A Reference Genome Assembly for the Spotted Flycatcher (Muscicapa striata)

Gaspard Baudrin, Jean-Marc Pons, Bertrand Bed'hom, Lisa Gil, Roxane Boyer, Yves Dusabyinema, Frédéric Jiguet, Jérôme Fuchs

A Reference Genome Assembly for the Spotted Flycatcher (*Muscicapa striata*)

Gaspard Baudrin [1], Jean-Marc Pons[1], Bertrand Bed'Hom[1], Lisa Gil[2], Roxane Boyer[2], Yves Dusabyinema[2], Frédéric Jiguet[3], and Jérôme Fuchs[1,*]

[1]Institut de Systématique, Evolution, Biodiversité (ISYEB), UMR7205, Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, Paris, France

[2]Plateforme Génomique (GeT-PlaGe), Genotoul, US1426, INRAE, Castanet-Tolosan, France

[3]Centre d'Ecologie et des Sciences de la Conservation (CESCO), UMR7204, Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, Paris, France

*Corresponding author: E-mail: jerome.fuchs@mnhn.fr.

Accepted: 21 July 2023

Abstract

The spotted flycatcher (*Muscicapa striata*) forms with the Mediterranean flycatcher (*Muscicapa tyrrhenica*) a newly recognized species pair of trans-Saharan migratory passerines. These flycatchers present a nested peripatric distribution, a pattern especially unusual among high dispersal species that questions the eco-evolutionary factors involved during the speciation process. Here, we present a genome assembly for *M. striata* assembled using a combination of Nanopore and Illumina sequences. The final assembly is 1.08 Gb long and consists of 4,779 contigs with an N50 of 3.2 Mb. The completeness of our *M. striata* genome assembly is supported by the number of BUSCO (95%) and ultraconserved element (UCE) (4889/5041; 97.0%) loci retrieved. This assembly showed high synteny with the *Ficedula albicollis* reference genome, the closest species for which a chromosome-scale reference genome is available. Several inversions were identified and will need to be investigated at the family level.

Key words: Muscicapidae, comparative genomics, speciation, migration.

Significance

The reference genome assembly presented here provides a major resource for eco-evolutionary studies dealing with the speciation process of recently separated bird species. More specifically, our new genome assembly will help to shed light on the possible role of migration in the differentiation process of peripatric bird species. In addition, it will be valuable in the context of phylogenomics and comparative genomic projects related to avian evolution.

Introduction

The continental spotted flycatcher (*Muscicapa striata*) and the insular Mediterranean flycatcher (*Muscicapa tyrrhenica*) are small-sized passerines with brown-grayish plumage, which annually migrate to sub-Saharan Africa. They form a recently recognized species pair (Pons et al. 2016) that can be discriminated morphologically by their wing formula (Viganò and Corso 2015). The two taxa are also strongly differentiated genetically and acoustically (Viganò and Corso 2015; Pons et al. 2016).

The spotted and Mediterranean flycatchers have a peripatric distribution with the former being distributed across the Palearctic mainland (Maghreb and from Western Europe eastward to Mongolia) while the latter is endemic to the Western Mediterranean islands (Corsica, Sardinia,

and Balearic islands) and to coastal Tuscany in Italy. The range of *M. tyrrhenica* is geographically nested within the range of *M. striata*, representing a very rare distribution pattern in birds. In the Western Mediterranean islands, the three other endemic terrestrial bird species are either not sister species with their continental congeneric species (Corsican *Sitta whiteheadi* vs. mainland *Sitta europaea*; Pasquet 1998), are sedentary (*Sylvia balearica*; Voelker and Light 2011), or are more widely distributed on the mainland (*Sylvia subalpina*; Brambilla et al. 2008). The unicity of this distribution pattern coupled to the potentially high dispersal capacities of the two *Muscicapa* species questions the evolutionary processes involved in their differentiation.

