

1 A high quality *de novo* genome assembly for clapper rail (*Rallus* 2 *crepitans*)

3 Elisa C. Elizondo^{1*}, Brant C. Faircloth² (ORCID: 0000-0002-1943-0217), Robb T. Brumfield² (ORCID: 0000-0003-2307-
4 0688), Subir B. Shakya², Vincenzo A. Ellis¹, Carl J. Schmidt³, Adrienne I. Kovach⁴, and W. Gregory Shriver¹.

5 ¹Department of Entomology and Wildlife Ecology, University of Delaware, Newark, Delaware 19716, USA. ²Museum of Natural Science and
6 Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA. ³Department of Animal and Food Sciences,
7 University of Delaware, Newark, Delaware 19716, USA. ⁴Department of Natural Resources, University of New Hampshire, Durham, New
8 Hampshire 03824, USA.
9

10 **Keywords:** Rallidae, reference genome, Hi-C Chicago, illumina shotgun

11 * Corresponding author: eelizo@udel.edu
12

13 **Abstract**

14 The clapper rail (*Rallus crepitans*), of the family Rallidae, is a secretive marsh bird species that is adapted
15 for high salinity habitats. They are very similar in appearance to the closely related king rail (*Rallus*
16 *elegans*), but while king rails are limited primarily to freshwater marshes, clapper rails are highly
17 adapted to tolerate salt marshes. Both species can be found in brackish marshes where they freely
18 hybridize, but the distribution of their respective habitats preclude the formation of a continuous hybrid
19 zone and secondary contact can occur repeatedly. This system, thus, provides unique opportunities to
20 investigate the underlying mechanisms driving their differential salinity tolerance as well as the
21 maintenance of the species boundary between the two species. To facilitate these studies, we
22 assembled a *de novo* reference genome assembly for a female clapper rail. Chicago and HiC libraries
23 were prepared as input for the Dovetail HiRise pipeline to scaffold the genome. The pipeline, however,
24 did not recover the Z chromosome so a custom script was used to assemble the Z chromosome. We
25 generated a near chromosome level assembly with a total length of 994.8 Mb comprising 13,226
26 scaffolds. The assembly had a scaffold N50 was 82.7 Mb, L50 of four, and had a BUSCO completeness
27 score of 92%. This assembly is among the most contiguous genomes among the species in the family
28 Rallidae. It will serve as an important tool in future studies on avian salinity tolerance, interspecific
29 hybridization, and speciation.
30

31 **Introduction**

32 Rallids (Aves: Rallidae) include 37 genera and 159 globally distributed species that occur primarily in
33 wetlands, jungle lowlands, and montane forests (Garcia-R et al. 2019, Winkler et al. 2020). Despite their
34 global distribution, most rallid species remain poorly understood because of their secretive nature. The
35 type genus *Rallus* includes thirteen species of slim bodied, long-billed rails that occur in the Americas,
36 Eurasia, Africa, and Madagascar (Winkler et al. 2020). Clapper rail (*Rallus crepitans*) and king rail (*Rallus*
37 *elegans*) are two closely related species that occur along the eastern coast of North America, south to
38 the Caribbean (Figure 1; del Hoyo et al. 2016, Rush et al. 2020). Clapper and king rails are similar in
39 plumage, vocalization, and morphology (Maley and Brumfield 2013), but they exhibit different habitat
40 preferences for saltwater (clapper rail) and freshwater (king rail) wetlands. The internal nasal salt glands
41 of clapper rails are larger than those of king rails, and this adaptation is believed to contribute to the
42 salinity tolerance (Conway et al. 1988) of clapper rails, although salt gland size is known to be a plastic

© The Author(s) 2023. Published by Oxford University Press on behalf of The Genetics Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

1 trait that varies based on the water salinity to which the birds are exposed (Conway et al. 1988, Olson
2 1997). Osteologically, a narrower interorbital bridge in clapper rails accommodates its larger salt gland,
3 and this species difference does not appear to be plastic, at least to the same extent as the salt gland
4 (Olson 1997). Clapper and king rail populations hybridize where they co-occur in brackish marsh (Olson
5 1997, Maley 2012).

6 Avian hybridization along salinity gradients in North American marshes occurs not only between clapper
7 and king rails, but also between Nelson's sparrows (*Ammospiza nelsoni*) and saltmarsh sparrows
8 (*Ammospiza caudacuta*) (Shriver et al. 2005, Walsh et al. 2016b, 2019). Similar to king rails, Nelson's
9 sparrows are more closely associated with fresh and brackish wetlands while saltmarsh sparrows, like
10 clapper rails, are considered salt marsh obligates (Greenberg et al. 2006, Greenlaw et al. 2018). In the
11 Nelson's/saltmarsh sparrow hybrid zone, genes associated with osmoregulation and salinity tolerance
12 exhibit increased introgression, leading to improved fitness when hybrids are compared to Nelson's
13 sparrows nesting in brackish and salt marshes (Walsh et al. 2016a). This observation suggests that for
14 some organisms, hybridization may facilitate expansion into increasingly saline environments and
15 additional work is warranted to explore these dynamics in other taxa. As climate change and sea level
16 rise alter tidal marsh salinity gradients, it is increasingly important to understand how organisms can
17 adapt to these changes in salinity.

