scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Draft genome assemblies of four manakins

Xuemei Li[1,2,10], Rongsheng Gao[1,2,10], Guangji Chen[1,2], Alivia Lee Price[3], Daniel Bilyeli Øksnebjerg[4], Peter Andrew Hosner[5,6], Yang Zhou[2], Guojie Zhang[3,7,8,9] & Shaohong Feng[7,8,9] ✉

Manakins are a family of small suboscine passerine birds characterized by their elaborate courtship displays, non-monogamous mating system, and sexual dimorphism. This family has served as a good model for the study of sexual selection. Here we present genome assemblies of four manakin species, including *Cryptopipo holochlora*, *Dixiphia pipra* (also known as *Pseudopipra pipra*), *Machaeropterus deliciosus* and *Masius chrysopterus*, generated by Single-tube Long Fragment Read (stLFR) technology. The assembled genome sizes ranged from 1.10 Gb to 1.19 Gb, with average scaffold N50 of 29 Mb and contig N50 of 169 Kb. On average, 12,055 protein-coding genes were annotated in the genomes, and 9.79% of the genomes were annotated as repetitive elements. We further identified 75 Mb of Z-linked sequences in manakins, containing 585 to 751 genes and an ~600 Kb pseudoautosomal region (PAR). One notable finding from these Z-linked sequences is that a possible Z-to-autosome/PAR reversal could have occurred in *M. chrysopterus*. These *de novo* genomes will contribute to a deeper understanding of evolutionary history and sexual selection in manakins.

Background & Summary

Manakins (Aves: Pipridae), a family of Passeriformes, contain 17 genera and about 50 species distributed across the Neotropics, and have some unique behavioral and morphological features[1]. Most species in the family have sexual dimorphism in plumage color[2] and are polygynous[3,4]. Moreover, the complex courtship displays of males, which include high-speed movements, sophisticated acrobatics, coordinated movements of multiple males, mechanical and vocal sounds and constructed display site construction[5,6], makes this lineage a fascinating model for studying sexual selection. During mating periods, males hold territories or aggregate for competitive displays to attract females for the chance to mate[7]. Courtship varies substantially among genera and species[8–11]. For example, in genus *Chiroxiphia*, one male forms a partnership with another male and they perform elaborate courtship dances and sing common songs together[12]. In contrast, *Corapipo gutturalis* does not cooperate with other males during courtship displays[13]. *Xenopipo atronitens* males elaborate courtship displays by making mechanical sounds through flapping their wings[14], whereas *Lepidothrix coronata* males sing to attract females in addition to acrobatic displays[15].

Courtship behavior plays an important role in attracting the opposite sex, increasing the chance of producing offspring and improving the reproductive rate of birds[2,16]. At present, the courtship display of manakin species has been studied from the aspect of behavior observation[17,18], neuroendocrine[14,19,20] and physiology[2]. The genetic mechanisms have also been discussed[21–24], yet insights are lacking due to a lack of comparative genomic and transcriptomic data. As courtship displays are derived from sexual selection[25,26], we expect that investigating the evolution of their genomes, particularly the sex chromosomes, could bring insights to the understanding of

[1]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China. [2]BGI-Shenzhen, Shenzhen, 518083, China. [3]Villum Centre for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200, Copenhagen, Denmark. [4]GLOBE Institute, Section for Evolutionary Genomics, University of Copenhagen, Copenhagen, Øster Farimagsgade 5, 1014, Copenhagen, Denmark. [5]Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, 2100, Copenhagen, Denmark. [6]Villum Center for Global Mountain Biodiversity, Biodiversity Section, GLOBE Institute, University of Copenhagen, Universitetsparken 15, 2100, Copenhagen, Denmark. [7]Evolutionary & Organismal Biology Research Center, Zhejiang University School of Medicine, Hangzhou, 310058, China. [8]Liangzhu Laboratory, Zhejiang University Medical Center, 1369 West Wenyi Road, Hangzhou, 311121, China. [9]Innovation Center of Yangtze River Delta, Zhejiang University, Jiashan, 314102, China. [10]These authors contributed equally: Xuemei Li, Rongsheng Gao. ✉e-mail: fengshaohong@zju.edu.cn

species		*Cryptopipo holochlora*	*Dixiphia pipra*	*Machaeropterus deliciosus*	*Masius chrysopterus*
Intitution acronym		NHMD	NHMD	NHMD	NHMD
SpecimenCode		B-126359	B-126493	B-125026	B-125031
Genus		*Cryptopipo*	*Dixiphia*	*Machaeropterus*	*Masius*
SpeciesName		*holochlora*	*pipra*	*deliciosus*	*chrysopterus*
subspecies		*holochlora*	*discolor*	/	*pax*
DateCollected		6-Jul-94	6-Jul-94	15-Apr-91	12-Sep-90
Sex		Not recorded	Not recorded	Not recorded	Not recorded
Field number		NK4-15.11.94	NK6-13.7.94	NK10-15.4.91	NK17-12.9.90
Sample in		EDTA	EDTA	EDTA	EDTA
voucher		skin; MECN:6936	Not recorded	skeleton; Salango Museum	skeleton; Museo Arqueológico Salango
CollectedBy		Niels Krabbe	Niels Krabbe	Niels Krabbe	Niels Krabbe
Location		Yasuní, Napo, Ecuador	Parque Nacional Yasuní, Napo, Ecuador	9 km west of Piñas, El Oro, Ecuador	Above Chinapinza, Zamora-Chinchipe, Ecuador
LocLatitude		−0.63333	−0.63333	−3.65	−4.039
LocLongitude		−76.43333	−76.43333	−79.75	−78.583
Elevation		300	300	900	1700
strategy		stlFR	stlFR	stlFR	stlFR
Sequencing platform		DNBseq	DNBseq	DNBseq	DNBseq
Library Insert Size (bp)		200~2000	200~2000	200~2000	200~2000
Raw reads	Total Data (Gb)	155.92	119.28	152.54	169.08
	Reads Length (bp)	100 + 100 + 30	100 + 100 + 30	100 + 100 + 30	100 + 100 + 30
	Sequence depth (×)	113	81	109	112
Clean reads	Total Data (Gb)	111.62	90	108.28	117.25
	Reads Length (bp)	100 + 100	100 + 100	100 + 100	100 + 100
	Sequence depth (×)	105	79	101	101

Table 1. Sequencing reads statistics.

underlying genetic mechanisms. To address this knowledge gap, we conducted whole genome sequencing of four species representing four manakin genera: *C. holochlora, D. pipra, M. deliciosus* and *M. chrysopterus*[27,28]. Genome sizes of these four manakin species were estimated to be 1.15 Gb, the contig N50 ranged from 125 Kb to 212 Kb, and the scaffold N50 ranged from 18.4 Mb to 36.6 Mb. We annotated about 12,055 protein-coding genes on each manakin genome. On average, 99.97% of the predicted protein-coding genes were successfully annotated by three functional databases (SwissProt, InterPro, and KEGG). About 75 Mb of Z-linked sequences, including an ~600 Kb PAR, were identified from the available female manakin genomes, including two published species (*Corapipo altera* and *Neopelma chrysocephalum*). These genomic resources will benefit research on genetic mechanisms of manakin courtship displays, and other behavioral and ecological aspects.

