Journal of Heredity, 2023, **114**, 549–560
https://doi.org/10.1093/jhered/esad042
Advance access publication 3 July 2023
Genome Resources

Genome Resources

A highly contiguous reference genome for the Steller's jay (*Cyanocitta stelleri*)

Phred M. Benham[1,2], Carla Cicero[1], Devon A. DeRaad[3], John E. McCormack[4], Robert K. Wayne[5,†], Merly Escalona[6], Eric Beraut[7], Mohan P.A. Marimuthu[8], Oanh Nguyen[8], Michael W. Nachman[1,2] and Rauri C.K. Bowie[1,2]

[1]Museum of Vertebrate Zoology, University of California Berkeley, Berkeley, CA, United States,
[2]Department of Integrative Biology, University of California Berkeley, Berkeley, CA, United States,
[3]Biodiversity Institute and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS, United States,
[4]Moore Laboratory of Zoology, Occidental College, Los Angeles, CA, United States,
[5]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA, United States,
[6]Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, CA, United States,
[7]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA, United States,
[8]DNA Technologies and Expression Analysis Core Laboratory, Genome Center, University of California-Davis, Davis, CA, United States

[†]Deceased.
Address correspondence to P.M. Benham at the address above, or e-mail: phbenham@gmail.com.

Corresponding Editor: Arun Sethuraman

Abstract

The Steller's jay is a familiar bird of western forests from Alaska south to Nicaragua. Here, we report a draft reference assembly for the species generated from PacBio HiFi long-read and Omni-C chromatin-proximity sequencing data as part of the California Conservation Genomics Project (CCGP). Sequenced reads were assembled into 352 scaffolds totaling 1.16 Gb in length. Assembly metrics indicate a highly contiguous and complete assembly with a contig N_{50} of 7.8 Mb, scaffold N_{50} of 25.8 Mb, and BUSCO completeness score of 97.2%. Repetitive elements span 16.6% of the genome including nearly 90% of the W chromosome. Compared with high-quality assemblies from other members of the family Corvidae, the Steller's jay genome contains a larger proportion of repetitive elements than 4 crow species (*Corvus*), but a lower proportion of repetitive elements than the California scrub-jay (*Aphelocoma californica*). This reference genome will serve as an essential resource for future studies on speciation, local adaptation, phylogeography, and conservation genetics in this species of significant biological interest.

Key words: California Conservation Genomics Project, Corvidae, transposable elements

Introduction

A striking blue body and conspicuous crest, coupled with raucous calls, make the Steller's jay (*Cyanocitta stelleri*) one of the most iconic birds of forested landscapes in western North America and Mesoamerica (Fig. 1a). The species is found primarily in mountainous regions from Alaska to Nicaragua across broad thermal and precipitation gradients (Fig. 1b; Walker et al. 2016), where it occupies habitat varying from oak woodlands to Douglas fir and redwood forest, mixed conifer forest, yellow pine forest, and pinyon–juniper woodlands (Fig. 1c). This broad ecological distribution has motivated important research on ecomorphological variation (Cicero et al. 2022) and how signaling behavior interacts with the environment to shape geographic variation in avian plumage (Brown 1963a). Further, the conspicuousness of the Steller's jay in many localities has made it an important model for studying the evolution of complex vocal communication (Hope 1980; Billings et al. 2017) and social behavior in birds (Brown 1963b; Kalinowski et al. 2015).

The Steller's jay is one of 128 species of crows, magpies, and jays in the globally distributed songbird family Corvidae (Winkler et al. 2020). The Steller's jay belongs to a monophyletic clade of jay genera distributed throughout the Americas that are characterized by extensive blue plumage. Within this group, mitochondrial (mtDNA) and nuclear sequence data place *Cyanocitta* (Steller's and blue jays) as sister to other North American jays in the genera *Aphelocoma* and *Gymnorhinus* (Ericson et al. 2005; Bonaccorso and Townsend Peterson 2007). Within Steller's jays, extensive intraspecies variation has resulted in the recognition of 16 subspecies which can be divided into 3 main groups (Walker et al. 2016). The coastal group includes 4 subspecies distributed from Alaska south through California and generally west of the Cascade and Sierra Nevada mountain ranges. The interior group comprises 4 subspecies found east of the Cascades and in the Rocky Mountains south to northern Mexico. Finally, the Central American group includes 8 subspecies distributed in the highlands of Mexico, Guatemala,

Received April 14, 2023; Accepted June 29, 2023

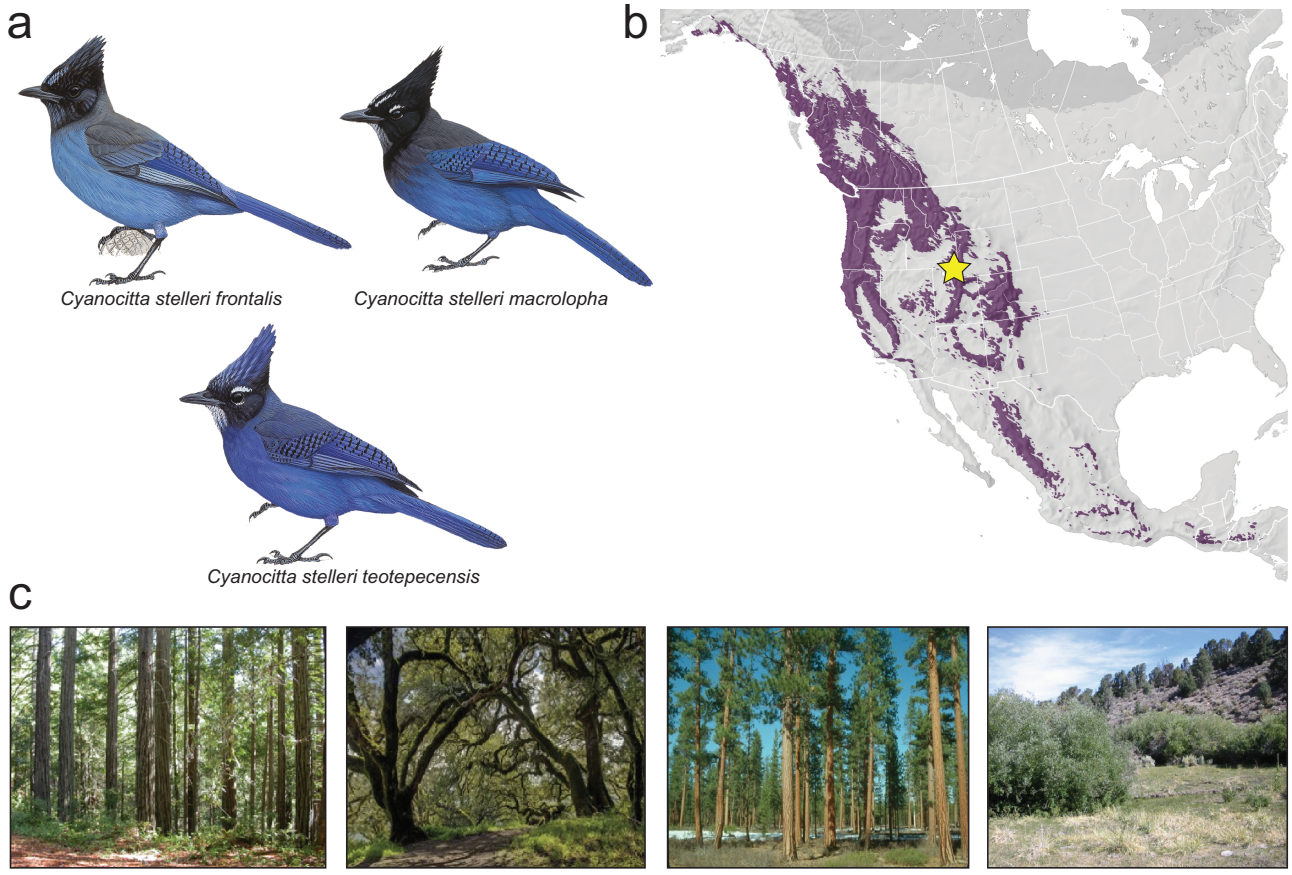

Fig. 1. a) Geographic variation among different Steller's jay (*Cyanocitta stelleri*) subspecies. *C. s. frontalis* is characteristic of the short-crested birds with blue forehead streaks found along the Pacific coast and in the Sierra Nevada of California. The *C. s. macrolopha* group of the Rocky Mountains and northern Mexico includes long-crested birds with white forehead stripes and a white line above the eye. Birds of Central American mountains possess bluer crests that are similar to *C. s. teotepecensis* illustrated here. b) Distribution map of the species (map data from https://ebird.org/). The yellow star marks the location of the contact zone between coastal plus northwest interior and Rocky Mountain populations. c) Range of habitats occupied by different Steller's jay populations in western North America. From left to right: redwood forest, oak woodland, yellow pine forest, and pinyon–juniper habitats (photos by Carla Cicero). Illustrations reproduced from https://birdsoftheworld.org with permission from Lynx Edicions (artist Brian Small).