18 To facilitate molecular investigations of the underlying mechanisms of saltwater tolerance and adaptive
19 divergence between clapper and king rail populations, we completed the first genome assembly for
20 clapper rail using DNA from a vouchered, wild female bird collected in Louisiana. To produce a
21 chromosome-level assembly, we scaffolded contigs assembled using Meraculous (Chapman et al. 2011)
22 and Spades (Bankevich et al. 2012) using Chicago and Hi-C libraries (Dovetail Genomics LLC). The
23 resulting reference genome will be foundational to future studies investigating adaptation to high
24 salinity environments, species limits in actively hybridizing populations, the conservation of *Rallus*
25 species, and the genetic effects of sea level rise on marsh taxa.

26 **Methods**

27 **Specimen collection and DNA extraction**

28 Because we were interested generating data from both sex chromosomes, we collected a female
29 clapper rail from Barataria Bay (saltwater), Plaquemines Parish, Louisiana (LSU IACUC 18-054; Louisiana
30 Department of Wildlife and Fish Permit 18-022; US Fish and Wildlife Service Permit MB02467D);
31 prepared a voucher specimen (Buckner et al. 2021) for the Louisiana State University Museum of
32 Natural Science (LSUMNS) Collection of Birds (LSUMZ 199649); and archived muscle, liver, and other
33 tissues in the LSUMNS Collection of Genetic Resources (LSUMZ B-95207). We shipped liver tissue to
34 Dovetail Genomics, LLC (Scotts Valley, CA) where Dovetail staff performed high molecular weight (HMW)
35 DNA extraction using the Blood and Cell Culture Midi Kit (Qiagen, Gmbh).

36 **Library preparation, sequencing, and assembly**

37 Following HMW DNA extraction, Dovetail staff fragmented the DNA, prepared short insert sequencing
38 libraries using an Illumina TruSeq DNA PCR-free kit, and sequenced the DNA using paired-end (PE) 150
39 base pair (BP) sequencing on an Illumina HiSeq X. The resulting data were trimmed to remove bases
40 with quality scores lower than 20 using Trimmomatic (Bolger et al. 2014), and we used meryl 1.3 (k=20;
41 <https://github.com/marbl/meryl>) and Genomescope (Vurture et al. 2017) to estimate the genome size
42 and heterozygosity of the sampled individual.

1 Dovetail staff used in-house software to profile the trimmed reads at a variety of k-mer values (19, 31,
2 49, 79, 109) and fit negative binomial models to the data to determine the best k-mer value for
3 assembly. The constrained heterozygous model with 49-mers and a homozygous peak depth of 42.0 was
4 selected as optimal for the assembly. Dovetail staff then assembled contigs using Meraculous with a k-
5 mer value of 49, a minimum k-mer frequency of seven, and the diploid nonredundant haplotigs mode.

6 Following contig assembly, Dovetail staff used remaining tissue to prepare a single, proprietary
7 “Chicago” library following the methods described in Putnam et al. (2016) and summarized in Salter et
8 al. (2019). They sequenced the resulting Chicago library on an Illumina HiSeq X using PE, 150 bp reads to
9 a depth of approximately 70X. Similarly, Dovetail staff prepared one HiC library from remaining tissue
10 following the methods described in Lieberman-Aiden et al. (2009) and summarized in Salter et al. (2019).
11 Dovetail staff sequenced the resulting HiC library to a depth of approximately 45X using PE, 150 bp reads
12 on an Illumina HiSeq X. After preparing and sequencing Chicago and HiC libraries, Dovetail staff used
13 HiRise (Putnam et al, 2016) to conduct two rounds of scaffolding: (1) using the Chicago reads to scaffold
14 the Meraculous contigs, and (2) using the HiC reads to scaffold the Chicago scaffolds. We refer to the
15 resulting assembly as the “Dovetail HiC Assembly”.

