Chromosome-Level Genome Assembly of the Common Chaffinch (Aves: Fringilla coelebs): A Valuable Resource for Evolutionary Biology

María Recuerda 1,*,†, Joel Vizueta 1,2,†, Cristian Cuevas-Caballé 2, Guillermo Blanco 1, Julio Rozas 2, and Borja Milá 1

1National Museum of Natural Sciences, Spanish National Research Council (CSIC), Madrid, Spain
2Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain

†These authors contributed equally to this work.
*Corresponding author: E-mail: mariarecuerdacarrasco@gmail.com.
Accepted: 16 February 2021

Abstract

The common chaffinch, Fringilla coelebs, is one of the most common, widespread, and well-studied passerines in Europe, with a broad distribution encompassing Western Europe and parts of Asia, North Africa, and the Macaronesian archipelagos. We present a high-quality genome assembly of the common chaffinch generated using Illumina shotgun sequencing in combination with Chicago and Hi-C libraries. The final genome is a 994.87-Mb chromosome-level assembly, with 98% of the sequence data located in chromosome scaffolds and a N50 statistic of 69.73 Mb. Our genome assembly shows high completeness, with a complete BUSCO score of 93.9% using the avian data set. Around 7.8% of the genome contains interspersed repetitive elements. The structural annotation yielded 17,703 genes, 86.5% of which have a functional annotation, including 7,827 complete universal single-copy orthologs out of 8,338 genes represented in the BUSCO avian data set. This new annotated genome assembly will be a valuable resource as a reference for comparative and population genomic analyses of passerine, avian, and vertebrate evolution.

Key words: common chaffinch, Fringilla coelebs, reference genome, whole genome assembly.

Significance

High-quality reference genomes of wild, nonmodel species are very useful tools to understand how organisms evolve. If genomes are annotated, so that the specific genes are identified, we can make progress toward associating specific physical or behavioral traits with the genes that code for them, and thus further understand the evolutionary process. Here, we provide a high-quality, annotated genome of the common chaffinch, a common and widespread Eurasian finch, that will be a useful resource in studies related to evolution, phylogenomics, biogeography, and adaptation genomics, among others.

Introduction

The decreasing costs of DNA sequencing, along with advances in computational genomics, are promoting a rapid increase in the availability of high-quality reference genomes of nonmodel species, which greatly improves our capacity to address a range of biological questions from a genomic perspective. Among them, the correct annotation of protein-coding genes in whole genomes allows to identify new genes involved in the process of evolutionary adaptation and provides a better understanding of the evolutionary mechanisms involved in the speciation process. Avian genomes are

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Genome Biol. Evol. 13(4) doi:10.1093/gbe/evab034 Advance Access publication 22 February 2021 1
particularly suited for studying the molecular basis of speciation as they have a relatively simple architecture and are among the smallest within amniotes, ranging from 0.91 to 1.3 Gb (Gregory 2002). In the last decade, the number of bird reference genomes has increased dramatically (e.g. Dalloul et al. 2010; Warren et al. 2010; Zhang et al. 2012; Jarvis et al. 2014; Poelstra et al. 2014; Frankl-Vilches et al. 2015; Friis et al. 2018; Louha et al. 2020; Peñalba et al. 2020; Ducrest et al. 2020, Wang et al. 2020), providing major scientific breakthroughs in phylogenetics (Alström et al. 2018; Ducrest et al. 2020, Wang et al. 2020), comparative genomics (Alström et al. 2018; Braun et al. 2019; Jarvis et al. 2015), comparative genomics (Zhang et al. 2014, Feng et al. 2020), adaptation genomics (Wirthlin et al. 2014; Lawson and Petren 2017), and genomic architecture (Poelstra et al. 2014; Vijay et al. 2016), among others. Moreover, the Ten-Thousand Bird Genomes (B10K) consortium has generated and analyzed over 300 avian genomes from 92.4% of bird families, providing an unprecedented genomic resource for avian comparative studies (Zhang et al. 2015, Feng et al. 2020).

