A Chromosome-level Genome Assembly of the Reed Warbler (*Acrocephalus scirpaceus*)

Camilla Lo Cascio Sætre[1*], Fabrice Eroukhmanoff[1], Katja Rönkä[2,3], Edward Kluen[2,3], Rose Thorogood[2,3], James Torrance[4], Alan Tracey[4], William Chow[4], Sarah Pelan[4], Kerstin Howe[4], Kjetill S. Jakobsen[1], Ole K. Tørresen[1].

[1]Centre for Ecological and Evolutionary Synthesis, University of Oslo, Norway

[2]HiLIFE Helsinki Institute of Life Sciences, University of Helsinki, Helsinki 00011, Finland

[3]Research Programme in Organismal and Evolutionary Biology, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki 00011, Finland

[4]Wellcome Sanger Institute, Cambridge, UK

[*]Corresponding author: E-mail: c.l.c.satre@ibv.uio.no

Abstract

The reed warbler (*Acrocephalus scirpaceus*) is a long-distance migrant passerine with a wide distribution across Eurasia. This species has fascinated researchers for decades, especially its role as host of a brood parasite, and its capacity for rapid phenotypic change in the face of climate change. Currently, it is expanding its range northwards in Europe, and is altering its migratory behaviour in certain areas. Thus, there is great potential to discover signs of recent evolution and its impact on the genomic composition of the reed warbler. Here we present a high-quality reference genome for the reed warbler, based on PacBio, 10X and Hi-C sequencing. The genome has an assembly size of 1,075,083,815 bp with a scaffold N50 of 74,438,198 bp and a contig N50 of 12,742,779 bp. BUSCO analysis using aves_odb10 as a model showed that 95.7% of genes in the assembly were complete. We found unequivocal evidence of two separate macrochromosomal fusions in the reed warbler genome, in addition to the previously identified fusion between chromosome Z and a part of chromosome 4A in the Sylvioidea superfamily. We annotated 14,645 protein-coding genes, of which 97.5% were complete BUSCO orthologs. This reference genome will serve as an important resource, and will provide new insights into the genomic effects of evolutionary drivers such as coevolution, range expansion, and adaptations to climate change, as well as chromosomal rearrangements in birds.

Keywords: genome assembly, Hi-C sequencing, long reads, reference genome, *Acrocephalus scirpaceus*.

Significance statement

The reed warbler (*Acrocephalus scirpaceus*) has been lacking a genomic resource, despite having been broadly researched in studies of coevolution, ecology and adaptations to climate change. Here, we generated a chromosome-length genome assembly of the reed warbler, and found evidence of macrochromosomal fusions in its genome, which are likely of recent origin. This genome will provide the opportunity for a deeper understanding of the evolution of genomes in birds, as well as the evolutionary path and possible future of the reed warbler.

Introduction

The ecology and evolution of the reed warbler (*Acrocephalus scirpaceus*) has been of interest for over 40 years (Thorogood et al. 2019) as it is one of the favourite host species of the brood-parasitic common cuckoo (*Cuculus canorus*) (Davies and Brooke 1989; Stokke et al. 2018). Decades of field experiments have demonstrated behavioural coevolution and spatial and temporal variation in species interactions (e.g., Thorogood and Davies 2013). However, the reed warbler's response to climate change has begun to attract increasing attention. Reed warblers are experiencing far less severe declines in population size than is typical for long-distance migrants (Both et al. 2010; Vickery et al. 2014). In fact, they are expanding their breeding range northwards into Fennoscandia (Järvinen and Ulfstrand 1980; Røed 1994; Stolt 1999; Brommer et al. 2012), and have generally increased their productivity following the rise in temperature (Schaefer et al. 2006; Eglington et al. 2015; Meller et al. 2018). They are also showing rapid changes in phenology (Halupka et al. 2008), and migratory behaviour; instead of crossing the Sahara, monitoring suggests that some reed warblers now remain on the Iberian Peninsula over winter (Chamorro et al. 2019). Morphological traits such as body mass and wing shape have been shown to change rapidly in reed warbler populations, indicating possible

local adaptation (Salewski et al. 2010; Kralj et al. 2010; Sætre et al. 2017). Genetic differentiation is generally low between reed warbler populations, but moderate levels of differentiation have been connected to both migratory behaviour (Procházka et al. 2011) and wing shape (Kralj et al. 2010). Reed warblers thus provide a promising system to study population, phenotypic, and genetic responses to climate change.

