

RESEARCH

Open Access



Centromeric satellite expansion drives genome evolution in the snowy owl

Helle Tessand Baalsrud^{1,2*}, Benedicte Garmann-Aarhus^{1,3}, Emily Louise Gilbert Enevoldsen¹, Anders Kristian Krabberød¹, Dominik Fischer⁴, Ave Tooming-Klunderud¹, Morten Skage¹, Mariann Árnýasi², Simen Rød Sandve², Kjetill Sigurd Jakobsen¹, Rasmus Nielsen^{5,6}, Sanne Boessenkool¹ and Ole Kristian Tørresen¹

*Correspondence:
helle.tessand.baalsrud@nmbu.no

¹ Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway

² Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, Norway

³ Natural History Museum, University of Oslo, Oslo, Norway

⁴ Clinic for Birds, Reptiles, Amphibians and Fish, Justus Liebig University Giessen, Frankfurter Str. 114, Giessen 35392, Germany

⁵ Department of Integrative Biology and Department of Statistics, University of California Berkeley, Berkeley, CA, USA

⁶ Globe Institute, University of Copenhagen, Copenhagen, Denmark

Abstract

Background: Comparative genomic studies in birds have revealed that bird genomes are relatively repeat-poor and stable in terms of karyotype, size, and gene synteny/collinearity compared to other vertebrates. One notable exception is the owls, with cytogenetic studies demonstrating large variations in karyotypes and the evolution of unusual centromeric satellite repeats in some species. However, there has so far not been an investigation into genome architecture and repeat landscape of owls.

Results: Here, we present a chromosome-level genome assembly for the snowy owl (*Bubo scandiacus*). We find that the repeat DNA content in the relatively large snowy owl genome (1.6 Gb) is among the highest reported for any bird genome to date (28.34% compared to an average of ~10% in other birds). The bulk of the snowy owl genomic repeat landscape consists of extensive centromeric satellite arrays that form large contiguous blocks across multiple chromosomes. These arrays represent one of the largest centromeric satellite expansions reported in birds and are consistent with rapid turnover and amplification of centromeric repeats. Using gene collinearity analyses we show that the position of the centromeres are not homologous with chicken centromeres, and are located in regions with collinearity breaks to other bird genomes due to chromosomal rearrangements.

Conclusions: Our results support rapid satellite turnover at lineage-specific centromeres, which could have played a role in centromeric drive, reproductive isolation and speciation of the snowy owl. The extensive centromeric satellite arrays uncovered in this genome highlight centromere evolution as a potentially important driver of genome evolution in owls.

Background

The recent rapid accumulation of high-quality genome assemblies across the Tree of Life has significantly advanced our understanding of genome evolution. However, even within well-studied taxonomic groups like birds [1–4], there are still lineages that lack



© The Author(s) 2026. **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/>.

high-quality long-read reference genomes and thereby represent gaps in our understanding of chromosomal architecture and genomic repeats [3, 5]. The owls (Strigiformes) are one such understudied group.

Earlier cytogenetic studies of the owls have demonstrated fascinating aspects of chromosome evolution within both of its two families, the barn owls (Tytonidae) and the true owls (Strigidae). Tytonidae seem to lack the classic avian genome organization of micro- and macro chromosomes [6]. Strigidae are also unusual by having large genome sizes compared to other birds (average 1.6 Gb), with several species displaying large heterochromatin blocks in the centromeric regions of their chromosomes [7]. These regions have been shown to harbor satellite repeats unique to the true owls, which cluster phylogenetically by species [7]. Centromeric regions are known hotspots for chromosomal rearrangements in other systems [8, 9]. In owls, however, the genomic impact of these centromeric repeats is still unknown.

Here we present a chromosome-level reference genome assembly for the iconic snowy owl (*Bubo scandiacus*) as part of the Earth Biogenome Project Norway initiative (<https://www.ebpnor.org/english/>). We have used a combination of chromatin interaction mapping (Hi-C) and long reads from both Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT), which allows us to characterize the genomic repeat landscape of this species. By comparing the snowy owl assembly with representative chromosome-level assemblies from other birds, we investigate the following three questions: What is the overall repeat content in the snowy owl genome compared to other birds? How are the centromeric satellite DNA and other classes of repeats distributed across the snowy owl repeat landscape? And what is the impact of the unique centromere-associated repeat landscape on synteny to other avian genomes? This in-depth genomic investigation of the repeat landscape of the snowy owl will lift our knowledge of bird genome evolution. Furthermore, this study can facilitate future research and management on the snowy owl (*Bubo scandiacus*), which is a key apex predator and an indicator species for the health of the Arctic ecosystems.

Results

A chromosome-level genome assembly for the snowy owl

We obtained a highly contiguous, phased, diploid snowy owl assembly (Fig. 1) by leveraging the high accuracy of PacBio HiFi reads, the length of ONT reads, and linkage information in Hi-C reads (Additional file 1: Fig. S1). While the expectation is that the assembly is mostly phased within each chromosome, the chromosomes themselves are not phased with respect to each other, that is, chromosomes one and two in haplotype one might not be from the same parent. Recognizing that the two assemblies are technically pseudo-haplotypes if defining haplotypes as unions of segments of DNA inherited from the same parent, we refer to them as haplotypes (hap1 and hap2). The resulting assembly was 1.6 Gb and had a contig N50 of 16.5 Mb and scaffold N50 of 51 Mb for hap1 (Fig. 1). Both sex chromosomes were assigned to hap1 for convenience. Chromosomes were numbered according to size in hap1, with the homologous chromosome in hap2 given the same number. Completeness based on k-mers was 96.5% for hap1, 86.4% for hap2 and 98.4% for both combined. Hap1 had 97.3% complete genes based on BUSCO, while hap2 had 90.1% BUSCO completeness. Using a manual curation process,

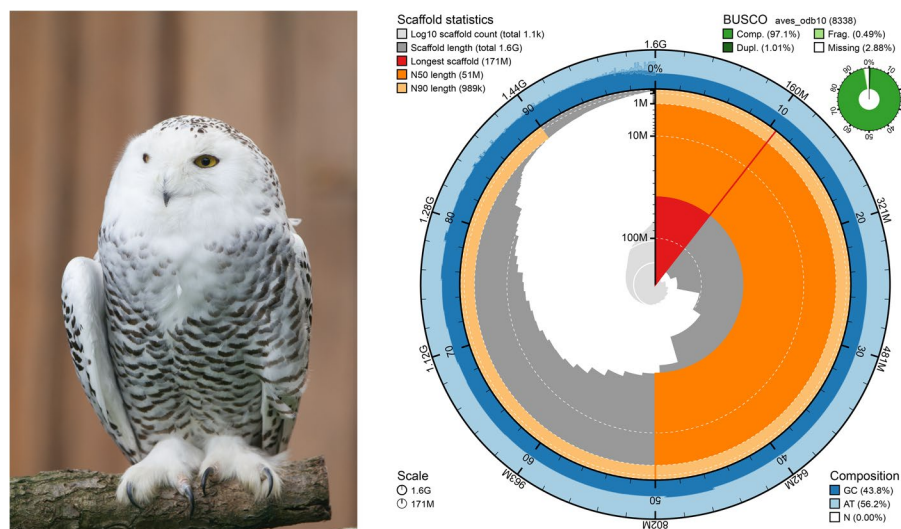


Fig. 1 Genome assembly of *Bubo scandiacus*, bBubSca1.1: metrics. The BlobToolKit Snailplot shows a graphical representation of the assembly metrics for hap1 presented in Additional file 2: Table S1. The snailplot circle represents the assembly, with the red line represents the largest scaffold and all other scaffolds arranged clockwise after this, with scaffold N50 and N90 shown according to legend. The blue areas show base composition, and the smaller green circle shows BUSCO gene completeness. The picture of the snowy owl female sequenced in this project was taken by Wolfgang Holtmeier at the Raptor Center & Wildlife Zoo, Hellenenthal

which included homology comparisons to the assemblies of chicken and zebra finch (see [Methods](#) for details), we identified 32 and 30 autosomes in hap1 and hap2, respectively. Cytogenetic studies have previously found that the chromosome number is $2N = 82$ in the snowy owl [10], which indicates that 10 pairs of the microchromosomes were not distinguished from unplaced scaffolds in our assembly. This is typical for vertebrate species with microchromosomes and dot chromosomes, which are notoriously difficult to characterize [11, 12].

