







Complex histories of gene flow and a mitochondrial capture event in a nonsister pair of birds

Michael J. Andersen¹  | Jenna M. McCullough¹  | Ethan F. Gyllenhaal¹  |
Xena M. Mapel^{1,2}  | Tri Haryoko³ | Knud A. Jønsson⁴  | Leo Joseph⁵ 

¹Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico, USA

²Animal Genomics, ETH Zürich, Lindau, Switzerland

³Museum Zoologicum Bogoriense, Research Centre for Biology, Indonesian Institute of Sciences (LIPI), Cibinong, Indonesia

⁴Natural History Museum of Denmark, University of Copenhagen, Copenhagen Ø, Denmark

⁵Australian National Wildlife Collection, CSIRO National Research Collections, Canberra, Australian Capital Territory, Australia

Correspondence

Michael J. Andersen, Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, NM, USA.
Email: mjandersen@unm.edu

Funding information

National Science Foundation, Grant/Award Number: DEB-1557051 and DGE-1650114; National Geographic Society, Grant/Award Number: 8853-10

Abstract

Hybridization, introgression, and reciprocal gene flow during speciation, specifically the generation of mitonuclear discordance, are increasingly observed as parts of the speciation process. Genomic approaches provide insight into where, when, and how adaptation operates during and after speciation and can measure historical and modern introgression. Whether adaptive or neutral in origin, hybridization can cause mitonuclear discordance by placing the mitochondrial genome of one species (or population) in the nuclear background of another species. The latter, introgressed species may eventually have its own mtDNA replaced or “captured” by other species across its entire geographical range. Intermediate stages in the capture process should be observable. Two nonsister species of Australasian monarch-flycatchers, Spectacled Monarch (*Symposiachrus trivirgatus*) mostly of Australia and Indonesia and Spot-winged Monarch (*S. guttula*) of New Guinea, present an opportunity to observe this process. We analysed thousands of single nucleotide polymorphisms (SNPs) derived from ultraconserved elements of all subspecies of both species. Mitochondrial DNA sequences of Australian populations of *S. trivirgatus* form two paraphyletic clades, one being sister to and presumably introgressed by *S. guttula* despite little nuclear signal of introgression. Population genetic analyses (e.g., tests for modern and historical gene flow and selection) support at least one historical gene flow event between *S. guttula* and Australian *S. trivirgatus*. We also uncovered introgression from the Maluku Islands subspecies of *S. trivirgatus* into an island population of *S. guttula*, resulting in apparent nuclear paraphyly. We find that neutral demographic processes, not adaptive introgression, are the most likely cause of these complex population histories. We suggest that a Pleistocene extinction of *S. guttula* from mainland Australia resulted from range expansion by *S. trivirgatus*.

KEYWORDS

introgression, Maluku Islands, mitonuclear discordance, Monarchidae, *Symposiachrus*, ultraconserved elements

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Two inter-related phenomena are increasingly recognized as elements of the speciation process: hybridization and mitonuclear discordance. Hybridization is now seen as a pervasive (Nosil, 2008), almost inevitable, element of speciation (reviews in Edwards et al., 2016; Joseph, 2018; Ottenburghs et al., 2017; Rheindt & Edwards, 2011; Smadja & Butlin, 2011; Taylor & Larson, 2019) and mitochondrial discordance is being increasingly detected in speciation studies (Bonnet et al., 2017; Good et al., 2008; Kearns et al., 2014; Pons et al., 2014; Toews & Brelsford, 2012). Consequently, genomic approaches to study both phenomena are now commonplace (e.g., McElroy et al., 2020; Payseur & Rieseberg, 2016). One outcome of hybridization during speciation is introgression (i.e., one-way gene flow through repeated backcrossing). Introgression, in turn, may generate mitonuclear discordance. Bonnet et al. (2017) introduced the term “massively discordant mitochondrial introgression (MDMI)” to describe extreme cases of mitonuclear discordance, whereby all or nearly all individuals of a species have the mitochondrial DNA (mtDNA) of another species but show very little nuclear introgression. This phenomenon is also termed mitochondrial capture. Hybridization and MDMI can be thought of as leading to one genome replacing another (Mallet, 2005).

Massively discordant mitochondrial introgression can arise through multiple mechanisms. First is through a suite of mitonuclear neutral or nonadaptive processes. These include sex-related asymmetry in interspecific gene flow or negative selection on multiple nuclear loci (discussed extensively in Bonnet et al., 2017), or a geographical range expansion following introgressive hybridization (Currat et al., 2008). Second are adaptive processes that center on positive selection for the introgressed mitochondrial genome. For example, enhanced cellular-level physiology can arise from introgression that places a novel mitochondrial genome in the nuclear genomic background of the introgressed species (Hill, 2019). For theoretical and empirical discussions of these mechanisms, see Borge et al. (2005), Currat et al. (2008), Petit and Excoffier (2009), Rheindt and Edwards (2011), Toews and Brelsford (2012), Toews et al. (2014); Sloan et al. (2017); Hill (2019), and Zhang et al., (2019). Bates-Muller-Dobzhansky incompatibilities represent yet a third mechanism that can lead to mitonuclear discordance. These incompatibilities arise from hybridization between two populations each having coadapted mitochondrial and nuclear genomes. This mechanism, however, is probably less likely to lead to mitochondrial capture (see Burton & Barreto, 2012; Burton et al., 2013). The introgressing and introgressed species need not be sister species and mitochondrial capture can occur long after speciation has been completed (Drovetski et al., 2015; Kearns et al., 2014; Shipham et al., 2016). In addition, two other factors could explain mitochondrial discordance: incorrect taxonomy and occasional gene flow, either historical or contemporary. Here, we explore these hypotheses and their difficulties in a case of mitochondrial discordance in birds.

An opportunity to study massively discordant mitochondrial introgression is offered by two nonsister species of Australasian

monarch-flycatchers. The Spectacled Monarch *Symposiachrus trivirgatus* (Aves: Passeriformes: Monarchidae) comprises eight geographically disjunct subspecies in three regions (Figure 1; Beehler & Pratt, 2016; del Hoyo & Collar, 2016; Schodde & Mason, 1999): (a) in the Lesser Sundas and the Maluku Islands of Indonesia, (b) in the Louisiade Archipelago, south-east of mainland New Guinea, and (c) in northeastern Australia, Torres Strait Islands, and southern New Guinea. Within Australia, three subspecies are currently recognized (see Beehler & Pratt, 2016; and Schodde & Mason, 1999 for details; Figure 1). *Symposiachrus t. albiventris* is sedentary and occurs in northernmost Cape York Peninsula and the Torres Strait Islands. Sedentary *S. t. melanorrhous* and migratory *S. t. gouldii* have largely allopatric breeding distributions in north-eastern Australia but overlap in the nonbreeding season. A nonsister species, the Spot-winged Monarch *S. guttula*, ranges across all of New Guinea and some of its eastern and western satellite islands. *Symposiachrus guttula* does not occur in Australia (Andersen et al., 2015); the Torres Strait narrowly separates the two species. They co-occur in a small part of southern New Guinea when migrant *S. trivirgatus* from Australia are present seasonally. They closely approach each other, however, on outlying islands east and west of New Guinea. Phenotypically, *S. trivirgatus* males from Indonesian populations resemble *S. guttula*.

Earlier studies that included up to six individuals of *S. trivirgatus* identified two nonsister haplogroups in that species' mtDNA (Andersen et al., 2015; Filardi & Smith, 2005). One haplogroup was sister to mtDNA of *S. guttula* and the other was phylogenetically separated from it by several other *Symposiachrus* species (Andersen et al., 2015). Prior to the present study, more extensive sampling of Australian *S. trivirgatus* individuals (reported below) expanded the numbers of *S. trivirgatus* individuals in these two haplogroups and affirmed their phylogenetic relationships and composition being solely of *S. trivirgatus*. The key point here is that neither haplogroup of *S. trivirgatus* mtDNA included any individuals of *S. guttula*. This is potentially consistent with the “allophyly” stage of incomplete lineage sorting (Omland et al., 2006); however, the nonsister species relationships of the two species coupled with one *S. trivirgatus* haplogroup being sister to one lineage of *S. guttula* rule out this explanation.

The aims of this study were to further investigate the apparent massively discordant mitochondrial introgression of *S. trivirgatus* with additional sampling of all *S. trivirgatus* subspecies and *S. guttula*. We tested four hypotheses that might explain apparent MDMI in Australian *S. trivirgatus*: cryptic speciation, gene flow, neutral capture, and adaptive capture. We applied phylogenetic, population genetics, and demographic modelling approaches of both nuclear DNA and mitochondrial genomes. Cryptic speciation is considered because Indonesian and Australian *S. trivirgatus* may not be closely related (del Hoyo & Collar, 2016; Eaton et al., 2016). It predicts that species-specific clades of nuclear DNA diversity will mirror the pattern of relationships among mtDNA haplotypes. We applied tests for signals of contemporary and historical gene flow across all of our samples and used single nucleotide polymorphisms (SNPs) derived from ultraconserved elements (UCEs) in the nuclear genome and