The common chaffinch (Aves, Passeriformes, Fringillidae, Fringilla coelebs) is a widely distributed species, ranging from Eurasia to the north of Africa, and has colonized three Macaronesian archipelagos in the Atlantic Ocean (Azores, Madeira, and the Canary Islands) (Collar et al. 2020). With about 15 currently recognized subspecies, the common chaffinch is an ideal system for testing hypotheses on the evolutionary process given its distribution across the continent and the colonization of several oceanic islands, recognized as excellent natural laboratories for studying evolution (Brown et al. 2013). Island systems have inspired the development of biogeographical theories (MacArthur and Wilson 1967) and are of central importance for understanding the role of area and isolation in colonization, extinction, and speciation rates (Valente et al. 2020), which are processes influencing global patterns of species richness (Losos and Schluter 2000). Species that have colonized insular environments, like the common chaffinch, are also excellent systems for the study of demographic events, such as bottlenecks leading to small effective population size (Ne) (Leroy et al. 2021), or the roles of drift and selection in the divergence process (Barton 1996). The common chaffinch has been intensively studied using molecular tools, so that the availability of a reference genome represents a valuable resource to improve our understanding of avian evolution, biogeography, and demography (Illera et al. 2018).

Results and Discussion

Assembly and Quality Control

The total length obtained by the HiRise software for the common chaffinch assembly was 994.87 Mb. Nevertheless, the estimate from k-mer metrics is 1.2 Gb. The discrepancy between these estimates could be caused by the presence of repetitive elements given the assembly strategy used, which could have been improved including long-read sequencing technologies. This final assembly consists of 3,255 scaffolds, 3,239 over 1 kb, and an N50 of 69.73 Mb (see supplementary table S1, Supplementary Material online) with a sequence coverage of 249 x. The use of Chicago and Hi-C libraries provided a clear improvement in quality by increasing 917 times the scaffold N50, reducing the number of scaffolds from 38,666 to 3,255 (Supplementary table S1, see supplementary methods for details, Supplementary Material online). In fact, 98% of the total genome sequence maps in the 30 described chromosomes.

The chaffinch genome showed high synteny with the zebra finch genome (fig. 1), evidencing the completeness of the assembly, with all micro-chromosomes and the Z chromosome present in the assembly. In addition, the alignment between these genomes suggests the presence of several inversions in chromosomes 1, 1A, 2, 3, 5, 7, 8, and 9. Several studies have documented that inversions are very common in birds (Aslam et al. 2010; Völker et al. 2010; Skinner and Griffin 2012; Zhang et al. 2014). For instance, Hooper and Price (2017) identified 319 inversions on the 9 largest autosomes combined in 81 independent clades. No putative contaminations were detected and 89.6% of the reads were mapped in the genome assembly (see Supplementary fig. S1, Supplementary Material online). The mean GC content of the assembly was 41.86% (± 11 SD). The common chaffinch genome assembly included 7,832 complete copies (93.9%) out of the 8,338 BUSCO data set from avian genomes, among which 7,816 were single-copy orthologs and 16 were duplicated. Only 1.8% of the gene models were fragmented, and 4.3% were missing in the genome. These few missing gene models could represent divergent or lost genes in our species, but also could be related with putative errors during the assembly process or missing data.

Repetitive Regions

Overall, 7.82% of the genome assembly are repeats (~78 Mb), of which 85.4% are transposable elements (TEs). The most abundant TEs are LINEs (53.5%) followed byLTR (29.4%), DNA elements (4.1%), and SINEs (1.4%), with the remaining 11.6% unclassified. The rest of repeats (14.6%) contained simple repeats (75.4%), low complexity repeats (18.5%), satellites (4.2%) and small RNA (1.9%) (see supplementary table S2, Supplementary Material online). The number of repetitive regions is within the expected range in birds, which is at 4–10% of the genome (Zhang et al. 2014).

