Chromosome-level genome of the wood stork (*Mycteria americana*) provides insight into avian chromosome evolution

Authors: Richard Flamio Jr.[1] and Kristina M. Ramstad[1]

[1]Biological, Environmental, and Earth Sciences, University of South Carolina Aiken, 471 University Parkway, Aiken, SC, USA

Corresponding author: richard.flamio@gmail.com

Running title: Chromosome-level wood stork (*Mycteria americana*) genome

Abstract

Despite being quite specious (~10,000 extant species), birds have a fairly uniform genome size and karyotype (including the common occurrence of microchromosomes) relative to other vertebrate lineages. Storks (Family Ciconiidae) are a charismatic and distinct group of large wading birds with nearly worldwide distribution but few genomic resources. Here we present an annotated chromosome-level reference genome and chromosome orthology analysis for the wood stork (*Mycteria americana*), a species that has been federally protected under the Endangered Species Act since 1984. The annotated chromosome-level reference assembly was produced using the blood of a wild female wood stork chick, has a length of 1.35 Gb, a contig N50 of 37 Mb, a scaffold N50 of 80 Mb, and a BUSCO score of 98.8%. We identified 31 autosomal pairs and two sex chromosomes in the wood stork genome, but failed to identify four additional autosomal microchromosomes previously found via karyotyping. Orthology analyses confirmed reported synapomorphies unique to storks and identified the chromosomes participating in these fusions. This study highlights the difficulty and potential problems associated with delineating microchromosomes in reference genome assemblies. It also provides a foundation for studying karyotype evolution in the core water bird clade that includes penguins, albatrosses, storks, cormorants, herons, and ibises. Finally, our reference genome will allow for numerous genomic studies, such as genome-wide association studies of local adaptation, that will aid in wood stork conservation.

Keywords: Ciconiiformes, microchromosomes, sequencing, scaffolding, vertebrate, core water birds, orthology

Introduction

Bird genomes feature a relatively small and stable genome size (nuclear DNA content) compared to other vertebrate taxa (Tiersch & Wachtel, 1991). The reason for small genome sizes in birds is unknown, but it has been hypothesized to be adaptive for the energy requirements associated with flight (Hughes & Piontkivska, 2005). Genome organization within Aves is also quite consistent. Birds have a diploid number of $2n \approx 80$ which includes ~10 pairs of macrochromosomes (large chromosomes that can be flow-sorted) and many somewhat smaller microchromosomes, although the transition in size between macrochromosomes and microchromosomes is more gradual than this binary classification implies (Griffin et al., 2007). Compared to macrochromosomes, microchromosomes have high recombination rates and G+C content and are gene dense with little repetitive sequence (International Chicken Genome Sequencing Consortium, 2004). Other vertebrate taxa, such as reptiles (Olmo, 2008), amphibians (Morescalchi, 1980), and fish (Ohno et al., 1969), contain species with microchromosomes as well. However, the presence of many microchromosomes seems to be particularly associated with avian genomes, which may have retained this feature from an original chordate ancestor (Waters et al., 2021).

Storks (Order Ciconiiformes, Family Ciconiidae) are a distinct lineage of large wading birds that constitute the only family in their order. Current molecular evidence places storks within the clade Pelecanimorphae as sister to Pelecanes, a clade that contains Order Suliformes (frigatebirds, gannets, boobies, darters, cormorants, and shags) and Order Pelecaniformes (ibises, spoonbills, herons, bitterns, shoebill, hamerkop, and pelicans) (Burleigh et al., 2015; Hackett et al., 2008; Kimball et al., 2019;

Kuhl et al., 2021; Kuramoto et al., 2015; Prum et al., 2015). More broadly, storks are members of the core water bird clade, Aequornithes, which includes Gaviiformes (loons) and Feraequornithes (Burleigh et al., 2015; Sangster & Mayr, 2021). Feraequornithes contains the Pelecanimorphae (ciconiiforms, suliforms, and pelecaniforms) and the Procellariimorphae (albatrosses, petrels, and penguins) (Burleigh et al., 2015; Sangster & Mayr, 2021).

Traditionally, storks have been classified into three distinct lineages, tribes Mycteriini (genera *Anastomus* and *Mycteria*), Ciconiini (genus *Ciconia*), and Leptoptilini (genera *Leptoptilos*, *Jabiru*, and *Ephippiorhynchus*), based on morphology and behavior (Kahl, 1987). Several lines of evidence, including karyotype analysis by cell staining (de Boer & van Brink, 1982), a DNA-DNA hybridization study (Slikas, 1997), comparison of cytochrome b sequences (Slikas, 1997), and chromosome painting (Seligmann et al., 2019) suggest non-monophyly of the tribe Leptoptilini. The recent stork phylogeny of (Rodríguez-Rodríguez & Negro, 2021) supports this claim. Within this phylogeny, storks are divided into four groups: 1) *Jabiru* and *Ephippiorhynchus*, 2) Mycteriini, 3) Ciconiini, and 4) Leptoptilos. Groups 1 and 2 form a clade sister to a clade consisting of groups 3 and 4.

One member of tribe Mycteriini, the wood stork (*Mycteria americana*), is a species of conservation concern in the United States. The wood stork's range includes the southeastern United States, Mexico, Central America, Cuba, and South America. In 1984, the U.S. government listed the wood stork as an endangered species ("Endangered and Threatened Wildlife and Plants; U.S. Breeding Population of the Wood Stork Determined to be Endangered; Final Rule," February 28, 1984) due to the

loss of suitable feeding habitat in southern Florida, the historical stronghold of the U.S. wood stork population (Ogden & Patty, 1981). Northward range expansion and a concomitant increase in stork numbers in the succeeding decades motivated downlisting of the species in the U.S. from endangered to threatened status ("Endangered and Threatened Wildlife and Plants; Reclassification of the U.S. Breeding Population of the Wood Stork from Endangered to Threatened; Final Rule," June 30, 2014). It has been recently proposed to delist the wood stork completely from the Endangered Species Act due to recovery, including the perception of sufficient numbers and productivity to guarantee long-term viability of the U.S. wood stork population. However, the adaptative potential for the species remains unclear amidst climate change related threats including changes in seasonal rainfall patterns, warming temperatures, and sea level rise ("Endangered and Threatened Wildlife and Plants; Removal of the Southeast U.S. Distinct Population Segment of the Wood Stork From the the List of Endangered and Threatened Wildlife ", February 15, 2023).

