



Using historical genome-wide DNA to unravel the confused taxonomy in a songbird lineage that is extinct in the wild

Pratibha Baveja¹ | Kritika M. Garg^{1,2} | Balaji Chattopadhyay¹ | Keren R. Sadanandan^{1,3} | Dewi M. Prawiradilaga⁴ | Pramana Yuda⁵ | Jessica G. H. Lee⁶ | Frank E. Rheindt¹

¹Department of Biological Sciences, National University of Singapore, Singapore, Singapore

²Institute of Bioinformatics and Applied Biotechnology, Bangalore, India

³Max Planck Institute for Ornithology, Seewiesen, Germany

⁴Bidang Zoologi, Puslit Biologi - LIPI, Cibinong, Indonesia

⁵Fakultas Teknobiologi, Universitas Atma Jaya Yogyakarta, Yogyakarta, Indonesia

⁶Department of Conservation and Research, Wildlife Reserves Singapore, Singapore, Singapore

Correspondence

Frank E. Rheindt, Department of Biological Sciences, National University of Singapore, 16 Science Drive 4, Singapore 117558, Singapore.

Email: dbrfe@nus.edu.sg

Funding information

Ramanujan Fellowship; South East Asian Biodiversity Genomics, Grant/Award Number: WBS R-154-000-648-646 and WBS R-154-000-648-733; Croeni Foundation, Grant/Award Number: WBS R-154-000-A05-592; American Museum of Natural History; University of Michigan; Wildlife Reserves Singapore, Grant/Award Number: WBS R-154-000-A99-592; Science and Engineering Research Board

Abstract

Urgent conservation action for terminally endangered species is sometimes hampered by taxonomic uncertainty, especially in illegally traded animals that are often cross-bred in captivity. To overcome these problems, we used a genomic approach to analyze historical DNA from museum samples across the Asian Pied Starling (*Gracupica contra*) complex in tropical Asia, a popular victim of the ongoing songbird crisis whose distinct Javan population (“Javan Pied Starling”) is extinct in the wild and subject to admixture in captivity. Comparing genomic profiles across the entire distribution, we detected three deeply diverged lineages at the species level characterized by a lack of genomic intermediacy near areas of contact. Our study demonstrates that the use of historical DNA can be instrumental in delimiting species in situations of taxonomic uncertainty, especially when modern admixture may obfuscate species boundaries. Results of our research will enable conservationists to commence a dedicated ex situ breeding program for the Javan Pied Starling, and serve as a blueprint for similar conservation problems involving terminally endangered species subject to allelic infiltration from close congeners.

KEYWORDS

Asian Pied Starling, Asian Songbird Crisis, conservation genetics, museum samples, South-East Asia, target enrichment, wildlife trade

1 | INTRODUCTION

Conservation action is ideally based on adequate scientific data (Rose et al., 2019). A lack of clarity over species boundaries may often lead to failed conservation action (Ely et al., 2017). Critically endangered taxa, or those extinct in the wild, pose a particular dilemma, as surviving individuals are often inaccessible to taxonomists

and susceptible to genetic admixture, making historical museum material the only reliable source of taxonomic information (Payne & Sorenson, 2002). With the emergence and continuous advancement of next-generation sequencing and related bioanalytical pipelines (Derkarabetian et al., 2019; Linck et al., 2017; Prüfer, 2018; Rowe et al., 2011; Schubert et al., 2014; Schuster, 2007), it is now possible to study trace amounts of DNA from historical samples,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

thereby overcoming the problems of low sample size (Nedoluzhko et al., 2020; Tsai et al., 2019), modern genetic admixture (Prost et al., 2020) and unavailability of live individuals (Wood et al., 2018).

The Asian Pied Starling (*Gracupica contra*) in South and South-East Asia (Figure 1a, Craig et al., 2020) is one such species complex in which both taxonomic uncertainty and the threat of genomic infiltration have hampered conservation. Traditionally, this complex has comprised five taxa, *contra*, *sordida*, *superciliaris*, *floweri*, and *jalla*, recognized at the subspecies level based on morphological differences (Figure 1a,c, Gill et al., 2020). The nominate subspecies *contra* from the Indian subcontinent and the widely synonymized subspecies *sordida* from Assam have a largely black crown and a red-dish-orange bill base, with extremely limited bare skin around the eye (Figure 1a). In contrast, neighboring *superciliaris* from Myanmar

has white streaks on the forehead and a wider extent of orange bare skin around the eye (Figure 1c). The subspecies *floweri*, centered in Thailand, has even more extensive white streaking on the forehead and more bare facial skin (Figure 1a,c). Finally, the morphologically most distinctive subspecies *jalla*, originally from Java and Bali, lacks any reddish coloration on its bill base and sports the most extensive patch of deep-orange bare skin around the eye (Figure 1a, Horsfield, 1821). Due to this distinctive morphology, it has sometimes been considered a separate species, the Javan Pied Starling (Birdlife International, 2020; del Hoyo et al., 2020).

The distinct subspecies *jalla* is extinct in the wild following unsustainable and illegal wildlife trade, and the few surviving captive populations are subject to an ever-increasing market demand (Chng et al., 2015; Eaton et al., 2015; Marshall et al., 2019). Because of

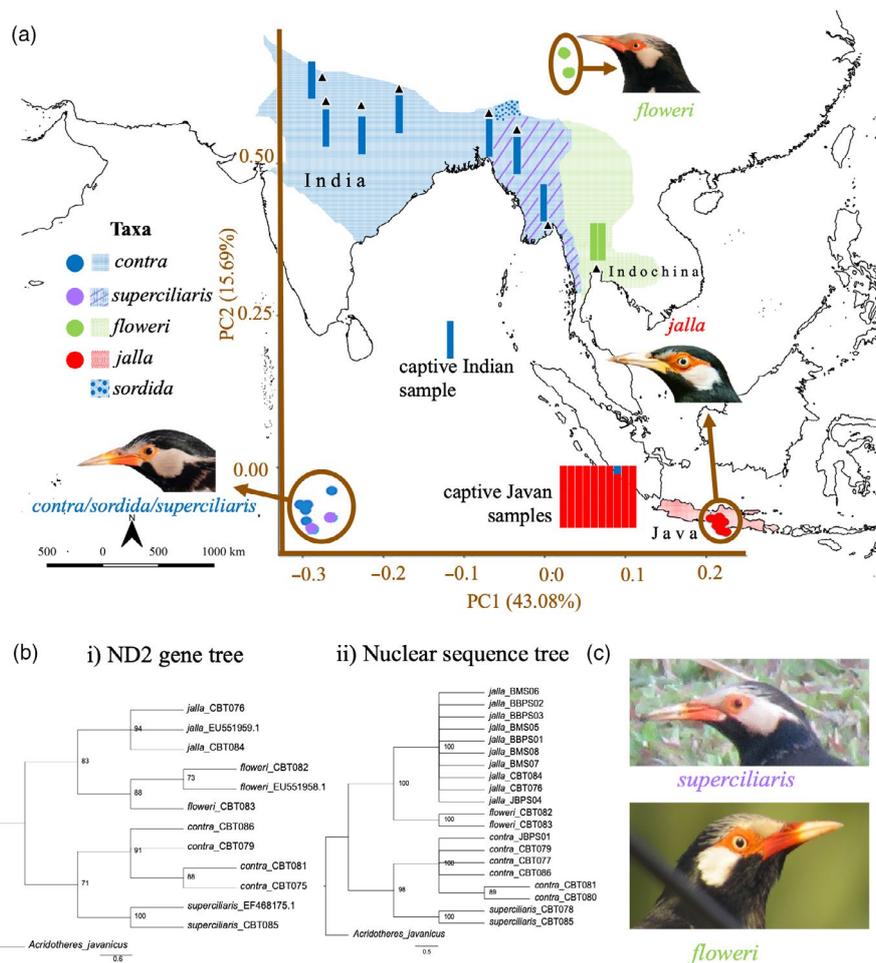


FIGURE 1 (a) Map showing the Asian distribution, sampling localities, and population structure for the five taxa of the Asian Pied Starling (*Gracupica contra*) complex: nominate *contra*, *sordida*, *superciliaris*, *floweri* and *jalla*. Black triangles depict wild sampling localities. Principal component analysis (PCA) is superimposed on the map with brown axes and based on an unlinked variant set of 936 SNPs (SNP set A2). STRUCTION results are displayed as colored columns adjacent to collection localities (one column per individual) and are based on 412 unlinked SNPs (SNP set B2) at $K = 3$, with each color representing a different ancestral contribution. Legend depicts the color of PCA points (circles) and of distribution ranges (squares) for each taxon. Coastline map sourced from Natural Earth (www.naturalearthdata.com). Bird photo credits: *contra*—Mustafa Sozen, *floweri*—Peter Ericsson, *jalla*—Boas Emmanuel. (b) Maximum-likelihood consensus trees based on (i) mitochondrial ND2 gene with bootstrap support $\geq 70\%$ shown beside nodes and (ii) concatenated nuclear sequences of 117 loci (each with >12 SNPs per locus) with bootstrap support $\geq 95\%$ shown beside nodes. GenBank accession numbers or sample codes are given behind each taxon name. (c) Pictures of *superciliaris* and *floweri* demonstrating differences in eye color and extent of bare facial skin across the two taxa. Photo credits: *superciliaris*—Thomas Brooks, *floweri*—Tom Wheatley