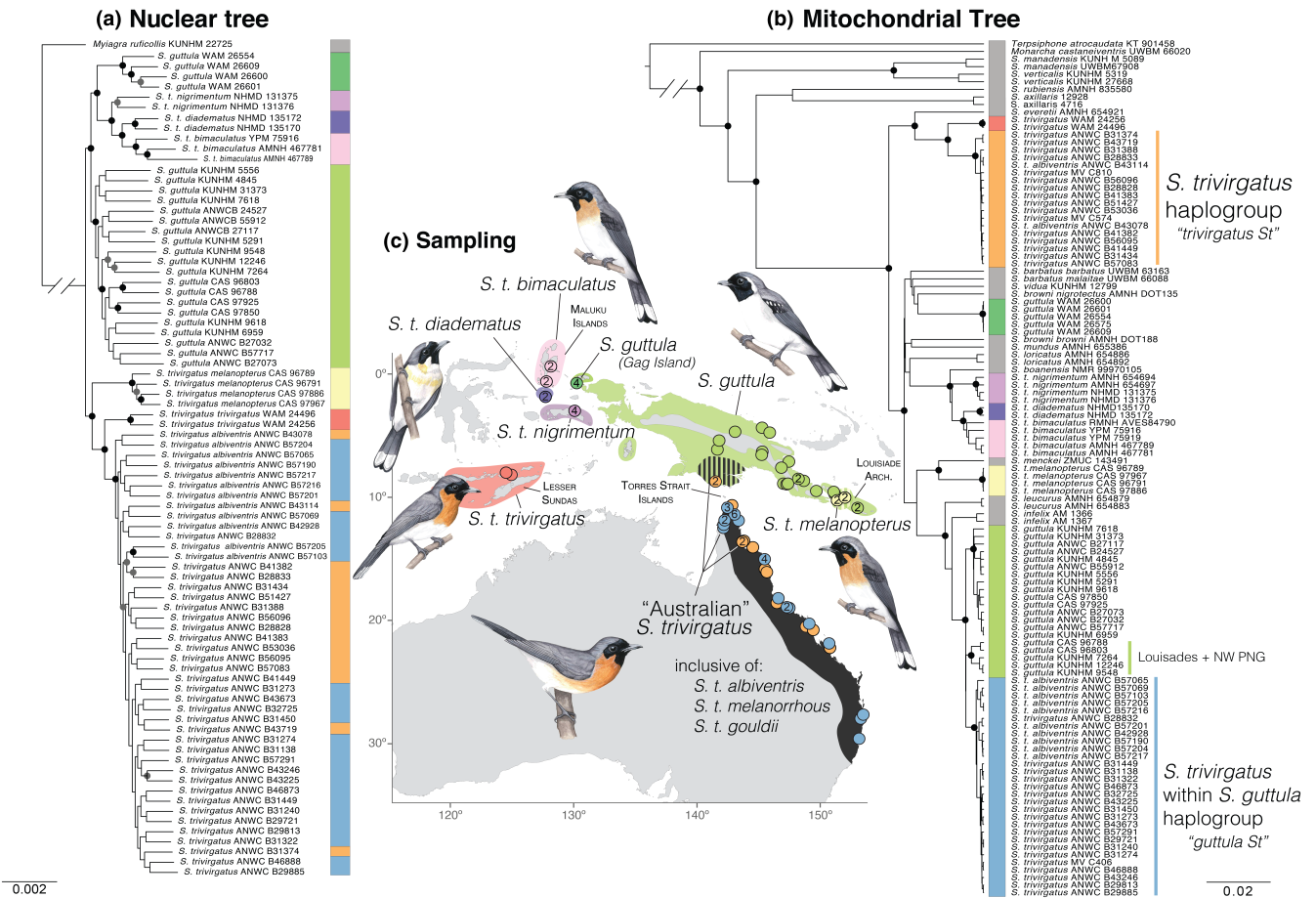


FIGURE 1 Mitochondrial discordance within the *Symposiachrus trivirgatus/guttula* complex. (a) Maximum likelihood phylogeny from a concatenated analysis of 4,269 ultraconserved elements in the 75%-complete matrix in RAxML. Black circles indicate bootstrap values of 95–100; grey circles represent 70–94; no circle means that BS values were ≤ 69 . (b) Maximum clade-credibility tree generated in BEAST, based on 13 mtDNA coding genes of all ingroup samples from this study plus ND2 sequences from additional *Symposiachrus* samples downloaded from GenBank. Black circles at major nodes represent those with posterior probabilities ≥ 0.95 . Coloured clade tips match sampling locations on the map for populations in the *S. trivirgatus/guttula* complex; grey tips denote samples outside of the complex. (c) Sampling map. Numbers on some sampling points refer to the number of individuals from that location; points without numbers represent one individual. Note the nonmonophyly of blue and orange Australian populations of *S. trivirgatus* in the mitogenome tree compared to the monophyly of these samples in the UCE tree. See Figure 3 for a more detailed illustration of the distributions of Australian *S. trivirgatus* subspecies. Original artwork of representative *Symposiachrus* taxa by Madison E. Mayfield (clockwise from top: *S. t. bimaculatus*, *S. guttula*, *S. t. melanopterus*, *S. t. albiventris*, *S. t. trivirgatus*, and *S. t. diadematus*)

whole mitogenomes of samples spanning the ranges of both species. Relationships among component populations and subspecies of *S. trivirgatus* are also reported. Ultimately, we hope to increase our understanding of speciation in the biologically important Indo-Pacific region.

2 | MATERIALS AND METHODS

2.1 | Sampling and DNA extraction

Our data set comprised 81 ingroup taxa (57 *S. trivirgatus sensu lato* and 24 *S. guttula*; Table S1). We sampled broadly across the distributions of *S. guttula* and *S. trivirgatus* (Table S1, Figure 1), including all

eight subspecies of *S. trivirgatus* (Clements et al., 2019; del Hoyo & Collar, 2016; Dickinson & Christidis, 2014). All mainland New Guinea samples were from Papua New Guinea because we lacked tissue samples from the west, but for simplicity, we refer to this population as “Mainland New Guinea.” We rooted phylogenetic analyses to *Myiagra ruficollis* (sequenced for this study), but used a more closely related outgroup, *Symposiachrus verticalis*—from Moyle et al. (2016)—for analyses of gene flow. Most samples were derived from frozen or ethanol-preserved, specimen-vouchered tissues; however, three samples of Indonesian island populations of *S. trivirgatus* were sourced from specimen toepad clippings (Table S1). For tissue-derived libraries, we extracted genomic DNA following standard protocols with a Qiagen DNeasy Blood and Tissue Kit and quantified extracts with a Qubit 3.0 Fluorometer (ThermoFisher Scientific).

2.2 | Target capture of ultraconserved elements

We performed target capture of UCEs (Faircloth et al., 2012) at the University of New Mexico. Prior to library preparation, we visualized DNA extraction quality via gel electrophoresis to determine shearing protocol. We sheared 250 ng of genomic DNA to 500 base pair (bp) fragments with a Covaris M220 focused-ultrasonicator (50 W peak incident power, 10% duty factor, and 200 cycles per burst for 55–75 s, depending on extraction quality). We prepared Illumina libraries using 1/2 the volume per reaction of the Kapa Biosystems Hyper Prep kit and the iTruStub dual-indexing system (baddna.org), following established protocols (Faircloth et al., 2012). We enriched pooled libraries (500 ng of 8 equimolar samples per pool) for 24 h at 65°C using the Arbor Biosciences (formerly MYcroarray) myBaits tetrapod 5 K v1 probe kit, targeting 5060 UCE loci. UCE-enriched pooled libraries were quality checked with a bioanalyser (Agilent Bioanalyzer 2100), qPCRred, and combined at equimolar ratios prior to sequencing on a single lane of an Illumina HiSeq 3000 with a PE-150 flow cell at the Oklahoma Medical Research Foundation (OMRF). We handled the six toepad-derived samples differently than the tissue-derived samples. This is because DNA from museum study skins tend to be fragmented (McCormack et al., 2016; Mundy et al., 1997; Tsai et al., 2019). For toepads, we followed protocols we have used previously (Andersen et al., 2019; McCullough et al., 2019). Toepad-derived libraries were pooled with samples from other projects and sequenced at OMRF using the same sequencing parameters as the tissue-derived libraries.

2.3 | UCE data processing and phylogenetic analysis

We processed UCE data, built contigs, trimmed, and aligned loci with the Python bioinformatics package “PHYLUCE” v1.5.0 (Faircloth, 2016; see <https://github.com/faircloth-lab/phyluce>). First, we demultiplexed raw reads from BCL files using BCL2FASTQ2 v2.17.1.14 (Illumina Inc.) then removed adaptor sequences and low-quality bases with ILLUMIPROCESSOR v1 (Bolger et al., 2014; Faircloth, 2013; Lohse et al., 2012). We used TRINITY v2.0.6 (Grabherr et al., 2011) to assemble contigs from cleaned reads then extracted contigs for each taxon that matched UCE loci. We aligned each locus specific data sets with MAFFT (Kato & Standley, 2013) and trimmed them with GBLOCKS v0.91 (Castresana, 2000). We used default parameters for gblocks except for the minimum number of sequences allowed for flank positions, b2, which we set to 65%. These steps produced a final 75%-complete UCE data set in which all locus-specific alignments had data from at least 61 of the 81 individuals.

Prior to phylogenomic analysis of UCE data, we selected the optimal partitioning schemes and substitution models with PARTITIONFINDER2 v2.1.1 using the rcluster algorithm (Lanfear et al., 2014, 2017). Then, we analysed the partitioned, concatenated matrix with RAXML v8.2.4 (Stamatakis, 2014) using the CIPRES Science Gateway (Miller et al., 2010). We evaluated support of the ML

topology with 900 rapid bootstrap replicates. To take into account gene tree heterogeneity in our UCE data set, we inferred species trees using SVDQuartets (Chifman & Kubatko, 2014) invoked in PAUP v4.0a (build 163; Swofford, 2003). For this analysis, we considered the nine clades that we inferred from mitochondrial data (i.e., the nine coloured clades in Figure 1b; see mitogenome section below), we evaluated 139,492 quartets, and assessed support with 1,000 bootstraps.

2.4 | Mitogenome data processing and analysis

We recovered off-target sequencing reads of the mitochondrion and generated mitogenomes by mapping the illumiprocessor cleaned reads to a mitogenome reference of *Terpsiphone atrocaudata* (GenBank: KT901458) in Geneious R9.1 (Biomatters Ltd). We checked these assemblies against known mitochondrial sequences that were generated with standard PCR and Sanger sequencing (Andersen et al., 2015; Filardi & Moyle, 2005; Filardi & Smith, 2005) to guard against the presence of NUMTs. We also checked these alignments by eye for internal stop codons and failing to find any, we continued with our analyses. We inferred an unrooted haplotype network of the ND2 gene using the unrooted TCS (Clement et al., 2000) algorithm with default settings in PopART (Leigh & Bryant, 2015). We analysed the 13 mitochondrial protein-coding genes using three partitions, one for each codon position. To expand our mtDNA alignment, we downloaded from GenBank 28 additional ND2 sequences of *Symposiachrus* spp., one *Monarcha castaniaventris*, and one *T. atrocaudata*. This additional sampling included all samples from Andersen et al. (2015), plus *S. menckei* (GenBank: KP726921) and *S. trivirgatus bimaculatus* (GenBank: KP726923). We used this expanded mitogenome matrix of 113 samples to estimate a phylogeny in BEAST v2.5.2 (Bouckaert et al., 2014). We assigned the HKY+I + G model to each partition and used a strict clock model and a birth-death tree model. We ran four independent MCMC chains of 10 million generations each, sampling every 2,000 generations. We assessed convergence of parameter estimates and determined that ESS values were all >200 by visualizing trace files in TRACER v1.7.1 (Rambaut et al., 2018). We discarded the first quartile of MCMC generations before summarizing the remaining trees in a maximum clade-credibility tree using LOGCOMBINER v2.5.2 and TREEANNOTATOR v2.5.2 (Bouckaert et al., 2014).