Honduras, El Salvador, and Nicaragua. Recent phylogeographic work on the species points to a major genetic break in both mtDNA (7.8% divergent) and microsatellites between coastal plus northwest interior populations and those from the Rocky Mountains, with birds from mountain ranges in southeastern Idaho and northeastern Utah occurring in a narrow secondary contact zone (Cicero et al. 2022). Based on genetic, morphological, and ecological differences, the authors suggested that Rocky Mountain and coastal/northwest interior jay populations likely comprise distinct species. Other genetically distinct Steller's jay lineages include the subspecies from the Haida Gwaii archipelago in British Columbia (*C. s. carlottae*; Burg et al. 2005) and birds from the central California coast (*C. s. carbonacea*; Cicero et al. 2022).

The Steller's jay is widespread and generally common throughout its range, making it a species of least conservation concern (BirdLife International 2023). However, relative abundance during the breeding season varies widely across the species' range (Walker et al. 2016), and data from the citizen science project eBird show declines in certain portions of the range, notably central coastal to southern California and east of the Cascade and Sierra Nevada ranges (Fink et al. 2022; https://science.ebird.org/en/status-and-trends/species/stejay/trends-map). Although nonthreatened, the Steller's jay has been included in the California Conservation Genomics Project (CCGP; Shaffer et al. 2022) because the inclusion of common and widespread species is important for contextualizing genetic diversity patterns in species of greater conservation concern. Here, we present a draft assembly of the Steller's jay genome generated by the CCGP using PacBio HiFi long-read and Omni-C chromatin-proximity sequencing data. This draft assembly, the first for the genus *Cyanocitta*, represents a significant expansion of the taxonomic breadth of high-quality reference genomes available within the family Corvidae, which has become a model for studies of avian genome evolution in recent years (e.g. Weissensteiner et al. 2020; Warmuth et al. 2022). The public availability of this high-quality assembly will catalyze future projects investigating genomic evolution at multiple scales ranging from intraspecific (*C. stelleri*) to family (Corvidae) levels.

Methods

Biological materials

A liver tissue sample was obtained from a female Steller's jay (*Cyanocitta stelleri frontalis*) collected on 5 September 2020. This bird was captured at Mitsui Ranch, Sonoma Mountain, Sonoma Co., California (38.33334°N; 122.58060°W). The

individual was collected with approval of the California Department of Fish and Wildlife (permit #: SCP-458) and the U.S. Fish and Wildlife Service (permit #: MB153526-0). The bird was captured and euthanized using methods approved by the University of California, Berkeley IACUC (AUP-2015-10-8045-1). A voucher specimen has been deposited at the Museum of Vertebrate Zoology, Berkeley, California (https://arctos.database.museum/guid/MVZ:Bird:193367).

PacBio HiFi library preparation

High-molecular-weight (HMW) genomic DNA (gDNA) was extracted from 30 mg of liver tissue using the Nanobind Tissue Big DNA kit as per the manufacturer's instructions (Pacific BioSciences [PacBio], Menlo Park, California). The extracted HMW DNA was further purified using the high-salt-phenol-chloroform method (PacBio). The DNA purity was estimated using absorbance ratios (260/280 = 1.76 and 260/230 = 2.01) on a NanoDrop ND-1000 spectrophotometer and a final DNA yield of 4.7 µg was quantified with a Quantus Fluorometer (QuantiFluor ONE dsDNA Dye assay; Promega, Madison, Wisconsin). A Femto Pulse system (Agilent, Santa Clara, California) was used to estimate the size distribution of the HMW DNA with 62% of DNA fragments found to exceed 100 kb in length.

The HiFi SMRTbell library was constructed using the SMRTbell Express Template Prep Kit v2.0 (PacBio, Cat. #100-938-900) according to the manufacturer's instructions. HMW gDNA was sheared to a target DNA size distribution of 15 to 20 kb using the Diagenode Megaruptor 3 system (Diagenode, Belgium; Cat. B06010003). The sheared gDNA was concentrated using 0.45X of AMPure PB beads (PacBio, Cat. #100-265-900) before the removal of single-strand overhangs at 37 °C for 15 min. This was followed by further enzymatic steps that performed DNA damage repair at 37 °C for 30 min, end repair and A-tailing at 20 °C for 10 min then 65 °C for 30 min, ligation of overhang adapter v3 at 20 °C for 60 min and ligase inactivation at 65 °C for 10 min, and the final step of nuclease treatment at 37 °C for 1 h. The SMRTbell library was purified and concentrated with 0.45× Ampure PB beads prior to size selection of fragments greater than 7 to 9 kb using the BluePippin/PippinHT system (Sage Science, Beverly, Massachusetts; Cat. #BLF7510/HPE7510). The 15 to 20 kb average HiFi SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, California) using two 8 M SMRT cells, Sequel II sequencing chemistry 2.0, and 30-h movies each on a PacBio Sequel II sequencer.

Omni-C library preparation and sequencing

The Omni-C library was prepared using the Dovetail Omni-C Kit (Dovetail Genomics, Scotts Valley, California) according to the manufacturer's protocol with slight modifications. First, specimen tissue (liver, ID: RCKB 2456) was thoroughly ground with a mortar and pestle while cooled with liquid nitrogen. Subsequently, chromatin was fixed in place in the nucleus through DNA crosslinking. Crosslinking is achieved by incubating the ground tissue for 10 min at room temperature while rotating the tube in phosphate-buffered saline and disuccinimidyl glutarate followed by an additional 10 min after adding formaldehyde. The suspended chromatin solution was then passed through 100 and 40 µm cell strainers to remove large debris. Fixed chromatin was digested for 30 min at 30 °C in an agitating thermal mixer set at 1,250 rpm using 5× the manufacturers recommended volume of DNase 1 until a suitable fragment length distribution of DNA molecules was obtained. Chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter containing ends. After proximity ligation crosslinks were reversed and the DNA was purified from proteins through a proteinase K and SDS incubation on an agitating thermal mixer set at 1,250 rpm for 15 min at 55 °C followed by 45 min at 68 °C. Purified DNA was treated to remove biotin that was not internal to ligated fragments. An NGS library was generated using an NEB Ultra II DNA Library Prep kit (New England Biolabs [NEB], Ipswich, Massachusetts) with an Illumina compatible y-adaptor. Biotin-containing fragments were then captured using streptavidin beads. The post capture product was split into 2 replicates prior to PCR enrichment to preserve library complexity with each replicate receiving unique dual indices. The library was sequenced at the Vincent J. Coates Genomics Sequencing Lab (Berkeley, California) on an Illumina NovaSeq 6000 platform (Illumina, San Diego, California) to generate approximately 100 million 2 × 150 bp read pairs per GB genome size.

