99–101 (2016).
25. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760 (2009).
26. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. (2013).
27. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 30, 2503–2505 (2014).
28. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075 (2013).
29. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31,

- 3210–3212 (2015).
30. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* 38, 4647–4654 (2021).
 31. Cabanettes, F. & Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6, e4958 (2018).
 32. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100 (2018).
 33. Smit, AFA, Hubley, R. RepeatModeler Open-1.0. 2008–2015 %3C <http://www.repeatmasker.org> %3E. Smit, AFA, Hubley, R. (2008).
 34. Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013–2015 %3C <http://www.repeatmasker.org> %3E.
 35. Platt, R. N., Vandewege, M. W. & Ray, D. A. Mammalian transposable elements and their impacts on genome evolution. *Chromosome Res.* 26, 25–43 (2018).
 36. Keilwagen, J., Hartung, F. & Grau, J. GeMoMa: Homology-Based Gene Prediction Utilizing Intron Position Conservation and RNA-seq Data. in *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) 161–177 (Springer New York, New York, NY, 2019). doi:10.1007/978-1-4939-9173-0_9.
 37. Wang, C. *et al.* A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. *Commun. Biol.* 4, 185 (2021).
 38. Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34, W435–W439 (2006).
 39. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics Bioinforma.* 2, lqaa026 (2020).
 40. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP + and AUGUSTUS supported by a protein database. *NAR Genomics Bioinforma.* 3, lqaa108 (2021).
 41. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, R7 (2008).
 42. Aubert, G. & Lansdorp, P. M. Telomeres and Aging. *Physiol. Rev.* 88, 557–579 (2008).
 43. Nurk, S. *et al.* The complete sequence of a human genome. *Science* 376, 44–53 (2022).
 44. Brown, M. R., Manuel Gonzalez De La Rosa, P. & Blaxter, M. tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics* 41, btaf049 (2025).
 45. Srinivas, N., Rachakonda, S. & Kumar, R. Telomeres and Telomere Length: A General Overview. *Cancers* 12, 558 (2020).
 46. Steinig, E. & Coin, L. Nanoq: ultra-fast quality control for nanopore reads. *J. Open Source Softw.* 7, 2991 (2022).

Figures

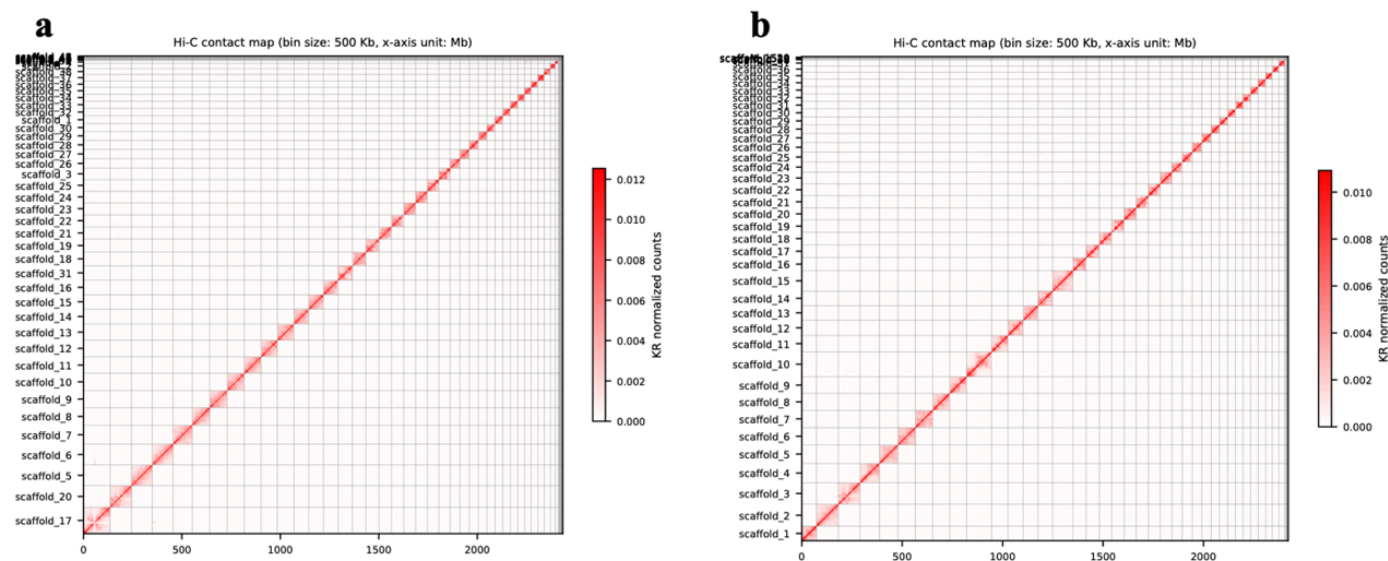


Figure 1

Chromosome-scale Hi-C interaction maps. **(a)** Polar bear and **(b)**Chichagof brown bear assemblies. Hi-C contact matrices showing intra- and inter-chromosomal interaction frequencies. Heatmap intensity (red gradient) represents normalized contact frequencies between genomic loci, with stronger signals indicating higher interaction probabilities. Chromosomes reveal clear spatial partitioning of chromosomal territories. Prominent diagonal patterns reflect strong intra-chromosomal interactions, confirming successful chromosome-scale scaffolding of both genome assemblies.

