Title

A high-quality reference genome for the Ural Owl (*Strix uralensis*) enables investigations of cell cultures as a genomic resource for endangered species

Authors

Ioannis Chrysostomakis[1], Annika Mozer[1], Camilla Bruno Di-Nizo[1], Dominik Fischer[2], Nafiseh Sargheini[3], Laura von der Mark[1], Bruno Huettel[3], Jonas J. Astrin[1], Till Töpfer[1], *Astrid Böhne[1]

*corresponding author

Affiliations

[1]Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Adenauerallee 127, 53113 Bonn, Germany

[2]Zoo Wuppertal, Wuppertal, Germany

[3]Max Planck Genome-Centre Cologne, Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, 50829 Cologne, Germany

Emails

I.Chrysostomakis@leibniz-lib.de, a.mozer@leibniz-lib.de, C.DiNizo@leibniz-lib.de, fischer@zoo-wuppertal.de, nsargheini@mpipz.mpg.de, L.vonderMark@leibniz-lib.de, huettel@mpipz.mpg.de, J.Astrin@leibniz-lib.de, T.Toepfer@leibniz-lib.de, a.boehne@leibniz-lib.de

Abstract

Background

Reference genomes have a wide range of applications. Yet, we are from a complete genomic picture for the tree of life. We here contribute another piece to the puzzle by providing a high-quality reference genome for the Ural Owl (*Strix uralensis*), a species of conservation concern and efforts affected by habitat destruction and climate change.

Results

We generated a reference genome assembly for the Ural Owl based on high-fidelity (HiFi) long reads and chromosome conformation capture (Hi-C) data. It figures amongst the best avian genome assemblies currently available (BUSCO completeness of 99.94 %). The primary assembly had a size of 1.38 Gb with a scaffold N50 of 90.1 Mb, while the alternative assembly had a size of 1.3 Gb and a scaffold N50 of 17.0 Mb. We show an exceptionally high repeat content (21.07 %) that is different from those of other bird taxa with repeat extensions. We confirm a *Strix* characteristic chromosomal fusion and support the observation that bird microchromosomes have a higher density of genes, associated with a reduction in gene length due to shorter introns. An analysis of gene content provides evidence of changes in the keratin gene repertoire as well as modifications of metabolism genes of owls. This opens an avenue of research if this is related to flight adaptations. The population size history of the Ural Owl decreased over long periods of time with increases during the Eemian interglacial and stable size during the last glacial period. Ever since it is declining to its currently lowest effective population size. We also investigated cell culture of progressive passages as a tool for genetic resources. Karyotyping of passages confirmed no large variants, while a SNP analysis revealed a low presence of short variants across cell passages.

Conclusions

The established reference genome is a valuable resource for ongoing conservation efforts,

but also for (avian) comparative genomics research. Further research is needed to determine whether cell culture passages can be safely used in genomic research.

Keywords

Strix uralensis, Strigidae, karyotyping, genome sequence, genome annotation, cell culture, SNP, variant

Background

High-quality reference genomes are rapidly becoming available for many branches of the tree of life (https://www.earthbiogenome.org) [1]. These data are now increasingly used for comparative genomic studies on large evolutionary timescales trying to link phenotypes to genotypes [2]. However, even in genomically and traditionally well-studied groups such as birds, most lineages still lack high-quality reference genome assemblies that would allow for detailed studies of genome evolution.

Typical avian karyotypes are composed of macro- and microchromosomes (but see [3,4]). Compared to macrochromosomes, which are typically between 30 and 250 mega base pairs (Mb) in size, microchromosomes have an average size of 12 Mb, although microchromosomes as small as 3.4 Mb have been observed [5,6]. Despite recent efforts to characterise avian genomes and understand their karyotype evolution, less than 10% of all known bird species have a characterized karyotype [7]. The diploid number of about half of these varies between 78 to 82 chromosomes [1]. Regarding the family Strigiformes (owls), karyotype information is available for 13 % of species [7]. Interestingly, microchromosomes encode half of the genes in birds, although they account for only about a quarter of the genome sequence [5,8]. Moreover, the mutation rate of microchromosomes is significantly higher than that of macrochromosomes [9]. Therefore, avian karyotypes, genome structure and especially the microchromosomes deserve more cytogenetic and molecular attention.

To this aim, we here provide a first high-quality reference genome for the Ural Owl (*Strix uralensis*). This species is one of the largest Eurasian owls, inhabiting the Palaearctic lowlands up to the treeline, mainly in the taiga forest belt over a large uninterrupted range from Scandinavia through Siberia to Sakhalin and the Japanese islands. It also occurs in geographically isolated, mixed and deciduous forests of southeastern and central Europe (southern Germany, Czech Republic, Austria, Slovenia and Poland; partly supported by reintroductions). So far, 11 subspecies have been described from its vast distribution based

on differences in size and colouration [10]; although not all of these have been widely accepted [11]. Furthermore, the molecular data at hand (i.e., mitochondrial and nuclear marker genes) do not support morphology-based taxonomic distinctions [12].

Ural Owls are nocturnal hunters of small mammals and birds and usually stay in their territories throughout the year [11,13]. As the Ural Owl is sedentary and nests in hollow stumps or tree holes [14,15], it is affected by ecosystem degradation [16]. Nesting sites have been reduced by intensive logging activities, agricultural use, and forestry management [10]. While globally still considered under the IUCN Red List category “Least Concern”, *S. uralensis* went extinct in Austria, southern Germany, and the Czech Republic in the last century, mainly due to direct persecution [17–19]. Successful reintroductions have taken place in these countries (e.g. [17–19]). These central European reintroductions have restored gene flow between the remaining Alpine and European populations [12,20]. The Ural Owl will likely further be affected by climate change, potentially shifting its range to more northern regions [21] and altering breeding times [22]. Correspondingly, the Ural Owl is a species of conservation measures in the European Union under the EU Birds Directive and Nature Habitats Directive. It is also part of the Bern Convention (Convention on the Conservation of European Wildlife and Nature Habitats). International trade of all Strigiformes is regulated by the Convention on International Trade in Endangered Species of Wild Fauna and Flora CITES.

Cryobanking, defined as the preservation of viable cells and tissues at ultracold temperatures, typically using liquid nitrogen, is considered paramount in preserving the genetic variability of species, especially those facing population decline as the Ural Owl, to ensure population health and persistence [23,24]. Although some instances have been reported where long-term cell culture generated genetic instability and heteroploidy [25,26], it is still unclear how frequent such a phenomenon is and at which stage of cell cultivation it occurs.

Herein, we generated a reference genome for the Ural Owl as a genomic resource to facilitate further research on this species and on Strigidae more generally. We assess the genome assembly quality and provide a first analysis of its gene content. As a species of potential

conservation concern and as a proof of principle, we assessed the application of cell culture to produce sufficient DNA in terms of quantity and quality to allow genomics for species with limited biological material. We investigated mutation as a function of passage number (i.e., the transfer of cells from vessel to vessel). To this end, we obtained a cell culture from the same individual that was genome-sequenced, and cultivated the cell lines until passage 10 and subsequently sequenced replicates of passages 5 and 10 (Figure 1).

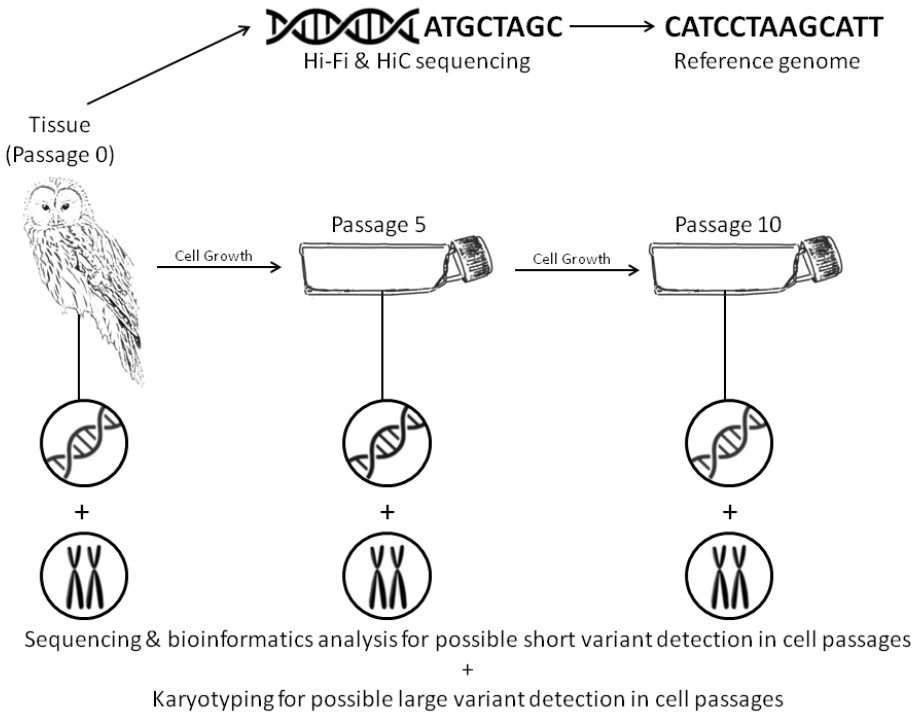


Figure 1: Reference genome and cell passage variant detection workflow. Tissue from a male Ural Owl (*Strix uralensis*) is extracted and sequenced, assembled and annotated to provide a reference genome. Additionally, a cell culture is established from the primary tissue. From passage 0 (primary tissue), passage 5 (three independent replicates) and passage 10 (four independent replicates) cells are harvested for short-read sequencing and karyotyping.

Data Description

In order to provide valuable genomic resources to the scientific community studying avian ecology and phylogenomics and to investigate the potential of lab-grown cells for use in DNA sequencing, skin cells were harvested from a 10-year-old, recently deceased male Ural Owl individual. The skin samples were, originally, frozen at -80°C and later grown in an appropriate medium and used for DNA sequencing. We performed PACBIO long-read sequencing of muscle tissue, which produces high-quality, long DNA fragments. We used cultured cells for Hi-C sequencing, which allows us to estimate physical proximity of DNA molecules inside the cell to create a bird genome assembly with the highest gene completeness score to date. Next, we grew the harvested skin cells for multiple generations to understand whether this process causes damage to chromosome structure and the accumulation of DNA mutations. In the future, this data can be used to study avian phylogenomics and diversity as well as further understand the unique traits of owls. All sequence data of this study can be accessed from INSDC under the BioProject ID PRJNA1212906. Processed data are available from Zenodo under DOI 10.5281/zenodo.14676512.

Analyses

Read quality control and estimation of genome size and heterozygosity

After quality control, filtering, and decontamination the final set of HiFi reads used was composed of 5,078,732 reads with a total length of ~58 Gb and the Hi-C reads used were composed of 79.7 million reads with a total length of ~ 20 Gb.

Using a k-mer size of 21, GenomeScope was able to predict a genome size of 1,292,799,460 bp, a repeat length of 188,615,362 bp, a heterozygosity of 0.2 % (this would translate to 2 heterozygous sites per 1 kb, a commonly reported heterozygosity indicator for birds) and a read error rate of 0.14 % (Supplementary Table S1). Smudgeplot and GenomeScope both verified the diploid status of the individual. (Supplementary Figure S1; Figure 2). The genome did not reveal any large runs of homozygosity (ROH).