16 After receiving the Dovetail HiC Assembly, we computed contiguity statistics using assembly-stats
17 (<https://github.com/sanger-pathogens/assembly-stats>) and estimated assembly completeness using
18 BUSCO v5.1.3 (Manni et al. 2021) and aves_odb10. While evaluating this version of the assembly, we
19 noticed that the Z chromosome appeared to be missing. Specifically, after aligning scaffolds and contigs
20 from the Dovetail HiC Assembly to the chicken genome assembly (UCSC galGal6; NCBI
21 GCF_000002315.5) using ragtag v1.0.1 (Alonge et al. 2019), we did not recover any contigs or scaffolds
22 that aligned to the chicken Z chromosome, suggesting Z chromosome contigs and scaffolds were not
23 present. This problem has been observed in other Dovetail assemblies of birds (Del-Rio et al. 2021,
24 Recuerda et al. 2021, Shakya et al. 2021) and may have resulted from the coverage parameters used by
25 Dovetail during the Meraculous assembly process inadvertently excluding contigs representing sex
26 chromosomes.

27 We addressed this problem by maintaining the macrochromosomes (scaffolds > 20 Mbp) from the
28 Dovetail HiC Assembly while re-assembling and re-scaffolding contigs representing the
29 microchromosomes. To start the microchromosome reassembly process, we trimmed the short-insert
30 sequencing reads with trimmomatic v0.39 and corrected the trimmed reads using Musket v1.1 (Liu et al.
31 2013) and a kmer value of 61. We then performed a second *de novo* assembly using spades v3.14.0
32 (Andrey et al. 2020) with error correction turned off (--only-assembler) on a high-memory (1.5 TB)
33 compute node, and we filtered the resulting assembly using faFilter (Kent et al. 2002) to remove contigs
34 < 1 kbp. We extracted macrochromosomes (scaffolds > 20 Mbp) from the Dovetail HiC Assembly using
35 faSize (Kent et al. 2002) and custom Python code, concatenated each into a single file, and used ragtag
36 to align the contigs output by spades to this file macrochromosomes. Because of the way that ragtag
37 formats output files, we were able to separate the contigs that aligned to macrochromosomes from
38 those that did not, and we used custom Python code to create a file of contigs that did not align to the
39 macrochromosomes. We provided this file of contigs to Dovetail staff, who re-ran the Chicago and HiC
40 scaffolding processes using their proprietary HiRise pipeline.

41 After rescaffolding, we merged the resulting scaffolds (many representing microchromosomes) into the
42 file of macrochromosomes to produce an assembly representing the entire genome, and we sorted the
43 file by descending scaffold length using sortbyname in BMap 38.78 (Bushnell 2014). We used custom
44 Python code to rename all scaffolds, and we used faFilter to remove contigs/scaffolds shorter than 1000

1 bp in length. To ensure that the updated assembly contained scaffolds representing the Z chromosome,
2 we performed a second alignment of the updated assembly to the chicken genome assembly (galGal6).

3 After validating that the updated assembly contained a large scaffold representing the Z chromosome,
4 we used BWA v0.7.17 (Li 2013) to align reads from the short-insert libraries to the assembly, SAMtools
5 v1.1.0 (Li et al. 2009) to sort and index the resulting BAM file, and Pilon 1.23 (Walker et al. 2014) to
6 polish the assembly by fixing `--all` of the issues identified. We modeled repeats in the polished
7 assembly using RepeatModeler v2.0.1 (Smith and Hubley 2008), and we soft-masked repeats using the
8 output of RepeatMasker v 4.1.0 (Smith et al. 2013) with BedTools (Quinlan and Hall 2010). After
9 polishing and repeat-masking, we checked the resulting assembly for adapter and other contamination
10 using the NCBI Foreign Contamination Screen (FCS) tool (<https://github.com/ncbi/fcs>), we removed
11 scaffolds/contigs that represented contamination, and we removed bases from scaffolds/contigs that
12 matched adapter sequences. We also identified scaffolds/contigs that represented mitochondrial
13 contamination by mapping the assembly to the mtDNA genome of *Rallus limicola* (CM040152.1) using
14 minimap2 (v2.17-r941; Li 2018) and removing those scaffolds/contigs that matched (>90% length, >90%
15 identity) portions of this mtDNA sequence. After making these changes, we sorted the remaining
16 scaffolds/contigs by size and renamed them in order of decreasing length using custom Python code,
17 and we used meryl 1.3 and Merqury 1.3 (Rhie et al. 2020) to compute reference-free estimates of k-mer
18 completeness and consensus quality.

19 To produce a contact map of the resulting assembly, we removed adapters and low-quality bases from
20 the HiC reads using trimmomatic, and we mapped trimmed reads to the assembly using BWA (v0.7.17)
21 and SAMtools (v1.10). We used Picard (v.2.27.5; <http://broadinstitute.github.io/picard>) to sort and
22 deduplicate properly aligned reads, and we produced a contact map of the deduplicated data using
23 PretextView (v0.1.9; <https://github.com/wtsi-hpag/PretextView>) and
24 PretextView (v0.2.5; <https://github.com/wtsi-hpag/PretextView>). We also assembled the mitochondrial genome by inputting
25 trimmed reads from the short-insert libraries to MitoFinder v1.4.1 (Allio et al. 2020, Li et al. 2016) along
26 with the NCBI reference sequence of *Rallus indicus* (NC_068741.1), which MitoFinder uses to identify
27 mitochondrial reads during the initial stages of assembling a mitochondrial genome.

