A reference genome for the nectar-robbing Black-throated Flowerpiercer (*Diglossa brunneiventris*)

Anna E. Hiller, Robb T. Brumfield, and Brant C. Faircloth (iD) *

Museum of Natural Science and Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA

*Corresponding author: 282 Life Sciences Building, Baton Rouge, LA 70806, USA. Email: brant@lsu.edu

Abstract

Black-throated Flowerpiercers (*Diglossa brunneiventris*) are one species representing a phenotypically specialized group of tanagers (Thraupidae) that have hooked bills which allow them to feed by stealing nectar from the base of flowers. Members of the genus are widely distributed in montane regions from Mexico to northern Argentina, and previous studies of *Diglossa* have focused on their systematics, phylogenetics, and interesting natural history. Despite numerous studies of species within the genus, no genome assembly exists to represent these nectivorous tanagers. We described the assembly of a genome sequence representing a museum-vouchered, wild, female *D. brunneiventris* collected in Peru. By combining Pacific Biosciences Sequel long-read technology with 10× linked-read and reference-based scaffolding, we produced a 1.08 Gbp pseudochromosomal assembly including 600 scaffolds with a scaffold N50 of 67.3 Mbp, a scaffold L50 of 6, and a BUSCO completeness score of 95%. This new assembly improves representation of the diverse species that comprise the tanagers, improves on scaffold lengths and contiguity when compared to existing genomic resources for tanagers, and provides another avenue of research into the genetic basis of adaptations common to a nectivorous lifestyle among vertebrates.

Keywords: Thraupidae; *Diglossa brunneiventris*; Flowerpiercers; long-read sequencing; genome assembly

Introduction

Flowerpiercers in the genus *Diglossa* are a phenotypically specialized group of tanagers known for their unique adaptation for feeding—a hooked bill that they use to pierce the base of flowers—and an associated behavior called nectar-robbing (Inouye 1983; Bock 1985; Mauck and Burns 2009). The 18 species of flowerpiercer are phenotypically diverse and show variation in body size, bill shape, plumage color and patterning, and eye color (Zimmer 1929; Isler and Isler 1987; del Hoyo *et al.* 1992; Mauck and Burns 2009; Figure 1). *Diglossa* species vary in their degree of sexual dichromatism, and there are several examples of phenotypic convergence in the group (*e.g.*, "leapfrog" variation where disjunct populations are phenotypically similar; Zimmer 1929; Vuilleumier 1969; Remsen 1984). *Diglossa* are widely distributed in montane regions from Central Mexico to northern Argentina with peak species diversity in the equatorial Andes (Vuilleumier 1969; Isler and Isler 1987; del Hoyo *et al.* 1992; Figure 1). Two flowerpiercers, *D. gloriossisima* and *D. venezuelensis*, are species of conservation concern (del Hoyo *et al.* 1992; IUCN 2020).

Previous studies of *Diglossa* have focused on their systematics (Zimmer 1929; Zimmer *et al.* 1952; Bock 1985; Graves 1990), phylogenetics (Hackett 1995; Mauck and Burns 2009), biogeography (Chapman 1926; Vuilleumier 1969; Graves 1982, 1985, 1988; Navas Berdugo 2008; Cuervo 2013; Gutiérrez-Zuluaga *et al.* 2021), natural history (Lyon and Chadek 1971; Colwell *et al.* 1974; Graves 1982; Stiles *et al.* 1992; Rojas-Nossa 2007; Solano-Ugalde and Brinkhuizen 2012; Vaicenbacher *et al.* 2014; Hazlehurst and Karubian 2016), and

relationships to other taxonomic groups within the Thraupidae (Burns 1997; Burns *et al.* 2003, 2014), one of the most species-rich bird families. Despite the ecological and evolutionary uniqueness of *Diglossa*, there are no genome assemblies available for these nectivorous tanagers, and only four genome assemblies (Supplementary Table S1) represent the 98 genera and 371 named species within the tanager family (Dickinson and Christidis 2014).