Migration is a complex behavior under strong genetic control (e.g., Berthold et al. 1990; Helbig 1991; Delmore and Irwin 2014). Recent works demonstrated that phenological differences across a migratory divide for the Swainson's thrush (*Catharus ustulatus*) constitute a strong prezygotic isolation mechanism (Ruegg et al. 2012; Battey et al. 2018) and that genes putatively linked to migratory behavior were significantly differentiated (Ruegg et al. 2014). Consequently, it has been suggested that migratory behavior could be a speciation force by counter-selecting hybrids that cannot achieve a full migratory cycle and by favoring homogamy among breeding pairs (Irwin 2009; Delmore and Irwin 2014). Interestingly, differences exist between the migration phenotypes of the spotted and Mediterranean flycatchers. For instance, the two species appear to cross the Sahara in different ways: Corsican *tyrrhenica* crosses the Sahara by the shortest distance ("straight flight") whereas the westernmost population of the mainland *M. striata* appears to preferentially cross the Mediterranean through or close to the Gibraltar Strait and then move along the Western African coast (Jiguet et al. 2019). Furthermore, it appears that Corsican *M. tyrrhenica* flycatchers arrive and settle in breeding territories while mainland *M. striata* birds are still migrating (Faggio and Jolin 2008; Piacentini et al. 2023). Interestingly, continental *M. striata* individuals that stop over on the island do not seem to form mixed pairs with local birds. Because these differences in migratory behavior most probably act on the breeding behavior, we hypothesize that migration may play a key role in preventing gene flow between the spotted and Mediterranean flycatchers.

Here, we provide a first reference genome for the spotted flycatcher (*M. striata*) to further assess the genome-wide differentiation between the two taxa. This genome represents a determinant tool for the identification of putative regions and candidate genes linked to ecological trait variations and more generally for genomic studies focusing on the evolution and speciation in birds.

Results and Discussion

Genomes Assembly Features and Quality

The genome sequencing resulted in 96.0 Gb of reads used in the assembling procedure. This includes 40.8 Gb of data for Nanopore sequencing (33 and 7.8 Gb for PromethION and GridION runs respectively; see table 1) leading to an estimated long-read depth of 32×. Additionally, we obtained 369,699,970 Illumina paired-end reads (55.5 Gb), of which 367,938,486 (55.2 Gb) were retained after Trimmomatic cleaning, corresponding to a 44× depth for short reads.

The *M. striata* genome size estimated by SGA preqc using short reads was 1.25 Gb. This estimate is smaller than the one calculated by Wright et al. (2014), using Feulgen image analysis densitometry (1.45 Gb), and slightly larger than the *Ficedula albicollis* genome estimated at 1.13 Gb (Ellegren et al. 2012).

Our initial *M. striata* wtdbg2 long-read assembly resulted in 5,098 contigs (see supplementary table S1, Supplementary Material online). After the six Pilon and the single Racon polishing rounds, the final number of contigs was 4,779 for a total assembly length of 1.08 Gb (contig N50: 3.25 Mb; largest contig: 12.05 Mb). Fifty percent of the assembly was recovered in only 102 contigs (L50).

The assembly has a guanine–cytosine (GC) content of 42.17% and on average one single nucleotide polymorphism (SNP) every 152 bp (7,125,734 SNPs including indels were identified by our variant calling protocol).

Besides, there was no sign of an important bacterial DNA contamination as only 37 contigs (see supplementary table S2, Supplementary Material online) presented at least one Blast hit with an identified bacterial sequence (39 hits in total with a mean length of 33 bp). These contigs represent only 0.02% of the assembly's total size with the longest potentially contaminated contig being 19,778 bp long.