28 To ensure that repeat annotations exactly matched the names and coordinates in this final version of
29 the assembly, we removed the soft-masking from the assembly, re-ran RepeatMasker with the repeat
30 models we created, and soft-masked repeats using BEDTools. We assigned the Tree of Life Identifier
31 (ToLID; <https://id.tol.sanger.ac.uk>) bRalCre1.1 to this version of the assembly, computed final set of
32 contiguity statistics (assembly-stats) and BUSCO scores (aves_odb10) for this assembly version, and
33 archived bRalCre1.1 with NCBI Genome. To compare bRalCre1.1 with genome assemblies from other
34 rallids, we downloaded all assemblies for the family, and we computed contiguity statistics using
35 assembly-stats and completeness estimates using BUSCO (aves_odb10) for each.

36 Results and Discussion

37 Short-insert library sequencing produced 325 million read pairs with an approximate insert size of 382
38 bp, and Genomescope results suggested that the *Rallus* genome was ~1.3 Gb with a relatively low
39 heterozygosity of 0.75 to 0.76%. Meraculous assembly using a k-mer value of 49 output 55,528 contigs
40 with a total length of 990.8 Mb, a N50 of 50 kb (L50 = 5,380), and a maximum contig length of 606.9 kb.

41 Chicago library sequencing produced 254 million read pairs, and HiRise made 27,838 joins and 24 breaks
42 to the Meraculous assembly, producing an intermediate Chicago assembly including 19,218 scaffolds
43 and having a total length of 994.3 Mb, a N50 of 1.8 Mb (L50 = 128), a N90 of 0.06 Mb (L90 = 1384), and a
44 maximum scaffold length of 13.8 Mb. HiC library sequencing produced 170 million read pairs, and HiRise

1 made 5,992 joins and zero breaks to the Chicago assembly. Fifty-seven gaps in the resulting assembly
2 were closed using short-insert reads to produce the Dovetail HiC Assembly that included 13,226
3 scaffolds having a total length of 994.9 Mb, a N50 of 82.7 Mb (L50 = 4) scaffolds, a N90 of 10.8 Mb (L90 =
4 18), and a maximum scaffold length of 204 Mb. BUSCO completeness estimates for the Dovetail HiC
5 Assembly are provided in Table 2.

6 Contig re-assembly using spades output 55,026 contigs having a total length of 1.1 Gb, a N50 of 58.0 kb
7 (L50 = 4,904), a N90 of 9.5 kb (L90 = 22,795), and a maximum contig length of 907 kb. We identified
8 24,773 contigs that did not align to macrochromosomes in the Dovetail HiC assembly and we submitted
9 these to Dovetail for re-scaffolding, which output a set of 12,193 scaffolds having an N50 of 15.3 Mb
10 (L50 = 5) and a N90 of 8 Kb (L90 = 673). The longest scaffold in the re-assembly was 76.1 Mb in length
11 and primarily aligned to the chicken Z chromosome. After merging the macrochromosomes from the
12 Dovetail HiC Assembly with these scaffolds representing the microchromosomes and unplaced contigs
13 and polishing the assembly, we removed 4 contigs identified by the NCBI FCS tools as
14 alphaproteobacteria or eukaryotic viruses, masked 44 bases that corresponded to known adapter
15 sequences, and removed 5 contigs identified as mitochondrial contamination. The contact map
16 illustrated that HiRise performed well when scaffolding large (>100 kb) macro- and micro-chromosomes
17 (Supplemental Figure 1), although we could not discern a shift in the distribution of scaffold lengths that
18 potentially differentiated microchromosomes from unplaced scaffolds (Supplemental Data). MitoFinder
19 assembled a contig representing the mitochondrial genome that was similar in length (17.1 kb) to other
20 rail species.

21 The final version of the assembly, bRalCre1.1, included 12,159 scaffolds/contigs having a total length of
22 1.1 Gb, a N50 of 82.9 Mb (L50 = 4), a N90 of 12.2 Mb (L90 = 20), and a maximum scaffold length of 204.6
23 Mb. BUSCO completeness estimates for bRalCre1.1 improved on the results from the Dovetail HiC
24 Assembly (Table 2), although several BUSCOs remained fragmented (n=216; 2.6%) or were not detected
25 (n=451; 5.4%). Merqury results suggested that bRalCre1.1 was relatively complete (kmer
26 completeness=91.4%) and accurate (consensus quality = 55.2 or > 99.999% accuracy). Repetitive
27 elements comprised ~9% of the assembly (Supplemental Table 2), and a majority of these repeats were
28 retroelements.

