



OPEN

A reference genome for the Harpy Eagle reveals steady demographic decline and chromosomal rearrangements in the origin of Accipitriformes

Lucas Eduardo Costa Canesin¹, Sibelle T. Vilaça¹, Renato R. M. Oliveira¹, Farooq Al-Ajli^{2,4}, Alan Tracey², Ying Sims², Giulio Formenti², Olivier Fedrigo², Aureo Banhos⁵, Tania M. Sanaiotti⁶, Izeni P. Farias⁷, Erich D. Jarvis^{2,3}, Guilherme Oliveira¹, Tomas Hrbek^{7,8}, Vera Solferini⁹ & Alexandre Aleixo¹✉

The Harpy Eagle (*Harpia harpyja*) is an iconic species that inhabits forested landscapes in Neotropical regions, with decreasing population trends mainly due to habitat loss, and currently classified as vulnerable. Here, we report on a chromosome-scale genome assembly for a female individual combining long reads, optical mapping, and chromatin conformation capture reads. The final assembly spans 1.35 Gb, with $N50_{\text{scaffold}}$ equal to 58.1 Mb and BUSCO completeness of 99.7%. We built the first extensive transposable element (TE) library for the Accipitridae to date and identified 7,228 intact TEs. We found a burst of an unknown TE ~13–22 million years ago (MYA), coincident with the split of the Harpy Eagle from other Harpiinae eagles. We also report a burst of solo-LTRs and CR1 retrotransposons ~31–33 MYA, overlapping with the split of the ancestor to all Harpiinae from other Accipitridae subfamilies. Comparative genomics with other Accipitridae, the closely related Cathartidae and Galloanserae revealed major chromosome-level rearrangements at the basal Accipitriformes genome, in contrast to a conserved ancient genome architecture for the latter two groups. A historical demography reconstruction showed a rapid decline in effective population size over the last 20,000 years. This reference genome serves as a crucial resource for future conservation efforts towards the Harpy Eagle.

Keywords Accipitridae, Genome assembly, Transposable elements, Demographic decline, Genome architecture

The Harpy Eagle, *Harpia harpyja* (Aves: Accipitridae; Linnaeus, 1758), a critical apex predator, is one of the largest extant eagles on Earth. It occurs widely in the Neotropics from Central to South America, with half of its distribution in the Brazilian Amazon¹. The Harpy Eagle occurs in low population density throughout its range and is characterized by a sizable territory and a patchy distribution². Despite showing reversed sexual dimorphism in body size, they have a positive female-biased sex ratio at birth³. As an apex predator, it feeds mainly on medium-sized and tree-dwelling mammals, especially sloths and small monkeys^{4,5}. The Harpy Eagle is one of two raptor species with documented cases of persecution (i.e., direct killing), which, in conjunction with habitat decrease, disturbance, and reduction of food sources⁶, increases its extinction risk.

The IUCN classification of the Harpy Eagle as vulnerable is reflected in the low genetic diversity documented for the species¹, with signs of recent bottlenecks reducing heterozygosity, even though the species retains high

¹Instituto Tecnológico Vale - Desenvolvimento Sustentável (ITV-DS), Belém, Brazil. ²Rockefeller University, New York, USA. ³Howard Hughes Medical Institute (HHMI), New York, USA. ⁴Katara Biodiversity Genomics Program, Katara Cultural Village Foundation, Doha, Qatar. ⁵Universidade Federal do Espírito Santo (UFES), Alegre, Brazil. ⁶Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, Brazil. ⁷Universidade Federal do Amazonas (UFAM), Manaus, Brazil. ⁸Trinity University, San Antonio, USA. ⁹Universidade Estadual de Campinas (Unicamp), Campinas, Brazil. ✉email: alexandre.aleixo@itv.org

genetic diversity in mitochondrial markers throughout its distribution⁷. Individuals from the Brazilian southern Amazon and Atlantic Forest populations show high gene flow with low population structure, while the northern Amazon population shows greater differentiation⁴. However, estimates of population dynamics through time and effective population size (N_e), two key features for assessment of genetic health of any species and development of high-impact conservation strategies, still need to be improved for the Harpy Eagle⁷.

Previous studies showed that the Harpy Eagle has many large-scale rearrangements in its genome architecture when compared to the more distantly related Galloanserae (i.e., chicken, geese, ducks, quails, and pheasants)^{8,9}. Transposable elements (TEs) have been associated with large-scale structural variants, including chromosomal rearrangements, transcriptional regulation, and methylation¹⁰. Within Aves, the chicken and zebra finch genomes are the best annotated and have the most in-depth descriptions of TEs¹¹, even though robust descriptions of the TE landscape for a diversity of non-model bird species have been recently done^{12–16}. However, as studies of TEs are heavily impacted by genome assembly contiguity and quality, TE landscapes are still restricted to a few model species for which chromosome-level genomes produced using long-read data are available¹¹. In general, genome-wide patterns indicate a reduced activity of TEs across birds, with a strikingly low presence of TEs (lower than 15%, except Piciformes; see¹⁷) possibly limiting genome size expansion and rearrangements^{10–12}. Consequently, birds have a remarkably stable genome size and large-scale synteny in macrochromosomes^{12,18}. Despite the overall trend of genome architecture conservation, genome assemblies of diverse Aves groups are adding evidence of extensive variation in some lineages¹⁶. Studies of comparative genomics benefit from high-quality genomes to elucidate structural variants and large- and micro-scale synteny.

Here, we report on a high-quality, chromosome-level genome assembly of the Harpy Eagle and bring evidence to a long-standing enigma in the apparent genome architecture (*bauplan*) reshuffling within Accipitridae. We provide structural coding gene annotation, identify transposable elements, and compare the genome architecture of the Harpy Eagle to species with chromosome-level genome assemblies available for other Accipitridae, the Cathartidae (New World vultures and condors), and the basal Galloanserae. Our findings support a major reshuffling event at the base of the Accipitridae-Pandionidae split or earlier. We also used the genome reported herein to reconstruct the Harpy Eagle effective population size through time, contextualizing the current genetic diversity documented for the species. Given this apex predator's ecological importance, cultural relevance, and conservation needs, the reference genome for the Harpy Eagle is a valuable asset for future studies on this species and will heavily contribute to population and conservation studies.

Results

Genome sequencing and assembly

We generated the near-complete chromosome-level genome assembly for the Harpy Eagle (named bHarHar1.0) (Fig. 1; Table 1). We obtained 37.68 Gb of reads and a coverage of 31.41X for PacBio reads, 707.27X for Bionano optical mapping, and 157.04X for Hi-C. Coverages of HiFi-CCS reads and Hi-C libraries were considered for all reads prior to filtering and trimming. We estimated the genome size at 1.23 Gb and heterozygosity at 0.306% through the k-mer analysis from raw read data using Genomescope (Fig. 1a). The repetitive content was initially estimated from sequencing reads at 9.4% of the total genome length (Fig. 1a). The final curated assembly was 9% larger at 1.35 Gb (1,351,447,071 bases), distributed in 322 scaffolds (664 contigs), with $N50_{contigs} = 16.8$ Mb and $N50_{scaffolds} = 58.1$ Mb. Over 96% of the assembled sequence was assigned to 30 pseudo-chromosomes (28 autosomes plus Z and W sex chromosomes). The assembled chromosome set is consistent with a previously reported karyotype of $2n = 58^{8,19}$. Macrochromosome length varied from 18.49 Mb to 109.61 Mb, the 5 microchromosomes varied from 0.13 Mb to 1.1 Mb, while the sex chromosomes had lengths of 117.05 Mb (Z chromosome) and 37.4 Mb (W chromosome) (Fig. 1b). We observed a negative correlation between chromosome length and inter-chromosomal chromatin contacts, with more contacts occurring between the 15 smallest macrochromosomes, both Z and W sex chromosomes, and all five microchromosomes (Fig. 1c). The read k-mer completeness was estimated at 97.86%, and the consensus quality value was equal to 60.9562 (0.8 errors/Mb). BUSCO estimates showed a complete genome retrieving 99.7% of single copy genes using the Aves reference database (aves_odb10) and only 0.4% of duplicated sequences remaining (Fig. 1d). The complete circularized mitogenome was recovered from the final assembly and had all 37 expected coding genes distributed in 17.74 Kb. Hereafter, we renamed the scaffold 439, identified as the mitochondrial genome, as chrM (NW_026293368.1 in NCBI's deposited assembly). The nuclear genome is deposited under assembly accession number GCF_026419915.1 in NCBI.

Transposable elements annotation

We built a custom intact TE library for *H. harpyja*, the first for the Accipitiformes order to date. The library of intact elements included 327 LTRs, 1,703 LINEs/SINEs, and 4843 DNA elements and miniature inverted-repeat transposable elements (MITEs) (7228 intact transposable elements in total) (Fig. 2A). The main class of intact elements found was the CACTA DNA transposon, with 2852 elements, followed by the CR1 LINE retrotransposons (1397 elements) and the Mutator DNA transposon class (1102 elements). These intact elements represent the TEs potentially still active in the *H. harpyja* genome, directly impacting the evolution of the genomic architecture of the species.