A total of 111,076 microsatellites, with motif length ranging between 2 and 20 bp, were identified in the common chaffinch genome (see supplementary fig. S3, Supplementary Material online; their genomic locations are shown in supplementary file S1 in the Figshare repository). The most common k-mer sizes conforming the microsatellites were 2 (68.2%), 3 (15.9%) and 4 (8.2%) (see supplementary...
The most common length of the microsatellites was 10 bp (40.4%), followed by 12 bp (13%) and 15 bp (8.8%) (see supplementary file S1 for the length distribution of microsatellites). In addition, the number of microsatellites was positively correlated with the sequence length (Supplementary fig. S3, Supplementary Material online; see supplementary file S1 for the frequency of occurrence in every scaffold).

**Gene Annotation and Function Prediction**

Our annotation pipeline combining both de novo and homology-based predictions inferred 21,831 proteins encoded by 17,703 genes in the common chaffinch genome with a mean length of 15,818 bp (Table 1). The common chaffinch genome annotation (see supplementary file S2 in Figshare) included 7,850 complete copies (94.2%) out of the 8,338 of BUSCO avian data set used, retrieving all expected copies with a slight increase from that estimated in the unannotated genome (see above). Among the complete BUSCO genes, 7,827 were single-copy orthologs (99.7%) and 23 were duplicated (0.3%). Around 1.9% (162) of the gene models were fragmented and 3.9% showed no significant matches (326).

Over all predicted proteins, 19,458 (89.1%) provided positive BLASTP hits against the Uniprot SwissProt database, and...
In addition, InterproScan identified 18,551 (85%) specific protein-domain signatures in the predicted peptides. The combination of the annotation from these databases allowed assigning a functional annotation with GO terms to 19,425 proteins (89%) assigned to 15,309 genes (86.5%; supplementary file S3 in Figshare).

**tRNAs and Other Noncoding RNA Prediction**

The search by tRNAscan-SE (supplementary file S4 in Figshare) identified 325 tRNAs in the common chaffinch genome, of which 167 decode for the standard twenty amino acids. Among all the tRNAs detected, 131 presented low scores and therefore were categorized as pseudogenes (i.e., lacking tRNA-like secondary structures). There were no suppressor tRNAs, 1 had undetermined isotopes, 25 were chimeric and 15 included introns within their sequences. One of the tRNAs was predicted to code for selenocysteine (sequences and structures of the predicted tRNAs are available in File S5 in Figshare). In addition, the search against both tRNA databases (GtRNAdb and tRNAdb) yielded positive results in many other species, suggesting that tRNA prediction in our assembly was correct. Moreover, our searches using Infernal identified 354 ncRNAs, which were classified as follows: 39 CREs, 2 Ribozymes, 7 Gene, 140 miRNAs, 126 snoRNAs, 18 snRNAs, 5 rRNAs, and 17 lncRNAs (File S6 in Figshare).

19,617 (89.9%) against the annotated proteins from the zebra finch genome. In addition, InterproScan identified 18,551 (85%) specific protein-domain signatures in the predicted peptides. The combination of the annotation from these databases allowed assigning a functional annotation with GO terms to 19,425 proteins (89%) assigned to 15,309 genes (86.5%; supplementary file S3 in Figshare).