29 The bRalCre1.1 assembly we produced is the second for a species in the genus *Rallus* and one of six
30 assemblies representing taxa within the Rallidae. Our assembly is among the most contiguous for the
31 taxonomic family (Supplemental Table 1), and the availability of a genome assembly representing this
32 genus will facilitate investigations of salinity tolerance, interspecific hybridization, and mechanisms of
33 speciation in clapper and king rails.

34 **Data availability**

35 All short-insert, Chicago, and HiC sequencing data are available as part of NCBI BioProject PRJNA926626.
36 The Whole Genome Shotgun project for bRalCre1.1 has been deposited at DDBJ/ENA/GenBank under
37 the accession JAQOTC000000000. The version described in this paper is version JAQOTC010000000.
38 Supplemental Table 1, Supplemental Figure 1, a list of steps used to assemble the genome that includes
39 the Python code used, Genomescope results, the PretextMap, Merqury results, RepeatMasker
40 annotations, and results from BUSCO analyses of other rallid genomes are available from FigShare
41 (<https://doi.org/10.6084/m9.figshare.21983261>).

42

43

1 **Acknowledgements**

2 We thank the Dovetail Genomics staff members for facilitating this work. We thank Andre Moncrieff for
3 assistance with specimen collection and voucher specimen preparation and Donna Dittmann for
4 assistance with the tissue loan. We thank LSUMZ staff for curation of the voucher specimen and
5 associated samples.

6 **Funding**

7 Funding for this project was provided by the University of Delaware College of Agriculture and Natural
8 Resources Seed Grant. ECE was supported by the University of Delaware. BCF and RTB were partially
9 supported by DEB-1655624. Portions of this research were conducted with high performance
10 computing resources provided by Louisiana State University (<http://www.hpc.lsu.edu>) and the Louisiana
11 Optical Network Infrastructure (<https://loni.org>).

12 **Conflicts of interest**

13 None declared.

14 **Literature Cited**

- 15 Allio, R., A. Schomaker-Bastos, J. Romiguier, F. Prosdocimi, B. Nabholz, F. Delsuc. 2020. MitoFinder:
16 Efficient automated large-scale extraction of mitogenomic data in target enrichment
17 phylogenomics. *Molecular Ecology Resources* 20:892-905.
- 18 Alonge, M., S. Soyk, S. Ramakrishnan, X. Wang, S. Goodwin, F. J. Sedlazeck, Z. B. Lippman, and M. C.
19 Schatz. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome*
20 *Biology* 20:224.
- 21 Andrey, P., A. Dmitry, M. Dmitry, L. Alla, and K. Anton. 2020. Using SPAdes De Novo Assembler. *Current*
22 *protocols in bioinformatics* 70:e102.
- 23 Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S.
24 Pham, A. D. Prjibelski, A. V Pyshkin, A. V Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A.
25 Pevzner. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell
26 Sequencing. *Journal of Computational Biology* 19:455–477. Mary Ann Liebert, Inc., publishers.
27 <<https://doi.org/10.1089/cmb.2012.0021>>.
- 28 Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
29 data. *Bioinformatics* 30:2114–2120. <<https://doi.org/10.1093/bioinformatics/btu170>>.
- 30 Buckner, JC, Sanders, RC, Faircloth, BC, Chakrabarty, P. 2021. The critical importance of vouchers in
31 genomics. *Elife* 10:e68264.
- 32 Bushnell, B. 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner.
33 <<https://sourceforge.net/projects/bbmap/>>.
- 34 Chapman, J. A., I. Ho, S. Sunkara, S. Luo, G. P. Schroth, and D. S. Rokhsar. 2011. Meraculous : De Novo
35 Genome Assembly with Short Paired-End Reads. 6.
- 36 Conway, G. L., M. R. Hucmcst, and R. R. Moldenhaltr. 1988. Extra-renal salt excretion in Clapper and King
37 Rails. *Comparative Biochemistry and Physiology* 9:671–674. <[https://ac-els-cdn-
38 com.udel.idm.oclc.org/0300962988909462/1-s2.0-0300962988909462-main.pdf?_tid=94b45dae-
39 e759-4d81-bf06-da97ea3202de&acdnt=1539636642_5c82ba83f9ba86f5b019616dbfd84260](https://ac-els-cdn-com.udel.idm.oclc.org/0300962988909462/1-s2.0-0300962988909462-main.pdf?_tid=94b45dae-e759-4d81-bf06-da97ea3202de&acdnt=1539636642_5c82ba83f9ba86f5b019616dbfd84260)>.
40 Accessed 15 Oct 2018.
- 41 Del-Rio, G., M. A. Rego, B. M. Whitney, F. Schunck, L. F. Silveira, B. C. Faircloth, and R. T. Brumfield. 2021.
42 Displaced clines in an avian hybrid zone (Thamnophilidae: Rhegmatorhina) within an Amazonian

1 interfluvial. *Evolution* n/a. John Wiley & Sons, Ltd. <<https://doi.org/10.1111/evo.14377>>.