This library was subsequently used to identify ancient genome-wide transpositions (Fig. 2a). We found that 17.26% of the total genome was masked as repetitive sequences. Transposable elements represent the most significant portion of non-redundant masked sequences (16.33%), with 480,259 independent TE insertion events. Among the annotated elements, the Unknown class represents 36.5% of the masked bases (Fig. 2a). A single element (TE_00000121, Supplementary Table S1) accounted for 25.3% of all masked bases. We searched this TE sequence for TE-related motifs in DFAM and protein motifs in Uniprot and only found disordered domains. The

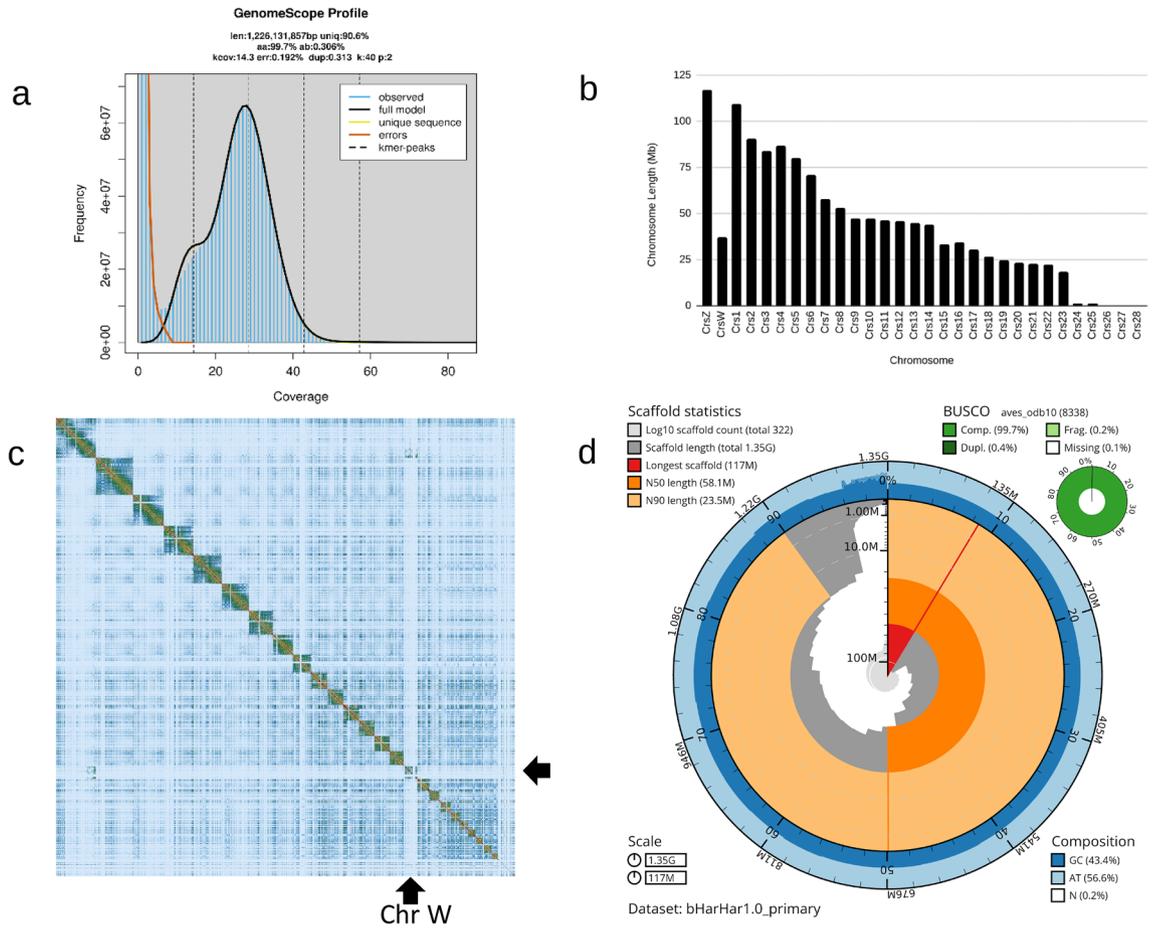


Figure 1. Genome assembly statistics. (a) GenomeScope k-mer profile; (b) Length distribution for each of the assembled pseudo-chromosomes; (c) Hi-C interaction heatmaps for the curated bHarHar1.0 assembly. Chromosomes are displayed in size order from top to bottom and from left to right. Arrows indicate the chromosome W; (d) Snail plot for the genome assembly of *Harpia harpyja*, bHarHar1.0 with summary statistics. The plot shows the N50 metrics, base pair composition, and BUSCO gene completeness, generated with Blobtoolkit v2.6.4 (<https://github.com/blobtoolkit/blobtoolkit>)⁶⁵.

	Statistics
Total assembly length (Gb)	1,351,447,071
Number of scaffolds/contigs	322/664
Longest scaffold (Mb)	117
Scaffold N50/L50 (Mb/count)	58.1/8
Contig N50/L50 (Mb/count)	16.8/26
Gaps	342
Average Gap/Longest gap length (Kb)	7.75/484.506
Total gap length (Mb)	2.65
QV	60.9562
BUSCO Completeness (%)	99.7
Single copy (%)	99.3
Duplicated (%)	0.4
Fragmented (%)	0.2
Missing (%)	0.1
Total (Aves_odb10)—single copy genes	8338

Table 1. Summary statistics for the Harpy Eagle (*Harpia harpyja*) genome assembly (bHarHar1.0).

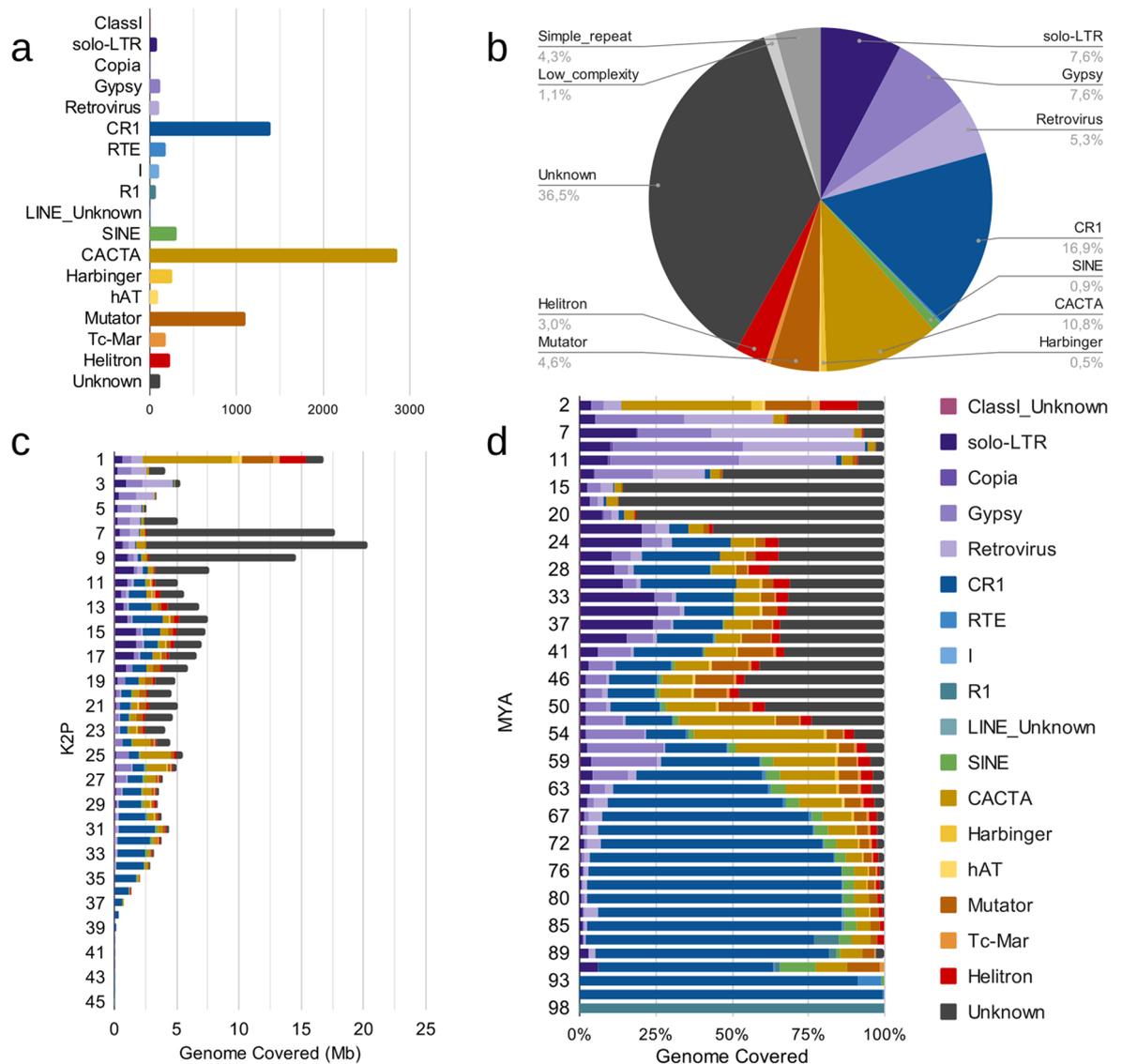


Figure 2. Overall annotation of repeat elements considering (a) intact elements and (b) annotated insertions found on the Harpy Eagle genome; (c) repeat landscape transposable elements in the Harpy Eagle genome in terms of K2P divergence; (d) relative repeat landscape in terms of divergence time. TE annotation was done with EDTA v2.0.1 and summary output was plotted with LibreOffice's spreadsheet editor v24.2.5 (<https://libreoffice.com.br/>).

second most prevalent TE was the CR1 retroelement, with 16.9% of insertions. The third most enriched class (10.8%) is the CACTA transposon class (Fig. 2b).