2.5 | SNP calling from UCEs

We called SNPs from UCEs to determine individual population assignment and explore gene flow between *S. trivirgatus* and *S. guttula*. We largely followed the seqcap_pop pipeline (Harvey et al., 2016). We de novo assembled our cleaned reads from 15 individuals into a consensus reference assembly using VelvetOptimiser (Zerbino & Birney, 2008). We selected these specific individuals for their geographic and taxonomic breadth within our data set and because they represented samples with the highest number of cleaned reads (see

Table S1). To avoid including markers of differing ploidy, we used BLASTN v2.2.31 (Zhang et al., 2000) to identify Z-linked loci from the Zebra Finch (*Taeniopygia guttata*) reference genome from ENSEMBL 95 (Zerbino et al., 2018) and removed them from our reference assembly using a custom python script. We then mapped contigs to targeted UCE loci, and mapped reads to those contigs using BWA v0.7.7 (Li & Durbin, 2009). To convert SAM to BAM files and sort, clean, add read groups to, call PCR duplicates in, and merge BAM files across individuals, we used PICARD v2.18.17 (<http://broadinstitute.github.io/picard/>). We also used this program to create a dictionary with the reference assembly and SAMTOOLS v1.5 (Li et al., 2009) to index the BAM files and reference. With GATK v3.8.1 (McKenna et al., 2010), we called indels using RealignerTargetCreator and realigned them using IndelRealigner. Subsequently, we called SNPs with UnifiedGenotyper, annotated variants with VariantAnnotator, removed variants with a quality score of less than 30 or a quality by depth <5 with VariantFiltration, and finally phased the data with ReadBackedPhasing to produce a VCF file for downstream analyses. With VCFTOOLS v0.1.15 (Danecek et al., 2011), we restricted the data set to biallelic SNPs with no missing data and used custom python scripts to remove possible paralogs based on a cutoff of ≥ 0.8 of all individuals heterozygous for a given SNP. To account for physical linkage, we restricted the final VCF file to one random SNP per locus. We generated six SNP data sets (Table 1): (a) full taxon sampling ($n = 82$), (b) all taxa but downsampled to have a maximum sample size of four per population ($n = 25$), (c) as the former but including two *S. trivirgatus* subpopulation haplogroups ($n = 29$), (d) for the *guttula* haplogroup of Australian *S. trivirgatus* and southern New Guinean *S. guttula* ($n = 24$), (e) for only Australian *S. trivirgatus* ($n = 44$), and (f) for only New Guinean *S. guttula* ($n = 20$).

2.6 | Population structure and assignment

We used two approaches to determine basic population structure. First, we used sNMF (Frichot et al., 2014), implemented in the

R package LEA 2.2.0 (Frichot & François, 2015), to infer the best-fit number of populations (k) and construct assignment plots with admixture coefficients. We checked for consistency of k under varying values of the regularization parameter alpha (1, 10, and 100), in which higher values penalize admixture more than lower values. Once we chose an optimal value of alpha based on a preliminary investigation of which value minimized the cross-entropy criterion, we performed 100 replicates for each value of k tested. Due to the unstable nature of estimating k (Kalinowski, 2011) and given uneven sampling, we include and interpret values of k greater than the one inferred by the cross-entropy criterion. Second, we used the gPca method within ADEGENET (Jombart, 2008) to conduct a principal component analysis on the data sets including only New Guinean *S. guttula* and Australian *S. trivirgatus*, respectively.

2.7 | Tests of gene flow

We generated two species trees with the SNP data. First, in SNAPP v1.4.2 (Bryant et al., 2012), we included up to four individuals per population, except in cases where less than four individuals were available (i.e., $n = 3$ of *S. t. bimaculatus*, and $n = 2$ for each of *S. t. diadematus*, *S. t. nigrimentum*, and *S. t. trivirgatus*). We ran two million generations and discarded the first quartile as burnin; we calculated mutation parameters automatically rather than inferring them from population size estimates. This tree enabled an alternate topology to better assess the validity of ABBA/BABA tests (described in detail below). Second, we used TREEMIX v1.13 (Pickrell & Pritchard, 2012), which uses the covariance of populations' allele frequencies to estimate an ML species tree and then identifies species that are more closely related than can be explained by the relationships in the species tree. These species pairs are thus candidates for migration events (i.e., migration edges in the phylogeny) and we added migration edges until we observed an increase in likelihood of <1. First, we performed an unsupervised run of TREEMIX, where no prior migration edges were tested. To observe the effects of migration events

TABLE 1 Numbers of samples included and SNPs recovered in each dataset used for population-level analyses. Note that although 4,051 total SNPs were used to create input for ABBA/BABA analysis, only 2,756 were informative

Dataset	N samples	All SNPs	Analyses using all SNPs	Excluding singletons	Analyses with SNPs & no singletons
Full sampling (with <i>S. verticalis</i> outgroup)	82	4,051	ABBA/BABA (figure 5 Table S2); TREEMIX (Figure S1)	—	—
All major populations ($n \leq 4$)	25	1,855	SNAPP (Figure 2d)	—	—
All major populations and a second Australian <i>S. trivirgatus</i> group ($n \leq 4$)	29	—	—	1,140	sNMF (Figure 2b)
One subpopulation each of <i>S. guttula</i> (southern New Guinea) and <i>S. trivirgatus</i> (<i>S. guttula</i> haplogroup sans <i>S. t. albiventris</i>)	24	2,793	DADI (Table 2, S3)	1,853	sNMF (Figure 3)
All samples of Australian <i>S. trivirgatus</i>	44	4,217	PCA (Figure 4, S2)	1,069	sNMF (Figure 4b)
All samples of New Guinea <i>S. guttula</i>	20	4,216	PCA (Figure S3c)	830	sNMF (Figure S3c)

supported by ABBA/BABA tests and mitonuclear discordance, we also performed supervised runs of TREEMIX by adding a priori migration edges using the `-cor_mig` and `-climb` flags. Specifically, we manually tested four edges: *S. t. diadematus* to Gag Island *S. guttula*; *S. t. melanopterus* to *S. t. diadematus*; southern New Guinea *S. guttula* to Australian *S. trivirgatus*; and *S. t. melanopterus* to *S. guttula* from the Louisiades. This allowed for testing the hypothesis of introgression to explain phylogenetic uncertainty between analyses.

To test for evidence of modern or historic gene flow between *S. guttula* and Australian *S. trivirgatus*, we used ABBA/BABA tests (Durand et al., 2011). Specifically, we tested a priori migration edges of interest, as well as instances of nonsister introgression identified by our TREEMIX analyses. The ABBA/BABA test generates a *D* statistic with a ladderized tree with four tips (tips labelled P1–4), where positive values imply that there is gene flow between P2 and P3, and negative values imply gene flow between P1 and P3. The magnitude of *D* is proportional to the amount of gene flow between P3 and the other population, but this observation can be confounded by gene flow between P1 and P2 (Durand et al., 2011). We focused on four major migration edges: (a) between New Guinean *S. guttula* and Australian *S. trivirgatus*, (b) between Louisiade *S. guttula* and *S. t. melanopterus*, (c) between Gag Island *S. guttula* and *S. t. diadematus*, and (d) between *S. t. melanopterus* and *S. t. diadematus*. In addition to these focused tests, we tested additional ABBA/BABA patterns to demonstrate the local nature of this gene flow in cases where taxa had multiple subspecies or populations (see Table S2 for a full list). We generated input files using scripts by Simon Martin (https://github.com/simonmartin/genomics_general) and ran the ABBA/BABA tests in R (R Development Core Team, 2013). We performed 1,000 bootstrap replicates to generate a distribution of values for both statistics, and we calculated *p*-values with the proportion of replicates less than or equal to 0.

2.8 | Demographic analyses

We performed demographic analyses and model testing with DADI v2.0.3 (Gutenkunst et al., 2009) to test different models of gene flow and admixture between Australian *S. trivirgatus* and New Guinean *S. guttula* (visualized in Figure S1). Due to notable population structure within both taxa, we chose the largest and most relevant populations from each: *S. guttula* from southern New Guinea and *S. trivirgatus* (excluding *S. t. albiventris*) within the *S. guttula* mitochondrial haplogroup. We generated input files for this analysis using EASY-SFS (Oswald et al., 2017; <https://github.com/isaacovercast/easySFS>) to generate and downsample (for even sample size) a site frequency spectrum (SFS) for DADI. Once the SFS was generated, we used it to test ten different models (visualized in Figure S1). The null model did not allow migration or admixture between populations after divergence. The next three models allowed for continuous, unidirectional (in either direction) or bidirectional migration. The final six models included a discrete unidirectional or bidirectional admixture

event, as well as either including or not including continuous gene flow between populations. We performed 50 optimization runs with perturbed parameters, whose bounds were universally set across models to ensure that no model was limited by its parameters. We chose the highest likelihood run from a given model for comparison and compared these models using an Akaike information criterion (AIC) weight framework using likelihoods calculated in DADI (Rougeux et al., 2019). Following Winker et al. (2018), we blasted our Velvet Optimiser reference assembly FASTA (see SNP calling from UCEs section above) against genomes of *Melopsittacus undulatus* and *Acanthisitta chloris*, and used the date of their most recent common ancestor (60.5 and 53 Ma, respectively; Claramunt & Cracraft, 2015) to estimate substitution rates for the sequences from which SNPs were called to interpret demographic parameters. The final effective sequence length and mutation rate were 834,259 bp (the length of the entire reference FASTA) and 5.6×10^{-10} mutations per site per generation, respectively. A generation time of two years was used, as in *Ficedula* (Smeds et al. 2016), due to the uncertainty in actual generation time of our study system. Note that our primary parameter of interest, admixture proportion, is not affected by these parameters.

