*Journal of Heredity*, 2025, **116**, 540–547
https://doi.org/10.1093/jhered/esaf019
Advance access publication 24 April 2025
**Genome Resources**



Genome Resources

# A reference genome for boat-tailed grackles (*Quiscalus major*)

**Eamon C. Corbett[1,2,*], Andre E. Moncrieff[2], Robb T. Brumfield[1,2] and Brant C. Faircloth[1,2]**

[1]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, United States
[2]Museum of Natural Science, Louisiana State University, Baton Rouge, LA, United States

[*]Corresponding author: Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, United States. Email: eamonccorbett@gmail.com

Corresponding Editor: William Murphy

## Abstract

Boat-tailed Grackles (*Quiscalus major*) are marsh-dwelling blackbirds that are endemic to the eastern United States. Various aspects of their biology have been studied extensively, including their mating system, plumage and molt patterns, diet, and interspecific interactions. Boat-tailed Grackles are also interesting because they exhibit variation in their iris color that is associated with geography. However, resources that enable genomic studies of Boat-tailed Grackles and other related grackle species are few. Here, we combined Pacific Biosciences long-read, HiFi data with short-read Illumina data from a HiC library to produce haplotype-phased, chromosome-scale genome assemblies for Boat-tailed Grackles. The final version of the assembly, bQuiMaj1, includes two, contiguous haplotypes with total lengths of ~1 Gbp, N50s of ~70 Mbp, and L50s of 5-6. BUSCO and merqury analyses suggest both haplotypes are also relatively complete (95-99%) with respect to gene and k-mer content. The resulting assemblies will significantly enhance our understanding of Boat-tailed Grackle biology and physiology, as well as contribute to the growing number of genomes representing species belonging to the taxonomic family Icteridae (the New World blackbirds).

**Key words:** Boat-tailed Grackle, Icteridae, reference genome, *Quiscalus major*

## Introduction

Boat-tailed Grackles (Icteridae: *Quiscalus major*) are marsh-dwelling blackbirds that are endemic to the eastern United States (Post et al. 2020). They occur along the Atlantic and Gulf Coasts from Connecticut to Texas, as well as throughout the Florida peninsula. Various elements of Boat-tailed Grackle biology have been studied extensively, including their general life history (Coues 1870; McIlhenny 1937), mating system (female-defense polygyny, unusual in North American birds) (Selander and Giller 1961; Bancroft 1987; Post 1995; Poston 1997), nesting biology (Bancroft 1986), development (Bancroft 1984; Clum 1991), plumage and molt (Selander 1958; Pratt 1974), vocalizations (Melman and Searcy 1999), diet (Snyder and Snyder 1969), interspecific interactions (Sprunt 1941; Jackson 1985; Post and Seals 1993), and distribution and conservation (Remsen et al. 2019; Summers et al. 2023). Boat-tailed Grackles are also remarkable because they exhibit pronounced geographic variation in iris coloration (Fig. 1) between the four described subspecies (Stevenson 1978), making them a good candidate system for examining the genetic basis of intraspecific eye color variation in birds (Corbett et al. 2024).

A high-quality genome assembly will enable genetic studies of Boat-tailed Grackle eye color variation, among other life history characteristics. Yet, genome assemblies for Icteridae are relatively few. Boat-tailed Grackles and Great-tailed Grackles (*Q. mexicanus*) have long been the focus of research regarding their genetic differences (Avise and Zink 1988), zones of secondary contact (Pratt 1974; Pratt et al. 1977; Wehtje 2003), phylogenetic relationships (DaCosta et al. 2008; Powell et al. 2008, 2014), and degree of reproductive isolation (Selander and Giller 1961; Pratt 1991), underscoring the need for a contiguous high-quality reference genome. However, the available assembly for Great-tailed Grackles (GCA_013399035.1) is relatively fragmented (scaffold N50 = 93.2 kb), and there is no genome assembly available for Boat-tailed Grackles.

Here, we announce a contiguous reference genome assembly constructed from tissues of a museum-vouchered (Buckner et al. 2021), female Boat-tailed Grackle of the nominate subspecies *Q. m. major* that we collected in southeast Louisiana.