a lack of fresh recruitment of wild captures, and possibly in order to increase market value, illegally captive *jalla* individuals are frequently cross-bred with other taxa, especially *floweri* from Thailand, compromising the genetic integrity of Javan Pied Starlings. At the recent Songbird Crisis Summit held by the IUCN Specialist Group for Asian Songbird Trade in Singapore in 2019, *jalla* was identified as one of the top 10 taxa warranting immediate conservation action to preclude final extinction. However, a full resolution of the taxonomic status of *jalla* and other taxa within the Asian Pied Starling complex has never been achieved, and the genetic integrity of some of the last remaining captive samples of *jalla* has so far been left unaddressed. As *jalla* is deemed critically endangered (Birdlife International, 2017) and pure *jalla* individuals must now be scarce in the markets, historical museum samples are the only credible source of DNA to resolve the identity and affinities of the various components of the Asian Pied Starling complex and design appropriate conservation strategies.

In this study, we used the Asian Pied Starling complex as a model to address the complicated conservation status of endangered organisms that are subject to taxonomic uncertainty and susceptible to an erosion of genomic integrity through allelic infiltration from closely related taxa. Our use of historical samples ensures that we can reach appropriate taxonomic conclusions regardless of modern admixture practices in captivity. We used a target-enrichment approach to capture a genome-wide array of loci for analysis (Chattopadhyay et al., 2019; McCormack et al., 2016; Wood et al., 2018). Our work can serve as a blueprint for future studies in which historical samples can shed light on similar taxonomically ambiguous cases to facilitate conservation action.

2 | METHODS

2.1 | Sample collection

We obtained 12 historical tissue samples (collection dates range from 1878 to 1983) from four subspecies of the Asian Pied Starling complex (*contra*, *jalla*, *superciliaris*, and *floweri*) from the American Museum of Natural History (AMNH), University of Michigan Museum of Zoology (UMMZ), and Royal Ontario Museum (ROM) (Table S1). We were unable to procure material for *sordida*, a subspecies that is thought to be geographically restricted to eastern Assam in northeast India, although we did include DNA of a specimen collected immediately adjacent to the traditionally accepted range of *sordida* in central Assam. However, note that *sordida*'s purported distinguishing features are doubtfully distinct from nominate *contra* (Ripley, 1950), and it is variably synonymized with *contra*. We also procured nine fresh blood samples (Table S1) across two subspecies (*contra* and *jalla*) through Bali Bird Park (Indonesia) and Jurong Bird Park (Singapore) whose identity was not in doubt as they displayed the typical morphological traits of these two distinct subspecies (Figure 1a). However, as the surviving captive population of *jalla* is

known to be subject to breeding practices exposing birds to admixture, genomic comparison with historic DNA samples allows us to test their purity. We strictly adhered to the Institutional Animal Care and Use Committee's Biodiversity protocol B17-0459 while carrying out our project.

2.2 | DNA extraction

We used Qiagen DNeasy Blood and Tissue kits (Qiagen, Germany) for DNA extraction. The manufacturer's protocol was modified for the historical samples according to Chattopadhyay et al. (2019), and DNA extraction was carried out in a dedicated historical DNA facility, strictly isolated from post-PCR workspace (Cooper & Poinar, 2000). All historical DNA extractions were carried out under sterile conditions in a separate Class II Biosafety cabinet in batches. Each batch had its own negative control to monitor contamination (Krings et al., 1997). We used the Qubit® 2.0 High Sensitivity DNA Assay (Invitrogen, USA) for DNA quantification. For historical samples, we additionally used an Agilent High Sensitivity DNA kit (Agilent Technologies, USA) for fragment analysis and detection of degraded DNA.

2.3 | Library preparation

Library preparation of historical samples was carried out in a sterile environment in a biosafety cabinet with laminar airflow. The cabinet was located in an isolated space between the historical DNA clean facility and post-PCR area. Historical samples were first subjected to NEB Next FFPE DNA Repair Mix (New England BioLabs) to repair degraded and damaged regions (Pääbo et al., 2004). DNA of fresh samples was first sheared using a Bioruptor Pico (Diagenode, Belgium) with 12 cycles of sonication (conditions: 30s ON and 30s OFF) to obtain an average fragment size of ~250 bp. Subsequently, we prepared dual indexed libraries using a NEB Next Ultra DNA Library Prep kit (New England BioLabs) for sheared fresh and FFPE-treated historical DNA samples. NEB Next Multiplex Oligos for Illumina (New England BioLabs) were used for indexing each sample. As an additional safeguard against cross-contamination, we carried out historical DNA library preparation in two batches; geographically proximate specimens were preferably processed in different batches to facilitate easy detection of cross-contamination (Tonnis et al., 2005). We had separate negative controls for library preparation but we also carried forward negative controls from DNA extraction through library preparation to monitor contamination at every step of the laboratory work.

We performed DNA quantification and fragment analysis for all the sample DNA libraries and negative controls. DNA quantification and fragment analysis of negative controls revealed no quantifiable DNA but only primer-dimers and adapter sequences in the fragment profile.

2.4 | Target locus design

Target regions for bait design were chosen to account for both recent and deep divergence within the Asian Pied Starling complex and hence comprised both variable intronic regions and conserved exons (Chattopadhyay et al., 2019). We used EvolMarkers (Li et al., 2012) to identify single-copy exons (hence, filtering out paralogs) conserved across the genomes of the Javan Myna (*Acridotheres javanicus*, GCA_002849675.1), Collared Flycatcher (*Ficedula albicollis*, GCA_000247815.1), and Zebra Finch (*Taeniopygia guttata*, GCF_003957565.1). Javan Myna (Low et al., 2018) was chosen for its phylogenetic proximity to Asian Pied Starlings (Lovette et al., 2008; Lovette & Rubenstein, 2007; Zuccon et al., 2008). We set a minimum of 55% identity and an e-value <10E-15 for the BLAST (Altschul et al. 1990) search within the EvolMarkers database to identify conserved exons. Only exons >500 bp were used for downstream analysis. We used BEDtools 2.28.0 (Quinlan & Hall, 2010) to extend these targeted coding regions to include flanking intronic regions 500 bp upstream and downstream on the Javan Myna genome in order to account for shallow divergence. We further accounted for overlapping targets by merging overlapping loci (Quinlan & Hall, 2010). In addition, we only retained loci with 40%–60% GC content. Finally, we filtered for repeat regions in RepeatMasker 4.0.7 (Smit et al., 2015) to retain a total of 983 loci ranging from 1,502–8,173 bp in length. MYcoarray (USA) used these target loci to design 75,292 100-bp-long RNA baits with a 4X tiling density for in-solution hybridization.

2.5 | Sequence capture

We used these specifically designed RNA baits to enrich our previously generated DNA libraries with target-specific endogenous loci through in-solution hybridization capture. The myBaits protocol (myBaits manual version 3) was followed with modifications (Chattopadhyay et al., 2019). We used a Qubit® 2.0 High Sensitivity DNA Assay (Invitrogen, USA) to quantify the enriched libraries followed by fragment analysis (Agilent Technologies, USA). Samples were pooled at equimolar ratios and sent for paired-end sequencing (Illumina HiSeq 4000 2 × 151 bp). Fresh and historical samples were sequenced separately.