The Kimura two-parameter (K2P) distance between the intact elements and each identified insertion showed seven activity peaks of transposable elements in the Harpy Eagle genome (Fig. 2c). The oldest peak is composed mainly of the CR1 retrotransposon, predominating in the most ancient transposition events within 26% K2P or older (Fig. 2c) at over 98 million years ago (MYA; Fig. 2d). The CR1 insertions represented more than 50% of all TE insertions until ~63 MYA when the insertion events of DNA transposons (CACTA elements) and LTR Gypsy started to take place. CACTA elements comprise most TE insertions ~54 MYA (43%). Noteworthy, we detected the spread of an unknown element (TE_00000018) also at ~54 MYA, reaching a peak of 47.9% of all insertions at 48 MYA, simultaneous to the increase of Mutator DNA transposon insertion events (10.9% of all insertions at the same time). This was followed by a period (41–33 MYA) of relative stability for all TE classes, except for solo-LTRs (peak ~33–37 MYA with mean of 24.67% of all insertions) and the Helitron rolling-circle elements. After this period, we could observe a small increase in CR1 elements, peaking at ~31 MYA, but that was followed by silencing, reducing its relative content to ~1.8% of total insertions ~20 MYA. Solo-LTRs peak again at ~22–24 MYA. Concomitant to this, started the most pronounced spread, in relative and absolute terms, of the unknown element described above (TE_00000121, Supplementary Table S1), hereafter named *Harpia harpyja* 1 TE (HarHar1), which covered 87.1% of all insertions by ~17 MYA. This TE had no hits to the DFAM curated TE hidden Markov models (*Ficedula albicollis* and *Uraeginthus cyanocephalus* data) and no hits to UniProtKb and SwissProt protein databases. After translating the HarHar1 sequence in the six frames, we also searched for

protein domains in the *Harpia harpyja* 1 TE, but none were detected. The second most recent event of elevated transposition activity showed the proliferation of the Retrovirus-like and Gypsy LTRs ~4–11 MYA, summing 73.9–58.3% of all insertions. Curiously, the prevailing elements in the most recent insertion events (nearly intact elements, transposition events in the last 2 million years) are the DNA transposons CACTA (42.4%) and Mutator (15.3%) and the Helitron rolling-circle DNA TE (12.6%). Retrotransposons and the group of unknown elements compose only a minor fraction of nearly intact TEs.

Coding gene annotation

The NCBI eukaryotic annotation pipeline annotated 16,803 unique protein-coding loci (which span 44,808 mRNAs) and 174 pseudo-genes; and 4251 lncRNAs, 41 snRNAs, 268 snoRNAs, and 367 tRNAs among non-coding RNAs. Most of the annotated coding loci (16,363) were also identified as known protein coding loci deposited in public sequence databases (only 440 were uncharacterized proteins), of which 99.8% (16,341 loci) were located on the thus far named chromosome scaffolds. When considering the distribution of coding loci across the genome, microchromosomes had higher gene density than macrochromosomes (Fig. 3a) despite the linear scaling of gene number with chromosome length (Fig. 3b).

We compared the annotated Harpy Eagle coding sequences with those of the Eurasian Goshawk (*Accipiter gentilis*), the Golden Eagle (*Aquila chrysaetos*), the California Condor (*Gymnogyps californianus*), and a new de novo assembly of chicken (*Gallus gallus*). We assigned orthology based on reciprocal best hits in all-vs-all pairwise comparisons. We found 15,636 orthogroups (OGs), of which 12,838 had genes from all species considered (12,646 OGs had exactly 1:1 relationships and are thus the most suitable gene set for phylogenetic inference). The Harpy Eagle coding genes were distributed in 14,987 OGs, 16 species-specific (total 121 coding genes). Moreover, 189 Harpy Eagle genome coding genes were not assigned to any orthogroup. Thus, we have identified 310 coding genes specific to this lineage.

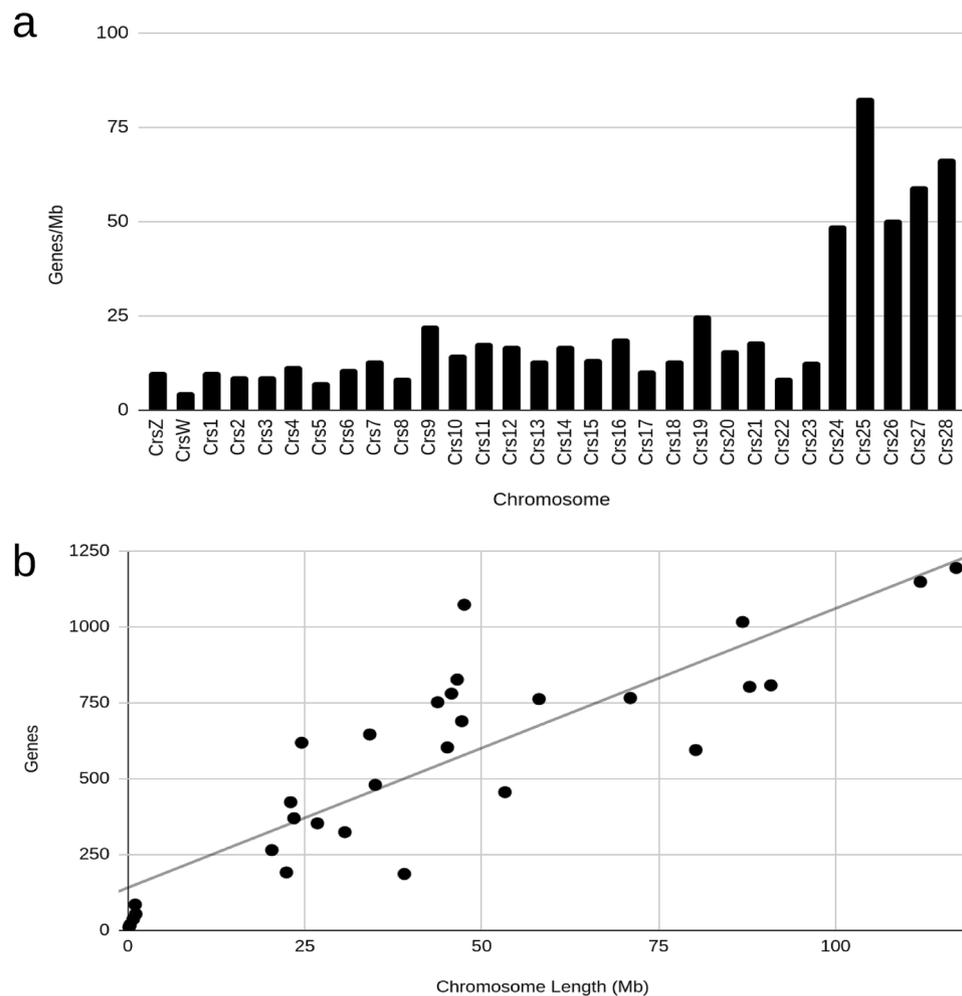


Figure 3. Gene density per chromosome. (a) Genes per kilobase for each assembled chromosome; (b) Number of coding loci per chromosome length.