Methods

Sample collection, library construction, and sequencing. Tissue samples of four manakin species (*C. holochlora, D. pipra, M. deliciosus* and *M. chrysopterus*) were provided by the Natural History Museum of Denmark. High-molecular-weight genomic DNA of these samples was extracted with the Kingfisher Cell and Tissue DNA Kit Protocol. Single tube-Long Fragment Read (stLFR) technology[29] was used to construct the libraries for each sample. The resulting libraries underwent DNA Nanoball (DNB™) generation and DNBSEQ sequencing in 100 + 100 + 30 mode. On average, 149 Gb raw reads were produced for each species (Table 1).

Genome assembly and quality evaluation. A series of filtering steps was applied to these stLFR reads prior to the downstream analyses using SOAPfilter2 package (v2.2).

1. Remove reads with more than 10% of N bases;
2. Remove reads with more than 40% low quality bases (Phred score < = 10);
3. Remove reads with undersize insert size;
4. Filter out the PCR duplicates.

All cleaned stLFR library reads were transformed into 10X Genomics linked-reads format and passed into Supernova software (v2.0.1)[30] to assemble the genome under the "pseudohap" mode for each species. After removing scaffolds with "N" >80%, GapCloser (v1.12)[31] was used to close the intra-scaffold gaps.

The size of the four assembled genomes are about 1.15 Gb, similar to the sizes of other avian genomes[32] (Fig. 1a, Table 2). The scaffold N50 of all species is higher than 18 Mb, with the largest scaffold N50 found in *M. chrysopterus* (36 Mb). The contig N50 of all species is higher than 124 Kb. (Fig. 1b, Table 2).

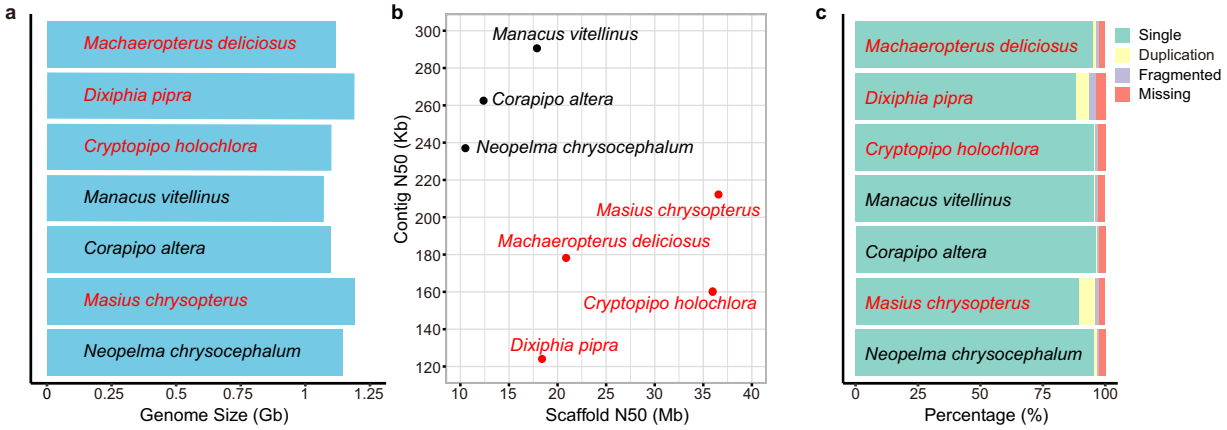

Fig. 1 Genome assembly statistics of four manakin genomes assembled in this study and three previously published genomes. (**a**) Comparison of genome sizes. (**b**) Distribution of N50 statistics of the manakin genomes. Each dot represents a manakin species, with the x-axis representing the value of scaffold N50 and the y-axis representing the value of contig N50. (**c**) BUSCO analysis of the seven manakin genomes. Assembly completeness is shown as the percentage of single, duplicated, fragmented and missing genes. Four newly assembled manakin genomes in this study were marked in red, while three published ones in black. Three published species are *Corapipo altera* (GCF_003945725.1), *Manacus vitellinus* (GCF_001715985.3) and *Neopelma chrysocephalum* (GCF_003984885.1).

Species	Genome assembly			BUSCO			
	Contig N50 (bp)	Scaffold N50 (bp)	Genome Size (bp)	Single (%)	Duplication (%)	Fragmented (%)	Missing (%)
Corapipo altera	262,501	12,385,833	1,095,745,976	96.10	0.60	0.90	2.40
Cryptopipo holochlora	160,303	35,963,530	1,099,400,137	95.40	0.30	1.40	2.90
Dixiphia pipra	124,585	18,411,073	1,187,686,032	88.40	5.10	2.80	3.70
Machaeropterus deliciosus	178,415	23,705,376	1,116,150,816	95.20	0.90	1.30	2.60
Masius chrysopterus	212,472	36,568,189	1,189,220,944	89.30	6.50	1.40	2.80
Manacus vitellinus	290,580	17,883,582	1,072,328,541	95.50	0.40	1.10	3.00
Neopelma chrysocephalum	237,029	10,517,223	1,142,796,179	95.50	0.90	1.00	2.60

Table 2. Genome assembly and BUSCO statistics.

We applied BUSCO (v5.2.2)[33] to evaluate the completeness of these seven manakin genomes using aves_odb10 as the reference gene set. On average 92% of the core genes were assembled as complete single-copy genes in the four manakin genomes and only about 3% of the core genes could not be annotated on the four manakin genomes (Fig. 1c, Table 2). Therefore, the overall quality of the newly assembled genomes was high and comparable to other published manakin assemblies.

Repeat annotation. Tandem repeats were identified by Tandem Repeat Finder (TRF, v4.09.1)[34], and transposable elements (TEs) were annotated using a combination of homology-based RepeatMasker (v4.1.2)[35], and *de novo* methods with RepeatModeler (v2.0.2a)[36] and LTR_Finder(v1.07)[37]. The homology-based annotation of TEs was performed by RepeatMasker with its built-in library. RepeatModeler and LTR_Finder methods were used to build the *de novo* repeat library for each species, which was further used by RepeatMasker to predict repeats for each species.

We found that the four species contained an average of 9.79% TEs in the genomes, with the proportions of each type being similar across these species (Fig. 2, Table 3). Long Interspersed Nuclear Elements (LINEs) accounted for most TEs, occupying about 6.79% of the genome.

Protein-coding gene annotation. We applied the homolog-based approach to annotate the protein-coding genes by using the protein sequences of *Gallus gallus*, *Taeniopygia guttata* and *Homo sapiens* downloaded from Ensembl release 105 as the reference gene sets. The protein sequences of these reference genes were aligned to each genome using TBLASTN (v2.2.26)[38] with an e-value cut off 1e-5, and multiple adjacent hits of the same query were connected by genBlastA (v1.0.4)[39]. Homologous blocks with length greater than 30% of the query protein length were retained. The connected hit region was later extended to include its 2 Kb upstream and downstream flanking regions, on which gene structure was predicted by Genewise (v2.4.1)[40]. MUSCLE (v3.8.31)[41] was then used to align the annotated protein with the reference protein. Predicted proteins with length $\geq$30 amino acids and identity value $\geq$40% were retained. Pseudogenes (annotated genes containing $>$2 frame shifts or $>$1 premature stop codon) and retrogenes were further removed.