2.6 | Quality filtering

We checked the quality of the raw DNA sequence data using FastQC v0.11.7 (Babraham Bioinformatics). We used Trimmomatic v0.33 (Bolger et al., 2014) to remove adapters, reads with low quality (Phred score <20), and reads below a minimum length of 36 bp. Duplicate reads and low complexity reads were removed using FastUniq v1.1 (Xu et al., 2012) and fastp v0.20.0 (Chen et al., 2018), respectively. Subsequently, the quality of the reads was again assessed in FastQC v0.11.7 (Babraham Bioinformatics) to ensure a complete removal of adapters and a high sequence quality.

2.7 | SNP-calling pipeline for population-genomic analysis

We mapped the cleaned reads to the target loci using BWA-MEM v 0.7.12 (Li, 2013). The aligned sam files were converted to bam format using Samtools v1.9 (Li et al., 2009). Then, we added read groups to the bam files using AddOrReplaceReadGroups in PICARD v2.20.0 (Picard tools, Broad Institute, Cambridge, MA, USA) and marked duplicate reads using MarkDuplicates in PICARD v2.20.0. The reads were then realigned around indels using RealignerTargetCreator and IndelRealigner in GATK v3.8 (McKenna et al., 2010) to obtain improved local alignment. For historical samples, we processed the resultant files in mapDamage v.2.0.9 (Jónsson et al., 2013) to check for historical DNA-specific damage patterns. When we compared the results for historical and fresh samples, we found no significant damage in the historical samples. We rescaled the quality scores for further downstream analyses. We sorted the final mapped reads for historical and fresh samples in Samtools v1.9 (Li et al., 2009) and assessed their quality in Qualimap v2.2.1 (García-Alcalde et al., 2012). Finally, we obtained genotype likelihoods for the alignments using "mpileup" and called the variants from genotype likelihoods using the "call" utility in bcftools v1.9 (Li, 2011), obtaining a total of 43,050 SNPs. The variants were filtered iteratively using "vcffilter" in vcflib v1.0.0 (Garrison, 2012) and vcftools v0.1.16 (Danecek et al., 2011) to retain only reliable, well-represented, paralog-filtered, and high-quality variants by means of the following parameters: minimum depth of 20 reads per individual, minor allele count 3, minimum quality 30, minor allele frequency 0.05, minimum mean depth 30, and maximum mean depth 175. We removed individuals with >50% missing data, losing three historical samples (2 *contra* from Assam and one *jalla*; Figure S1).

To check for contamination, we carried the DNA reads from negative controls through the above pipeline. After the above steps, final negatives (i.e., those that had been carried through to in-solution hybridization) produced reads of three loci. In contrast, non-enriched negatives (i.e., those that had only been carried through library preparation) did not produce any reads mapping to target loci. This suggests that reads obtained from the DNA of final negatives were derived from nondegraded RNA baits added to negative controls during in-solution hybridization. As a precautionary step, we removed these three loci from all analyses.

We generated multiple final variant sets to account for different biases. Firstly, variants were filtered to retain only those loci with <5% missing data, resulting in a total of 6333 SNPs (SNP set A1) across 18 individuals. Secondly, to account for linkage disequilibrium between SNPs located on the same locus, we selected one random SNP from each locus and retained a final SNP set of 936 variants (SNP set A2). To further check for linkage across loci, we ran PLINK v1.9 (Chang et al., 2015) with the following parameters: sliding window of 50 SNPs, step size of 5, and pairwise linkage disequilibrium <0.1. We did not detect evidence for linkage disequilibrium between our filtered variants even at such strict parameters.

Our quality filters for missing data led to a loss of both historical samples from central Assam (northeast India), collected from close to the range of *sordida* (from eastern Assam) and possibly containing an allelic profile representative of the unsampled subspecies *sordida*. Therefore, to account for this bias, we constructed another SNP set allowing for no missingness in loci and keeping all other parameters equal. This alternative SNP-calling regime resulted in a set of 871 SNPs (SNP set B1) across 20 individuals. To account for linkage, we again chose one random SNP from each locus, resulting in a total of 412 SNPs (SNP set B2) while managing to salvage one historical sample from Assam and one historical sample of *jalla*. Last, we constructed a third SNP set only containing a selection of loci that displayed the highest levels of polymorphism; that is, each locus contained >20 SNPs to examine whether these highly divergent loci carry an underlying genomic signal that corroborates the genome-wide phylogenetic signal. For this final set, we obtained a total of 749 SNPs (SNP set C) across 25 loci.

We investigated potential pairwise kinship relationships among individuals within the same population using SNPRelate (Zheng et al., 2012) and checked for outlier loci under positive selection across all SNP sets using Bayescan v2.1 (Fischer et al., 2011) at a false discovery rate of 5%. We did not detect kinship or positive selection in any of our variant sets.

2.8 | Population-genomic analyses

We used the R package SNPRelate (Zheng et al., 2012) in RStudio v1.0.143 (RStudio Team, 2015) to construct principal component analysis (PCA) plots to visualize population structure based on all sets of linked and unlinked SNPs (A1, A2, B1, B2, and C). We further

investigated population structure using the Bayesian population clustering program STRUCTURE (Pritchard et al., 2000) and associated programs Structure Harvester (Earl, 2012) and CLUMPP (Jakobsson & Rosenberg, 2007). We ran 10 replications of our unlinked SNP sets A2 and B2 for a population cluster size of $K = 1-6$ using a burn-in period of 200,000 and 1,000,000 Monte Carlo Markov chain iterations for each replicate.

We performed an analysis to inspect the relationship between geographic distance and genomic distance specifically across the boundary between the two adjacent taxa *floweri* and *superciliaris*. We first visualized their boundary along the Thai–Burmese border mountain range (Tenasserim Hills, Dawna range, and Karen Hills) on a terrain map obtained from Natural Earth (www.naturalearthdata.com; Figure 2a). We then measured the geographic distances of the collection localities of all mainland Asian individuals (including *contra*, *superciliaris*, and *floweri*) from the line marking the terrain boundary and plotted these geographic distances against the PC2 values of our PCA plot from SNP set B2, which is representative of mainland Asian genomic differentiation in Asian Pied Starlings. The insular taxon *jalla* was excluded from this analysis as its distribution is widely disjunct and not geographically contiguous.

2.9 | Morphological inspection

We performed a morphological inspection of Asian Pied Starling photographic material from Thailand and Myanmar deposited on two online photo repositories, Oriental Bird Images (www.orientalbirdimages.org) and eBird (www.ebird.org). We only included adult birds and we eliminated replicate pictures of identical birds posted multiple times. We counted multiple birds in the same

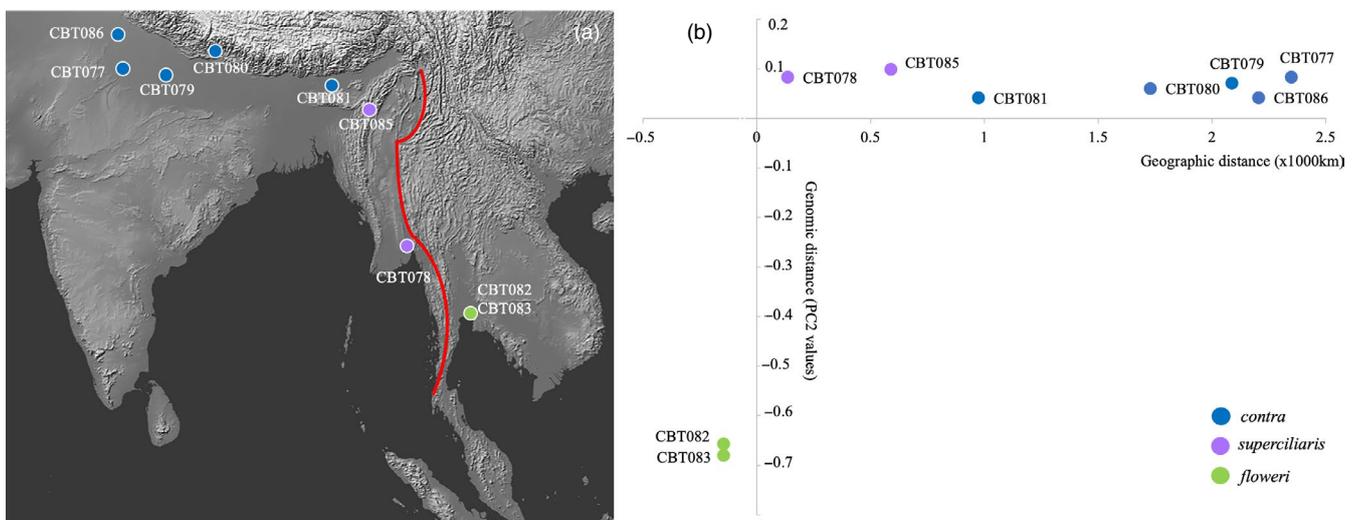


FIGURE 2 (a) Terrain map of South and South-East Asia. Each individual point marks a wild sample locality and the red line depicts the terrain boundary of the Thai–Burmese border mountain range, which separates taxa *superciliaris* (from Myanmar) and *floweri* (from Thailand). Terrain map sourced from Natural Earth (www.naturalearthdata.com); (b) Graph showing relationship between geographic and genomic distances. The x-axis refers to the geographic distance of the collection localities of *contra*, *superciliaris*, and *floweri* to the nearest point along the terrain boundary (red line in panel a), with positive values reflecting points west of the terrain boundary and negative values east of the boundary. The y-axis reflects genomic distance expressed in terms of PC2 values from the PCA plot based on SNP set B2 (Fig. S3b)