Our comparisons between hap1 and hap2 revealed several putative inversions, duplications, and translocations (Additional file 1: Fig. S2). In total, we found 159 syntenic regions (~1 Gb), 16 inversions (~45 Mb), 155 translocations (~48 Mb), 121 duplications in hap1 (~1 Mb), and 194 duplications in hap2 (~2 Mb). The largest inversions were found on chromosomes 2, 3, 4, and 5.

Phylogenetic analyses based on orthologs

In order to study genome evolution in the snowy owl, we selected five other species with high-quality assemblies in the Telluraves lineage for comparison: downy woodpecker (*Dryobates pubescens*), Northern Carmine bee-eater (*Merops nubicus*), Northern goshawk (*Accipiter gentilis*), California condor (*Gymnogyps californianus*), and barn owl (*Tyto alba*), as well as two outgroup species: chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*). Our phylogenetic tree placed Strigiformes as a sister clade to Accipitriformes (Fig. 2 and Additional file 1: Fig. S10), which is concordant with the topology in the most recently published phylogenetic tree for birds [4].

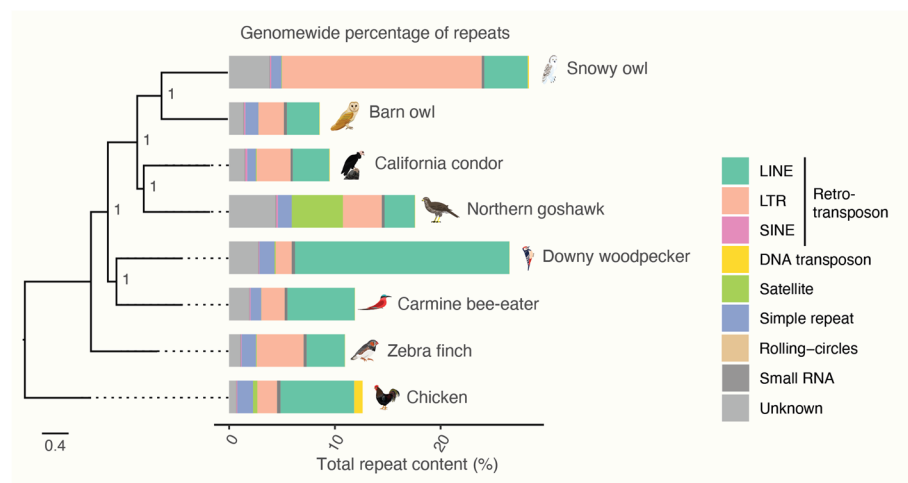


Fig. 2 Repeat content of a selection of avian genome assemblies. Repeat content for all species, with different classes colored according to legend. The phylogenetic tree was generated using Astral. Posterior probabilities are displayed at each node, branch lengths in coalescent units are according to scale bar. This plot was made in R and modified in Illustrator

The snowy owl repeat landscape

De novo repeat modeling and detection revealed that 28.34% of the snowy owl genome consists of repeated elements (Fig. 2, Additional file 2: Table S5), which is significantly higher than the average reported repeat content of ~10% across birds [13]. This makes the snowy owl genome a notable outlier among birds in company with the downy woodpecker [13, 14] and Bell's sparrow (*Artemisiospiza belli*) [15]. Unlike the two latter species, the size of the snowy owl genome is significantly larger than the average avian genome size (1.6 Gb vs an average of ~1 Gb), suggesting that repeat expansions in the snowy owl are driving genome size evolution. To date, snowy owl has the largest avian genome assembly if we consider only those species with chromosome-level assemblies (Additional file 1: Fig. S4). To explore this further, we annotated repeat DNA across all Telluraves and outgroup species in our study and found that the snowy owl had the highest absolute repeat content (Fig. 2, Additional file 2: Table S4). The type of repeat DNA also differed extensively between snowy owl and the other species. In terms of base pairs, most of the repeats in the snowy owl genome were annotated as LTR retrotransposons (LTRs) (Fig. 2 and Additional file 1: Fig. S3). A staggering 255 Mb of the LTRs consist of one element, annotated as endogenous retrovirus 1 (ERV1), class family-0, hereby referred to as class family-0. The overall repeat content of the snowy owl genome goes from 28.42% to 12.56% without family-0 (9.19% in autosomes; Additional file 2: Table S5). In contrast, in the other bird species with high repeat content, the downy woodpecker and Bell's sparrow, there has been an expansion of CR1 LINE elements [13, 15]. The barn owl, California condor, and Northern goshawk genomes do not appear to have similar expansions of LTRs (Fig. 2), which seems to be unique to the snowy owl genome. The Northern goshawk genome is also quite repeat-rich (17.53%) by avian standards, with a large proportion of satellite DNA.

Centromere satellite turnover in the snowy owl genome

To understand the significance of the extensive expansion of class family-0 in the snowy owl genome we investigated this family in more detail. We compared the family-0 consensus sequence (length = 305 bp) to the NCBI nucleotide database using BLAST and found 74 highly significant hits (e-value > 1e27, percentage identity > 83.24) against centromeric repeat region from six species within Strigidae (including snowy owl) [7], and no hits against any other species. This suggests the highly abundant family-0 repeats found in our genome assembly are the centromeric satellite DNA identified in [7], and moreover, that these centromeric repeats are unique to the Strigidae family. Yamada et al. [7] used probes designed with cloned sequences from another species (*Bubo blakistoni*) and estimated that 9.6% of the snowy owl genome consists of this repeat, which is 60% lower than what we found in our high-quality genome sequence (15.8% of hap1). Although family-0 was initially classified as an ERV1-like element during repeat annotation, we did not find independent evidence supporting a specific ERV origin and therefore treat family-0 primarily as a Strigidae-specific centromeric satellite repeat in the analyses below.

To investigate the physical location of centromeric repeats in the snowy owl genome, we estimated the density of different classes of repeats in 200 kb windows, including a separate class for family-0. The family-0 repeats are not evenly spread across the genome but form large, contiguous blocks, consistent with the approximate position of centromeres found in [7] (Fig. 3). Cross-referencing the cytogenetic images in [7] with the presence of family-0 repeats suggests that chromosomes 3, 4, and 5 are metacentric/submetacentric (Additional file 3: Support Note 1). The remaining chromosomes have family-0 repeats at the end and appear to be acrocentric. The two medium-sized chromosomes with a very weak fluorescence signal for the centromeric satellite DNA [7] may correspond to chromosomes 6 and 8 (Fig. 3), which have very small regions of family-0 repeats. Some of the smaller chromosomes (16, 17, 19, and 29) have no family-0 sequence and are difficult to match to the cytogenetically identified chromosomes. The absence of family-0 sequences in these smaller chromosomes may be due to misassembly, as there is 206 Mb of family-0 in unplaced scaffolds of a total 254 Mbp in the assembly.

Size and relative age distribution of snowy owl repeats

We also wanted to look at the relative timing of repeat expansions, especially the family-0 elements, in the snowy owl genome. The distribution of the percentage divergence to the consensus sequence is a good proxy for the relative age of a repeat family. We found that the distribution of percentage divergence varied both within and between repeat families in the snowy owl genome (Additional file 1: Fig. S6). In relative terms, the most recent repeat expansions seem to be SINE 5S, and also some of the simple repeats, rRNAs and LINE/L1 elements. However, there are few of the SINE 5S and L1 copies (Additional file 1: Fig. S3B), consistent with the known paucity of these elements in birds [13]. The LTR annotations had a wide span of percentage divergence to the consensus. Family-0 showed a bimodal distribution of percentage divergence from the consensus (Additional file 1: Fig. S7B). The regions