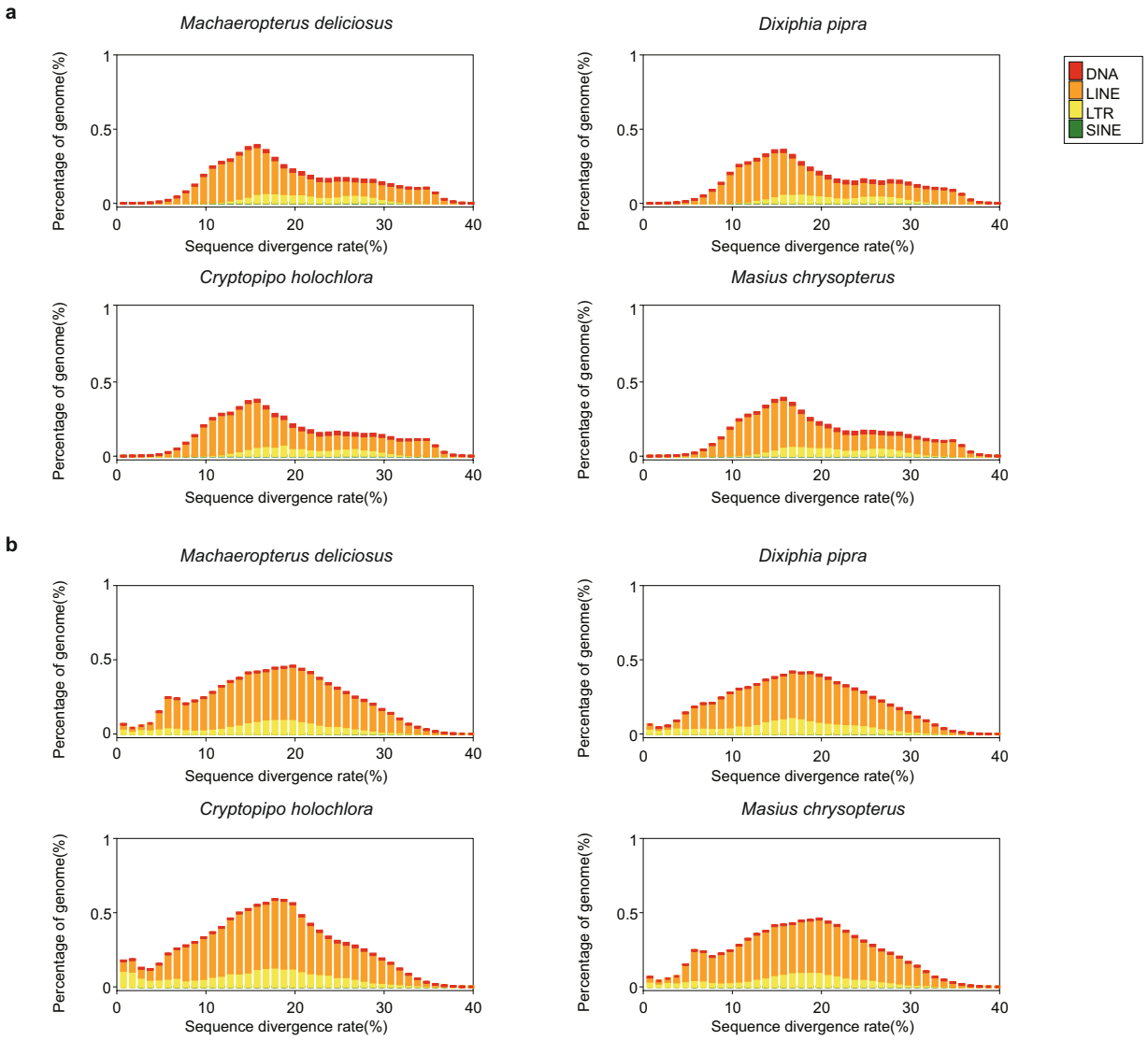

Fig. 2 Distribution of divergence rate of four types of transposable elements (TEs) in the four manakin genomes. (**a**) The divergence rate was calculated between the identified TEs in the genome by homology-based method and the consensus sequence in the built-in RepeatMasker TE library. (**b**) The divergence rate was calculated between the identified TEs in the genome by *de novo* and the consensus sequence in the *de novo* TE library.

species	DNA		LINE		SINE		LTR		Unknown		total	
	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome
Cryptopipo holochlora	3,878,915	0.35	76,847,348	6.99	1,170,656	0.11	23,590,172	2.15	8,465,390	0.77	106,843,742	9.72
Dixiphia pipra	3,856,289	0.32	73,600,596	6.20	1,420,049	0.12	25,019,052	2.11	9,522,408	0.80	107,604,264	9.06
Machaeropterus deliciosus	3,354,142	0.30	77,377,782	6.93	1,362,779	0.12	35,341,855	3.17	8,143,000	0.73	118,332,726	10.60
Masius chrysopterus	3,596,970	0.30	83,835,605	7.05	1,338,816	0.11	23,337,458	1.96	10,145,405	0.85	116,163,527	9.77

Table 3. Repeats statistic.

To build a non-redundant gene set, we first used hierarchical clustering[42] to combine the homologous-based gene sets of *G. gallus* and *T. guttata*. The gene model with the highest identity to the query was preserved if a locus has been annotated with more than one gene model. By doing so, we obtained 8,250 protein-coding genes on average after removing the highly duplicated genes (genes had >10 duplicates, were single-exon genes, and overlapped with the repeats in >70% of coding region). In the end, the newly annotated loci from the human gene set, i.e., the gene model did not overlap with the above combined one, were added into the results. In

species	# Total gene	# Single exon gene	Mean gene length (bp)	Mean coding sequence length (bp)	# Mean exons per gene	Mean exon length (bp)	Mean intron length (bp)
Cryptopipo_holochlora	11,681	702	22,859	1,677	9.89	170	2,384
Dixiphia pipra	11,770	703	22,821	1,669	9.84	170	2,393
Machaeropterus_deliciosus	11,985	730	22,880	1,689	9.91	170	2,378
Masius_chrysopterus	12,785	761	23,247	1,685	9.88	171	2,428

Table 4. Protein-coding gene statistics.

	Swissprot		KEGG		Interpro		Overall	
species	#gene	%	#gene	%	#gene	%	#gene	%
Corapipo altera	15,772	96.29	14,835	90.57	15,858	96.81	15,987	97.60
Cryptopipo holochlora	11,652	99.75	10,972	93.93	11,669	99.90	11,677	99.97
Dixiphia pipra	11,742	99.76	11,036	93.76	11,762	99.93	11,768	99.98
Machaeropterus deliciosus	11,958	99.75	11,276	94.07	11,980	99.93	11,985	99.97
Manacus vitellinus	14,204	98.03	13,299	91.79	14,254	98.38	14,317	98.81
Masius chrysopterus	12,754	99.76	12,027	94.07	12,770	99.88	12,781	99.97
Neopelma chrysocephalum	16,229	96.01	15,287	90.44	16,302	96.44	16,447	97.30

Table 5. Function annotation results.

summary, we predict an average of 12,055 protein-coding genes for each manakin with an average gene length of 22,952 bp. (Table 4).

Gene function annotation. The translated gene coding sequences were aligned to the SwissProt database (release-2020_05)[43] using BLASTP (v2.2.26)[38] with e-value cutoff 1e-5. The best match was assigned as the function annotation for each gene. Motifs and domains of each gene was annotated with modules PRINTS, SMART, PANTHER, ProSiteProfiles, ProSitePatterns, CDD, SFLD, Gene3D, SUPERFAMILY, and TMHMM of InterPro (v5.52–86.0)[44]. To identify the pathways in which genes may be involved, we also aligned the protein sequence of each gene to the KEGG database (release-93)[45] using BLASTP (v2.2.26)[38] with e-value cutoff 1e-5. Overall, 99.97% of the protein-coding genes of the four manakin genomes were annotated by the functional databases (Table 5).