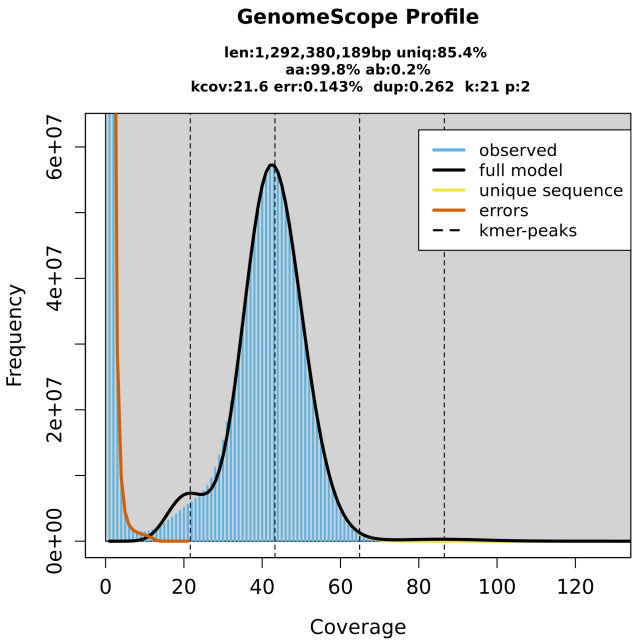
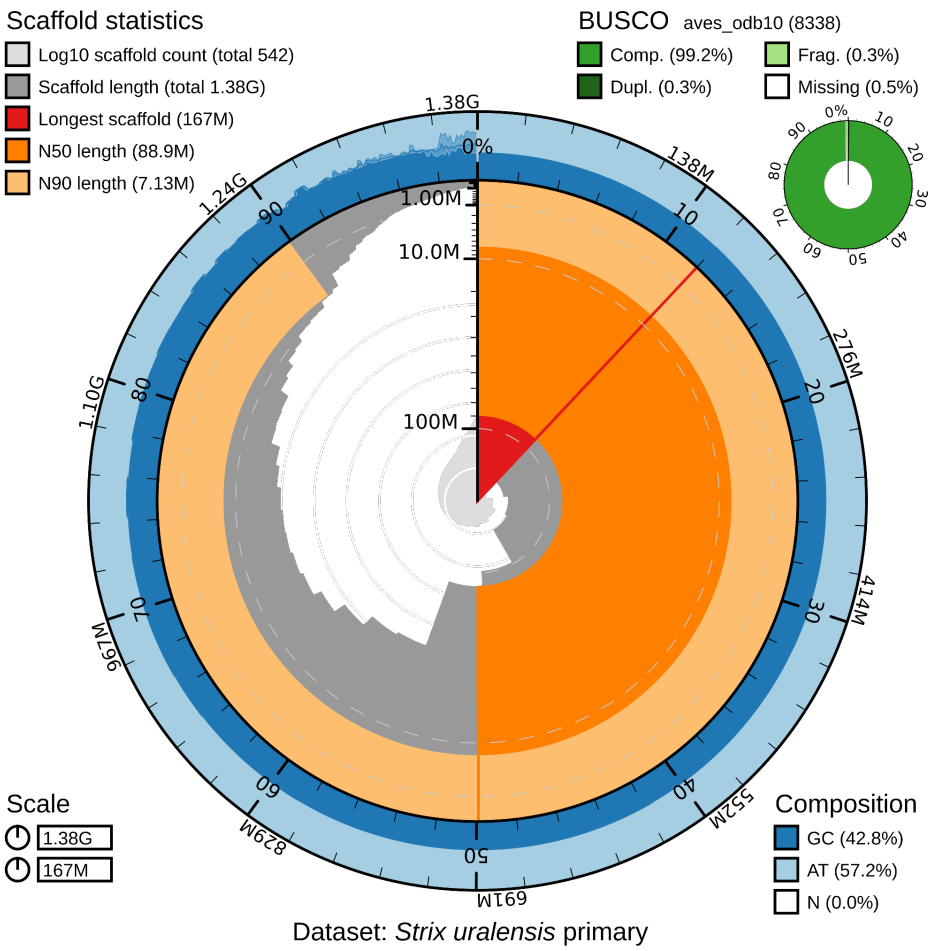


Figure 2: K-mer genome profile of *Strix uralensis* generated from PacBio HiFi reads with GenomeScope2.The y-axis shows the k-mer counts and the x-axis shows sequencing depth. The first peak corresponds to heterozygous k-mers and the second larger peak to homozygous k-mers with a coverage of ~42 x.

Reference genome

The optimal assembly was created with Hifiasm parameters “-l2 --n-weight 5 --n-perturb 50000 --f-perturb 0.5 -D 10 -N 150 -s 0.4” (Supplementary Table S1).



Dataset: *Strix uralensis* primary

Figure 3: Snail plot summary of assembly statistics for *Strix uralensis* primary assembly. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 1,381,000,783 bp assembly. The distribution of sequence lengths is shown in dark grey with the plot radius scaled to the longest sequence present in the assembly (166,530,430 bp, shown in red). Orange and pale-orange arcs show the N50 and N90 sequence lengths (88,922,949 and 7,132,230 bp), respectively. The pale grey spiral shows the cumulative sequence count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT and N percentages in the same bins as the inner plot. A summary of complete, fragmented, duplicated and missing BUSCO genes in the aves_odb10 set is shown in the top right.

Genome Quality Metrics

We could place 93.6 % of assembled scaffolded genome sequence data into 41 chromosomes, which is consistent with the karyotype of the species (Figure 3). We also detected no contamination as all scaffolds aligned to sequences of other avian genomes (Supplementary Figure S2). Our Hi-C contact map further supported the high contiguity of the primary assembly, by showing no remaining conflicts and little to no scaffolds with strong contacts to non-repeat regions (Figure 4, Supplementary Figures S3 and S4).

The Merqury Quality Value (QV) score, which is the proportion of the assembly sequence supported by HiFi reads, was estimated for both haplomes. We obtained a score of 64.2 (equivalent to an error probability of 3.8238e-07 %) for the primary and 57.4 (equivalent to an error probability of 1.80919e-06 %) for the alternate assembly (Supplementary Table S2). We also find a completeness score of 98.36 % for the primary and a combined 99.81 % for the two haplomes, representing the fraction of high-quality k-mers from the reads present in the assembly. This further supports the completeness and accuracy of the assembly (Supplementary Table S2).

Aligning the PacBio HiFi, and Illumina Hi-C reads to both haplomes revealed comparable coverage levels (primary: 41.78 ± 12.18, 14.18 ± 91.47-fold respectively; alternate: 34.15 ± 23.24, 12.10 ± 107.37 respectively), mapping rates (primary: 99.85 %, 99.92 % respectively; alternate: 80.52 %, 86.6 % respectively) and mapping quality scores (primary: 36.93, 28.90 respectively; alternate: 28.11, 8.2 respectively). These results further indicate that the assembly is well-phased with a minimal amount of assembly bias and errors (Supplementary Table S3).

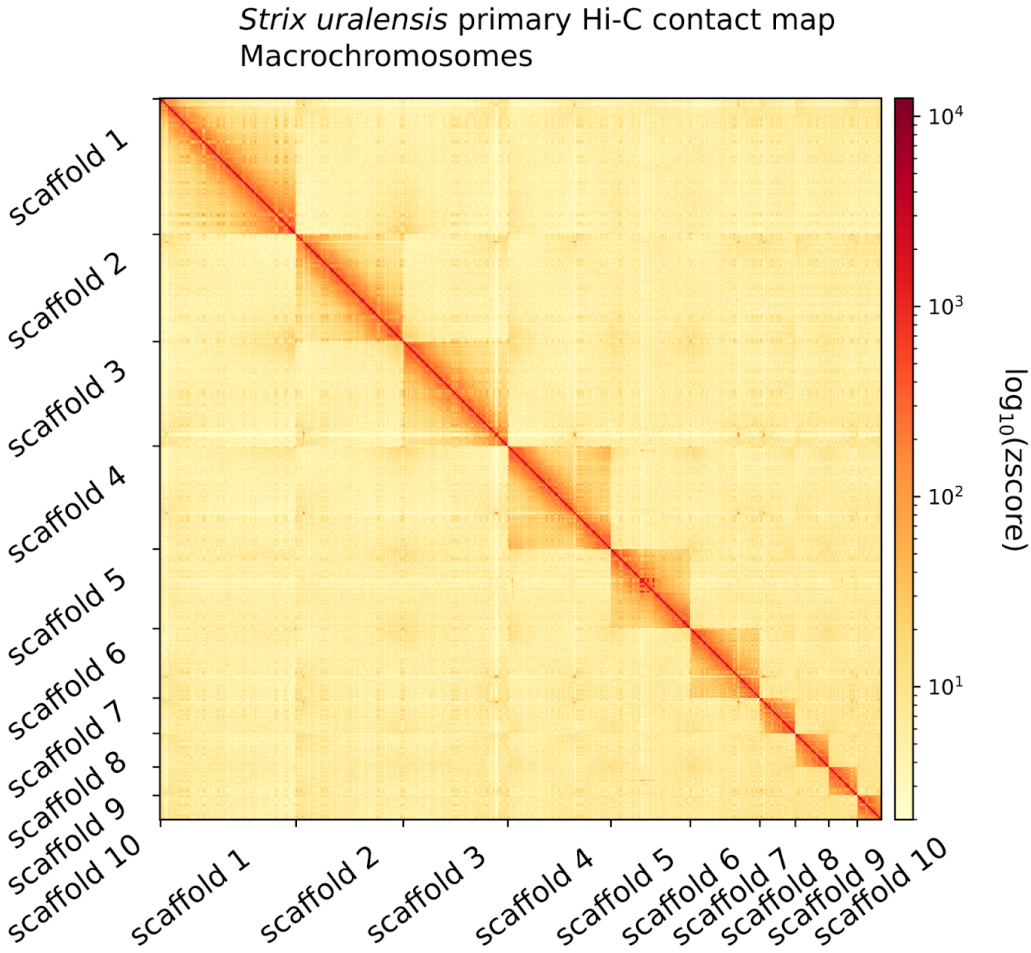


Figure 4: *Strix uralensis* primary haplome Hi-C contact map showing spatial interactions between the ten largest chromosomes. Chromosomes are ordered by size from left to right and from top to bottom. The red diagonal corresponds to intra-chromosomal contacts and depicts chromosome boundaries. The frequency of contacts is shown on a logarithmic heatmap scale. Plot generated with HiCExplorer.

191 **Table 1**: Assembly statistics of the primary and alternate genome assembly of *Strix uralensis*.

Assembly statistics	**Primary**	**Alternate**
Assembly size [bp]	1,381,008,983	1,262,176,999
GC content [%]	42.77	42.81
Contigs	512	15,615
N50	90,173,155	17,018,198
L50	6	18
L90	28	8,171
Ns per 100 kb	2.94	68.29
Merqury Error (%)	3.82794e-07	1.80919e-06
Merqury QV score	64.17	57.43
Complete BUSCOs [%]	99.94	70.82

Genome Repeat and Gene Annotation

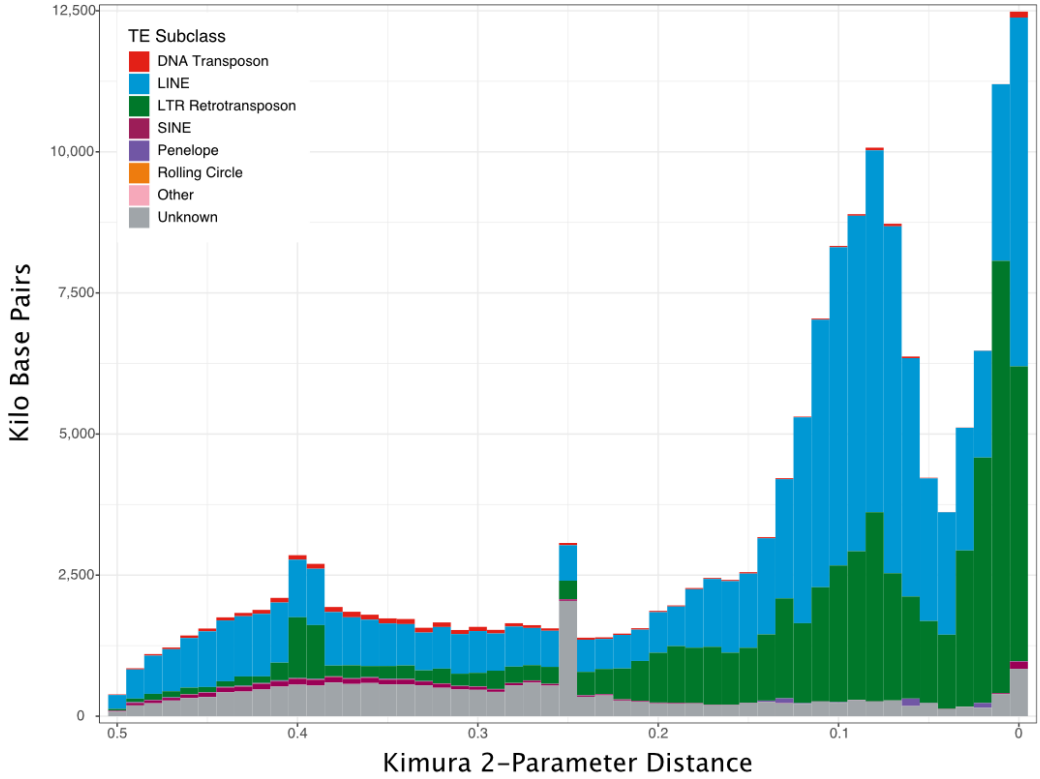


Figure 5: *Strix uralensis* primary haplome repeat landscape. The x axis shows the Kimura substitution of detected repeat categories and the y axis the number of repeats detected for each TE family in kilo base pairs. Detected subclasses are colour-coded as indicated in the inset. The genome assembly was masked using EarlGrey.