Comparative genomics

We found strong synteny between the genomes of the Harpy Eagle and the other two species of Accipitridae (the more closely related Eurasian Goshawk, *A. gentilis*, and the more distantly related Golden Eagle, *A. chrysaetos*; Fig. 4; Supplementary Figure S1), preserving the same collinear gene blocks. Despite the overall trend of macrosynteny conservation within Accipitridae, assuming the *G. gallus* genome architecture as the ancestral *bauplan*, we identified 60 major rearrangements (Fig. 5; Supplementary Table S2) comparing the five species considered

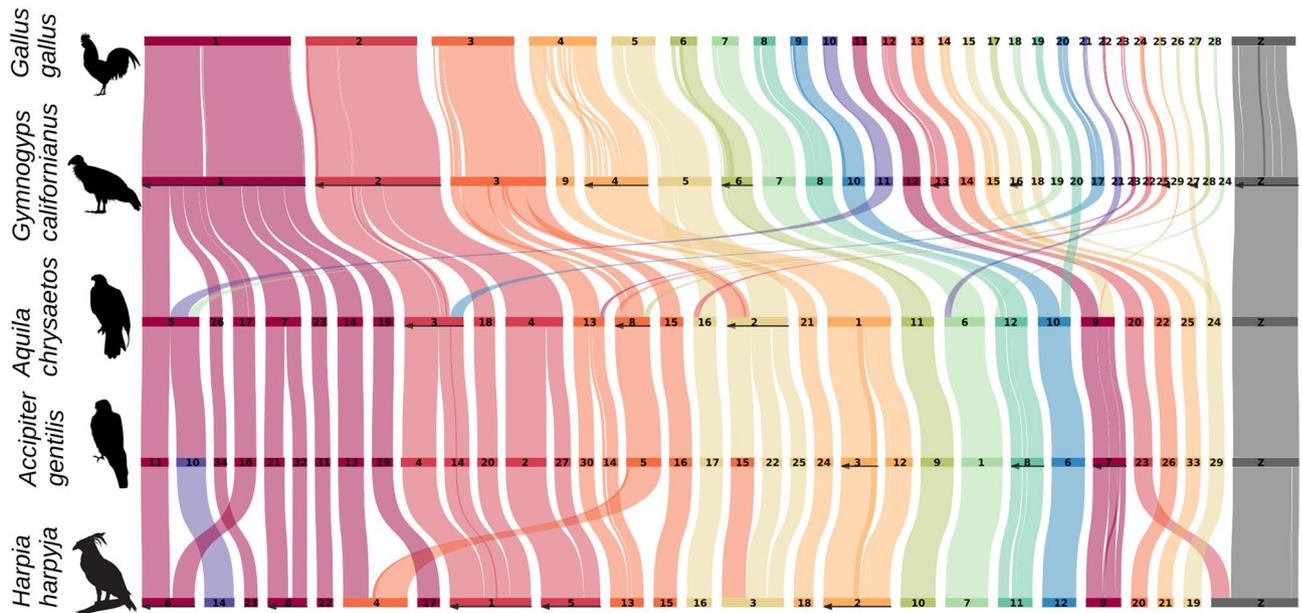


Figure 4. Patterns of chromosome evolution within Accipitriformes and Galloanserae. Pairwise macrosynteny based on 20 gene collinear blocks as reciprocal best hits orthologs. Numbered bars indicate putative chromosomes for each analyzed species. Links indicate the boundaries of syntenic gene blocks identified by MCScanX. Color codes are based on the *Gallus gallus* chromosome organization. Data was plotted using SynVisio web application (<https://synvisio.github.io/>)⁸³ and edited manually for better readability. Arrows denote inverted sequence orientation in alignment. Chromosomes not shown in this figure did not show any syntenic gene block.

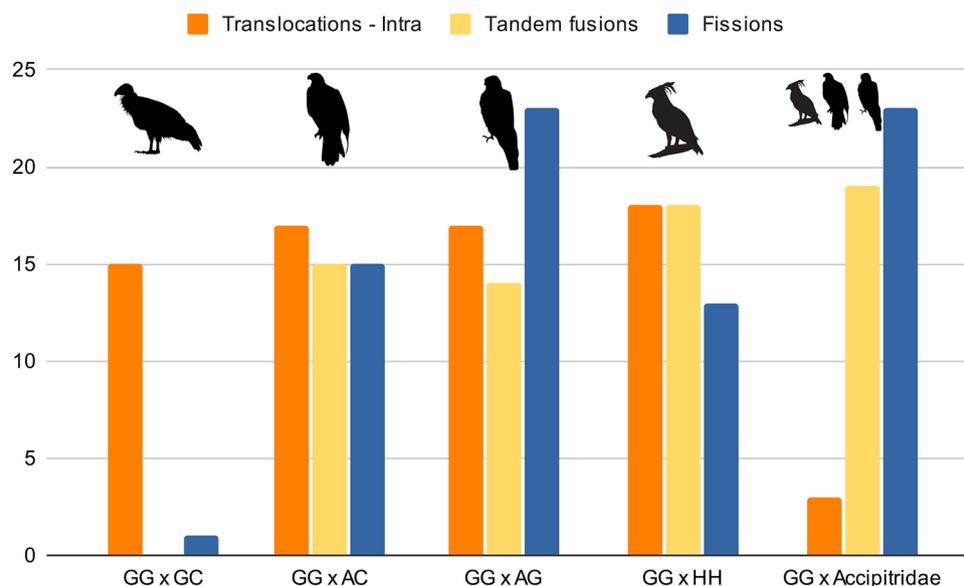


Figure 5. Chromosomal rearrangement counts within Accipitriformes in comparison with the *Gallus gallus* ancestral *bauplan*. Counts were computed based on the synteny analysis shown in Fig. 4. Any rearrangement was considered exclusive when found only in a single species. Abbreviations: GG *Gallus gallus*, GC *Gymnogyps californianus*, AC *Aquila chrysaetos*, AG *Accipiter gentilis*, and HH *Harpia harpyja*.

here (all rearrangements were manually checked with cross-species HiC mapping, using bHarHar1 as reference; Supplementary Figure S2). Of these, 15 small intrachromosomal translocations and one fission (all occurring on the ancestral linkage group equivalent to the chicken chromosome 4) are evident when comparing the *G. gallus* and *G. californianus* genomes. These rearrangements persisted in the same configuration in the other Accipitriformes analyzed (see Supplementary Table S2 for a detailed description of all rearrangements identified) and might be characteristic of the Accipitriformes genome *bauplan*.

Within Accipitriformes, the synteny alignment of the three Accipitridae species and the California Condor (Cathartidae) revealed 23 chromosome fissions occurring in the largest autosomes (equivalent to gg1 to gg5), with 19 tandem fusions of microchromosomes with macrochromosome fragments, and 3 intrachromosomal translocations (Fig. 5). Thus, most chromosomal rearrangements detected ($n = 45$) were found exclusively within the Accipitridae family. *H. harpyja* shows four exclusive end-to-end (tandem) chromosome translocations, originating chromosomes 1, 4, 6, and Z, by fusing gene blocks that are apart in the other four species considered here (e. g., the orthologous sequence to the autosomes ag23, ac20, gc13, and gg12, from *A. gentilis*, *A. chrysaetos*, *G. californianus*, and *G. gallus*, respectively, are fused to the sex chromosome Z in the Harpy Eagle). In turn, the *A. gentilis* genome shows seven exclusive chromosome breaks relative to the other Accipitridae (Fig. 4; Supplementary Table S2). *Aquila chrysaetos* has the most conserved genome architecture amongst the Accipitridae sampled, when compared to the ancestral genome *bauplan* of *G. gallus* and *G. californianus*, with only one exclusive tandem translocation, originating the chromosome ac5 (Fig. 4).

Past effective population size dynamics

We reconstructed the past demographic history of the Harpy Eagle (Fig. 6). The PSMC analysis revealed an overall declining trend in the effective population size over the past 1 million years, with a steeper reduction in the last 20 thousand years, encompassing the Last Glacial Maximum (LGM, 18–23 thousand years ago—KYA), and continuing until the onset of the Holocene. This effective population decline went from ~8000 individuals (~0.5–1 million years ago—MYA) to about ~4,000 individuals (~200 KYA). This was followed by periods of increase and decrease in N_e around a mean (from 200 to 20 KYA) of 4000 individuals again until ~20 KYA. Nevertheless, the more recent N_e intervals in the PSMC for the Harpy Eagle are especially concerning since they reach $N_e < 1000$ and represent the lowest effective population size throughout its evolutionary history (Fig. 5). We observed only minor N_e variations over warming (~147–122 KYA) and cooling (~123–65 KYA) periods of climatic change²⁰. Notwithstanding, starting ~20 KYA, the continuous downturn in genetic diversity of the Harpy Eagle spanned the Last Glacial Maximum and the Holocene, two sequential periods characterized by opposing cooling and warming phases, respectively, and which did not impact the continuous and steep N_e reduction observed during this period.

Discussion

Here, we have assembled a reference genome for the Harpy Eagle and characterized its genome in the context of other members of the same order (Accipitriformes) for which chromosomal level genomes are available and a more distantly related basal avian lineage (Galloanserae). We inferred the TE landscape from a custom repeat library for the Harpy Eagle genome, the first such library for the order Accipitriformes, and described overall genomic features. We also reconstructed the past demography of the species to infer changes in its effective

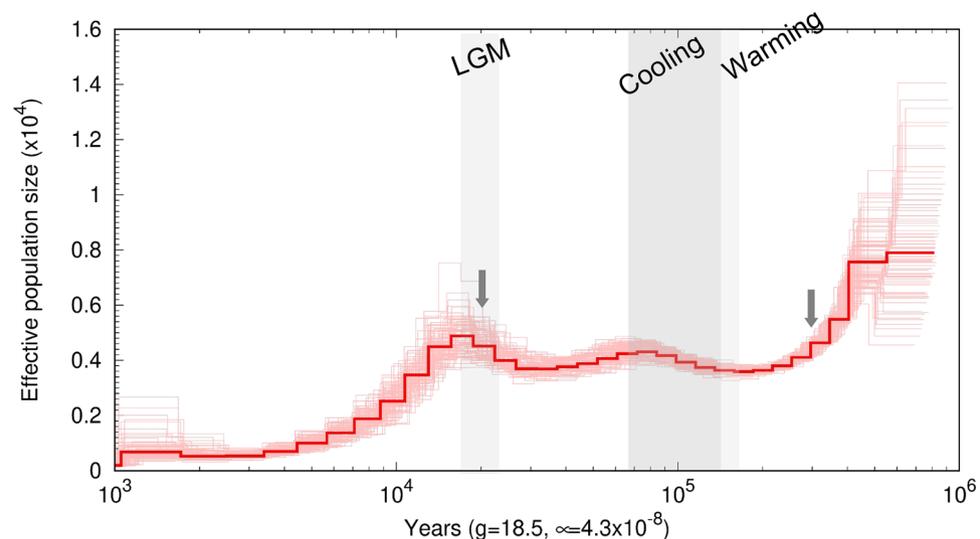


Figure 6. Past effective population size (N_e) estimates of the Harpy Eagle. The darker line depicts the N_e through time, while lighter colors represent bootstrap estimates over 100 replicates. Gray arrows denote the inflection points of effective population size. Shaded areas denote the LGM, and the warming and the cooling period discussed by Germain et al²⁰.

population size. Below, we discuss how the genomic features combined with its evolutionary history might have contributed to the evolution of the Harpy Eagle and other Accipitridae.