2 Garcia-R, J. C., C. E. Gonzalez-Orozco, and S. A. Trewick. 2019. Contrasting patterns of diversification in a
3 bird family (Aves: Gruiformes: Rallidae) are revealed by analysis of geospatial distribution of
4 species and phylogenetic diversity. *Ecography* 42:500–510. John Wiley & Sons, Ltd.
5 <<https://doi.org/10.1111/ecog.03783>>.

6 Greenberg, R., J. E. Maldonado, S. Droege, and M. V McDonald. 2006. Tidal Marshes: A Global
7 Perspective on the Evolution and Conservation of Their Terrestrial Vertebrates. *BioScience* 56:675–
8 685. <[https://doi.org/10.1641/0006-3568\(2006\)56\[675:TMAGPO\]2.0.CO](https://doi.org/10.1641/0006-3568(2006)56[675:TMAGPO]2.0.CO)>.

9 Greenlaw, J. S., C. S. Elphick, W. Post, and J. D. Rising. 2018. Saltmarsh Sparrow. P. G. Rodewald, editor.
10 *Birds of the World*.

11 del Hoyo, J., N. Collar, and D. A. Christie. 2016. Clapper Rail (*Rallus crepitans*). *Handbook of the Birds of*
12 *the World Alive*.

13 Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Haussler. 2002. The
14 human genome browser at UCSC. *Genome research* 12:996–1006.

15 B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakhthikumar, C. A. Cuomo, Q. Zeng, J.
16 Wortman, S. K. Young, A. M. Earl. 2014. Pilon: An integrated tool for comprehensive microbial
17 variant detection and genome assembly improvement. *PLoS ONE* 9: e112963.
18 <<https://doi.org/10.1371/journal.pone.0112963>>.

19 Li, D., R. Luo, C. M. Liu, C. M. Leung, H. F. Ting, K. Sadakane, H. Yamashita, T. W. Lam. 2016. MEGAHIT
20 v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and
21 community practices. *Methods* 102:3-11.

22 Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*
23 preprint arXiv:1303.3997.

24 Li, H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18): 3094–3100.
25 <<https://doi.org/10.1093/bioinformatics/bty191>>.

26 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and
27 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and
28 SAMtools. *Bioinformatics* 25:2078–2079. <<https://doi.org/10.1093/bioinformatics/btp352>>.

29 Lieberman-Aiden, E., N. L. van Berkum, L. Williams, M. Imakaev, T. Ragozy, A. Telling, I. Amit, B. R.
30 Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A.
31 Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker. 2009. Comprehensive
32 Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*
33 326:289 LP – 293. <<http://science.sciencemag.org/content/326/5950/289.abstract>>.

34 Liu, Y., J. Schröder, and B. Schmidt. 2013. Musket: a multistage k-mer spectrum-based error corrector for
35 Illumina sequence data. *Bioinformatics* 29:308–315.
36 <<https://doi.org/10.1093/bioinformatics/bts690>>.

37 Maley, J. M. 2012. Ecological speciation of King rails (*Rallus elegans*) and Clapper rails (*Rallus*
38 *longirostris*). Ph.D. dissertation, Louisiana State University.
39 <http://digitalcommons.lsu.edu/gradschool_dissertations/1773>.

40 Maley, J. M., and R. T. Brumfield. 2013. Mitochondrial and Next-Generation Sequence Data Used to Infer
41 Phylogenetic Relationships and Species Limits in the Clapper/King Rail Complex - Datos
42 Mitochondriales y de la Próxima Generación Usados para Inferir Relaciones Filogenéticas y Límites
43 de Especi. *The Condor* 115:316–329. American Ornithological Society.
44 <<http://www.jstor.org/stable/10.1525/cond.2013.110138>>.