Although there has been an increasing number of avian genome assemblies in recent years (e.g., Feng et al. 2020), many non-model species, including the reed warbler, are still lacking a genome resource. To date, the closest relative to the reed warbler with a published reference genome is the great tit (*Parus major*) (GCA_001522545.3, deposited in NCBI; Laine et al. 2016), but the unpublished genome of the garden warbler (*Sylvia borin*) is available in public databases (GCA_014839755.1, deposited in NCBI). There is also a genome in preprint from the *Acrocephalus* genus, the great reed warbler (*A. arundinaceus*) (Sigeman et al. 2020a), but the scaffolds are not chromosome-length.

Here, we present the first genome assembly of the reed warbler, based on PacBio, 10X and Hi-C sequencing, with descriptions of the assembly, manual curation and annotation. This genome will be a valuable resource for a number of studies, including studies of coevolution, population genomics, adaptive evolution and comparative genomics. For reduced-representation sequencing (e.g., RAD-seq) studies, it will help produce a more robust SNP set than with a *de novo* approach (Shafer et al. 2017). It will facilitate the detection of selective sweeps, and provide the physical localization of variants (Manel et al. 2016), thus giving insight into the potential genes involved in adaptation. Furthermore, the genome will be an important resource in the study of chromosomal rearrangements in birds.

Materials and Methods

Sampling and isolation of genomic DNA

Blood was collected from a brachial vein of a female reed warbler (subspecies *A. scirpaceus scirpaceus*, NCBI Taxonomy ID: 126889) in Storminnet, Porvoo (60°19'24.9"N 25°35'23.0"E), Finland, on May 22, 2019. Catching and sampling procedures complied with the Finnish law on animal experiments and permits were licenced by the National Animal Experiment Board (ESAVI/3920/2018) and Southwest Finland Regional Environment Centre (VARELY/758/2018). Reed warblers were trapped with a mist net, ringed and handled by E.K. under his ringing licence.

The blood (~80 ml) was divided and stored separately in 500 ml ethanol, and in 500 ml SET buffer (0.15M NaCl, 0.05M Tris, 0.001M EDTA, pH 8.0). The samples were immediately placed in liquid nitrogen, and kept at -80 °C when stored. We performed phenol-chloroform DNA isolation on the sample stored in SET buffer, following a modified protocol from Sambrook et al. (1989).

Library preparation and sequencing

DNA quality was checked using a combination of a fluorometric (Qubit, Invitrogen), UV absorbance (Nanodrop, Thermo Fisher) and DNA fragment length assays (HS-50 kb fragment kit from AATI, now part of Agilent Inc.). The PacBio library was prepared using the Pacific Biosciences Express library preparation protocol. DNA was fragmented to 35 kb. Size selection of the final library was performed using BluePippin with a 15 kb cut-off. Six single-molecule real-time (SMRT) cells were sequenced using Sequel Polymerase v3.0 and Sequencing chemistry v3.0 on a PacBio RS II instrument. The 10X Genomics Chromium linked-read protocol (10X Genomics Inc) was used to prepare the 10X library, and due to the reed warbler's smaller sized genome, only 0.7 ng/µl of high molecular weight DNA was used as input. A high-throughput chromosome conformation capture (Hi-C) library was constructed using 50 µl of blood, following step 10 and onwards in the Arima Hi-C (Arima Genomics) library protocol for whole blood. Adaptor ligation with Unique dual indexing (Illumina), were chosen to match the indexes from the 10X linked-read library for simultaneous paired-end sequencing (150 bp) on the same lane on an Illumina HiSeq X platform. Both libraries were quality controlled using a Fragment analyzer NGS kit (AATI) and qPCR with the Kapa library quantification kit (Roche) prior to sequencing.

The sequencing was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the South-Eastern Regional Health Authorities.