## Methods

### Biological materials

We collected an adult female Boat-tailed Grackle representing the nominate subspecies (*Q. m. major*) on March 1, 2021 at Manchac Wildlife Management Area, St. John the Baptist Parish, Louisiana (30.27° N, 90.38° W). We immediately (<15 min) preserved pectoral muscle in liquid nitrogen and blood in ethanol. We later prepared a museum study skin, with associated data, and collected additional tissue samples

Received November 11, 2024; Accepted April 2, 2025

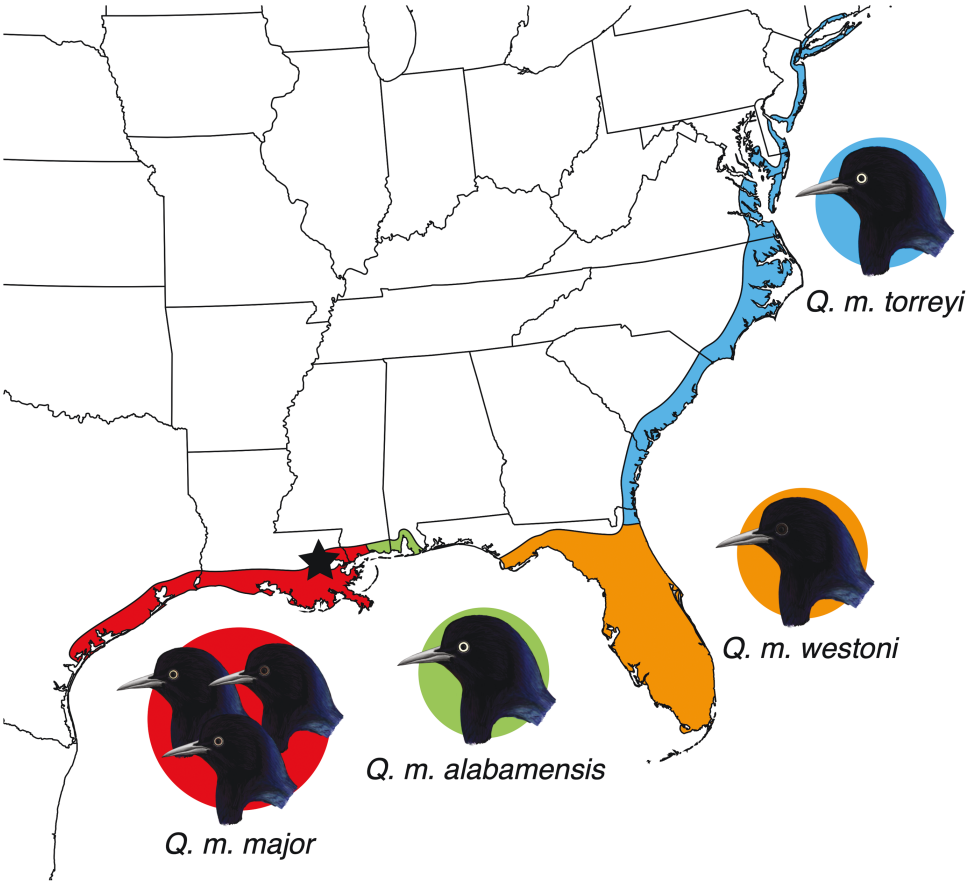


**Fig. 1.** Map of the distribution of Boat-tailed Grackle (*Quiscalus major*) subspecies (Stevenson 1978; Post et al. 2020), with typical eye colors shown following Stevenson (1978) and Pratt (1974): pale yellow in *Q. m. alabamensis* and *Q. m. torreyi*, dark brown in *Q. m. westoni*, and variable in the nominate *Q. m. major*. The locality of the reference genome individual is marked with a star.

(muscle, heart, liver, intestines, lungs, stomach contents, iris, and eye), and we added these to the Louisiana State University Museum of Natural Science ornithology collection (skin LSUMZ 230566, tissue B-96265).

Specimen collection was approved as part of LSU IACUC Protocol 21-042 and conducted under federal collecting permit number MB02467D-0, Louisiana state collecting permit number WDP-21-068, and WMA special use permit number WL-Research-2020-16.

## Nucleic acid library preparation and DNA sequencing

We extracted DNA from ~25 mg muscle tissue using a MagAttract High Molecular Weight (HMW) DNA Kit (Qiagen, Gmbh), and we shipped extracted DNA to the University of Maryland School of Medicine Institute for Genome Sciences (IGS) for library preparation and sequencing. IGS staff performed sample quality control before shearing the DNA extract, selected sheared fragments in the range of 10 to 20 kbp using a BluePippen (Sage Science, Inc.), and prepared a SMRTBell library for HiFi sequencing. The library was sequenced using two PacBio Sequel II 8M SMRT Cell runs to generate circular consensus (HiFi) reads.

After receiving the data from IGS, we used cutadapt (Table 1) (Martin 2011) to remove adapter contamination from sequence reads, and we generated a temporary assembly using hifiasm (Cheng et al. 2021, 2022). Then, we sent muscle tissue to Phase Genomics (PG), who prepared a HiC (Lieberman-Aiden et al. 2009; Van Berkum et al. 2010) library from the tissue using their Proximo kit v4.0. PG staff performed quality control (QC) of the HiC library by generating a small number of reads using an Illumina iSeq100 and aligning those reads to the temporary assembly. After successful QC, PG Staff sent the library to Azenta Life Sciences for deeper, paired-end, 150 bp sequencing using an Illumina NovaSeq 6000. After receiving the short-read sequence data from PG, we removed adapter contamination and trimmed low-quality reads using Trimmomatic (Bolger et al. 2014).