There are currently few genomic resources for storks including only one stork chromosome-level assembly (the maguari stork (*Ciconia maguari*); NCBI BioProject PRJDB4709). The objective of this study is to build an annotated chromosome-level genome for the wood stork that will provide a detailed map of what genes are present on each chromosome and serve as a resource for conservation and evolutionary studies. In this paper, we additionally test for genome-level synapomorphies unique to storks to improve our understanding of genome evolution in birds.

Methods

Biological Materials

The Jacksonville Zoo and Aquarium in northern Florida contains a wood stork rookery that was naturally established in 1999 (Bear-Hull et al., 2005). In May 2021, fresh blood samples from ten of the colony's chicks were collected in tubes pre-coated with the anticoagulant EDTA and stored at -80˚C. DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen., Valencia, CA, USA) following the manufacturer's protocol. Birds were sexed genetically according to Griffiths et al. (1998) and Lee et al. (2010) to identify a female individual for genomic sequencing. In birds, female is the heterogametic sex (ZW).

Nucleic Acid Library Preparation

Following genetic sexing, a blood sample from a single female wood stork was sent to the commercial provider Cantata Bio (Scotts Valley, CA, USA) for nucleic acid library preparation. Two genomic libraries were produced: 1) a PacBio high-fidelity (HiFi) library (~20 kb) for long read sequencing, and 2) a Dovetail Omni-C library for short read sequencing and continuity ligation. Additionally, an RNA-Seq library was produced for genome annotation.

For HiFi library preparation, high-quality double stranded DNA was extracted from stork blood and purified using the Blood & Cell Culture DNA Mini Kit (Qiagen). Following purification, DNA was quantified using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and the Qubit dsDNA Broad Range Assay Kit (Thermo Fisher Scientific). The library was prepared using the SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) following the manufacturer's

protocol. The library was bound to DNA polymerase using the Sequel II Binding Kit 2.0 (PacBio).

Dovetail Omni-C library preparation followed methods described in Putnam et al. (2016). Briefly, chromatin was cross-linked using formaldehyde and extracted. Cross-linked chromatin was subsequently fragmented using DNAse I, a sequence-independent endonuclease. The ends of the chromatin fragments were blunted and tagged with biotin, followed by proximity ligation to create chimeric molecules. Crosslinks were reversed and DNA was purified from protein. Next, DNA was treated to remove biotin that was not internalized within ligated fragments and sheared to ~350 bp mean fragment size. Sequencing libraries were generated using NEBNext Ultra enzymes (New England Biolabs, Ipswich, MA, USA) and Illumina-compatible adapters. Streptavidin beads were used to isolate biotin-containing fragments, which were subsequently amplified using polymerase chain reaction (PCR).

Extraction of total RNA was performed using the RNeasy Plus Mini Kit (Qiagen) following the manufacturer's protocol. After extraction, RNA was quantified using: 1) the Qubit 2.0 Fluorometer (Thermo Fisher Scientific) with the Qubit RNA Broad Range Assay Kit (Thermo Fisher Scientific), and 2) the 4200 TapeStation system (Agilent, Santa Clara, CA, USA). DNase treatment, AMPure bead cleanup (Beckman Coulter Life Sciences, Indianapolis, IN, USA), and Qiagen FastSelect HMR rRNA (Qiagen) depletion were performed prior to library preparation. Library preparation was performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) following the manufacturer's protocol.

Sequencing and Genome Assembly

Genome Assembly

Genomic assembly and annotation were executed by Cantata Bio. A list of all programs and versions used throughout the assembly process is available in Table 1.

The HiFi library was loaded onto a Sequel II 8M SMRT cell (PacBio) using the MagBindKit v2 (PacBio) and sequenced to 67x coverage using circular consensus sequencing (CCS) mode (Wenger et al., 2019). A draft genome assembly was built from the subsequent reads using default parameters in Hifiasm v0.15.4 (Cheng et al., 2021). The Hifiasm output assembly (hifiasm.p_ctg.fa) was then compared to the BLAST (Basic Local Alignment Search Tool) v2.9.0 (Altschul et al., 1990) nucleotide database (nt). The resulting file was used as input for BlobTools v1.1.1 (Laetsch & Blaxter, 2017), and scaffolds identified as possible contamination were removed from the assembly (filtered.asm.cns.fa). Finally, Purge_dups v1.2.5 (Guan et al., 2020) was used to remove haplotigs and contig overlaps (purged.fa).

The Omni-C library was sequenced on a HiSeqX platform (Illumina, San Diego, CA, USA) to ~30x coverage using 2 x 150 bp paired-end reads. The input *de novo* assembly and the Omni-C library reads were used as input data for HiRise v2.1.1 (Putnam et al., 2016), a software pipeline designed to scaffold genome assemblies using proximity ligation data. Briefly, the Omni-C reads were first mapped to the Hifiasm assembly using BWA v0.7.17 (Li & Durbin, 2009). Only reads with mapping quality scores ≥ 50 were retained. Then, the separations of Omni-C read pairs mapped within draft scaffolds were used by HiRise to produce a likelihood model that identified and broke putative misjoins, scored prospective joins, and made novel joins.