Repeat landscapes depict the clustering of transposable elements (TEs) in relation to their Kimura substitution rates, which measures the divergence of TEs from their respective consensus sequence. Lower Kimura substitution rates indicate recent transposition events, while higher rates suggest older events. From the landscape of the primary haplome (Figure 5, Supplementary Figures S5 and S6), a strong signal for a recent repeat expansion of LINEs (long interspersed nuclear elements) and an even more recent expansion of LTR (long terminal repeat) retrotransposons as well as a well-maintained large number of older repeats are visible. This might support a repeat expansion in the Ural Owl or the genus *Strix*. A third and older expansion is dominated by LINEs and unknown repeats suggesting that they are either a new or unique feature of *Strix* and a reference or consensus might not yet exist in the reference databases.

Gene annotation

For the primary haplome, we were able to annotate a total of 17,977 protein-coding genes which cover ~33.6 % of the total size of the assembly (Supplementary Table S4). We detected 182,313 exons and 164,373 introns. Compared to the Swiss-Prot and UniProt databases we were able to match 16,461 and 17,511 of our genes to annotations respectively (Supplementary Table S5).

We next investigated gene distribution along the genome. Using a 30 Mb cutoff [5,6], we identified ten macrochromosomes and 31 microchromosomes based on our assembly (Figure 4 and Supplementary Figure S7). Despite their size, microchromosomes had a higher gene density than macrochromosomes. While there are comparatively more genes on microchromosomes, these genes are shorter than those on macrochromosomes, mainly due to shorter introns (Supplementary Figure S7).

To shed light on the genome annotation content, we compared the Ural Owl genome to other high-quality genomes of the Aves lineage, including several owl species. Gene expansion (gain) and contraction (loss) among our selected species found 316 gene family gains in the Ural Owl and 207 losses, 168 of which were, presumably, completely lost and, thus, have no representative in the Ural Owl genome assembly (Figure 6).

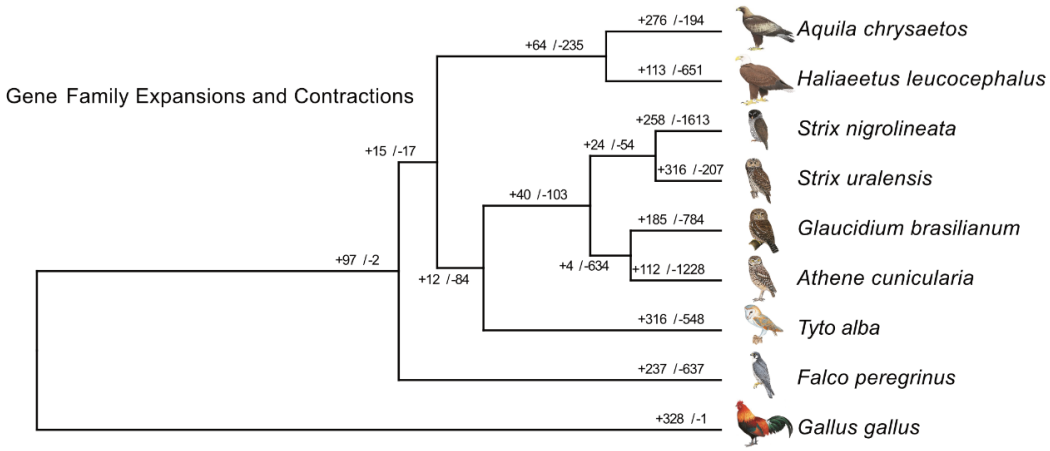


Figure 6: Ultrametric phylogenetic tree of selected Neoaves species and *Strix uralensis*. Numbers indicate gene family expansions (+) and contractions (-). Bird drawings from https://birdsoftheworld.org/.

231

232 Additionally, we found 81 gene families unique to the Ural Owl that do not have orthologs in
233 the other species (Supplementary Table S5). We further found that genomes of lower quality,
234 such as those of the Ferruginous pygmy owl, *Glaucidium brasilianum,* and the Black-and-white
235 Owl, *Strix nigrolineata*, had more gene losses, which are hence probably not biologically true
236 but represent technical limitations. A Gene Ontology (GO) term analysis of the genes unique
237 to the Ural Owl revealed many interesting gene families that due to the low quality of the Black-
238 and-white Owl genome might also be interpreted as partially representing the *Strix* genus
239 (Supplementary Figure S8). Among these categories we note several GO terms relevant to
240 characteristic traits of the Ural Owl, namely its adaptation to dim-light conditions and a
241 sedentary and predatory hunting strategy. The “animal organ morphogenesis" parent GO term
242 groups the child GO terms “eye development”, “sensory organ development”, “neurogenesis”
243 and “heart development”, all of which point to adaptations of *Strix*, either to their environment
244 or lifestyle.

245 Next, we investigated gene gain and loss at nodes that are supported by more than one
246 reference genome which would make them more robust and at the same time informative
247 about clade-specific genomic changes.

248 We observed 15 gene family gains and 17 losses in the last common ancestor of Strigiformes
249 and Accipitriformes (hawks, eagles, vultures, kites), both characterized by a predatory
250 lifestyle. Overarching GO terms among the gained genes included “behavior”, metabolic,
251 cellular and developmental processes. Notably, the child GO terms contained many terms
252 related to general and cellular metabolism (e.g., “ATP metabolic process”, “carbohydrate
253 derivative metabolic process”, “cellular lipid catabolic process", "cellular lipid metabolic
254 process”). We identified three gains in keratin genes (feather and scale keratin), two related
255 to histones/histone modification and two related to skeletal muscle functioning (BEST3, CKB).

256 The gene losses comprised several mitochondrial genes which we attribute to lower quality of
257 mitochondrial gene annotation of the used genomes since contrastingly to the ortholog based

258 results, we could annotate 36 out of 37 mitochondrial genes in our assembly.

259 The other gene losses concerned uncharacterized gene families as well as a ribonucleoprotein
260 (IMP4), the claudin gene family encoding for tight junction proteins and a DNA polymerase.

261 We identified 12 gene gains and 84 losses reconstructed for the common ancestor of owls.
262 We again found an expansion of the keratin gene repertoire (gain of one keratin and one scale-
263 keratin like gene). GO parental terms of gains pointed again to metabolic changes but also
264 those associated with the immune system. The gains contained also an olfactory receptor.
265 The much more numerous losses were associated even at the higher level with many different
266 GO categories again often related to metabolism (e.g., “regulation of amide metabolic
267 process”, “pyridine−containing compound metabolic process”).

268 Chromosome Scale Syntenies

269 Synteny with the chromosome-level assemblies of *Strix aluco* and *Bubo scandiacus* identified
270 the Z chromosome of our male individual. It is the fifth largest chromosome in the Ural Owl
271 assembly. The synteny between the two *Strix* genomes shows no major syntenic differences
272 (Figure 7). This is also mostly true in the comparison to the Snowy Owl with the exception of
273 the Z chromosome, which shows some internal rearrangements compared to the two *Strix*
274 species. Whether this is caused by assembly quality and accuracy remains to be investigated.
275 Compared to chicken and zebrafinch however, we identified several large-scale changes. We
276 detected a fusion of chromosomes 5 and 6 of the Snowy Owl (corresponding to parts of
277 chicken chromosome 4 and chromosome 5 and zebrafinch chromosome 4 and parts of 5,
278 Figure 7 and Supplementary Table S6) into chromosome 4 of the two *Strix* assemblies. This
279 is supported by previous cytogenetic analyses [27]. The remaining part of chicken
280 chromosome 4 corresponds to Ural Owl chromosome 13, Snowy Owl chromosome 12 and
281 zebrafinch 4a. The remaining part of chicken and zebrafinch chromosome 5 corresponds to
282 two chromosomes in the Ural Owl (chromosomes 16 and 30). As in the zebrafinch, chicken
283 chromosome 1 corresponds to two chromosomes in the Snowy Owl (chromosomes 2 and 6).

For the remaining chromosomes of our Ural Owl assembly, we could mostly identify 1:1 relationships with chromosomes of Snowy Owl, chicken and zebrafinch with the exception of the Ural Owl microchromosomes 31, 35, 39 and 41 for which we could not unambiguously identify a corresponding chicken or zebrafinch chromosome but homology to Snowy Owl scaffolds and Tawny Owl chromosomes.

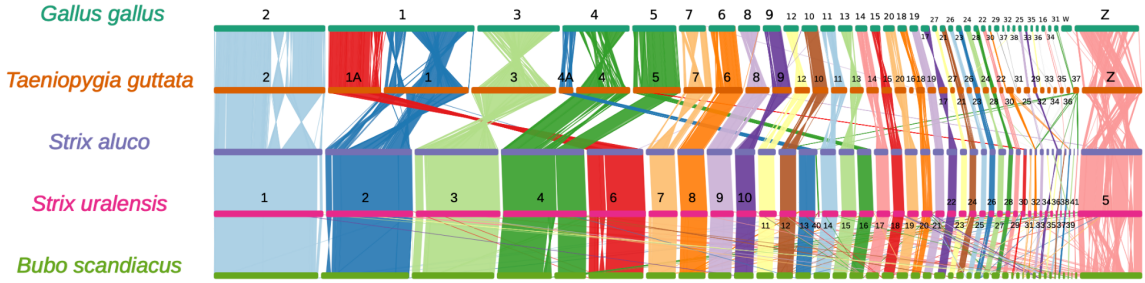


Figure 7: Chromosome scale synteny analysis. Synteny of chromosomes of *Taeniopygia guttata*, *Gallus gallus*, *Strix aluco* and *Bubo scandiacus* compared to the newly sequenced *Strix uralensis*. Syntenic regions amongst the species are indicated with a unique colour. Assignment of sex chromosomes was based on the *S. aluco* genome annotation. The chromosomes of all species have been reordered to highlight synteny relationships to *S. uralensis*. Plot made with NGenomeSynt.

Demographic history

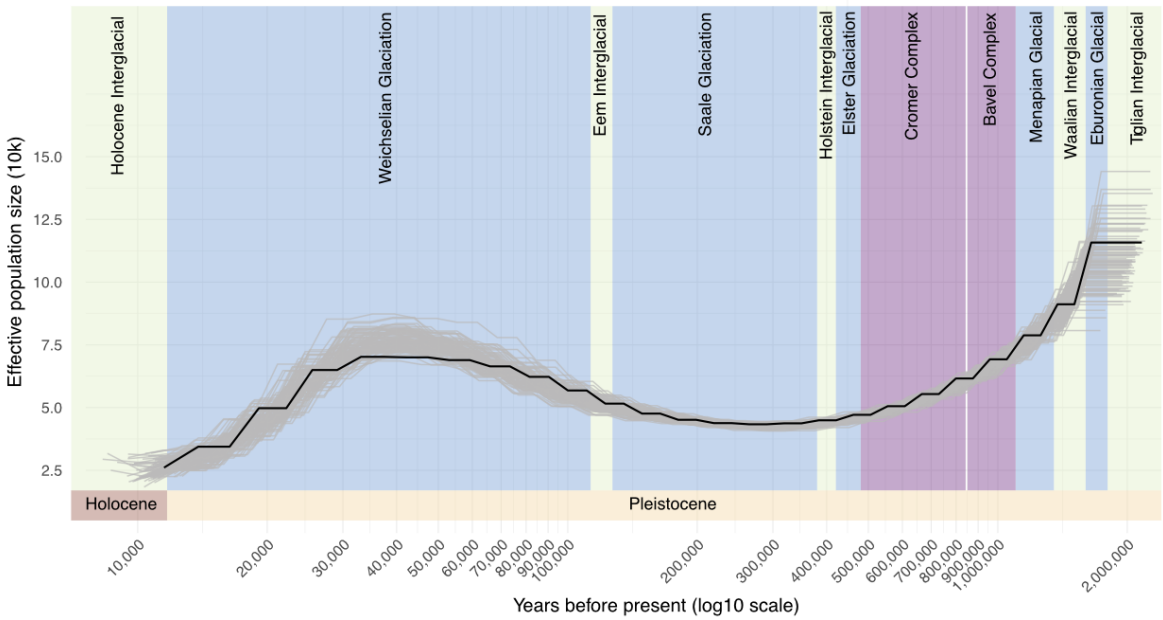


Figure 8. Inferred demographic history of *Strix uralensis.* The plot shows a Pairwise Sequentially Markovian Coalescent (PSMC) analysis based on the primary genome assembly. The x-axis shows years before present (ya) on a logarithmic scale and the y-axis shows the estimated effective population size. Bootstrap results are shown in light grey.