Here, we announce the assembly of a genome from a wild caught, museum-vouchered (Buckner *et al.* 2021), female Black-throated Flowerpiercer (*Diglossa brunneiventris*) collected in the Department of Lima, Peru. We selected this species among other *Diglossa* because of its role in foundational studies of biogeography (Chapman 1926; Hazzi *et al.* 2018), ecogeographic variation in plumage and body size (Remsen 1984; Graves 1991), secondary contact and hybridization (Graves 1982), and the genetics of high-altitude blood physiology (Natarajan *et al.* 2016). A reference assembly for the Black-throated Flowerpiercer will enable related studies on the role of genetic architecture in reproductive isolation and adaptation, the genetics of color and patterning in birds, and the impact of hybridization on speciation processes.
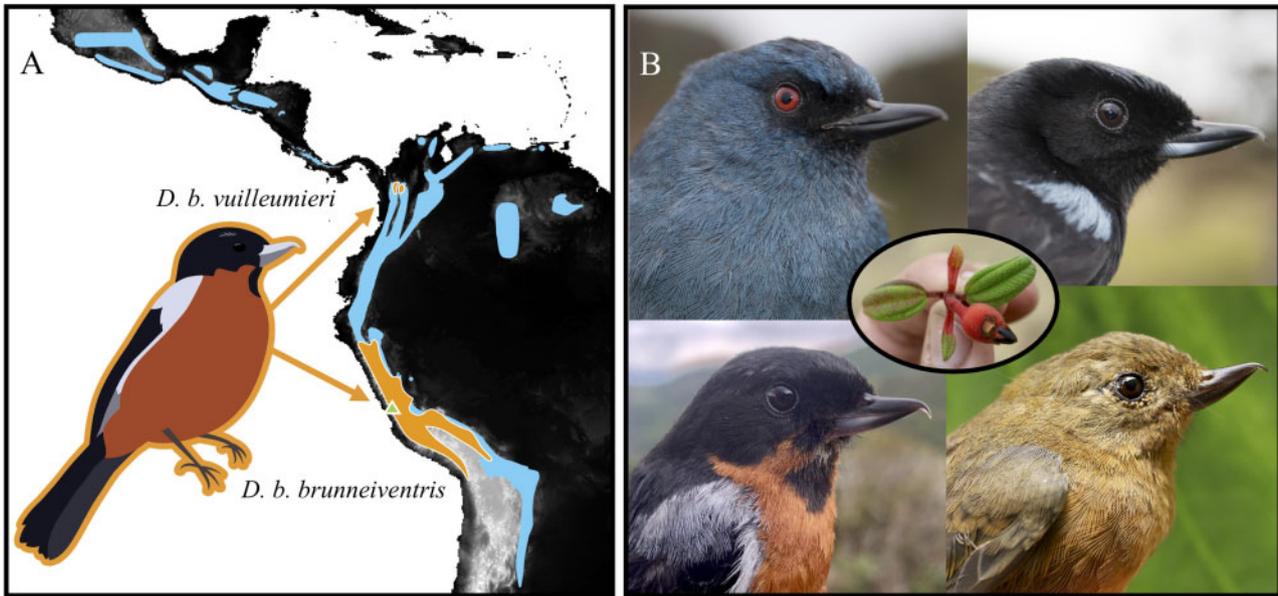
Materials and methods
Specimen collection and high-molecular-weight DNA extraction

We collected blood and other tissues from a wild, female bird captured 6.3 km east-southeast of San Mateo, Department of

Figure 1 (A) A range map for the *Diglossa* flowerpiercers with the distribution of the genus shown in blue, and a range map for both subspecies of *D. brunneiventris* shown in orange accompanied by an illustration of the focal species (adapted from BirdLife International and Handbook of the Birds of the World 2019). Locality for the reference genome individual *D. brunneiventris* is indicated by the green triangle. (B) Examples of the phenotypes observed among *Diglossa* flowerpiercers. Top left: *D. caerulescens*, top right: *D. humeralis* (photos by Diego Cueva), bottom left: *D. brunneiventris*, bottom right: *D. albilatera* female (photos by Anna Hiller). The central photo shows a flower of *Brachyotum* sp. with a "pierced" corolla (photo by Anna Hiller).