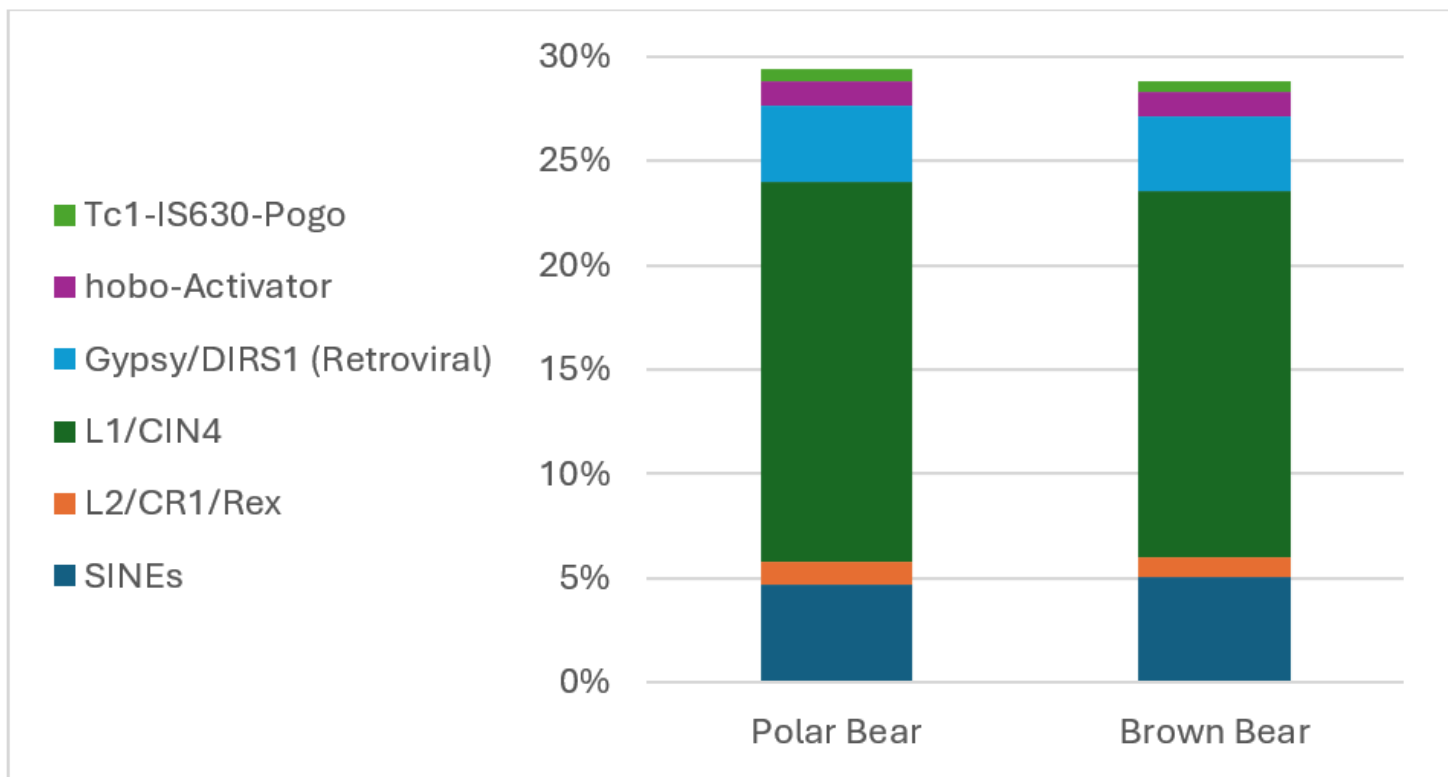


Figure 2

Genome-wide repeat composition by hierarchical subclass in the polar bear and Chichagof brown bear assemblies. Stacked bar plots show the proportion of each genome occupied by major classes and hierarchical subclasses of repetitive elements, as identified by RepeatModeler and RepeatMasker. Percentages indicate the fraction of the total genome attributed to each repeat category. Retrotransposons comprise the largest component in both genomes, dominated by LINE elements, with additional contributions from SINEs, LTR retrotransposons, DNA transposons, and other repeat subclasses.