picture separately. For the taxon *floweri* from Thailand, we examined a total of 13 birds on Oriental Bird Images and a total of 318 birds on eBird. For the taxon *superciliaris* from Myanmar, we examined one bird on Oriental Bird Images and a total of 23 birds on eBird. We focused on two important morphological traits during inspection: (a) extent of bare red/orange facial skin and (b) eye color. If the length of bare facial skin extending from the posterior end of the eye toward the nape was greater than the length of bare skin extending from the anterior end of the eye to the base of the bill, the bare facial skin patch was considered large, and vice versa. The eye color was examined to check whether it was pale or dark. Traits were recorded as “indeterminate” in photographs of insufficient quality or those in which the bird was too distant or depicted at an unsuitable angle.

2.10 | Sequence assembly for phylogenetic analysis

We assembled the cleaned DNA reads using HybPiper v1.3.1 (Johnson et al., 2016) to obtain sequences of genomic target loci across all individuals. We also assembled reads for the mitochondrial gene ND2 obtained as a nonspecific product from target-enrichment sequencing, adding sequences available from the NCBI database (<https://www.ncbi.nlm.nih.gov/>; EF468175.1) as a reference.

We ran the python script *reads_first.py* from HybPiper v1.3.1, specifying “bwa” for mapping (Li, 2013) and inputting a “cov_cutoff” value of 18 for subsequent de novo assembly at each locus (Bankevich et al., 2012). We ignored HybPiper's translation output for exons. We also ran the above pipeline for the DNA output from our negative controls, which produced sequence for three loci, probably due to nondegraded RNA baits (see above). Again, we excluded these three loci from all further analyses. We used *get_seq_lengths.py* to summarize the total loci recovered and visualized this information using *gene_recovery_heatmap.R* in RStudio v1.0.143 (RStudio Team, 2015). We also ran *hybpiper_stats.py* to summarize target enrichment and locus recovery efficiency across samples (Table S2). As there was limited locus recovery (~23%) for one historical *contra* sample (CBT075) due to poor sequence quality (Table S2), we excluded this sample from further analyses. We also excluded 15 nuclear loci that were not retrieved across all samples. In summary, we retrieved a total of 965 nuclear loci for 20 samples and recovered the mitochondrial ND2 gene for 12 samples. We then used the python script *retrieve_sequences.py* in HybPiper v1.3.1 (Johnson et al., 2016) to collate sequences from all individual fasta files as a multifasta file for each locus.

2.11 | Tree construction

We constructed gene trees for all nuclear sequence loci containing >20 SNPs in RAxML-NG v.0.9.0 (Kozlov et al., 2019). This resulted in 24 individual gene trees (minus one gene that was excluded due

to missing data during HybPiper assembly). We also constructed a gene tree for the mitochondrial gene ND2 using all samples in which the recovery length was >960 bp. We supplemented our dataset with mitochondrial sequences available on the NCBI database (EF468175.1, EU551958.1, EU551959.1). We used sequences from the NCBI database belonging to the Javan Myna (*Acridotheres javanicus*) as an outgroup (nuclear loci—GCA_002849675.1, ND2—EU403593.1).

The resultant multifasta files were aligned in MAFFT v7.388 (Katoh & Standley, 2013). We then used Phyutility v.2.2.6 (Smith & Dunn, 2008) to clean aligned multifasta files removing sites with >50% missing data. We ran jModelTest v.2.1.10 (Darriba et al., 2012) to identify the best nucleotide substitution model for each locus based on the corrected Akaike information criterion (AICc). Finally, we constructed maximum-likelihood (ML) gene trees in RAxML-NG v.0.9.0 (Kozlov et al., 2019) using the best nucleotide substitution model as identified by jModelTest (Table S3) with 5,000 bootstrap iterations. Trees were visualized in FigTree v1.4.4 (Rambaut, 2018). We additionally computed uncorrected pairwise distances for the ND2 gene in MEGA7 (Kumar et al., 2016).

To concentrate phylogenetic signal, we selected all our sequence loci with >12 SNPs per locus to construct a concatenated sequence tree across constituent taxa, resulting in an alignment of 299,371 bp consisting of a total of 117 loci for which sequences had been assembled through HybPiper. We obtained the outgroup sequence for these loci from the Javan Myna genome on NCBI (GCA_002849675.1). We first aligned the multifasta files for each locus (Katoh & Standley, 2013) and cleaned the aligned multifasta files to remove sites with >50% missing data (Smith & Dunn, 2008). We concatenated the resultant files in SequenceMatrix (Vaidya et al., 2011) and exported the concatenated file in phylip format. We constructed an ML tree for the concatenated sequence file in RAxML-NG v.0.9.0 (Kozlov et al., 2019) using the general time-reversible (GTR) model with 5,000 bootstrap iterations. The tree was visualized in FigTree v1.4.4 (Rambaut, 2018).

2.12 | Gene ontology

For the 25 selected nuclear sequence loci (i.e., those with >20 SNPs per locus), we further performed gene ontology (GO) enrichment tests to identify whether functional loci associated with molecular, biological, or cellular processes might be particularly represented in these stretches of sequence. We annotated the entire set of target loci using Blast2GO (Conesa et al., 2005). We then performed the GO enrichment test in Blast2GO for the 25 selected loci using—as a reference set—all 980 loci excluding the three loci that had been removed from analyses because of negative control bias (see above). The test was performed using corrected *p*-values at a false discovery rate of .05. We also performed over-representation analysis (ORA) using the web-based tool WebGestalt (Zhang et al., 2005).

TABLE 1 Pairwise genetic distance matrix between *contra*, *superciliaris*, *floweri* and *jalla* (within the Asian Pied Starling [*Gracupica contra*] complex) and Javan Myna (*Acridotheres javanicus*) based on the mitochondrial ND2 gene

	<i>contra</i>	<i>superciliaris</i>	<i>floweri</i>	<i>jalla</i>
<i>contra</i>				
<i>superciliaris</i>	1.6%–1.9%			
<i>floweri</i>	2.6%–3.1%	2.9%–3.2%		
<i>jalla</i>	2.8%–3.2%	3.0%–3.1%	1.7%–2.1%	
<i>Acridotheres javanicus</i>	10.3%–10.6%	9.8%	10.2%–10.4%	10.4%–10.5%

Note: Mega7 was used to calculate raw, uncorrected pairwise distances using the pairwise deletion option for missing data.

3 | RESULTS

We found no significant DNA damage in our historical samples, with a deamination incidence at 0%–1% at read termini (Table S4). Based on the number of reads that mapped to their target compared to the total number of reads, target-enrichment efficiency ranged between 24%–60% for historical samples (except for two poor-quality samples) and 58%–75% for fresh samples (Table S2). Based on the number of loci retrieved compared to the total number of target loci, target-enrichment efficiency was >98% except for one poor-quality historical sample of *contra* that was excluded from further analysis (Table S2). The length of loci recovered was >80% of total locus length for most samples except for a few degraded historical samples (Figure S2).