Lima, Peru (11.775 S, 76.245 W, 3900 m) during October 2018 under permits from the Servicio Nacional Forestal y de Fauna Silvestre (SERFOR; collecting permits 203-2015-SERFOR-DGGSPFFS, 222-2015-SERFOR-DGGSPFFS, export permit 003336 SERFOR) and following LSU's Institutional Animal Care and Use Committee (IACUC) protocol number 18-054. We selected this population because it was far from any potential zone of secondary contact and/or hybridization with other *Diglossa* species (Graves 1982). We stored blood in a cryotube with 95% ethanol, placed the blood in liquid nitrogen, and flash froze all remaining tissues (liver, pectoral and cardiac muscle, ovary, eyes, and brain) in liquid nitrogen. Following tissue collection, we prepared a specimen for the LSU Museum of Natural Science (LSUMNS) Collection of Birds (LSUMZ 227748), and we deposited tissue samples from this specimen into the LSUMNS Collection of Genetic Resources (LSUMZ B-95988) as well as the tissue collection of the Centro de Ornitología y Biodiversidad (CORBIDI) in Lima, Peru.

To minimize contaminating mitochondrial sequences, we extracted high-molecular weight (HMW) DNA from the blood we collected using a modified phenol-chloroform protocol (Supplementary File S1). After extraction, we performed a bead cleanup using an inexpensive alternative to commercial SPRI reagents (Glenn *et al.* 2019) to remove remaining impurities and short DNA fragments. We quantified the extracted DNA using a Qubit fluorometer (Thermo Fisher Scientific, Inc.) with the broad spectrum kit, and we determined the size distribution of the DNA by analyzing a 100 ng/μL aliquot using an Agilent TapeStation and the Genomic DNA ScreenTape (Kansas University Medical Center, Kansas City, KS, USA). We shipped two aliquots totaling 8.5 μg HMW DNA on dry ice to the Georgia Genomics Facility (GGF; Athens, GA, USA) for library preparation.

Library preparation and sequencing

We were interested in comparing an assembly from Pacific Biosciences (hereafter PacBio) Sequel long-reads technology to an assembly from 10× Genomics Chromium linked-reads technology, so we had the GGF staff prepare: (1) a single 10× linked-read library from <0.5 μg HMW DNA using the Chromium Genome Library Kit (v2) and Chromium Genome Chip Kit (v2), and (2) a single SMRTbell library from 5.0 μg input DNA using the SMRTbell Express Template Prep Kit (v1.0) following the guidelines for preparing 20 kb templates.

After preparing the 10× linked-read library, GGF staff determined the insert size of the library using an Agilent Fragment Analyzer and quantified the library using a Qubit Fluorometer. We sequenced the library using a partial Illumina HiSeq X (Novogene, Inc.) lane, targeting 250 million paired-end, 150 bp reads to achieve ~56X coverage assuming a final assembly of ~1.1 Gbp (Feng *et al.* 2020). The libraries sharing the remainder of the HiSeq X lane had nonoverlapping 10× barcodes.

After preparing the SMRTbell Express 20 kb library and performing quality control steps following the manufacturer's instructions, GGF staff sequenced an aliquot of the resulting library using one SMRT Cell 1M (v3) on a PacBio Sequel System with software version 6.0.x and chemistry version 3.0 (movie time: 600 m; immobilization time: 120 m). Because the initial SMRT Cell performed nominally, we had GGF staff sequence remaining aliquots of the same library on two additional SMRT Cells using the same run parameters in order to approach ~30X coverage.

Genome assembly

We assembled all 10× linked-read sequencing data (–maxreads="all") with the Supernova (v2.2.1) software on a compute node using 24 compute cores and 745 GB RAM. After successfully running the Supernova pipeline, we output all versions of the assembly, although subsequent comparisons were performed using the pseudohaploid version of the assembly. Specifically, we computed assembly contiguity statistics with Quast (Gurevich *et al.* 2013; v5.0.2) and estimated assembly completeness using BUSCO (Seppey *et al.* 2019; v4.0.6). Because the linked-read, 10× sequences can also be used for other purposes (like assembly polishing) once the internal barcodes are trimmed, we processed the raw, linked-read FASTQ files using the

"basic" analysis of the 10× Genomics Long Ranger analysis pipeline (v2.2.2) to output reads that were trimmed of their 10× barcodes.