Table 1
Summary Statistics of the *M. striata* Genome Assembly

Item	Category	*M. striata*
Sequencing data	Nanopore PromethION reads (Gb)	33
	Nanopore GridION reads (Gb)	7.8
	Illumina reads (Gb)	55.5
Genome assembly	Estimated genome size (Gb)	1.25
	Assembled genome size (Gb)	1.08
	Contigs number	4,779
	Contigs L50	102
	Contigs N50 (Mb)	3.25
	Longest Contig (Mb)	12.05
Genome completness (ODB10)	BUSCO complete [single; duplicated]	7,946 (95.3%) [7,923; 23]
	BUSCO fragmented	113 (1.4%)
	BUSCO missing	279 (3.3%)

RepeatMasker identified 12.85% of the genome as repeats (see supplementary table S3, Supplementary Material online). From this fraction, 82.2% were interspersed elements, which situates *M. striata* within the higher range of bird genomes for transposable element fraction; Sotero-Caio et al. (2017) defined a "typical" fraction in bird genomes to be within the 4.1–9.8% interval. Then, our predictive annotation procedure using GeMoMa and BRAKER2 initially identified 13,653 and 39,871 putative genes, respectively. After filtering out the BRAKER2 transcripts containing no feature supported by extrinsic evidence (which were also on average approximatively three times shorter and composed of three times less features) and the ones overlapping GeMoMa transcripts, our annotation protocol resulted in 21,100 transcripts representing 20,520 protein-coding genes. This result is in line with the average gene number of other bird genome assemblies recently annotated (Zhang et al. 2022; Black et al. 2023) and includes 87.6% of the ortholog genes of the ODB10 Aves database (either complete, duplicated, or fragmented).

The *M. striata* mitochondrial genome (18,027 bp) includes a duplicated control region (see supplementary table S4, Supplementary Material online), as noted in other *Muscicapa* species. The sequences of the control regions differ substantially among *Muscicapa* species (average pairwise distance: 5.35%) but are identical within an individual, suggesting that concerted evolution is involved for the two control regions copies in *Muscicapa*.

Muscicapa striata Genome Completeness and Synteny

Our final *M. striata* genome assembly resulted in 1.08 Gb with an estimated genome size of 1.25 Gb (86.4% of the SGA preqc estimate). This 1.08 Gb mapped on 87% of the closest reference genome available (*F. albicollis*, 1.13 Gb; Ellegren et al. 2012), a species from which the spotted flycatcher diverged about 17–18 Ma (Zhao et al. 2023).

Results of the BUSCO analyses for *M. striata* indicated that 7,946 (95.3%) of the 8,338 loci in the ODB10 database were retrieved as complete (7,923 single-copy and 23 duplicated), 113 (1.4%) were incomplete, and 279 (3.3%) were missing. Comparison with a selection of other available avian genomes indicates that our assembly is in line with recently published genomes (see supplementary fig. S1 and table S5, Supplementary Material online). Moreover, the highest number of retrieved UCEs, 4,889 UCE loci (out of 5,041), among all Muscicapoidea taxa, was found in our *M. striata* assembly confirming its high completeness. A summary of all loci retrieved during the UCEs searching procedure is given in supplementary table S6, Supplementary Material online.

Muscicapa striata pseudochromosomes comprised 703 contigs of the assembly mapped on all the 30 chromosomes of the *F. albicollis* reference genome (see fig. 1*a*). Whole-genome mapping led to the identification of some putative chromosomal rearrangements. Five contigs concentrated the primary identified inverted regions (see fig. 1*b*). These regions may have played an important role in the speciation process because they could have favored reproductive isolation when not shared by all individuals (Sanchez-Donoso et al. 2022).

Besides, four candidate translocations were initially identified but they were not supported by unique long reads overlapping the regions containing putative breakpoints. Thus, they were considered as artifacts resulting from possible assembling errors.

Phylogenetic Relationships Using the UCE and Mitochondrial DNA Sets

The Muscicapoidea genomes permitted ultimately us to obtain 3,110 UCE sequences satisfying our requirements and leading to a global alignment of 681,320 bp (see supplementary table S6, Supplementary Material online). The topology resulting from the concatenated maximum likelihood (ML) analyses (5,024 variable sites) was well supported (bootstrap support >95%, supplementary fig. S2, Supplementary Material online) except for two nodes (position of *Alethe castanea* with respect to the *Cercotrichas/Copsychus* and *Muscicapa* clades [bootstrap: 51%] and the monophyly of *M. striata* [bootstrap: 87%]).