The demographic history of the Ural Owl derived from our genome assembly appears to have a complex relationship to glacial and interglacial periods. The effective population was

predicted to have decreased until around the Holstein interglacial period (3.74 × 10^5 – 4.24 × 10^5 ya) where its population size stabilized but remained low during the Saalian glacial period (4-1.3 × 10^5 ya) and began to increase as the Eemian interglacial period (1.3-1.15 × 10^5 ya) began to emerge. It continued to increase and reached a plateau during the last glacial period (Weichselian glaciation, 1.15-0.117 × 10^5 ya). Before the end of the last glacial period, at around 0.3 ×10^5 ya, the Ural Owl population began to decrease until it reached the current lowest effective population size (Figure 8).

Variation analysis over progressive cell passages

Karyotype confirms chromosome numbers and reveals no large variants caused by passaging

Chromosomal analyses detected 2n = 82 in both passages 5 and 10, corroborating the diploid chromosome number described for the Ural Owl previously (subspecies *S. uralensis uralensis* and *S. u. japonica*, [28]). No large-scale chromosomal rearrangements were observed between both passages (Figure 9).

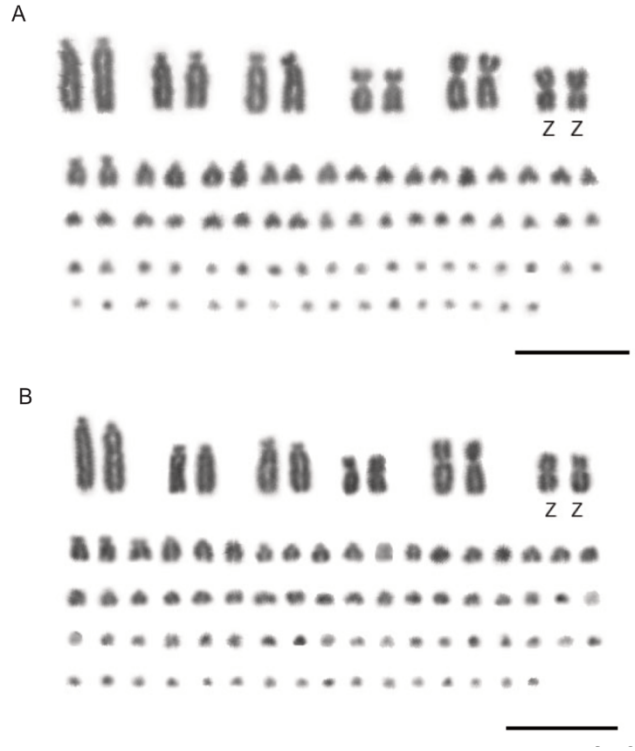


Figure 9: Karyotype analysis. Karyotype of *S. uralensis* male with 2n = 82 after passage 5

331 (a) and passage 10 (b). Bar = 10 µm.

332 Short-read variants

333 After quality filtering, we identified 885,159 variant sites (in the following referred to as SNPs)
334 (Supplementary Figure S9). Out of these, the vast majority (i.e., 670,463 SNPs) were fixed
335 variant sites across all samples and hence mostly represented heterozygous sites of the
336 individual which are represented with just one of the two alleles in the reference genome or
337 sites that had a wrong allele in the reference assembly.

338 The remaining 214,696 SNPs varied across samples, indicative of potential mutations, and
339 were analysed in the following.

340 A comparison of all SNPs across all samples revealed that variant amount and type differed.
341 The biggest differences resulted from SNPs called from the HiFi data as well as the passages
342 5.1 and 10.1 which appear to have more SNPs than the other passages. The majority of these
343 are heterozygous first-alternate sites (Supplementary Figure S9, Table 2) and we suspect for
344 many of those that they are false heterozygous calls rather than true mutations. To investigate
345 this pattern further, we inspected genotype quality and depth focusing on sites with a genotype
346 in a sample not found in any other sample (i.e., private sites) compared to the same metrics
347 at all other sites of that individual (i.e., common variants and conserved heterozygous sites).
348 This analysis revealed that all passages had similar median depth per variant site (DP ~26.78
349 ± 6.55) and genotype quality (GQ ~99), suggesting rather consistent data quality across
350 samples (Table 2). It further showed that median depth and quality of private SNPs
351 consistently had a significantly lower depth and quality than the average non-private site,
352 suggesting that these alleles are to some extent wrongly called (Supplementary Figure S10).
353 To account for patterns potentially driven by sequencing technology, we also assessed in each
354 replicate at how many positions it differed compared to passage 0. This revealed the same
355 pattern of an increase of SNPs in samples passage 5.1 and 10.1

356 **Table 2**: SNP statistics over progressive cell culture passages.

Sample	Median depth of variant sites	Median GQ of variant sites	SNPs other than shared heterozygous sites/private to individual	SNPs other than shared heterozygous sites as a percentage of the total genome size [%]	SNPs that are 0/1	SNPs that are not 0/1 nor 1/1	SNPs other than fixed heterozygous sites [%]	SNPs different from passage 0	SNPs different from passage 0 [%]
HiFi Reads	39	99	78,983/35,645	0.0057	77,682	1,265	0.14	-	-
Passage 0	27	99	57,929/8,212	0.0042	56,232	816	0.922	-	-
Passage 5.1	27	99	78,193/23,889	0.0057	76,709	716	0.809	58,620	6.62
Passage 5.2	29	99	55,225/3,280	0.0040	54,037	615	0.695	38,765	4.38
Passage 5.3	28	99	57,664/5,581	0.0042	56,267	781	0.882	41,301	4.67
Passage 10.1	20	99	103,086/63,058	0.0075	101,450	25	0.028	96,768	10.93
Passage 10.2	28	99	54,404/3,410	0.0039	53,373	420	0.475	38,921	4.40
Passage 10.3	28	99	55,976/3,758	0.0041	54,679	642	0.725	39,843	4.50
Passage 10.4	15	99	54,215/15,521	0.0039	52,153	750	0.847	50,033	5.65

357

Discussion

Reference genomes are accumulating across the tree of life and here birds have seen special attention fuelled by initiatives such as B10K (https://b10k.com/) [29,30]. Still, many of these genomes remain incomplete in terms of chromosomal-scale assembly type as well as gene annotation comprehensiveness. Genome assembly quality can impact phylogenomic inferences, analyses of gene prediction, gene family expansion and contraction, and most importantly structural evolution. In an effort to allow such analyses for the vastly understudied bird order Strigiformes, we here present a reference genome for the Ural Owl, that is among the best bird genome assemblies currently, reflected by assessments of sequence and gene completeness. We could place most of the genome into chromosomal-scale scaffolds, which are in line with the species karyotype, that we also confirm by cytogenetics. We further located the supposedly *Strix*-specific chromosomal fusion which distinguishes it from the genus *Bubo* [4]. Besides this, we identified several other chromosomal rearrangements between owl genomes (which are overall very syntenic) and those of chicken and zebrafinch. For four microchromosomes we could not identify a reliable homologous chromosome in chicken nor zebrafinch yet syntenic scaffolds in the Snowy Owl and microchromosomes in the Tawny Owl assemblies. In addition, the microchromosomes we assembled were also well supported by our HiC data. We thus hypothesize that the lack of synteny to chicken and zebrafinch might reflect more structural changes specific to owls. Matchingly, a first analysis of the Ural Owl genome content also indicates an important increase in repetitive sequences compared to most other non-owl bird genomes. Birds on average have rather compact genomes compared to other vertebrate lineages (average ~1.1 Gb [31]), mostly owing to a low content of repetitive elements (10-15 %). Until now, owls were seemingly no exception to this with an average genome size of 1.2 Gb and a repeat content of ~8,6 % [31]. Still, cytogenetic studies already suggested differently and hint at owls being rather an exception in the avian lineage with large scale variations in karyotypes, which our assembly also supports. For example, the barn owls have chromosomes of a more homogenous smaller size while the true owls have a more

385 classical avian karyotype with macro- and microchromosomes, suggesting chromosome
386 fusion and fission in owls [4]. Interestingly, the recently published genome (1.6 Gb) of the
387 Snowy Owl further extends this observation by demonstrating that it has one of the highest
388 reported repeat contents for birds (28.34 %), mainly composed of centromeric satellite DNA
389 [32]. Our assembly's total repeat content, at 21.07 % (1.5 % of which is unidentified), follows
390 this pattern. While both owl genomes' repeat expansions are largely driven by
391 retrotransposons, the Snowy Owl had a stronger increase of LTRs compared to LINEs than
392 the Ural Owl. Nevertheless, LTR retrotransposons are the largest repeat class also in the Ural
393 Owl, and, especially the youngest repeat expansion is also driven by LTRs suggesting this
394 pattern to be more broadly present in true owls. In the Snowy Owl, the repeats are suspected
395 to have driven the evolution of novel centromeres. Accordingly, cytogenetic analyses already
396 identified large centromeric satellite blocks shared among and unique to true owls [27]. Other
397 bird lineages with increased repeat content are woodpeckers and the Common Scimitarbill
398 (*Rhinopomastus cyanomelas*) [31]. The cause and consequences of the repeat extensions in
399 the genera *Strix* and *Bubo* remain unclear at this point, which is also true for the woodpecker
400 [33].

401 The number of microchromosomes identified in *S. uralensis* (n = 31) is consistent with the
402 average number of microchromosomes reported by Tegelström and Ryttman [34] from
403 karyotypes of over 230 bird species. However, there is no established rule for distinguishing
404 between macrochromosomes and microchromosomes (e.g. [35]). We here confirm that in
405 birds, microchromosomes have a higher gene density than macrochromosomes (e.g. [5,35–
406 37]). We could further show that the higher density of genes on microchromosomes is
407 associated with a reduction of gene length, which in turn is due to correspondingly shorter
408 introns. A similar pattern has been reported in chickens, where the size of the chromosome
409 correlates with the length of the genes it harbours [38]. Thus, in true owls, microchromosomes
410 hold up the crucial role they supposedly have played throughout vertebrate evolution [39].

411 Our assembly seems to be particularly well-suited for an analysis of gene content due to a

412 high completeness of gene annotation. However, due to vastly varying assembly qualities, the
413 correctness of our gene family expansion analysis should be taken with a grain of caution.
414 Still, this preliminary analysis suggests that in the future we will be able to connect changes in
415 gene content to adaptations of owls. This is supported by Ural Owl specific gene gains in the
416 GO-term derived function of e.g. “eye development”, “sensory organ development” and
417 “neurogenesis”, which could be linked to adaptations required for a nocturnal, predatory
418 lifestyle [40].

419 We also offer candidate genes for further investigation that characterize predatory lifestyle,
420 i.e., genes gained in the common ancestor of Strigiformes and Accipitriformes that acquired
421 this lifestyle. We especially observed gains of genes with a metabolic function which could
422 relate to the change in diet in the ancestor of these two bird orders. We also found gains of
423 keratin genes, in the common ancestor of the two predatory bird lineages and also in the
424 ancestor of owls. Feathers are epidermal appendages. Vertebrate skin appendages consist
425 of two fibrous proteins, alpha and beta keratins. Interestingly, β-keratins are exclusively found
426 in reptiles and birds. Both keratin gene families show expansions in different lineages. The
427 Barn Owl had the lowest number (6) of β-keratins in a study comparing 48 bird (draft)
428 genomes. The zebra finch in comparison had 149 genes [41]. This comparison further showed
429 that the proportion of claw β-keratins and keratinocyte β-keratins is higher in predatory birds.
430 We support the latter finding with the detection of three gains of feather and scale keratins in
431 the common ancestor of Accipitriformes and Strigiformes and two further gains in the ancestor
432 of all owls. These keratin genes deserve more attention as potential candidates that could
433 underlie morphological adaptations of feathers in predatory birds in general but also more
434 specifically in the mostly nocturnally hunting owls. Their silent flight is made possible by
435 physical characteristic fringes of the feathers on the leading edge of the wings [42]. The
436 genomic basis of this adaptation remains to be identified.