Prior to assembling the PacBio Sequel long-read data, we converted the subread BAM files to FASTQ format using bam2fastq (v1.3.0) from the PacBio BAM2fastx package (https://github.com/PacificBiosciences/bam2fastx; accessed 2021 August 2). We assembled the FASTQ files using canu (Koren *et al.* 2017; v1.8, r9465) on a 48-core compute node with 1.5 TB RAM and default options, other than setting the genome size parameter to 1.1 Gbp. Following this initial round of contig assembly, we computed contiguity statistics with Quast and estimated completeness using BUSCO, so that we could compare both metrics to the 10× linked-read assembly.

Based on the results of this initial comparison, we decided to move forward with the PacBio assembly. We aligned the short-read data, trimmed of 10× barcodes, to the PacBio contigs using bwa-mem (Li 2013; v0.7.17-r1188) and samtools (Li *et al.* 2009; v1.9) with default options, and we polished the contigs with Pilon (Walker *et al.* 2014; v1.23) using default options on a 48-core compute node with 1.5 TB RAM. We performed an initial round of misassembly correction using tigmint (v1.1.2) followed by contig scaffolding using arks+links (Warren *et al.* 2015; Coombe *et al.* 2018; Jackman *et al.* 2018; arks v1.0.4) with the Long Ranger-processed 10× reads and a file of barcode multiplicities we generated using the "calcBarcodeMultiplicities.pl" helper script from the arks software package. After the first round of scaffolding, we cleaned the contig names using custom Python code and removed contigs <1000 bp in length using "faFilter" from the UCSC Genome Browser software package (Kent *et al.* 2002, v25-Mar-2014-Kent-Source-LINUX-x86_64). Then, we computed contiguity statistics with Quast and estimated completeness using BUSCO. The BUSCO results showed a moderate level of duplication among BUSCOs, suggesting that the assembly contained a number of haplotigs. In order to identify and remove these, we modeled repeats in the scaffolded assembly using RepeatModeler2 (Flynn *et al.* 2020, v2.0.1) and RMBlast (v2.9.0), and we identified repeats using RepeatMasker (Smit *et al.* 2019; v4.1.0) and RMBlast (v2.9.0) with the sensitive option (-s) and output repeats in GFF format (-gff). We converted the GFF file of repeat locations to BED using awk (GNU Software Foundation 2020), and we removed putative haplotigs from the assembly using purge_haplotigs (Roach *et al.* 2018; v1.1.1) with this BED-formatted file of repeat locations. After haplotig purging, we computed contiguity statistics with Quast and estimated assembly completeness using BUSCO.

We converted the gaps in the arks-scaffolded assembly from 100 to 50 bp using custom Python code, and we performed a round of reference-based scaffolding using ragtag (Alonge *et al.* 2019; v1.0.0) and the *Camarhyncus parvulus* assembly (GCF_901933205.1; last common ancestor ~12–10 Ma; Barker *et al.* 2013). We filtered the *Camarhyncus* assembly to contain only chromosomes, and we scaffolded using ragtag with default options (setting gaps to 100 bp so that pseudochromosomal scaffolds can be easily identified and/or split). We renamed the pseudochromosomal scaffolds to reflect the reference genome used to scaffold them using custom Python code, uppercased the entire assembly, and ran the unmasked, ragtag-scaffolded assembly through RepeatMasker using the repeat models created above. We computed a final round of contiguity statistics with Quast and estimated completeness using BUSCO.

Results and discussion

Illumina sequencing of the 10× linked-read library generated 290,045,347 read pairs. Supernova software trimmed reads to a mean length of 139 bp and estimated raw coverage as 60X (41X "effective coverage"). After completion of the Supernova assembly pipeline, the pseudohaploid assembly included 26,480 contigs with an N50 of 6.1 Mbp (L50 50; Table 1). Total assembly length was 1.10 Gbp. BUSCO scores suggested the 10× assembly was reasonably complete (Table 2), although the Supernova assembly was less contiguous than any version of the canu assembly (Table 1; see below).