Phylogenetic relationships within Muscicapidae, as inferred from 15 mitochondrial loci, did not support the division between Saxicolinae and Niltavinae as the position of the *Erithacus/Cossypha* with respect to the Niltavinae is not strongly supported (see supplementary fig. S3, Supplementary Material online; Sangster et al. 2010). The sampled *Muscicapa* species form a clade that is sister to the genus *Melaenornis*. Within *M. striata* ($n = 2$, France and Mongolia), 46 out of the 11,361 bp (0.4%) in protein-coding regions were found to be variable (8 nonsynonymous and 38 synonymous substitutions) whereas 395 fixed differences (3.5%) were found between *M. striata* and *M. tyrrhenica*.

Materials and Methods

Sample Collection and Sequencing

We extracted DNA from muscle tissues of a *M. striata* breeding male (symmetric testes 10 × 8 mm) prepared as a study skin (MNHN-ZO-2018-449, https://science.mnhn.fr/institution/mnhn/collection/zo/item/2018-449), with a standard Phenol-Chloroform protocol.

Short-read DNA sequencing libraries were prepared with the Illumina TruSeq Nano DNA HT Library Prep Kit. Briefly,

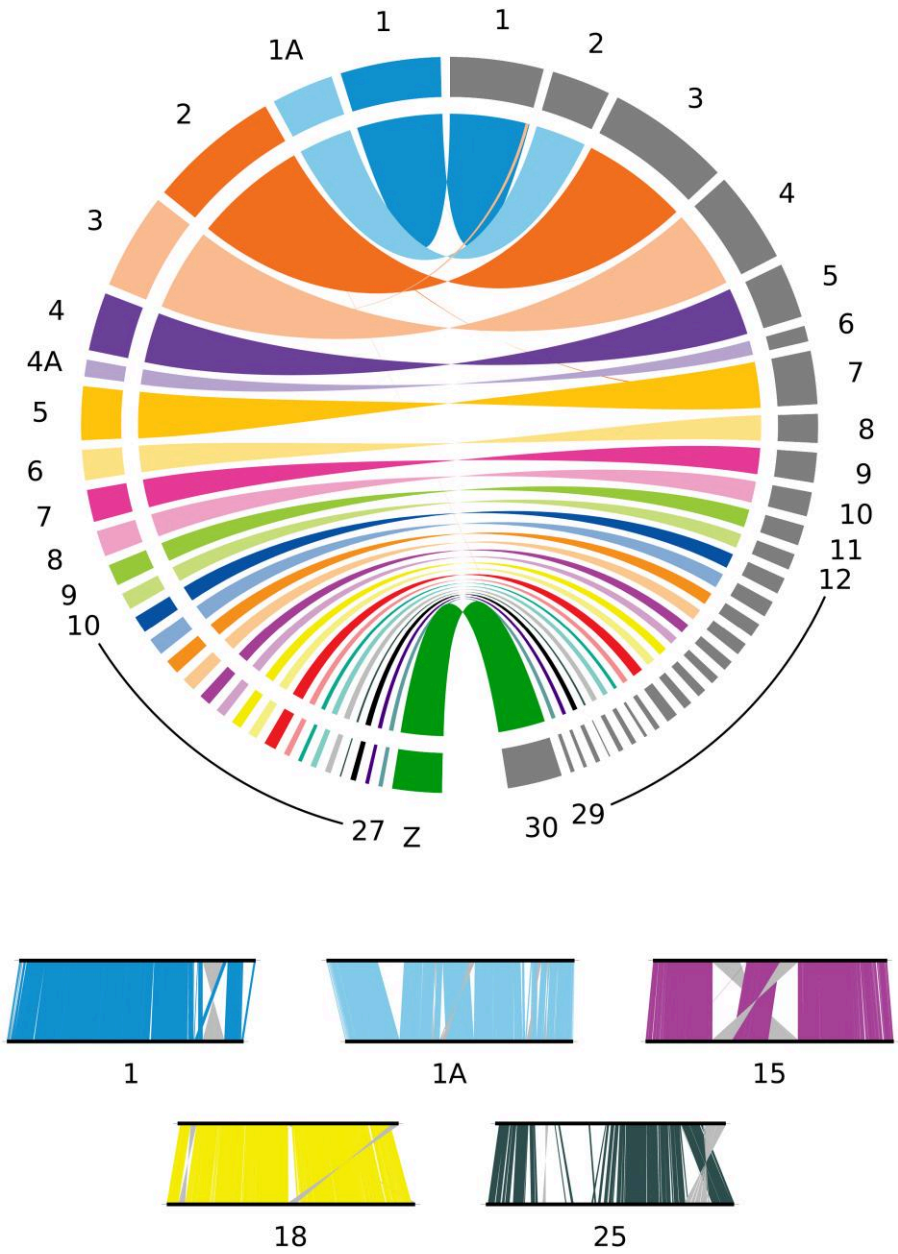

Fig. 1.—(*top*) Circos plot showing the comparison between the chromosomes of *F. albicollis* (left side) with the *M. striata* pseudochromosomes (right side). (*bottom*) Linear syntenic plots generated with the R package genoPlotR (Guy et al. 2010) of the chromosomes displaying the larger proportions of inverted mapping hits (>5% of their total length). Inversions are displayed in gray.

DNA was fragmented by sonication, and size selection (average insert size: 400 bp) was performed using Sample Purification Beads before adaptor ligation. Library quality was assessed on an Advanced Analytical Fragment Analyzer, and libraries were quantified by quantitative PCR with the KAPA Library Quantification Kit. DNA sequencing (paired-end read length: 2 × 150 bp) was performed on an Illumina HiSeq3000 using the Illumina HiSeq3000 Reagent Kits.