Orthology assignment and phylogeny inference. To reconstruct the phylogenetic history of the seven genera in manakins, we chose one representative species for each genus, including the four species in this study and three published species (*C. altera*: GCF_003945725.1, *M. vitellinus*: GCF_001715985.3 and *N. chrysocephalum*: GCF_003984885.1). *T. guttata* (GCF_003957565.2) and *Calypte. anna* (GCF_003957555.1) were used as outgroups. The protein-coding gene sets of these species were obtained from NCBI. We used the *T. guttata* gene sets as the reference and performed a BLASTP (v2.2.26)[38] search on the protein sequences with an e-value cut-off of 1e-5. The reciprocal best hit (RBH) orthologs between *T. guttata* and every other species were identified following the published literature[46] but without the evidence of genomic synteny. In total, we obtained 9,654 one-to-one orthologs of these nine species by merging pairwise orthologs according to the reference *T. guttata* gene set.

The phylogeny of nine species was inferred based on the coalescent-based method, ASTRAL-III (v5.14.2)[47]. First, ortholog alignments were generated as follows: (1) we aligned the protein sequences with MAFFT L-INS-I (v7.487)[48]; (2) we used trimAl (v1.4.rev15)[49] to achieve a column-based alignment filtering with the parameter "automated", i.e., a heuristic selection of the automatic method based on similarity statistics; and (3) the nucleic acid alignments were back-translated from the trimmed protein alignments. After these steps, we obtained 9,653 trimmed ortholog alignments containing 805,481 parsimony informative sites in total. Then, we inferred the gene tree for each ortholog alignment using IQ-TREE (v1.6.12)[50] with ModelFinder[51] function to determine the best-fit model. The output gene trees were next used as the input for ASTRAL-III (v5.14.2)[47] with default parameters to infer the species tree shown in Fig. 3. As ASTRAL-III measures the branch lengths in coalescent units, we further ran RAxML (v8.2.12)[52] under GTR + GAMMA substitution model to estimate the branch lengths in substitution per site for the concatenated ortholog alignments by specifying the ASTRAL species tree (Fig. 4a). We also used DiscoVista[53] to analyze the discordance frequencies between the ASTRAL species tree and the 9,653 gene trees (Fig. 4b). The frequency of three potential topologies is inferred based on the focal internal branches of the species tree with the main topology (in red) and alternative topologies (in blues). More phylogenetic discordance can be observed in branch 5. Specifically, the frequency of the gene trees that support *C. holochlora* or *M. vitellinus* as the sister clade to *D. pipra* and *M. deliciosus* is close (Fig. 4c). In contrast to our species tree based on the coding regions, the UCEs-based topology published by Leite *et al.* 2021 concluded *M. vitellinus* as the sister clade to *D. pipra* and *M. deliciosus*[54]. Previous studies have suggested that such topological differences could result from data-type effects[55,56]. As in Leite *et al.* 2021 study, the UCE-based and exon-based topologies were not consistent either. Considering that our result still differed from their reported tree even based on coding regions, we assumed that such conflicts of *C. holochlora* and *M. vitellinus* could be caused by their limited parsimony informative sites, our restricted number of species, or the evolutionary forces (e.g.

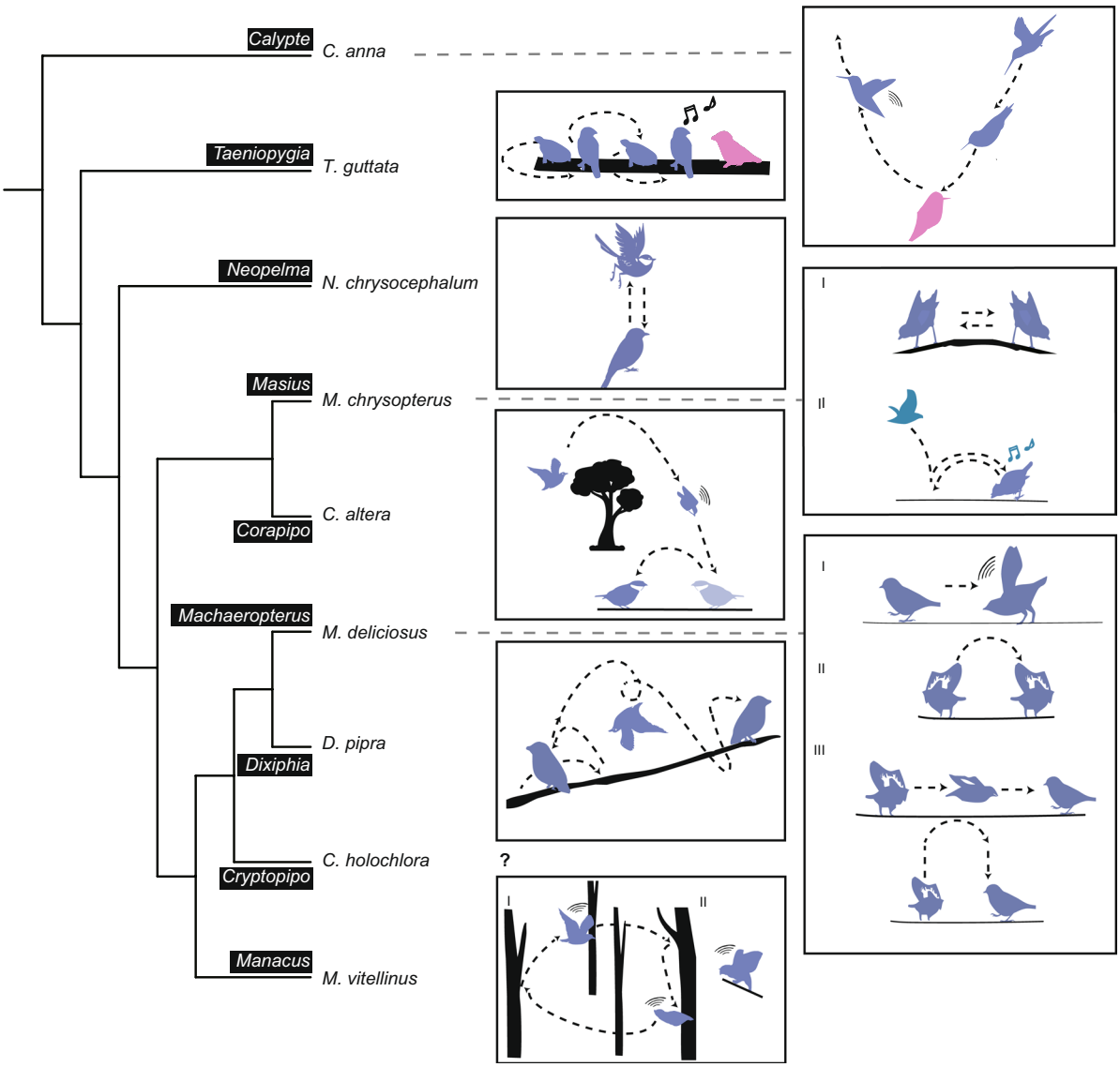

Fig. 3 Species tree with courtship displays. The ASTRAL species tree has the local posterior probabilities of all branch support as 1.0 across the tree. *C. anna*: The male ascends and swoops over the female. As the male nears the bottom of the dive, it flies upwards and its tail feathers make a sound[90]. *T. guttata*: The male jumps in the direction of the female, rotating 180° with each hop, moving its head and tail, and singing. When facing a female, the male sings and rhythmically shakes its head[91]. *N. chrysocephalum*: The male flaps its wings in a vertical leap[92]. *M. chrysopterus*: I. The male performs a side-to-side bow, with his head down and his tail up, turning his body 90°–180° degrees as he bows. II. The male flies to the log, then jumps to another place, lands on the log and sings. This can be done by two males working together[93]. *C. altera*: The male flies up from the display log, following by a high speed descent, wings making a sound, turns in the air, and lands facing the original landing site[94]. *M. deliciosus*: I. The male produces mechanical sounds by flapping their wings. II. When the male stands perpendicular to the perch, he bends forward, jumps from side to side, as if to display the black and white markings on the wings. but makes no sound. III. The male first flies along the perch in a short distance and then flies vertically upward, turning its body 180° in the process[95]. *D. pipra*: The male low jumped forward and high jumped back, spins his body in the air in a somersault-like motion, then flies to the perch and lands on it[10]. *C. holochlora*: We don't have much information about its courtship. *M. vitellinus*: The male performs snap-jump displays, jumping from one sapling to another, shaking its wings in midair. II. The male flaps its wings to produce mechanical sounds[96]. In the silhouette males are in blue and females are in pink.

introgression and incomplete lineage sorting). More whole genome resources are needed to solve the phylogenomics of these genera.