2.9 | Detecting positive selection in the mitogenome

We tested for positive selection in each of the 13 mitochondrial genes with several methods. Using the McDonald-Kreitman test (MK test; McDonald & Kreitman, 1991), we derived *p*-values from a Chi-squared analysis of a contingency table of fixed and polymorphic mutations. These mutations can either be synonymous or nonsynonymous between the focal group and the outgroup, *S. verticalis*. We examined the three mitochondrial clades: *S. trivirgatus* with the *S. guttula* mtDNA, *S. trivirgatus* with the “true” *trivirgatus* mtDNA, and *S. guttula* from southern New Guinea. For each population, we computed the Chi-squared value and Neutrality Index (NI; Rand & Kann, 1996) with the standard MK test web portal (<http://mkt.uab.es/mkt/>; Egea et al., 2008). Taken together, these two statistics test whether any mitochondrial group showed evidence of gene-level positive selection that significantly deviated from neutrality. Furthermore, the proportion of fixed amino acid substitutions was calculated to see if the genes in one group were enriched for substitutions relative to others. In addition, we used CODEML (Yang, 2007), as implemented in EASYCODEML (Gao et al., 2019), to test for natural selection in each of the 13 individual mitochondrial coding genes and accompanying gene trees with the site-specific model. This model allows the ratio of nonsynonymous to synonymous substitution rates (ω) to vary among sites (but not branches) to test for site-specific evolution in the context of a phylogeny. For each gene region, we evaluated which of six codon substitution models (M0, M1a, M2a, M3, M7, M8) was favored using likelihood ratio tests.

To address the adaptive capture hypothesis, we also performed three similar statistical tests to detect a selective sweep in the mtDNA. These three tests are Tajima's D (Tajima, 1989) and Fu and Li's F^* and D^* (Fu & Li, 1993). All three tests were performed in DNASP 6 (Rozas et al., 2017). These tests should return a significantly negative value if a historical bottleneck or selective sweep has occurred. In order to correct for demography, we performed this test on all three mitochondrial clades tested in the prior section. If adaptive capture has occurred, we would expect a significantly negative value of captured *S. guttula* mtDNA that is also less than that of native *S. trivirgatus* and *S. guttula* mtDNA.

3 | RESULTS

3.1 | Sampling

We recovered an average of 4,186 UCEs (median 4,242, range: 3,685–4,441) per sample. The 75% complete matrix (i.e., 75% of taxa were present per UCE alignment) comprised 4,269 loci and 3,970,857 bp, of which 101,808 bp were variable and 54,002 bp were parsimony-informative characters. We recovered full mitogenomes from 79 individuals totalling 16,983 bp, but we reduced our final mitogenome data matrix for phylogenetic analysis to only the 13 coding genes (11,404 total bp). We removed two toepad-derived samples (*S. t. nigriumentum* AMNH654697 and AMNH654694) from all downstream UCE- and SNP-based analyses due to excessively long branch lengths and incorrect topologies that were apparent in preliminary RAxML analyses; however, we kept these samples in the ND2 haplotype network because all recovered mtDNA was suitable for analysis. The six SNP data sets that we produced for this study, including numbers of SNPs and their specific uses, are outlined in Table 1. Note that we produced data sets without singletons for sNMF analyses because their inclusion has been shown to decrease the effectiveness of model-based analyses of population structure (Linck & Battey, 2019).

3.2 | Phylogenetics

Mitochondrial and nuclear data showed conflicting relationships across the *Symposiachrus trivirgatus* and *S. guttula* species complexes. The mitochondrial tree (Figure 1) and haplotype network (Figure S2) indicated that there are two haplogroups within Australian *S. trivirgatus*: those with the *S. guttula* mitochondrion and those we infer to have the “true” *S. trivirgatus* mitochondrion (hereafter referred to as the “*guttula St*” and “*trivirgatus St*” haplogroups within *S. trivirgatus*, respectively; Figure 1b,c). Median uncorrected p-distances of mitochondrial ND2 and nuclear UCEs are reported in Table 2. ND2 genetic distances ranged from 0.96% to 5.8% and UCE genetic distances ranged from 0.18% to 0.34%. As expected, *guttula St* and *trivirgatus St* were 5.0% divergent in ND2, but only 0.18% divergent in the nuclear UCE data set. Overall, the mitogenome tree was largely similar to that of Andersen et al.

(2015) in that *S. trivirgatus* was paraphyletic and comprised five clades (colours as in Figure 1b):

1. *S. t. trivirgatus* of the Lesser Sundas (red), which was sister to
2. Australian *trivirgatus St* haplogroup clade (orange).
3. Maluku *S. t. nigriumentum*, *S. t. diadematus*, and *S. t. bimaculatus* (light purple, dark purple, and pink, respectively).
4. *S. t. melanopterus*, endemic to offshore islands of far south-eastern Papua New Guinea (yellow), distinct from all other *S. trivirgatus* and sister to *S. menckei*, an endemic species from Mussau Island in the Bismarck Archipelago.
5. Remaining Australian *S. trivirgatus* samples (*guttula St* haplogroup; blue) as sister to *S. guttula* from New Guinea.

Within *S. guttula*, we found some phylogenetic structure, including a clade of two samples from the Louisiade Archipelago and three samples from East Sepik and Madang provinces in northwest Papua New Guinea, to the exclusion of all other *S. guttula* samples (green clades; Figure 1b); however, recall that we did not sample from western mainland New Guinea. Finally, *S. guttula* from Gag Island—offshore from New Guinea's Vogelkop Peninsula (hereafter Gag *guttula*)—formed a clade that was not sister to other *S. guttula*. The ND2 haplotype network (Figure S2) mirrored the relationships in the mitogenome phylogeny.

Patterns of paraphyly in the mitogenome tree largely disappeared in concatenated and species trees inferred from nuclear data. Our RAxML tree, based on a partitioned, concatenated analysis of UCEs (Figure 1a) indicates that all *S. trivirgatus* (barring the Maluku subspecies) formed a clade. Samples belonging to the two mtDNA haplogroups of Australian *S. trivirgatus* (paraphyletic in the mtDNA tree) were indistinguishable in the UCE tree. Nominate *S. t. trivirgatus* was sister to all Australian individuals and *S. t. melanopterus* was the first branch in this clade, sister to the rest. We also recovered novel relationships in the nuclear tree with respect to Gag *guttula*, which was strongly supported as sister to the Maluku *S. trivirgatus* samples, rather than as an independent lineage subtended from a polytomy within the tree. And, unlike the mitogenome tree, all four far south-eastern Papua New Guinea *S. guttula* samples formed a clade to the exclusion of those from northwest Papua New Guinea.

Species trees, estimated with either full loci or SNPs, show similar results to the concatenated analysis: (a) both *S. trivirgatus* mitochondrial haplogroups are monophyletic, and (b) the Maluku subspecies are more closely related to *S. guttula* than nominate and Australian *S. trivirgatus* (Figure 2, S3). Yet, the only significant differences between species tree analyses was the equivocal topology regarding the sister lineage of the closest relative of the clade of Maluku *S. trivirgatus* subspecies. The SVDquartets analysis indicated with moderate support (72%–73% bootstrap values) that the clade of Maluku subspecies was closest to Gag *guttula*, which was in turn sister to the rest of *S. guttula* (Figure 2) and thus supported the RAxML topology. The most likely SNAPP species tree supported all of *S. guttula* (including Gag Island samples) as the sister clade to the Maluku subspecies (88.8% of trees in the posterior distribution);

Mitochondrial group	1	2	3	4	5	6
1 <i>S. t. trivirgatus</i>	–	0.25%	0.29%	0.34%	0.26%	0.33%
2 <i>trivirgatus</i> St haplogroup	2.40%	–	0.28%	0.32%	0.18%	0.31%
3 <i>S. t. bimaculatus</i> , <i>S. t. diadematus</i> , <i>S. t. nigrimentum</i>	5.20%	5.80%	–	0.30%	0.28%	0.26%
4 <i>S. t. melanopterus</i>	4.70%	4.80%	4.30%	–	0.32%	0.34%
5 <i>guttula</i> St haplogroup	4.50%	5.00%	4.00%	3.70%	–	0.31%
6 <i>S. guttula</i>	4.37%	4.71%	4.04%	3.55%	0.96%	–

TABLE 2 Median uncorrected p-distances between six groups of *Symposiachrus* monarch-flycatchers. Numbers above the diagonal are distances in the nuclear UCE data; numbers below the diagonal are based on the mitochondrial ND2 gene