### Table 1

|                   | F. coelebs | M. melodia | T. guttata | F. albicollis | M. vitellinus | G. fortis |
|-------------------|------------|------------|------------|---------------|---------------|-----------|
| Number of genes   | 17,703     | 15,086     | 17,561     | 16,763        | 18,976        | 14,399    |
| Mean gene length (bp) | 15,818     | 14,457     | 26,458     | 31,394        | 27,847        | 30,164    |
| Number of CDSs    | 17,703     | 15,086     | 17,561     | 16,763        | 18,976        | 14,399    |
| Mean CDS length (bp) | 1,679      | 1,325      | 1,677      | 1,942         | 1,929         | 1,766     |
| Number of exons   | 221,872    | 131,940    | 171,767    | 189,043       | 190,390       | 164,721   |
| Mean exon length (bp) | 165        | 153        | 255        | 253           | 264           | 195       |
| Mean number of exons/gene | 10.16      | 8.67       | 10.25      | 12.22         | 11.51         | 11.41     |
| Number of introns | 200,041    | 116,724    | 153,909    | 171,236       | 171,089       | 149,563   |
| Mean intron length (bp) | 1,902      | 1,695      | 2,930      | 3,257         | 3,294         | 3,294     |
| Total proteins    | 21,831     |            |            |               |               |           |
| ncRNA             |            |            |            |               |               |           |
| tRNA              | 325        | 267        | 184        | 179           |               |           |
| miRNA             | 140        | 166        | 302        | 510           |               |           |
| snRNA             | 18         | 16         | 44         | 32            |               |           |
| snoRNA            | 126        | 154        | 241        | 199           |               |           |
| rRNA              | 0          | 8          | 100        | 22            |               |           |
| lncRNA            | 17         | 20         | 908        | 1473          |               |           |

19,617 (89.9%) against the annotated proteins from the zebra finch genome. In addition, InterproScan identified 18,551 (85%) specific protein-domain signatures in the predicted peptides. The combination of the annotation from these databases allowed assigning a functional annotation with GO terms to 19,425 proteins (89%) assigned to 15,309 genes (86.5%; supplementary file S3 in Figshare).

**Conclusions**

We provide here a high-quality assembly for the common chaffinch, a valuable resource as a reference genome to address a range of biological questions from a genomic perspective. Moreover, our annotation provides useful information to detect candidate genes involved in adaptation and divergence processes. The combination of the Chicago and shotgun sequencing with the HiRise assembly approach lead to a highly contiguous chromosome-level genome assembly. The genome assembly size was 994.87 Mb, with the 30 chromosomes accounting for 98% of it. Although the expected length of the genome was 1.2 Gb, closer to those obtained in other avian species by flow cytometry (Gregory 2002), the BUSCO analyses showed that both the assembly and structural annotation encode 93.9% and 94.2% complete copies out of the 8,338 orthologous conserved genes in avian species, respectively. This discrepancy of the genome size could be caused by the absence of large repetitive elements in the assembly. The structural annotation predicted 17,703 coding genes, with most of them (86.5%) assigned to functional annotation and GO terms.

**Materials and Methods**

**Sample Collection and Genome Assembly**

A blood sample was extracted from a common chaffinch female captured in Torreiglesias, Segovia, Spain, in 2017 and frozen immediately in liquid nitrogen. The sample was
Identification of Repetitive Regions and Gene Annotation

Repetitive regions were identified and masked prior to gene prediction. First, repeats were modelled ab initio using Repeat Modeler 1.0.11 (Smit and Hubley 2019) in scaffolds longer than 100 Kb with default options. The repeats obtained were merged with known bird repeat libraries from the RepBase database (RepBase-Version181026) (Bao et al. 2015), Dfam_Consensus-20181026 and repeats from the zebra finch (obtained from B10K). The resulting repeat library was compared against the complete assembly with Repeat Masker 4.0.7 (Smit et al. 2015) and the identified regions were soft-masked. For the identification and description of microsatellites in the common chaffinch genome assembly we used GMATA v.2.01 (Wang and Wang 2016), with sequence motif length between 2 and 20bp.

Gene prediction was conducted with BRAKER v2.1.5 (Hoff et al. 2016) and GeMoMa v1.7.1 (Keilwagen et al. 2016, 2018). First, the conserved orthologous genes from BUSCO Aves_odb10 were used as proteins from short evolutionary distance to train Augustus (Gremme et al. 2005; Stanke et al. 2018). First, the conserved orthologous genes from BUSCO Aves_odb10 were used as proteins from short evolutionary distance to train Augustus (Gremme et al. 2005; Stanke et al. 2018). The functional annotation, including Gene Ontology terms, was integrated from all searches providing a curated set of chaffinch coding genes (see supplementary methods, Supplementary Material online, for details).