437 The Ural Owl is protected under the CITES convention Annex II and the Bern Convention on
438 the Conservation of European Wildlife and Natural Habitats. While globally not (yet) under

concern, the species went extinct in Germany and Austria due to habitat destruction but also direct persecution. Reintroduction programmes have been started, however due to low availability of breeding couples, individuals of various origins are used for these actions [18]. An analysis of marker genes neither supported morphological subspecies nor did it reveal a phylogeographic population structure for the Ural Owl; yet, it revealed genetic clusters that could be informative for breeding programs [12]. The here generated reference genome will facilitate future genomic studies in this direction of *S. uralensis*.

With an estimated genome wide heterozygosity of 0.2 % (2 het/kb), the here sequenced individual shows a higher level of heterozygosity than genomes of endangered bird species (red list status accessed March 2025, /www.iucnredlist.org/) such as the white-eared night heron (*Gorsachius magnificus*, Endangered, 0.49 het/kb) [43], Andean condor (*Vultur gryphus*; Vulnerable, 0.75 het/kb) and California condor (*Gymnogyps californianus*; Critically Endangered, 1.34 het/kb) [44], and Crested ibis (*Nipponia nippon*, Endangered, 0.043 het/kb) [45]. A similar heterozygosity level as the one we estimated for the Ural Owl was detected in e.g., Wild Turkey (0.24 %) and Mallard (0.26 %) [46]. It is somewhat lower than levels reported for other Strigiformes such as little owl (*Athene noctua*; 0.593) [47], Tawny Owl (*S. aluco*, 0.57 to 0.70) [48] and Barn Owl (*Tyto alba*, 0.59 to 0.71) [49] yet twice as high as in the Burrowing Owl (*Athene cunicularia*; 0.1%) [50].

Overall, species with a threat of going extinct show reduced levels of heterozygosity compared to non-threatened related taxa [51]. Related taxa of the same bird order with and without risk of extinction, differ quite drastically in genome-wide heterozygosity [45]. These differences likely result in lower evolutionary potential, reduced reproductive fitness and may contribute to species extinction [51]. It remains to be assessed at which level heterozygosity reduction causes an issue for a particular species. By generating genomic data for the Ural Owl, we contribute to the required knowledge for genetic monitoring of biodiversity.

Further, our reference genome already sheds light on the demographic history of the species, indicating both population contractions and expansions, apparently related to ecological

466 effects of the glacial-interglacial cycle. In particular, the pattern over the last 120,000 years
467 not only demonstrates the Ural Owl's tolerance to lower temperatures, but more importantly
468 reflects its flexible habitat choice of semi-open woodlands with a mixed composition of
469 broadleaf and coniferous species [10]. From the Eemian interglacial through the Weichselian
470 glaciation, climatic changes caused fluctuations in ice sheet extent and associated changes
471 in vegetation composition, including a gradual and/or repeated reduction in forest cover
472 leading to a treeless shrubby or grassy tundra from the mid-Weichselian (e.g., [52–54]). While
473 the open or semi-open structure of the woodland habitats favoured the Ural Owl's preference
474 for breeding and hunting grounds [11] until the mid-Weichselian, the expanding tundra
475 substantially reduced suitable habitats, leading to a marked decrease in effective population
476 size.

477 In the light of species preservation, protection and restoration, *ex situ* efforts are gaining more
478 attention. Cell culturing is a valuable and widely used technique, spanning applications from
479 basic science to biotechnology research [25]. However, there is no consensus regarding the
480 number of passages considered "safe" before cells experience metabolic changes, DNA
481 damage, and chromosomal instability. What is deemed "high passage" for one cell culture
482 may not lead to significant passage effects in another [26]. Thus, the effects of prolonged
483 culture are complex and depend on the individual cell culture, tissue, and species.

484 The first criterion for identifying healthy and stable cells is observing cell morphology.
485 Chromosome content serves as another critical benchmark, as normal cells maintain a stable
486 chromosome number. Some studies using non-model species, such as felines [55] and fishes
487 [56], showed no heteroploidy on karyotypes obtained by cell cultures. However, to the best of
488 our knowledge, this is the first study that addresses genomic and chromosome changes in
489 wild birds and compares the effect of different cell passages on genome integrity.

490 Cryopreservation of cells has increasingly been considered a strategy for conservation as new
491 technologies using genetic material from somatic cells (e.g., somatic cell nuclear transfer or
492 induced pluripotent stem cells) are evolving [23,57]. One of the prerequisites for nuclear donor

493 cells and *ex situ* conservation is the stability of chromosomes [55]. Studies that investigate cell
494 line passage and age effects are still scarce in non-model organisms and are crucial since
495 altered metabolism and genomic instability no longer represent reliable models of their original
496 source of material.

497 Herein, comparison between karyotypes of passage 5 and passage 10 showed no differences,
498 suggesting that no large structural rearrangement occurred during the progressive number of
499 passages and that it is safe to establish the diploid number of (this) bird species until at least
500 the 10th passage. Genetic instability is well-documented in cells that have undergone more
501 than 20 passages, particularly in transformed continuous cell lines (e.g., [58]) and tumour cell
502 lines [59,60]. However, for primary cell cultures, a straightforward method to determine the
503 safest passage number before cells develop mutations or genetic instability is lacking. We
504 opted to cultivate cells up to passage 10 based on two factors: first, the uncertainty
505 surrounding the exact passage limit at which primary cells may enter senescence (as non-
506 continuous cell line has a limited *in vitro* life time); and second, technical challenges observed
507 during later passages as cells began to exhibit signs of morphological decline, including the
508 presence of granules and debris, difficulty detaching, and a reduced growth rate, all of which
509 would complicate further subculturing beyond passage 10. This seems to suggest rather safe
510 cell culturing for this species until passage 10. To some extent, this is supported by our SNP
511 analysis of several passage replicates which indicated no general pattern for increased
512 genomic changes between passages 5 and 10. However, we detected outlier samples with
513 respect to SNP numbers among replicates of both passage numbers. At this point, we lack
514 any point of reference expectation as to how many (potential) mutations are to be expected in
515 a cell culture system as the one we applied. Compared to overall levels of variant sites, the
516 number of SNPs in the individual samples which could be mutations is lower and represented
517 between 0.0039-0.0075 % of the genome assembly length. The effects of these variants
518 remain to be determined as well as the reason for between replicate differences. We further
519 suspect that several mutations are variant-calling artefacts, supported by lower SNP calling

quality, which asks for an exploration of mutation identification and more importantly validation for the type of cell culture we have set up here.

Potential implications

We were able to assemble a reference genome for the Ural Owl of gold standard quality which is open to the community to be used for broader comparative genomic studies and phylogenomic analysis but also serves immediately to researchers interested in the Ural Owl for taxonomic and conservation aspects. With the data generated, we contribute to the endeavour of sequencing all life on Earth https://www.earthbiogenome.org/. Our analysis of genomic data derived from cell passages opens space for discussion of cell cultures as material for genomics especially for species with limited biological material available. The workflows applied by us could be used on similar data from other species.

Methods

Species origin and sampling strategy

Skin and muscle tissue samples from a ten-year-old male individual of *S. uralensis* (ring ID ZG-14.0-10-0234) were obtained from the Raptor Center & Wildlife Park Hellenthal (Wildfreigehege und Greifvogelstation Hellenthal, Hellenthal, Germany) during necropsy in 2020. The procedure was performed by Dominik Fischer, who is a veterinarian and approved to handle animals. No further approval was needed for this study. DNA barcoding was performed (collection ID ZFMK-TIS-50475) to ensure species identity using primers for COI from Astrin and Stüben [61] and sequences matched against BOLD (Barcode of Life Data System) [62]. The barcode sequence has been uploaded to BOLD as FOGS049-22.

Reference genome

Sequencing

DNA was extracted from the skin biopsy (collection ID ZFMK-TIS-50482, stored at LIB Biobank in liquid nitrogen vapor phase) using the Monarch HMW DNA Extraction Kit (NEB, Ipswich, USA). High-molecular weight status was validated by quality control with capillary electrophoresis (Agilent Femto Pulse) and a SPK 3.0 PacBio HiFi library was prepared according to the recommendations by the vendor. Next, HiFi SMRT sequencing was performed on two SMRT cells on a PacBio Sequel IIe (Pacific Biosciences, Menlo Park, USA) at the Max-Planck Genome-centre Cologne (MP-GC; Cologne, Germany). Also, a chromatin-capture library was prepared from cryopreserved cells generated for the analysis over progressive cell passages as described below with an Arima-Hi-C Kit according to the protocol for Mammalian Cell Lines followed by sequencing on an Illumina NextSeq 2000 in paired-end read mode.

Read processing

HiFi data

Contaminant sequences were filtered from the HiFi reads using Kraken2 v2.1.3 [63,64] with the Kraken database kraken2 PlusPFP downloaded in March 2023 and parameters "--confidence 0.51 --use-names". HiFi read quality was assessed using seqkit v2.8.2 [65,66] and a k-mer-based approach. K-mers were calculated with Meryl v1.4.1 [67] using the parameters "count k=21", and the counts were converted into a histogram with the *meryl histogram* command.

To verify the ploidy of the individual, Smudgeplot v0.2.5 [68] was used. First, k-mers within a specific range (lower-upper), determined with the *smudgeplot.py* cutoff function, were extracted using the meryl *print less-than* command. These filtered k-mers were then processed with *smudgeplot.py hetkmers* to calculate the coverage of unique heterozygous k-mer pairs. The resulting coverage was plotted using smudgeplot_plot.R.

GenomeScope2 v2.0.1 [68] with default parameters was used to estimate genome size, heterozygosity, and the homozygous and heterozygous coverage peaks.

ROHan v1.0.1 [69] was used to identify large (>1 Mb) runs of homozygosity.

Hi-C data

Adapter removal and quality filtering of the raw Hi-C reads were performed using Fastp v0.23 [70] with parameters "--length_required 95, --qualified_quality_phred 20 --adapter_fasta", with a curated adapter list of the most common adapters used as input.

Error correction was done using *Tadpole* from BBMap v39.01 (https://sourceforge.net/projects/bbmap/), with parameters "k=50, reassemble=t, mode=correct, minprob=0.6, prefilter=1, prehashes=2, and prealloc=t". To remove contamination from the short, Hi-C reads, Kraken2 v2.1.3 was used similarly to the HiFi reads but in paired-read mode, with parameter "--paired".

Initial Genome Assembly

The HiFi reads were used with Hifiasm v0.19.5 [71] to generate a phased genome assembly. In order to obtain an optimal, phased genome we tested several Hifiasm parameters before choosing the ones that provided us with the assembly of the highest contiguity and completeness with both phased haplotypes having a similar length. We tested all possible combination of different purging level (0, 2, 3), increasing run-time, and number of iterations ("--n-weight 5 --n-perturb 50000 --f-perturb 0.5 -D 10 -N 150 -s 0.2") and explicitly providing the homozygous peak to Hifiasm which was estimated by GenomeScope ("--hom-cov 40") (for more details see Supplementary Table S1).

Genome Scaffolding

The selected haplomes from Hifiasm were split at positions containing Ns using *split_fa* from the Purge_Dups package v1.2.6 [72]. The resulting sequences were mapped to themselves using Minimap2 v2.26 [73] with parameters "-x asm5 -DP" and to the HiFi reads using Minimap2 with parameters "-x map-hifi". These mappings were used to remove assembly duplicates with Purge_Dups.

Mitochondrial Genome Detection

To identify and extract the mitochondrial genome, we utilized MitoHiFi v3.2.1 [74,75] referencing the sequence NC_038218.1 from *S. uralensis* (isolate C5 mitochondrial genome, complete, https://www.ncbi.nlm.nih.gov/nuccore/NC_038218.1). The most likely scaffold was kept and identified as the mitochondrial chromosome (MT) and all other candidate scaffolds were removed from the assembly.

Assembly Manual Curation

Hi-C reads were aligned to the final assemblies and a Hi-C contact map was created using PretextMap v0.0.2 (https://github.com/sanger-tol/PretextMap). A HiFi coverage track was generated from the aligned HiFi reads using bedtools *genomecov* and integrated into the Pretext map with *PretextGraph.*

Manual curation was performed within PretextView v0.0.2 (https://github.com/sanger-tol/PretextView), where scaffolds were reordered and oriented based on Hi-C interaction frequencies. Following curation, the final assembly scaffolds were processed with AGP tools from the Vertebrate Genomes Project (VGP) using the rapid manual curation protocol (https://gitlab.com/wtsi-grit/rapid-curation/-/tree/main) established by the Darwin Tree of Life consortium (https://www.darwintreeoflife.org/) to create the curated assembly. Scaffold names were further sorted and renamed by size using a combination of seqkit v2.8.2 and SAMtools v1.19.2 [76]. The final Hi-C contact map was visualized with HiCExplorer [77].