Nuclear genome assembly

We assembled the *Cyanocitta stelleri* genome following the CCGP assembly pipeline Version 5.0 (see Table 1 for all software used, alongside version numbers and nondefault parameters used). First, we removed remnant adapter sequences from the PacBio HiFi dataset using HiFiAdapterFilt (Sim et al. 2022). The filtered PacBio HiFi reads and the Omni-C dataset were used to generate an initial dual assembly in HiFiasm in Hi-C mode (Cheng et al. 2021). HiFiasm (Hi-C) produces dual, partially phased assemblies (http://lh3.github.io/2021/10/10/introducing-dual-assembly) that we tagged as a primary and an alternate assembly based on contig-level contiguity metrics. We next identified sequences corresponding to haplotypic duplications, contig overlaps, and repeats on the primary assembly with purge_dups (Guan et al. 2020) and transferred these sequences to the alternate assembly.

Alignment of the Omni-C data to both assemblies followed the Arima Genomics Mapping Pipeline (https://github.com/ArimaGenomics/mapping_pipeline) and both assemblies were scaffolded with SALSA (Ghurye et al. 2017, 2018). Both genome assemblies were curated manually by iteratively generating and analyzing their corresponding Omni-C contact maps. To generate the contact maps we aligned the Omni-C data with BWA-MEM (Li 2013), identified ligation junctions, and generated Omni-C pairs using pairtools (Goloborodko et al. 2018). We generated a multiresolution Omni-C matrix with cooler (Abdennur and Mirny 2020) and balanced it with hicExplorer (Ramírez et al. 2018). We used HiGlass (Kerpedjiev et al. 2018) and the PretextSuite (https://github.com/wtsi-hpag/PretextView; https://github.com/wtsi-hpag/PretextMap; https://github.com/wtsi-hpag/PretextSnapshot) to visualize contact maps and identify misassemblies and misjoins. Some of the remaining gaps (joins generated during scaffolding) were closed using the PacBio HiFi reads and YAGCloser (https://github.com/merlyescalona/yagcloser). Finally, we checked for contamination using the BlobToolKit Framework (Challis et al. 2020).

Mitochondrial genome assembly

We assembled the mitochondrial genome of *C. stelleri* from PacBio HiFi reads using MitoHiFi (Allio et al. 2020; Uliano-Silva et al. 2021). The mitochondrial sequence of *Pica pica melanotos* (NCBI:MT792356.1; Kryukov et al. 2020) was used as the starting reference sequence. After completion of

Table 1. Assembly pipeline and software used.

Assembly	Software and options[a]	Version
Filtering PacBio HiFi adapters	HiFiAdapterFilt	Commit 64d1c7b
K-mer counting	Meryl (k = 21)	1
Estimation of genome size and heterozygosity	GenomeScope	2
De novo assembly (contiging)	HiFiasm (Hi-C mode, –primary, output p_ctg.hap1, p_ctg.hap2)	0.16.1-r375
Identification of haplotypic duplications	purge_dups (custom cutoffs: 1,18,28,29,67,144)	1.2.5
Scaffolding		
Omni-C data alignment	Arima Genomics Mapping Pipeline	Commit 2e74ea4
Omni-C scaffolding	SALSA (-DNASE, -i 20, -p yes)	2
Gap closing	YAGCloser (-mins 2 -f 20 -mcc 2 -prt 0.25 -eft 0.2 -pld 0.2)	Commit 0e34c3b
Omni-C contact map generation		
Short-read alignment	BWA-MEM (-5SP)	0.7.17-r1188
SAM/BAM processing	samtools	1.11
SAM/BAM filtering	pairtools	0.3.0
Pairs indexing	pairix	0.3.7
Matrix generation	cooler	0.8.10
Matrix balancing	hicExplorer (hicCorrectmatrix correct --filterThreshold -2 4)	3.6
Contact map visualization	HiGlass	2.1.11
	PretextMap	0.1.4
	PretextView	0.1.5
	PretextSnapshot	0.0.3
Genome quality assessment		
Basic assembly metrics	QUAST (--est-ref-size)	5.0.2
Assembly completeness	BUSCO (-m geno, -l aves)	5.0.0
	Merqury	2020-01-29
Circos Assembly Consistency (Jupiter) plot	JupiterPlot (ng = 80, m = 1,000,000)	1
Identification of repetitive elements	RepeatMasker	4.1.2
	RepeatModeler	2
	Minimap	2
	cd-hit-est	4.8.1
	mafft	7.49
	Aliview	1.28
	TE-Aid	0-dev
	EMBOSS	6.6.0
Gene annotation	Liftoff	1.6.3
Organelle assembly		
Mitogenome assembly	MitoHiFi (-r, -p 50, -o 1)	2 Commit c06ed3e
Contamination screening		
Local alignment tool	BLAST+ (-db nt, -outfmt "6 qseqid staxids bitscore std," -max_target_seqs 1, -max_hsps 1, -evalue 1e-25)	2.1
General contamination screening	BlobToolKit	2.3.3

Software citations are listed in the text.
[a]Options detailed for nondefault parameters.

the nuclear genome, we searched for matches of the resulting mitochondrial assembly sequence in the nuclear genome assembly using BLAST+ (Camacho et al. 2009) and filtered out contigs and scaffolds from the nuclear genome with a percentage of sequence identity >99% and size smaller than the mitochondrial assembly sequence.

Genome assembly assessment

We generated k-mer counts from the PacBio HiFi reads using meryl (https://github.com/marbl/meryl). Approximations of genomic features including genome size, heterozygosity, and repeat content were made by optimizing model fit to the k-mer count profile in GenomeScope2.0 (Ranallo-Benavidez et al. 2020). General contiguity metrics were obtained in QUAST (Gurevich et al. 2013). Genome quality and functional completeness were evaluated in BUSCO (Manni et al. 2021) using the 8,338 genes contained within the Aves ortholog database, aves_odb10. Assessment of base level accuracy (QV) and k-mer completeness was performed using the previously generated meryl database and merqury (Rhie et al. 2020). We

further estimated genome assembly accuracy via the BUSCO gene set frameshift analysis using the pipeline described in Korlach et al. (2017).

Measurements of the size of the phased blocks are based on the size of the contigs generated by HiFiasm on Hi-C mode. We follow the quality metric nomenclature established by Rhie et al. (2021), with the genome quality code $x{\cdot}y{\cdot}P{\cdot}Q{\cdot}C$, where $x = \log_{10}$[contig NG50]; $y = \log_{10}$[scaffold NG50]; $P = \log_{10}$[phased block NG50]; Q = Phred base accuracy QV (quality value); C = % genome represented by the first "n" scaffolds, following a karyotype of 2n = 78 estimated from other species in the same genus, *Cyanocitta cristata* (Jovanović and Atkins 1969). Quality metrics for the notation were calculated on the primary assembly.

We visualized higher level synteny between the Steller's jay scaffolds and chromosomes of the New Caledonian crow (*Corvus moneduloides*) genome (Feng et al. 2020; NCBI accession: GCA_009650955.1) and generated a Jupiter plot using the JupiterPlot pipeline (https://github.com/JustinChu/JupiterPlot). We mapped the longest scaffolds representing 80% of the draft assembly to crow chromosomes exceeding 1 Mb in length. We additionally performed a liftover using Liftoff with default parameters (Shumate and Salzberg 2021) to transfer gene annotations from the New Caledonian crow assembly to the Steller's jay assembly.