Genome size estimation and genome assembly

The genome size of the reed warbler was estimated by a k-mer analysis of 10X reads using Jellyfish v. 2.3.0 (Marçais and Kingsford 2011) and Genome Scope v. 1.0 (Vurture et al. 2017), with a k-mer size of 21. The estimated genome size was 1,130,626,830 bp.

We assembled the long-read PacBio sequencing data with FALCON and FALCON-Unzip (falcon-kit 1.5.2 and falcon-unzip 1.3.5) (Chin et al. 2016). Falcon was run with the following parameters: length_cutoff = -1; length_cutoff_pr = 1000; pa_HPCdaligner_option = –v –B128 –M24; pa_daligner_option = –e0.8 –l2000 –k18 –h480 –w8 –s100; ovlp_HPCdaligner_option = –v –B128 –M24; ovlp_daligner_option = –k24 –e.94 –l3000 –h1024 –s100; pa_DBsplit_option = –x500 –s200; ovlp_DBsplit_option = –x500 –s200; falcon_sense_option = –output-multi –min-idt 0.70 –min-cov 3 –max-n-read 200; overlap_filtering_setting = –max-diff 100 –max-cov 100 –min-cov 2. Falcon-unzip was run with default settings. The purge_haplotigs pipeline v. 1.1.0 (Roach et al. 2018) was used to curate the diploid assembly, with -l5, -m35, -h190 for the contig coverage, and -a60 for the purge pipeline. Next, we scaffolded the curated assembly with the 10X reads using Scaff10X v. 4.1 (https://github.com/wtsi-hpag/Scaff10X), and the Hi-C reads using SALSA v. 2.2 (Ghurye et al. 2017). Finally, we polished the assembly (combined with the alternative assembly from Falcon-Unzip), first with PacBio reads using pbmm2 v. 1.2.1, which uses minimap2 (Li 2018) internally (v. 2.17), and then with 10X reads for two rounds with Long Ranger v. 2.2.2 (Marks et al. 2019) and FreeBayes v. 1.3.1 (Garrison and Marth 2012).

Curation

The assembly was decontaminated and manually curated using the gEVAL browser (Chow et al. 2016; Howe et al. 2021), resulting in 521 corrections (breaks, joins and removal of erroneously

duplicated sequence). HiGlass (Kerpedjiev et al. 2018) and PretextView (https://github.com/wtsi-hpag/PretextView) were used to visualize and rearrange the genome using Hi-C data, and PretextSnapshot (https://github.com/wtsi-hpag/PretextSnapshot) was used to generate an image of the Hi-C contact map. The corrections made reduced the total length of scaffolds by 0.5% and the scaffold count by 44.6%, and increased the scaffold N50 by 20.2%. Curation identified and confirmed 29 autosomes and the Z and W chromosomes, to which 98.6% of the assembly sequences were assigned.

Genome quality evaluation

We assessed the quality of the assembly with the assemblathon_stats.pl script (Bradnam et al. 2013) and investigated the completeness of the genome with Benchmarking Universal Single-Copy Orthologs (BUSCO) v. 5.0.0 (Simão et al. 2015), searching for 8338 universal avian single-copy orthologs (aves_odb10).

We aligned the assembly against the great tit (*Parus major*) and the garden warbler (*Sylvia borin*) genome assemblies with minimap2 v. 2.18-r1015 and extracted only alignments longer than 5000 bp. The bundlelinks from circos-tools v. 0.23 was used to merge neighbouring links using default options and a plot was created using circos v. 0.69-8.

Genome annotation

We used a repeat library provided by Alexander Suh called bird_library_25Oct2020 and described in Peona et al. (2020) to softmask repeats in the reed warbler genome assembly. Softmasked genome assemblies for golden eagle (*Aquila chrysaetos*), chicken (*Gallus gallus*), great tit (*Parus major*), Anna's hummingbird (*Calypte anna*), zebra finch (*Taeniopygia guttata*), great reed warbler (*Acrocephalus arundinaceus*), icterine warbler (*Hippolais icterina*), collared flycatcher (*Ficedula albicollis*) and New Caledonian crow (*Corvus moneduloides*) were downloaded from NCBI. The triangle subcommand from Mash v. 2.3 (Ondov et al. 2016) was used to estimate a lower-triangular distance matrix, and a Python script (https://github.com/marbl/Mash/issues/9#issuecomment-

509837201) was used to convert the distance matrix into a full matrix. The full matrix was used as input to RapidNJ v. 2.3.2 (Simonsen et al. 2008) to create a guide tree based on the neighbour-joining method. Cactus v. 1.3.0 (Armstrong et al. 2020) was run with the guide tree and the softmasked genome assemblies as input.