Assembly Validation

The qualities of the initial Hifiasm assembly and the scaffolded HiRise assembly were assessed for genome completeness using the program BUSCO v4.05 (Manni et al., 2021). BUSCO uses universal single-copy orthologs; for this project, the eukaryota_odb10 database, which includes 70 species and 255 single-copy orthologous genes, was used.

Genome Annotation

First, repetitive regions (e.g., transposable elements) within the genome were identified *de novo* using the pipeline RepeatModeler v2.0.1 (Flynn et al., 2020). This pipeline used two distinct discovery algorithms to accomplish this task: RECON v1.08 (Bao & Eddy, 2002) and RepeatScout v1.0.6 (Price et al., 2005). The repeat library produced from the pipeline was input into the program RepeatMasker v4.1.0 (Smit et al., 2013-2015) which annotated and masked the repeats in the assembly file.

The RNA-Seq library was run on the NovaSeq6000 platform (Illumina) in 2 x 150 bp configuration. RNA-Seq reads were mapped onto the genome using the RNA-Seq aligner STAR v2.7 (Dobin et al., 2013). The bam2hints tool within Augustus v2.5.5 (Stanke & Waack, 2003) was used to generate intron-exon boundary hints. Coding sequences from four avian species, the little egret (*Egretta garzetta*), the crested ibis (*Nipponia nippon*), the chicken (*Gallus gallus*), and the zebra finch (*Taeniopygia guttata*), were used to train initial *ab initio* models for the gene prediction programs SNAP v2006-07-28 (Korf, 2004) and Augustus; for Augustus, this included six rounds of

prediction optimization. Following model training, gene prediction was performed in the repeat-masked assembly file using SNAP and Augustus. Gene prediction was also performed in the repeat-masked assembly file using the annotation pipeline MAKER v3.01.03 (Cantarel et al., 2008). UnitProKB/Swiss-Prot peptide sequences from the UniProt Knowledgebase (http://www.uniprot.org) and the protein sequences from the four avian species above (i.e., little egret, crested ibis, chicken, and zebra finch) were used when running MAKER. Annotation edit distance (AED) scores for each of the predicted genes were generated by MAKER to assess gene prediction quality.

The final genome annotation contained the intersection of the genes predicted by SNAP and Augustus. Genes were further characterized for their putative function by performing a BLAST search of the peptide sequences against the UniProt Knowledgebase. tRNAs were predicted using tRNAscan-SE v2.05 (Chan et al., 2021).

Determination of chromosome-level scaffolds including sex verification

After scaffolding with HiRise, a genomic contact matrix was produced to visualize chromosomal-level scaffolds. First, the command parse in the Pairtools v1.0.2 (Open2C et al. 2023) pipeline was used to identify valid ligation events present in the Omni-C data. Then, the pairs were sorted and PCR duplicates were removed using the Pairtools commands sort and dedup, respectively. Pairtools split was used to produce a .pairs file, and the .pairs file was indexed with Pairix v0.3.7 (Lee et al., 2022). A single resolution cool file was generated using the command cload pairix in Cooler v0.8.11 (Abdennur & Mirny, 2020). Subsequently, a multi-resolution mcool file was generated

using the command zoomify in Cooler. The genomic contact matrix, in the form of the mcool file, was visualized in the software HiGlass v1.11.7 (Kerpedjiev et al., 2018).

Six different avian genomes with chromosome-level assemblies were downloaded from the GenBank database (Benson et al., 2013): 1) chicken (WGS master accession JAENSK000000000; BioProject PRJNA660757), 2) zebra finch (WGS master accession RRCB00000000; BioProject PRJNA489098), 3) common cuckoo (*Cuculus canorus*; WGS master accession JAGIYT000000000; BioProject PRJNA562015), 4) Humboldt penguin (*Spheniscus humboldti*; WGS master accession JAPZLJ000000000; BioProject PRJNA838343), 5) plumbeous ibis (*Theristicus caerulescens*; WGS master accession JAJGSR000000000; BioProject PRJNA774297), and 6) maguari stork (WGS master accession JAGFVN000000000; BioProject PRJNA715733) for chromosome orthology analysis. To assess reciprocity in orthology, each of these genomes were aligned with the wood stork genome twice using the program D-GENIES v1.4.0 (Cabanettes & Klopp, 2018). One alignment used the wood stork genome as the query and the other genome as the target and the other used the wood stork genome as the target and the other genome as the query. The program utilized Minimap2 v2.24 (Li, 2018), with the option for few repeats, for genome alignment and then generated dot-plots. Additionally, D-GENIES produced an association table for each comparison that included the best matching chromosome in the target for each scaffold in the query (or vice-versa) as well as PAF (Pairwise mApping Format) files that consisted of alignments between sequences.

The package pafr v0.0.2 (Winter et al., 2020) in R v4.2.1 (R Core Team, 2022) was used in conjunction with the PAF files to visualize and confirm orthology between

wood stork chromosome scaffolds and chromosomes of chicken, finch, cuckoo, penguin, ibis, and maguari stork (see R Markdown). Briefly, we first visualized the coverage of the wood stork sex chromosome scaffolds (wood stork Z chromosome = MAMZ and wood stork W chromosome = MAMW) with the sex chromosomes of chicken (GGAZ and GGAW) and maguari stork (CMAZ and CMAW). Coverage plots between wood stork chromosomes and chicken and/or penguin chromosomes were produced when dot plots and association tables indicated non 1:1 chromosome orthology between species. Next, we used pafr to confirm orthology between microchromosomes MAM25-29 and their complements in Humboldt penguin, maguari stork, and chicken. Finally, we used the leftover microchromosomes in chicken, zebra finch, common cuckoo, and Humboldt penguin that were not orthologous to MAM1-29 to identify if any other scaffolds in the wood stork genome were chromosomes.