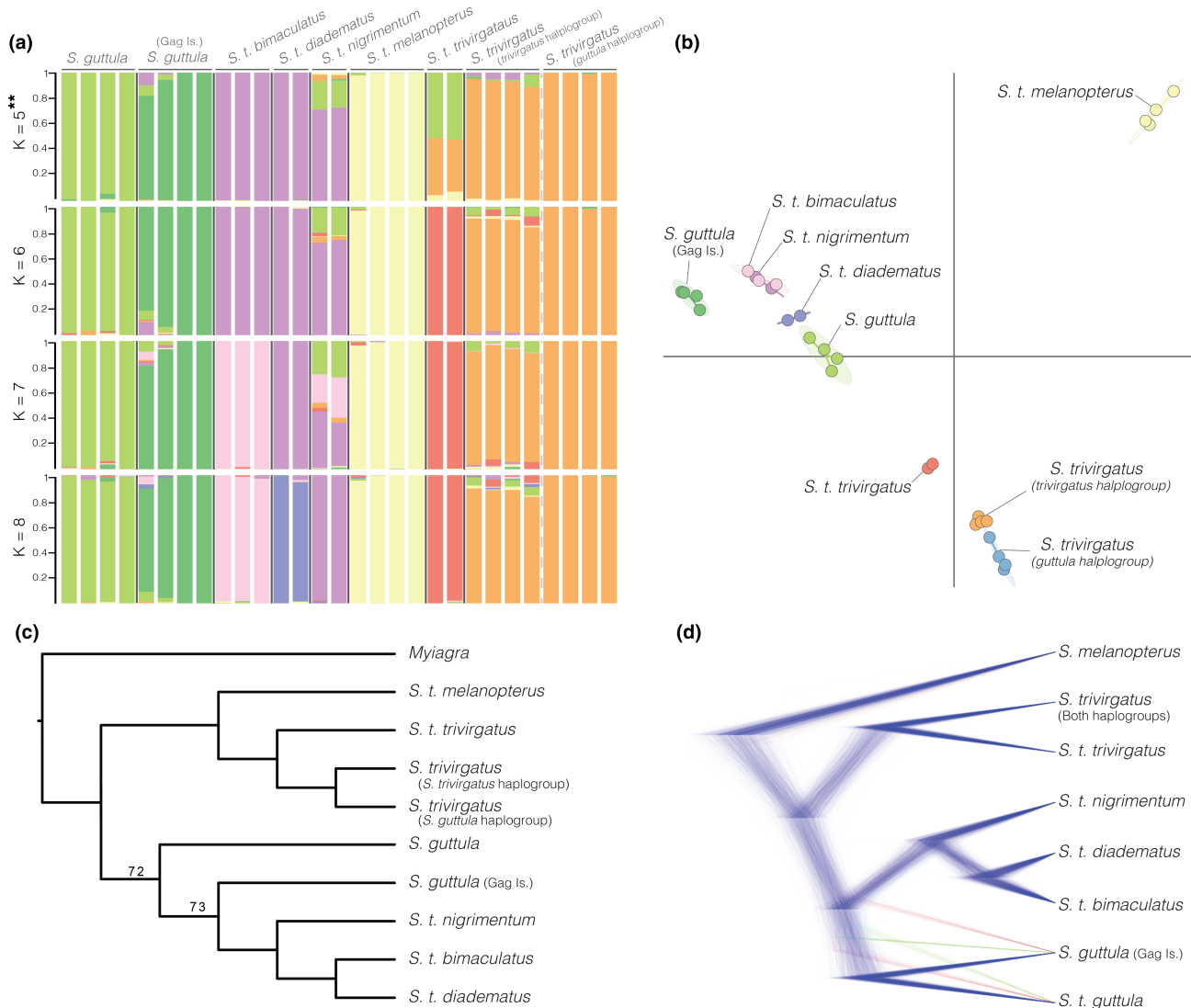


FIGURE 2 Genomic exploration of population structure within the *Symposiachrus trivirgatus/guttula* complex. (a) Population assignment plots generated in sNMF for a range of k values from 5–8. (b) Principal components analysis of SNP data, showing the first two PC components (contribution of PC1 and two axes are 17.6% and 15.8%, respectively. A full table of the first four PC axes is reported in Table S3). (c) Species tree based on 4,269 UCE loci and estimated in SVDQuartets, all nodes are fully supported unless noted otherwise. (d) Species tree based on 1,855 SNPs and estimated in SNAPP

however, two alternative topologies were represented in the posterior distribution—either Gag *guttula* or mainland *S. guttula* were the sister lineage to Maluku *S. trivirgatus* (5.6% or 5.3% of trees, respectively). Note that *S. t. melanopterus* was inferred as the outgroup

lineage in SNAPP; however, this result may be unreliable due to our inability to include an actual outgroup given insufficient outgroup sample size. Finally, our TREE-MIX results mostly supported the sister relationship of Gag *guttula* and Maluku *S. trivirgatus* populations

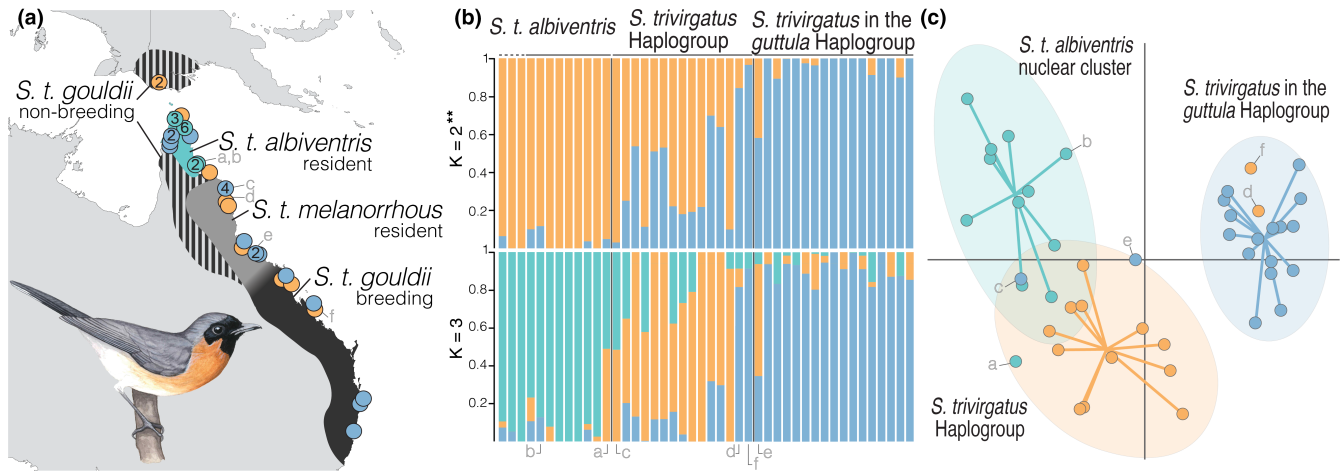


FIGURE 3 Population structure of Australian *Symposiachrus trivirgatus*. (a) Sampling map and ranges of the three *S. trivirgatus* subspecies in Australia, including two immature birds (possible migrants from Australia) captured in Papua New Guinea. Colours represent the three genetic clusters found in our analyses: teal = *S. t. albiventris* (all but two of which have the *S. guttula* mtDNA, exceptions are noted below); orange = individuals with “true” *S. trivirgatus* mtDNA; blue = nonalbinventris individuals with *S. guttula* mtDNA. The latter two nuclear clusters are not delineated by subspecies, as *gouldii* and *melanorrhous* are morphologically indistinguishable, and they separate genetically in all analyses of population structure. (b) Population structure of *S. trivirgatus* based on sNMF analyses, including results from $k = 2$ – 3 . When $k = 2$, the *guttula* haplogroup is determined to be distinct (i.e., blue individuals), but when $k = 3$, *S. t. albiventris* (teal) and the orange group separated. Subspecific identity of the three left-most samples indicated by a dashed line was unclear based on phenotype of specimens; however, they were assigned genetically to the *S. t. albiventris* population (from left to right: ANWC B57069, ANWC B57291, and ANWC B57103). (c) Principal components of Australian *S. trivirgatus* based on the SNP dataset. Six points are noted in this figure: a (ANWC B43114) and b (ANWC B43078) are identified as the only sampled *albiventris* individuals with the “correct”; *trivirgatus* mitochondrion; points a, c (ANWC B28832), d (ANWC B31374), e (ANWC B31273), and f (ANWC B43719) are hybrid individuals. See Figure S5 for a similar PCA with these outliers included in their a priori groups. Original artwork of *S. t. albiventris* by Madison E. Mayfield

when a migration edge was not manually added (see Tests for gene flow section below; Figure S3).

3.3 | Population structure and assignment

Our sNMF analyses revealed population structure across this study system and that all Australian *S. trivirgatus* are distinct in nuclear DNA from *S. guttula*, regardless of mitochondrial haplogroup (Figure 2). Because k optimization is often conservative and can be unreliable (Kalinowski, 2011), we included analyses from a range of k (5–8), which parsed ever more geographically circumscribed populations as k increased. For example, at $k = 5$, all populations except the Maluku, Lesser Sundas, and Australian *S. trivirgatus* were distinct; whereas at $k = 8$, all but two populations were assigned as distinct (Figure 2b). This analysis did not find clear population structure within Australian *S. trivirgatus*. However, the *S. trivirgatus* with the “true” *trivirgatus* mitochondrial DNA showed a mixed assignment, which suggested that more detailed analysis within Australian *S. trivirgatus* was needed.

When testing for recent admixture between *S. guttula* and *guttula St*, we inferred minor levels of admixture (all proportions ≤ 0.1). This was probably due to noise in the data, introduced by population structure within *S. trivirgatus*, not higher order backcrosses (Figure S4). Intraspecific population structure analyses between *S.*

t. albiventris, *trivirgatus St*, and *guttula St* (Figure 3) confirms population structure in Australian birds. Our results indicated three genetic clusters with few admixed individuals: one with the *guttula St* haplogroup, one with the *trivirgatus St* haplogroup, and a separate cluster of *S. t. albiventris* from the Cape York Peninsula. This latter subspecies, which primarily groups together in both nuclear and mitochondrial trees and are mostly assigned to the *guttula St* haplogroup (Figure 1), is a genetically distinct population based on both sNMF and genetic PCAs (Table S3, Figure 3, S5). The two other genetic groups are not obviously delineated by Australian *S. trivirgatus* subspecies, *S. t. gouldii* and *S. t. melanorrhous*, but rather by mitochondrial haplogroup (Fisher's exact test, $p < .001$). The *trivirgatus St* haplogroup cluster was more similar to *S. t. albiventris* (overlapping 95% inertia ellipses; Figure 3c), for which all but two *S. t. albiventris* were in the *guttula St* haplogroup clade (Figure 1b); the most distinctive grouping was the *guttula St* haplogroup nuclear cluster (Figure 3c). Despite their distinctiveness, there is still ongoing gene flow between all three genetic clusters, as shown by three putative F1 or F2 hybrids (see labeled individuals in Figure 3), which indicates that these populations are not reproductively isolated. Both the *S. t. albiventris* and *guttula St* haplogroup clusters differed significantly from an expectation of random mtDNA haplogroup distributions (Chi-squared test, $p < .01$), but did not differ from each other (Fisher's exact test, $p = 1$). Although the 95% inertia ellipse of the *trivirgatus St* haplogroup cluster did not separate entirely from the *S. t. albiventris*