## Genome assembly and annotation

To estimate genome size, we generated a kmer histogram from the HiFi data using meryl (Rhie et al. 2020) and a kmer length of 21, and we input the histogram to GenomeScope (Ranallo-Benavidez et al. 2020). Then, we generated haplotype resolved assemblies by inputting the HiFi data and the HiC data to hifiasm in HiC partitioned mode. We converted the resulting haplotype 1 and 2 assemblies (hap1 and hap2 hereafter) to FASTA format, and we screened each haplotype for contaminants using the NCBI Foreign Contamination Screen (FCS) (Astashyn et al. 2024) to identify foreign and adapter contamination. We computed assembly statistics for each haplotype using gfastats (Formenti et al. 2022), and we used

**Table 1.** List of programs used to assemble and scaffold the *Quiscalus major* genome.

| Assembly | Software and options | Version |
|---|---|---|
| Long-read trimming | cutadapt | 4.3 |
| Temporary assembly | hifiasm | 0.15 |
| Short-read trimming | trimmomatic | 0.39 |
| Kmer histogram | meryl | 1.4 |
| Genome size estimate | GenomeScope | 2.0 |
| Haplotype assembly | hifiasm | 0.19.8 |
| Contamination screen | NCBI FCS | 0.5.0 |
| Assembly statistics | gfastats | 1.3.6 |
| Assembly completeness (BUSCO) | compleasm | 0.2.4 |
| Duplicate removal | purge_dupes | 1.2.6 |
| Mitogenome assembly | mitohifi | 3.2.1 |
| Mitogenome contamination removal | minimap | 2.24-r1122 |
| Mitogenome contamination removal | Python | 3.9.18 |
| Mitogenome contamination removal | BioPython | 1.79 |
| **Scaffolding** | | |
| HiC read alignment | Arima Mapping Pipeline | 02-08-2019 |
| HiC read alignment | bwa | 0.7.17-r1188 |
| HiC read alignment | samtools | 1.17 |
| HiC read alignment | Picard | 2.27.5 |
| Scaffolding | YaHS | 1.2a.1 |
| Z chromosome identification | minimap | 2.24-r1122 |
| **Annotation** | | |
| Repeat modeling | Dfam TETools Container | 1.88 |
| Repeat masking | Dfam TETools Container | 1.88 |
| Soft-masking scaffolds | Bedtools | 2.30.0 |
| Gene annotation | Braker3 container | 3.0.7 |
| GTF to GFF conversion | agat container | 1.0.0 |
| GFF sanitization | **GFFtk** | 0.1.7 |
| Functional annotation | interproscan | 5.66-98.0 |
| Functional annotation | eggNOG-mapper | 2.1.11 |
| Annotation integration | funannotate container | 1.8.15 |
| **Final assembly statistics** | | |
| Assembly statistics | gfastats | 1.3.6 |
| Assembly completeness (BUSCO) | compleasm | 0.2.4 |
| Assembly completeness (kmer) | merqury | 1.3 |

compleasm (Huang and Li 2023) to compute completeness statistics with the aves_odb10 database. After checking completeness, we removed duplicates from each haplotype assembly using purge_dupes (Guan 2022), and we computed another round of assembly and completeness statistics with gfastats and compleasm. We assembled a mitogenome for *Q. m. major* by inputting the HiFi data to the mitohifi (Uliano-Silva et al. 2023) container using a *Quiscalus mexicanus* mitogenome (NCBI GenBank MN356197.1) as the reference, and we used minimap (Li 2018) and custom Python (van Rossum 2009) and BioPython (Cock et al. 2009) code (see Supplemental Files) to identify and remove contigs from the haplotype assemblies that partially or completely overlapped the mitogenome. We aligned the trimmed, HiC data to each haplotype using bwa (Li and Durbin 2010), samtools (Danecek et al. 2021), and Picard (Broad Institute 2019) within the Arima Genomics Mapping Pipeline (Arima Genomics 2019), and we scaffolded each haplotype assembly (and produced contact maps) using YaHS (Zhou et al. 2023).