We compared the visualized alignment of the scaffold pertaining to the wood stork mitogenome with that of the mitogenomes of the other species. Based on an unusually large sequence length (33,032 bp) and incongruence with the other mitogenomes, the wood stork mitochondrion scaffold was trimmed using Geneious Prime v.2023.1.2 (https://www,geneious.com). First, the scaffold was aligned to the white stork (*Ciconia ciconia*; GenBank accession NC_002197) mitogenome using a Geneious alignment (parameters included a global alignment with free end gaps and a cost matrix of 70% similarity) and trimmed. Geneious alignments between the trimmed sequence and chicken (CM028585.1), oriental stork (*Ciconia boyciana*; NC_002196.1), and black stork (*Ciconia nigra*; KF906246.1) mitogenomes available on GenBank were performed to confirm accurate trimming.

To verify using bioinformatics that our stork was a female, we mapped the sequencing reads back to the assembled wood stork genome using BWA-MEM2 v2.2.1 (Vasimuddin et al., 2019). We then used the function coverage in SAMtools v1.16.1 (Danecek et al., 2021) to determine coverage or read depth of each scaffold. SAMtools coverage also provided the mapping quality values for the wood stork reads mapped onto the wood stork reference; this reinforced confidence in which scaffolds could be assigned as chromosome-level scaffolds.

Results

Sequencing and Genome Assembly

Genome Assembly and Assembly Validation

We obtained ~4.7 million (67 gigabase-pairs (Gbp)) PacBio CCS reads, which resulted in 67x coverage for the initial de novo genome assembly using Hifiasm. The initial *de novo* genome assembly using Hifiasm had a total length of 1.35 Gbp across 359 contigs (342 scaffolds) and a contig (and scaffold) N50 of 36,845,572 base-pairs (bp) after primary filtering. The initial assembly had a longest contig length of 131,390,114 bp and 17 gaps. After scaffolding this assembly with HiRise using the Dovetail Omni-C library, the final assembly retained a total length of ~1.35 Gbp. The final assembly contained 280 scaffolds; the number of contigs in the final assembly remained at 359 because the scaffolding process does not change the number (or length) of contigs. The scaffold N50 for the final assembly was 80,020,930 bp. The final assembly contained 70 gaps and a BUSCO score of 98.8%.

Genome Annotation

Annotation predicted 28,238 genes in the assembly, and these genes accounted for 38.87 Mb or 2.88% of the length of the final assembly. The average gene length was 1.38 kb, and there were 2,731 single-exon genes identified. At least 50% of predicted genes had AED scores <2.5 and at least 80% of predicted genes has AED scores <0.5, signifying that most predicted genes in the annotation were well supported by external evidence (Figure S1). Out of 255 BUSCO genes searched, BUSCO analysis of predicted genes identified 226 (88.6%) complete single-copy BUSCOs and 18 (7.1%) fragmented BUSCOs. Eleven (4.3%) BUSCOs were missing. In terms of repeats masked in the genome, 15.4% of the total genome was masked. Within the genome, 5.9% were Class I TEs repeats, 0.1% were Class II TEs repeats, 0.2% were low complexity repeats, and 0.9% were simple repeats.

Determination of chromosomal-level scaffolds including sex verification

Visualization with HiGlass of the largest 53 scaffolds in the wood stork reference assembly identified 26 well-defined scaffolds (Figure 1A). Based on orthology analyses, these 26 scaffolds corresponded to wood stork chromosomes 1-24 (MAM1-24) and two MAMZ scaffolds (a major MAMZ scaffold that accounted for most of the chromosome and a minor MAMZb that was at a distal end of the Z chromosome; Table S1 and Figure S2). After the 26th scaffold, scaffolds became less defined, with some scaffolds having little intra-scaffold contact (Figure 1B). Synteny plots with maguari stork, Humboldt penguin, and chicken confirmed five scaffolds after scaffold 26 in the contact matrix were most likely wood stork chromosomes 25-29 (MAM25-29; Figure 2). At least 50% of

the orthologous maguari stork chromosome had synteny with the wood stork scaffold and at least 60% of the wood stork scaffold was aligned to the orthologous penguin microchromosome. One exception was MAM27, which was orthologous to maguari stork chromosome 29 (CMA29) but showed little orthology to chicken or penguin microchromosomes. Two scaffolds pertaining to MAMW (defined MAMWa and MAMWb) and an additional MAMZ scaffold (defined MAMZc) were also present after scaffold 26 (Figure S2). The two MAMW scaffolds covered about a quarter of maguari stork chromosome W (CMAW) but did not have significant alignments with chicken chromosome W (GGAW; Figure S2). MAMZc aligned to distal portions of both chicken chromosome Z (GGAZ) and maguari stork chromosome Z (CMAZ; Figure S2).

Two additional microchromosomes were defined, wood stork chromosomes 30 and 31 (MAM30 and MAM31), based on synteny plotting between wood stork scaffolds and additional microchromosomes in the Humboldt penguin genome (Figure 2). These two microchromosomes were not previously identified in the maguari stork genome assembly, but unique unplaced scaffolds in the maguari stork assembly were identified that had at least 60% synteny with these microchromosomes. Additional scaffolds in the wood stork genome were suspected of being orthologous to additional penguin microchromosomes due to substantial sequence alignments (>30% of wood stork scaffold aligned to a penguin microchromosome). However, there were no substantial sequence alignments between these wood stork scaffolds and maguari stork unplaced scaffolds (see R Markdown).