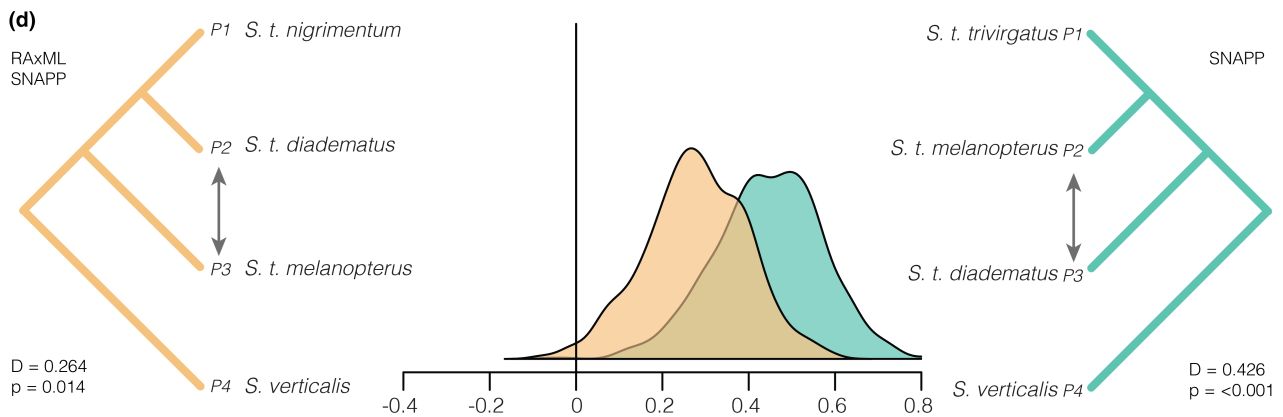
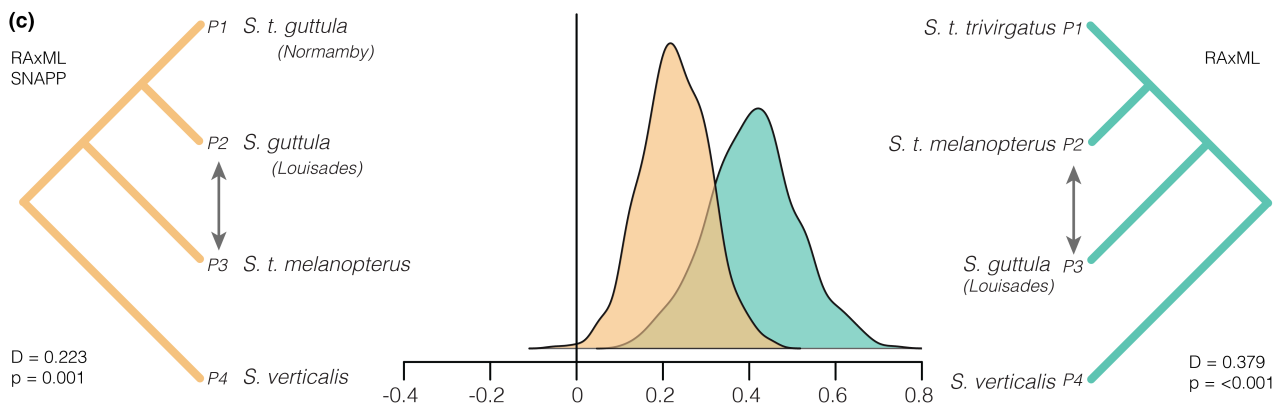
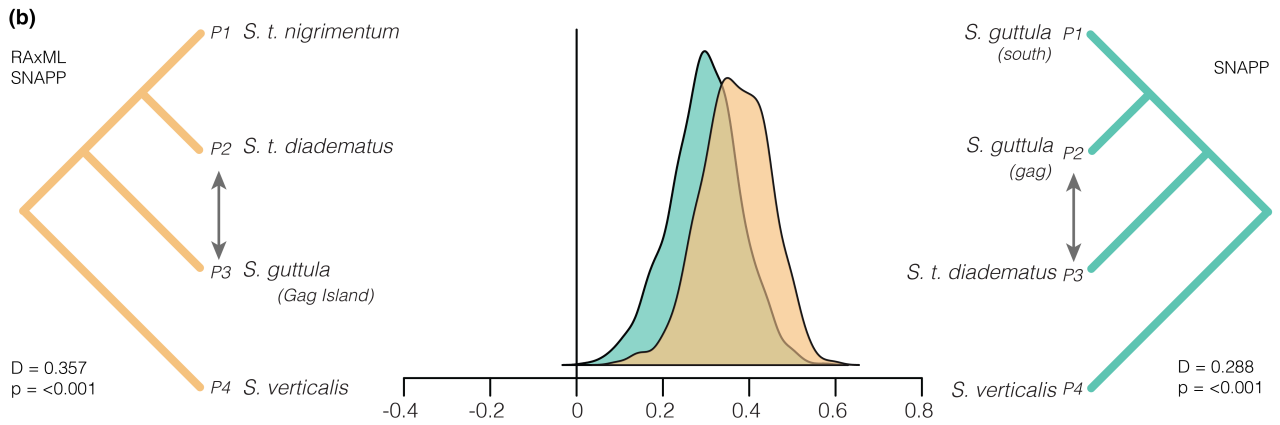
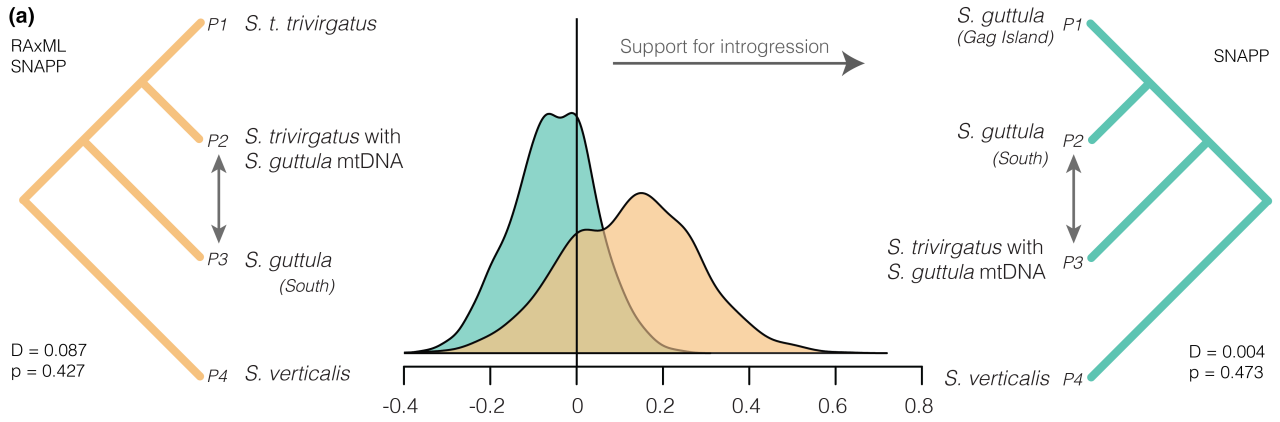


FIGURE 4 ABBA/BABA tests shown by topologies and visualized using density plots of values of D obtained by bootstrapping. A value greater than zero is indicative of either gene flow or an incorrect topology tested. Topologies that are supported by our RAxML and/or SNAPP trees are denoted above the trees and p and D values are listed below. Note that certain tests, denoted by the green topologies (right) are supported by one method of phylogenetic analysis, but the orange topology (left) was supported by all phylogenetic methods used in this study. Therefore, significant results from those topologies are strong evidence for gene flow, as incorrect topologies can result in falsely rejecting the null hypothesis of no gene flow. (a) The test for gene flow between Australian *S. trivirgatus* (the *guttula* St haplogroup) and New Guinea *S. guttula* (southern population), demonstrating no statistically significant nuclear introgression despite mitochondrial capture. (b) The test for gene flow between *S. t. diadematus* and Gag Island *S. guttula*, showing significance for both reciprocal tests. (c) The test for gene flow between *S. t. melanopterus* and *S. guttula* from the Louisiade Archipelago, where the two species co-occur. (d) Test for gene flow between *S. t. melanopterus* and *S. t. diadematus*

group in the PCA, their distinctness was supported by sNMF and significantly different (Fisher's exact test, $p < .001$) distributions of mitochondrial haplogroups between the two clusters in PCA space (Figure 3c, S5). Superficially, the mitochondrial haplogroups do not correspond well to our sampling map (Figure 1). However, when only breeding-season specimens are considered, both *trivirgatus* St haplogroup and *guttula* St haplogroup largely correspond to the breeding ranges of *melanorrhous* and *gouldii*, respectively (Figure S6).

Despite no phenotypic variation or subspecies being recognized in *S. guttula* (Beehler & Pratt, 2016), we observed population structure across its populations in eastern mainland New Guinea and offshore islands. Conversely, in both mitochondrial and nuclear trees, Gag *guttula* consistently grouped on its own and was separate from mainland New Guinea samples (Figures 1–2, S7). Though $k = 1$ was our most supported assignment, further exploration of New Guinean *S. guttula* uncovered five geographically relevant subpopulations (Figure S7). We first observed distinct population structure in samples from the Louisiade Archipelago, which are sympatric with *S. t. melanopterus*, then with Normanby Island in the D'Entrecasteaux Archipelago (yet this population was not distinct in the PCA; Figure S7). At $k = 4$, we found a divide north and south of the New Guinea Central Cordillera, and weaker population structure on either side of the Huon Peninsula at $k = 5$. Lacking samples from the Huon Peninsula, we cannot determine where this break lies.

3.4 | Tests of gene flow and demographic analyses

Despite evidence of a mitochondrial capture event, we did not detect significant modern gene flow between the Australian *guttula* St haplogroup or mainland New Guinea *S. guttula* (Figure 4a). This negative result in the face of evidence for mitochondrial capture can probably be explained by the reduced power of ABBA/BABA tests in cases with low levels of admixture and reduced representation of the genome as in this study (Durand et al., 2011). Instead, we revealed a complicated history of gene flow elsewhere within this species complex. For example, ABBA/BABA detected significant gene flow between *S. t. diadematus* and Gag *guttula* (Figure 4b) and also in the Louisiade Archipelago, where *S. t. melanopterus* and a population of *S. guttula* occur sympatrically (Figure 4c); however, gene flow was not detected between *S. t. melanopterus* and other *S. guttula* (Table S2). Unexpectedly, we also detected significant gene flow between *S. t. melanopterus* and *S. t. diadematus* (Figure 4d).