Compleasm suggested that hap2 was the slightly more complete assembly, likely because it included the avian Z chromosome, so we used minimap2 to align both haplotypes to the Zebra Finch (*Taeniopygia guttata*) genome (NCBI GCF_003957565.2). Minimap2 results confirmed hap2 included a scaffold orthologous to the Zebra Finch Z chromosome, and we used RepeatModeler (Flynn et al. 2020) within the Dfam Transposable Element Tools (TETools) container (Dfam Consortium 2023) to model repeats for this haplotype with the DFAM 0th and 3rd partitions (Storer et al. 2021) and RepBase Repeat Masker libraries v20181026 (Bao et al. 2015). With a custom repeat library created, we used RepeatMasker (Smit et al. 2013) within the TETools container to generate a general feature format (GFF) file of repeats for each haplotype, and

**Table 2.** Assembly statistics and completeness at different stages of the process (white background), and for the final assemblies submitted to the NCBI (gray background).

| | Haplotype 1 (contigs, pre-purge) | Haplotype 1 (contigs, post-purge) | Haplotype 1 (scaffolds) | Haplotype 2 (contigs, pre-purge) | Haplotype 2 (contigs, post-purge) | Haplotype 2 (scaffolds) |
|---|---|---|---|---|---|---|
| # scaffolds | 1,732 | 1,068 | 401 | 1,659 | 1,139 | 385 |
| Total scaffold length | 1,079,634,103 | 1,004,708,711 | 1,004,317,150 | 1,202,340,140 | 1,090,522,921 | 1,090,631,418 |
| Average scaffold length | 623,345 | 940,738 | 2,504,532 | 724,738 | 957,439 | 2,832,809 |
| Scaffold N50 | 3,000,474 | 3,214,539 | 70,945,146 | 2,745,311 | 3,130,792 | 70,258,100 |
| Scaffold auN | 4,405,444 | 4,683,967 | 72,153,392 | 4,113,045 | 4,431,783 | 71,588,675 |
| Scaffold L50 | 95 | 83 | 5 | 108 | 90 | 6 |
| Largest scaffold | 20,784,452 | 20,784,452 | 151,780,972 | 16,326,315 | 16,326,315 | 153,946,839 |
| Smallest scaffold | 6,775 | 9,537 | 1,000 | 9,988 | 11,184 | 1,000 |
| # contigs | 1,732 | 1,068 | 1,066 | 1,659 | 1,139 | 1,163 |
| Total contig length | 1,079,634,103 | 1,004,708,711 | 1,004,184,150 | 1,202,340,140 | 1,090,522,921 | 1,090,475,818 |
| Average contig length | 623,345 | 940,738 | 942,011 | 724,738 | 957,439 | 937,640 |
| Contig N50 | 3,000,474 | 3,214,539 | 3,214,539 | 2,745,311 | 3,130,792 | 3,072,548 |
| Contig auN | 4,405,444 | 4,683,967 | 4,677,973 | 4,113,045 | 4,431,783 | 4,395,166 |
| Contig L50 | 95 | 83 | 83 | 108 | 90 | 91 |
| Largest contig | 20,784,452 | 20,784,452 | 20,784,452 | 16,326,315 | 16,326,315 | 16,326,315 |
| Smallest contig | 6,775 | 9,537 | 1,000 | 9,988 | 11,184 | 1,000 |
| # gaps in scaffolds | 0 | 0 | 665 | 0 | 0 | 778 |
| Total gap length in scaffolds | 0 | 0 | 133,000 | 0 | 0 | 155,600 |
| Average gap length in scaffolds | 0 | 0 | 200 | 0 | 0 | 200 |
| Gap N50 in scaffolds | 0 | 0 | 200 | 0 | 0 | 200 |
| Gap auN in scaffolds | 0 | 0 | 200 | 0 | 0 | 200 |
| Gap L50 in scaffolds | 0 | 0 | 333 | 0 | 0 | 389 |
| Largest gap in scaffolds | 0 | 0 | 200 | 0 | 0 | 200 |
| Smallest gap in scaffolds | 0 | 0 | 200 | 0 | 0 | 200 |
| GC content % | 43 | 43 | 43 | 43 | 43 | 43 |
| # soft-masked bases | 0 | 0 | 118,046,428 | 0 | 0 | 145,310,361 |
| # segments | 1,732 | 1,068 | 1,066 | 1,659 | 1,139 | 1,163 |
| Total segment length | 1,079,634,103 | 1,004,708,711 | 1,004,184,150 | 1,202,340,140 | 1,090,522,921 | 1,090,475,818 |
| Average segment length | 623,345 | 940,738 | 942,011 | 724,738 | 957,439 | 937,640 |
| # gaps | 0 | 0 | 665 | 0 | 0 | 778 |
| # paths | 1,732 | 1,068 | 401 | 1,659 | 1,139 | 385 |
| **Compleasm results** | | | | | | |
| Single copy complete genes | S:94.77%, 7902 | S:94.75%, 7900 | S:94.96%, 7918 | S:98.78%, 8236 | S:98.86%, 8243 | S:99.00%, 8255 |
| Duplicated complete genes | D:0.20%, 17 | D:0.17%, 14 | D:0.14%, 12 | D:0.28%, 23 | D:0.16%, 13 | D:0.16%, 13 |
| Fragmented genes, subclass 1 | F:0.60%, 50 | F:0.59%, 49 | F:0.42%, 35 | F:0.36%, 30 | F:0.36%, 30 | F:0.26%, 22 |