Noncoding RNA Prediction and Identification

For the prediction and functional classification of Transfer RNAs (tRNAs) in the common chaffinch genome we used tRNAscan-SE v2.0 (Lowe and Chan 2016). The tRNA search across the genome and the identification of ncRNA (noncoding RNA) homologues was conducted using the software package Infernal v.1.1.1 (Nawrocki 2014)(see supplementary methods, Supplementary Material online, for details). For comparative purposes, we added our results to those from Louha et al. (2020), which compared different genome assemblies of avian species.

Supplementary Material

Supplementary data are available at Genome Biology and Evolution online.

Acknowledgments

This research was supported by the Spanish Ministry of Economy and Competitiveness (CGL2015-66381P to B.M. and G.B.) and the Spanish Ministry of Science and Innovation (PGC2018-098897-B-I00 from to B.M.). M.R. was supported by a doctoral fellowship from the Spanish Ministry of Education, Culture, and Sport (FPU16/05724).

Data Availability

The chaffinch genome assembly has been deposited at NCBI under BioProject PRJNA674347 with accession number JADKPM000000000, the raw data are available at SRA NCBI database with accession numbers SRR12998620–SRR12998622, and the annotation and all described data sets are publicly accessible in Figshare (https://doi.org/10.6084/m9.figshare.13296122.v3).

Literature Cited

Alström P, et al. 2018. Complete species-level phylogeny of the leaf warbler (Aves: Phylloscopidae) radiation. Mol Phylogenet Evol. 126:141–152.

Aslam ML, et al. 2010. A SNP based linkage map of the turkey genome reveals multiple intrachromosomal rearrangements between the turkey and chicken genomes. BMC Genomics 11(1):647–611.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 6:11.

Barton NH. 1996. Natural selection and random genetic drift as causes of evolution on islands. Philos Trans R Soc Lond B Biol Sci. 351(1341):785–795.

Braun EL, Cracraft J, Houde P. 2019. Resolving the avian tree of life from top to bottom: the promise and potential boundaries of the phylogenomic era. In: Kraus R, editor. Avian genomics in ecology and evolution. Cham (Switzerland): Springer. p. 151–210.

Brown RM, et al. 2013. Evolutionary processes of diversification in a model island archipelago. Annu Rev Ecol Evol Syst. 44(1):411–435.

Collar N, Newton I, Bonan A. 2020. Finches (Fringillidae). In: del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E, editors. Handbook of the birds of the world alive. Barcelona (Spain): Lynx Edicions. Available from: https://www.hbw.com/node/52376 [accessed 2020 Jan 15].
Ducrest AL, et al. 2020. New genome assembly of the barn owl (Tyto alba alba). Ecol Evol. 10(5):2284–2298.

Feng S, et al. 2020. Dense sampling of bird diversity increases power of comparative genomics. Nature 587(7833):252–257.

Franki-Vilches C, et al. 2015. Using the canary genome to decipher the evolution of hormone-sensitive gene regulation in seasonal singing birds. Genome Biol. 16(1):19.

Friis G, et al. 2018. Genome-wide signals of drift and local adaptation during rapid lineage divergence in a songbird. Mol Ecol. 27(24):5137–5153.

Gregory TR. 2002. Animal genome size database. Available from: http://www.genomesize.com.

Gremme G, Brendel V, Sparks ME, Kurtz S. 2005. Engineering a software tool for gene structure prediction in higher organisms. Inf Softw Technol. 47(15):965–978.

Guy L, Roat Kultima J, Andersson SG. 2010. genoPlotR: comparative gene and genome visualization in R. Bioinformatics 26(18):2334–2335.

Hillier LW, et al. 2014. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 423:695–777.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32(5):767–769.

Hillier LW, et al. 2014. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 423:695–777.

Hopper DM, Price TD. 2017. Chromosomal inversion differences correlate with range overlap in passerine birds. Nat Ecol Evol. 1(10):1526–1534.