PacBio sequencing of the SMRTbell Express 20 kb library across three SMRT Cells produced 1.6 million reads having a mean value for longest subread of 20,319 bp (95 CI ± 1980 bp) and a mean value for longest subread N50 of 33,144 (95 CI ± 1369). Sequencing produced a total of 32.7 Gbp of data. Canu produced an assembly including 3087 contigs having an N50 of 6.9 Mbp (L50 46), a total length of 1.17 Gbp, and BUSCO scores that indicated a reasonable level of completeness with a moderate level of duplicated BUSCOs (Table 2). Polishing improved the detection of several missing BUSCOs, while it also increased the number of duplicated BUSCOs (Table 2). Assembly correction and scaffolding with tigmint, arks, and links increased the N50 by 3 Mbp (Table 1) with negligible effects on assembly completeness (Table 2). Visual analysis of the coverage histogram produced by purge_haplotigs suggested that appropriate coverage cutoffs were 5 (low read depth cutoff), 15 (the point between haploid and diploid peaks), and 190, and haplotig purging considering repeat regions removed a number of low coverage contigs ($n = 1632$) from the assembly and reduced the number of duplicated BUSCOs to a low level (~1%) with minor effects (~0.4% reduction) on the number of complete and single copy BUSCOs detected. Removal of putative haplotigs reduced the assembly length from 1.17 to 1.08 Gbp (Table 1). Reference-guided scaffolding using the *C. parvulus* assembly increased assembly contiguity to chromosome-level while retaining BUSCO scores similar to previous assembly versions (Tables 1 and 2).

The highly contiguous assembly we produced for *Diglossa* contributes to the growing number of assemblies representing the incredible diversity of this important avian family, enabling future studies of how, when, and where tanagers diversified. More broadly, sequencing and assembling a genome from a nectivorous tanager adds to the collection of obligate nectivores sequenced among the vertebrate Tree of Life (Feng *et al.* 2020; Supplementary Table S2), providing another avenue of research into the genetic basis of adaptations common to a nectivorous lifestyle, including: sweet taste perception (*e.g.*, Baldwin *et al.* 2014; Toda *et al.* 2021), sugar metabolism (*e.g.*, Schondube and Del Rio 2004; Workman *et al.* 2018), spatial memory of floral resources (*e.g.*, Araya-Salas *et al.*, 2018), frilled channeled tongues (*e.g.*, Rico-Guevara *et al.* 2019), and whether such adaptations have evolved through similar or different molecular mechanisms.

Data availability

Raw sequencing data are available from NCBI (PRJNA629984). 10× Supernova Assemblies are available from https://doi.org/10.5281/zenodo.4439775, while the pseudochromosomal, PacBio assembly has been deposited at DDBJ/ENA/GenBank under the accession JAFCGX000000000. The version described in this study is version JAFCGX010000000. Two supplemental tables describe available tanager genome assemblies (Supplementary Table S1), and assemblies for other vertebrate organisms that are obligate nectivores (Supplementary Table S2), one supplemental file describes our DNA extraction protocol (Supplementary File S1), a

Table 1 Quast summary statistics for different versions of the *Diglossa brunneiventris* assemblies

Assembly	Supernova	Canu	Canu+*pol*	Canu+*pol*+*scaf1*	Canu+*pol*+ scaf1+*purg*	Canu+*pol*+ scaf1+purg+scaf2
# contigs (≥0 bp)	26,480	3087	3087	2960	1328	601
# contigs (≥1000 bp)	26,480	3087	3087	2960	1328	601
# contigs (≥5000 bp)	5499	2971	2971	2762	1260	555
# contigs (≥10,000 bp)	1967	2880	2880	2619	1197	513
# contigs (≥25,000 bp)	803	2628	2628	2344	1049	426
# contigs (≥50,000 bp)	596	1960	1962	1714	829	285
Total length (≥0 bp)	1.11E+09	1.17E+09	1.17E+09	1.17E+09	1.08E+09	1.08E+09
Total length (≥1000 bp)	1.11E+09	1.17E+09	1.17E+09	1.17E+09	1.08E+09	1.08E+09
Total length (≥5000 bp)	1.06E+09	1.17E+09	1.17E+09	1.17E+09	1.08E+09	1.08E+09
Total length (≥10,000 bp)	1.04E+09	1.17E+09	1.17E+09	1.17E+09	1.08E+09	1.08E+09
Total length (≥25,000 bp)	1.02E+09	1.17E+09	1.17E+09	1.17E+09	1.08E+09	1.08E+09
Total length (≥50,000 bp)	1.02E+09	1.14E+09	1.14E+09	1.14E+09	1.07E+09	1.07E+09
Largest contig	22,027,751	28,875,242	28,916,091	30,446,970	30,446,970	155,774,234
N50	6,056,454	6,979,468	6,984,183	9,220,893	10,495,952	67,281,049
N75	1,734,222	937,141	934,919	1,297,763	2,420,771	22,837,129
L50	50	46	46	37	32	6
L75	129	164	165	123	83	12
# N's per 100 kbp	1593	—	—	2	2	8