Chromosome orthology analyses based on dot plots and association tables revealed fissions and fusions of chromosomes in the wood stork and the other avian

345 species studied (Table S1 and Figure S3). Wood stork chromosomes 4 and 10 (MAM4
346 and MAM10) were orthologous to different sections of chicken chromosome 4 (GGA4;
347 Figure S4), but orthologous to separate chromosomes in the other avian species (Table
348 S1). Several wood stork chromosomes were fusions of two chromosomes in chicken: a)
349 MAM6 was a fusion of GGA6 and GGA10, b) MAM7 was a fusion of GGA8 and GGA9,
350 and c) MAM8 was a fusion of GGA11 and GGA13 (Figure 3; Figure S5; Table S1).
351 These fusions were shared with maguari stork, but not with zebra finch, common
352 cuckoo, Humboldt penguin, or plumbeous ibis (Table S1).
353 Orthology analyses additionally identified fissions and fusions of chromosomes in
354 chicken and the other avian species. Chicken chromosome 1 (GGA1) was orthologous
355 to two chromosomes in zebra finch (TGU1 and TGU1A, data not shown) and two
356 chromosomes in plumbeous ibis (TCA2 and a portion of TCA3; Table S1; Figure 3c).
357 Also of note were the fusions observed within the clade Feraequornithes (Figure 3).
358 These fusions were not shared between Humboldt penguin, plumbeous ibis, and storks,
359 but did involve complements of the same chicken chromosomes 6-14 (GGA6-14). In
360 Humboldt penguin, chromosome 5 (SHU5) was a fusion of GGA6 and GGA8 and
361 chromosome 6 (SHU6) was a fusion of GGA7 and GGA9 (Figure 3a). In wood stork
362 (and maguari stork), as previously described, MAM6 was a fusion of GGA6 and GGA10,
363 MAM7 was a fusion of GGA8 and GGA9, and MAM8 was a fusion of GGA11 and
364 GGA13 (Figure 3b). Substantial chromosomal rearrangements occurred in the
365 plumbeous ibis genome (Figure 3c). Plumbeous ibis chromosome 5 (TCA5) was a
366 fusion of GGA7, GGA8, and GGA14, chromosome 8 was a fusion of GGA9 and
367 GGA11, and chromosome 9 was a fusion of GGA10 and GGA12. It appears that GGA1

experienced a fission event in plumbeous ibis, in which part of the chromosome became plumbeous ibis chromosome 2 (TCA2) and the other part of the chromosome fused with GGA6 to become plumbeous ibis chromosome 3 (TCA3).

Originally, it appeared that the scaffold pertaining to the wood stork mitochondrion sequence was the composite of duplicated contigs (Figure S6). After alignment and trimming, the final wood stork mitochondrion sequence was 17,347 bp in length and appeared orthologous to chicken, white stork, black stork, and oriental stork mitogenome sequences.

The mean coverage of wood stork autosomes 1-24 (MAM1-24) was 31.39. Coverage of the major Z scaffold (MAMZ) was approximately half that of the autosomes (17.63) validating that the individual sequenced was the heterogametic sex (female).

Discussion

Despite new technologies and diminishing costs facilitating the production of more accurate and complete genomic resources for non-model organisms, challenges remain in assembling genomes with complex architecture. Unlike human and other mammalian genomes, most avian genomes, many reptile genomes, and some fish and amphibian genomes include microchromosomes, or chromosomes < 0.5 μm in size, in addition to macrochromosomes (Srikulnath et al., 2021; Waters et al., 2021). These microchromosomes are often gene-rich, have little repetitive sequence, and comprise up to a third of the total genome content in avian genomes. Recent genome assemblies for chicken characterize all chromosomes including microchromosomes (Masabanda et al., 2004), but many avian genomes are being deposited into public repositories (e.g.,

391 GenBank) with discrepancies between karyotype number from cytological studies and
392 the number of chromosome sets identified in genome assemblies. This discordance is
393 likely due to the difficulty of distinguishing some microchromsomes from unplaced
394 scaffolds and thus results in the number of chromosomes in the final genomic
395 assemblies being underestimated.

396 We produced a highly contiguous genome assembly of the wood stork and were
397 able to identify 31 autosomes in the wood stork genome. Francisco and Galetti Junior
398 (2000), and more recently de Sousa et al. (2023), determined 2n = 72 for wood stork
399 based on karyotype analysis. Based on this diploid number, we would expect 35
400 autosomes in the wood stork genome and thus our assembly and subsequent analysis
401 has either failed to assemble or failed to identify four microchromosomes. This is not
402 surprising as microchromosomes are small and can be hard to both identify and
403 assemble. Future studies aimed at producing complete sequences for the wood stork
404 genome may utilize alternative techniques such as isolating and sequencing individual
405 microchromosomes.

406 Chromosome-level assemblies of wood stork and maguari stork identified
407 chromosome synapomorphies that are most likely unique to Order Ciconiiformes.
408 Previous research partially characterized these synapomorphies using chicken probes
409 for GGA1-9 on the jabiru and the maguari stork (Seligmann et al., 2019) and chicken
410 probes for GGA1-11 on the wood stork (de Sousa et al., 2023). Our data support the
411 identified GGA8/GGA9 fusion in storks, and our comparisons with other members of the
412 clade Feraequornithes (Humboldt penguin and plumbeous ibis) provide additional
413 support that this fusion is unique to storks. Additionally, our study identified that GGA10

414 was the unidentified chromosome in Seligmann et al. (2019) that was fused with GGA6
415 in storks. This is in partial disagreement with de Sousa et al. 2023, who stated that
416 GGA6 was fused with an unidentified chromosome in wood stork, but characterized
417 GGA10 as an independent chromosome in the wood stork karyotype.
418 The two stork fusions (GGA8/GGA9) and (GGA6/GGA10) were not found in a
419 chromosome painting study of three members of Pelecaniformes, the grey heron (*Ardea*
420 *cinerea*), the little egret (*Egretta garzetta*), and crested ibis (*Nipponia nippon*),
421 suggesting these fusions may also be stork specific (Wang et al., 2022). The inclusion
422 of additional taxa from within the clade Feraequornithes in studies of karyotype
423 evolution will be informative in determining if these chromosomal fusions are stork
424 specific.
425 Our study also identified the fusion, GGA11/GGA13, that was present in both
426 stork species but none of the other species tested, including Humboldt penguin and
427 plumbeous ibis. The fusion GGA11/GGA13 has also been identified in the neotropic
428 cormorant (*Nannopterum brasilianum*) (Kretschmer et al., 2021) and was partially
429 characterized for the wood stork by de Sousa et al. 2023 who found GGA11 fused to an
430 unidentified chromosome. Further research needs to be done to determine if the
431 GGA11/GGA13 fusion in some members of Feraequornithes occurred independently in
432 separate lineages or was the product of the same event.
433 The reference genome described here will allow for broader studies of genome
434 evolution in the core water bird clade (including penguins, albatrosses, storks,
435 cormorants, herons, and ibises) and numerous genomic studies of wood storks
436 specifically. In particular, the genomic population structure of wood storks has not been

adequately tested and evidence of local adaptation of wood storks to nesting and feeding sites has yet to be assessed. Such studies will provide a solid foundation for wood stork management and conservation.