Genome Quality Control

The completeness of the final curated assembly was assessed using BUSCO v5.8 [78,79]] and compleasm v0.2.6 [80] with the aves_odb10 lineage. Assembly contiguity and general assembly metrics were calculated using Quast v5.2.0 [81].

For k-mer-based analysis, k-mer counts were generated for each assembly using Meryl. These counts were analysed with Merqury v1.3 [82] to estimate assembly completeness and accuracy. The analysis yields Merqury's consensus QV, which is estimated by comparing the read and assembly k-mer counts and then transformed to a log-scaled probability of base-call errors. A higher QV indicates a more accurate assembly. We also obtained a Merqury completeness percent, which reflects the proportions of high-quality HiFi read k-mers present in the assembly.

HiFi reads were mapped to each assembly using Minimap2 with parameters “-ax map-hifi”. Alignment quality and coverage distribution were assessed using Qualimap v2.3 [83].

627 Potential contamination and quality was also assessed using the blobtoolkit pipeline v3.5.4
628 [84] and visualized using the interactive Blobtoolkit viewer in the Galaxy EU server [85] .

629 Genome Annotation

630 Repeat annotation

631 Repetitive elements in the primary assembly were identified and annotated using EarlGrey
632 v5.1.1 [86], which was run with RepeatMasker v4.1.5 [87] and RepeatModeler v2.0.6 [88]. In
633 addition to the RepeatModeler library, we used a previously-created, custom avian TE library
634 to mask repetitive elements [89]. The softmasked genome was used for protein-coding gene
635 prediction.

636 Protein-Coding Gene Annotation

637 To perform protein annotation, we created two reference protein sets. Set one contained only
638 the merged proteomes of the following publicly available genomes, downloaded using the
639 NCBI dataset cli v16.3.0 [90]*: S. nigrolineata* (GCA_013396715.1), *Gallus gallus*
640 (GCF_016699485.2), *Glaucidium brasilianum* (GCA_013399595.1), *Falco peregrinus*
641 (GCF_023634155.1), *Athene cunicularia* (GCF_003259725.1), *Aquila chrysaetos*
642 (GCF_900496995.4) [91], *Taeniopygia guttata* (GCF_003957565.2), and *Anas platyrhynchos*
643 (GCF_015476345.1).

644 Protein set two was created by merging all proteomes in set one with proteins from the
645 following public and curated databases: i) proteins from the BUSCO v5.4 aves_odb10 dataset,
646 ii) aves proteins from OrthoDB v11 [92] were obtained using Tomas Bruna's orthodb-clades
647 pipeline (https://github.com/tomasbruna/orthodb-clades) and iii) proteomes were also
648 extracted from the UniProt database for the following species: *Calypte anna* (UP000054308),
649 *Steatornis caripensis* (UP000516988), *Cnemophilus loriae* (UP000517678), *Dasyornis*
650 *broadbenti* (UP000521322), *Corythaixoides concolor* (UP000526942), *Irena cyanogastra*
651 (UP000530962), *Bucco capensis* (UP000534107), *Cephalopterus ornatus* (UP000543364),

652 *Molothrus ater* (UP000553862), *Ptilonorhynchus violaceus* (UP000584880), *Promerops cafer*
653 (UP000587587), *Vidua chalybeata* (UP000634236), and *Urocolius indicus* (UP000654395).

654 Protein-coding genes in the *S. uralensis* genome were annotated using a combination of *ab*
655 *initio*, protein similarity, and transcriptome-based protein prediction models. BRAKER3 v3.0.3
656 [93,94] was run in EP mode using the protein set two described above. GALBA v1.0.11.2 [95]
657 was run using protein set one to annotate genes. The outputs from GALBA and BRAKER3
658 v3.0.2 were combined using *TSEBRA* from the BRAKER3 package. To ensure high-quality
659 annotations, only the longest gene orthologs for each locus were retained using the
660 *agat_sp_keep_longest_isoform.pl* script from the AGAT package v1.4.1 [96].

661 Demographic History of *S. uralensis*

662 The demographic history of the Ural Owl was reconstructed using PSMC v0.6.5 as
663 implemented by [97]. Variants were called per chromosome using a combination of BCFtools
664 v1.21 (http://github.com/samtools/bcftools) *mpileup* with parameters "-Q 30 -q 30" and bcftools
665 call using the "-c" option. The resulting VCF file was converted to a consensus fastq format
666 using the *vcfutils.pl* vcf2fq script with parameters "-d 10, -D 60, and -Q 30". The PSMC model
667 was run with the following parameters: -N25 -t15 -r5 -p "2+2+25*2+4+6" and 100 bootstraps,
668 a generation time of 3 years [98,99] and an assumed mutation rate of 4.6x10-9 [100,101].

669 Genome Synteny

670 In order to identify the Z sex chromosome within our genome assembly and to assess the
671 synteny of different bird genomes we used the *GetTwoGenomeSyn.pl* built-in script of
672 NGenomeSyn [102] with options: "-MappingBin minimap2 -MinLenA 100000 -MinLenB
673 100000 -NumThreads 5 -MappingPara '-Lx asm10 --eqx -I 200G --MD -N 1'" to estimate
674 chromosome-scale alignments between *G. gallus* (GCA_024206055.2), *T. guttata*
675 (GCF_048771995.1), *S. aluco* (GCA_031877795.1), *S. uralensis,* and *B. scandiacus*
676 (GCA_965212795.1). To improve visualisation and reduce noisiness in the synteny plot, we

677 used the get.synteny.blocks.multi function of the asyntr R package
678 (https://github.com/simonhmartin/asynt) to aggregate single alignments into synteny blocks of
679 incrementally increasing minimum block size of 100 to 4000 and filtered for chromosomes
680 smaller than 1000 bp and for alignment lengths smaller than 200 bp. For visualization
681 purposes, we manually removed alignments between *G. gallus* and *T. guttata* that were not
682 supported by Luo et al. [103].

683 Functional Gene Annotation

684 Predicted genes were functionally annotated by performing sequence similarity searches
685 against the Swiss-Prot database using *BLASTP* from BLAST v2.13.0+ with default
686 parameters. As with our own genome's annotation, we used
687 *agat_sp_keep_longest_isoform.pl* to only keep the longest isoform of each gene locus from
688 the *Aquila chrysaetos, Gallus gallus, S. nigrolineata*, *Athene cunicularia, Glaucidium*
689 *brasilianum, Falco peregrinus,* proteomes of protein set one together with *Tyto alba*
690 (GCA_018691265.1) and *Haliaeetus leucocephalus* (GCA_000737465.1) and used
691 OrthoFinder v2.5.5 [104] to estimate orthologous gene families among those species. This
692 analysis identified orthogroups and genes that have undergone expansion or contraction in
693 the Ural Owl, as well as orthogroups unique to this species.

694 Gene Ontology Term Analysis

695 Gene Ontology Term analysis was performed by mapping all Ural Owl genes to the Vertebrate
696 Eggnog database using eggnog mapper v2.1.12 [105,106]. Missing GO Terms were filled in
697 with the GO Terms of the previously found Swiss-Prot gene symbols associated with each
698 gene. Next, the genes belonging to gene families unique to the Ural Owl (found with
699 Orthofinder) were analysed by Revigo v1.8.1 [107] with default settings and choosing the
700 *Large* subset option. The resulting GO Terms were analysed with Revigo again with default
701 settings and this time with the *Small* subset option. The full *Biological Process* Revigo table

702 was plotted in R v4.4.2 using an edited version of the Revigo treemap plotting script.

703 Variation analysis over progressive cell passages

704 Cell Culture

705 Primary cells were grown from a skin biopsy of the same individual as used for genome
706 sequencing (collection ID ZFMK-TIS-51054) previously stored at LIB Biobank in liquid
707 nitrogen, following standard protocols. Skin tissues were rapidly thawed, minced into small
708 fragments, and transferred to cell culture flasks. Flasks were incubated at 37°C with 5 % CO_2
709 in Fibroblast Growth Basal Medium (FBM; Lonza, Cologne, Germany) supplemented with 20%
710 Fetal Bovine Serum (FBS; Biowest, Nuaillé, France) including antibiotics (100 U/mL penicillin
711 and 100 g/mL streptomycin; Sigma-Aldrich, St. Louis, United States). Cells were visually
712 inspected in inverted microscope Nikon Eclipse TS2 for contamination and cell media was
713 changed every 2-3 days. After reaching ~80 % confluence (determined visually), cells were
714 propagated using 0.125 % trypsin solution (Biowest), at subculture ratio 50:50. Cells were
715 harvested for DNA extraction and chromosome analysis at passages 5 (three different
716 replicates) and passage 10 (four different replicates).

717 Chromosome sampling for large variant analysis

718 In order to investigate the stability of the karyotype composition through different passages,
719 chromosome preparations were obtained from cells for passages 5 and 10, according to [108],
720 with modifications. Chromosomes were harvested after treatment with colchicine 0.01% for
721 one hour, followed by hypotonic treatment with 0.075 M KCl, and cell fixation in methanol /
722 acetic acid (3:1). Slides were stained with Giemsa 5 %. At least 20 metaphases for each
723 passage were analysed to define the diploid number (2n) in a Zeiss microscope Axio Imager
724 Z2m.

725 ## DNA extraction of primary tissue and cell culture passages

726 Passage samples were extracted using the DNeasy Blood & Tissue Kit (Qiagen, Hilden,
727 Germany) following the manufacturer's protocol for cultured cells, while muscle tissue of the
728 same individual (collection ID ZFMK-TIS-50476) previously stored at LIB Biobank in 96 %
729 ethanol (passage 0) was extracted using the standard protocol of the same kit.

730 ## Sequencing of primary tissue and passages

731 After DNA extraction, samples were sent for purification (Vahtstm DNA Clean Beads; Vazyme
732 Biotech, Nanjing, China), PCR-free library preparation (NEBNext Ultra II FS DNA PCR-free
733 Library Prep Kit for Illumina; NEB) and subsequent paired-end sequencing on a NovaSeq
734 6000 (Illumina, San Diego, USA) using the NovaSeq 6000 S4 Reagent Kit (Illumina) to
735 Biomarker Technologies (bmkgene; Beijing, China).

736 ## Read Mapping

737 The Illumina reads of all passages (passage 0, i.e. the primary tissue, three passage 5
738 replicate samples and four passage 10 replicate samples) were processed with fastp v0.20.0
739 [70] with parameters "--length_required 95, --qualified_quality_phred 20 --adapter_fasta", with
740 a curated adapter list of the most common adapters used as input and decontaminated with
741 Kraken2 v2.1.3 with the Kraken database kraken2 PlusPFP database downloaded in March
742 2023 in paired-read mode, with parameter "--paired --confidence 0.51 --use-names". These
743 reads were then mapped to the reference genome using BWA-MEM2 v2.2.1 [109,110] with
744 the *mem* command and options "-M -R", where a read group (RG) specific to each sample
745 was used for "-R". The resulting output was sorted using SAMtools v1.19.2, and additional
746 processing steps (SAMtools' *fixmate*, *sort*, and *markdup*) were performed to generate the final
747 mapping files for each sample.

SNP Calling

SNP calling for each cell passage sample and the HiFi reads (“reference”) was performed individually using GATK HaplotypeCaller v4.2.6.1 [111,112] with the options “-ERC GVCF --min-base-quality-score 30 --pcr-indel-model NONE”. Joint SNP calling was performed by first combining samples using GATK *GenomicsDBImport* with the option “--batch-size 3”. The combined database was then used for joint SNP calling with GATK’s *GenotypeGVCFs*.

Variant quality recalibration was conducted in three rounds. GATK’s *BaseRecalibrator* was run with the option “--maximum-cycle-value 50000”, followed by GATK’s *ApplyBQSR* for each sample before re-calling variants individually and collectively. From the final set of called genotypes, SNPs were extracted using GATK’s *SelectVariants* with the option “-select-type SNP” and filtered with GATK VariantFiltration using the filters: “QD < 2.0, FS > 60.0, MQ < 40.0, SOR > 3.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, QUAL < 30.0”.

Variants were filtered for depth, minor allele frequency (MAF) and the fraction of missing genotypes using BCFtools filter v1.21 (https://github.com/samtools/bcftools) with the options “-e "INFO/DP<$MIN_DEPTH || INFO/DP>$MAX_DEPTH" ” and “-i "MAF>$MAF && F_MISSING<=$MISS".