PCAs generally indicated three genomic clusters within the Asian Pied Starling complex; *contra* and *superciliaris* were genomically proximate within one cluster, whereas *floweri* and *jalla* formed two distinct clusters (Figure 1a, Figure S3a–c). For both *jalla* and *contra*, the contemporary and historical samples of each taxon formed a single spatial unit, confirming that our dataset did not suffer from ancient DNA artifacts and indicating an absence of admixture in our contemporary samples (Figure 1a, Figure S3a–c).

STRUCTURE analysis revealed the same tripartition of Asian Pied Starlings across all variant sets even at $K = 4, 5,$ and 6 (Figure 1a, Figure S4). When enforcing $K = 2$, *contra* and *superciliaris* formed one genomic division and *jalla* formed another genomic division, whereas *floweri* displayed genomic admixture between these two groups with a slightly greater proportion of a *jalla* contribution in its genome (Figure S4). One historical *jalla* sample, CBT076, showed traces of potential *contra* admixture in its STRUCTURE profile at $K = 3$ (Figure 1a). However, we interpret this pattern as a likely artifact due to heavy DNA damage as this was one of the samples with

the poorest DNA quality (high missing data; Figure S1), which was specifically salvaged in SNP set B. The three distinct genomic Pied Starling clusters became less well defined, with more blurred boundaries, when only selected loci with a high SNP count (>20 SNPs per locus) were analyzed (Figure S3d).

We constructed gene trees for the 25 loci with the highest SNP count (>20 variants per locus) to investigate whether loci of high divergence carry a phylogenetic signal that reflects the overarching genome-wide phylogenetic signal. In general, the phylogenetic signal across these 25 loci was mutually incongruent, and we failed to identify cohorts of candidate loci that would have converged on a particular phylogenetic signal—whether equal or different from the full dataset's signal (Figure S5). Some gene trees retrieved high support for a distinct clade of *jalla*, explaining its separation in the PCA of SNP set C (Figure S3d). To examine whether this distinct placement of *jalla* may be driven by adaptive processes reflected in a subset of loci, we carried out GO enrichment tests and ORA analyses for these selected loci. We found no statistically significant GO terms (Figure S6; Tables S5 and S6), which is consistent with expectations from our low locus sample size (Mishra et al., 2014). Some of the biological processes that were most frequently identified as overrepresented among the genes adjacent to our candidate loci include metabolic processes and biological regulation (Figure S7).

All four taxa analyzed emerged as monophyletic lineages with strong support in both the concatenated nuclear tree and the mitochondrial ND2 tree (Figure 1b). Geographic distances notwithstanding, the taxon *floweri* emerged as sister to *jalla*, not to *contra*/*superciliaris* (Figure 1b). In accordance with the population-genetic results, the taxon *superciliaris* emerged as most closely related to *contra* in the concatenated nuclear tree (Figure 1b). The mitochondrial ND2 sequence divergence between the *superciliaris*-*contra*

cluster and the *floweri-jalla* cluster is approximately 3%, whereas *jalla* and *floweri* differ by approximately 2% (Table 1). The taxa *superciliaris* and *contra* differ by approximately 1.7% (Table 1).

A graph plotted between geographic distance versus genomic distance (as represented by PC2 in SNP set B2; Fig. S3b) among all geographically contiguous individuals belonging to the mainland Asian taxa illustrates that genomic variation in Asian Pied Starlings is not arranged in a geographically clinal fashion, but that there is a deep genomic division between *floweri* and *superciliaris* along a short geographic distance across the Thai–Burmese border range (Figure 2b).

Inspection of morphological traits revealed that all *floweri* individuals from Thailand for which conclusive visual material was available had a pale iris color (Table S7a). In contrast, none of the *superciliaris* individuals examined from Myanmar showed a pale iris color (Table S7a). Among individuals for which conclusive visual confirmation was possible, almost 86% of *floweri* but only 8% of *superciliaris* showed a facial morphology consistent with our definition of a large extent of bare red/orange skin around the eye (Table S7b).

4 | DISCUSSION

4.1 | Genomic tripartition of the Asian Pied Starling complex

Our use of genome-wide DNA markers established three distinct evolutionary lineages within the Asian Pied Starling complex (Figure 1a, b): (a) the *contra-superciliaris* cluster; (b) *floweri*; and (c) *jalla*. Mitochondrial DNA information corroborated this result, furnishing pairwise sequence distances between the three lineages that are within the confines of sequence divergence thresholds used for species delimitation by the general bird DNA barcoding movement (Table 1; Hajibabaei et al., 2006; Hebert et al., 2004; Johnston et al., 2011; Tizard et al., 2019), although the divergence between *jalla* (from Java) and *floweri* (from Thailand)—at ~2%—is at the lower end of divergences typically displayed by good species.

While species status for insular *jalla* from Java has been previously proposed (Birdlife International, 2020; del Hoyo et al., 2020), mainland taxa of the Asian Pied Starling are geographically contiguous across a vast range spanning ~4,000 km from easternmost Pakistan to eastern Cambodia (Figure 1a) and have never been divided into two different species in modern times. However, the deepest divergences in the complex are found between Thai *floweri* and adjacent Burmese *superciliaris* (Table 1), and our genome-wide markers refute the hypothesis that the substantial genomic differentiation across mainland Asia is arranged in a geographic cline (Figure 2). Instead, when analyzing genomic profiles across an east–west gradient, we observe a clear spike of differentiation between samples in relatively close geographic proximity: one *superciliaris* from Bago, Myanmar, and two *floweri* samples from Bangkok, Thailand (Figure 2b). This leap in genomic differentiation across the Thai–Burmese border range (Tenasserim Hills, Dawna range, and Karen Hills) suggests a strong

barrier to gene flow for these starlings (Figure 2). These mountains are a known barrier to dispersal across multiple vertebrate lineages, with recent demonstrations of deep, species-level divergence in bats (Puechmaile et al., 2011) and bulbuls (Garg et al., 2016). In contrast, the genomic proximity of *contra* and *superciliaris* suggests that the Indo-Burmese mountain range (Arakan mountains, Chin Hills, Naga Hills) is not an effective barrier to gene flow in these starlings (Figure 2), corroborating results from studies on other lowland vertebrates (e.g., Iyengar et al., 2005; Sadanandan et al., 2018).

4.2 | Genomic–morphological congruence

Morphologically, the insular Javan form *jalla* has long been interpreted as the most distinct component of the Asian Pied Starling complex (e.g., Horsfield, 1821), accounting for its occasional treatment as a separate species (Birdlife International, 2020; del Hoyo et al., 2020). Our genomic result of a primary divide of Asian Pied Starlings across the Thai–Burmese border range (Figure 2a) may therefore be seen as contradicting morphological evidence. However, previous morphological inquiry has been restricted to broad qualitative assessments of plumage traits on dead museum skins. Close inspection of a collection of photographs of live individuals suggests that a previously overlooked suite of phenotypic characters may be disproportionately important in sexual advertisement and mate choice and consequently also in species delimitation across these starlings. Chief among these neglected traits is the extent of bare red facial skin and eye color. The extent of red facial skin is much larger to the east of the Thai–Burmese border range than to the west, and the eye color is invariably pale to the east of the Thai–Burmese border range but dark to the west (Figure 1c, Table S7). Crucially, such traits are not visible on dead museum skins, which only preserve the plumage of a bird but do not allow for an assessment of bare parts. Our study corroborates the importance of eye color and other bare parts as long-neglected traits in avian species delimitation, as recently shown by another genomic analysis on South-East Asian passerines with deep divergences across the Thai–Burmese border range (Garg et al., 2016). Ultimately, *floweri*'s closer genomic affinity with *jalla*, rather than with *superciliaris/contra* (Figure 1a,b), can easily be accounted for by the Pleistocene land bridges that have connected Java and mainland Asia during Quaternary periods of sea-level change (Peterson et al., 2015; Voris, 2000; Whittaker & Fernández-Palacios, 2007).

One sample from central Assam (CBT081) collected adjacent to the type locality of *sordida* in eastern Assam clustered with other individuals of *contra* in all our genomic analyses (Figure 1a,b). Due to its geographic proximity, this sample probably provides a close representation of the genomic profile of the unsampled subspecies *sordida*. Noting the dubious morphological distinction of *sordida* from the nominate *contra* and discrepancies regarding the exact circumscription of its distribution (Ripley, 1950), our genomic results (Figure 1a,b) are consistent with a synonymization of *sordida* with *contra* in future taxonomic classifications.