Repeat annotation

We performed de novo repeat annotation of the draft Steller's jay reference assembly using the program RepeatModeler2 with the ltrstruct option selected to improve identification of long terminal repeat retrotransposons (LTR) elements (Flynn et al. 2020). We next prioritized LTR and unknown elements for manual curation that were at least 1,000 bp in length and had at least 5 blast hits in the genome. We focused on these classes of elements as RepeatModeler is known to generate incomplete LTR sequences and misidentify LTRs as unknown (e.g. Suh et al. 2018). For each LTR consensus sequence, we used blastn (Camacho et al. 2009) to identify other members of each transposable element (TE) family in the genome, added 2,000 bp of flanking sequence to both ends of each blastn hit, aligned extended sequences with mafft (Katoh and Standley 2013), and visualized the alignment in Aliview (Larsson 2014). We confirmed the completeness of LTR elements based on the presence of canonical 5′ TG and 3′ CA dinucleotides at the termini of LTRs. A consensus sequence of the trimmed multiple sequence alignment was then generated using the cons tool in EMBOSS (Rice et al. 2000). For sequences labeled as unclassified, we used blastn to check for significant hits with protein coding genes and removed elements from the repeat library that had high sequence similarity with a known gene for over 80% of their length. The program TE-Aid (https://github.com/clemgoub/TE-Aid) was also used to explore structural properties and presence of open reading frames for the expected proteins characteristic of each class of TE element. Finally, we used cd-hit-est (Li and Godzik 2006) to cluster sequences belonging to the same family within the Steller's jay repeat library following the 80-80-80 rule (Wicker et al. 2007). Specifically, this rule considers consensus sequences to be the same family if they are 80 bp in length and share >80% similarity over >80% of their length. This produced a final repeat library with consensus sequences for 436 elements.

The TE sequences identified using repeat modeler were then compared against a library of avian TE elements downloaded from repbase using cd-hit-est with the same settings as above. The resulting repeat library was used to annotate TE diversity in the Steller's jay assembly using RepeatMasker (Smit et al. 2013–2015). We next assigned scaffolds to the Z, W, or autosomal chromosomes based on homology with the New Caledonian crow chromosomes using Minimap2 (Li 2018). We then ran RepeatMasker separately for the autosomes, the Z chromosome, and the W chromosome. For autosomes and sex chromosomes, the calcDivergenceFromAlign.pl script in RepeatMasker was used to estimate Kimura 2-parameter (K2P) distances of each TE element from their consensus sequence. To visualize temporal patterns of TE activity, the K2P distances were used to generate barplots for the LTR, SINE, LINE, and DNA classes of TEs. Finally, we used the merged Steller's jay and avian TE library to annotate repeat content in the highly contiguous genomes (contig N_{50} >1 Mb) of 5 other Corvidae species using RepeatMasker. These genomes included the Eurasian jackdaw (*Corvus monedula*; Weissensteiner et al. 2020), New Caledonian crow (*C. moneduloides*; Feng et al. 2020), Hawaiian crow (*C. hawaiiensis*; Rhie et al. 2021), 2 separate hooded crow (*C. cornix*) assemblies (Weissensteiner et al. 2020; Warmuth et al. 2022), and the California scrub-jay (*Aphelocoma californica*) also generated by the CCGP (DeRaad et al. in review).

Results

Sequencing data

The Omni-C and PacBio HiFi sequencing libraries generated 131.4 million read pairs and 2.8 million reads, respectively. The latter yielded 48.7-fold coverage (N_{50} read length 19,545 bp; minimum read length 56 bp; mean read length 18,405 bp; maximum read length of 58,267 bp) based on the GenomeScope2.0 genome size estimation of 1.07 Gb. A sequencing error rate of 0.214% for PacBio HiFi reads was estimated using GenomeScope2.0. The k-mer spectrum based on PacBio HiFi reads show a bimodal distribution with 2 major peaks at ~24- and 48-fold coverage, where lower and higher coverage peaks correspond to heterozygous and homozygous states of a diploid species, respectively (Fig. 2a).

Assembly metrics

The primary Steller's jay assembly includes 352 scaffolds with an assembly length of 1.16 Gb (Table 2). All descriptive statistics including contig N_{50} (7.85 Mb), scaffold N_{50} (25.8 Mb), and the average number of Ns per 100 kb (1.92), point to a highly contiguous genome assembly (see Table 2 for detailed summary statistics). Graphical representation of quality metrics and the Hi-C contact maps for the primary assembly are shown in Fig. 2b and c, respectively (see Supplementary Fig. S1 for the alternate assembly). BUSCO scores further underscore the high degree of completeness, with 97.2% of the 8,338 orthologs in the aves_odb10 database present in the jay genome; 96.7% of these genes were found as a single copy, 0.5% had duplicated copies, and 0.5% were fragmented. Only 2.3% of avian orthologs were found to be completely missing from the assembly. Similarly, 98.3% of the 20,680 genes annotated in the New Caledonian crow genome were