Selection analysis of plumage color related genes. Manakins are characterized by a variety of plumage colors[57–59]. To explore the possible genetic mechanism of the color diversity, we investigated the signatures of selection on 37 orthologous genes related to plumage color reported in previous studies[60–69]. With

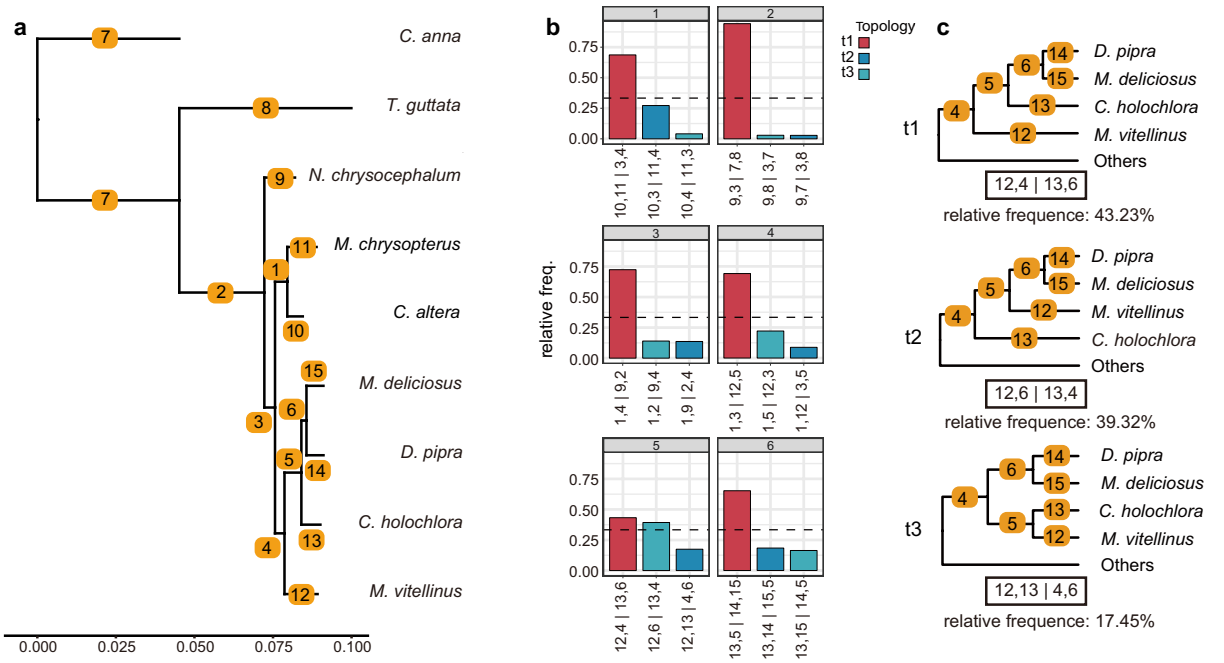

Fig. 4 Discordance of gene trees with the species tree. (**a**) The branch lengths are estimated by RAxML based on the concatenated ortholog alignments. The branch length scale refers to substitutions per site. (**b**) Frequency of three potential topologies around focal internal branches of the ASTRAL species tree. Main topology (species tree) is shown in red, and the other two alternative topologies are shown in light and dark blue. The dotted line indicates the 1/3 threshold. The title of each subfigure indicates the label of the corresponding branch on the tree in panel a. Each internal branch has four neighboring branches which could be used to represent quartet topologies. On the x-axis the exact definition of each quartet topology is shown using the neighboring branch labels separated by "|". (**c**) Example of the three topologies with the relative frequency values for internal branch 5, corresponding to panel 5 in b. The alternative topologies for other branches can be found in the Figshare database[89].

our phylogenetic tree, the maximum likelihood estimation of *dN* (non-synonymous substitution rate), *dS* (synonymous substitution rate), and ω (*dN*/*dS*) for each gene was performed under two branch models, one-ratio model (H0) and free-ratio model (H1), by using codeml program in PAML package (v4.9)[70]. Likelihood ratio test was used to test if H1 was significantly better than H0, and the output p-values were next corrected with the false-discovery rate (FDR) method. Under FDR-corrected p-value cutoff 0.05, if a branch showed $\omega > 1$ in the branch model analysis, the gene was considered to be positively selected at this branch. We further filtered results with abnormally high ω values ($\omega > 3$)[71]. We finally obtained four genes, *TBC1D22A*, *EDA*, *SLC45A2* and *GOLGB1*, that were likely to have undergone positive selection during manakin evolution. Among them, *SLC45A2* was found to be positively selected in *M. deliciosus*. The gene encodes a transporter protein that mediates melanin synthesis[66]. As pheomelanin is responsible for brown and reddish coloration[72,73], the positive selection signal in *M. deliciosus* may explain its unique reddish-brown body plumage among other studied manakin species. The other three genes were found under positive selection in the internal branches. *TBC1D22A* was positively selected in the most recent common ancestor (MRCA) of *M. deliciosus, D. pipra* and *C. holochlora* (branch 5 in Fig. 4a). *EDA* was positively selected in the MRCA of *M. deliciosus, D. pipra, C. holochlora* and *M. vitellinus* (branch 4 in Fig. 4a). *GOLGB1* was positively selected in both *M. chrysopterus* and MRCA of *M. deliciosus, D. pipra* and *C. holochlora* (branch 5 in Fig. 4a).

Sex chromosomes. Unlike mammals where males are heterogametic (XY system), in birds the females are heterogametic (ZW system). The avian ZW chromosomes are evolved from a pair of ancestral autosomes about 102 million years ago[74]. During evolution, the differentiation of sex chromosomes is caused by recombination arrests on the W chromosome, resulting in the reduction of functional genes on the chromosome and the accumulation of repetitive elements. The Z and W chromosomes of the extant Neoaves are of great differences in length and gene content[74]. Only a small PAR remains for recombination during cell division in females[74].