In summary, we propose a separation of the complex into the following three species: (a) *Gracupica contra* (Indian Pied Starling), with subspecies *contra* (most of the Indian subcontinent, including a synonymized *sordida* from eastern Assam) and *superciliaris* (mostly Myanmar and adjacent Yunnan, with a southeastern range boundary here confirmed to include southern Myanmar); (b) *Gracupica floweri* (Thai Pied Starling), monotypic (mostly Thailand and Cambodia); and (c) *Gracupica jalla* (Javan Pied Starling), monotypic (historically Java and Bali, now extinct in the wild).

4.3 | Use of historical DNA in the conservation of endangered taxa affected by introgression

Conservation management often hinges upon clear taxonomic resolution (Ely et al., 2017; McCarthy, 2014; Regan et al., 2002). In many terminally endangered organisms, taxonomic research is impaired by the rarity of the species involved (Tsai et al., 2019) exacerbated by potential genomic admixture affecting the last surviving populations (Pilot et al., 2019). Human-mediated hybridization, either in captivity or in the wild via habitat shifts (Mallet et al., 2011), invasions by alien species (Bleeker et al., 2007), wildlife trade (Nijman et al., 2018), and accidental release (Baveja et al., 2019), among other factors, may compromise the genetic integrity of the last survivors of a terminally endangered population and confound taxonomic research. For instance, one of our study lineages, the Javan Pied Starling (*G. jalla*), is heavily traded in Indonesian markets and frequently admixed with *floweri* and *superciliaris* in illegal captivity due to a lack of recruitment of new individuals resulting from their extinction in the wild. Our study was only possible because of the availability of museum samples from a time preceding such introgression.

4.4 | Conservation recommendations

The captive *jalla* individuals from Bali Bird Park emerged with an identical genomic profile to historical *jalla* samples, indicating a fortunate lack of admixture (Figure 1a). They are therefore suitable founding individuals for a conservation breeding program of Javan Pied Starlings, which—at the time of writing—does not yet exist. In 2018, the Ministry of Environment and Forestry of Indonesia removed Javan Pied Starlings (local name: *jalak suren*) and four other species of songbirds from the list of protected species because of advocacy work by the bird-keepers' lobby (Jakarta Hidayat, 2020; Post, 2018). Despite the Javan Pied Starling's extinction in the wild, the previous lack of a scientific consensus regarding its taxonomic status has probably contributed to hesitations in affording this taxon full protection. Urgent conservation attention is hence imperative to prevent the species from going extinct even in captivity, for example, through rampant cross-breeding with *G. floweri*.

At a time when the survival of more and more species hinges on conservation breeding (Chattopadhyay et al., 2019; Çilingir et al., 2017; Galla et al., 2020; Wright et al., 2020), our study is a

unique demonstration of the successful use of historical DNA material to derive genome-wide DNA marker sets that clarify the taxonomic status of a species that is extinct in the wild and increasingly admixed in captivity. Historical DNA is a potentially underutilized resource (Bi et al., 2013; Payne & Sorenson, 2002; Shaffer et al., 1998; Suarez & Tsutsui, 2004). In conservation problems such as the pied starlings of Asia, a definitive conservation strategy would likely be impossible to design without the clear evidence provided by recourse to historical DNA. We encourage a broader utilization of historical samples using the laboratory and bioinformatic methods implemented in our study for a lasting resolution of controversial dilemmas in conservation involving terminally endangered species.

DATA ARCHIVING STATEMENT

The raw sequence data for this study are available at Sequence Read Archive, NCBI under BioProject PRJNA669910. Additional details and results of this study are available in the supporting information of this article.

ACKNOWLEDGEMENTS

We thank the American Museum of Natural History (AMNH, New York), University of Michigan Museum of Zoology (UMMZ, Ann Arbor), Royal Ontario Museum (ROM, Canada), Jurong Bird Park (Singapore), and Bali Bird Park (Indonesia) for providing historical and fresh samples. We also thank G.W. Low, M.Y. Wu, and C.Y. Gwee for help and support during research. This work was funded by Wildlife Reserves Singapore (WRS) and cofunded by the Croeni Foundation (grant numbers WBS R-154-000-A99-592 and WBS R-154-000-A05-592). B.C. acknowledges funding from South East Asian Biodiversity Genomics (SEABIG) (grant numbers WBS R-154-000-648-646 and WBS R-154-000-648-733). K.M.G. acknowledges support from the Ramanujan Fellowship of Science and Engineering Research Board (SERB).

ORCID

Frank E. Rheindt  <https://orcid.org/0000-0001-8946-7085>

REFERENCES

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Baveja, P., Tang, Q., Lee, J. G., & Rheindt, F. E. (2019). Impact of genomic leakage on the conservation of the endangered Milky Stork. *Biological Conservation*, 229, 59–66. <https://doi.org/10.1016/j.biocon.2018.11.009>
- Bi, K., Linderoth, T., Vanderpool, D., Good, J. M., Nielsen, R., & Moritz, C. (2013). Unlocking the vault: Next-generation museum population genomics. *Molecular Ecology*, 22(24), 6018–6032. <https://doi.org/10.1111/mec.12516>
- BirdLife International. (2017). *Gracupica jalla* (amended version of 2017 assessment). The IUCN Red List of Threatened Species 2017: e.T103890801A118590020. <https://doi.org/10.2305/IUCN.UK.2017-3.RLTS.T103890801A118590020.en>