Our TREEMIX analysis did not support gene flow between mainland New Guinean *S. guttula* and the Australian *guttula* St haplogroup. Though not supporting gene flow between the expected taxa, TREEMIX did indicate historical gene flow in this system elsewhere. When unsupervised, TREEMIX produced a topology concordant with RAxML (Figure S3A) and inferred gene flow between *S. t. melanopterus* and the ancestor of the Maluku *S. trivirgatus* populations (Figure S3B; not significant by the jackknifing algorithm, $p = .18$). Our supervised analysis supported gene flow from *S. t. diadematus* to Gag *guttula* ($p = 0.03$; admixture proportion = 0.217), but resulted in a major topological rearrangement (Figure S3c). Other supervised tests of historical gene flow were not statistically significant and resulted in no topological change, which highlights the authenticity of the inferred gene flow from *S. t. diadematus* to Gag *guttula*, as opposed to an artifact due to manually adding a migration edge.

However, our demographic analyses supported at least one recent, discrete admixture event between *S. guttula* and *guttula* St, the mitochondrially introgressed individuals of *S. trivirgatus*. DADI indicated the best fit model was one with recent, bidirectional admixture without ongoing migration (Table 3). All well-supported models (AICw of admixture models = .991) included a discrete gene flow event; however, we acknowledge that it was difficult to distinguish between different models that include discrete gene flow events. For models with discrete and continuous gene flow in the same direction, the amount of inferred gene flow in that direction was nearly nonexistent (less than 1 migrant every 100 generations). The *S. guttula* population size was approximately 10x larger than the included *S. trivirgatus* population (Table 3), and DADI indicated low admixture proportions for each direction of admixture (~1.0–2.5% for each model) that occurred approximately 5,000 years ago. However, estimates of divergence time and population size in DADI for regions with high linked selection can perform poorly, whereas estimates of gene flow tend to perform well (Ewing & Jensen, 2016).

3.5 | Selection in the mitogenome

All tests for selection in the mitogenome did not indicate positive selection or a sweep for *S. trivirgatus* individuals with *S. guttula* mtDNA (Figure S8). Specifically, the McDonald-Kreitman (MK) tests and associated calculation of the neutrality indices for all mitochondrial genes found no evidence for positive selection (Table S4). All significant values of the MK test had a neutrality index greater than

TABLE 3 Results of DADI analyses, including estimates of parameters for population sizes, migration rates (in number of migrants), admixture proportion, and admixture timing

Model	<i>S. guttula</i> population	<i>S. trivirgatus</i> population	Divergence time (years)	Time of admixture (years)	Admixture proportion of <i>S. guttula</i> into <i>S. trivirgatus</i> genome	Admixture proportion of <i>S. trivirgatus</i> into <i>S. guttula</i> genome	Number of <i>S. trivirgatus</i> migrants into <i>S. guttula</i> per generation	Number of <i>S. guttula</i> migrants into <i>S. trivirgatus</i> per generation	AIC	AICw
No migration	1,418,680	117,277	490,141	-	-	-	-	-	508.0	1.0E-06
Continuous gene flow from <i>S. trivirgatus</i> to <i>S. guttula</i>	1,307,927	114,211	529,005	-	-	4.7E-01	-	-	491.1	4.7E-03
Continuous gene flow from <i>S. guttula</i> to <i>S. trivirgatus</i>	1,341,420	106,649	511,957	-	-	-	6.3E-02	-	498.2	1.4E-04
Continuous bidirectional gene flow between both species	1,294,535	109,161	532,143	-	-	3.6E-01	3.0E-02	-	491.6	3.7E-03
Single admixture event from <i>S. guttula</i> to <i>S. trivirgatus</i>	1,439,112	103,767	493,886	7,589	2.5E-02	-	-	-	489.4	1.1E-02
Single admixture event from <i>S. guttula</i> to <i>S. trivirgatus</i> and continuous bidirectional gene flow	1,383,458	101,336	523,241	3,046	2.5E-02	-	3.4E-01	5.7E-06	483.8	1.8E-01
Single admixture event from <i>S. trivirgatus</i> to <i>S. guttula</i>	1,327,902	114,725	503,848	19,361	-	2.0E-02	-	-	490.4	6.8E-03
Single admixture event from <i>S. trivirgatus</i> to <i>S. guttula</i> and continuous bidirectional gene flow	1,314,031	110,140	486,650	42,425	-	2.0E-02	3.2E-03	2.3E-02	493.4	1.5E-03
Single bidirectional admixture event	1,405,950	101,890	515,915	4,957	2.5E-02	1.3E-02	-	-	481.0	7.3E-01
Single bidirectional admixture event, continuous bidirectional gene flow	1,401,649	101,280	520,412	4,627	2.5E-02	1.0E-03	2.9E-01	5.6E-04	486.0	6.0E-02

one or undefined (due to a lack of fixed nonsynonymous substitutions). Some loci had a neutrality index less than one, but it was not statistically significant. Furthermore, the proportion of amino acid substitutions per locus barely differed between groups, suggesting none had an excess of changes relative to others. Similarly, the MK and NI tests for mitochondrial sweeps barely deviated from neutral demographic expectations. Values across all loci and mtDNA groups were negative, but the values for captured *S. guttula* mtDNA were both lower and less significant than those of *trivirgatus* *St* or *S. guttula* itself. In particular, results for *S. trivirgatus* suggested a recent bottleneck followed by population expansion, as all three statistics were significantly negative ($p < .05$) at five loci (Table S5), and marginally significant ($0.05 < p < .10$) at up to three other loci depending on the test. Additionally, CodeML analyses showed little support for a selective sweep in the mitochondria of *S. trivirgatus* individuals with *S. guttula* mtDNA. Of the 13 mitochondrial protein-coding genes, only two genes, ATP6 and ND2, indicated positive selection in both the M7 versus M8, and M8a versus M8 model comparisons and each contained one positively selected site (ATP6: Position 41, N to S; ND2: Position 31, T to A) with a posterior probability over 0.95 (Table S6, Figure S9). One other gene, COX1, showed statistical support ($p < .05$) for the M7 versus M8 model comparison, but not M8a versus M8, which yields fewer false positives (Gao et al., 2019; Swanson et al., 2003; Wong et al., 2004), and had one positively-selected site with a posterior probability over 0.95 (Position 473; V to I). However, there was no phylogenetic pattern relating to the capture of *S. guttula* mitochondria into Australian *S. trivirgatus* (Figure S9). Four other genes (CYTB, ND1, ND4, ND5) also showed statistically significant results for the M7 versus M8 model comparisons but not the M8a versus M8 comparisons; these did not have strongly supported positively-selected amino acid sites (Table S6).

4 | DISCUSSION

Our genomic investigation of the Spectacled Monarch *Symposiachrus trivirgatus* and Spot-winged Monarch *S. guttula* complex in Australia, Papua New Guinea, and Indonesia revealed a complex evolutionary history. This includes a clear instance of massively discordant mitochondrial introgression (MDMI, or mitochondrial capture) from *S. guttula* into its nonsister species *S. trivirgatus*. Our key findings were: (a) strong support for mitochondrial capture of *S. guttula* by its nonsister species *S. trivirgatus* (Figure 1), (b) no evidence of a selective sweep or selection for *S. trivirgatus* individuals that possess the *S. guttula* mitochondrial genome (Table S4, Figure S8), (c) nuclear DNA is only weakly associated with mtDNA haplogroups with low levels of admixture, (d) some signal of historical gene flow ~5,000 years ago between Australian of *S. trivirgatus* and *S. guttula*, and (e) unexpected evidence of two other instances of gene flow between allopatric island subspecies of *S. trivirgatus* and *S. guttula* at opposite ends of the latter's range: Indonesia in the west and offshore islands of far southeastern Papua New Guinea in the east. We now address

the details of our findings in terms of the four hypotheses proposed at the outset: cryptic species, gene flow, neutral capture, and adaptive capture. We first discuss interactions between Australian *S. trivirgatus* and *S. guttula* and then between Indonesian *S. trivirgatus* and *S. guttula*.

4.1 | Australian *S. trivirgatus* and *S. guttula*: multiple hypotheses supported?

Symposiachrus trivirgatus, sampled from across Australia, was assigned to either of two highly divergent nonsister haplogroups across the *Symposiachrus* mitogenome tree (Figure 1). UCE-based analyses of nuclear DNA found no similar patterns, *S. trivirgatus* and *S. guttula* being reciprocally monophyletic with respect to each other (with the exception of the Maluku taxa, which are a separate monophyletic clade). There is no signal of nuclear introgression between *S. guttula* and *S. trivirgatus* that possess the *S. guttula* mitochondria (Figure 2, S4). They are affirmed as nonsister species and we can reject the cryptic species hypothesis in Australian *S. trivirgatus*.

To address the neutral hypothesis, we first consider our results on the timing of the capture. Our results are consistent with Pleistocene glacial cycles that facilitated geographic and genetic contact between *S. guttula* and *S. trivirgatus*. During these cycles, lowered sea levels exposed the vast Sahul Shelf connecting northern Australia and New Guinea up to eight times (Lambeck et al., 2002; Lambeck & Chappell, 2001; Voris, 2000) and for much of the last 100,000 years (review in Joseph et al., 2019; Voris, 2000). By promoting connectivity between faunas of Australia and New Guinea (Heinsohn & Hope, 2006; Peñalba et al., 2019), repeated episodes of range expansion, secondary contact, and gene flow between *S. guttula* and *S. trivirgatus* would have been possible. Based on ND2 substitution rates derived from birds (2.1%–3.3%; Lerner et al., 2011; Weir & Schluter, 2007), our results suggest gene flow could have occurred between Australian *S. trivirgatus* and New Guinea *S. guttula* as recently as 5,000 years ago (Table 3) or as long ago as up to 200,000–400,000 years ago. The uncertainty is attributed to repeated bouts of exposure of the Sahul Shelf as sea levels rose and fell. Nonetheless, our most recent inferred estimate for an admixture event is consistent with simulation studies suggesting that genomic data will reflect only the most recent glacial cycles (Linck & Battey, 2019). The best model supported bidirectional historic admixture but the strongest supported directionality from *S. guttula* into *S. trivirgatus*. This again is consistent with *S. trivirgatus* having expanded into the range of *S. guttula*. On balance, a neutral capture implies a novel corollary: extinction of *S. guttula* from an unknown proportion of its range on the Australian mainland or the now submerged Sahul Shelf. To the extent that we cannot reject neutrality, this capture is consistent with palaeoenvironmental data allowing for the expansion of the range of *S. trivirgatus* into that of *S. guttula* (Currat et al., 2008; Rheindt & Edwards, 2011). This could have resulted in a high frequency of *S. guttula* mitochondrial DNA in Australian *S. trivirgatus*

populations and its near fixation in two genetic clusters. One of these clusters comprises all samples of *S. t. albiventris* in northern-most Australia and the other is widespread across the eastern coast of Australia (Figures 1 and 3).