45 Manni, M., M. R. Berkeley, M. Seepey, F. A. Simão, and E. Zdobnov. 2021. BUSCO update: novel and
46 streamlined workflows along with broader and deeper phylogenetic coverage for scoring of

- 1 eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* 38:4647–4654.
- 2 Olson, S. L. 1997. Toward a less imperfect understanding of the systematics and biogeography of the
3 Clapper and King Rail Complex. Pages 93–111 in R. W. Dickerman, editor. *The Era of Allan R.*
4 *Phillips*. Horizon Communications, Albuquerque, NM.
5 <<https://pdfs.semanticscholar.org/cfb072bb341640a1ecea849f77707890406.pdf>>.
- 6 Putnam, N. H., B. L. O’Connell, J. C. Stites, B. J. Rice, M. Blanchette, R. Calef, C. J. Troll, A. Fields, P. D.
7 Hartley, C. W. Sugnet, D. Haussler, D. S. Rokshar, R. E. Green. 2016. Chromosome-scale shotgun
8 assembly using an in vitro method for long-range linkage. *Genome Research* 26: 342–350.
9 <<https://doi.org/10.1101/gr.193474.115>>.
- 10 Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.
11 *Bioinformatics* 26:841–842. <<https://doi.org/10.1093/bioinformatics/btq033>>.
- 12 Recuerda, M., J. Vizueta, C. Cuevas-Caballé, G. Blanco, J. Rozas, and B. Milá. 2021. Chromosome-Level
13 Genome Assembly of the Common Chaffinch (Aves: *Fringilla coelebs*): A Valuable Resource for
14 Evolutionary Biology. *Genome Biology and Evolution* 13:evab034.
15 <<https://doi.org/10.1093/gbe/evab034>>.
- 16 Rhie, A., B. P. Walenz, S. Koren, A. M. Phillipy. 2020. Merqury: reference-free quality, completeness, and
17 phasing assessment for genome assemblies. *Genome Biology* 21:245.
- 18 Rush, S. A., K. F. Gaines, W. R. Eddleman, and C. J. Conway. 2020. Clapper Rail (*Rallus crepitans*), version
19 1.0. in *Birds of the World*. P. G. Rodewald, editor. *Birds of the World*. Cornell Lab of Ornithology,
20 Ithaca, NY, USA.
- 21 Salter J.F., O. Johnson, N. J. Stafford, III, W. F. Herrin, Jr., D. Schilling, C. Cedotal, R. T. Brumfield, B. C
22 Faircloth. 2019. A highly contiguous reference genome for Northern Bobwhite (*Colinus*
23 *virginianus*), G3 Genes|Genomes|Genetics 9:3929–3932.
24 <<https://doi.org/10.1534/g3.119.400609>>.
- 25 Shakya, S. B., T. Haryoko, M. Irham, Suparno, D. M. Prawiradilaga, and F. H. Sheldon. 2021. Genomic
26 investigation of colour polymorphism and phylogeographic variation among populations of black-
27 headed bulbul (*Brachypodius atriceps*) in insular southeast Asia. *Molecular Ecology* 30:4757–4770.
28 John Wiley & Sons, Ltd. <<https://doi.org/10.1111/mec.16089>>.
- 29 Shriver, W. G., J. P. Gibbs, P. D. Vickery, H. L. Gibbs, T. P. Hodgman, P. T. Jones, and C. N. Jacques. 2005.
30 Concordance Between Morphological and Molecular Markers in Assessing Hybridization Between
31 Sharp-Tailed Sparrows in New England. *The Auk* 122:94–107.
32 <<https://doi.org/10.1093/auk/122.1.94>>.
- 33 Smith, A., and R. Hubley. 2008. RepeatModeler Open-1.0. <<http://www.repeatmasker.org>>.
- 34 Smith, A., R. Hubley, and P. Green. 2013. RepeatMasker Open-4.0. <<http://www.repeatmasker.org>>.
- 35 Sullivan, B. L., C. L. Wood, R. E. Illif, R. E. Bonney, D. Fink, and S. Kelling. 2009. eBird: a citizen-based bird
36 observation network in the biological sciences. *Biological Conservation* 142:2282–2292.
- 37 Vurture, G. W., , F. J. Sedlazeck, M. Nattestad, C. J. Underwood, H. Fang, J. Gurtowski, M. C. Schatz.
38 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33:
39 2202–22.
- 40 Walsh, J., P. M. Benham, P. E. Deane-Coe, P. Arcese, B. G. Butcher, Y. L. Chan, Z. A. Cheviron, C. S.
41 Elphick, A. I. Kovach, B. J. Olsen, W. G. Shriver, V. L. Winder, and I. J. Lovette. 2019. Genomics of
42 rapid ecological divergence and parallel adaptation in four tidal marsh sparrows. *Evolution Letters*
43 3:324–338. John Wiley & Sons, Ltd. <<https://doi.org/10.1002/evl3.126>>.
- 44 Walsh, J., B. J. Olsen, K. J. Ruskin, W. Gregory Shriver, K. M. O’Brien, and A. I. Kovach. 2016a. Extrinsic
45 and intrinsic factors influence fitness in an avian hybrid zone. *Biological Journal of the Linnean*
46 *Society* 119:890–903. <<https://doi.org/10.1111/bij.12837>>.