We also downloaded the annotation for chicken, and used it as input to the Comparative Annotation Toolkit (CAT) v. 2.2.1-36-gfc1623d (Fiddes et al. 2018) together with the hierarchical alignment format file from Cactus. Chicken was used as reference genome, reed warbler as the target genome and the AUGUSTUS (Stanke et al. 2008) species parameter was set to 'chicken'. InterProScan v. 5.34-73 (Jones et al. 2014) was run on the predicted proteins to find functional annotations, and DIAMOND v. 2.0.7 (Buchfink et al. 2021) was used to compare the predicted proteins against UniProtKB/Swiss-Prot release 2021_03 (The UniProt Consortium 2021). AGAT v. 0.5.3 (Dainat 2021) was used to generate statistics from the GFF3 file with annotations and to add functional annotations from InterProScan and gene names from UniProtKB/Swiss-Prot. BUSCO v. 5.0.0 was used to assess the completeness of the annotation.

Results and Discussion

Genome assembly

We generated 3,810,665 reads with PacBio, with an average read length of 16 kb at 61x coverage. We further obtained 277,617,608 paired-end reads (2 x 150) with 10X Genomics, and 185,974,525 paired-end reads (2 x 150) with Hi-C, at 83x and 56x coverage, respectively. The final genome assembly was 1.08 Gb in length, and contains 1081 contigs (contig N50 of 13 Mb) and 200 scaffolds (scaffold N50 of 74 Mb) (Table 1).

Genome quality evaluation

The completeness of the assembled genome is high: of the 8338 universal avian single-copy orthologs, we identified 7978 complete BUSCOs (95.7%), including 7920 single-copy (95.0%) and

193 58 duplicated BUSCOs (0.7%). 59 BUSCOs (0.7%) were fragmented, and 301 BUSCOs (3.6%) were
194 missing.

195 The reed warbler genome showed high synteny with the great tit genome, though with some
196 notable differences (Figure 1). The reed warbler chromosome 6 is a fusion of great tit chromosomes 7
197 and 8, and reed warbler chromosome 8 is a fusion of great tit chromosomes 6 and 9. Interestingly,
198 these chromosomes are not fused in the garden warbler genome (Supplementary figure 1), but
199 correspond to the great tit chromosomes. This suggests that the fusions evolved relatively recently,
200 perhaps at the base of the Acrocephalidae branch within Sylvioidea, but further research is needed to
201 determine this. Hi-C contact maps confirm that the chromosomes assembled in the reed warbler
202 genome are unbroken (Supplementary figure 2). Interchromosomal rearrangements are rare in avian
203 evolution (Ellegren 2010; Skinner and Griffin 2012), with some exceptions, such as in the orders
204 Falconiformes (Damas et al. 2017) and Psittaciformes (Furo et al. 2018). In fact, in all or most species
205 of Psittaciformes, chicken chromosomes 6 and 7, and 8 and 9 are fused (Furo et al. 2018; Kretschmer
206 et al. 2018) – the same chromosomes involved in the fusions discovered in the reed warbler genome.
207 We can only speculate about the significance of this without more data. Passeriformes, the sister
208 group of Psittaciformes, exhibit much lower rates of interchromosomal rearrangements, despite being
209 a large, highly diverse order (Kretschmer et al. 2021). There is still a large knowledge gap in the
210 cytogenetics of birds (Degrandi et al. 2020), and more research is needed to determine the rarity of the
211 fusions we discovered in the reed warbler genome.

212 We furthermore confirm the previously identified neo-sex chromosome (Pala et al. 2012;
213 Sigeman et al. 2020b), a fusion between the ancestral chromosome Z and a part of chromosome 4A
214 (according to chromosome naming from the zebra finch). This fusion is thought to have occurred at
215 the base of the Sylvioidea branch (Pala et al. 2012), and is shared with all species of Sylvioidea
216 studied so far (Sigeman et al. 2020b). Figure 1 clearly shows that reed warbler chromosome Z
217 corresponds to great tit chromosome Z, plus a part of great tit chromosome 4A, whereas reed warbler
218 chromosome Z corresponds to garden warbler chromosome Z (Supplementary figure 1).