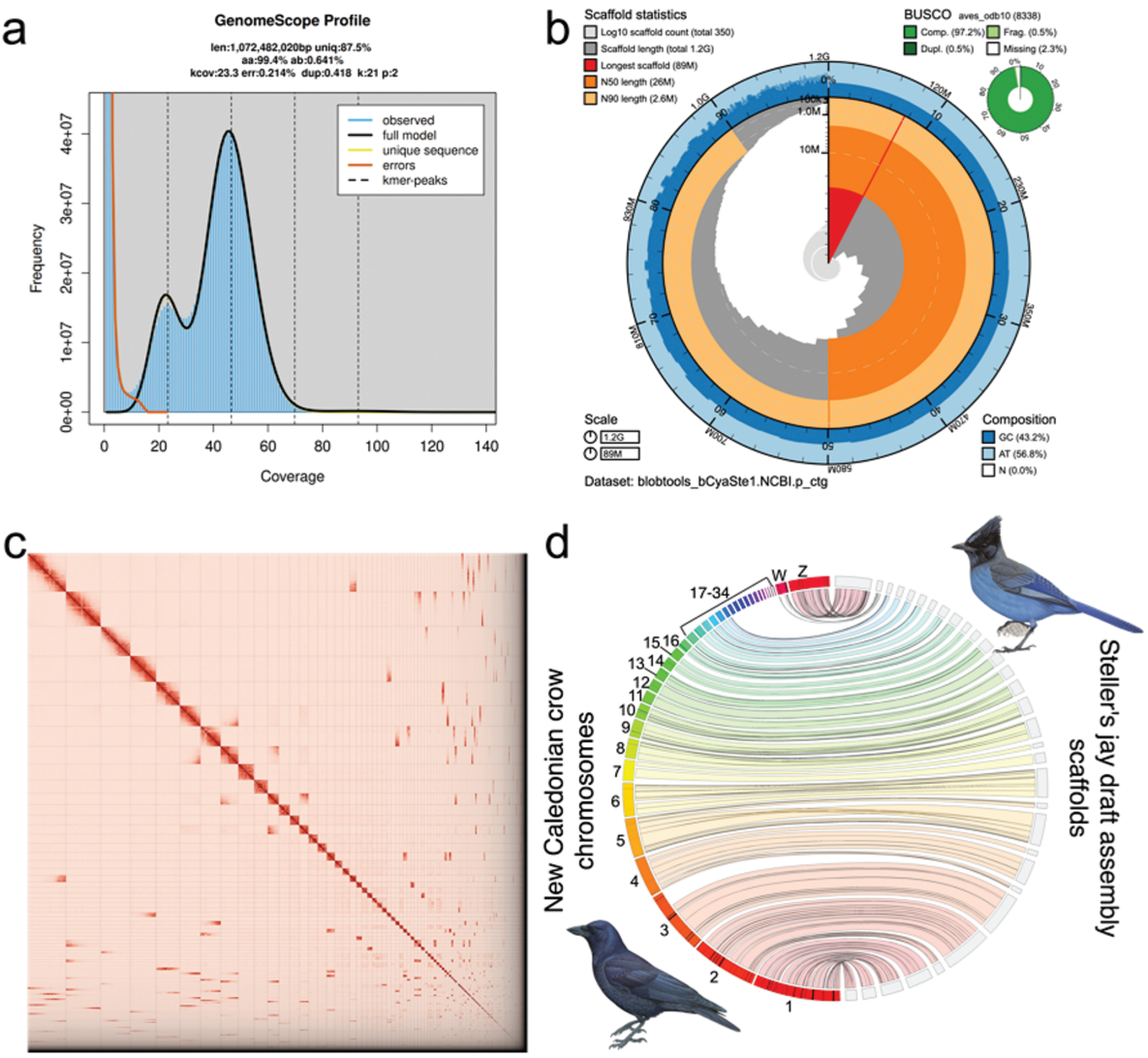

Fig. 2. Visual overview of genome sequencing and assembly quality. a) K-mer spectra output generated from PacBio HiFi data without adapters using GenomeScope2.0. The bimodal pattern observed corresponds to a diploid genome and the k-mer profile matches that of low (<1%) heterozygosity. The left-hand peak at lower coverage and frequency derives from k-mers in heterozygous regions, while the higher coverage right hand peak k-mers stem from homozygous regions. b) BlobToolKit Snail plot shows quality metrics for the primary assembly (also see Table 2) of the Steller's jay (*Cyanocitta stelleri*) genome. The plot circle represents the full size of the assembly. From the inside-out, the central plot covers length-related metrics. The red line represents the size of the longest scaffold; all other scaffolds are arranged in size-order moving clockwise around the plot and drawn in gray starting from the outside of the central plot. Dark and light orange arcs show the scaffold N_{50} and scaffold N_{90} values. The central light gray spiral shows the cumulative scaffold count with a white line at each order of magnitude. White regions in this area reflect the proportion of Ns in the assembly; the dark versus light blue area around it shows mean, maximum, and minimum GC vs. AT content at 0.1% intervals. c) Hi-C contact map for the primary assembly generated using PretextSnapshot. Hi-C contact maps translate proximity of genomic regions in 3D space to contiguous linear organization. Each cell in the contact map corresponds to sequencing data supporting the linkage (or join) between 2 such regions. d) Jupiter plot comparing higher level synteny and degree of completeness between the New Caledonian crow (*Corvus moneduloides*) genome and the Steller's jay genome. Crow chromosomes are on the left (colored) and jay scaffolds are on the right (light gray). Twists represent reversed orientation of scaffolds between assemblies. Illustrations reproduced from https://birdsoftheworld.org with permission from Lynx Edicions (jay artist: Brian Small; crow artist: David Quinn).

successfully mapped to the Steller's jay genome. A Jupiter plot shows jay scaffolds mapping to most chromosomes of the New Caledonian crow genome, with little evidence for inversions or translocations that may be indicative of misassemblies (Fig. 2d). Compared with other genome assemblies from the family Corvidae, the contiguity metrics for this assembly places it in the class of the most contiguous crow (*Corvus*) genomes and the California scrub-jay genome also generated by the CCGP (Table 3).

Repeat annotation

16.72% of the Steller's jay genome was masked by RepeatMasker v. 4.1.2 (Table 4). This included 137.2 Mb (11.79%) of the genome masked by interspersed repeats and 41.4 Mb (3.56%) masked by satellites and other simple repeats. Endogenous retroviral elements (7.85%) and LINE elements of the CR1 family (3.82%) were the 2 most abundant TEs annotated in the jay genome. DNA transposons and SINE elements were each found in less than 1% of the genome.

Table 2. Metrics for the primary and alternate assemblies of the Steller's jay (*Cyanocitta stelleri*) genome.

BioProjects and Vouchers	**CCGP NCBI BioProject**		**PRJNA720569**			
	Genera NCBI BioProject		PRJNA765809			
	Species NCBI BioProject		PRJNA777159			
	NCBI BioSample		SAMN30523672			
	Specimen identification		RCKB 2456			
	NCBI Genome accessions		Primary		Alternate	
	Assembly accession		JANXIP000000000		JANXIQ000000000	
	Genome sequences		GCA_026167965.1		GCA_026168045.1	
Genome Sequence	PacBio HiFi reads	Run	1 PACBIO_SMRT (Sequel II) run: 2.8 M spots, 52.2 G bases, 39.3 Gb			
		Accession	SRX19143332			
	Omni-C Illumina reads	Run	2 ILLUMINA (Illumina NovaSeq 6000) runs: 131.4 M spots, 39.7 G bases, 12.6 Gb			
		Accession	SRX19143333, SRX19143334			
Genome Assembly Quality Metrics	Assembly identifier (quality code[a])				bCyaSte1(6.7.6.Q63.C79)	
	HiFi read coverage[b]				48.71×	
			Primary		Alternate	
	Number of contigs		575		1,162	
	Contig N_{50} (bp)		7,854,909		6,633,807	
	Contig NG50[b]		8,422,260		7,718,774	
	Longest contigs		48,674,044		40,161,822	
	Number of scaffolds		352		962	
	Scaffold N_{50}		25,842,608		27,786,026	
	Scaffold NG50[b]		31,261,391		38,692,504	
	Largest scaffold		88,620,470		86,024,699	
	Size of final assembly		1,162,964,953		1,177,709,819	
	Phased block NG50[b]		31,261,391		38,692,504	
	Gaps per Gbp (# gaps)		192 (223)		170 (200)	
	Indel QV (frameshift)		41.155972		41.20408138	
	Base pair QV		63.879		62.7316	
			Full assembly = 63.2639			
	k-mer completeness		94.3357		88.477	
			Full assembly = 99.6011			
	BUSCO completeness (aves_odb10) n = 8,338		C	S	D	F	M
		P[c]	97.20%	96.70%	0.50%	0.50%	2.30%
		A[c]	93.10%	92.60%	0.50%	0.50%	6.40%
	Organelles		# Partial mitochondrial sequence			JANXIP010000352.1	

[a]Assembly quality code $x{\cdot}y{\cdot}P{\cdot}Q{\cdot}C$ derived notation, from Rhie et al. (2021). $x = \log_{10}$[contig NG50]; $y = \log_{10}$[scaffold NG50]; $P = \log_{10}$[phased block NG50]; Q = Phred base accuracy QV (quality value); C = % genome represented by the first "n" scaffolds, following a known karyotype for *C. stelleri* of 2n = 78. Quality code for all the assembly denoted by primary assembly (bCyaSte1.0.p).
[b]Read coverage and NGx statistics have been calculated based on the estimated genome size of 1.07 Gb.
[c](P)rimary and (A)lternate assembly values.

Repeat content spanned 15.63% of autosomal DNA and was characterized by an older proliferation of LINE elements (K2P divergence of 20% to 25%) followed by more recent TE activity dominated by LTR elements (Fig. 3). The Z chromosome exhibited a similar landscape of TE activity with a slightly larger percentage of repeat DNA (21.93%). Repetitive DNA was densest on the W chromosome with 89.86% of our assembled scaffold spanned by repetitive elements. Relatively recent activity of LTR elements corresponds to the bulk of W chromosome repeat content (85.58%).

Compared with crows in the genus *Corvus*, repetitive elements comprised a greater percentage of the Steller's jay and California scrub-jay genomes (Fig. 4). California scrub-jay exhibited both the highest percentage (24.6%; Fig. 4a) and length of repeat sequence (331.7 Mb) as well as the longest assembly length (1.35 Gb) of all species analyzed. Much of the difference in repetitive content between crows and jays was driven by expansions of LTR and satellite DNA in jays. In contrast, a near identical proportion of LINE DNA was observed across the genomes of all analyzed corvid species.

Table 3. Assembly metrics of genome assemblies for other Corvidae species deposited on GenBank as of May 2023.

English name	Scientific name	C-value (Gb)	Length (Gb)	GC (%)	# contigs	Contig N_{50} (Mb)	Contig L_{50}	# scaffolds	Scaffold N_{50} (Mb)	Scaffold L_{50}
Steller's jay	*Cyanocitta stelleri*	NA	1.16	43.18	575	7.85	45	352	25.84	11
Florida scrub-jay	*Aphelocoma coerulescens*	1.56	1.06	41.3	22,008	0.13	1,701	1,701	7.63	42
California scrub-jay	*Aphelocoma californica*	NA	1.35	44.6	1,665	11.53	28	1,393	66.14	7
Eurasian magpie	*Pica pica*	1.25	1.06	41.5	72,252	0.04	6,712	24,674	73.42	5
hooded crow	*Corvus cornix*	NA	1.03	42.12	472	8.55	33	47	73.73	5
American crow	*Corvus brachyrhynchos*	1.26	1.09	42	89,646	0.03	10,509	10,547	6.95	49
Mariana crow	*Corvus kubaryi*	NA	1.06	42.2	12,022	0.21	1,437	1,443	17.89	18
New Caledonian crow	*Corvus moneduloides*	NA	1.11	42.15	394	11.52	30	105	74.69	5
Large-billed crow	*Corvus Macrorhynchos*	NA	1.02	41.9	48,959	0.05	5,217	16,523	0.13	2,023
Eurasian jackdaw	*Corvus monedula*	1.25	1.04	42.15	552	23.26	15	41	73.77	5
Hawaiian crow	*Corvus hawaiiensis*	NA	1.15	43.32	482	23.06	15	188	76.27	6

Table 4. Estimates of the total number of elements, total length spanned, and the percentage of the Steller's jay genome occupied by each class of repetitive element.