- BirdLife International. (2020). Species factsheet: *Gracupica jalla*. Retrieved from <http://www.birdlife.org> (accessed March 2020).
- Bleeker, W., Schmitz, U., & Ristow, M. (2007). Interspecific hybridisation between alien and native plant species in Germany and its consequences for native biodiversity. *Biological Conservation*, 137(2), 248–253. <https://doi.org/10.1016/j.biocon.2007.02.004>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*, 4(1), s13742–s14015. <https://doi.org/10.1186/s13742-015-0047-8>
- Chattopadhyay, B., Garg, K. M., Mendenhall, I. H., & Rheindt, F. E. (2019). Historic DNA reveals Anthropocene threat to a tropical urban fruit bat. *Current Biology*, 29(24), R1299–R1300. <https://doi.org/10.1016/j.cub.2019.11.013>
- Chattopadhyay, B., Garg, K. M., Soo, Y. J., Low, G. W., Frechette, J. L., & Rheindt, F. E. (2019). Conservation genomics in the fight to help the recovery of the critically endangered Siamese crocodile *Crocodylus siamensis*. *Molecular Ecology*, 28(5), 936–950.
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chng, S. C. L., Eaton, J. A., Krishnasamy, K., Shepherd, C. R., & Nijman, V. (2015). *In the market for extinction: An inventory of Jakarta's bird markets*. Petaling Jaya, Malaysia: TRAFFIC.
- Çilingir, F. G., Rheindt, F. E., Garg, K. M., Platt, K., Platt, S. G., & Bickford, D. P. (2017). Conservation genomics of the endangered Burmese roofed turtle. *Conservation Biology*, 31(6), 1469–1476. <https://doi.org/10.1111/cobi.12921>
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18), 3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Cooper, A., & Poinar, H. N. (2000). Ancient DNA: Do it right or not at all. *Science*, 289(5482), 1139. <https://doi.org/10.1126/science.289.5482.1139b>
- Craig, A., Feare, C., & Christie, D. A. (2020). Asian Pied Starling (*Gracupica contra*). In J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, & E. de Juana (Eds.), *Handbook of the birds of the world alive*. Lynx Edicions. Retrieved from <https://www.hbw.com/node/60858> (accessed March 2020).
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772. <https://doi.org/10.1038/nmeth.2109>
- del Hoyo, J., Collar, N., Sharpe, C. J., & Christie, D. A. (2020). Javan Pied Starling (*Gracupica jalla*). In J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, & E. de Juana (Eds.), *Handbook of the birds of the world alive*. Lynx Edicions. Retrieved from <https://www.hbw.com/node/60858>
- Derkarabetian, S., Benavides, L. R., & Giribet, G. (2019). Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. *Molecular Ecology Resources*, 19(6), 1531–1544. <https://doi.org/10.1111/1755-0998.13072>
- Earl, D. A., & von Holdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Eaton, J. A., Shepherd, C. R., Rheindt, F. E., Harris, J. B. C., van Balen, S., Wilcove, D. S., & Collar, N. J. (2015). Trade-driven extinctions and near-extinctions of avian taxa in Sundaic Indonesia. *Forktail*, 31, 1–12.
- Ely, C. V., de Loreto Bordignon, S. A., Trevisan, R., & Boldrini, I. I. (2017). Implications of poor taxonomy in conservation. *Journal for Nature Conservation*, 36, 10–13.
- Fischer, M. C., Foll, M., Excoffier, L., & Heckel, G. (2011). Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology*, 20(7), 1450–1462. <https://doi.org/10.1111/j.1365-294X.2011.05015.x>
- Galla, S. J., Moraga, R., Brown, L., Cleland, S., Hoepfner, M. P., Maloney, R. F., Richardson, A., Slater, L., Santure, A. W., & Steeves, T. E. (2020). A comparison of pedigree, genetic and genomic estimates of relatedness for informing pairing decisions in two critically endangered birds: Implications for conservation breeding programmes worldwide. *Evolutionary Applications*, 13(5), 991–1008. <https://doi.org/10.1111/eva.12916>
- García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T. F., & Conesa, A. (2012). Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics*, 28(20), 2678–2679. <https://doi.org/10.1093/bioinformatics/bts503>
- Garg, K. M., Tizard, R., Ng, N. S., Cros, E., Dejtaradol, A., Chattopadhyay, B., Pwint, N., Päckert, M., & Rheindt, F. E. (2016). Genome-wide data help identify an avian species-level lineage that is morphologically and vocally cryptic. *Molecular Phylogenetics and Evolution*, 102, 97–103. <https://doi.org/10.1016/j.ympev.2016.05.028>
- Garrison, E. (2012). *Vcflib: A C++ library for parsing and manipulating VCF files*. GitHub. Retrieved from <https://github.com/ekg/vcflib>
- Gill, F., Donsker, D., & Rasmussen, P. (Eds). (2020). IOC world bird list (v10.2). <https://doi.org/10.14344/IOC.ML.10.2>
- Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., & Hebert, P. D. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences*, 103(4), 968–971. <https://doi.org/10.1073/pnas.0510466103>
- Hebert, P. D., Stoeckle, M. Y., Zemlak, T. S., & Francis, C. M. (2004). Identification of birds through DNA barcodes. *PLoS Biology*, 2(10), e312. <https://doi.org/10.1371/journal.pbio.0020312>
- Hidayat, D. (2020). Demi Kontes Burung Dalam Sangkar. *Tempo 8 Maret 2020*, 64–65.
- Horsfield, T. (1821). XIV. Systematic arrangement and description of birds from the Island of Java. *Transactions of the Linnean Society of London*, 13(1), 133–200.
- Iyengar, A., Babu, V. N., Hedges, S., Venkataraman, A. B., Maclean, N., & Morin, P. A. (2005). Phylogeography, genetic structure, and diversity in the dhole (*Cuon alpinus*). *Molecular Ecology*, 14(8), 2281–2297. <https://doi.org/10.1111/j.1365-294X.2005.02582.x>
- Jakarta Post. September 2018. Retrieved from <https://www.thejakartapost.com/news/2018/09/08/three-songbirds-removed-from-protected-list.html> (accessed March 2020).
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zerega, N. J., & Wickett, N. J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences*, 4(7), 1600016. <https://doi.org/10.3732/apps.1600016>
- Johnston, A. R., Morikawa, M. K., Ntie, S., & Anthony, N. M. (2011). Evaluating DNA barcoding criteria using African duiker antelope (Cephalophinae) as a test case. *Conservation Genetics*, 12(5), 1173. <https://doi.org/10.1007/s10592-011-0220-2>

- Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *35*(21), 4453–4455. <https://doi.org/10.1093/bioinformatics/btz305>
- Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M., & Pääbo, S. (1997). Neandertal DNA sequences and the origin of modern humans. *Cell*, *90*(1), 19–30. [https://doi.org/10.1016/S0092-8674\(00\)80310-4](https://doi.org/10.1016/S0092-8674(00)80310-4)
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874.
- Li, C., Riethoven, J. J. M., & Naylor, G. J. (2012). EvolMarkers: A database for mining exon and intron markers for evolution, ecology and conservation studies. *Molecular Ecology Resources*, *12*(5), 967–971. <https://doi.org/10.1111/j.1755-0998.2012.03167.x>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Linck, E. B., Hanna, Z. R., Sellas, A., & Dumbacher, J. P. (2017). Evaluating hybridization capture with RAD probes as a tool for museum genomics with historical bird specimens. *Ecology and Evolution*, *7*(13), 4755–4767. <https://doi.org/10.1002/ece3.3065>
- Lovette, I. J., McCleery, B. V., Talaba, A. L., & Rubenstein, D. R. (2008). A complete species-level molecular phylogeny for the “Eurasian” starlings (Sturnidae: *Sturnus*, *Acridotheres*, and allies): Recent diversification in a highly social and dispersive avian group. *Molecular Phylogenetics and Evolution*, *47*(1), 251–260. <https://doi.org/10.1016/j.ympev.2008.01.020>
- Lovette, I. J., & Rubenstein, D. R. (2007). A comprehensive molecular phylogeny of the starlings (Aves: Sturnidae) and mockingbirds (Aves: Mimidae): Congruent mtDNA and nuclear trees for a cosmopolitan avian radiation. *Molecular Phylogenetics and Evolution*, *44*(3), 1031–1056. <https://doi.org/10.1016/j.ympev.2007.03.017>
- Low, G. W., Chattopadhyay, B., Garg, K. M., Irestedt, M., Ericson, P. G. P., Yap, G., Tang, Q., Wu, S., & Rheindt, F. E. (2018). Urban landscape genomics identifies fine-scale gene flow patterns in an avian invasive. *Heredity*, *120*, 138–153.
- Mallet, J., Wynne, I. R., & Thomas, C. D. (2011). Hybridisation and climate change: Brown argus butterflies in Britain (*Polyommatus* subgenus *Aricia*). *Insect Conservation and Diversity*, *4*(3), 192–199. <https://doi.org/10.1111/j.1752-4598.2010.00122.x>
- Marshall, H., Collar, N. J., Lees, A. C., Moss, A., Yuda, P., & Marsden, S. J. (2019). Spatio-temporal dynamics of consumer demand driving the Asian Songbird Crisis. *Biological Conservation*, *241*, 108237.
- McCarthy, M. A. (2014). Contending with uncertainty in conservation management decisions. *Annals of the New York Academy of Sciences*, *1322*(1), 77. <https://doi.org/10.1111/nyas.12507>
- McCormack, J. E., Tsai, W. L., & Faircloth, B. C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, *16*(5), 1189–1203. <https://doi.org/10.1111/1755-0998.12466>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Mishra, P., Törönen, P., Leino, Y., & Holm, L. (2014). Gene set analysis: Limitations in popular existing methods and proposed improvements. *Bioinformatics*, *30*(19), 2747–2756. <https://doi.org/10.1093/bioinformatics/btu374>
- Nedoluzhko, A. V., Sharko, F. S., Tsygankova, S. V., Boulygina, E. S., Barmintseva, A. E., Krasivskaya, A. A., Ibragimova, A. S., Gruzdeva, N. M., Rastorguev, S. M., & Mugue, N. S. (2020). Molecular phylogeny of one extinct and two critically endangered Central Asian sturgeon species (genus *Pseudoscaphirhynchus*) based on their mitochondrial genomes. *Scientific Reports*, *10*, 722. <https://doi.org/10.1038/s41598-020-57581-y>
- Nijman, V., Langgeng, A., Birot, H., Imron, M. A., & Nekaris, K. A. I. (2018). Wildlife trade, captive breeding and the imminent extinction of a songbird. *Global Ecology and Conservation*, *15*, e00425. <https://doi.org/10.1016/j.gecco.2018.e00425>
- Pääbo, S., Poinar, H., Serre, D., Jaenicke-Després, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., & Hofreiter, M. (2004). Genetic analyses from ancient DNA. *Annual Review of Genetics*, *38*, 645–679. <https://doi.org/10.1146/annurev.genet.37.110801.143214>
- Payne, R. B., & Sorenson, M. D. (2002). Museum collections as sources of genetic data. *Bonner Zoologische Beiträge*, *51*, 97–104.
- Peterson, A. T., Moyle, R. G., Campillo, L. C., Hosner, P. A., Klicka, L. B., Lim, H. C., Nyári, Á. S., Reddy, S., Sheldon, F. H., & Zou, F. S. (2015). Avian evolution and speciation in the Southeast Asian tropics. *Current Zoology*, *61*(5), 898–900. <https://doi.org/10.1093/czoolo/61.5.898>
- Pilot, M., Moura, A. E., Okhlopkov, I. M., Mamaev, N. V., Alagaili, A. N., Mohammed, O. B., Yavruyan, E. G., Manaseryan, N. H., Hayrapetyan, V., Kopaliani, N., Tsingarska, E., Krofel, M., Skoglund, P., & Bogdanowicz, W. (2019). Global phylogeographic and admixture patterns in grey wolves and genetic legacy of an ancient Siberian lineage. *Scientific Reports*, *9*(1), 1–13. <https://doi.org/10.1038/s41598-019-53492-9>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.
- Prost, S., Machado, A. P., Zumbroich, J., Preier, L., Mahtani-Williams, S., Guschanski, K., Brealey, J. C., Fernandes, C., Vercammen, P., Godsall-Bottriell, L., Bottriell, P., Dalton, D. L., Kotze, A., & Burger, P. A. (2020). Conservation genomic analyses of African and Asiatic cheetahs (*Acinonyx jubatus*) across their current and historic species range. *bioRxiv* <https://doi.org/10.1101/2020.02.14.949081>
- Prüfer, K. (2018). snpAD: An ancient DNA genotype caller. *Bioinformatics*, *34*(24), 4165–4171. <https://doi.org/10.1093/bioinformatics/bty507>
- Puechmaile, S. J., Gouilh, M. A., Piyapan, P., Yokubol, M., Mie, K. M., Bates, P. J., Satasook, C., Nwe, T., Bu, S. S. H., Mackie, I. J., Petit, E. J., & Teeling, E. C. (2011). The evolution of sensory divergence in the context of limited gene flow in the bumblebee bat. *Nature Communications*, *2*(1), 1–9. <https://doi.org/10.1038/ncomms1582>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Rambaut, A. (2018). FigTree v1. 4.4, a Graphical Viewer of Phylogenetic Trees. 2014. Java. Retrieved from <https://github.com/rambaut/figtree>
- Regan, H. M., Colyvan, M., & Burgman, M. A. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications*, *12*(2), 618–628. [https://doi.org/10.1890/1051-0761\(2002\)012\[0618:ATATOU\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2002)012[0618:ATATOU]2.0.CO;2)