We prepared four libraries for the long-read sequencing based on an initial amount of 20 μg of DNA: one library without size selection, one library with a size selection (5-kb cutoff using the BluePippin size selection system, Sage Science) and two libraries with Short Read Eliminator XS Kit (Circulomics). Libraries were loaded on three R9.4.1 flow cells and sequenced on GridION (8 fmol within 48 h) and on one R9.4.1 flow cell sequenced on PromethION (25 fmol within 64 h). All short- and long-read

libraries were prepared and sequenced at the GeT-PlaGe core facility, INRAE, Toulouse (https://get.genotoul.fr/).

Assembly Pipeline

Illumina adapters and contaminated reads were removed prior to the assembling using Trimmomatic v0.36 (Bolger et al. 2014). The *M. striata* genome was first assembled from Nanopore reads and wtdbg2 v2.1 (Ruan and Li 2019), followed by one round of polishing with the same long reads. Three consecutive rounds of polishing using Pilon v1.22 (Walker et al. 2014) and the Illumina paired-end reads were performed next. We then repolished the assembly with the long reads using Racon v1.3.1 (Vaser et al. 2017) before three final polishing rounds with Pilon v1.22 and the Illumina paired-end reads.

The mitochondrial genome was assembled with the paired-end Illumina reads using NOVOPlasty (Dierckxsens et al. 2017). We used one CO1 barcode fragment from Pons et al. (2016) as a seed. *Muscicapa* species possess two control regions. We confirmed the sequence of the ND5-12S region, encompassing the loci Cytb-CR1-ND6-CR2, by PCR and Sanger sequencing because in the case of duplicated control regions, short reads only could not enable an accurate assembly. The primers used for amplification and sequencing are indicated in supplementary table S7, Supplementary Material online. Mitochondrial genome annotation was performed automatically using MITOS2 (Donath et al. 2019).

Variant Calling

PCR duplicates were removed with Picard (Picard Toolkit 2019). We performed variant calling using samtools (mpileup) and bcftools (options –multiallelic-caller –variants-only -Ob) packages (Li and Durbin 2009) after mapping Illumina reads on the assembly.