Short-read variant analysis in passages

To assess the quality of the DNA contained in cell cultures and to understand its potential for being used as an amplified genomic resource, these filtered SNPs were then analysed in R v4.4.2. Passages 5 and 10 were compared to passage 0 and the HiFi reads at sites where passages 5 and 10 differ from either of the reference passages. A Wilcoxon test from rstatix v0.7.2 was performed to test whether the depth and GQ of these SNPs of each sample were significantly different from the average DP or GQ. The variant calls were re-coded so that: "0|0" = 0, "0|1" = 1, "1|0" = 1, "1|1" = 2, "0|2" = 3, "2|0" = 3, "1|2" = 4, "2|1" = 4, "2|2" = 5, "0|3" = 6, "3|0" = 6 and plotted with ggplot2.

Availability of source code and requirements

Project: Strix uralensis assembly, annotation and comparative analysis

Location: Zenodo DOI: 10.5281/zenodo.15100180

Operating system(s): e.g. Platform independent

Licence: CC0

Data Availability

The sequencing reads, assembly and BioSample data supporting the results of this article are available in the INSDC under the BioProject number PRJNA1212906. Further datasets and code supporting the results of this article are available from Zenodo under DOI: 10.5281/zenodo.14676512. Code is available from Zenodo under DOI: 10.5281/zenodo.15100180.

List of abbreviations

2n: diploid chromosome number; b: bases; bp: base pair; BLAST: basic local alignment search tool; BOLD: Barcode of Life Data System; BUSCO: Benchmarking Universal Single-Copy Orthologs; C: Celsius; CBD: convention on biological diversity; CITES: convention on international trade in endangered species of wild fauna and flora; DP: depth; EU: European Union; FBS: Fetal Bovine Serum; GATK: Genome Analysis Toolkit; Gb: gigabases; GO: gene ontology; GQ: genotype quality; Hi-C: high-throughput chromosome conformation capture; HiFi: high-fidelity; HMW: high molecular weight; INSDC: International Nucleotide Sequence Database Collaboration; kb: kilobases; LINE: long interspersed nuclear element; LTR: long terminal repeat transposable element; M: molar; MAF: major allele frequency; Mb: megabases; MT: mitochondrial chromosome; PSMC: Pairwise Sequentially Markovian Coalescent; QV: quality value; RG: read group; ROH: runs of homozygosity; SNP: single nucleotide polymorphism; TE; transposable element; VCF: variant call format; VGP: Vertebrate Genomes Project; ya: years ago.

Declarations

The primary tissue used for this work was derived from a naturally deceased bird and provided by a veterinarian. We did not perform animal experimentation.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Leibniz Gemeinschaft Leibniz Association Network grant CollOmic K419/2021 to AB and LIB innovation fund to AB and JJA.

Authors' contributions

IC: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing (original draft; review & editing); AM: conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing (original draft; review & editing); CBDN: conceptualization, formal analysis, investigation, methodology, writing (original draft; review & editing); DF: resources, writing (review & editing); NS: investigation, writing (review & editing); LvdM: investigation, writing (review & editing); BH: investigation, writing (review & editing); JJA: conceptualization, funding acquisition, supervision, writing (original draft; review & editing); TT: validation, supervision, writing (original draft; review & editing); AB: conceptualization, data curation, validation, visualization, funding acquisition, supervision, writing (original draft; review & editing).

Acknowledgements

We thank Juliane Vehof and Benjamin Wipfler for enabling us to use their microscope.

References

1. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, et al.. Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences*; 2018; doi: 10.1073/pnas.1720115115.

2. Blaxter M, Archibald JM, Childers AK, Coddington JA, Crandall KA, Di Palma F, et al.. Why sequence all eukaryotes? *Proceedings of the National Academy of Sciences USA*. 2022; doi: 10.1073/pnas.2115636118.

3. Ellegren H. Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology & Evolution*. 2010; doi: 10.1016/j.tree.2009.12.004.

4. Rebholz WER, Boer LEMD, Sasaki M, Belterman RHR, Nishida-Umehara C. The chromosomal phylogeny of owls (Strigiformes) and new karyotypes of seven species. *cytologia*. 1993; doi: 10.1508/cytologia.58.403.

5. Burt DW. Origin and evolution of avian microchromosomes. *Cytogenetic Genome Research*. 2002; doi: 10.1159/000063018.

6. Pichugin AM, Galkina SA, Potekhin AA, Punina EO, Rautian MS, Rodionov AV. Estimation of the minimal size of chicken *Gallus gallus domesticus* microchromosomes via pulsed-field electrophoresis. *Russian Journal of Genetics*. 2001; doi: 10.1023/A:1016622816552.

7. Degrandi TM, Barcellos SA, Costa AL, Garnero ADV, Hass I, Gunski RJ. Introducing the bird chromosome database: An overview of cytogenetic studies in birds. *Cytogenetic Genome Research*. 2020; doi: 10.1159/000507768.

8. Smith J, Bruley CK, Paton IR, Dunn I, Jones CT, Windsor D, et al.. Differences in gene density on chicken macrochromosomes and microchromosomes. *Animal Genetics*. 2000; doi: 10.1046/j.1365-2052.2000.00565.x.

9. Axelsson E, Webster MT, Smith NGC, Burt DW, Ellegren H. Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Research*. 2005; doi: 10.1101/gr.3021305.

10. Roselaar C. *Strix uralensis* Ural Owl. In: Cramp S, editor. *Handbook of the Birds of Europe, the Middle East, and North Africa The Birds of the Western Paleartic*. Oxford University Press; p. 550–60.

11. Able KP. Handbook of the Birds of the World, Volume 5, Barn-owls to Hummingbirds. *The Auk*. 2000; doi: 10.1093/auk/117.2.532.

12. Hausknecht R, Jacobs S, Müller J, Zink R, Frey H, Solheim R, et al.. Phylogeographic analysis and genetic cluster recognition for the conservation of Ural Owls (*Strix uralensis*) in Europe. *Journal of Ornithology*. 2014; doi: 10.1007/s10336-013-0994-8.

13. Cramp S. Handbook of the birds of Europe, the Middle East, and north Africa: the birds of the western Paleartic. Oxford London New York: Oxford university press;

14. König C, Weick F. Owls of the World (2nd ed.). Helm identification Guides. A & C Black Publishers Ltd.;

15. Mikkola H, Willis I. Owls of Europe. Calton, Waterhouses, Staffordshire, England: T & A D Poyser;

16. IUCN. *Strix uralensis*: Westrip, J.R.S. & BirdLife International: The IUCN Red List of Threatened Species 2022: e.T22689108A209840432.

17. Kopij G. Population and range expansion of forest boreal owls (Glaucidium passerinum, Aegolius funereus, Strix uralensis, Strix nebulosa) in East-Central Europe. *Vogelwelt*. 132:207–142011;

18. Scherzinger W. Die Wiederbegründung des Habichtskauz-Vorkommens *Strix uralensis* im Böhmerwald. *Ornithologischer Anzeiger*. 45:97–1562006;

19. Soorae PS. Global re-introduction perspectives, 2011: more case studies from around the globe. Abu Dhabi, UAE: IUCN/SSC Re-introduction Specialist Group & Environment Agency - Abu Dhabi;

20. Scope A, Schwendenwein I, Stanclova G, Vobornik A, Zink R. Exploratory plasma biochemistry reference intervals for Ural Owls (*Strix uralensis*, Pallas 1771) from the Austrian reintroduction project. *Journal of Zoo and Wildlife Medicine*. 2016; doi: 10.1638/2015-0200.1.

21. Huntley B, Green R, Collingham YC, Willis SG. A climatic atlas of European breeding birds. Barcelona: Lynx ed;

22. Lehikoinen A, Ranta E, Pietiäinen H, Byholm P, Saurola P, Valkama J, et al.. The impact of climate and cyclic food abundance on the timing of breeding and brood size in four boreal owl species. *Oecologia*. 2011; doi: 10.1007/s00442-010-1730-1.

23. Mooney A, Ryder OA, Houck ML, Staerk J, Conde DA, Buckley YM. Maximizing the potential for living cell banks to contribute to global conservation priorities. *Zoo Biology*. 2023; doi: 10.1002/zoo.21787.

24. Ryder OA, Onuma M. Viable cell culture banking for biodiversity characterization and conservation. *Annual Review Animal Bioscience*. 2018; doi: 10.1146/annurev-animal-030117-014556.

25. Freshney RI. Culture of animal cells: a manual of basic technique and specialized applications. 6th edition. John Wiley & Sons;

26. Hughes P, Marshall D, Reid Y, Parkes H, Gelber C. The costs of using unauthenticated, over-passaged cell lines: How much more data do we need? *BioTechniques*. Taylor & Francis; 2007; doi: 10.2144/000112598.

27. Yamada K, Nishida-Umehara C, Matsuda Y. A new family of satellite DNA sequences as a major component of centromeric heterochromatin in owls (Strigiformes). *Chromosoma*. 2004; doi: 10.1007/s00412-003-0267-z.

28. Takagi N, Sasaki M. A phylogenetic study of bird karyotypes. *Chromosoma*. 1974; doi: 10.1007/BF00332341.

29. Zhang G. Bird sequencing project takes off. *Nature*. 2015; doi: 10.1038/522034d.

30. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al.. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*. 2014; doi:

10.1126/science.1251385.

31. Feng S, Stiller J, Deng Y, Armstrong J, Fang Q, Reeve AH, et al.. Dense sampling of bird diversity increases power of comparative genomics. *Nature*. 2020; doi: 10.1038/s41586-020-2873-9.

32. Baalsrud HT, Garmann-Aarhus B, Enevoldsen ELG, Krabberød AK, Fischer D, Tooming-Klunderud A, et al.. Evolutionary new centromeres in the snowy owl genome putatively seeded from a transposable element. *BioRxiv*.2024; doi: 10.1101/2024.07.05.602039.

33. Forest T, Achaz G, Marbouty M, Bignaud A, Thierry A, Koszul R, et al.. Chromosome-level genome assembly of the European green woodpecker *Picus viridis*. Campbell P, editor. *G3: Genes, Genomes, Genetics*. 2024; doi: 10.1093/g3journal/jkae042.

34. Tegelström H, Ryttman H. Chromosomes in birds (Aves): evolutionary implications of macro-and microchromosome numbers and lengths. *Hereditas*. 1981; doi: 10.1111/j.1601-5223.1981.tb01757.x.

35. Fillon V. The chicken as a model to study microchromosomes in birds: a review. *Genetics Selection Evolutiom*. 1998; doi: 10.1186/1297-9686-30-3-209.

36. McQueen HA, Fantes J, Cross SH, Clark VH, Archibald AL, Bird AP. CpG islands of chicken are concentrated on microchromosomes. *Nature Genetics*. 1996; doi: 10.1038/ng0396-321.

37. Schmid M, Nanda I, Guttenbach M, Steinlein C, Hoehn M, Schartl M, et al.. First report on chicken genes and chromosomes. *Cytogenetic Genome Research*. 2000; doi: 410.1159/000056772.

38. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*. 2004; doi: 10.1038/nature03154.

39. Waters PD, Patel HR, Ruiz-Herrera A, Álvarez-González L, Lister NC, Simakov O, et al.. Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proceedings of the National Academy of Sciences USA*. 2021; doi: 10.1073/pnas.2112494118.

40. Borges R, Khan I, Johnson WE, Gilbert MTP, Zhang G, Jarvis ED, et al.. Gene loss, adaptive evolution and the co-evolution of plumage coloration genes with opsins in birds. *BMC Genomics*. 2015; doi: 10.1186/s12864-015-1924-3.

41. Greenwold MJ, Bao W, Jarvis ED, Hu H, Li C, Gilbert MTP, et al.. Dynamic evolution of the alpha (α) and beta (β) keratins has accompanied integument diversification and the adaptation of birds into novel lifestyles. *BMC Evolutionary Biology*. 2014; doi: 10.1186/s12862-014-0249-1.

42. Wagner H, Weger M, Klaas M, Schröder W. Features of owl wings that promote silent flight. *Interface Focus*. 2017; doi: 10.1098/rsfs.2016.0078.

43. Luo H, Lin Q, Fang W, Chen X, Zhou X. Genomic insights into the endangered white-eared night heron (*Gorsachius magnificus*). *BMC Genomic Data*. 2024; doi: 10.1186/s12863-024-01194-1.

44. Robinson JA, Bowie RCK, Dudchenko O, Aiden EL, Hendrickson SL, Steiner CC, et al..

943 Genome-wide diversity in the California condor tracks its prehistoric abundance and decline.
944 *Current Biology*. 2021; doi: 10.1016/j.cub.2021.04.035.