219

220 **Genome annotation**

The GC content of the reed warbler genome assembly was 41.9%. The total repeat content of the assembly was 10.94%, with LTR elements as the most common type of repeat (4.50%) followed by LINEs (4.11%).

Using the Comparative Annotation Toolkit, based on a whole-genome multiple alignment from Cactus, we predicted 14,645 protein coding genes, with an average Coding DNA Sequence (CDS) length of 1782 bp, and an average intron length of 2918 bp (Table 1). The annotated genes had 97.5% completeness (based on predicted proteins).

Conclusion

In this study, we present the first assembled and annotated genome for the reed warbler *A. scirpaceus*. We have accomplished this through utilizing long read PacBio sequencing, and scaffolding with paired-end 10X and Hi-C reads. In addition to the previously identified autosome-sex chromosome fusion shared by all members of Sylvioidea, we found unequivocal evidence of two novel macrochromosomal fusions in the reed warbler genome. Further research is needed to determine the evolutionary age of these fusions, especially because they are not present in the garden warbler genome, suggesting they are relatively new. This genome will serve as an important resource to increase our knowledge of chromosomal rearrangements in birds, both their prevalence and their significance for avian evolution. Furthermore, the genome will, through the identification of genetic variants and information of the function of associated genes, provide a deeper insight into the evolution of the reed warbler, a bird which will continue to fascinate researchers for years to come.

249 Table 1. Summary statistics of the reed warbler genome assembly and annotation.

Genome Assembly	Estimated genome size	1.13 Gb	
	Guanine and Cytosine content	41.91%	
	N50 length (contig)	13 Mb	
	Longest contig	48 Mb	
	Total length of contigs	1.07 Gb	
	N50 length (scaffold)	74.44 Mb	
	Longest scaffold	153.80 Mb	
	Total length of scaffolds	1.08 Gb	
Transposable elements	Annotation	Percent (%)	Total length
	DNA	0.22	2.35 Mb
	LINE	4.11	44.2 Mb
	SINE	0.09	0.98 Mb
	LTR	4.50	48.4 Mb
	Unknown	0.55	5.9 Mb
	Other (satellites, simple repeats and low complexity)	1.49	16 Mb
	Total	10.94	117.6 Mb
Protein-coding genes	Predicted genes	14,645	
	Average coding sequence length (bp)	1782	
	Average exon length (bp)	284	
	Average intron length (bp)	2918	