Harpy eagles have a similar genome organization pattern observed in other birds and non-avian reptiles, with macro- and microchromosomes, a greater concentration of genes on the latter, and higher microchromosome recombination rates²¹. Indeed, microchromosomes have been considered genome building blocks due to their frequent rearrangements across species²¹. Previous studies have shown a high degree of genome synteny between Galloanserae (galliform and anseriform birds) and Neoaves (most other birds, including passerines)^{17,22}. Birds are known for having highly syntenic genomes across the entire group²³ and even share large syntenic blocks with turtles²⁴. However, our results show an interesting pattern of genomic rearrangement within Accipitriformes. Following the elementary algebraic operations in chromosome rearrangements proposed by Simakov et al.²⁵, we show that only reversible macrosyntenic changes (tandem or end-to-end translocations and fissions) occurred within the Accipitridae and between the more distantly related Cathartidae and Galloanserae. These chromosomal rearrangements occur when whole chromosomes are fused end-to-end or break on intergenic sequence and thus preserve the overall gene order of the syntenic blocks comprising each chromosome. In contrast, most rearrangements between Accipitridae *versus* Cathartidae plus Galloanserae are macrochromosome fissions combined with microchromosome fusions, followed in some cases by small synteny-breaking rearrangements within gene space, which are irreversible changes in genomic architecture.

While birds are known for having an overall stable chromosome order¹⁸, we observed that several reshuffling events happened throughout the evolutionary history of Accipitriformes, followed by restabilization of genome architecture in its most diverse and widespread lineage, the Accipitridae. Previous studies based on chromosome painting had already identified major syntenic disruptions within Accipitriformes, demonstrating a similar genomic *bauplan* amongst Accipitridae and Pandionidae (ospreys), but one that differed from the basal-most Cathartidae, which, in turn, conserved the same overall syntenic association present in the basal Galloanserae within Aves^{9,26,27}. However, in the only study to analyze these chromosomal reshuffling events in Accipitriformes from a phylogenetic perspective, species of Falconidae (falcons) were included in the analyses and were assumed as closely related to Accipitridae (eagles, hawks, harriers, kites, and Old-World vultures) and Pandionidae (ospreys) to the exclusion of Cathartidae (New-World vultures⁹). The conclusions drawn from this study, however, need to be re-evaluated given that whole-genome data has confirmed the paraphyly of diurnal birds of prey, with Cathartidae, Sagittariidae (secretary birds), Pandionidae, and Accipitridae grouping together in one clade (Accipitriformes) which, in fact, is sister to nocturnal birds of prey (owls, Strigiformes) and therefore totally separated from the Falconidae²⁶. Unfortunately, no chromosome painting data or chromosomal level genomes are available for Sagittariidae (which includes a single species, the Secretarybird *Sagittarius serpentarius*). Still, these two combined datasets support the evolution of a more conserved ancestral genome architecture until at least the base of the Accipitriformes, given the basal divergence of Cathartidae from all other Accipitriformes²⁶, contrasting strongly to a derived “reshuffled” *bauplan* shared at least by Pandionidae and Accipitridae^{9,26,27}. Chromosome-level genomes or chromosome painting data for Sagittariidae should elucidate whether this unique lineage shares a genomic *bauplan* that is either more similar to that of Cathartidae and Galloanserae or to those found in Pandionidae and Accipitridae^{9,26,27}.

Another organizational characteristic is that repeat content in birds is generally low when compared with other animals (lower than 15%¹⁷), except for Piciformes (woodpeckers and toucans), which have > 20% repeat content¹² and sparrow songbirds (Passerellidae) with > 30% repeat content¹⁶. For the Harpy Eagle, we found more than twice the repeat content (16.33%) reported for most songbirds (7.8%) and parrots (9.8%,²⁸), a closer figure to Piciformes. Over 36% of all TE insertions in the Harpy Eagle genome are from unknown elements. Two unknown elements comprise over 88% of all insertions identified (Supplementary Table S1). The most prevalent of these two unknown elements is solely responsible for the peak of activity at ~ 17 MYA (mean K2P = 8.7%). CR1 nonLTR retrotransposons are the second most prevalent superfamily but are concentrated in the ancient transposition events, fading out to become virtually absent in the last 3 MYA in the Harpy Eagle. As shown by Kapusta and Suh¹¹, LTR retrotransposons are especially active in Accipitriformes, but we found more Gypsies (17.8 Mb masked) than ERVs (12.3 Mb) in the Harpy Eagle genome.

We found evidence of significant changes in the TE content of the Harpy Eagle genome over the past 98 MYA. Major shifts in TE content are coincident or precede phylogenetic diversification events in the Accipitridae family. The most prevalent element in avian species, the LINE CR1 elements^{29–31} are replaced by a concomitant expansion of Gypsy LTR TEs and CACTA DNA transposons after the Cretaceous–Paleogene extinction event ~ 65 MYA, and the split between Cathartidae and Accipitridae families. Bursts of solo-LTRs precede major expansions within the Accipitridae family. The occurrence of solo LTRs results from ectopic recombination events between very similar LTRs, usually associated with recent expansions of these elements. The split of the *H. harpyja* from the other Harpiinae at ca. 16.4 MYA³² is also coincident with an unknown TE burst ~ 13 to 22 million years ago (Fig. 2C and D). Taken together, these TE expansions may have played a role in diversification events within this group. Bursts of TE activity have been associated with speciation events^{33,34}, as chromosome rearrangements can lead to postzygotic (reproductive) isolation. Widespread transposition events, often provoked due to recent invasion through horizontal transfer or reactivation of previously silenced ancient elements, can alter drastically the genome *bauplan* in a population, especially those with low effective population size, since natural selection is relatively less effective than genetic drift in small populations. Studies using high-quality genomes and repeat landscapes have supported the idea that increased TE activity might correlate with speciation timing across species³³. In mammals, ERVs have been associated as mediators of genomic plasticity by facilitating recombination and DNA rearrangements³⁵. In birds, CR1 LINEs were associated with speciation events in songbirds³³. Here, we show that major changes in the relative TE content of different classes, namely Gypsy retrotransposons and solo-LTRs, DNA transposons CACTA and an unknown element occurred concomitantly to the reshuffling in the ancestral genomic *bauplan* to all Accipitridae. Following the concept of “ecology of the genome”³⁶, similar

to the deer mouse³⁷, Harpy eagles have a recent expansion of an unknown TE (HarHar1) that caused a shift in its TE landscape and drove LINEs (common in other avian species) to lower relative abundances in the genome. The causes and consequences of this shift in abundance and how competition for insertion sites between LINEs, LTRs, and unknown TEs³⁷ shaped the genome architecture of Accipitridae, remains to be elucidated following the availability of more high-quality genomes for this and related families.

The past demographic history of the Harpy Eagle provides evidence of a very concerning scenario. Between ~300 and 20 KYA, N_e estimates fluctuated relatively little in the Harpy Eagle. However, this trend was followed by a steep decline starting at approximately ~20 KYA to less than 10% of the N_e initially calculated. Despite the lack of reliability in N_e estimates younger than ~10 KYA, characteristic of PSMC analysis, the contemporary population size is estimated to be declining fast with rapid habitat loss (as discussed in Kaizer et al.⁵). The overall declining trend since ~20 KYA is thus supported by independent evidence. Given the species' deep cultural significance to Amerindian societies of the Amazon region³⁸ and hunting reports, the N_e decrease since the LGM may have multiple causes. Many species had associated decreasing population sizes due to past climate change compounded by anthropogenic impact, possibly from the first waves of human migration to South America^{39–41}. Recent evidence for earlier presence of human populations in the American continent pushes back the occupation to even before the LGM, ~30 KYA^{42–45} (as an open debate, see also Surovell et al.⁴⁶). Thus, considering recent analysis on the impact of human activity in the extinction of megafauna, mammals, and birds^{47–49} (but see also Stewart et al.^{50,51}), we can assume the decline in population size of the Harpy Eagle overlapped the human occupation of South America.