Repetitive element class		Total elements	Length (bp)	Percentage genome
Retroelements		277,846	137,180,455	11.79
SINEs		6,374	802,539	0.07
	MIRs	3,700	434,282	0.04
LINEs		149,655	45,064,039	3.87
	L3/CR1	143,593	44,472,559	3.82
LTR elements		121,817	91,313,877	7.85
	ERVL	28,826	19,002,084	1.63
	ERV-Class I	33,879	33,505,233	2.88
	ERV-Class II	34,358	25,750,296	2.21
DNA transposons		4,894	1,419,621	0.12
	hAT-Charlie	951	682,599	0.06
Other repetitive elements				
Small RNA		1,019	796,462	0.07
Satellites		34,381	25,910,028	2.23
Simple repeats		220,253	12,022,023	1.03
Low complexity		44,096	2,702,753	0.23
Unclassified		33,690	15,030,475	1.29
Total bases masked			194,407,346	16.72

Repeat elements were identified by RepeatMasker using a TE library that included all bird repeats from the RepeatMasker database and consensus sequences of elements generated de novo for the Steller's jay genome with RepeatModeler2.

Discussion

We generated the first highly contiguous reference genome assembly for the Steller's jay (*C. stelleri*). In the context of a recent survey of over 500 avian genomes (Bravo et al. 2021), the contig N_{50} and scaffold N_{50} for this assembly were in the 94th and 58th percentile of all surveyed genomes, respectively. Prior to this study, genome sequencing efforts within the family Corvidae focused heavily on the genus *Corvus* (Table 3); with only 3 publicly available non-*Corvus* genome assemblies existing within Corvidae (*Pica pica*, *Aphelocoma coerulescens*, and *A. californica*). The most recent genomes

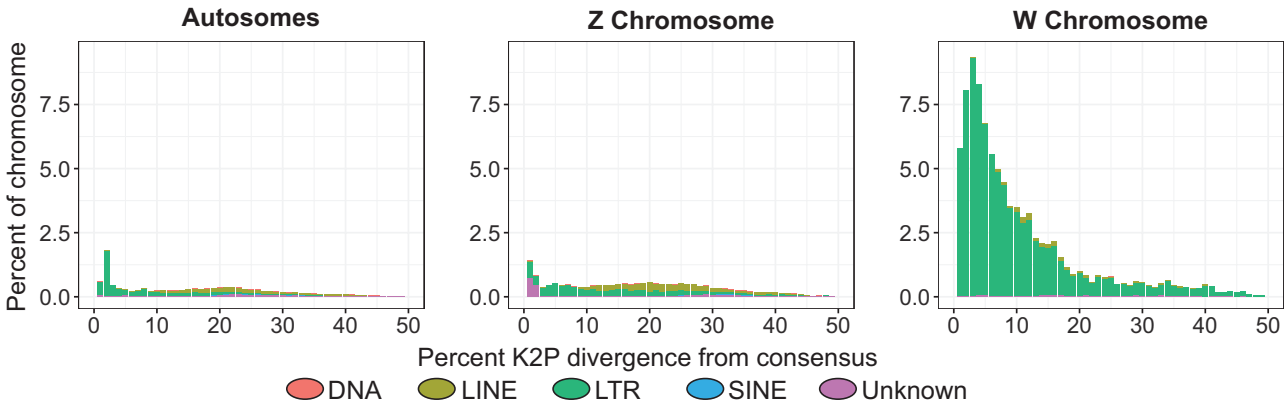

Fig. 3. TE landscapes for the autosomal, Z, and W chromosomes. Percent divergence on the *x* axis was calculated as the percent Kimura 2-parameter (K2P) distance excluding CpG sites. The abundance of TEs in each percent divergence bin was normalized as a percentage of the chromosome length on the *y* axis.

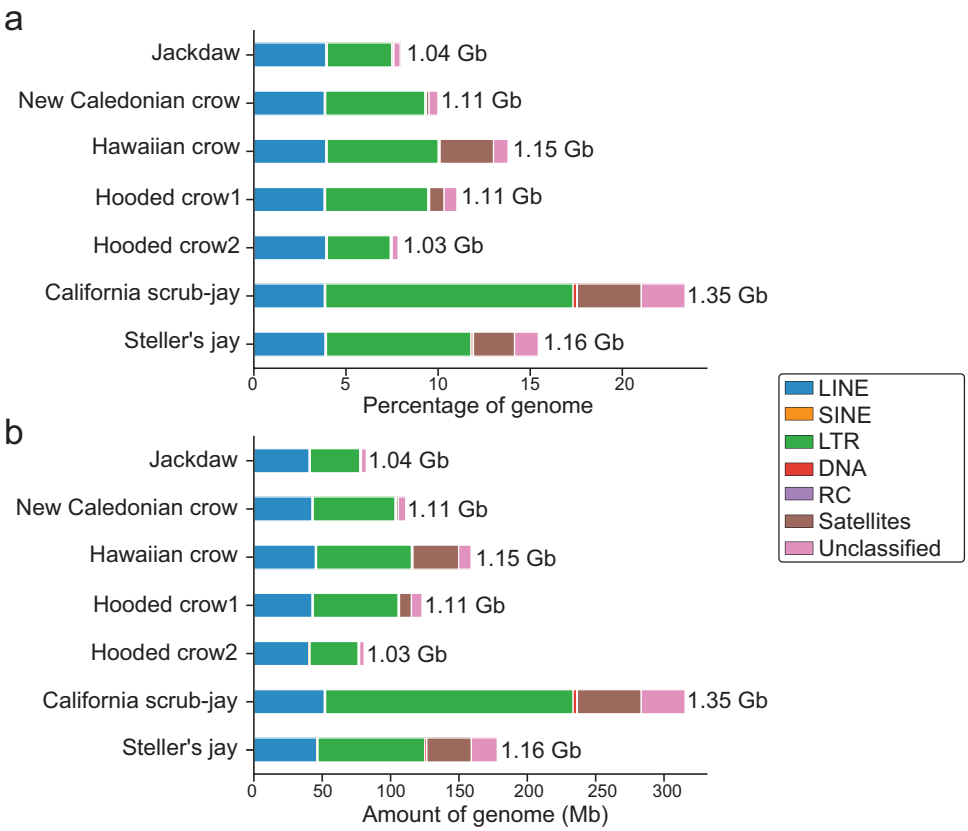

Fig. 4. a) Percent of genome comprising interspersed repeats, including retroelements (LINE, SINE, and LTR), DNA transposons (DNA), rolling-circles (RC), satellites, and unclassified elements. Draft assemblies produced by the CCGP are the 2 bottom genomes in the bar chart. b) Total number of base pairs in the genome spanned by each class of TE. Each bar is annotated with the length of each genome assembly in Gb. Hooded crow1 is an Oxford Nanopore assembly of a female bird. Hooded crow2 is an assembly based on PacBio and Hi-C sequencing of a male bird.

generated for *Corvus* crows are among the highest quality genomes generated for birds to date. Assemblies of hooded, Hawaiian, and New Caledonian crows plus the Eurasian jackdaw exhibit contig N_{50}s ranging from 8.5 to 23.2 Mb and scaffold N_{50}s exceeding 73.7 Mb. The Steller's jay assembly presented here comes in just under the contiguity range of these exceptionally high-quality *Corvus* genomes, and represents a marked improvement (>50× increase in contig N_{50} length and 3-fold increase in scaffold N_{50} length) over the previous most contiguous assembly for a jay species (Florida scrub-jay; Feng et al. 2020). This Steller's jay genome assembly expands the taxonomic breadth of high-quality genomes available within Corvidae, a family that has become a model for understanding the dynamics of avian genome evolution.