250

251

252

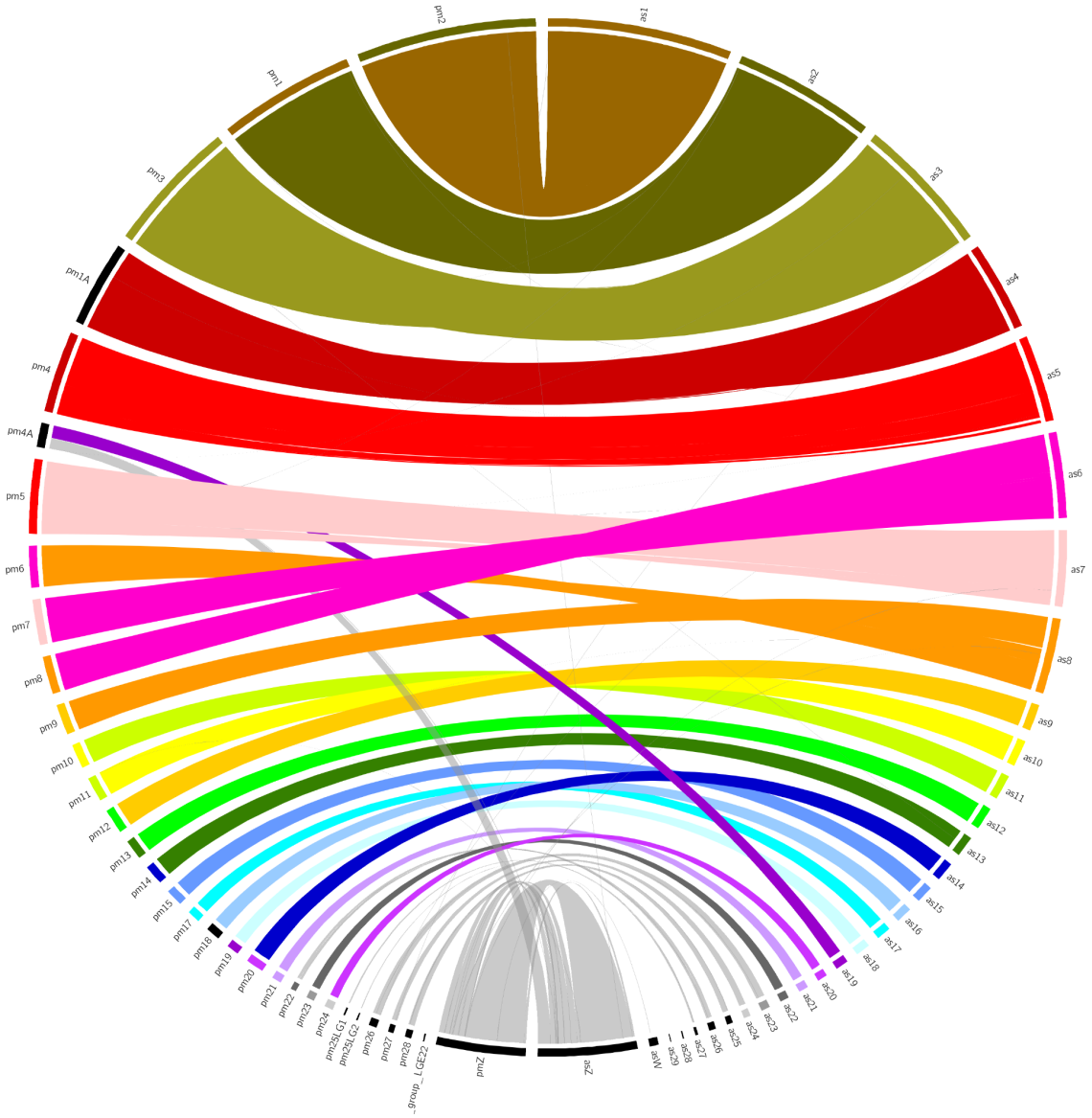

Figure 1. Circos plot showing the synteny between the reed warbler (on the right side, denoted with the prefix as [*Acrocephalus scirpaceus*]) and the great tit (left side, prefix pm [*Parus major*]) genome assemblies. The reed warbler chromosome 6 is a fusion of great tit chromosomes 7 and 8, while reed warbler chromosome 8 is a fusion of great tit chromosomes 6 and 9 (see Hi-C contact maps in supplementary figure 2). The reed warbler chromosome Z corresponds to great tit chromosome Z, and a part of great tit chromosome 4A.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgements

We would like to thank Marjo Saastamoinen, Suvi Sallinen, Paolo Momigliano and the Molecular Ecology and Systematics laboratory in the University of Helsinki for facilitating DNA extraction. We would like to thank Ave Tooming-Klunderud and the Norwegian Sequencing Centre for performing the sequencing. We also thank Pasi Rastas from the HiLIFE BioData Analytics Service unit, for his assistance with preliminary analyses. The computations were performed on resources provided by UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway. This work was supported by Research Council of Norway by grants # 251076 and 300032 to KSJ and a HiLIFE Start-up grant and a University of Helsinki Faculty of Biological and Environmental Sciences travel grant to R.T.

Author Contributions

C.L.C.S., F.E., K.R., K.S.J., O.K.T. and R.T. designed the research. E.K., K.R. and R.T. collected the sample. K.R. extracted DNA. C.L.C.S. and O.K.T. performed the research and/or analysed the data. A.T., J.T., K.H., S.P. and W.C. curated the assembly. C.L.C.S. drafted the manuscript. All authors read and approved the final manuscript.

Data Availability

The reference genome of *Acrocephalus scirpaceus* (bAcrSci1), and the raw sequence data, have been deposited in the European Nucleotide Archive under the BioProject number PRJEB45715.