945 45. Li S, Li B, Cheng C, Xiong Z, Liu Q, Lai J, et al.. Genomic signatures of near-extinction
946 and rebirth of the crested ibis and other endangered bird species. *Genome Biology*. 2014;
947 doi: 10.1186/s13059-014-0557-1.

948 46. Li B-P, Kang N, Xu Z-X, Luo H-R, Fan S-Y, Ao X-H, et al.. Transposable elements shape
949 the landscape of heterozygous structural variation in a bird genome. *Zoological Research*.
950 2025; doi: 10.24272/j.issn.2095-8137.2024.237.

951 47. Pellegrino I, Negri A, Boano G, Cucco M, Kristensen TN, Pertoldi C, et al.. Evidence for
952 strong genetic structure in European populations of the little owl *Athene noctua*. *Journal of*
953 *Avian Biology*. 2015; doi: 10.1111/jav.00679.

954 48. Brito PH. Contrasting patterns of mitochondrial and microsatellite genetic structure
955 among Western European populations of tawny owls (*Strix aluco*). *Molecular Ecology*. 2007;
956 doi: 10.1111/j.1365-294X.2007.03401.x.

957 49. Antoniazza S, Burri R, Fumagalli L, Goudet J, Roulin A. Local adaptation maintains clinal
958 variation in melanin-based coloration of European barn owls (*Tyto alba*). *Evolution*. 2010;
959 doi: 10.1111/j.1558-5646.2010.00969.x.

960 50. Mueller JC, Kuhl H, Boerno S, Tella JL, Carrete M, Kempenaers B. Evolution of genomic
961 variation in the burrowing owl in response to recent colonization of urban areas. *Proceedings*
962 *of the Royal Society B*. 2018; doi: 10.1098/rspb.2018.0206.

963 51. Spielman D, Brook BW, Frankham R. Most species are not driven to extinction before
964 genetic factors impact them. *Proceedings of the National Academy of Sciences USA*. 2004;
965 doi: 10.1073/pnas.0403809101.

966 52. Novenko EYu, Seifert-Eulen M, Boettger T, Junge FW. Eemian and Early Weichselian
967 vegetation and climate history in Central Europe: A case study from the Klinge section
968 (Lusatia, eastern Germany). *Review of Palaeobotany and Palynology*. 2008; doi:
969 10.1016/j.revpalbo.2008.02.005.

970 53. Velichko AA, Novenko EY, Pisareva VV, Zelikson EM, Boettger T, Junge FW. Vegetation
971 and climate changes during the Eemian interglacial in Central and Eastern Europe:
972 comparative analysis of pollen data. *Boreas*. 2008; doi: 10.1111/j.1502-
973 3885.2005.tb01016.x.

974 54. Malkiewicz M. A Late Saalian Glaciation, Eemian Interglacial and Early Weichselian
975 pollen sequence at Szklarka, SW Poland – Reconstruction of vegetation and climate.
976 *Quaternary International*. 2018; doi: 10.1016/j.quaint.2016.09.026.

977 55. Song J, Hua S, Song K, Zhang Y. Culture, characteristics and chromosome complement
978 of Siberian tiger fibroblasts for nuclear transfer. *In Vitro Cellular & Developmental Biology -*
979 *Animal*. 2007; doi: 10.1007/s11626-007-9043-3.

980 56. Alvarez MC, Otis J, Amores A, Guise K. Short-term cell culture technique for obtaining
981 chromosomes in marine and freshwater fish. *Journal of Fish Biology*. John Wiley & Sons,
982 Ltd; 1991; doi: 10.1111/j.1095-8649.1991.tb04411.x.

983 57. Bolton RL, Mooney A, Pettit MT, Bolton AE, Morgan L, Drake GJ, et al.. Resurrecting
984 biodiversity: advanced assisted reproductive technologies and biobanking. *Reproduction and*

985 *Fertility*. Bristol, UK: Bioscientifica Ltd; 2022; doi: 10.1530/RAF-22-0005.

986 58. Odoemelam E, Raghavan N, Miller A, Bridger JM, Knight M. Revised karyotyping and
987 gene mapping of the Biomphalaria glabrata embryonic (Bge) cell line. *International Journal*
988 *for Parasitology*. 2009; doi: 10.1016/j.ijpara.2008.11.011.

989 59. He Z, Wilson A, Rich F, Kenwright D, Stevens A, Low YS, et al.. Chromosomal instability
990 and its effect on cell lines. *Cancer Reports*. John Wiley & Sons, Ltd; 2023; doi:
991 10.1002/cnr2.1822.

992 60. Wenger SL, Senft JR, Sargent LM, Bamezai R, Bairwa N, Grant SG. Comparison of
993 established cell lines at different passages by karyotype and comparative genomic
994 hybridization. *Bioscience Reports*. 2005; doi: 10.1007/s10540-005-2797-5.

995 61. Astrin JJ, Stüben PE. Phylogeny in cryptic weevils: molecules, morphology and new
996 genera of western Palaearctic Cryptorhynchinae (Coleoptera: Curculionidae). *Invertebrate*
997 *Systematics*. 2008; doi: 10.1071/IS07057.

998 62. Ratnasingham S, Hebert PDN. BOLD : The Barcode of Life Data System
999 (http://www.barcodinglife.org). *Molecular Ecology Notes*. 2007; doi: 10.1111/j.1471-
1000 8286.2007.01678.x.

1001 63. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome*
1002 *Biology*. 2019; doi: 10.1186/s13059-019-1891-0.

1003 64. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using
1004 exact alignments. *Genome Biology*. 2014; doi: 10.1186/gb-2014-15-3-r46.

1005 65. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q
1006 File Manipulation. Zou Q, editor. *PLoS ONE*. 2016; doi: 10.1371/journal.pone.0163962.

1007 66. Shen W, Sipos B, Zhao L. SeqKit2: A Swiss army knife for sequence and alignment
1008 processing. *iMeta*. 2024; doi: 10.1002/imt2.191.

1009 67. Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley A, et al.. Aggressive
1010 assembly of pyrosequencing reads with mates. *Bioinformatics*. 2008; doi:
1011 10.1093/bioinformatics/btn548.

1012 68. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for
1013 reference-free profiling of polyploid genomes. *Nature Communications*. 2020; doi:
1014 10.1038/s41467-020-14998-3.

1015 69. Renaud G, Hanghøj K, Korneliussen TS, Willerslev E, Orlando L. Joint estimates of
1016 heterozygosity and runs of homozygosity for modern and ancient samples. *Genetics*. 2019;
1017 doi: 10.1534/genetics.119.302057.

1018 70. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
1019 *Bioinformatics (Oxford, England)*. 2018; doi: 10.1093/bioinformatics/bty560.

1020 71. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo
1021 assembly using phased assembly graphs with hifiasm. *Nature Methods*. 2021; doi:
1022 10.1038/s41592-020-01056-5.

1023 72. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing
1024 haplotypic duplication in primary genome assemblies. Valencia A, editor. *Bioinformatics*.

2020; doi: 10.1093/bioinformatics/btaa025.

73. Li H. Minimap2: pairwise alignment for nucleotide sequences. Birol I, editor. *Bioinformatics*. 2018; doi: 10.1093/bioinformatics/bty191.

74. 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. *Molecular Ecology Resources*. 2020; doi: 10.1111/1755-0998.13160.

75. Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Darwin Tree of Life Consortium, Blaxter M, Mieszkowska N, et al.. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics*. 2023; doi: 10.1186/s12859-023-05385-y.

76. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al.. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; doi: 10.1093/bioinformatics/btp352.

77. Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, et al.. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Research*. 2020; doi: 10.1093/nar/gkaa220.

78. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; doi: 10.1093/bioinformatics/btv351.

79. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Kelley J, editor. *Molecular Biology and Evolution*. 2021; doi: 10.1093/molbev/msab199.

80. Huang N, Li H. Compleasm: a faster and more accurate reimplementation of BUSCO. Marschall T, editor. *Bioinformatics*. 2023; doi: 10.1093/bioinformatics/btad595.

81. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013; doi: 10.1093/bioinformatics/btt086.

82. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*. 2020; doi: 10.1186/s13059-020-02134-9.

83. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016; doi: 10.1093/bioinformatics/btv566.

84. Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit – Interactive Quality Assessment of Genome Assemblies. *G3 Genes|Genomes|Genetics*. 2020; doi: 10.1534/g3.119.400908.

85. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, et al.. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*. 2018; doi: 10.1093/nar/gky379.

86. Baril T, Galbraith J, Hayward A. Earl Grey: A fully automated user-friendly transposable element annotation and analysis pipeline. *Molecular Biology and Evolution*. 2024; doi: 10.1093/molbev/msae068.

87. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *CP in Bioinformatics*. 2009; doi: 10.1002/0471250953.bi0410s25.

88. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al.. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences USA*. 2020; doi: 10.1073/pnas.1921046117.

89. Kapusta A, Suh A. Evolution of bird genomes—a transposon's-eye view. *Annals of the New York Academy of Sciences*. 2017; doi: 10.1111/nyas.13295.

90. O'Leary NA, Cox E, Holmes JB, Anderson WR, Falk R, Hem V, et al.. Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. *Sciemtific Data*. 2024; doi: 10.1038/s41597-024-03571-y.

91. Mead D, Ogden R, Meredith A, Peniche G, Smith M, Corton C, et al.. The genome sequence of the European golden eagle, *Aquila chrysaetos chrysaetos* Linnaeus 1758. *Wellcome Open Research*. 2021; doi: 10.12688/wellcomeopenres.16631.1.

92. Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, Pozdnyakov IA, et al.. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Research*. 2015; doi: 10.1093/nar/gku1220.

93. Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, et al.. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *BioRxiv*. 2023; doi: 10.1101/2023.06.10.544449

94. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: Unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*. 2016; doi: 10.1093/bioinformatics/btv661.

95. Brůna T, Li H, Guhlin J, Honsel D, Herbold S, Stanke M, et al.. Galba: genome annotation with miniprot and AUGUSTUS. *BMC Bioinformatics*. 2023; doi: 10.1186/s12859-023-05449-z.

96. Jacques Dainat, Darío Hereñú, Dr. K. D. Murray, Ed Davis, Ivan Ugrin, Kathryn Crouch, et al.. NBISweden/AGAT: AGAT-v1.4.1. Zenodo;

97. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. 2011; doi: 10.1038/nature10231.

98. Béziers P, Roulin A. Sexual maturity varies with melanic plumage traits in the barn owl. *Journal of Avian Biology*. 2021; doi: 10.1111/jav.02715.

99. Brommer JE, Pietiäinen H, Kolunen H. Reproduction and survival in a variable environment: Ural owls (Strix uralensis) and the three-year vole cycle. Marti C, editor. *The Auk*. 2002; doi: 10.1093/auk/119.2.544.

100. Fujito NT, Hanna ZR, Levy-Sakin M, Bowie RCK, Kwok P-Y, Dumbacher JP, et al.. Genomic variation and recent population histories of Spotted (*Strix occidentalis*) and Barred (*Strix varia*) Owls. Lohmueller K, editor. *Genome Biology and Evolution*. 2021; doi: 10.1093/gbe/evab066.

101. Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*. 2017; doi: 10.1038/ng.3748.

102. He W, Yang J, Jing Y, Xu L, Yu K, Fang X. NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. Marschall T, editor. *Bioinformatics*. 2023; doi: 10.1093/bioinformatics/btad121.

103. Luo H, Jiang X, Li B, Wu J, Shen J, Xu Z, et al.. A high-quality genome assembly highlights the evolutionary history of the great bustard (Otis tarda, Otidiformes). *Communications Biology*. 2023; doi: 10.1038/s42003-023-05137-x.

104. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*. 2019; doi: 10.1186/s13059-019-1832-y.

105. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al.. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*. 2019; doi: 10.1093/nar/gky1085.

106. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Tamura K, editor. *Molecular Biology and Evolution*. 2021; doi: 10.1093/molbev/msab293.

107. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. Gibas C, editor. *PLoS ONE*. 2011; doi: 10.1371/journal.pone.0021800.

108. Raxworthy M. Animal Cell Culture: A Practical Approach. *Biochemical Education*. 1987; doi: 10.1016/0307-4412(87)90173-7.

109. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009; doi: 10.1093/bioinformatics/btp324.

110. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

111. Van Der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al.. From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *CP in Bioinformatics*. 2013; doi: 10.1002/0471250953.bi1110s43.

112. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van Der Auwera GA, et al.. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv.* 2017; doi: 10.1101/201178