**Table 2.** Continued

| | Haplotype 1 (contigs, pre-purge) | Haplotype 1 (contigs, post-purge) | Haplotype 1 (scaffolds) | Haplotype 2 (contigs, pre-purge) | Haplotype 2 (contigs, post-purge) | Haplotype 2 (scaffolds) |
|---|---|---|---|---|---|---|
| Fragmented genes, subclass 2 | I:0.00%, 0 | I:0.00%, 0 | I:0.00%, 0 | I:0.01%, 1 | I:0.01%, 1 | I:0.00%, 0 |
| Missing genes | M:4.43%, 369 | M:4.50%, 375 | M:4.47%, 373 | M:0.58%, 48 | M:0.61%, 51 | M:0.58%, 48 |
| Total genes (aves_odb10) | N:8338 | N:8338 | N:8338 | N:8338 | N:8338 | N:8338 |
| Merqury completeness | | | 87.70% | | | 93.65% |

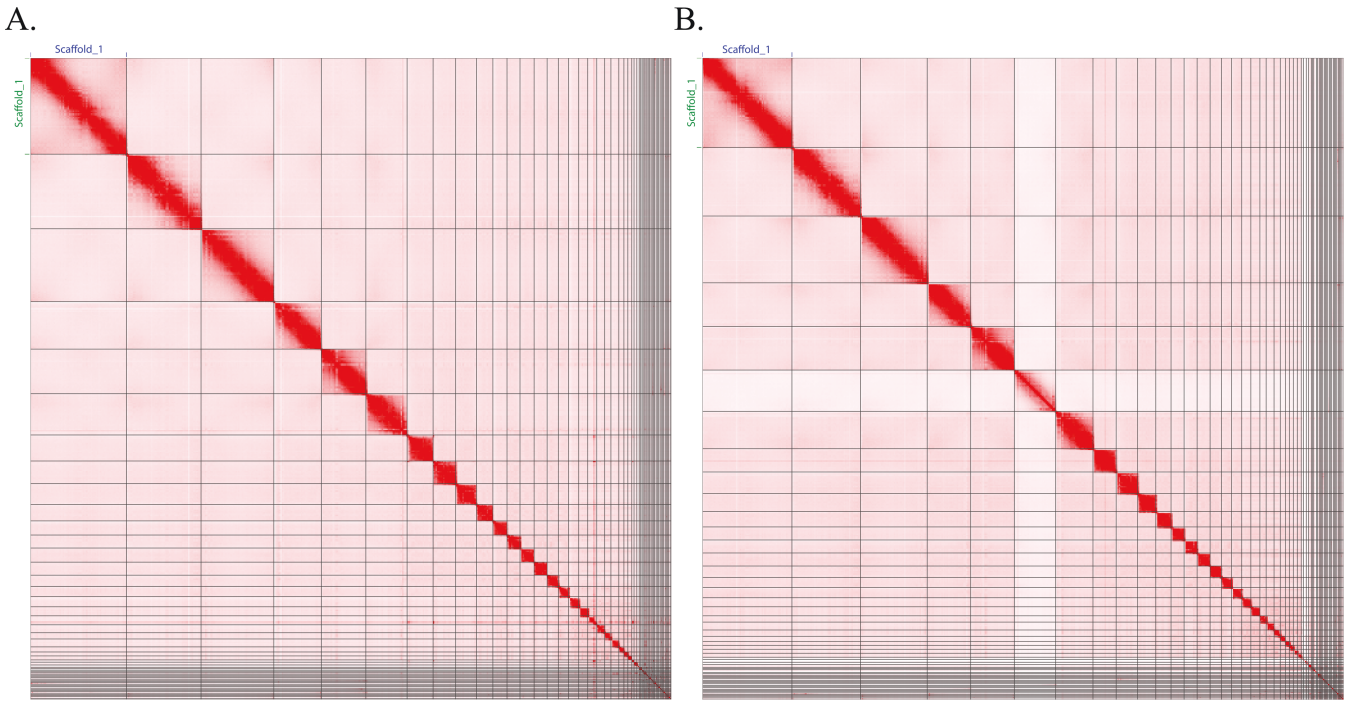


**Fig. 2.** Contact maps of the haplotype 1 (A) and haplotype 2 (B) assemblies after scaffolding with YaHS.

we soft-masked repeats in each using bedtools (Quinlan and Hall 2010).