Literature cited

Armstrong J et al. 2020. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature. 587:246–251.

Both C et al. 2010. Avian population consequences of climate change are most severe for long-distance migrants in seasonal habitats. Proceedings of the Royal Society B. 277:1259–1266.

Bradnam KR et al. 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. GigaScience. 2:10.

Brommer JE, Lehikoinen A, Valkama J. 2012. The breeding ranges of central European and arctic bird species move poleward. PLoS ONE 7(9): e43648.

Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods. 18:366–368.

Chamorro D, Nieto I, Real R, Muñoz AR. 2019. Wintering areas on the move in the face of warmer winters. Ornis Fennica. 96:41–54.

Chin CS et al. 2016. Phased diploid genome assembly with single molecule real-time sequencing. Nature Methods. 13:1050–1054.

Chow W et al. 2016. gEVAL — a web-based browser for evaluating genome assemblies. Bioinformatics. 32:2508–2510.

Dainat J. 2021. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. (Version v0.5.3). Zenodo. https://www.doi.org/10.5281/zenodo.3552717.

Damas J et al. 2017. Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. Genome Research. 27:875–884.

Davies NB, Brooke ML. 1989. An experimental study of co-evolution between the cuckoo, *Cuculus canorus*, and its hosts. I. Host egg discrimination. Journal of Animal Ecology. 58:207–224.

Degrandi TM et al. 2020. Introducing the bird chromosome database: an overview of cytogenetic studies in birds. Cytogenetic and Genome Research. 160:199–205.

Eglington SM et al. 2015. Latitudinal gradients in the productivity of European migrant warblers have not shifted northwards during a period of climate change. Global Ecology and Biogeography. 24:427–436.

Ellegren H. 2010. Evolutionary stasis: the stable chromosomes of birds. Trends in Ecology & Evolution. 25:283–291.

Feng S et al. 2020. Dense sampling of bird diversity increases power of comparative genomics. Nature. 587:252–257.

Fiddes IT et al. 2018. Comparative Annotation Toolkit (CAT) – simultaneous clade and personal genome annotation. Genome Research. 28:1029–1038.

Furo IDO et al. 2018. Chromosome painting in Neotropical long- and short-tailed parrots (Aves, Psittaciformes): phylogeny and proposal for a putative ancestral karyotype for tribe Arini. Genes. 9. https://doi.org/10.3390/genes9100491.

348 Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv
349 preprint arXiv:1207.3907 [q-bio.GN].
350
351 Ghurye J, Pop M, Koren S, Bickhart D, Chin CS. 2017. Scaffolding of long read assemblies using
352 long range contact information. BMC genomics. 18:527.
353
354 Halupka L, Dyrcz A, Borowiec M. 2008. Climate change affects breeding of reed warblers
355 *Acrocephalus scirpaceus*. Journal of Avian Biology. 39:95–100.
356
357 Howe K et al. 2021. Significantly improving the quality of genome assemblies through curation.
358 GigaScience. 10. giaa153.
359
360 Järvinen O, Ulfstrand S. 1980. Species turnover of a continental bird fauna: Northern Europe, 1850-
361 1970. Oecologia. 46:186–195.
362
363 Jones P et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics.
364 30:1236–1240.
365
366 Kerpedjiev P et al. 2018. HiGlass: web-based visual exploration and analysis of genome interaction
367 maps. Genome Biology. 19, 125. https://doi.org/10.1186/s13059-018-1486-1.
368
369 Kralj J, Procházka P, Fainová D, Patzenhauerová H, Tutiš V. 2010. Intraspecific variation in the wing
370 shape and genetic differentiation of reed warblers *Acrocephalus scirpaceus* in Croatia. Acta
371 Ornithologica. 45:51–58.
372
373 Kretschmer R, Ferguson-Smith MA, de Oliveira EHC. 2018. Karyotype evolution in birds: from
374 conventional staining to chromosome painting. Genes. 9:181.
375