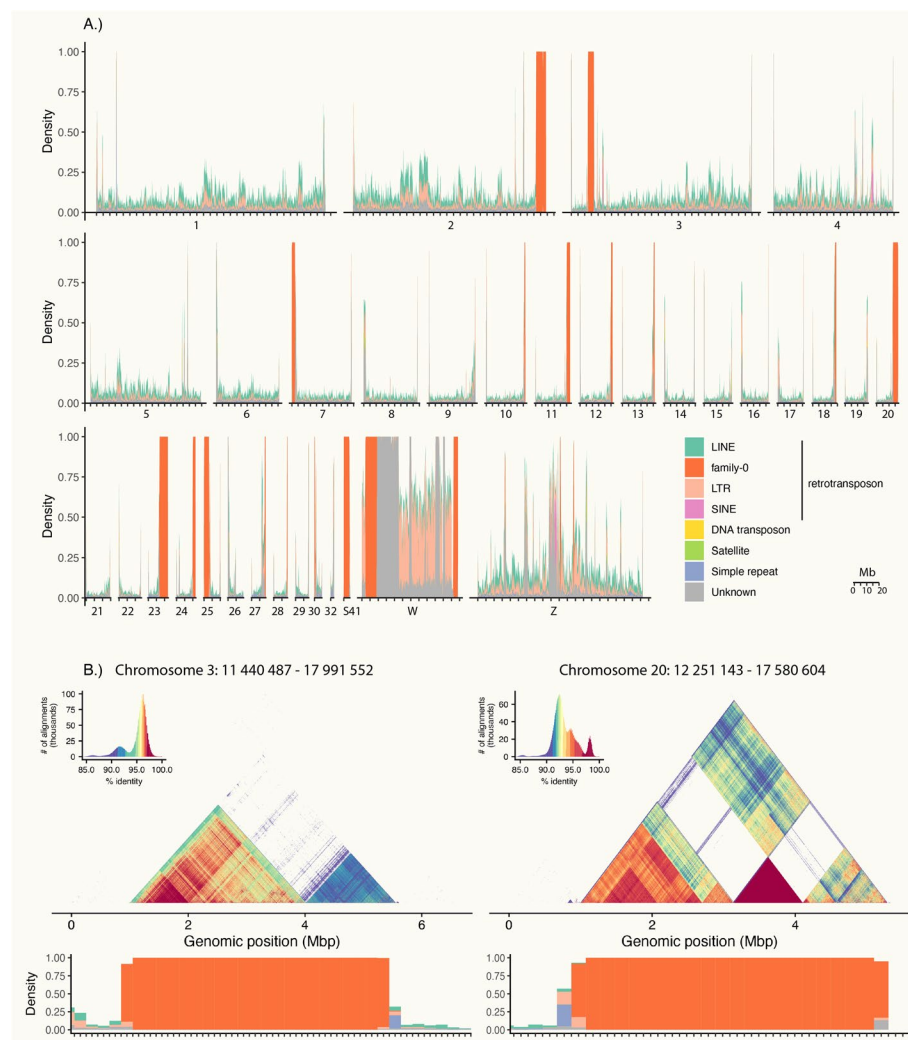


Fig. 3 The chromosomal landscape of repeats in the snowy owl genome. **A** Densities of repeats in 200 kb windows across all curated, putative chromosomes. Different repeat classes are colored according to legend. **B** Examples of higher-order repeat structure of the putative centromere/pericentromere regions at chromosomes 3 and 20. Each region is delimited by the position of family-O elements in addition to 1 Mb of flanking sequence on either side (except for the right region on chromosome 20 where the flanking region is 40 kbp because the chromosome ends). The heatmaps display all-by-all present identity colored according to legend, calculated for 2 kb windows using StainedGlass. The lower panel is a zoom-in of repeat density as in **A** for the same region. This plot was made in R and modified in Illustrator

of family-0 elements ranged in size from 41 to 5 055 234 bp, with a median size of 133 006 bp (Additional file 1: Fig. S7A). This is several magnitudes larger than what would be expected for TEs [16], but is consistent with long stretches of satellite DNA. The family-0 elements form complex higher-order repeat structures in the putative centromeric regions (Fig. 3B, Additional file 1: Fig. S9). In the flanking regions there is no such higher-order repeat structure, however, there are other repeats such as TEs and simple repeats (Fig. 3B). The larger elements are more similar to the consensus sequence than the smaller elements (Additional file 1: Fig. S7C).

Genomic rearrangements and syntenic collinearity

To investigate whether the family-0 expansion has affected genome-wide collinearity between the snowy owl and other birds as regions surrounding centromeres are known hotspots for rearrangements [8, 9, 17], we investigated collinearity based on ortho-groups using GENESPACE [18]. The overall gene collinearity including genomic regions flanking centromere sequences was generally conserved, with a few notable exceptions (Fig. 4). Although collinearity was, as expected, fairly conserved between the snowy owl genome and the other avian genomes included in this study, we also found extensive genomic rearrangements between species (Fig. 4). The number of rearrangements do not necessarily reflect evolutionary distance alone. For instance, there seems to have been many fissions in the barn owl genome since the split with the snowy owl, in addition to translocations indicated by the collinearity between snowy chr2 and barn owl 282.1. Some of these fission events may reflect assembly artifacts, as this assembly is not based on any contact map data. However, cytogenetic studies have shown that the barn owl does not have the classic division between macro- and microchromosomes and instead has a more uniform chromosome distribution. Compared to the barn owl, the snowy owl genome has more conserved collinearity with chicken and zebra finch (Fig. 4). The snowy owl has a higher degree of collinearity with the two species belonging to Accipitriformes, especially the California condor, than with the carmine bee-eater and the downy woodpecker (Fig. 4). Since the split between the snowy owl and the California condor there have been two fission events; California condor's chromosome 1 is split into snowy owl chromosomes 2 and 4, and California condor's chromosome 5 is

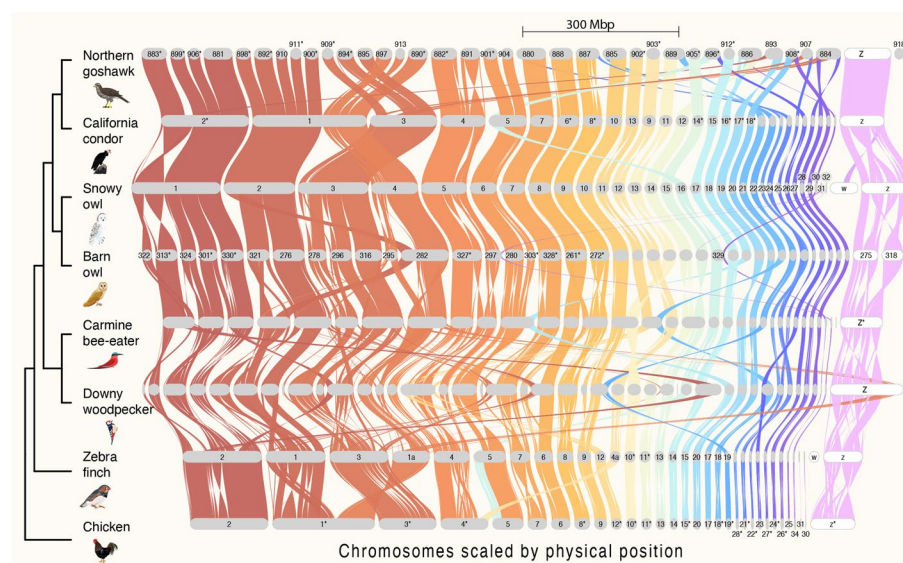


Fig. 4 Chromosome-level syntenic network of snowy owl and representative avian assemblies. The plot was generated by GENESPACE [18]. Genomes are ordered vertically according to phylogenetic relationships. The phylogenetic tree was generated by OrthoFinder [22]. Chromosomes are ordered horizontally to maximize collinearity with the snowy owl genome, with the sex chromosomes (Z and W) shown to the left in white. Only chromosomes or scaffolds with > 100 genes and collinearity blocks > 5 genes were included in the plot. Chromosome segment sizes are scaled by Mbs according to legend. Braids illustrate gene order along the chromosome sequence. Some chromosome names were abbreviated or omitted for clarity, see supporting information for pairwise syntenic comparisons with full names. *Indicate inverted chromosomes. This plot was made in R and modified in Illustrator

split into snowy owl chromosomes 5 and 1. Northern goshawk has undergone several fissions and therefore has smaller macrochromosomes than the California condor. Part of the snowy owl W chromosome is homologous to the Z chromosome in California condor and Northern goshawk, which is most likely the PAR region or gametologous regions between Z and W. The snowy owl W chromosome is 60 Mb, which is relatively large compared to other birds [19, 20]. Northern goshawk also has a large W chromosome (~40 Mb), whereas in the California condor W appears to be missing from the assembly even though a female was sequenced. As W is a refugium for repeats [21] it is often difficult to assemble, and collinearity comparisons of W become challenging.