Funding

This material is based upon work supported by the National Science Foundation under Grant No. 2129600.

Acknowledgements

The authors thank Donna Bear (Jacksonville Zoo and Gardens) for supplying wood stork blood samples, which were collected under US Fish and Wildlife Service migratory bird permit TE08606C, US Geological Survey banding permit 23946, and IACUC approval USCA-IACUC-006. We also thank Jordan Zhang and Mark Daly (Cantata Bio) for genome assembly, annotation, answering our questions about assembly statistics, and connecting us with other avian genome researchers. Nicolas Alexandre (University of California Berkeley) provided advice on which programs to use to analyze the genome post-assembly and Linelle Abueg (Rockefeller University) provided us with genomes from the Vertebrate Genomes Project and answered our questions regarding VGP assembly. Finally, we thank Philippe Bordron (Centre INRA de Toulouse) for providing help with the D-Genies program.

Data Availability

We have deposited the primary data and code underlying these analyses as follows:

- Description of the biological source material deposited at GenBank under BioSample SAMN36274196.
- Annotated chromosome-level assembly deposited at GenBank under accession JAUNZN000000000 in BioProject PRJNA990753.
- Raw DNA and RNA sequence reads were deposited under BioProject PRJNA990753 in NIH's Sequence Read Archive (SRA).
- Pairwise mApping Format (PAF) files and chromosome orthology code can be found in https://github.com/rflamio/stork_chrs.

References

Abdennur, N., & Mirny, L. A. (2020). Cooler: Scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics*, *36*(1), 311-316. https://doi.org/10.1093/bioinformatics/btz540

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bao, Z., & Eddy, S. R. (2002). Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Research*, *12*(8), 1269-1276. https://doi.org/10.1101/gr.88502

Bear-Hull, D., Brooks, W., Jr., & Maharaj, A. M. (2005, October 12-15). *Tracking movements of an adult male wood stork Mycteria americana during the breeding season at the Jacksonville Zoo and Gardens* [Conference session]. 29th Waterbird Society Meeting, Jekyll Island, GA, United States.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *41*(D1), D36-42. https://doi.org/10.1093/nar/gks1195

Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life using a large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*, *84*, 53-63. https://doi.org/10.1016/j.ympev.2014.12.003

Cabanettes, F., & Klopp, C. (2018). D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, *6*, Article e4958. https://doi.org/10.7717/peerj.4958

Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., & Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, *18*(1), 188-196. https://doi.org/10.1101/gr.6743907

Chan, P. P., Lin, B. Y., Mak, A. J., & Lowe, T. M. (2021). tRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research*, *49*(16), 9077-9096. https://doi.org/10.1093/nar/gkab688

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods*, *18*(2), 170-175. https://doi.org/10.1038/s41592-020-01056-5

International Chicken Genome Sequencing Consortium (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, *432*(7018), 695-716. https://doi.org/10.1038/nature03154

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2), 1-4. https://doi.org/10.1093/gigascience/giab008

de Boer, L. E. M., & van Brink, J. M. (1982). Cytotaxonomy of the Ciconiiformes (Aves), with karyotypes of eight species new to cytology. *Cytogenetics and Cell Genetics*, *34*(1-2), 19-34. https://doi.org/10.1159/000131791

de Sousa, R. P. C., Campos, P. S. B., dos Santos, M. d. S., O'Brien, P. C., Ferguson-Smith, M. A., & de Oliveira, E. H. C. (2023). Cytotaxonomy and molecular analyses of *Mycteria americana* (Ciconiidae: Ciconiiformes): Insights on stork phylogeny. *Genes*, *14*(4), Article 816. https://doi.org/10.3390/genes14040816

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-Seq aligner. *Bioinformatics*, *29*(1), 15-21. https://doi.org/10.1093/bioinformatics/bts635

Endangered and Threatened Wildlife and Plants; Reclassification of the U.S. Breeding Population of the Wood Stork from Endangered to Threatened; Final Rule, 79 Fed. Reg. 37078-37103 (June 30, 2014).

Endangered and Threatened Wildlife and Plants; Removal of the Southeast U.S. Distinct Population Segment of the Wood Stork From the the List of Endangered and Threatened Wildlife, 88 Fed. Reg. 9830-9850 (February 15, 2023).

Endangered and Threatened Wildlife and Plants; U.S. Breeding Population of the Wood Stork Determined to be Endangered; Final Rule, 49 Fed. Reg. 7332-7335 (February 28, 1984).