Repeat and Gene Annotations

Repeat annotation and predictive gene annotations were performed on the *M. striata* assembled genome. For de novo repeat discovery, we ran RepeatModeler v1.10.11 (Smit and Hubley 2008–2015) on the final assembly to create a *M. striata* repeat library. Then, the repeat content was characterized by running RepeatMasker v4.0.7 (Smit et al. 2013–2015) with the database containing our specific repeat library combined with those from other passerine species (*F. albicollis*, *Taeniopygia guttata*, and *Corvus cornix*) (Vijay et al. 2016).

We conducted an annotation of predicted genes on the soft-masked assembly in two steps. First, we ran the homology-based gene prediction implemented in GeMoMa v1.8 (Keilwagen et al. 2016) using *F. albicollis* reference genome and its annotation downloaded from the Ensembl genome browser (release 109; Cunningham et al. 2021). Then, we performed a BRAKER2 v2.1.6 (Brůna et al. 2021) automated prediction, which is based on successive GeneMark-Ep+ (Brůna et al. 2020) and AUGUSTUS (Stanke et al. 2006) runs using the extrinsic information of homologous protein sequences (OrthoDB Vertebrata database). BRAKER2 annotated transcripts not supported by any external evidence or overlapping GeMoMa predicted transcripts were filtered out. The proteins sequences were generated using the AGAT agat_sp_extract_sequences.pl and AUGUSTUS *getAnnoFastaFromJoingenes.py* scripts for GeMoMa and BRAKER2 annotation files, respectively, and the annotation completeness was assessed using BUSCO v4.0.2 protein mode (Simão et al. 2015).

Genome Quality and Completeness Assessment

Genome size was estimated with the Illumina paired-end reads with SGA preqc (Simpson 2014). We used BBMap v38-31 (Bushnell 2016) to obtain the descriptive statistics of the spotted flycatcher genome (number of contigs, N50, GC content).

We used BUSCO v4.0.2 and the Aves ODB10 database to search in *M. striata* genome for 8,338 single-copy orthologs shared among avian species in order to assess the completeness and quality of the genome assembly. We also ran BUSCO v3.0.2 with Aves ODB9 and compared the scores (either published or generated in this study) of genome assemblies from other available Passeriformes genomes (with a focus on Muscicapoidea or model species) non-Passeriformes sequenced using a similar sequencing strategy (*Melanerpes aurifrons*) and model species (e.g., *Gallus gallus*).

We also assessed genome completeness by estimating the number of UCEs, a set of loci commonly used in phylogenetics that could be retrieved from the genomes. We extracted the UCEs with *Phyluce* (Faircloth 2016) according to the online tutorial (https://phyluce.readthedocs.io/en/latest/tutorials/tutorial–3.html).

The absence of bacterial contamination within the assembly was assessed by comparing all the contigs with sequences identified as bacteria within the NCBI nucleotide database using Blast v2.10.1 (Altschul et al. 1990).

Finally, in order to ensure the coherence of the genome assembly with previous knowledge on our focal species, we reconstructed the Muscicapidae phylogeny based on both UCE sequences and mitochondrial genomes. We included *M. tyrrhenica* sequences in both analyses derived from Illumina short reads of a *M. tyrrhenica* individual extracted and sequenced the same way as *M. striata* (see Supplementary Methods, Supplementary Material online).

Syntenic Estimates

We generated pseudochromosomes of *M. striata* (masked assembly) with the reference genome-assisted approach implemented in RaGOO v1.1 (Alonge et al. 2019). The scaffold-level assemblies of *Muscicapa comitata* and

Muscicapa adusta (accessions GCA_027123655.1 and GCA_026546585.1) comprising 16,925 and 19,801 scaffolds, respectively, are organized in pseudochromosomes using the primarily and ab initio assembled *F. albicollis* genome (GCA_000247815.2) as a guide. We thus chose to use the latter as a reference. The validity and completeness of the *M. striata* pseudochromosomes were then measured by estimating synteny against the *F. albicollis* genome. For this purpose, we used the nucmer module in MUMMER v4.0.0b2 (Kurtz et al. 2004). Results were plotted on a circular plot generated with Circos (Krzywinski et al. 2009) using the *F. albicollis* assigned chromosomes (unplaced scaffolds were not considered).