- 1 Walsh, J., R. J. Rowe, B. J. Olsen, W. G. Shriver, and A. I. Kovach. 2016b. Genotype-environment
 2 associations support a mosaic hybrid zone between two tidal marsh birds. *Ecology and Evolution*
 3 6:279–294. John Wiley & Sons, Ltd. <<https://doi.org/10.1002/ece3.1864>>.
 4 Winkler, D. W., S. M. Billerman, and I. J. Lovette. 2020. Rails, Gallinules, and Coots (Rallidae), version 1.0
 5 in *Birds of the World*. B. K. Keeney, P. G. Rodewald, and T. S. Schulenberg, editors. Cornell Lab of
 6 Ornithology, Ithaca, NY, USA.

7
 8 **Figure 1.** Clapper rail (*Rallus crepitans*) range produced using observational data from eBird (Sullivan et
 9 al. 2009) along with an image of a clapper rail captured in Woodland Beach, Delaware (photo credit:
 10 Elisa Elizondo).

11
 12 **Table 1.** Contiguity statistics for *Rallus crepitans* assemblies comparing the Dovetail HiC Assembly and
 13 the bRalCre1.1 assembly.

	Dovetail HiC Assembly	bRalCre1.1
Scaffolds	13,226	12,159
Total length (Mb)	994.8	1,107.5
N50 (Mb)	82.7	82.9
N90 (Mb)	10.8	12.2
L50	4	4
L90	18	20
Longest Scaffold (Mb)	204.0	204.6
# N's	4,085,069	3,899,784
# Gaps	42,269	41,488

14
 15 **Table 2.** Estimates of assembly completeness using the BUSCO aves_odb10 database (n = 8338 BUSCOs)
 16 showing the improvements in completeness between the Dovetail HiC Assembly and the bRalCre1.1
 17 assembly, which includes the Z chromosome.

	Dovetail HiC Assembly		bRalCre1.1	
	Count	Percentage	Count	Percentage
Complete BUSCOs	7130	85.6	7671	92.0
Complete and single-copy BUSCOs	7117	85.4	7616	91.3
Complete and duplicated BUSCOs	13	0.2	55	0.7
Fragmented BUSCOs	314	3.8	216	2.6
Missing BUSCOs	894	10.6	451	5.4

18

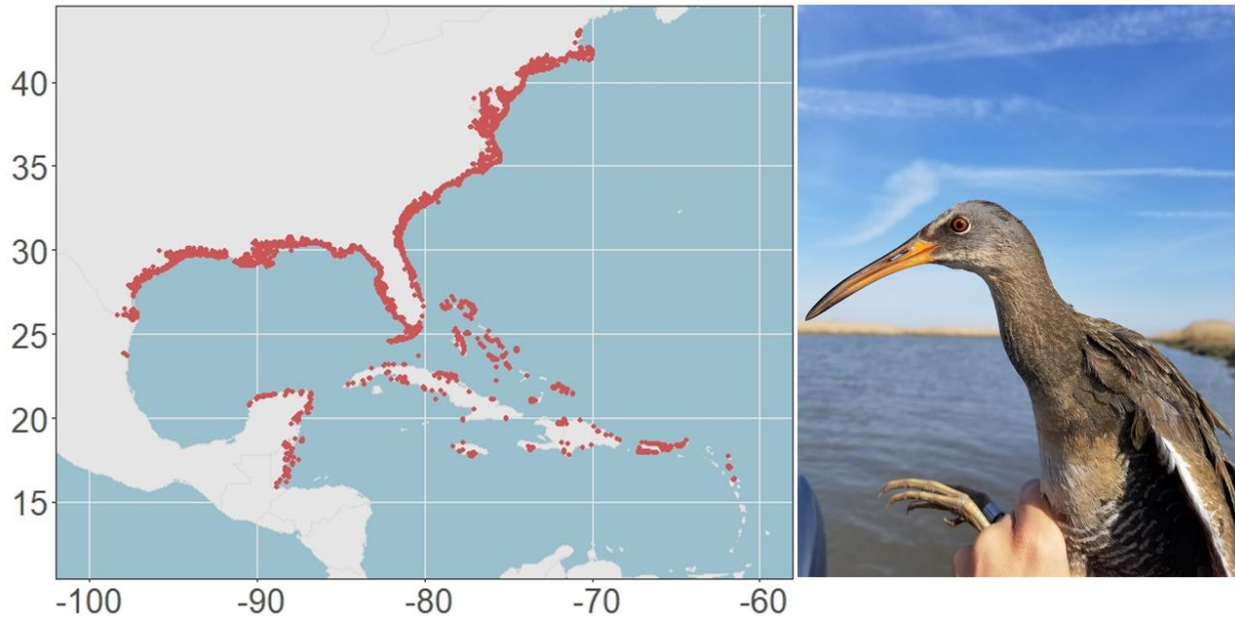


Figure 1
339x173 mm (x DPI)

1
2
3