The lack of large macro chromosomes also seems to characterize carmine bee-eater and downy woodpecker. In the latter species, there have been massive rearrangements, including inversions, fission/fusion events, and translocations (Fig. 4). Interestingly, a large chunk of the downy woodpecker Z chromosome (~23 Mb) is homologous to autosomes in the other birds (e.g. chicken chromosome 1). This could be the result of a fusion between an autosome and Z, potentially forming a neo-sex chromosome.

Centromere position homology between the snowy owl and chicken

Centromeres are known hotspots for chromosome rearrangements [8, 9]. To investigate the link between centromeres and karyotype evolution in the snowy owl, we therefore compared the snowy owl centromere positions and gene order collinearity to the well-characterized chicken genome [12]. Chicken has 7 chromosomes that are submetacentric, metacentric, or sub-telocentric and all of these have large structural rearrangements in comparison to the snowy owl, including fusion/fission events and inversions (Fig. 4). We then compared the chicken centromere position to collinearity breakpoints and found a complete overlap for chicken chromosomes 1, 4, 7, and 8, and near overlap for chicken chromosomes 2, 3, and 5 (Additional file 1: Fig. S11). For the snowy owl we used the position of the family-0 repeats to approximately characterize the centromere locations on the genome. The putative positions of the centromeres in the snowy owl metacentric/submetacentric chromosomes 3, 4, and 5 are not homologous with the centromere position in chicken (Additional file 1: Fig. S8). However, they do overlap with or are very close to breaks in collinearity between these two species in chromosomal regions with many rearrangements. We also investigated collinearity between the snowy owl centromeres on chromosomes 3, 4, and 5 with barn owl and found that the snowy owl centromere position overlaps with collinearity breaks, but not with fission breakpoints in the barn owl (Additional file 1: Fig. S12). There are three putative fission events in the snowy owl compared with chicken, two of which occurred in the chicken centromeric region (Chicken chromosomes 1 and 4)(Fig. 4 and Additional file 1: Fig. S11). Only snowy owl chromosomes 2 and 12 have centromeres located in the fission breakpoints resulting in acrocentric chromosomes, whereas chromosomes 4 and 5 have centromeres in a genomic region not homologous to the fission breakpoints. The regions flanking centromeres at chromosomes 4 and 5 has little collinearity with chicken, which is not surprising given the large span of family-0 repeats in the snowy owl in this region. We also investigated the presence of known centromere repeats from chicken in the snowy owl genome to investigate if there are traces of centromeres present in ancestor of chicken and snowy owl. The chicken genome has different centromere repeats

at different chromosomes, and these repeats are placed in two main categories, CNM (Chicken Nuclear Membrane-associated) and non-CNM [12]. We therefore carefully investigated the presence of specific centromere repeats for each pair of syntenic chromosomes between snowy owl and chicken (Fig. 4 and Additional file 2: Table S7). We found that the chicken centromere repeats often flanked the family-0 repeats (Additional file 2: Table S7). This indicates that the position of centromeres in the snowy owl is syntenic with chicken, although there has been a slight shift, and there has been a turnover of satellite DNA. For snowy owl chromosomes 3, 4, 7, 9, 10, 13, 15, 22, 23, 26 and 27 there seems to have also been a shift in the position of the centromere (Additional file 2: Table S7) but needs to be verified with functional centromere annotation.

Discussion

We have produced a chromosome-level assembly for the snowy owl using a combination of PacBio HiFi, Oxford Nanopore Technologies and Hi-C reads (Fig. 1). The assembly includes the W chromosome, which is notoriously difficult in birds because it is a refugium for repeats [21]. We found that 28.34% of the snowy owl genome consists of repeats, which is one of the highest recorded for any bird and on par with that of the woodpeckers [13, 14]. It is also significantly higher than in the recently published Ural owl (*Strix uralensis*) chromosome-level assembly, where the repeat content is 12.48% [23]. The lion's share of the repeats consists of a recent expansion of centromeric satellite repeats in Strigidae owls and is most likely the reason why the owl genomes are relatively large with C-values spanning from 1.4 to 2.0 (Animal genome size database, genomesize.com, accessed May 6th 2024). The snowy owl assembly was produced using two types of long reads (ONT and HiFi), which was crucial to resolve some of the complex repeat DNA regions, such as in the large regions of centromeric satellite repeats and the W chromosome (Fig. 3). Currently neither technology results in complete representation of all DNA sequences. In chicken, ONT was better at resolving the euchromatic regions with high GC content on the smallest microchromosomes ('dot chromosomes') [12]. Here we have identified some, but not all microchromosomes from unplaced scaffolds using both ONT and HiFi. Increasing ONT coverage in the future might amend this issue, as well as lead to placing more of the centromeric satellites in chromosomes.

A main result in our investigations of the snowy owl repeat landscape is the presence of large centromere satellite arrays (Fig. 3.) The largest snowy owl centromere array of 5.272 Mb is several orders of magnitude larger than the largest reported for zebra finch of 677 kb [24]. For most bird species centromere associated sequences are uncharacterized, limiting the potential for comparative insights. What we do know, however, is that the presence of centromeres is almost ubiquitous in eukaryotes, yet the centromeric associated sequence is not conserved, and in most organisms centromeres are determined epigenetically [25]. Centromere evolution is therefore quite dynamic, with the appearance of neocentromeres, evolutionary new centromere (shifting of centromeric position on a chromosome) and turnover of the centromeric satellite DNA. It is therefore not unusual that the centromeres have different sequences even between chromosomes in the same species, including in birds like chicken [26]. A previous study has reported that the ancestor of the Strigidae family likely evolved new centromere satellites that replaced all previous centromere sequences, as indicated by the presence of lineage-specific

centromere satellites on all chromosomes [7]. Our snowy owl genome assembly supports these previous findings of complete centromere satellite turnover, similar to findings from other birds [27] and *Drosophila* [28]. In addition to their exceptional size, the family-0 arrays exhibit a complex internal organization (Fig. 3B). The most homogeneous arrays form large higher-order repeat structures in the putative centromeric regions, whereas more divergent copies occur towards the flanking regions. Consistent with this pattern, family-0 displays a bimodal divergence distribution (Additional file 1: Fig. S7), suggesting that different parts of the arrays have experienced distinct evolutionary histories. One possibility is that the highly homogeneous repeats correspond to the functional centromere, while the more divergent repeats form surrounding pericentromeric regions. Testing this hypothesis will require functional annotation of owl centromeres, for example through mapping of CENP-A binding sites. The consensus sequence for the centromeric satellite DNA (family-0 element) is 305 bp and consist of two repeat copies (Additional file 1: Fig. S8). There are typically 146 bp surrounding a nucleosome [25, 29], corresponding to the repeat motif of family-0. At some point in the evolutionary history of snowy owl the centromeres have been invaded by this satellite repeat. We do not have the absolute timing of this event. However, our finding that the putative centromere satellites (family-0 elements) are present in the snowy owl and not the barn owl genome supports that these satellite elements are lineage-specific to the Strigidae owls and are not present in Tytonidae. It is possible that the family-0 repeats could have been lost in the barn owl genome, but the more parsimonious explanation is that these centromere-associated repeats expanded in Strigidae after the split with Tytonidae ~40 million years ago [30]. We do not conclusively know if the position of the centromeres have shifted to form Evolutionary New Centromeres (ENCs) [27] in the snowy owl genome. Some of the snowy owl chromosomes have sequences similar to the chicken centromere repeats that do not overlap with the family-0 repeat loci (Additional file 2: Table S7). This could imply that some of the centromeres have shifted positions relative to chicken centromeric loci. To further elucidate the shifts in centromere position and centromere sequence content in owls requires more detailed investigations into the epigenomic centromere positions, i.e. where CENP-A binds to the genome in owl species with and without fam0 centromere repeats. The evolutionary origin of the Strigidae-specific centromeric satellite also remains unclear. Future comparative analyses of repeat landscapes and transposable element evolution across Strigidae and Tytonidae may help elucidate how these lineage-specific satellites originated and expanded.