- Ripley, S. D. (1950). *Notes on Indian Birds III.: Birds from Assam*. New Haven, CT: Yale Peabody Museum of Natural History.
- Rose, D. C., Amano, T., González-Varo, J. P., Mukherjee, N., Robertson, R. J., Simmons, B. I., Wauchope, H. S., & Sutherland, W. J. (2019). Calling for a new agenda for conservation science to create evidence-informed policy. *Biological Conservation*, 238, 108222. <https://doi.org/10.1016/j.biocon.2019.108222>
- Rowe, K. C., Singhal, S., Macmanes, M. D., Ayroles, J. F., Morelli, T. L., Rubidge, E. M., Bi, K. E., & Moritz, C. C. (2011). Museum genomics: Low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources*, 11(6), 1082–1092. <https://doi.org/10.1111/j.1755-0998.2011.03052.x>
- RStudio Team (2015). *RStudio: Integrated development for R*. RStudio, Inc.
- Sadanandan, K. R., Olsson, U., Tizard, R., Khan, A. A., Naing, T. Z., Lin, N., Pwint, N., & Rheindt, F. E. (2018). Genetic and biometric variation across the fragmented range of Jerdon's Babbler, *Chrysomma altirostre*, a threatened Oriental grassland specialist. *Journal of Ornithology*, 159(2), 575–579. <https://doi.org/10.1007/s10336-018-1530-7>
- Schubert, M., Ermini, L., Sarkissian, C. D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., & Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, 9(5), 1056. <https://doi.org/10.1038/nprot.2014.063>
- Schuster, S. C. (2007). Next-generation sequencing transforms today's biology. *Nature Methods*, 5(1), 16. <https://doi.org/10.1038/nmeth1156>
- Shaffer, H. B., Fisher, R. N., & Davidson, C. (1998). The role of natural history collections in documenting species declines. *Trends in Ecology & Evolution*, 13(1), 27–30. [https://doi.org/10.1016/S0169-5347\(97\)01177-4](https://doi.org/10.1016/S0169-5347(97)01177-4)
- Smit, A. F. A., Hubley, R., & Green, P. (2015). RepeatMasker Open-4.0. 2013–2015. Retrieved from <http://www.repeatmasker.org>
- Smith, S. A., & Dunn, C. W. (2008). Phyutility: A phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*, 24(5), 715–716. <https://doi.org/10.1093/bioinformatics/btm619>
- Suarez, A. V., & Tsutsui, N. D. (2004). The value of museum collections for research and society. *BioScience*, 54(1), 66–74. [https://doi.org/10.1641/0006-3568\(2004\)054\[0066:TVOMCF\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2)
- Tizard, J., Patel, S., Waugh, J., Tavares, E., Bergmann, T., Gill, B., Norman, J., Christidis, L., Scofield, P., Haddrath, O., Baker, A., Lambert, D., & Millar, C. (2019). DNA barcoding a unique avifauna: An important tool for evolution, systematics and conservation. *BMC Evolutionary Biology*, 19(1), 52. <https://doi.org/10.1186/s12862-019-1346-y>
- Tonnis, B., Grant, P. R., Grant, B. R., & Petren, K. (2005). Habitat selection and ecological speciation in Galápagos warbler finches (*Certhidea olivacea* and *Certhidea fusca*). *Proceedings of the Royal Society of London B: Biological Sciences* 272(1565), 819–826.
- Tsai, W. L., Mota-Vargas, C., Rojas-Soto, O., Bhowmik, R., Liang, E. Y., Maley, J. M., Zarza, E., & McCormack, J. E. (2019). Museum genomics reveals the speciation history of *Dendrortyx* wood-partridges in the Mesoamerican highlands. *Molecular Phylogenetics and Evolution*, 136, 29–34. <https://doi.org/10.1016/j.ympev.2019.03.017>
- Vaidya, G., Lohman, D. J., & Meier, R. (2011). SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, 27(2), 171–180. <https://doi.org/10.1111/j.1096-0031.2010.00329.x>
- Voris, H. K. (2000). Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations. *Journal of Biogeography*, 27(5), 1153–1167. <https://doi.org/10.1046/j.1365-2699.2000.00489.x>
- Whittaker, R. J., & Fernández-Palacios, J. M. (2007). *Island biogeography: Ecology, Evolution, and Conservation*. London, UK: Oxford University Press.
- Wood, H. M., González, V. L., Lloyd, M., Coddington, J., & Scharff, N. (2018). Next-generation museum genomics: Phylogenetic relationships among palpimanoid spiders using sequence capture techniques (Araneae: Palpimanoidea). *Molecular Phylogenetics and Evolution*, 127, 907–918. <https://doi.org/10.1016/j.ympev.2018.06.038>
- Wright, B. R., Farquharson, K. A., McLennan, E. A., Belov, K., Hogg, C. J., & Grueber, C. E. (2020). A demonstration of conservation genomics for threatened species management. *Molecular Ecology Resources*, 00, 1–16. <https://doi.org/10.1111/1755-0998.13211>
- Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J., & Chen, S. (2012). FastUniq: A fast de novo duplicates removal tool for paired short reads. *PLoS One*, 7(12), e52249. <https://doi.org/10.1371/journal.pone.0052249>
- Zhang, B., Kirov, S., & Snoddy, J. (2005). WebGestalt: An integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Research*, 33(suppl_2), W741–W748. <https://doi.org/10.1093/nar/gki475>
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), 3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>
- Zuccon, D., Pasquet, E., & Ericson, P. G. (2008). Phylogenetic relationships among Palearctic-Oriental starlings and mynas (genera *Sturnus* and *Acridotheres*: Sturnidae). *Zoologica Scripta*, 37(5), 469–481.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Baveja P, Garg KM, Chattopadhyay B, et al. Using historical genome-wide DNA to unravel the confused taxonomy in a songbird lineage that is extinct in the wild. *Evol Appl*. 2020;00:1–12. <https://doi.org/10.1111/eva.13149>