Canu assemblies are named according to the steps used to produce each (*pol*: polished; *scaf1*: arks scaffolded; *purg*: haplotig purged; *scaf2*: ragtag scaffolded).

Table 2 BUSCO statistics for different versions of the *Diglossa brunneiventris* assemblies

Assembly	Supernova		Canu		Canu+pol		Canu+pol+scaf		Canu+pol+scaf+purg		Canu+pol+scaf+purg+scaf	
Complete BUSCOs (C)	237	93%	238	93%	242	95%	241	94.5%	242	95%	241	95%
Complete and single-copy BUSCOs (S)	234	92%	228	89%	227	89%	226	88.6%	240	94%	239	94%
Complete and duplicated BUSCOs (D)	3	1%	10	4%	15	6%	15	5.9%	2	1%	2	1%
Fragmented BUSCOs (F)	11	4%	6	2%	2	1%	3	1.2%	3	1%	4	2%
Missing BUSCOs (M)	7	3%	11	4%	11	4%	11	4.3%	10	4%	10	4%
Total BUSCO groups searched	255		255		255		255		255		255	

Canu assemblies are named according to the steps used to produce each (*pol*: polished; *scaf1*: arks scaffolded; *purg*: haplotig purged; *scaf2*: ragtag scaffolded).

second supplemental file describes all processing steps and commands (Supplementary File S2), and a final set of supplemental files provides the RepeatModeler and RepeatMasker output (Supplementary File S3). Supplementary tables and files are available from FigShare: https://doi.org/10.25387/g3.15030714. Samples were collected under collecting permits 203-2015-SERFOR-DGGSPFFS and 222-2015-SERFOR-DGGSPFFS, exported under permit 003336 SERFOR, and collected following LSU IACUC protocol 18-054.

Acknowledgments

Funding

Conflicts of interest
The authors declare that there is no conflict of interest.

Literature cited
Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, et al. 2019. RaGOO: fast and accurate reference-guided scaffold of draft genomes. Genome Biol. 20:224.

Araya-Salas M, Gonzalez-Gomez P, Wojczulanis-Jakubas K, Lœpez V, Wright TF. 2018. Spatial memory is as important as weapon and body size for territorial ownership in a lekking hummingbird. Sci Rep. 8: 2001.

Barker FK, Burns KJ, Klicka J, Lanyon SM, Lovette IJ. 2013. Going to extremes: contrasting rates of diversification in a recent radiation of new world passerine birds. Syst Biol. 62:298–320.

Baldwin MW, Toda Y, Nakagita T, O'Connell MJ, Klasing KC, et al. 2014. Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. Science 345: 929–933.

BirdLife International and Handbook of the Birds of the World 2019. Bird Species Distribution Maps of the World. v2019.1. http://datazone.birdlife.org/species/requestdis (Accessed: 2021 August 2).

Bock WJ. 1985. Is *Diglossa* (Thraupinae) monophyletic? Ornithol Monogr. 36:319–332.

Buckner JC, Sanders RC, Faircloth BC, Chakrabarty P. 2021. Science Forum: the critical importance of vouchers in genomics. eLife. 10: e68264.

Burns KJ. 1997. Molecular systematics of tanagers (Thraupinae): evolution and biogeography of a diverse radiation of neotropical birds. Mol Phylogenet Evol. 8:334–348.

Burns KJ, Hackett SJ, Klein NK. 2003. Phylogenetic relationships of Neotropical honeycreepers and the evolution of feeding morphology. J Avian Biol. 34:360–370.