The observation of large blocks of centromeric satellites (Fig. 3) combined with the knowledge that centromeric satellite sequences are highly divergent between Strigidae species could be explained by centromeric drive. Centromeres can behave as selfish genetic elements that compete to be one of the four meiotic products during female meiosis that becomes the final oocyte [31]. This centromeric drive selects for repeat expansions which can quickly be fixed in a population, but is counterbalanced by selection in the 'host' genome to prevent distortion of segregation during meiosis [27, 30]. This evolutionary arms race can ultimately lead to diversification and speciation, which has been demonstrated in several model systems including chicken [28], the *Drosophila* genus [32], and maize (reviewed by [33]). The presence of megabase-scale centromeric satellite arrays that are highly similar in the snowy

owl genome (Fig. 2, Additional file 1: Fig. S4) and phylogenetically within different species of Strigidae [7] raises the question whether centromeric drive was involved in the diversification of this lineage. Similar patterns were reported in crows [34] and voles [35]. Centromeric drive may therefore be quite prevalent in nature, but underreported as it requires 1.) known centromere positions and 2.) pedigree data to track the fate of centromeres in female meiosis. In birds, these data are not available for most species representing a significant knowledge gap in avian genome biology.

Rapid turnover of centromere satellites, repeat expansions, and putative centromeric drive could lead to genome instability and more frequent chromosome rearrangements and therefore cause breaks in collinearity with other species [36, 37]. However, we find that collinearity is highly conserved between the genomes of snowy owls and distant relatives such as chicken and zebra finch (Fig. 4). This is concordant with a study showing that repositioning of centromeres in reptiles did not disrupt synteny with other vertebrates [17]. In contrast, the downy woodpecker genome is repeat-rich and highly rearranged compared with the other species in this study (Fig. 3), which could be related to the recent high TE activity in this species [14]. The fusion event between the Z chromosome and an autosome (chromosomal region homologous to chicken chromosome 1) could represent the formation of a neo-sex chromosome in this lineage, which has also been observed in other birds [38–40]. Clearly, repeat expansions have had very different impacts on genome evolution in the snowy owl and downy woodpecker.

We have so far assumed adaptive evolution is not the cause of the high similarity between the family-0 elements but is rather due to a recent expansion of these satellite repeats influenced by centromeric drive. However, a recent study on the rapid evolution of new centromere satellites in the owl monkey posits the hypothesis that centromere replacement has been driven by adaptation to a nocturnal lifestyle [41]. Satellite sequences that once served a centromeric function appear to have been repurposed to form heterochromatic structures within the eye's microlens, potentially enhancing vision in low light. The possibility of satellite DNA being adaptive by having structural functions is a possible avenue of future research.

The genomic resources generated here are also a useful contribution to future phylogenetic investigations of birds. The phylogenetic placement of Strigiformes within the Telluraves (core landbirds) has been debated over the years [2, 5, 42, 43], as well as the phylogenetic relationships within Strigiformes [44]. Support for different topologies has depended on data type, method and/or number of taxa included in the analyses. A recent phylogenomic study of birds found that Strigiformes are phylogenetically closer to Accipitriformes than to Pittiformes + Coraciiformes [4]. This is concurrent with our results based on gene orthogroups (Fig. 2 and Additional file 1: Fig. S10). Stiller et al. [4] found that a sampling bias in favor of Passeriformes is the driver of not identifying Strigiformes and Accipitriformes as sister clades, highlighting the value of sequencing more species from these two groups. Given the highly conserved collinearity between Strigiformes and Accipitriformes (Fig. 4), it seems parsimonious to suggest that they are indeed sister groups [4].

Conclusion

We present a chromosome-level genome assembly for the snowy owl and show that its unusually large genome is largely explained by extensive expansion of Strigidae-specific centromeric satellite DNA. These megabase-scale satellite arrays represent one of the largest centromeric repeat expansions reported in birds and support rapid turnover of lineage-specific centromere-associated repeats. Together with the observed association between centromere positions and chromosomal rearrangements, our results highlight centromere evolution as a potentially important driver of genome architecture, and possibly speciation in owls.

Methods

Genome sample

We collected blood from a captive adult female snowy owl from the wildlife park Greifvogelstation & Wildfreigehege Hellenthal in Germany during a routine veterinary examination and put it on dry ice for preservation. Handling of the animal and blood collection were performed during routine hematological and blood chemical examinations and no additional handling, restraint or sampling was necessary for the purpose of this study. This individual was chosen for sequencing because obtaining a high-quality, snap-frozen tissue sample from a wild snowy owl is difficult in the wild. We further selected this individual because it is a female and thus has both sex chromosomes (W and Z, respectively). This snowy owl is captive-bred in Germany and believed to be of European ancestry.

DNA sequencing

DNA isolation started from 10–20 µl of ethanol-preserved blood per reaction. For ONT sequencing DNA extraction and library preparation was done using the Nanobind UL Library Prep Kit (Circulomics) with the combination of the Ultra-Long DNA Sequencing Kit (SQK-ULK001) (Oxford Nanopore) following the manufacture protocol. The library was loaded on an R9.4.1 flow cell and was sequenced using a Promethion24 device. To maximize the data output we performed a nuclease flush when the ratio of the sequencing pores dropped under 20%. Nuclease flush and library reloading were repeated twice. Guppy 5.1.13 was used for base calling with the High-accuracy base calling model.

For PacBio long-read sequencing, DNA extraction was performed according to the Circulomics protocol. First, ethanol was removed from the pelleted blood, and then washed twice in EtOH removal buffer (see Circulomics protocol). DNA was then isolated from the washed blood pellet using the Circulomics Nanobind CBB kit and protocol according to the manufacturer's recommendations (Circulomics, now PacBio company). High molecular weight DNA was eluted from the Nanodisk with 150 µl Tris-cl buffer, incubated overnight at room temperature before quality check of amount, purity, and integrity with UV-absorbance ratios (Nanodrop, Thermo Fisher), Qubit BR DNA quantification assay kit (Thermo Fisher) and Fragment Analyser using the DNA HS 50 kb large fragment kit (Agilent Tech.) Before PacBio HiFi library preparation, DNA was purified for an additional time using PacBio Ampure Beads (1:1 ratio). Then 7.5 µg of purified high molecular weight DNA was sheared to approx. 15–20 kb large fragments using the

Megaruptor3 (Diagenode) in 200 µl volume. We used 5 µg of fragmented DNA with the PacBio library preparation kit 2.0 (PacBio). The final HiFi library was size-selected with a 10 kb cut-off using a BluePippin instrument (Sage Biosciences) before sequencing on the PacBio Sequel II instrument at the Norwegian Sequencing Centre (NSC). We performed all QC steps with the same instrumentation as explained above.

HiC libraries were made using the Dovetail® Omni-C® kit, following the Omni-C proximity Ligation Assay version 1.0 (Dovetail Inc), starting from 20–30 µl of EtOH preserved blood. Final library quality was assayed using instrumentation described above as well as qPCR using the Kapa Library quantification kit for Illumina (Roche Inc.), before sequencing with other libraries on a quarter Illumina Novaseq S4 flowcell with 2*150 bp paired-end mode at the Norwegian Sequencing Centre.

Genome assembly and curation

We assembled the species using a pre-release of the EBP-Nor genome assembly pipeline (<https://github.com/ebp-nor/GenomeAssembly>). KMC v3.1.2rc1 was used to count k-mers of size 21 in the PacBio HiFi reads [45]. GenomeScope v2.0 was run on the k-mer histogram output from KMC to get estimates of genome size, heterozygosity, and repetitiveness [46]. HiFiAdapterFilt v2.0.0 was applied on the HiFi reads to remove remnant PacBio adapter sequences [47]. The filtered HiFi reads were assembled using hifiasm v0.16.1 with Hi-C integration [48], resulting in a pair of haplotype-resolved assemblies, hap1, and hap2. Unique k-mers in each assembly were identified using meryl v1.3.0 (<https://github.com/marbl/meryl>), and used to create two sets of Hi-C reads, one without any k-mers occurring uniquely in hap1 and the other without k-mers occurring uniquely in hap2.

The ONT reads were assembled with Flye 2.9-b1768 using default options [49]. The Flye assembly was used to scaffold both haplotype-resolved assemblies from hifiasm using RagTag v2.1.0 [50]. K-mer-filtered Hi-C reads were aligned to each scaffolded assembly using BWA-MEM v0.7.17 with the -5SPM option [51]. The alignments were sorted based on name using samtools 1.15.1, then samtools fixmate was applied, before default sorting and applying samtools markdup [52]. The resulting BAM file was used to further scaffold the two RagTag-scaffolded assemblies using YaHS v1.1a with default options [53]. The two resulting assemblies were then gap-filled using the Flye assembly with RagTag patch [50].