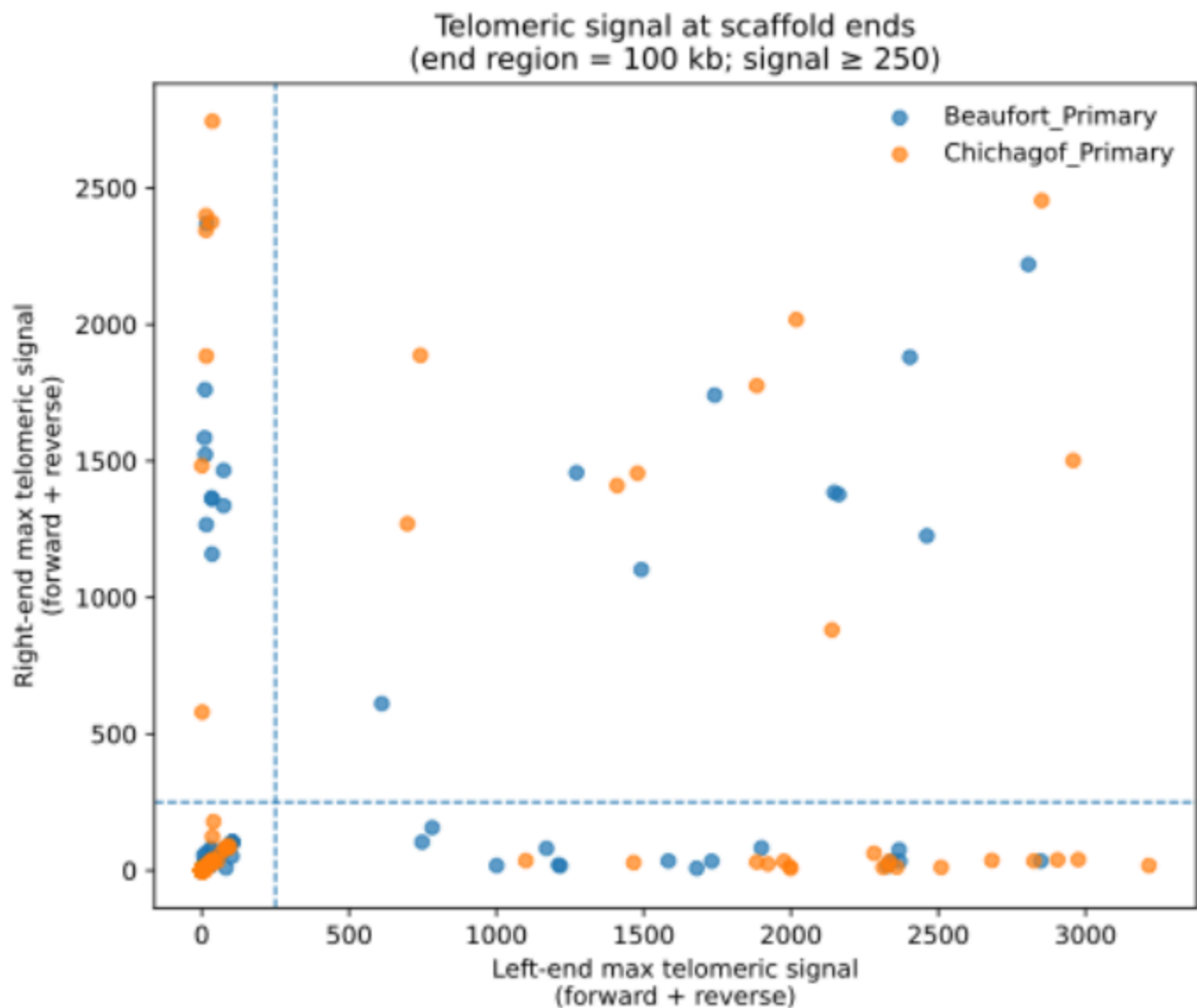


Figure 3

Terminal telomeric signal in the polar bear and brown bear genome assemblies. Scatter plots show the maximum telomeric repeat signal (forward + reverse) detected by TIDK within the terminal 200 kb of each scaffold. The new polar bear assembly (blue) is visualized alongside the new brown bear assembly (orange). Each point represents one scaffold, with the x- and y-axes indicating the strongest telomeric signal observed at the left and right scaffold ends, respectively. Dashed lines denote a conservative threshold (signal ≥ 250) used to identify strong telomeric enrichment. Scaffolds in the upper-right quadrant exhibit elevated telomeric signal at both ends, consistent with near telomere-to-telomere chromosome representation, whereas scaffolds with signal at only one or neither end likely correspond to truncated chromosome arms or internal fragments.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [UrsusScientificDataSuppl160926.docx](#)