TEs play a critical role in the evolution of genomic architecture (Ågren and Wright 2011). The abundance of TEs and other repeats was higher in the Steller's jay relative to 4

Corvus species. Hawaiian crow was the most similar in percent repeat content (15.01%) and also had a similar assembly length (1.15 Gb; Fig. 3). The California scrub-jay exhibited higher repeat content than any of the other Corvidae species analyzed (24.6%) and was also the longest assembly (1.35 Gb). The influence of repeat content on animal genome size is well established (Kidwell 2002; Elliott and Gregory 2015). Indeed, the association between assembly length and repeat content observed here is in line with variation in independent estimates of genome size derived from flow cytometry and other methods. On average, members of the genus *Corvus* have smaller genome size (1.25 Gb) than estimates of 1.33 Gb for the blue jay (*C. cristata*; no estimate available for Steller's jay) or 1.53 Gb for the Florida scrub-jay, a congener of the California scrub-jay (Gregory 2023). Differences in repeat content among Corvidae genomes were driven primarily by differences in the abundance of LTR elements and satellite DNA, with the scrub-jay genome exhibiting the highest percentages of both (9.46% and 6.38%, respectively). Meanwhile, there was little variation in LINE element abundance across all analyzed corvid species (Fig. 3). These patterns are broadly consistent with previous findings in birds, which suggest that LTR elements exhibit more recent and dynamic activity relative to LINEs (Kapusta and Suh 2016). This is especially true in the order Passeriformes, which tend to exhibit greater abundance and activity of LTR elements relative to non-Passeriformes (e.g. Warren et al. 2010; but see Zhu et al. 2021; Benham et al. 2023). Intriguingly, a prior study of another *Aphelocoma* species, the Mexican Jay (*A. ultramarina*) found evidence for a major, evolutionarily recent mutation to a microsatellite DNA locus that increased its length, in line with recent satellite activity in this clade (McCormack and Venkatraman 2013). Finally, TE abundance was greatest on the Steller's jay W chromosome (89.86%), consistent with recent results from the hooded crow where repetitive DNA spans ~84% of the W chromosome (Warmuth et al. 2022). These results contribute to the growing consensus that the W chromosome is a refugium for active LTR elements (Peona et al. 2021).

A high-quality genome for the Steller's jay will advance research on the biology of this iconic western American bird species, and will facilitate comparative evolutionary study of sociality (Brown 1974; Ekman and Ericson 2006) and memory (Emery et al. 2004; de Kort and Clayton 2006) in Corvids. Additionally, genomic analyses will aid investigations of the history of divergence, gene flow, and adaptation among the major lineages (and likely species) of the Steller's jay (Cicero et al. 2022). This range-wide genomic context will provide a framework for further understanding the extensive morphological and ecological variation both within and between the major subspecies groups of Steller's jays (Brown 1963a; Cicero et al. 2022). Future analyses leveraging genome-scale data will reveal the influence of specific evolutionary forces (e.g., gene flow) on geographic variation in phenotype among Steller's jay populations, and potentially help pinpoint loci associated with local adaptation in this ecologically widespread species. Additionally, demographic analyses of Steller's jays will advance understanding of how western forests and their animal communities responded to Pleistocene glacial cycles (e.g. Weir and Schluter 2004). Intriguingly, Steller's jays exhibit lower levels of microsatellite allelic diversity in California versus Rocky Mountain populations (Cicero et al. 2022). Aligning genome-wide sequencing data to the highly contiguous reference assembly presented here will be an essential step in teasing apart the role of glacial cycles and other historical processes in shaping range-wide patterns of genetic diversity in Steller's jays. Finally, despite few recognized conservation threats to this species, human activities such as urbanization (Vigallon and Marzluff 2005) and logging (Raphael et al. 1988) have been linked to local declines in abundance and reproductive success for this species. Recent declines in Steller's jay populations have been observed both in California and in the Rocky Mountains, in contrast with increasing abundance in the Pacific Northwest (Fink et al. 2022). The reference genome presented here will provide a key resource for ongoing efforts to document the effect of changing abundance across the range of Steller's jays on broad scale patterns of genetic diversity and gene flow in this iconic forest species of North America and Mesoamerica.

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Acknowledgments

We would like to thank Joshua Ho for assistance with tissue subsampling. Jeff Wilcox kindly provided access to the Mitsui Ranch (property of the Sonoma Mountain Ranch Preservation Foundation) for specimen collection. PacBio Sequel II library prep and sequencing were carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH shared instrumentation grant 1S10OD010786-01. Deep sequencing of Omni-C libraries used the Novaseq S4 sequencing platforms at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH shared instrumentation grant S10 OD018174. We thank the staff at the UC Davis DNA Technologies and Expression Analysis Cores, the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, and the UC Santa Cruz Paleogenomics Laboratory for their diligence and dedication to generating high-quality sequence data.

Funding

This work was supported by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224].

Data availability

Data generated for this study are available under NCBI BioProject PRJNA720569 for the CCGP. Raw sequencing data for the Steller's jay (NCBI BioSample: SAMN30523672) are deposited in the NCBI Short Read Archive (SRA) under SRX19143332–SRX19143334. The transposable element library used to mask repeats is included as a supplemental file. Assembly scripts and other data for the analyses presented can be found at the following GitHub repository: www.github.com/ccgproject/ccgp_assembly. Repeat library and gff files have been deposited in dryad: https://doi.org/10.6078/D15139.

References

Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*. 2020;36(1):311–316.

Ågren JA, Wright SI. Co-evolution between transposable elements and their hosts: a major factor in genome size evolution? *Chromosome Res*. 2011;19:777–786.

Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour*. 2020;20:892–905.

Benham PM, Cicero C, Escalona M, Beraut E, Marimuthu MPA, Nguyen O, Nachman MW, Bowie RCKB. A highly contiguous genome assembly for the California quail (*Callipepla californica*). *J Hered*. 2023;114(4):418–427. doi:10.1093/jhered/esad008

Billings AC, Greene E, Macarthur-Waltz D. Steller's jays assess and communicate about predator risk using detection cues and identity. *Behav Ecol*. 2017;28:776–783.

BirdLife International. Species factsheet: *Cyanocitta stelleri*. 2023 [accessed 2023 Jan 20]. http://www.birdlife.org

Bonaccorso E, Townsend Peterson A. A multilocus phylogeny of New World jay genera. *Mol Phylogenet Evol*. 2007;42:467–476.

Bravo GA, Schmitt CJ, Edwards SV. What have we learned from the first 500 avian genomes? *Annu Rev Ecol Evol Syst*. 2021;52:611–639. doi:10.1146/annurev-ecolsys-012121-085928

Brown JL. Ecogeographic variation and introgression in an avian visual signal: the crest of the Steller's Jay, *Cyanocitta stelleri*. *Evolution*. 1963a;17:23–39.

Brown J. Aggressiveness, dominance and social organization in the Steller Jay. *Condor*. 1963b;65:460–484.

Brown JL. Alternate routes to sociality in jays—with a theory for the evolution of altruism and communal breeding. *Am Zool*. 1974;14:63–80.

Burg TM, Gaston AJ, Winker K, Friesen VL. Rapid divergence and postglacial colonization in western North American Steller's jays (*Cyanocitta stelleri*). *Mol Ecol*. 2005;14:3745–3755.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. *G3*. 2020;10:1361–1374.

Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, Li H. Robust haplotype-resolved assembly of diploid individuals without parental data, arXiv, arXiv:2109.04785, 2021, preprint: not peer reviewed.