We aligned the *M. striata* assembly and the *F. albicollis* reference genome with minimap2 (Li 2018) on the D-GENIES online platform (Cabanettes and Klopp 2018), to highlight putative chromosomal rearrangements between the two species. In order to identify potential inversions and translocations, pseudoscaffolds were generated by concatenating *M. striata* assembly contigs according to the order of the median positions of their hits on *F. albicollis* chromosomes. Then, we investigated inverted regions within the pseudoscaffolds as well as contigs that presented hits on more than one of the reference chromosomes (i.e., putative translocations).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We are very grateful to the Centre de Réhabilitation de la Faune Sauvage de Rosenwiller and its staff (Emilie Dusausoy, Suzel Hurstel, Jade Oliva, and Lauriane Perraud) for collaboration on our salvage bird program. We warmly thank Jean-Claude Thibault who did the fieldwork in Corsica under a CRBPO (Muséum National d'Histoire Naturelle) program (number 647) directed by Georges Olioso to whom we are grateful. DNA extraction and quantification were performed at the Service de Systématique Moléculaire (UAR 2700 2AD), and we acknowledge the help of its staff during the laboratory work. This work was performed in collaboration with the GeT core facility, Toulouse, France (http://get.genotoul.fr). All bioinformatic analyses were performed on the Genotoul cluster (http://bioinfo.genotoul.fr/). This work was supported by France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09). We are also very grateful to Joshua Peñalba for the help with formatting the Circos files.

Data Availability

The *M. striata* genome assembly (accession GCA_030060385.1) and the raw reads sequences (accessions SRR23885830 and SRR23885831) are available at NCBI under the BioProject PRJNA936819. The supplementary material and annotation files have been deposited and are available on Figshare under the following project link: https://figshare.com/projects/A_reference_genome_assembly_for_the_Spotted_Flycatcher_Muscicapa_striata_/163867.

Literature Cited

Alonge M, et al. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 20:224.
Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215:403–410.
Battey CJ, et al. 2018. A migratory divide in the painted bunting (*Passerina ciris*). Am Nat. 191:259–268.
Berthold P, Wiltschko W, Miltenberger H, Querner U. 1990. Genetic transmission of migratory behavior into a nonmigratory bird population. Experientia 46:107–108.
Black AN, et al. 2023. A highly-contiguous and annotated genome assembly of the lesser prairie-chicken (*Tympanuchus pallidicinctus*). Genome Biol Evol. 15:evad043.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.
Brambilla M, et al. 2008. A molecular phylogeny of the *Sylvia cantillans* complex: cryptic species within the Mediterranean basin. Mol Phylogenet Evol. 48:461–472.
Brůna T, Hoff K, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 3:lqaa108.
Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform. 2:lqaa026.
Bushnell B. 2016. BBMap short read aligner. Available from: http://sourceforge.net/projects/bbmap.
Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ 6:4958.
Cunningham F, et al. 2021. Ensembl 2022. Nucleic Acids Res. 50:988–995.
Delmore KE, Irwin DE. 2014. Hybrid songbirds employ intermediate routes in a migratory divide. Ecol Lett. 17:1211–1218.
Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45:e18.
Donath A, et al. 2019. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 47:10543–10552.
Ellegren H, et al. 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. Nature 491:756–760.
Faggio and Jolin. 2008. Migration printanière des oiseaux au Cap Corse (1992–2007). AAPNRC/CEN Corse/section ornithologie. 32.
Faircloth BC. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32:786–788.
Guy L, Roat Kultima J, Andersson S. 2010. genoPlotR: comparative gene and genome visualization in R. Bioinformatics 26:2334–2335.
Helbig AJ. 1991. Inheritance of migratory direction in a bird species: a cross-breeding experiment with SE- and SW-migrating blackcaps (*Sylvia atricapilla*). Behav Ecol Sociobiol. 28:9–12.