Burns KJ, Shultz AJ, Title PO, Mason NA, Barker FK, et al. 2014. Phylogenetics and diversification of tanagers (Passeriformes: Thraupidae), the largest radiation of Neotropical songbirds. Mol Phylogenet Evol. 75:41–77.

Chapman FM, Cherrie GK, Richardson WB, Gill G, O'Connell GM, et al. 1926. The distribution of bird-life in Ecuador: a contribution to a study of the origin of Andean bird-life. Bulletin of the AMNH. v55.

Colwell RK, Betts BJ, Bunnell P, Carpenter FL, Feinsinger P. 1974. Competition for the Nectar of *Centropogon valerii* by the Hummingbird *Colibri thalassinus* and the Flower-Piercer *Diglossa plumbea*, and its evolutionary implications. Condor. 76:447–452.

Coombe L, Zhang J, Vandervalk BP, Chu J, Jackman SD, et al. 2018. ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers. BMC Bioinformatics. 19:234.

Cuervo AM. 2013. Evolutionary Assembly of the Neotropical Montane Avifauna. Ph.D. Dissertation. Baton Rouge, Louisiana: Louisiana State University.

del Hoyo J, Elliott A, Sargatal J, Christie D. 1992. Handbook of the Birds of the World. Barcelona: Lynx Edicions.

Dickinson EC, Christidis L. 2014. The Howard & Moore Complete Checklist of the Birds of the World, Vol. 2. 4th ed. Eastbourne: Aves Press.

Feng S, Stiller J, Deng Y, Armstrong J, Fang Q, et al. 2020. Dense sampling of bird diversity increases power of comparative genomics. Nature. 587:252–257.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 117:9451–9457.

Glenn TC, Nilsen RA, Kieran TJ, Sanders JG, Bayona-Vásquez NJ, et al. 2019. Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). PeerJ. 7:e7755.

GNU Software Foundation. 2020. The GNU Awk User's Guide. (Accessed: 2021 March 29). https://www.gnu.org/software/gawk/manual/gawk.html.

Graves GR. 1982. Speciation in the carbonated flower-piercer (*Diglossa carbonaria*) complex of the andes. Condor. 84:1.

Graves GR. 1985. Elevational correlates of speciation and intraspecific geographic variation in plumage in andean forest birds. Auk. 102:556–579.

Graves GR. 1988. Linearity of geographic range and its possible effect on the population structure of andean birds. Auk. 105:47–52.

Graves GR. 1990. A new subspecies of *Diglossa gloriosissima* (Aves, Thraupinae) from the Western Andes of Colombia. Proc Biol Soc Wash. 103:962–965.

Graves GR. 1991. Bergmann's rule near the equator: latitudinal clines in body size of an Andean passerine bird. Proc. Natl. Acad. Sci. U.S.A. 88: 2322–2325.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29: 1072–1075.

Gutiérrez-Zuluaga AM, González-Quevedo C, Oswald JA, Terrill RS, Pérez-Emán JL, et al. 2021. Genetic data and niche differences suggest that disjunct populations of *Diglossa brunneiventris* are not sister lineages. Ornithology. 138:1–14.

Hackett SJ. 1995. Molecular systematics and zoogeography of flower-piercers in the *Diglossa baritula* complex. Auk. 112:156–170.

Hazlehurst JA, Karubian JO. 2016. Nectar robbing impacts pollinator behavior but not plant reproduction. Oikos. 125:1668–1676.

Hazzi NA, Moreno JS, Ortiz-Movliav C, Palacio RD. 2018. Biogeographic regions and events of isolation and diversification of the endemic biota of the tropical Andes. Proc. Natl. Acad. Sci. U.S.A. 115: 7985–7990.

Inouye DW. 1983. The ecology of nectar robbing, In: B. Bentley and T. Elias (eds.). The Biology of Nectaries. New York: Columbia University Press. pp. 153–173.

Isler ML, Isler PR. 1987. The Tanagers: Natural History, Distribution, and Identification. Washington D.C: Smithsonian Institution Press.

IUCN 2020. The IUCN Red List of Threatened Species. Version 2020–1. https://www.iucnredlist.org (Accessed: 2021 August 2).