The recent fires in the Brazilian Amazon, Pantanal, and Cerrado biomes, compounded with extensive deforestation and climate change, pose dire threats to the survival of one of the world's largest extant eagles. The drastic reduction in population size under stressful environmental conditions diminishes the power of selection to purge and control transposable elements, old and new, and expose the Harpy genome to a reshuffling of syntenic gene blocks, higher rates of gene regulation-disrupting transposition events, and ectopic recombination among terminal repeats from TEs⁵². This, in combination with other factors such as low genetic diversity and higher rates of inbreeding, may contribute to a lower genetic health of the species. In summary, the high-quality genome presented here offers a robust foundation for informed conservation strategies of the Harpy Eagle. By providing a detailed genetic blueprint, this genome enables the development of targeted genetic management strategies for both in situ and ex situ conservation programs, ensuring the maintenance of genetic diversity of wild and captive populations. The development of molecular markers derived from the reference genome can facilitate the monitoring of wild and captive populations, enabling researchers to track genetic diversity, gene flow, and potential signs of inbreeding depression in real-time. Additionally, using the reference genome presented herein to support the re-sequencing of natural populations will allow for regional inferences of past demographic trends that might identify historical strongholds for the species. Finally, by mapping the TE landscape in the Harpy Eagle genome, it will be possible to monitor its influence on deleterious mutations in different natural and captive populations, guiding conservation programs. This comprehensive approach, integrating genomic insights with field monitoring, is essential for the long-term conservation of the Harpy Eagle and other threatened species.

Materials and methods

Sampling and sequencing

A blood sample was taken from a rescued female specimen shot illegally by hunters in a forest clearing near Parintins, Amazonas state, Brazil (2.625833 S, 56.649722 W); this individual was selected for whole genome sequencing to obtain both sex chromosomes. The specimen was taken to an animal care facility at “Centro de Triagem de Animais Silvestres” (Cetas) of the “Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis” (Ibama) in Manaus, where blood was sampled, stored in 96% ethanol, flash-frozen on dry ice at the Laboratório de Evolução e Genética Animal (LEGAL) of the Universidade Federal do Amazonas (UFAM) in Manaus, and kept frozen until sample processing. All experimental protocols were approved by the Brazilian National Council for Control of Animal Experimentation (CONCEA) and the Ethics Committee on Animal Use of the Federal University of Pará (protocol number #7335280722). All methods were carried out in accordance with relevant guidelines and regulations and are reported following the ARRIVE guidelines (<https://arriveguidelines.org>). The Harpy Eagle blood sample was collected under permit #31457 from the Brazilian “Sistema de Autorização e Informação em Biodiversidade” (SISBIO) issued to A. B. and accessed under the Brazilian “Conselho de Gestão do Patrimônio Genético” (permit CGen #A0341AE) and “Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado” (SISGEN #52231) issued to T. H. The Vertebrate Genomes Laboratory (VGL), a hub of the Vertebrate Genomes Project (VGP, <https://vertebrategenomesproject.org/>), performed all sequencing and assembly steps. High molecular weight (HMW) DNA was extracted using the Circulomics HMW DNA extraction standard TissueRuptor protocol with the Nanobind Tissue Big DNA Kit. After DNA extraction and sequencing library preparation, three different technologies were used for genome sequencing: Pacific Biosciences (PacBio) long-read HiFi-CCS using the Express Template Prep Kit 2.0, contact reads generated by Arima Genomics (Hi-C reads) using the v2.0 kit and sequenced on an Illumina NovaSeq 6000, and Bionano Genomics optical mapping using ultra-long DNA of high molecular weight (> 300 kb) and the Bionano Prep Direct Label and Stain (DLS) Protocol, which was run on a Saphyr flow cell.

Genome assembly

The assembly was done with the VGP pipeline v2.0 in HiFi-only mode. Overall HiFi read quality check was performed with FastQC⁵³ and Nanoplot feature from the NanoPack2 suite⁵⁴. Reads were discarded when at least 35 bp of adapter sequences (-O 35) were found in any position using CutAdapt⁵⁵, accepting a maximum error rate of 10% (-e 0.1). Contig assembly was performed with Hifiasm v. 0.15.1-r334 in primary mode^{56,57}

with default settings, followed by removal of haplotype duplicates with `purge_dups` v. 1.2.5⁵⁸ using two rounds chaining (-2). Scaffolding was performed in two steps. For the first step, the optical map was fed into Bionano's Solve software v3.6, setting the conflict filter level for genome maps (-B) and for contig sequences (-N) as two. The second step was done with Hi-C data⁵⁹ using the SALSA2 v2.3 software⁶⁰. Briefly, Hi-C reads were first mapped to the previous scaffolds using the BWA mem module⁶¹, setting the mismatch penalty to eight (-B 8). Chimeric reads alignments at the 3' end were filtered out with Arima Mapping pipeline scripts (https://github.com/ArmaGenomics/mapping_pipeline), and alignments for each pair were then combined into a single file and converted to BED format. Next, we executed the SALSA2 scaffolding pipeline with the default parameter set for the Arima Hi-C Mapping pipeline (-e GATC, GANTC, CTNAG, TTAA -m yes). The manual curation of the Hi-C contact map was done with PretextView⁸⁶. The final primary assembly was named bHarHar1.0 (NCBI accession number GCF_026419915.1). Assembly quality was evaluated using Merqury⁶², gfastats⁶³ and compleasm v0.2.2⁶⁴ software fed with Aves_odb10 and visualized with BlobToolKit⁶⁵. The complete mitochondrial genome assembly was generated from the whole genome sequence data, using the MitoHifi pipeline⁶⁶. The mitogenome of *Bustatur indicus*, the Grey-faced Buzzard, was retrieved from NCBI and used as a reference for searching the mitochondrial scaffolds. Circularization and gene completeness evaluation were also done using MitoHifi.

Transposable element and repeat annotation

We first retrieved all coding sequences (CDS) for all Accipitridae (except for the Harpy Eagle) with available assembled genomes in NCBI by September, 2023. Next, we filtered out CDS with BLAST hits to transposable element proteins deposited in the NCBI nr database, following the 80-80-80 rule (sequences are annotated as TE if they are longer than 80 base pairs, and share at least 80% sequence identity over 80% of their length; see^{67,68}). Then, we created a custom repeat library for the Harpy Eagle genome with the EDTA pipeline v2.0.1⁶⁹, filtering out false positives with the previously filtered CDS dataset (-cnds). We also supplemented annotation with Aves-restricted long interspersed elements (LINEs) and short interspersed elements (SINEs) from DFAM (-curatedlib). The library was fed to RepeatMasker⁷⁰. We grouped all CR1, L1, and L2 LINEs into a single group (referred to as CR1) and also R1 and R2 LINEs into a single group referred to as R1. The LTR TEs found were reclassified using TESorter to retrieve Endogenous RetroVirus (ERVs) from misclassified elements. The RepeatMasker output was parsed to obtain a non-redundant estimation of genome coverage by the repeats and the repeat landscape for the Harpy Eagle genome (<https://github.com/Aureliek/Parsing-RepeatMasker-Outputs>). We assumed the mutation rate of 2.3×10^{-9} substitutions/site/year, estimated from *Ficedula albicollis*⁷¹, recently used for *Aquila chrysaetos* populational studies⁷², and other birds⁷³⁻⁷⁶. TE insertion time was estimated as the divergence (Kimura-2-parameter distance) divided by twice the mutation rate per year. We generated hard and soft-masked versions of the assembled genome using the annotated repeats and BEDTools⁷⁷. Individual TE sequence searches were conducted manually, using online tools to scan DFAM, InterProScan, and UniProtKB (SwissProt).

Gene annotation

We submitted the bHarHar1.0 assembly to NCBI and requested annotation with the Eukaryotic Genome Annotation Pipeline v10.1^{78,79}. Since we did not sequence RNA-seq libraries for the Harpy Eagle, the annotation pipeline was run based on orthology to other birds already deposited in the NCBI Genome database. The RefSeq genome records can be found under accession number GCF_026419915.1-RS_2022_12.

Comparative genomics

Syntenic regions between the Harpy Eagle and other species were investigated using two complementary approaches. We included in our alignments three other Accipitiformes species with available chromosome-level high-quality genomes: the Eurasian Goshawk *Accipiter gentilis* (NCBI accession number: GCF_92944395.1), the Golden Eagle *Aquila chrysaetos* (GCF_900496995.4), the California Condor *Gymnogyps californianus* (GCF_018139145.2), and an outgroup from the Galliformes order represented by the chicken (*Gallus gallus* GCF_016699485.2). While *Accipiter gentilis* and *Aquila chrysaetos* belong to the same family clade as the Harpy Eagle (Accipitridae), *Gymnogyps californianus* belongs to the family Cathartidae, recovered as the sister group to all other Accipitiformes²⁶. First, collinear blocks were identified by directly aligning the genomic sequences using the minimap2 aligner⁸⁰ and represented with dot plots using the D-Genies interactive visualization software⁸¹. Second, macrosyntenic blocks of at least 20 genes were also identified using the MCSScanX method⁸² and visualized with SynVisio⁸³. Chromosomes in all genome assemblies used here were renamed following the NCBI numbering preceded by the first letter in the genus and the first in specific epithet (e.g., chromosome 1 in *Harpia harpyja* deposited genome assembly was renamed as hh1, chromosome 2 as hh2, and so on). The longest sequence for any gene locus was retrieved and used as the primary isoform for orthology inference, using OrthoFinder2 with default parameters⁸⁴. All chromosome translocations were validated by cross-species mapping the HiC data used to scaffold each genome assembly used in this study to the Harpy Eagle genome assembly presented here (Data available in the Genome Ark Database—entries bAccGen1, bAquChr1 and bGalGal1; and DNAzoo (entry *G. californianus*). Mapping was performed for each read pair file with BWA-MEM2⁸⁵, filtering and merging of BAMs with Bellerophon (<https://github.com/davebx/bellerophon>) and image generation with PretextView⁸⁶.