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451-9457. https://doi.org/10.1073/pnas.1921046117

Griffin, D. K., Robertson, L. B. W., Tempest, H. G., & Skinner, B. M. (2007). The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenetic and Genome Research*, *117*, 64-77. https://doi.org/10.1159/000103166

Griffiths, R., Double, M. C., Orr, K., & Dawson, R. J. G. (1998). A DNA test to sex most birds. *Molecular Ecology*, *7*(8), 1071-1075. https://doi.org/10.1046/j.1365-294x.1998.00389.x

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

Guttenbach, M., Nanda, I., Feichtinger, W., Masabanda, J. S., Griffin, D. K., & Schmid, M. (2003). Comparative chromosome painting of chicken autosomal paints 1–9 in nine different bird species. *Cytogenetic and Genome Research*, *103*(1-2), 173-184. https://doi.org/10.1159/000076309

Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C. K., Braun, E. L., Braun, M. J., Chojnowski, J. L., Cox, W. A., Han, K.-L., Harshman, J., Huddleston, C. J., Marks, B. D., Miglia, K. J., Moore, W. S., Sheldon, F. H., Steadman, D. W., Witt, C. C., & Yuri, T. (2008). A phylogenomic study of birds reveals their evolutionary history. *Science*, *320*(5884), 1763-1768. https://doi.org/10.1126/science.1157704

Hughes, A. L., & Piontkivska, H. (2005). DNA repeat arrays in chicken and human genomes and the adaptive evolution of avian genome size. *BMC Evolutionary Biology*, *5*, Article 12. https://doi.org/10.1186/1471-2148-5-12

Itoh, Y., & Arnold, A. P. (2005). Chromosomal polymorphism and comparative painting analysis in the zebra finch. *Chromosome Research*, *13*(2), 47-56. https://doi.org/10.1007/s10577-005-6602-x

Kahl, M. P. (1987). An overview of the storks of the world. *Colonial Waterbirds*, *10*(2), 131-134. https://doi.org/10.2307/1521251

583 Kerpedjiev, P., Abdennur, N., Lekschas, F., McCallum, C., Dinkla, K., Strobelt, H., Luber, J. M.,
584 Ouellette, S. B., Azhir, A., Kumar, N., Hwang, J., Lee, S., Alver, B. H., Pfister, H., Mirny, L.
585 A., Park, P. J., & Gehlenborg, N. (2018). HiGlass: Web-based visual exploration and
586 analysis of genome interaction maps. *Genome Biology*, *19*(1), Article 125.
587 https://doi.org/10.1186/s13059-018-1486-1
588 Kimball, R. T., Oliveros, C. H., Wang, N., White, N. D., Barker, F. K., Field, D. J., Ksepka, D. T.,
589 Chesser, R. T., Moyle, R. G., Braun, M. J., Brumfield, R. T., Faircloth, B. C., Smith, B. T., &
590 Braun, E. L. (2019). A phylogenomic supertree of birds. *Diversity*, *11*(7), Article 109.
591 https://doi.org/10.3390/d11070109
592 Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*, Article 59.
593 http://www.biomedcentral.com/1471-2105/5/59
594 Kretschmer, R., de Souza, M. S., Furo, I. O., Romanov, M. N., Gunski, R. J., Garnero, A. D. V., de
595 Freitas, T. R. O., de Oliveira, E. H. C., O'Connor, R. E., & Griffin, D. K. (2021). Interspecies
596 chromosome mapping in Caprimulgiformes, Piciformes, Suliformes, and Trogoniformes
597 (Aves): Cytogenomic insight into microchromosome organization and karyotype
598 evolution in birds. *Cells*, *10*(4). Article 826. https://doi.org/10.3390/cells10040826
599 Kuhl, H., Frankl-Vilches, C., Bakker, A., Mayr, G., Nikolaus, G., Boerno, S. T., Klages, S.,
600 Timmermann, B., & Gahr, M. (2021). An unbiased molecular approach using 3'-UTRs
601 resolves the avian family-level tree of life. *Molecular Biology and Evolution*, *38*(1), 108-
602 127. https://doi.org/10.1093/molbev/msaa191
603 Kuramoto, T., Nishihara, H., Watanabe, M., & Okada, N. (2015). Determining the position of
604 storks on the phylogenetic tree of waterbirds by retroposon insertion analysis. *Genome*
605 *Biology and Evolution*, *7*(12), 3180-3189. https://doi.org/10.1093/gbe/evv213
606 Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies [version
607 1; peer review: 2 approved with reservations]. *F1000Research*, *6*, Article 1287.
608 https://doi.org/10.12688/f1000research.12232.1
609 Lee, J. C., Tsai, L. C., Hwa, P. Y., Chan, C. L., Huang, A., Chin, S. C., Wang, L. C., Lin, J. T., Linacre,
610 A., & Hsieh, H. M. (2010). A novel strategy for avian species and gender identification
611 using the CHD gene. *Molecular and Cellular Probes*, *24*(1), 27-31.
612 https://doi.org/10.1016/j.mcp.2009.08.003
613 Lee, S., Bakker, C. R., Vitzthum, C., Alver, B. H., & Park, P. J. (2022). Pairs and Pairix: A file format
614 and a tool for efficient storage and retrieval for Hi-C read pairs. *Bioinformatics*, *38*(6),
615 1729-1731. https://doi.org/10.1093/bioinformatics/btab870
616 Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18),
617 3094-3100. https://doi.org/10.1093/bioinformatics/bty191
618 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
619 transform. *Bioinformatics*, *25*(14), 1754-1760.
620 https://doi.org/10.1093/bioinformatics/btp324
621 Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A., & Zdobnov, E. M. (2021). BUSCO update:
622 Novel and streamlined workflows along with broader and deeper phylogenetic coverage
623 for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and*
624 *Evolution*, *38*(10), 4647-4654. https://doi.org/10.1093/molbev/msab199
625 Masabanda, J. S., Burt, D. W., O'Brien, P. C. M., Vignal, A., Fillon, V., Walsh, P. S., Cox, H.,
626 Tempest, H. G., Smith, J., Habermann, F., Schmid, M., Matsuda, Y., Ferguson-Smith, M.