After masking, we conducted *ab initio* annotation of hap2 using the Braker3 container (Lomsadze 2005; Stanke et al. 2006, 2008; Gotoh 2008; Iwata and Gotoh 2012; Buchfink et al. 2015; Hoff et al. 2016, 2019; Brůna et al. 2020, 2021; Gabriel et al. 2021) with the OrthoDB11 Vertebrata data set (downloaded 2 August 2023) (Kuznetsov et al. 2023), we converted the Braker GTF file to GFF using the agat (Dainat 2022) container, and we sanitized the resulting GFF using gfftk (Chen et al. 2019). We generated functional annotations for the predicted genes/transcripts using interproscan (Jones et al. 2014) and eggNOG-mapper (Huerta-Cepas et al. 2019; Buchfink et al. 2021; Cantalapiedra et al. 2021), and we integrated both sources of functional annotation information to the braker GTF file using the funannotate container (Palmer 2023).

We added the mitochondrial contig assembled with mitohifi to the hap2 assembly, and we computed a final round of assembly statistics and BUSCO completeness for each haplotype using gfastats and compleasm, in addition to estimating kmer completeness for both haplotypes using merqury (Rhie et al. 2020).

## Results

Two PacBio Sequel II runs produced a total of 2.85e6 HiFi reads having an average length of 9.5 kbp and totaling 27.1e9 HiFi bases, and we processed these files individually with cutadapt to remove a total of 1,958 reads containing adapters. We discarded the temporary assembly that was used to perform quality control of the HiC sequencing library. Illumina sequencing of the HiC library produced 817.5 M read pairs, and 793.8 M reads pairs remained after trimming. GenomeScope2 estimated that the (maximum) haploid length of the *Q. m. major* genome was 1.13e9, suggesting that the realized HiFi coverage of the genome was approximately 24x.

The HiC partitioned mode of assembly in hifiasm produced haplotype assemblies that did not contain identifiable foreign- or adapter-contamination. Both haplotypes were relatively

contiguous and complete (Table 2). Because compleasm results indicated that there were a small number of apparently duplicate contigs, we ran purge_dupes on each of the haplotype assemblies, removing 664 contigs from hap1 and 520 contigs from hap2. This reduced the number of putatively duplicate BUSCOs in each assembly while slightly decreasing the number of complete, single-copy orthologs for hap1 and increasing the number of complete, single-copy orthologs for hap2 (Table 2). We assembled a mitochondrial genome sequence for *Q. m. major* that was 16,769 bp in length, circular, and included 37 mitochondrial genes. We used the mtDNA genome assembly to identify and remove 16 additional contigs from hap1 and one contig from hap2 because these contigs partially/entirely overlapped with the mitochondrial genome. Mapping reads from the HiC library to each haplotype assembly produced a total of 199 M and 123 M intra-contig read pairs (219 M/233 M inter-contig) for hap1 and hap2 (Supplementary Table S2), and yahs significantly increased the contiguity of each assembly, while slightly increasing the BUSCO assembly completeness (Table 2, Fig. 2). Assembly completeness estimated by merqury was 87.7% for hap1, 93.6% for hap2, and 98.6% for both haplotypes considered together.

The final version of hap1, bQuiMaj1.hap1, included 401 scaffolds having a total length of 1.0 Gbp, an N50 of 70.9 Mbp, and an L50 of 5, while the final version of hap2, bQuiMaj1.hap2, included 385 scaffolds having a total length of 1.1 Gbp, an N50 of 70.3 Mbp, and an L50 of 6. Repetitive elements comprised ~10-12% of the assembly, and a majority of these were either retroelements or unclassified (Supplementary Table S2). Braker3 identified 40,525 putative genes producing 43,069 predicted transcripts (including isoforms), and our annotation procedures assigned functional information to 9665 (22%) of these genes.

## Discussion

The haplotype assemblies we produced are among the most contiguous for the family Icteridae and will facilitate future genomic studies in the clade, which comprises over 100 species inhabiting the Americas. In particular, the Boat-tailed/Great-tailed Grackle complex is a promising system for studying the genomics of reproductive isolation in recently-diverged species pairs (Selander and Giller 1961; Avise and Zink 1988; Pratt 1991; DaCosta et al. 2008), as well as the genetic basis of rapid evolutionary changes in eye color (Pratt 1974; Stevenson 1978; Corbett et al. 2024). This high-quality reference genome will aid future work using short-read sequencing, such as phylogeographic and population genetic analyses with reduced-representation data, as well as whole-genome resequencing approaches to identify loci underlying phenotypic variation.

## Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

## Acknowledgments

We thank Subir Shakya for the illustrations used in Fig. 1, and we thank the staff members at University of Maryland School of Medicine Institute for Genome Sciences and Phase Genomics for their assistance with library preparation and sequencing. We thank Donna Dittmann, Steve Cardiff, and Jeffrey Duguay for assistance with collecting permits.