Population dynamics

To reconstruct population dynamics through time and infer the effects of past climatic events in effective population size (N_e) dynamics, we used the pairwise sequentially Markovian coalescent (PSMC) method⁸⁷. The generation time for the Harpy Eagle was calculated from estimates of age to maturity and reproductive longevity, as “age to maturity + ½ reproductive longevity”⁸⁸. We considered values from Watson et al.⁸⁹ (age to maturity = 4.5 years and 5.5 years, reproductive longevity = 30.5 years and 23.5 years, for males and females respectively) with an

estimated mean generation time of 18.5 years. Effective population size and times were scaled using the generation time and mutation rate estimated above, 4.26×10^{-8} substitutions/site/generation. Otherwise, we used the default settings of PSMC. Briefly, we mapped HiFi reads using bwa-mem⁶¹ to the assembled genome and produced a consensus sequence with SAMtools v0.1.19⁹⁰ and bcftools⁹¹. Reads that mapped to the Z and W chromosomes and unanchored contigs were excluded, as the sex chromosomes have different mutation and evolutionary rates. Confidence intervals were estimated using 100 bootstrap replicates⁸⁷.

Data availability

The *Harpia harpyja* (Harpy eagle) assembly is available on NCBI under accession number GCF_026419915.1 (BioProject PRJNA910140). All raw data and the assembly have been deposited on NCBI databases.

Received: 14 May 2024; Accepted: 14 August 2024

Published online: 12 September 2024

References

- Banhos, A., Hrbek, T., Sanaïotti, T. M. & Farias, I. P. Reduction of genetic diversity of the Harpy Eagle in Brazilian tropical forests. *PLoS ONE* **11**, e0148902 (2016).
- BirdLife International. *Harpia harpyja*. *The IUCN Red List of Threatened Species* e.T22695998A197957213 <https://doi.org/10.2305/IUCN.UK.2021-3.RLTS.T22695998A197957213.en>. (2021)
- Banhos, A. *et al.* Long-term female bias in sex ratios across life stages of Harpy Eagle, a large raptor exhibiting reverse sexual size dimorphism. *R. Soc. Open Sci.* **10**, 231443 (2023).
- Aguiar-Silva, F. H., Sanaïotti, T. M. & Luz, B. B. Food habits of the Harpy Eagle, a top predator from the Amazonian rainforest canopy. *J. Raptor Res.* **48**, 24–35 (2014).
- Kaizer, M. *et al.* The prey of the Harpy Eagle in its last reproductive refuges in the Atlantic Forest. *Sci. Rep.* **13**, 18308 (2023).
- Thiollay, J. M. Raptor community structure of a primary rain forest in French Guiana and effect of human hunting pressure. *J. Raptor Res.* **18**, 117–122 (1984).
- Lerner, H. R. L., Johnson, J. A., Lindsay, A. R., Kiff, L. F. & Mindell, D. P. It's not too late for the harpy eagle (*Harpia harpyja*): High levels of genetic diversity and differentiation can fuel conservation programs. *PLoS One* **4**, e7336 (2009).
- de Oliveira, E. H. C. *et al.* Chromosome reshuffling in birds of prey: The karyotype of the world's largest eagle (Harpy eagle, *Harpia harpyja*) compared to that of the chicken (*Gallus gallus*). *Chromosoma* **114**, 338–343 (2005).
- Nie, W. *et al.* Multidirectional chromosome painting substantiates the occurrence of extensive genomic reshuffling within Accipitriformes. *BMC Evol. Biol.* **15**, 1–17 (2015).
- Sotero-Caio, C. G., Platt, R. N., Suh, A. & Ray, D. A. Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* **9**, 161–177 (2017).
- Kapusta, A. & Suh, A. Evolution of bird genomes—a transposon's-eye view. *Ann. N. Y. Acad. Sci.* **1389**, 164–185 (2017).
- Manthey, J. D., Moyle, R. G. & Boissinot, S. Multiple and independent phases of transposable element amplification in the genomes of piciformes (woodpeckers and allies). *Genome Biol. Evol.* **10**, 1445–1456 (2018).
- Carotti, E. *et al.* LTR Retroelements and bird adaptation to arid environments. *Int. J. Mol. Sci.* **24**, 6332 (2023).
- Xu, L. *et al.* Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. *Nat. Ecol. Evol.* **3**, 834–844 (2019).
- Weissensteiner, M. H. *et al.* Discovery and population genomics of structural variation in a songbird genus. *Nat. Commun.* **11**, 3403 (2020).
- Benham, P. M. *et al.* Remarkably high repeat content in the genomes of sparrows: The importance of genome assembly completeness for transposable element discovery. *Genome Biol. Evol.* <https://doi.org/10.1093/gbe/evae067> (2024).
- Zhang, G. *et al.* Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
- Ellegren, H. Evolutionary stasis: The stable chromosomes of birds. *Trends Ecol. Evol.* **25**, 283–291 (2010).
- Tagliarini, M. M., Nagamachi, C. Y., Pieczarka, J. C. & De Oliveira, E. H. C. Description of two new karyotypes and cytotoxic considerations on Falconiformes. *Rev. Bras. Ornitol.* **15**(2), 261–266 (2007).
- Germain, R. R. *et al.* Species-specific traits mediate avian demographic responses under past climate change. *Nat. Ecol. Evol.* **7**, 862–872 (2023).
- Waters, P. D. *et al.* Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proc. Natl. Acad. Sci.* **118**, e2112494118 (2021).
- Nanda, I. *et al.* Synteny conservation of chicken Macrochromosomes 1–10 in different avian lineages revealed by cross-species chromosome painting. *Cytogenet. Genome Res.* **132**(3), 165–181 (2011).
- Feng, S. *et al.* Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
- FitzSimmons, N. N., Moritz, C. & Moore, S. S. Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Mol. Biol. Evol.* **12**(3), 432–440 (1995).
- Simakov, O. *et al.* Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci. Adv.* **8**, eabi5884 (2022).
- Nishida, C., Ishishita, S., Yamada, K., Griffin, D. K. & Matsuda, Y. Dynamic chromosome reorganization in the osprey (*Pandion haliaetus*, Pandionidae, Falconiformes): Relationship between chromosome size and the chromosomal distribution of centromeric repetitive DNA sequences. *Cytogenet. Genome Res.* **142**, 179–189 (2014).
- Kretschmer, R., Ferguson-Smith, M. A. & De Oliveira, E. H. C. Karyotype evolution in birds: From conventional staining to chromosome painting. *Genes (Basel)* **9**, 181 (2018).
- Huang, Z. *et al.* Recurrent chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. *Nat. Commun.* **13**, 944 (2022).
- Kaiser, V. B., van Tuinen, M. & Ellegren, H. Insertion events of CR1 retrotransposable elements elucidate the phylogenetic branching order in galliform birds. *Mol. Biol. Evol.* **24**, 338–347 (2007).
- Suh, A., Smeds, L. & Ellegren, H. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biol.* **13**, e1002224 (2015).
- Galbraith, J. D., Kortschak, R. D., Suh, A. & Adelson, D. L. Genome stability is in the eye of the beholder: CR1 retrotransposon activity varies significantly across avian diversity. *Genome Biol. Evol.* **13**, evab259 (2021).
- Catanach, T. A., Halley, M. R. & Pirro, S. Enigmas no longer: Using Ultraconserved elements to place several unusual hawk taxa and address the non-monophyly of the genus *Accipiter* (Accipitriformes: Accipitridae). *bioRxiv* <https://doi.org/10.1101/2023.07.13.548898> (2023).
- Prost, S. *et al.* Comparative analyses identify genomic features potentially involved in the evolution of birds-of-paradise. *Gigascience* **8**, giz003 (2019).