Cicero C, Mason NA, Oong Z, Title PO, Morales ME, Feldheim KA, Koo MS, Bowie RCK. Deep ecomorphological and genetic divergence in Steller's Jays (*Cyanocitta stelleri*, Aves: Corvidae). *Ecol Evol*. 2022;12:e9517. doi:10.1002/ece3.9517

De Kort SR, Clayton NS. An evolutionary perspective on caching by corvids. *Proc R Soc B Biol Sci*. 2006;273:417–423.

DeRaad DA, Escalona M, Phred Benham PM, Huffmeyer A, Marimuthu MPA, Sahasrabudhe RM, Nguyen O, Beraut E, Fairbairn CW, Seligmann W, et al. De novo assembly of a chromosome-level reference genome for the California Scrub-Jay. *Aphelocoma californica*. In review.

Ekman J, Ericson PG. Out of Gondwanaland; the evolutionary history of cooperative breeding and social behaviour among crows, magpies, jays and allies. *Proc R Soc B Biol Sci*. 2006;273:1117–1125.

Elliott TA, Gregory TR. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philos Trans R Soc B Biol Sci*. 2015;370:20140331.

Emery NJ, Dally JM, Clayton NS. Western scrub-jays (*Aphelocoma californica*) use cognitive strategies to protect their caches from thieving conspecifics. *Anim Cogn*. 2004;7:37–43.

Ericson PGP, Jansén AL, Johansson US, Ekman J. Inter-generic relationships of the crows, jays, magpies and allied groups (Aves: Corvidae) based on nucleotide sequence data. *J Avian Biol*. 2005;36:222–234.

Feng S, Stiller J, Deng Y, Armstrong J, Fang Q, Reeve AH, Xie D, Chen G, Guo C, Faircloth Brant C, et al. Dense sampling of bird diversity increases power of comparative genomics. *Nature*. 2020;587:252–257.

Fink D, Auer T, Johnston A, Strimas-Mackey M, Ligocki S, Robinson O, Hochachka W, Jaromczyk L, Rodewald A, Wood C, et al. *eBird status and trends, data version: 2021; released: 2022*. Ithaca (NY): Cornell Lab of Ornithology; 2022.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 2020;117:9451–9457.

Ghurye J, Pop M, Koren S, Bickhart D, Chin C-S. Scaffolding of long read assemblies using long range contact information. *BMC Genomics*. 2017;18:527.

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Koren S. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput Biol*. 2018;15:e1007273.

Goloborodko A, Abdennur N, Venev S, Brandao HB, Fudenberg G. *mirnylab/pairtools: v0.2.0*. 2018. doi:10.5281/zenodo.1490831

Gregory TR. Animal Genome Size Database. 2023. http://www.genomesize.com

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*. 2020;36(9):2896–2898. doi:10.1093/bioinformatics/btaa025

Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29:1072–1075.

Hope S. Call form in relation to function in the Steller's Jay. *Am Nat*. 1980;116:788–820.

Jovanović V, Atkins L. Karyotypes of four passerine birds belonging to the families Turdidae, Mimidae, and Corvidae. *Chromosoma*. 1969;26:388–394.

Kalinowski RS, Gabriel PO, Black JM. Who's watching influences caching effort in wild Steller's Jays (*Cyanocitta stelleri*). *Anim Cogn*. 2015;18:95–98.

Kapusta A, Suh A. Evolution of bird genomes—a transposon's-eye view. *Ann N Y Acad Sci*. 2016:1–22. doi:10.1111/nyas.13295

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

Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Gehlenborg N. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol*. 2018;19:125.

Kidwell MG. Transposable elements and the evolution of genome size in eukaryotes. *Genetica*. 2002;115:49–63.

Korlach J, Gedman G, Kingan SB, Chin C-S, Howard JT, Audet J-N, Cantin L, Jarvis ED, Jarvis ED. De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience*. 2017;6:1–16.

Kryukov AP, Spiridonova LN, Tyunin AP, Kryukov KA, Dorda BA. Complete mitochondrial genomes of five subspecies of the Eurasian magpie *Pica pica*, obtained with Oxford Nanopore MinION, and their interpretation regarding intraspecific taxonomy. *Mitochondrial DNA B Resour*. 2020;5:3810–3811.

Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 2014;30:3276–3278.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, *arXiv, arXiv:1303.3997*, 2013, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1303.3997

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094–3100.

Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22:1658–1659.

Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes, *Molecular Biology and Evolution*, 202138(10):4647–4654. https://doi.org/10.1093/molbev/ msab199

McCormack JE, Venkatraman MX. A distinctive genetic footprint of ancient hybridization. *Auk*. 2013;130:469–475.

Peona V, Blom MPK, Xu L, Burri R, Sullivan S, Bunikis I, Liachko I, Haryoko T, Jønsson KA, Zhou Q, et al. Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Mol Ecol Resour*. 2021;21:263–286.

Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T, Manke T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun*. 2018;9:189. doi:10.1038/s41467-017-02525-w

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11:1432.

Raphael MG, Rosenberg KV, Marot BG. Large-scale changes in bird populations of Douglas-fir forests, northwestern California. *Bird Conserv*. 1988;3:63–83.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021;592:737–746.

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;21:245.

Rice P, Longden L, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. 2000;16:276–277.

Shaffer HB, Toffelmier E, Corbett-Detig RB, Escalona M, Erickson B, Fiedler P, Wang IJ. Landscape genomics to enable conservation actions: the California Conservation Genomics Project. *J Hered*. 2022:1–12. doi:10.1093/jhered/esac020

Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics*. 2021;37:1639–1643.

Sim SB, Corpuz RL, Simmonds TJ, Geib SM. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics*. 2022;23:157.

Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. http://www.repeatmasker.org

Suh A, Smeds L, Ellegren H. Abundant recent activity of retrovirus-like retrotransposons within and among flycatcher species implies a rich source of structural variation in songbird genomes. *Mol Ecol*. 2018;27:99–111.

Uliano-Silva M, Nunes JGF, Krasheninnikova K, McCarthy SA. *marcelauliano/MitoHiFi: mitohifi_v2.0 (v2.0)*. Zenodo; 2021. doi:10.5281/zenodo.5205678

Vigallon SM, Marzluff JM. Abundance, nest sites, and nesting success of Steller's jays along a gradient of urbanization in Western Washington. *Northwest Sci*. 2005;79:22–27.

Walker LE, Pyle P, Patten MA, Greene E, Davison W, Muehter VR. Steller's Jay (*Cyanocitta stelleri*), version 1.0. In: Rodewald PG, editor. *Birds of the world*. Ithaca (NY): Cornell Lab of Ornithology; 2016.

Warmuth VM, Weissensteiner MH, Wolf JBW. Accumulation and ineffective silencing of transposable elements on an avian W Chromosome. *Genome Res*. 2022;32:671–681.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, Wilson RK. The genome of a songbird. *Nature*. 2010;464:757–762.

Weir JT, Schluter D. Ice sheets promote speciation in boreal birds. *Proc Biol Sci*. 2004;271:1881–1887.

Weissensteiner MH, Bunikis I, Catalán A, Francoijs KJ, Knief U, Heim W, Wolf JBW. Discovery and population genomics of structural variation in a songbird genus. *Nat Commun*. 2020;11:1–11.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Schulman AH. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 2007;8:973–982. http://www.nature.com/articles/nrg2165

Winkler DW, Billerman SM, Lovette IJ. Crows, Jays, and Magpies (*Corvidae*), version 1.0. In: Billerman SM, Keeney BK, Rodewald PG, Schulenberg TS, editors. *Birds of the world*. Ithaca (NY): Cornell Lab of Ornithology; 2020.

Zhu F, Yin ZT, Wang Z, Smith J, Zhang F, Martin F, Ogeh D, Hincke M, Lin F-B, Burt DW, et al. Three chromosome-level duck genome assemblies provide insights into genomic variation during domestication. *Nat Commun*. 2021;12:1–11.