Illera JC, et al. 2018. Acoustic, genetic, and morphological analyses of the Canarian common chaffinch complex Fringilla coelebs ssp. reveals cryptic diversification. J Avian Biol. 49(12):1–12.

Jervis ED, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346(6215):1320–1331.

Jervis ED, et al.; Avian Phylogenomics Consortium. 2015. Phylogenomic analyses data of the avian phylogenomics project. Gigascience 4(1):s13742-014.

Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.

Keilwagen J, et al. 2016. Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. 44(9):e89.

Keilwagen J, Hartung F, Paulini M, Twardzik SO, Grau J. 2018. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinformatics 19(1):189.

Lawson LP, Petren K. 2017. The adaptive genomic landscape of beak morphology in Darwin’s finches. Mol Ecol. 26(19):4978–4989.

Leroy T, et al. 2021. Forthcoming. Endemic island songbirds as windows into evolution in small effective population sizes. Curr Biol. Available from: 10.1016/j.cub.2020.12.040.

Losos JB, Schluter D. 2000. Analysis of an evolutionary species-area relationship. Nature 408(6814):847–850.

Louha S, Ray DA, Winkler K, Glenn TC. 2020. A high-quality genome assembly of the North American Song Sparrow, Melospiza melodia. G3 (Bethesda) 10(4):1159–1166.

Lowry TM, Chan PP. 2016. TRNAScan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44(W1):W54–W57.

MacArthur RH, Wilson EO. 1967. The theory of island biogeography. Princeton (NJ): Princeton University Press. 203 p.

Nawrocki EP. 2014. Annotating functional RNAs in genomes using Infernal. In: Gorodkin J, Russo W, editors. RNA sequence, structure, and function: computational and bioinformatic methods. Totowa (NJ): Humana Press. p. 163–197.

Peñalba JV, et al. 2020. Genome of an iconic Australian bird: High-quality assembly and linkage map of the superb fairy-wren (Malurus cyaneus). Mol Ecol Resour. 20(2):560–578.

Poelstra JV, et al. 2014. The genomic landscape underlying phenotypic integration in the face of gene flow in crows. Science 346(6190):1410–1414.

Putnam NH, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 26(3):342–350.

Seppøy M, Männi M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editor. Gene prediction. New York: Humana. p. 227–245.

Skinner BM, Griffin DK. 2012. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. Heredity (Edinb). 108(1):37–41.

Smit A, Hubley R. 2019. RepeatModeler-1.0. 11. Institute for Systems Biology. Available from: http://www.repeatmasker.org/RepeatModeler/.

Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. 2013–2015. Available from: http://www.repeatmasker.org.

Stanker M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 7:62.

UniProt Consortium. 2014. UniProt: a hub for protein information. Nucleic Acids Res. 43:D204–D212.

Valente L, et al. 2020. A simple dynamic model explains the diversity of island birds worldwide. Nature 579(7797):92–96.

Vijay N, et al. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nat Commun. 7(1):13195.

Völker M, et al. 2010. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. Genome Res. 20(4):503–511.

Wang X, Wang L. 2016. GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. Front Plant Sci. 7:1350.

Wang W, et al. 2020. First de novo whole genome sequencing and assembly of the barred-head goose. PeerJ. 8:e8914.

Warren WC, et al. 2010. The genome of a songbird. Nature 464(7289):757–762.

Wirthlin M, Lovell PV, Jarvis ED, Mello CV. 2014. Comparative genomics reveals molecular features unique to the songbird lineage. BMC Genomics 15:1082.

Zhang G, et al. 2015. Genomics: bird sequencing project takes off. Nature 522(7554):34.

Zhang G, et al.; Avian Genome Consortium. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346(6215):1311–1320.

Zhang G, Parker P, Li B, Li H, Wang J. 2012. The genome of Darwin’s finch (Geospiza fortis). GigaScience. Database. Available from: 10.5524/100040.

Associate editor: Bonnie Fraser