Irwin DE. 2009. Speciation: new migratory direction provides route toward divergence. Curr Biol. 19:1111–1113.
Jiguet F, et al. 2019. Desert crossing strategies of migrant songbirds vary between and within species. Sci Rep. 9:20248.
Keilwagen J, et al. 2016. Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. 44:e89.
Krzywinski M, et al. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19:1639–1645.
Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. doi:10.1186/gb-2004-5-2-r12
Li H. 2018. minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.
Pasquet E. 1998. Phylogeny of the nuthatches of the *Sitta canadensis* group and its evolutionary and biogeographic implications. Ibis 140:150–156.
Piacentini J, Jiguet F, Pons J-M, Thibault J-C. 2023. Note on the breeding biology of the Mediterranean flycatcher (*Muscicapa tyrrhenica tyrrhenica*) in Corsican villages, Western Mediterranean. Alauda 91:119–128.
Picard Toolkit. 2019. Broad Institute, GitHub Repository. Broad Institute. Available from: http://broadinstitute.github.io/picard/.
Pons J-M, et al. 2016. The role of western Mediterranean islands in the evolutionary diversification of the spotted flycatcher (*Muscicapa striata*), a long-distance migratory passerine species. J Avian Biol. 47:386–398.
Ruan J, Li H. 2019. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 17:155–158.
Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB. 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. Mol Ecol. 23:4757–4769.
Ruegg K, Anderson EC, Slabbekoorn H. 2012. Differences in timing of migration and response to sexual signalling drive asymmetric hybridization across a migratory divide. J Evol Biol. 25:1741–1750.
Sanchez-Donoso I, et al. 2022. Massive genome inversion drives coexistence of divergent morphs in common quails. Curr Biol. 32:462–469.
Sangster G, Alström P, Forsmark E, Olsson U. 2010. Multi-locus phylogenetic analysis of Old World chats and flycatchers reveals extensive paraphyly at family, subfamily and genus level (Aves: Muscicapidae). Mol Phylogenet Evol. 57:380–392.
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.
Simpson JT. 2014. Exploring genome characteristics and sequence quality without a reference. Bioinformatics 30:1228–1235.
Smit AFA, Hubley R. 2008–2015. RepeatModeler Open-1.0. Available from: http://www.repeatmasker.org.
Smit AFA, Hubley R, Green P. 2013–2015. RepeatMasker Open-4.0. Available from: http://www.repeatmasker.org.
Sotero-Caio CG, Platt RN, Suh A, Ray DA. 2017. Evolution and diversity of transposable elements in vertebrate genomes. Genome Biol Evol. 9:161–177.
Stanke M, et al. 2006. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. 34:W435–W439.
Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. Genome Res. 27:737–746.
Viganò M, Corso A. 2015. Morphological differences between two subspecies of spotted flycatcher *Muscicapa striata* (Pallas, 1764) (Passeriformes Muscicapidae). Biodivers J. 6:271–284.
Vijay N, et al. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nature Commun. 7:13195.
Voelker G, Light JE. 2011. Palaeoclimatic events, dispersal and migratory losses along the Afro-European axis as drivers of biogeographic distribution in *Sylvia* warblers. BMC Evol Biol. 11:163.
Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963.
Wright NA, Gregory TR, Witt CC. 2014. Metabolic 'engines' of flight drive genome size reduction in birds. Proc Biol Sci. 281:20132780.
Zhang X, et al. 2022. Chromosome-level genome assembly of the green peafowl (*Pavo muticus*). Genome Biol Evol. 14:evac015.
Zhao M, Gordon Burleigh J, Olsson U, Alström P, Kimball RT. 2023. A near-complete and time calibrated phylogeny of the Old World flycatchers, robins and chats (Aves, Muscicapidae). Mol Phylogenet Evol. 178:107646.

Associate editor: Bonnie Fraser