A., Croojimans, P. M. A., Groenen, M. A. M., & Griffin, D. K. (2004). Molecular cytogenetic definition of the chicken genome: The first complete avian karyotype. *Genetics*, *166*(3), 1367-1373. https://doi.org/10.1534/genetics.166.3.1367

Morescalchi, A. (1980). Evolution and karyology of the amphibians. *Italian Journal of Zoology*, *47*(sup1), 113-126. https://doi.org/10.1080/11250008009438709

Nanda, I., Benisch, P., Fetting, D., Haaf, T., & Schmid, M. (2011). Synteny conservation of chicken macrochromosomes 1-10 in different avian lineages revealed by cross-species chromosome painting. *Cytogenetic and Genome Research*, *132*(3), 165-181. https://doi.org/10.1159/000322358

Ogden, J. C., & Patty, B. W. (1981). The recent status of the wood stork in Florida and Georgia. In R. Q. Odom & J. W. Guthrie (Eds.), *Proceedings of the nongame and endangered wildlife symposium* (Technical Bulletin WL5, pp. 97-102). Georgia Department of Natural Resources Game and Fish Division.

Ohno, S., Muramoto, J., Stenius, C., Christian, L., & Kittrell, W. A. (1969). Microchromosomes in holocephalian, chondrostean and holostean fishes. *Chromosoma*, *26*(1), 35-40. https://doi.org/10.1007/bf00319498

Olmo, E. (2008). Trends in the evolution of reptilian chromosomes. *Integrative and Comparative Biology*, *48*(4), 486-493. https://doi.org/10.1093/icb/icn049

Open2C, Abdennur, N., Fudenberg, G., Flyamer, I. M., Galitsyna, A. A., Goloborodko, A., Imakaev, M., & Venev, S. V. (2023). Pairtools: From sequencing data to chromosome contacts. Preprint at *bioRxiv*. https://doi.org/10.1101/2023.02.13.528389

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). *De novo* identification of repeat families in large genomes. *Bioinformatics*, *21*(Suppl 1), i351-i358. https://doi.org/10.1093/bioinformatics/bti1018

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, *526*(7574), 569-573. https://doi.org/10.1038/nature15697

Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., Troll, C. J., Fields, A., Hartley, P. D., Sugnet, C. W., Haussler, D., Rokhsar, D. S., & Green, R. E. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*, *26*(3), 342-350. https://doi.org/10.1101/gr.193474.115

R Core Team. (2022). *R: A language and environment for statistical computing* (Version 4.2.2). R Foundation for Statistical Computing. https://www.R-project.org/

Rodríguez-Rodríguez, E. J., & Negro, J. J. (2021). Integumentary colour allocation in the stork family (Ciconiidae) reveals short-range visual cues for species recognition. *Birds*, *2*(1), 138-146. https://doi.org/10.3390/birds2010010

Sangster, G., & Mayr, G. (2021). Feraequornithes: A name for the clade formed by Procellariiformes, Sphenisciformes, Ciconiiformes, Suliformes and Pelecaniformes (Aves). *Vertebrate Zoology*, *71*, 49-53. https://doi.org/10.3897/vz.71.e61728

Seligmann, I. C. A., Furo, I. O., dos Santos, M. S., Tagliarini, M. M., Araujo, C. C. D., O'Brien, P. C. M., Ferguson-Smith, M. A., & de Oliveira, E. H. C. (2019). Comparative chromosome painting in two Brazilian stork species with different diploid numbers. *Cytogenetic and Genome Research*, *159*(1), 32-38. https://doi.org/10.1159/000503019

Slikas, B. (1997). Phylogeny of the avian family Ciconiidae (storks) based on cytochrome b sequences and DNA-DNA hybridization distances. *Molecular Phylogenetics and Evolution*, *8*(3), 275-300. https://doi.org/10.1006/mpev.1997.0431

Smit, A. F. A., Hubley, R., & Green, P. (2013-2015). *RepeatMasker Open-4.0*. Institute for Systems Biology. http://www.repeatmasker.org

Srikulnath, K., Ahmad, S. F., Singchat, W., & Panthum, T. (2021). Why do some vertebrates have microchromosomes? *Cells*, *10*(9), Article 2182. https://doi.org/10.3390/cells10092182

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, *19*(Suppl 2), ii215-ii225. https://doi.org/10.1093/bioinformatics/btg1080

Tiersch, T. R., & Wachtel, S. S. (1991). On the evolution of genome size in birds. *Journal of Heredity*, *82*(5), 363-368. https://doi.org/10.1093/oxfordjournals.jhered.a111105

Vasimuddin, M., Misra, S., Li, H., & Aluru, S. (2019). Efficient architecture-aware acceleration of BWA-MEM for multicore systems. *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 314-324, https://doi:10.1109/IPDPS.2019.00041

Wang, J., Su, W., Hu, Y., Li, S., O'Brien, P. C. M., Ferguson-Smith, M. A., Yang, F., & Nie, W. (2022). Comparative chromosome maps between the stone curlew and three ciconiiform species (the grey heron, little egret and crested ibis). *BMC Ecology and Evolution*, *22*(1), Article 23. https://doi.org/10.1186/s12862-022-01979-x

Waters, P. D., Patel, H. R., Ruiz-Herrera, A., Álvarez-González, L., Lister, N. C., Simakov, O., Ezaz, T., Kaur, P., Frere, C., Grützner, F., Georges, A., & Graves, J. A. M. (2021). Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proceedings of the National Academy of Sciences*, *118*(45), Article e2112494118. https://doi.org/10.1073/pnas.2112494118

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N. D., Töpfer, A., Alonge, M., Mahmoud, M., Qian, Y., Chin, C.-S., Phillippy, A. M., Schatz, M. C., Myers, G., Depristo, M. A., . . . Hunkapiller, M. W. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, *37*(10), 1155-1162. https://doi.org/10.1038/s41587-019-0217-9

Winter, D., Lee, K., & Cox, M. (2020). *pafr: Read, manipulate and visualize 'Pairwise mApping Format' data*. R package version 0.0.2. https://CRAN.R-project.org/package=pafr