We first confirmed the sex of the manakin samples by mapping the sequencing reads of the same individual to its genomes with BWA MEM (v0.7.17)[75]. Coverage information extracted by samtools (v1.9)[76] was calculated in 5 Kb non-overlapping windows with bedtools (v2.29.2)[77] and normalized by the peak coverage. We also softmasked the genomes and performed LASTZ(v1.04.00)[78] alignment with the manakin genomes using the *T. guttata* genome as a reference with parameter set '--step = 19 --hspthresh = 2200 --inner = 2000 --ydrop = 3400 --gappedthresh = 10000 --format = axt'. Based on the assumption that Z chromosomes are relatively conserved among avians, scaffolds mapped to the Z chromosome of *T. guttata* with the aligning rate >50% were treated as candidate Z-linked sequences. The distribution of normalized coverage of candidate Z and other (not_Z)

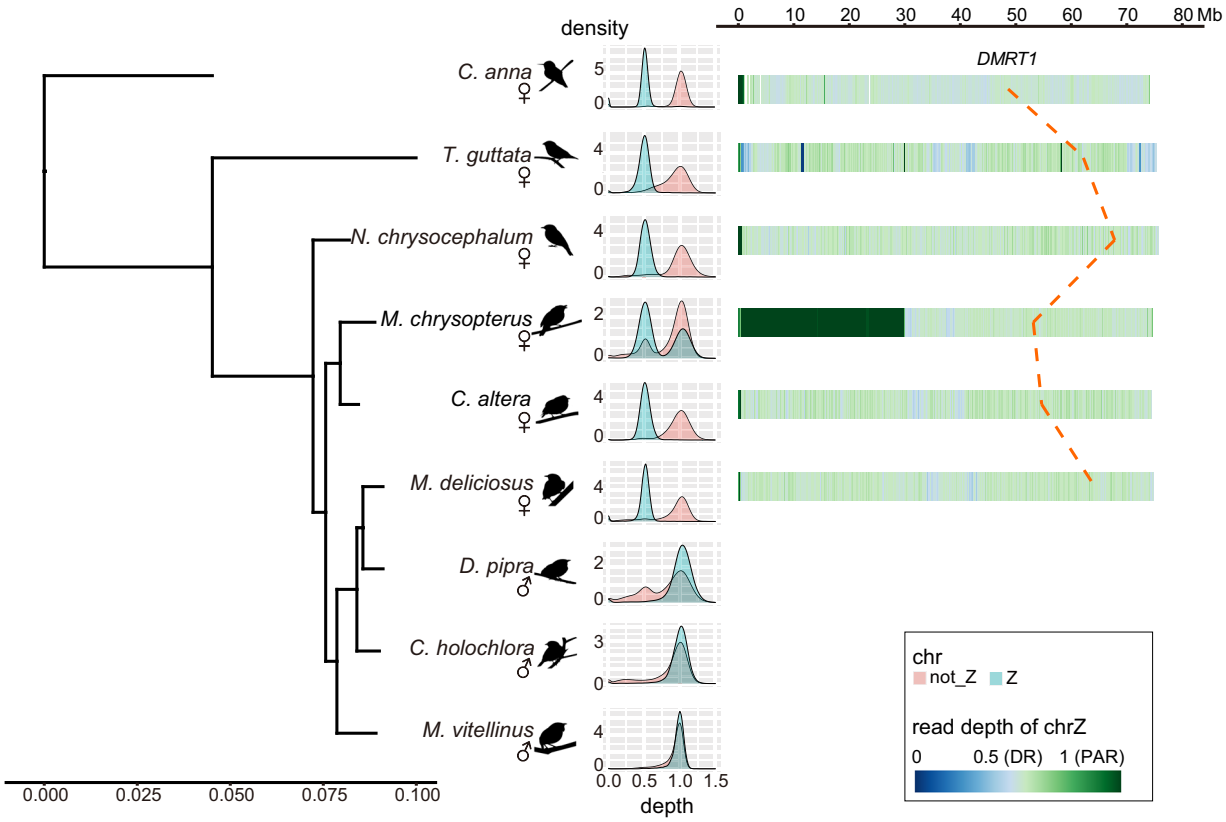

Fig. 5 Sex-related information of manakins. The sex of each sequenced bird is confirmed with the sequencing depth distribution of candidate Z (blue) and other sequence (not_Z, pink) shown in the middle. The putative Z chromosomes of species with a female sequenced are shown on the right. Each female avian species has a color-coded track showing the female read depth of chrZ under 100 Kb resolution, where PAR is shown in dark green, DR in light green and assembly gaps as blanks. One exception is found in *M. chrysopterus* where a large region shows normalized female sequencing depth around one, implying that a DR-to-autosome/PAR reversal might have happened in this species. The positions of the putative sex-determining gene *DMRT1* were traced with the dotted red line.

sequences were then visualized to check the sex of the sequenced individuals. We confirmed most of the sex information was consistent with records except *M. vitellinus* (BioSample SAMN02299332). This sample is more likely to be a male instead of a female. Its normalized coverage distribution was similar between the Z and not_Z sequences, with both peaks at around one and without a rise at 0.5 (Fig. 5).

With the above procedures we identified about 75 Mb of Z-linked scaffolds containing 585 to 751 genes in the manakins species where a female was sequenced (Fig. 5, Table 6, Supplementary Tables 1 and 2). We further constructed these Z-linked sequences into pseudo-Z-chromosomes for visualization with Ragtag (v2.1.0)[79] using *T. guttata* Z chromosomes as reference under parameter set "-q 10 -d 100,000 -i 0.2 -a 0.0 -s 0.0 -g 100 -m 100000 –aligner minimap2 –mm2-params '-x asm5'". To obtain the genomic coordinate of the avian candidate sex determining gene *DMRT1*[80], we used the *DMRT1* protein sequence of *G. gallus* downloaded from UniPort as a query, and annotated the orthologous genes on the pseudo-Z-chromosome of manakins using Genewise.

We also used the normalized coverage to identify PAR in the genomes assembled from female individuals. Z-linked scaffolds with normalized depth greater than 0.7 were identified as PAR candidates. We found that PAR is conserved between manakins and *T. guttata* with length of about 600 Kb and containing about 16 genes. However, one exception was found in *M. chrysopterus* where the candidate PAR is 30 Mb and contains 228 genes (Fig. 5, Table 6 & Supplementary Table 1). Most of the 30 Mb region has become differentiated region (DR) in the most recent common ancestor of Neoaves for about 69 million years[74], as well as the other manakins in this study. Thus, it is more likely that the region has reverted back to PAR or even autosome in *M. chrysopterus*. Such reversal is rare but has been found in other species[81,82]. Further exploration is required for the mechanism and explanation of this possible reversal.

Data Records

The genome sequencing data and assembly of the four manakin species has been deposited to CNSA (https://db.cngb.org/cnsa/) of CNGBdb[83] with accession number CNP0002887. The raw reads from DNBSEQ sequencing and the genome assembly of four manakins in this study was deposited to NCBI with SRA accession SRR19721507, SRR19721508, SRR19721509, SRR19721510, SRR19721511[84–88]. The annotation results of four manakin species, phylogenetic tree, discordance trees and the diploid assemblies were deposited in Figshare database[89].

species	PAR candidate		chrZ (PAR included)	
	length (bp)	# gene	length (bp)	# gene
Corapipo altera	674,217	16	74,797,643	744
Machaeropterus deliciosus	610,563	11	75,476,763	585
Masius chrysopterus	29,966,716*	228	74,940,613*	602
Neopelma chrysocephalum	660,109	17	76,014,776	751
Calypte anna	1,095,000	20	74,081,004	703
Taeniopygia Guttata	495,000	17	75,396,176	802

Table 6. Z chromosome statistics. *We suggest a differentiated region-to-autosome/pseudoautosomal region reversal has happened in this species.