## Author contributions

Eamon Corbett (Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Writing - original draft, Writing - review & editing), Andre Moncrieff (Data curation, Investigation, Methodology, Writing - review & editing), Rob Brumfield (Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing - review & editing), and Brant Faircloth (Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing)

Conflict of interest statement. None declared.

## Funding

This project was partially supported by National Science Foundation awards DEB-1655624 to BCF and RTB and IOS-1754417 to BCF, as well as funds from Louisiana State University to BCF and RTB. ECC was supported by a Werner and Hildegard Hesse Award from the American Ornithological Society. Portions of this research were conducted with high-performance computational resources provided by Louisiana State University (http://www.hpc.lsu.edu).

## Data availability

Sequence data are available from the NCBI as part of BioProject (PRJNA1176494). The haplotype assemblies have been deposited at DDBJ/ENA/GenBank under the accessions JBITNH000000000 (haplotype 1) and JBITNI000000000 (haplotype 2). The versions described in this paper are versions JBITNH010000000 and JBITNI010000000. The Supplementary Tables, a list of steps used to assemble and annotate the genome, RepeatMasker annotations, and gene predictions are available from FigShare (https://doi.org/10.6084/m9.figshare.27284436).

## References

Arima Genomics. Mapping_pipeline. [Online.]; 2019. https://github.com/ArimaGenomics/mapping_pipeline

Dfam Consortium. Dfam TE Tools. [Online.]; 2023. https://github.com/Dfam-consortium/TETools

Astashyn A, Tvedte ES, Sweeney D, Sapojnikov V, Bouk N, Joukov V, Mozes E, Strope PK, Sylla PM, Wagner L, et al. Rapid and sensitive detection of genome contamination at scale with FCS-GX. Genome Biol. 2024:25:60.

Avise JC, Zink RM. Molecular genetic divergence between avian sibling species: king and clapper rails, long-billed and short-billed dowitchers, boat-tailed and great-tailed grackles, and tufted and black-crested titmice. Auk. 1988:105:516–528.

Bancroft GT. Growth and sexual dimorphism of the boat-tailed grackle. Condor. 1984:86:423–432.

Bancroft GT. Nesting success and mortality of the boat-tailed grackle. Auk. 1986:103:86–99.

Bancroft GT. Mating system and nesting phenology of the boat-tailed grackle in central Florida. Fla Field Nat. 1987:15:1–28.

Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 2015: 6:11.
Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014:30:2114–2120.
Broad Institute. 2019. Picard. [Online.] https://broadinstitute.github.io/picard/
Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 2021:3:lqaa108.
Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform. 2020:2:lqaa026.
Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021:18:366–368.
Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015:12:59–60.
Buckner JC, Sanders RC, Faircloth BC, Chakrabarty P. The critical importance of vouchers in genomics. eLife 2021:10:e68264.
Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021:38:5825–5829.
Chen M-JM, Lin H, Chiang L-M, Childers CP, Poelchau MF. The GFF3toolkit: QC and merge pipeline for genome annotation. In Brown SJ and Pfrender ME, editors. Insect Genomics. NY, NY: Springer New York; 2019. pp. 75–87.
Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021:18:170–175.
Cheng H, Jarvis ED, Fedrigo O, Koepfli K-P, Urban L, Gemmell NJ, Li H. Haplotype-resolved assembly of diploid genomes without parental data. Nat Biotechnol. 2022:40:1332–1335.
Clum NJ. Proximate mechanisms producing sexual size dimorphism in the boat-tailed grackle (Quiscalus major). [Online.]; 1991. https://www.proquest.com/docview/303945356/abstract/3A63A8D3857F46E1PQ/1
Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009:25:1422–1423.
Corbett EC, Brumfield RT, Faircloth BC. The mechanistic, genetic and evolutionary causes of bird eye colour variation. Ibis. 2024:166:560–589.
Coues E. The natural history of quiscalus major. Ibis. 1870:12:367–378.
DaCosta J, Wehtje W, Klicka J. Historic genetic structuring and paraphyly within the Great-tailed Grackle. Condor. 2008:110:170–177.
Dainat J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format (version 1.0.0). [Online.]; 2022. https://zenodo.org/doi/10.5281/zenodo.3552717
Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021:10:giab008.
Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020:117:9451–9457.
Formenti G, Abueg L, Brajuka A, Brajuka N, Gallardo-Alba C, Giani A, Fedrigo O, Jarvis ED. Gfastats: conversion, evaluation and manipulation of genome sequences using assembly graphs. Bioinformatics. 2022:38:4214–4216.
Gabriel L, Hoff KJ, Brůna T, Borodovsky M, Stanke M. TSEBRA: transcript selector for BRAKER. BMC Bioinf. 2021:22:566.
Gotoh O. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. Nucleic Acids Res. 2008:36:2630–2638.
Guan D. Purge_Dups. [Online.]; 2022. https://github.com/dfguan/purge_dups
Hoff K, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. Methods Mol Biol (Clifton, N.J.). 2019:1962:65–95.
Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 2016:32:767–769.
Huang N, Li H. compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics. 2023:39:btad595.
Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019:47:D309–D314.
Iwata H, Gotoh O. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. Nucleic Acids Res. 2012:40:e161–e161.
Jackson JA. A Mutualistic feeding association between boat-tailed grackles and pied-billed grebes. Condor. 1985:87:147–148.
Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics (Oxford, England) 2014:30:1236–1240.
Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, Zdobnov EM. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res. 2023:51:D445–D451.
Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:34:3094–3100.
Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) 2010:26:589–595.
Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009:326:289–293.
Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 2005:33:6494–6506.
Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011:17:10.
McIlhenny EA. Life history of the boat-tailed grackle in Louisiana. Auk. 1937:54:274–295.
Melman DS, Searcy WA. Microgeographic song discrimination in a nonterritorial passerine, the boat-tailed grackle. Condor. 1999:101:845–848.
Palmer J. Funannotate. [Online.]; 2023. https://github.com/nextgenusfs/funannotate
Post W. Reproduction of female boat-tailed grackles: comparisons between South Carolina and Florida. J Field Ornithol. 1995:66:221–230.
Post W, Poston JP, Bancroft GT. Boat-tailed Grackle (*Quiscalus major*). *Birds of the World*. [Online.]; 2020. https://birdsoftheworld-org.proxy.birdsoftheworld.org/bow/species/botgra/cur/introduction
Post W, Seals CA. Nesting associations of least bitterns and boat-tailed grackles. Condor. 1993:95:139–144.
Poston JP. Dominance, access to colonies, and queues for mating opportunities by male boat-tailed grackles. Behav Ecol Sociobiol. 1997:41:89–98.
Powell AFLA, Barker FK, Lanyon SM. A Complete Species-Level Phylogeny of the Grackles (Quiscalus spp.), including the extinct slender-billed grackle, inferred from mitochondrial DNA. Condor. 2008:110:718–728.
Powell AFLA, Barker FK, Lanyon SM, Burns KJ, Klicka J, Lovette IJ. A comprehensive species-level molecular phylogeny of the New World blackbirds (Icteridae). Mol Phylogenet Evol. 2014:71:94–112.
Pratt HD. Field identification of boat-tailed and great-tailed grackles in their zone of overlap. Birding. 1974:6:217–223.
Pratt HD. Hybridization of great-tailed and boat-tailed grackles (*Quiscalus*) in Louisiana. J Louisiana Ornithol. 1991:2:2–14.

Pratt HD, Ortego B, Guillory HD. Spread of the great-tailed grackle in Southwestern Louisiana. Wilson Bull. 1977:89:483–485.
Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010:26:841–842.
Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020:11:1432.
Remsen JV, Wallace BP, Seymour MA, O'Malley DA, Johnson EI. The regional, national, and international importance of Louisiana's coastal avifauna. Wilson J Ornithol. 2019:131:221–434.
Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020:21:245.
Selander RK. Age determination and molt in the boat-tailed grackle. Condor. 1958:60:355–376.
Selander RK, Giller DR. Analysis of sympatry of great-tailed and boat-tailed grackles. Condor. 1961:63:29–86.
Smit A, Hubley R, Green P. RepeatMasker. [Online.]; 2013. www.repeatmasker.org
Snyder NFR, Snyder HA. A comparative study of mollusk predation by limpkins, Everglade kites, and boat-tailed grackles. Living Bird 1969:8:177–223.
Sprunt A Jr. Predation of boat-tailed grackles on feeding glossy ibises. Auk. 1941:58:587–588.
Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics. 2008:24:637–644.
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 Bioinf. 2006:7:62.
Stevenson HM. The populations of Boat-tailed Grackles in the south-eastern United States. Proc Biol Soc Wash. 1978:91:27–51.
Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mobile DNA 2021:12:2.
Summers J, Lukas D, Logan CJ, Chen N. The role of climate change and niche shifts in divergent range dynamics of a sister-species pair. Peer Commun J. 2023:3:e23.
Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Blaxter, Mieszkowska N, Hall N, Holland P, Durbin R, Richards T, Kersey P, et al; Darwin Tree of Life Consortium, M. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. BMC Bioinf. 2023:24:288.
Van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, Dekker J, Lander ES. Hi-C: a method to study the three-dimensional architecture of genomes. J Visual Experiments 2010:e1869.
van Rossum G. Python 3 Reference Manual. Scotts Valley, CA: Createspace; 2009.
Wehtje W. The range expansion of the great-tailed grackle (*Quiscalus mexicanus* Gmelin) in North America since 1880. J Biogeogr. 2003:30:1593–1607.
Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 2023:39:btac808.