FCS-Adaptor v0.2.2 (<https://github.com/ncbi/fcs>) was run on the assemblies, and any adaptor sequences found were masked using bedtools v2.30.0 maskfasta [54]. FCS-GX v0.2.2 (<https://github.com/ncbi/fcs>) was used to search for contamination. If a contaminant was found at the start or end of a sequence, the sequence was trimmed using a combination of samtools faidx, bedtools complement, and bedtools getfasta. If the contaminant was internal, it was masked using bedtools maskfasta. The mitochondrial genome was assembled using MitoHiFi v2.2, based on the contig assembly of pseudo-haplotype 1 [55].

Merqury v1.3 was used to assess the completeness and quality of the genome assemblies by comparing them to the k-mer content of the Hi-C reads [56]. BUSCO v5.3.1 was used to assess the completeness of the genome assemblies by comparing them against

the expected gene content in the Aves lineage (OrthoDB v10) [57]. Gfastats v1.2.2 was used to output different assembly statistics of the assemblies [58].

The assemblies were manually curated using the GRIT rapid curation suite [59] and the PretextView v0.2.5 (<https://github.com/wtsi-hpag/PretextView>, last accessed October 25, 2022). During curation 119 and 89 corrections were made to pseudo-haplotypes 1 and 2, respectively. Through this, we identified 32 putative chromosomes for pseudo-haplotype 1, and 30 for pseudo-haplotype 2 via visual inspection. The Z and W chromosomes were curated by mapping against the chicken sex chromosomes using minimap v2.1 [60].

The Hi-C contact maps for both pseudo-haplotypes were visualized before and after curation using PretextSnapshot (<https://github.com/wtsi-hpag/PretextSnapshot>, last accessed October 25, 2022).

Comparing pseudo-haplotypes 1 and 2

To compare the two pseudo-haplotype assemblies, we first aligned the homologous chromosomes using minimap2 v2.1 [60] with the asm5 setting. We then used SyRI v1.6.3 [61] to identify structural variation between the two assemblies and visualized the results using plotsr v1.1.0 [62]. For further comparisons, using the MUMmer toolkit v4.0.0rc1 [63], we aligned the sequences using nucmer and then processed the alignment with dnadiff, which produced a list of indels, SNPs, and insertions all of which are reported in Additional file 2: Table S1.

Comparative genomics data

For comparative genomic analyses, we selected five species from the Telluraves clade (core landbirds, which includes Strigiformes) which have chromosome-level assemblies; downy woodpecker (*Dryobates pubescens*), Northern Carmine bee-eater (*Merops nubicus*), Northern goshawk (*Accipiter gentilis*), California condor (*Gymnogyps californianus*) and barn owl (*Tyto alba*). The latter species is not chromosome-level but was included as it is in the owl order (Strigiformes) and the only long-read assembly for owls available. We also included the genomes of chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*), as these are well-studied avian reference genomes. Versions of assemblies and an overview of genome characteristics are listed in Additional file 2: Table S3.

Genome annotation

Gene annotations downloaded and used for the chicken, zebra finch, and California condor genomes are listed in Additional file 2: Table S3. For the other species, we annotated the genome assemblies using a pre-release version of the EBP-Nor genome annotation pipeline (<https://github.com/ebp-nor/GenomeAnnotation>). First, AGAT v1.0 agat_sp_keep_longest_isoform.pl and agat_sp_extract_sequences.pl were used on the chicken assembly and annotation to generate one protein (the longest isoform) per gene [64]. Miniprot v0.5 was used to align the proteins to the curated assemblies [65]. UniProtKB/Swiss-Prot release 2022_03 [66] in addition to the Vertebrata part of OrthoDB v10 [67] were also aligned separately to the assemblies. RED v2018.09.10 was run via redmask (<https://github.com/nextgenusfs/redmask>) on the assemblies to mask repetitive areas

[68]. In addition RepeatModeler v2.0.3 [69] was run on the hap1 genome assembly via the Dfam TE Tools container v1.6 (<https://github.com/Dfam-consortium/TETools>) to create a species-specific repeat library, and RepeatMasker 4.1.3-p1 (from the same container) was used to apply the library to the genome and discover repetitive elements [70]. GALBA (<https://github.com/Gaius-Augustus/GALBA>, commit: f4aaeca) was run with the chicken proteins using the miniprot mode on the masked assemblies [65, 71–74]. The funannotate-runEVM.py script from Funannotate v1.8.13 [75] was used to run EvidenceModeler v1.1.1 [76] on the alignments of chicken proteins, UniProtKB/Swiss-Prot proteins, vertebrata proteins and the predicted genes from GALBA. The resulting predicted proteins were compared to the protein repeats that Funannotate distributes using DIAMOND v2.0.15 [72] blastp and the predicted genes were filtered based on this comparison using AGAT. The filtered proteins were compared to the UniProtKB/Swiss-Prot release 2022_03 using DIAMOND blastp to find gene names and InterProScan v5.47–82 [77] was used to discover functional domains. AGATs agat_sp_manage_functional_annotation.pl was used to attach the gene names and functional annotations to the predicted genes. For the Downy woodpecker, Northern Carmine bee-eater, Northern goshawk, and barn owl the same annotation pipeline was used as for the snowy owl, with the exception that we used the already soft-masked genome assemblies available at NCBI (Additional file 2: Table S3). To compare the repeated content between these species, we obtained a bird specific repeat library from Alexander Suh (personal communication) that was used in [78]. We ran RepeatMasker [70] with the repeat library on all species.

Synteny and orthology inference

To investigate chromosome-level synteny, including collinearity, between the snowy owl genome and chicken, zebra finch, California condor, Downy woodpecker, Northern Carmine bee-eater, Northern goshawk, and barn owl we used GENESPACE v0.9.3 [18] with OrthoFinder v2.5.4 [22] in default mode using gene annotations and protein sequences from each species. GENESPACE cannot handle too many scaffolds for comparing collinearity blocks, thus we included only chromosomes if chromosome information was available, or scaffolds > 5 Mb and putative chromosomes for the snowy owl (see information above). We ran OrthoFinder separately on the entire dataset including unplaced scaffolds in order to assess genome-wide patterns of gene duplications/losses. We obtained the species tree from multi-copy gene family trees (as generated by OrthoFinder) using ASTRAL-Pro3 v1.16.2.4 [79] by optimizing the objective function of ASTRAL-Pro [80]. Branch lengths are computed using integrated CASTLES-Pro [81].

Repeat landscape

We used the output from RepeatMasker [70] to plot the size distribution and percentage divergence from consensus using custom R scripts. We estimated the repeat density of different repeat classes in 200 kb windows across the snowy owl genome using BED-Tools v2.30.0 coverage [54] and custom R script. We used StainedGlass [82] to create identity heatmaps of tandem repeat structures in the putative centromere regions, with a window size of 2000 bp.

We investigated ERV family-0 elements in more detail. First, we blasted [83] the consensus sequence against NCBI identify orthologous sequences. We aligned the hits in using ClustalX [84]. Second, we investigated all family-0 elements for putative target site duplications (TSDs) using a custom R script.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-026-04175-y>.

Additional file 1: Figs. S1–S12. Supplementary Figures cited in the main text.

Additional file 2: Tables S1–S3. Supplementary tables cited in the main text.

Additional file 3: Support Note 1 cited in the main text.

Acknowledgements

The authors are grateful to the wildlife park Greifvogelstation & Wildfreigehege Hellenthal, Germany, for the supply of biological material for this study. We thank The International Snowy Owl Working Group (ISOWG) for advice and support. We thank the Vertebrate Genomes Project (VGP) for earlier access to the genomes generated, particularly the Carmine bee-eater genome (Thomas Gilbert) and the Downy Woodpecker genome (Matthew Fuxjager), and the Chair Erich Jarvis for his help. We are grateful to Luohao Xu for providing the centromere positions in the chicken genome assembly. All laboratory work was carried out at the Norwegian Sequencing Centre (NSC) and at the Centre for Integrative Genetics (CIGENE) laboratory. The computations were performed on resources provided by Sigma2 - the National Infrastructure for High-Performance Computing and Data Storage in Norway.