Technical Validation

The assemblies of four manakins used in this study are the first version of the species. The average length of scaffold N50 and contig N50 were 29 Mb and 169 Kb, respectively. BUSCO analysis evaluated the genome assembly completeness. In total, about 95.23% core genes were assembled as complete genes of the four manakin genomes (single ~92.075%, duplicated ~3.200%, fragmented ~1.725%, missing ~ 3.000%). These results are comparable to those of three previously published manakins (*Corapipo altera*, *Manacus vitellinus*, and *Neopelma chrysocephalum*).

Code availability

The version and parameters of bioinformatic tools used in this study have been described in the Method section. If no parameter is described, the default is used.

Received: 20 June 2022; Accepted: 4 September 2022;
Published online: 13 September 2022

References

1. Kirwan, G. M., Green, G. & Barnes, E. *Cotingas and manakins*. (Princeton University Press, 2011).
2. Fusani, L., Barske, J., Day, L. D., Fuxjager, M. J. & Schlinger, B. A. Physiological control of elaborate male courtship: female choice for neuromuscular systems. *Neuroscience & Biobehavioral Reviews* **46**, 534–546 (2014).
3. Gaiotti, M. G., Webster, M. S. & Macedo, R. H. An atypical mating system in a neotropical manakin. *Royal Society open science* **7**, 191548 (2020).
4. Marçal, Bd. F. & Lopes, L. E. Non-monogamous mating system and evidence of lekking behaviour in the Helmeted Manakin (Aves: Pipridae). *Journal of Natural History* **53**, 2479–2488 (2019).
5. Johnsgard, P. A. Arena birds. Sexual selection and behavior. *Smithsonian Institution Press, Washington, DC(USA). 1994* (1994).
6. Prum, R. O. *The evolution of beauty: How Darwin's forgotten theory of mate choice shapes the animal world-and us*. (Anchor, 2017).
7. Bradbury, J. W. The evolution of leks. *Natural selection and social behavior*, 138–169 (1981).
8. Prum, R. O. Phylogenetic analysis of the evolution of display behavior in the Neotropical manakins (Aves: Pipridae). *Ethology* **84**, 202–231 (1990).
9. Prum, R. O. Sexual selection and the evolution of mechanical sound production in manakins (Aves: Pipridae). *Animal Behaviour* **55**, 977–994 (1998).
10. Castro-Astor, I. N., Alves, M. A. S. & Cavalcanti, R. B. Display behavior and spatial distribution of the White-crowned Manakin in the Atlantic Forest of Brazil. *The Condor* **109**, 155–166 (2007).
11. Prum, R. O. Phylogenetic analysis of the evolution of alternative social behavior in the manakins (Aves: Pipridae). *Evolution* **48**, 1657–1675 (1994).
12. DuVal, E. H. Cooperative display and lekking behavior of the lance-tailed manakin (Chiroxiphia lanceolata). *The Auk* **124**, 1168–1185 (2007).
13. Prum, R. O. The displays of the white-throated manakin Corapipo gutturalis in Suriname. *Ibis* **128**, 91–102 (1986).
14. Lindsay, W. R., Houck, J. T., Giuliano, C. E. & Day, L. B. Acrobatic courtship display coevolves with brain size in manakins (Pipridae). *Brain, behavior and evolution* **85**, 29–36 (2015).
15. Durães, R. Lek structure and male display repertoire of blue-crowned manakins in eastern Ecuador. *The Condor* **111**, 453–461 (2009).
16. Mitoyen, C., Quigley, C. & Fusani, L. Evolution and function of multimodal courtship displays. *Ethology* **125**, 503–515 (2019).
17. Prum, R. O. *Phylogenetic analysis of the morphological and behavioral evolution of the Neotropical manakins (Aves: Pipridae)*, University of Michigan, (1989).
18. Foster, M. S. Male Aggregation in Dwarf Tyrant-Manakins and What It Tells Us about the Origin of Leks. *Integrative and Comparative Biology* **61**, 1310–1318 (2021).
19. Schlinger, B. A., Fusani, L. & Day, L. Hormonal control of courtship in male Golden-collared manakins (Manacus vitellinus). *Ornitol Neotrop* **19**, 229–239 (2008).
20. Schlinger, B. A. & Chiver, I. Behavioral Sex Differences and Hormonal Control in a Bird with an Elaborate Courtship Display. *Integrative and Comparative Biology* **61**, 1319–1328 (2021).
21. Bennett, K. F., Lim, H. C. & Braun, M. J. Sexual selection and introgression in avian hybrid zones: spotlight on Manacus. *Integrative and Comparative Biology* **61**, 1291–1309 (2021).
22. Pennisi, E. *The genes behind the sexiest birds on the planet*, https://www.science.org/content/article/genes-behind-sexiest-birds-planet (2021).
23. Newhouse, D. J. & Vernasco, B. J. Developing a transcriptomic framework for testing testosterone-mediated handicap hypotheses. *General and Comparative Endocrinology* **298**, 113577 (2020).
24. Horton, B. M., Ryder, T. B., Moore, I. T. & Balakrishnan, C. N. Gene expression in the social behavior network of the wire-tailed manakin (Pipra filicauda) brain. *Genes, Brain and Behavior* **19**, e12560 (2020).
25. Andersson, M. & Iwasa, Y. Sexual selection. *Trends in ecology & evolution* **11**, 53–58 (1996).
26. Candolin, U. The use of multiple cues in mate choice. *Biological reviews* **78**, 575–595 (2003).
27. Dickinson, E. C. & Remsen, J. V. (eds) *The Howard and Moore Complete Checklist of the Birds of the World Volume 1: Non-passerines* 4th edn (Aves, 2013).