376 Kretschmer R et al. 2021. Karyotype evolution and genomic organization of repetitive DNAs in the
377 saffron finch, *Sicalis flaveola* (Passeriformes, Aves). Animals. 11, 1456.
378 https://doi.org/10.3390/ani11051456.
379
380 Laine VN et al. 2016. Evolutionary signals of selection on cognition from the great tit genome and
381 methylome. Nature Communications. 7, 10474. https://doi.org/10.1038/ncomms10474.
382
383 Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34:3094–3100.
384
385 Manel S et al. 2016. Genomic resources and their influence on the detection of the signal of positive
386 selection in genome scans. Molecular Ecology. 25:170–184.
387
388 Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of
389 occurrences of k-mers. Bioinformatics. 27:764–770.
390
391 Marks P et al. 2019. Resolving the full spectrum of human genome variation using linked-reads.
392 Genome Research, 29:635–645.
393
394 Meller K, Piha M, Vähätalo AV, Lehikoinen A. 2018. A positive relationship between spring
395 temperature and productivity in 20 songbird species in the boreal zone. Oecologia. 186:883–893.
396
397 Ondov BD et al. 2016. Mash: fast genome and metagenome distance estimation using MinHash.
398 Genome Biology. 17. https://doi.org/10.1186/s13059-016-0997-x.
399
400 Pala I et al. 2012. Evidence of a neo-sex chromosome in birds. Heredity. 108:264–272.
401

Peona V et al. 2020. The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. bioRxiv preprint doi: https://doi.org/10.1101/2020.07.31.230854.

Procházka P et al. 2011. Low genetic differentiation among reed warbler *Acrocephalus scirpaceus* populations across Europe. Journal of Avian Biology. 42:103–113.

Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 19:460.

Røed, T. 1994. "Rørsanger." In Norsk Fugleatlas, 382–383. Norsk Ornitologisk Forening.

Salewski V, Hochachka WM, Fiedler W. 2010. Global warming and Bergmann's rule: do central European passerines adjust their body size to rising temperatures? Oecologia. 162:247–260.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring. Harbor Laboratory Press, New York.

Schaefer T, Ledebur G, Beier J, Leisler B. 2006. Reproductive responses of two related coexisting songbird species to environmental changes: global warming, competition, and population sizes. Journal of Ornithology. 147:47–56.

Shafer ABA et al. 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods in Ecology and Evolution. 8:907–917.

Sigeman H et al. 2020a. Genomics of an avian neo-sex chromosome reveals the evolutionary dynamics of recombination suppression and sex-linked genes. bioRxiv preprint doi: https://doi.org/10.1101/2020.09.25.314088.

Sigeman H, Ponnikas S, Hansson B. 2020b. Whole-genome analysis across 10 songbird families within Sylvioidea reveals a novel autosome-sex chromosome fusion. Biology Letters. 16.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31:3210–3212.

Simonsen M, Mailund T, Pedersen CNS. 2008. Rapid Neighbour-Joining. In: Crandall KA, Lagergren J (eds) Algorithms in Bioinformatics. WABI 2008. Lecture Notes in Computer Science, vol 5251. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-87361-7_10.

Skinner BM, Griffin DK. 2012. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. Heredity. 108:37–41.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 24:637–644.

Stokke B et al. 2018. Characteristics determining host suitability for a generalist parasite. Scientific Reports. 8.

Stolt BO. 1999. The Swedish reed warbler *Acrocephalus scirpaceus* population estimated by a capture-recapture technique. Ornis Svecica. 9:35–46.

Sætre CL et al. 2017. Rapid adaptive phenotypic change following colonization of a newly restored habitat. Nature Communications. 8:14159.

457 The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids
458 Research. 49:D480–D489.
459
460 Thorogood R, Davies NB. 2013. Reed warbler hosts fine-tune their defenses to track three decades of
461 cuckoo decline. Evolution. 67:3545–3555.
462
463 Thorogood R, Spottiswoode CN, Portugal SJ, Gloag R. 2019. The coevolutionary biology of brood
464 parasitism: a call for integration. Philosophical Transactions of the Royal Society B. 374.
465
466 Vickery JA et al. 2014. The decline of Afro-Palaearctic migrants and an assessment of potential
467 causes. Ibis. 156:1–22.
468
469 Vurture GW et al. 2017. GenomeScope: fast reference-free genome profiling from short reads.
470 Bioinformatics. 33:2202–2204.
471