Peer review information

Andrew Cosgrove and Wenjing She were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions

The study was designed by HTB. Sampling was carried out by DF. DNA isolation, library preparation, and sequencing was carried out by AKT, MS, and MA. HTB, OKT, BGA, ELGE, SRS, KSJ, RN, SB, and AKK contributed to analyses and interpretation of results. Figures were made by HTB. Manuscript was written by HTB and OKT with input from all co-authors. All authors read and approved the final manuscript.

Authors' information

Ole Kristian Tørresen: @oletorresen.bsky.social.
Helle Tessand Baalsrud: @baalsrud.bsky.social.
Kjetill Sigurd Jakobsen: @kjetillsj.bsky.social.
Anders Kristian Krabberød: @krabberod.bsky.social.

Funding

This study was supported by the Earth Biogenome Project Norway, funded by the Research Council of Norway (grant # 326819 to KSJ, University of Oslo), and The Peder Sather Center for Advanced Study (to HTB and RN).

Data availability

All sequencing data for snowy owl are available in the European Nucleotide Archive (ENA) under BioProject PRJEB79518 [85]. The umbrella project for EBP-Nor is BioProject PRJEB65317. Raw PacBio HiFi reads for the snowy owl (BioSample SAMEA115805126) are deposited under run accessions ERR13993800, ERR13993792 and ERR13993789. Illumina Hi-C sequencing reads are available under run accession ERR13346484 and ERR13993872. The basecalled Oxford Nanopore Technologies Promethlon reads are found at ERR14002240, ERR14002242 and ERR14002241. The raw data Oxford Nanopore Technologies Promethlon reads are available at NIRD Research Data Archive, <https://doi.org/10.11582/2025.87azzuz2> [86]. The two assemblies are archived separately: pseudo-haplotype 1 under BioProject PRJEB77267 (assembly GCA_965212795) and pseudo-haplotype 2 under BioProject PRJEB79513 (assembly GCA_965212775). In addition, the annotations, repeat masking and assemblies produced in this study are available from Zenodo at <https://zenodo.org/records/17302240> (DOI: <https://doi.org/10.5281/zenodo.12643815>) [87]. Previously published genome assemblies used for comparative analyses were obtained from NCBI. These include *Gallus gallus* (bGalGal1), *Taeniopygia guttata* (bTaeGut1.4.pri), *Dryobates pubescens* (bDryPub1.pri), *Merops nubicus* (bMerNub1.pri), *Tyto alba* (T.alba_DEE_v4.0), *Accipiter gentilis* (bAccGen1.1), and *Gymnogyps californianus* (ASM1813914v2). Accession numbers and URLs are provided in Additional file: Table S3.

Declarations

Ethics approval and consent to participate

The Committee for Ethical Approval of Studies Involving Animal Patients at NMBU declared no ethical review and approval was required for this study. The blood sample used in this study was collected during a routine veterinary examination of a captive Snowy Owl (*Bubo scandiacus*) at Greifvogelstation & Wildfreigehege Hellenthal, Germany. Handling of the animal and blood collection were performed as part of routine veterinary care and no additional handling,

restraint, or sampling was conducted for the purposes of this research. Samples collected during routine veterinary diagnostic procedures falls outside the scope of Directive 2010/63/EU in accordance with Article 1(5)(b).

Competing interests

The authors declare no competing interests.

Received: 5 July 2024 Accepted: 25 June 2026

Published online: 30 June 2026

References

- 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;346(6215):1311–20.
- Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*. 2014;346(6215):1320–31.
- 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;587(7833):252–7.
- Stiller J, Feng S, Chowdhury A-A, Rivas-González I, Duchêne DA, Fang Q, et al. Complexity of avian evolution revealed by family-level genomes. *Nature*. 2024;629(8013):851–60.
- Braun EL, Cracraft J, Houde P. Resolving the avian tree of life from top to bottom: the promise and potential boundaries of the phylogenomic era. In: Kraus RHS, editor. *Avian genomics in ecology and evolution: from the lab into the wild*. Cham: Springer International Publishing; 2019. p. 151–210.
- Rebholz WER, Boer LEMD, Sasaki M, Belterman RHR, Nishida-Umehara C. The chromosomal phylogeny of owls (strigiformes) and new karyotypes of seven species. *Cytologia (Tokyo)*. 1993;58(4):403–16.
- 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;112(6):277–87.
- Ola M, O'Brien CE, Coughlan AY, Ma Q, Donovan PD, Wolfe KH, et al. Polymorphic centromere locations in the pathogenic yeast *Candida parapsilosis*. *Genome Res*. 2020;30(5):684–96.
- Smalec BM, Heider TN, Flynn BL, O'Neill RJ. A centromere satellite concomitant with extensive karyotypic diversity across the *Peromyscus* genus defies predictions of molecular drive. *Chromosome Res*. 2019;27(3):237–52.
- Sasaki M, Nishida-Umehara C, Tsuchiya K. A Comparative Study of G-banded Karyotypes in Eight Species of Owls. *Cytologia (Tokyo)*. 1994;59(2):183–5.
- 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. *Proc Natl Acad Sci U S A*. 2021. <https://doi.org/10.1073/pnas.2112494118>.
- Huang Z, Xu Z, Bai H, Huang Y, Kang N, Ding X, et al. Evolutionary analysis of a complete chicken genome. *Proc Natl Acad Sci USA*. 2023;120(8):e2216641120.
- Kapusta A, Suh A. Evolution of bird genomes—a transposon's-eye view. *Ann NY Acad Sci*. 2017;1389(1):164–85.
- Manthey JD, Moyle RG, Boissinot S. Multiple and independent phases of transposable element amplification in the genomes of piciformes (woodpeckers and allies). *Genome Biol Evol*. 2018;10(6):1445–56.
- Benham PM, Cicero C, Escalona M, Beraut E, Fairbairn C, Marimuthu MPA, et al. Remarkably High Repeat Content in the Genomes of Sparrows: The Importance of Genome Assembly Completeness for Transposable Element Discovery. *Gen Biol Evol*. 2024;16(4):evae067. <https://doi.org/10.1093/gbe/evae067>.
- Wells JN, Feschotte C. A field guide to eukaryotic transposable elements. *Annu Rev Genet*. 2020;54:539–61.
- Dobry J, Zhu Z, Zhou Q, Wapstra E, Deakin JE, Ezaz T. Fixed allele differences associated with the centromere reveal chromosome morphology and rearrangements in a reptile (*Varanus acanthurus* BOULENGER). *Mol Biol Evol*. 2023. <https://doi.org/10.1093/molbev/msad124>.
- Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, et al. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife*. 2022. <https://doi.org/10.7554/eLife.78526>.
- Xu L, Zhou Q. The female-specific W chromosomes of birds have conserved gene contents but are not feminized. *Genes (Basel)*. 2020. <https://doi.org/10.3390/genes11101126>.
- Xu L, Auer G, Peona V, Suh A, Deng Y, Feng S, et al. Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. *Nat Ecol Evol*. 2019;3(5):834–44.
- Peona V, Palacios-Gimenez OM, Blommaert J, Liu J, Haryoko T, Jönsson KA, et al. The avian W chromosome is a refuge for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. *Philos Trans R Soc Lond B Biol Sci*. 1833;376(1833):20200186.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20(1):238. <https://doi.org/10.1186/s13059-019-1832-y>.
- Winter S, Meißner R, Grethlein M, Wehrenberg G, Kiebler A, Silva AX, et al. Chromosome-Level genome assembly and transcriptome analysis of the ural owl, *Strix uralensis* Pallas, 1771. *J Hered*. 2025;117(1):107–14. <https://doi.org/10.1093/jhered/esaf038>.
- Formenti G, Jain N, Medico JA, Sollitto M, Antipov D, Barcellos S, et al. The complete genome of a songbird. *BioRxiv*. 2025. <https://doi.org/10.1101/2025.10.14.682431>.
- Talbert PB, Henikoff S. The genetics and epigenetics of satellite centromeres. *Genome Res*. 2022;32(4):608–15.
- Shang W-H, Hori T, Toyoda A, Kato J, Popendorf K, Sakakibara Y, et al. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res*. 2010;20(9):1219–28.
- Malik HS. The centromere-drive hypothesis: a simple basis for centromere complexity. *Prog Mol Subcell Biol*. 2009;48:33–52.
- Axelsson E, Albrechtsen A, Van AP, Li L, Megens HJ, Vereijken ALJ, et al. Segregation distortion in chicken and the evolutionary consequences of female meiotic drive in birds. *Heredity*. 2010;105(3):290–8.