28. Dickinson, E. C. & Christidis, L. (eds) *The Howard and Moore Complete Checklist of the Birds of the World Volume 2: Passerines* 4th edn (Aves, 2014).
29. Wang, O. *et al*. Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome research* **29**, 798–808 (2019).
30. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome research* **27**, 757–767 (2017).
31. Luo, R. *et al*. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 2047-2217X–2041-2018 (2012).
32. Feng, S. *et al*. Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
33. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular biology and evolution* **38**, 4647–4654 (2021).
34. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research* **27**, 573–580 (1999).
35. Smit, A., Hubley, R & Green, P. *RepeatMasker Open-4.0*, http://www.repeatmasker.org (2013–2015).
36. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451–9457 (2020).
37. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic acids research* **35**, W265–W268 (2007).
38. Altschul, S. F. *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**, 3389–3402 (1997).
39. She, R., Chu, J. S.-C., Wang, K., Pei, J. & Chen, N. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome research* **19**, 143–149 (2009).
40. Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome research* **14**, 988–995 (2004).
41. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**, 1792–1797 (2004).
42. Johnson, S. C. Hierarchical clustering schemes. *Psychometrika* **32**, 241–254 (1967).
43. Boeckmann, B. *et al*. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic acids research* **31**, 365–370 (2003).
44. Jones, P. *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
45. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27–30 (2000).
46. Zhang, G. *et al*. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
47. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics* **19**, 15–30 (2018).
48. Katoh, K. & Toh, H. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* **26**, 1899–1900 (2010).
49. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
50. Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* **32**, 268–274 (2015).
51. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature methods* **14**, 587–589 (2017).
52. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
53. Sayyari, E., Whitfield, J. B. & Mirarab, S. DiscoVista: Interpretable visualizations of gene tree discordance. *Molecular Phylogenetics and Evolution* **122**, 110–115 (2018).
54. Leite, R. N. *et al*. Phylogenomics of manakins (Aves: Pipridae) using alternative locus filtering strategies based on informativeness. *Molecular Phylogenetics and Evolution* **155**, 107013 (2021).
55. Reddy, S. *et al*. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Systematic biology* **66**, 857–879 (2017).
56. Braun, E. L. & Kimball, R. T. Data types and the phylogeny of Neoaves. *Birds* **2**, 1–22 (2021).
57. Schaedler, L. M., Taylor, L. U., Prum, R. O. & Anciães, M. Constraint and function in the predefinitive plumages of manakins (Aves: Pipridae). *Integrative and Comparative Biology* **61**, 1363–1377 (2021).
58. Hudon, J., Storni, A., Pini, E., Anciães, M. & Stradi, R. Rhodoxanthin as a characteristic keto-carotenoid of manakins (Pipridae). *The Auk* **129**, 491–499 (2012).
59. Igic, B., D'Alba, L. & Shawkey, M. D. Manakins can produce iridescent and bright feather colours without melanosomes. *Journal of Experimental Biology* **219**, 1851–1859 (2016).
60. Wang, X. *et al*. Combined transcriptomics and proteomics forecast analysis for potential genes regulating the Columbian plumage color in chickens. *PloS one* **14**, e0210850 (2019).
61. Hua, G., Chen, J., Wang, J., Li, J. & Deng, X. Genetic basis of chicken plumage color in artificial population of complex epistasis. *Animal Genetics* **52**, 656–666 (2021).
62. Mastrangelo, S. *et al*. Genome-wide analyses identifies known and new markers responsible of chicken plumage color. *Animals* **10**, 493 (2020).
63. Guo, Q. *et al*. Genome-Wide Analysis Identifies Candidate Genes Encoding Feather Color in Ducks. *Genes* **13**, 1249 (2022).
64. Funk, E. R. & Taylor, S. A. High-throughput sequencing is revealing genetic associations with avian plumage color. *The Auk* **136**, ukz048 (2019).
65. Li, S., Wang, C., Yu, W., Zhao, S. & Gong, Y. Identification of genes related to white and black plumage formation by RNA-Seq from white and black feather bulbs in ducks. *PLoS One* **7**, e36592 (2012).
66. Gunnarsson, U. *et al*. Mutations in SLC45A2 cause plumage color variation in chicken and Japanese quail. *Genetics* **175**, 867–877 (2007).
67. Davoodi, P., Ehsani, A., Vaez Torshizi, R. & Masoudi, A. New insights into genetics underlying of plumage color. *Animal Genetics* **53**, 80–93 (2022).
68. Yang, C.-w. *et al*. Polymorphism in MC1R, TYR and ASIP genes in different colored feather chickens. *3 Biotech* **9**, 1–8 (2019).
69. Domyan, E. T. *et al*. Epistatic and combinatorial effects of pigmentary gene mutations in the domestic pigeon. *Current Biology* **24**, 459–464 (2014).
70. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**, 1586–1591 (2007).
71. Uebbing, S. *et al*. Divergence in gene expression within and between two closely related flycatcher species. *Molecular ecology* **25**, 2015–2028 (2016).
72. Nordlund, J. J. *et al*. *The pigmentary system: physiology and pathophysiology*. (John Wiley & Sons, 2008).
73. Hill, G. E., Hill, G. E., McGraw, K. J. & Kevin, J. *Bird coloration, volume 2: function and evolution*. Vol. 2 (Harvard University Press, 2006).
74. Zhou, Q. *et al*. Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science* **346**, 1246338 (2014).
75. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* (2013).
76. Danecek, P. *et al*. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
77. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
78. Harris, R. S. *Improved pairwise alignment of genomic DNA*. (The Pennsylvania State University, 2007).

79. Alonge, M. *et al*. Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. *BioRxiv* (2021).
80. Smith, C. A. *et al*. The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature* **461**, 267–271 (2009).
81. Vicoso, B. & Bachtrog, D. Reversal of an ancient sex chromosome to an autosome in Drosophila. *Nature* **499**, 332–335 (2013).
82. Abbott, J. K., Nordén, A. K. & Hansson, B. Sex chromosome evolution: historical insights and future perspectives. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20162806 (2017).
83. *CNGB Sequence Read Archive and Genome Assembly* https://db.cngb.org/search/project/CNP0002887/ (2022).
84. *NCBI Sequence Read Archive (SRR19721507)* https://identifiers.org/ncbi/insdc.sra:SRR19721507 (2022).
85. *NCBI Sequence Read Archive (SRR19721508)* https://identifiers.org/ncbi/insdc.sra:SRR19721508 (2022).
86. *NCBI Sequence Read Archive (SRR19721509)* https://identifiers.org/ncbi/insdc.sra:SRR19721509 (2022).
87. *NCBI Sequence Read Archive (SRR19721510)* https://identifiers.org/ncbi/insdc.sra:SRR19721510 (2022).
88. *NCBI Sequence Read Archive (SRR19721511)* https://identifiers.org/ncbi/insdc.sra:SRR19721511 (2022).
89. Li, X. *et al*. Draft genome assemblies of four manakins (Aves: Pipridae), *figshare* https://doi.org/10.6084/m9.figshare.c.6128388.v3 (2022).
90. Clark, C. J. Courtship dives of Anna's hummingbird offer insights into flight performance limits. *Proceedings of the Royal Society B: Biological Sciences* **276**, 3047–3052 (2009).
91. Morris, D. The reproductive behaviour of the zebra finch (Poephila guttata), with special reference to pseudofemale behaviour and displacement activities. *Behaviour* **6**, 271–322 (1954).
92. Alonso, J. A. & Whitney, B. M. New distributional records of birds from white-sand forests of the northern Peruvian Amazon, with implications for biogeography of northern South America. *The Condor* **105**, 552–566 (2003).
93. Prum, R. O. & Johnson, A. E. Display behavior, foraging ecology, and systematics of the Golden-winged Manakin (Masius chrysopterus). *The Wilson bulletin (Wilson Ornithological Society)* **99**, 521–539 (1987).
94. Rosselli, L., Vasquez, P. & Ayub, I. The courtship displays and social system of the White-ruffed Manakin in Costa Rica. *The Wilson Bulletin*, 165–178 (2002).
95. Bostwick, K. S. Display behaviors, mechanical sounds, and evolutionary relationships of the Club-winged Manakin (Machaeropterus deliciosus). *The Auk* **117**, 465–478 (2000).
96. Fuxjager, M. J., Longpre, K. M., Chew, J. G., Fusani, L. & Schlinger, B. A. Peripheral androgen receptors sustain the acrobatics and fine motor skill of elaborate male courtship. *Endocrinology* **154**, 3168–3177 (2013).

Acknowledgements

We thank the Natural History Museum of Denmark, including Jan Bolding and Niels Krabbe for archiving and providing tissue samples and China National GeneBank for providing the computation resource. This work was supported by grants from National Natural Science Foundation of China grant (31901214 and 32170626) to S.F.

Author contributions

Shaohong Feng, Guojie Zhang, Yang Zhou designed and directed the project; Peter Andrew Hosner selected and provided samples for analysis; Daniel Bilyeli Øksnebjerg organized and managed the project; Alivia Lee Price extracted the DNA from the sample; Guangji Chen performed genome assembly; Xuemei Li and Rongsheng Gao analyzed the data; Xuemei Li and Rongsheng Gao wrote the manuscript with other authors' help; Shaohong Feng, Guojie Zhang, Yang Zhou and Peter Andrew Hosner revised the manuscript. All authors read and approved the final manuscript. Xuemei Li and Rongsheng Gao made the same contribution.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-022-01680-0.

Correspondence and requests for materials should be addressed to S.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.