Jackman SD, Coombe L, Chu J, Warren RL, Vandervalk BP, et al. 2018. Tigmint: correcting assembly errors using linked reads from large molecules. BMC Bioinformatics. 19:393.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. 2002. The human genome browser at UCSC. Genome Res. 12:996–1006.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722–736.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN].

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25: 2078–2079.

Lyon DL, Chadek C. 1971. Exploitation of nectar resources by hummingbirds, Bees (*Bombus*), and *Diglossa baritula* and its role in the evolution of *Penstemon kunthii*. Condor. 73:246–248.

Mauck WM, Burns KJ. 2009. Phylogeny, biogeography, and recurrent evolution of divergent bill types in the nectar-stealing flower-piercers (Thraupini: *Diglossa* and *Diglossopis*). Biol J Linn Soc Lond. 98:14–28.

Navas Berdugo AP. 2008. Phylogeography and patterns of genetic and morphological variation in *Diglossa albilatera* (Aves: Thraupidae): testing the influence of isolation and adaptive evolution on population differentiation. Thesis. Bogotá, Colombia: Universidad de los Andes.

Natarajan C, Hoffmann FG, Weber RE, Fago A, Witt CC, et al. 2016. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. Science 354: 336–339.

Remsen JV, Jr. 1984. High incidence of "leapfrog" pattern of geographic variation in andean birds: implications for the speciation process. Science. 224:171–173.

Rico-Guevara A, Rubega MA, Hurme KJ, Dudley R. 2019. Shifting Paradigms in the Mechanics of Nectar Extraction and Hummingbird Bill Morphology. Integr Org Biol. 1: oby006.

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

Rojas-Nossa SV. 2007. Estrategias de extracción de néctar por pinchaflores (Aves: *Diglossa* y *Diglossopis*) y sus efectos sobre la polinización de plantas de los altos Andes. Ornitología Colombiana. 5: 21–39.

Schondube JE, Martinez del Rio C. 2004. Sugar and protein digestion in flowerpiercers and hummingbirds: a comparative test of adaptive convergence. J. Comp. Physiol. B. 174: 263–273.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol. 1962: 227–245.

Smit AFA, Hubley R, Green P. 2019. 2013–2015. Repeat Masker Open-4.0. https://www.repeatmasker.org (Accessed: 2021 August 2).

Solano-Ugalde A, Brinkhuizen DM. 2012. Aves, Thraupidae, Indigo Flowerpiercer *Diglossa indigotica* (Sclater, 1856): new country records, natural history notes, and occurrence in northwest Ecuador. Check List. 8:245.

Stiles FG, Ayala AV, Girón M. 1992. Polinización de las flores de *Brachyotum* (Melastomataceae) por dos especies de *Diglossa* (Emberizidae). Caldasia. 17:47–54.

Toda Y, Ko MC, Liang Q, Miller ET, Rico-Guevara A, *et al.* 2021. Early origin of sweet perception in the songbird radiation. Science. 373: 226–231.

Vaicenbacher L, Grim T, Greeney H. 2014. The nest and eggs of Black-throated Flowerpiercer *Diglossa brunneiventris*. Cotinga. 36: 56–57.

Vuilleumier F. 1969. Systematics and Evolution in *Diglossa* (Aves, Coerebidae). American Museum novitates; no. 2381. New York.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 9:e112963.

Warren RL, Yang C, Vandervalk BP, Behsaz B, Lagman A, *et al.* 2015. LINKS: scalable, alignment-free scaffolding of draft genomes with long reads. Gigascience. 4:35.

Workman RE, Myrka AM, Wong GW, Tseng E, Welch Jr KC, et al. 2018. Single-molecule, full-length transcript sequencing provides insight into the extreme metabolism of the ruby-throated hummingbird Archilochus colubris. Gigascience. 7: giy009.

Zimmer JT. 1929. Variation and distribution in two species of *Diglossa*. Auk. 46:21–37.

Zimmer JT, Phelps WH, Gilliard ET. 1952. A new race of the honey-creeper, *Diglossa cyanea*, from Venezuela. American Museum novitates; no. 1603. New York.

Communicating editor: A. Sethuraman