29. van Holde KE. Chromatin. New York, NY: Springer New York; 1989.
30. Henikoff S, Ahmad K, Malik HS. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*. 2001;293(5532):1098–102.
31. Zwick ME, Salstrom JL, Langley CH. Genetic variation in rates of nondisjunction: association of two naturally occurring polymorphisms in the chromokinesin nod with increased rates of nondisjunction in *Drosophila melanogaster*. *Genetics*. 1999;152(4):1605–14.
32. Ferree PM, Barbash DA. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol*. 2009;7(10):e1000234.
33. Dawe RK. The maize abnormal chromosome 10 meiotic drive haplotype: a review. *Chromosome Res*. 2022;30(2–3):205–16.
34. Peona V, Kutschera VE, Blom MPK, Irestedt M, Suh A. Satellite DNA evolution in Corvoidea inferred from short and long reads. *Mol Ecol*. 2023;32(6):1288–305.
35. Romanenko SA, Serdyukova NA, Perelman PL, Trifonov VA, Golenishchev FN, Bulatova NS, et al. Multiple intrasyntenic rearrangements and rapid speciation in voles. *Sci Rep*. 2018;8(1):14980.
36. Plačková K, Bureš P, Lysak MA, Zedek F. Centromere drive may propel the evolution of chromosome and genome size in plants. *Ann Bot*. 2024;134(6):1067–76.
37. Barra V, Fachinetti D. The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. *Nat Commun*. 2018;9(1):4340.
38. Sigeman H, Zhang H, Ali Abed S, Hansson B. A novel neo-sex chromosome in *Sylvietta brachyura* (Macrosphenidae) adds to the extraordinary avian sex chromosome diversity among Sylvioidea songbirds. *J Evol Biol*. 2022. <https://doi.org/10.1111/jeb.14096>.
39. Huang Z, De O Furo I, Liu J, Peona V, Gomes AJB, Cen W, et al. Recurrent chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. *Nat Commun*. 2022;13(1):944.
40. Sigeman H, Strandh M, Proux-Wéra E, Kutschera VE, Ponnikas S, Zhang H, et al. Avian neo-sex chromosomes reveal dynamics of recombination suppression and W degeneration. *Mol Biol Evol*. 2021;38(12):5275–91.
41. Nishihara H, Stanyon R, Tanabe H, Koga A. Replacement of owl monkey centromere satellite by a newly evolved variant was a recent and rapid process. *Genes Cells*. 2021;26(12):979–86.
42. Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, et al. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*. 2015;526(7574):569–73.
43. Reddy S, Kimball RT, Pandey A, Hosner PA, Braun MJ, Hackett SJ, et al. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Syst Biol*. 2017;66(5):857–79.
44. Salter JF, Oliveros CH, Hosner PA, Manthey JD, Robbins MB, Moyle RG, et al. Extensive paraphyly in the typical owl family (Strigidae). *Auk*. 2020. <https://doi.org/10.1093/auk/ukz070>.
45. Kokot M, Dlugosz M, Deorowicz S. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics*. 2017;33(17):2759–61.
46. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11(1):1432.
47. Sim SB, Corpuz RL, Simmonds TJ, Geib SM. HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics*. 2022;23(1):157.
48. Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, et al. Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol*. 2022;40(9):1332–5.
49. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;37(5):540–46. <https://doi.org/10.1038/s41587-019-0072-8>.
50. Alonge M, Lebeigle L, Kirsche M, Aganezov S, Wang X, Lippman ZB, et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol*. 2022;23(1):258. <https://doi.org/10.1186/s13059-022-02823-7>.
51. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013.
52. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021. <https://doi.org/10.1093/gigascience/giab008>.
53. Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*. 2023. <https://doi.org/10.1093/bioinformatics/btac808>.
54. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
55. Uliano-Silva M, Ferreira JGRN, Krashenninnikova K, Darwin Tree of Life Consortium, Formenti G, Abueg L, et al. Mito-HiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics*. 2023;24(1):288.
56. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;21(1):245.
57. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. *Curr Protoc*. 2021;1(12):e323.
58. Formenti G, Abueg L, Brajuka A, Brajuka N, Gallardo-Alba C, Giani A, et al. Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. *Bioinformatics*. 2022;38(17):4214–6.
59. Howe K, Chow W, Collins J, Pelan S, Pointon D-L, Sims Y, et al. Significantly improving the quality of genome assemblies through curation. *Gigascience*. 2021. <https://doi.org/10.1093/gigascience/giaa153>.
60. Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*. 2021;37(23):4572–4.
61. Goel M, Sun H, Jiao W-B, Schneeberger K. SyRi: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol*. 2019;20(1):277.
62. Goel M, Schneeberger K. plots: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics*. 2022;38(10):2922–6.

63. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol*. 2018;14(1):e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>.
64. Dainat J, Hereñú D, Davis E, Crouch K, LucileSol, Agostinho N, et al. NBISweden/AGAT: AGAT-v1.1.0. Zenodo. 2023.
65. Li H. Protein-to-genome alignment with miniprot. *Bioinformatics*. 2023;39(1):btad014. <https://doi.org/10.1093/bioinformatics/btad014>.
66. UniProt Consortium. Uniprot: the universal protein knowledgebase in 2023. *Nucleic Acids Res*. 2023;51(D1):D523–31.
67. Manni M, Berkeley MR, Seppely 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. *Mol Biol Evol*. 2021;38(10):4647–54.
68. Girgis HZ. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinformatics*. 2015;16:227.
69. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA*. 2020;117(17):9451–7.
70. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0 [Internet]. RepeatMasker Open-4.0. 2015. Available from: <http://www.repeatmasker.org/>. Cited 2023 Jun 22.
71. Brůna T, Li H, Guhlín J, Honsel D, Herbold S, Stanke M, Nenasheva N, Ebel M, Gabriel L, Hoff KJ. Galba: genome annotation with miniprot and AUGUSTUS. *BMC Bioinformatics*. 2023;24(1):327. <https://doi.org/10.1186/s12859-023-05449-z>.
72. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12(1):59–60.
73. Hoff KJ, Stanke M. Predicting genes in single genomes with AUGUSTUS. *Curr Protoc Bioinformatics*. 2019;65(1):e57.
74. Stanke M, Schöffmann O, Morgenstern B, Waack S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*. 2006;7:62. <https://doi.org/10.1186/1471-2105-7-62>.
75. Palmer JM, Stajich J. Funannotate v1.8.1: Eukaryotic genome annotation. Zenodo. 2020.
76. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biol*. 2008;9(1):R7.
77. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30(9):1236–40.
78. Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. *Proc Natl Acad Sci U S A*. 2017;114(8):E1460–9.
79. Zhang C, Nielsen R, Mirarab S. ASTER: a package for large-scale phylogenomic reconstructions. *Mol Biol Evol*. 2025. <https://doi.org/10.1093/molbev/msaf172>.
80. Zhang C, Scornavacca C, Molloy EK, Mirarab S. ASTRAL-Pro: quartet-based species-tree inference despite paralogy. *Mol Biol Evol*. 2020;37(11):3292–307.
81. Tabatabaee Y, Zhang C, Warnow T, Mirarab S. Phylogenomic branch length estimation using quartets. *Bioinformatics*. 2023;39(39 Suppl 1):i185–93.
82. Vollger MR, Kerpedjiev P, Phillippy AM, Eichler EE. StainedGlass: interactive visualization of massive tandem repeat structures with identity heatmaps. *Bioinformatics*. 2022;38(7):2049–51.
83. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10.
84. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947–8.
85. Baalsrud HT, Tørresen OK. Snowy owl (*Bubo scandiacus*) genome sequencing data. Datasets. European Nucleotide Archive. 2025. <https://www.ebi.ac.uk/ena/browser/view/PRJEB79518>.
86. Árnyasi M. Oxford Nanopore sequencing data for *Bubo scandiacus*. Dataset. NIRD Research Data Archive. 2025. <https://doi.org/10.11582/2025.87azzuz2>.
87. Baalsrud HT, Tørresen OK. Gene and repeat annotation for snowy owl (*Bubo scandiacus*) and selected species. 2024. Zenodo. <https://doi.org/10.5281/zenodo.12643816>.

Publisher's Note

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