34. Kawakami, T. *et al.* A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol. Ecol.* **23**, 4035–4058 (2014).
35. Naville, M. & Volff, J.-N. Endogenous retroviruses in fish genomes: From relics of past infections to evolutionary innovations?. *Front. Microbiol.* **7**, 1197 (2016).
36. Brookfield, J. F. Y. The ecology of the genome—mobile DNA elements and their hosts. *Nat. Rev. Genet.* **6**, 128–136 (2005).
37. Gozashti, L., Feschotte, C. & Hoekstra, H. E. Transposable element interactions shape the ecology of the deer mouse genome. *Mol. Biol. Evol.* **40**, msad069 (2023).
38. Barreto, C. & Alves, M. Birds of prey in ancient Amazonia: Predation and perspective in ceramic iconography. In *The Art and Archaeology of Human Engagements with Birds of Prey: From Prehistory to the Present* (ed. Wallis, R. J.) (Bloomsbury Publishing Plc, 2023).
39. Lorenzen, E. D. *et al.* Species-specific responses of late quaternary megafauna to climate and humans. *Nature* **479**(7373), 359–364 (2011).
40. Farias, I. P., Santos, W. G., Gordo, M. & Hrbek, T. Effects of forest fragmentation on genetic diversity of the critically endangered primate, the pied tamarin (*Saguinus bicolor*): Implications for conservation. *J. Hered.* **106**, 512–521 (2015).
41. van der Kaars, S. *et al.* Humans rather than climate the primary cause of *Pleistocene megafaunal* extinction in Australia. *Nat. Commun.* **8**, 14142 (2017).
42. Bennett, M. R. *et al.* Evidence of humans in North America during the last glacial maximum. *Science* **1979**(373), 1528–1531 (2021).
43. Becerra-Valdivia, L. & Higham, T. The timing and effect of the earliest human arrivals in North America. *Nature* **584**, 93–97 (2020).
44. Ardelean, C. F. *et al.* Evidence of human occupation in Mexico around the last glacial maximum. *Nature* **584**, 87–92 (2020).
45. Pansani, T. R. *et al.* Evidence of artefacts made of giant sloth bones in central Brazil around the last glacial maximum. *Proc. R. Soc. B* **290**, 20230316 (2023).
46. Surovell, T. A. *et al.* Late date of human arrival to North America: Continental scale differences in stratigraphic integrity of pre-13,000 BP archaeological sites. *PLoS One* **17**, e0264092 (2022).
47. Bergman, J. *et al.* Worldwide late pleistocene and early holocene population declines in extant megafauna are associated with *Homo sapiens* expansion rather than climate change. *Nat. Commun.* **14**, 7679 (2023).
48. Cooke, R. *et al.* Undiscovered bird extinctions obscure the true magnitude of human-driven extinction waves. *Nat. Commun.* **14**, 8116 (2023).
49. Andermann, T., Faurby, S., Turvey, S. T., Antonelli, A. & Silvestro, D. The past and future human impact on mammalian diversity. *Sci. Adv.* **6**, eabb2313 (2020).
50. Stewart, M., Carleton, W. C. & Groucutt, H. S. Reply to: Accurate population proxies do not exist between 11.7 and 15 ka in North America. *Nat. Commun.* **13**, 4693 (2022).
51. Stewart, M., Carleton, W. C. & Groucutt, H. S. Climate change, not human population growth, correlates with late quaternary megafauna declines in North America. *Nat. Commun.* **12**, 965 (2021).
52. Sasaki, M., Lange, J. & Keeney, S. Genome destabilization by homologous recombination in the germ line. *Nat. Rev. Mol. Cell Biol.* **11**, 182–195 (2010).
53. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Preprint at (2010).
54. De Coster, W., Weissensteiner, M. H. & Sedlazeck, F. J. Towards population-scale long-read sequencing. *Nat. Rev. Genet.* **22**, 572–587 (2021).
55. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
56. 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).
57. Cheng, H. *et al.* Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* **40**, 1332–1335 (2022).
58. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
59. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
60. Ghurury, J. *et al.* Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Comput. Biol.* **15**, e1007273 (2019).
61. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14), 1754–1760 (2009).
62. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 1–27 (2020).
63. Formenti, G. *et al.* Gfastats: Conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics* **38**, 4214–4216 (2022).
64. Huang, N. & Li, H. Compleasm: A faster and more accurate reimplementation of BUSCO. *Bioinformatics* **39**, btad595 (2023).
65. Challis, R., Richards, E., Rajan, J., Cochrane, G. & Blaxter, M. BlobToolKit—interactive quality assessment of genome assemblies. *G3: Genes Genomes Genetics* **10**, 1361–1374 (2020).
66. Uliano-Silva, M. *et al.* MitoHiFi: A python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinform.* **24**, 1–13 (2023).
67. Wells, J. N. & Feschotte, C. A field guide to eukaryotic transposable elements. *Annu. Rev. Genet.* **54**, 539–561 (2020).
68. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
69. Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 1–18 (2019).
70. Smit, A., Hubley, R. & Grenn, P. RepeatMasker Open-4.0. *RepeatMasker Open* (2015).
71. Smeds, L., Qvarnström, A. & Ellegren, H. Direct estimate of the rate of germline mutation in a bird. *Genome Res.* **26**, 1211–1218 (2016).
72. Sato, Y. *et al.* Population history of the golden eagle inferred from whole-genome sequencing of three of its subspecies. *Biol. J. Linn. Soc.* **130**, 826–838 (2020).
73. Ericson, P. G. P. *et al.* A 14,000-year-old genome sheds light on the evolution and extinction of a Pleistocene vulture. *Commun. Biol.* **5**, 857 (2022).
74. 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.* **27**, 99–111 (2018).
75. Dussex, N. *et al.* Complete genomes of two extinct New Zealand passerines show responses to climate fluctuations but no evidence for genomic erosion prior to extinction. *Biol. Lett.* **15**, 20190491 (2019).
76. Hayes, K., Barton, H. J. & Zeng, K. A study of faster-Z evolution in the great tit (*Parus major*). *Genome Biol. Evol.* **12**, 210–222 (2020).
77. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
78. Pruitt, K. D. *et al.* RefSeq: An update on mammalian reference sequences. *Nucl. Acids Res.* **42**, D756–D763 (2014).
79. Rhie, A. *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).
80. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**(18), 3094–3100 (2018).

81. Cabanettes, F. & Klopp, C. D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).
82. Wang, Y. *et al.* MCSScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucl. Acids Res.* **40**, e49–e49 (2012).
83. Bandi, V. *et al.* Visualization tools for genomic conservation. In *Plant Bioinformatics. Methods in Molecular Biology*, Vol. 2443 (eds Edwards, D.) (Humana, New York, NY, 2022). https://doi.org/10.1007/978-1-0716-2067-0_16.
84. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome. Biol.* **20**, 238. <https://doi.org/10.1186/s13059-019-1832-y> (2019).
85. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* (eds Vasimuddin, M. *et al.*) 314–324 (IEEE, 2019).
86. Howe, K. *et al.* Significantly improving the quality of genome assemblies through curation. *Gigascience* **10**, gaa1153 (2021).
87. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
88. Vilaça, S. T. *et al.* Divergence and hybridization in sea turtles: Inferences from genome data show evidence of ancient gene flow between species. *Mol. Ecol.* **30**, 6178–6192 (2021).
89. Watson, R. T., McClure, C. J. W., Vargas, F. H. & Jenny, J. P. Trial restoration of the Harpy Eagle, a large, long-lived, tropical forest raptor Panama and Belize. *J. Raptor Res.* **50**, 3–22 (2016).
90. Li, H. *et al.* The sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16), 2078–2079 (2009).
91. Danecek, P. & McCarthy, S. A. BCFTools/csqr: Haplotype-aware variant consequences. *Bioinformatics* **33**, 2037–2039 (2017).

Acknowledgements

The *H. harpyja* genome sequencing initiative is part of the ICMBio / ITV Genomics of the Brazilian Biodiversity (GBB) consortium, the LEGAL/UFAM and the Instituto Nacional de Pesquisas da Amazônia's Harpy Eagle Project. It is also part of the Vertebrate Genomes Project (VGP) ordinal references program and the Earth Biogenome Project (EBP), with support to E. D. J. from Howard Hughes Medical Institute and the Rockefeller University. We thank Natália Aparecida Souza Lima (IBAMA) and Diogo Cesar Lagroteria Oliveira Faria (Instituto Chico Mendes de Conservação da Biodiversidade—ICMBio) for providing the sequenced Harpy Eagle sample from Cetas/IBAMA in Manaus (AM-Brazil). The Nuremberg Zoo provided critical support to LEGAL/UFAM. The following researchers are supported by a Productivity Research Fellowship from the Brazilian National and Technological Development Council—CNPq: A.A. (CNPq #309243/2023-8), I. P. F. (CNPq #316531/2023-5), and T. H. (CNPq #316532/2023-1).

Author contributions

A.A., A.B., G.O., I.P.F., T.H. and T.M.S. conceived and designed the study. A.B. and T.H. coordinated the collection and shipment of the sample. O.F. supervised DNA extraction, library construction and sequencing work. F.A.A. assembled the genome. A.T. performed manual curation. Y.S. generated the gEVAL database. E.D.J. and G.F. supervised the genome assembly. A.A. and V.S. supervised the analysis. L.E.C.C. analyzed the data and interpreted the results. A.A., G.O., R.R.M.O., S.T.V. and V.S. coordinated and provided guidance. L.E.C.C. and S.T.V. drafted the manuscript, with input from all other authors. L.E.C.C., S.T.V. and A.A. finalized the manuscript. The authors' initials are ordered alphabetically. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-70305-w>.

Correspondence and requests for materials should be addressed to A.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024