Next, we note that our tests for selection found little to no evidence for positive selection acting on the mitochondria since the capture event for the *guttula* *St* haplogroup. Superficially, this is consistent with a neutral capture. Again, however, we note the difficulties posed by Bonnet et al. (2017) in rejecting adaptive mechanisms completely. The near fixation in two different nuclear clusters suggests the need to still consider adaptive mechanisms (Figure 3). The few mitochondrial genes with statistically significant deviations from neutrality were consistent with weakly deleterious mutations (Tables S3; Charlesworth & Eyre-Walker, 2008). This pattern is consistent with little to no positive selection acting on entire genes since the mitochondrial capture event and thus does not support an adaptive explanation.

Next we consider our tests for gene flow. The low proportion of admixture inferred by our demographic modeling may suggest a low likelihood of the neutral capture hypothesis. However, multiple instances of secondary contact caused by glacial cycles, the feasibility of which we have noted, can explain this pattern. If the *guttula* *St* mitochondrial haplotype introgressed into Australian populations of *S. trivirgatus* at a frequency of 2.5% (Table 3) each of the eight times that the Sahul Shelf was exposed during our timing estimates (e.g., between 10,000 and 400,000 years ago), we would expect an 18% chance of fixation by purely neutral demographic processes. Drift alone could therefore explain this observed pattern without a selective sweep. As discussed earlier, the directionality of introgression could further promote a neutral capture. Further, fixation in certain genetic clusters could be more likely if effective population sizes were low. This would decrease the expected time to fixation of an allele. For a given gene introgressed into a population, the expectation is that one allele (or, in this case, mitochondrial haplotype) will fix after an average of N_e generations. This process is accelerated for mtDNA, whose effective population size (N_e) is smaller and fixation time is fourfold faster than that of the nuclear genome.

To summarize, our extra sampling has confirmed a case of massively discordant mitochondrial introgression from New Guinean *S. guttula* into Australian *S. trivirgatus*. Due to their reciprocal monophyly as shown by nuclear data, we can reject the cryptic species hypothesis. We find, however, that while there is support for a neutral capture model and some gene flow, we cannot say with certainty that selective capture is not also at play in this system.

4.2 | Complex patterns of gene flow

Here we discuss three observations concerning populations on satellite islands at opposite ends of New Guinea: *S. guttula* on Gag Island immediately west of New Guinea, *S. trivirgatus diadematus* of the Maluku Islands further west of New Guinea, and *S. trivirgatus*

melanopterus of the Louisiade Archipelago off New Guinea's south-eastern tip. We plan a separate, more detailed discussion of species limits and biogeography of this group elsewhere (McCullough et al. unpublished data).

First, concatenated UCE data supported Gag Island *S. guttula* as sister to the monophyletic Maluku *S. trivirgatus* subspecies (*bimaculatus*, *diadematus*, and *nigrimentum*) and this group is in turn sister to mainland New Guinean *S. guttula*. Thus, both mtDNA (Andersen et al., 2015) and UCEs suggested Gag and mainland New Guinean *S. guttula* are paraphyletic. Species-tree and network analyses, however, were equivocal in determining the relationships of Gag Island *S. guttula*. ABBA/BABA tests detected significant gene flow between Maluku *S. t. diadematus* and Gag *S. guttula* (Figure 4b). Simultaneous analysis of phylogenetic relationships and introgression supported the equivocal relationships of Gag *S. guttula* was due to high levels of gene flow from *S. t. diadematus* or a related taxon (Figure S3C). These lines of evidence suggest that Gag *S. guttula* is a hybrid population, but this hypothesis lacks evidence from plumage characters.

Second, *S. t. melanopterus* of the Louisiades at the opposite end of mainland New Guinea is phenotypically and genetically distinct but comes into contact with *S. guttula* (Figure S6). We inferred some gene flow between the two in sympatry (Figure 4c). This is consistent with other cases of gene flow within this complex. Whether such gene flow is ongoing requires further study. However, we note that the subspecies-defining features of *S. t. melanopterus* (white behind auriculars and dark wing coverts; Figure 1) are similar to features of *S. guttula*, which may suggest a hybrid origin of this taxon.

Third, we detected significant and unexpected gene flow between *S. t. melanopterus* and *S. t. diadematus*, which are currently isolated from each other by the entirety of mainland New Guinea (Figure 4d). This could reflect Pleistocene connections across the exposed Sahul Shelf as modelled in similarly distributed species by Peñalba et al. (2019). Alternatively, it may reflect a now extinct "ghost" taxon (Benham & Cheviron, 2019; Gopalakrishnan et al., 2018; Zhang et al., 2019) in the Louisiades. Such a ghost taxon would have been related to *S. t. diadematus* but would have to have been replaced by what is now *S. t. melanopterus*.

4.3 | Future directions and conclusions

Monarch-flycatchers are an iconic, species-rich radiation that span the Indo-Pacific. They have played a major role in the development of speciation, biogeography, and community assembly theory (Mayr & Diamond, 2001). Here, we studied one species complex within one genus, *Symposiachrus*. We find that hybridization between *S. guttula* and *S. trivirgatus*, two nonsister species, has itself had a complex history of different outcomes concerning gene flow and mitochondrial capture. Although we found no evidence for adaptive introgression, and the captured mitochondrial haplogroup has not yet gone to fixation in all Australian *S. trivirgatus*, theoretical findings suggest that an adaptive capture is difficult to rule out (Bonnet et al., 2017). Further study of this and other mitochondrial discordance systems needs to

consider joint modelling of nuclear and mitochondrial genomes and their interactions. This stresses that it may be inappropriate to consider a neutral model as the null to be rejected and that, instead, joint modelling and testing of neutral and adaptive scenarios will be more fruitful. One area of such future study has been raised by Rougemont et al. (2020) and Kim et al. (2018); that is, how past introgression may maintain maladaptive alleles in such modelling is an exciting dimension to future work and should be possible as genomic resources for these monarch-flycatchers increase. Additionally, higher resolution genomic data will allow for more accurate testing of possible introgressed nuclear loci associated with captured mitochondrial haplotypes. Finally, at a broader scale, we suggest that denser geographic and genomic sampling of all *Symposiachrus* taxa from Wallacea to Melanesia will offer excellent opportunities to explore new frontiers and build on the rich history of evolutionary theory that is central to this region of the world, and could add to the story of widespread admixture found here.

ACKNOWLEDGEMENTS

We thank the following museums and collections staff for providing tissue and toepad loans and other support: Paul Sweet, Tom Trombone, and Peter Capainolo, American Museum of Natural History; Robert Palmer and Alex Drew, Australian National Wildlife Collection, CSIRO, Canberra; Moe Flannery and Jack Dumbacher, California Academy of Sciences; Bulisa Iova, Papua New Guinea National Museum and Art Gallery; Mark Robbins, University of Kansas Biodiversity Institute; Ron Johnstone and Rebecca Bray, Western Australia Museum; Kristof Zyskowski, Yale Peabody Museum, and Jan Bolding Kristensen, Natural History Museum of Denmark. We thank the Ministry of Research and Technology/National Agency for Research and Innovation (RISTEK-BRIN; Research permit SIP No: 026/SIP/FRP/1/2012 to K.A.J); the Ministry of Environment and Forestry, Republic of Indonesia; the Research Center for Biology, Indonesian Institute of Sciences (RCB-LIPI) for providing permits to carry out fieldwork in Indonesia and to export select samples. In addition, we thank Fangluan Gao for assistance with Codeml analyses. LJ acknowledges numerous scientific research permits granted to the Australian National Wildlife Collection over several decades and that supported this work in Australia and PNG. We thank the UNM Center for Advanced Research Computing (CARC) and Cyberinfrastructure for Phylogenetic Research (CIPRES), supported in part by the National Science Foundation, for providing high-performance computing resources used in this study. JMM acknowledges the UNM Grove Summer Research Scholarship. EFG acknowledges NSF's Graduate Research Fellowship [DGE-650114]. KAJ acknowledges a National Geographic Research and Exploration Grant (8853-10), the Dybron Hoffs Foundation, and the Corrit Foundation for financial support for fieldwork in Indonesia. KAJ is also grateful for the financial support received from the Villum Foundation (Young Investigator Programme, project no. 15560). This work was supported by the University of New Mexico and a National Science Foundation award to MJA [DEB-1557051]. We gratefully acknowledge the comments of four reviewers and helpful discussions with Timothée Bonnet.

AUTHOR CONTRIBUTIONS

M.A. and L.J. conceived the study. T.H., K.A.J., and L.J. collected genetic samples. J.M. and X.M. collected the sequence data. M.A., J.M., and E.G. analysed the data and generated figures. M.A., J.M., E.G., and L.J. were principal writers; all authors contributed to writing and editing the manuscript.

DATA AVAILABILITY STATEMENT

Population genetic and phylogenetic input files will be uploaded to Dryad (<https://doi.org/10.5061/dryad.hx3ffbgd7>). Raw Illumina sequencing reads are available from the NCBI SRA (BioProject PRJNA698220). Mitogenomes are accessioned in NCBI GenBank (Accession #'s MW623660–MW623735). All custom scripts generated for this study will be uploaded to Dryad, with updated versions available on github: https://github.com/ethangyllenhaal/assorted_scripts.

ORCID

Michael J. Andersen  <https://orcid.org/0000-0002-7220-5588>
 Jenna M. McCullough  <https://orcid.org/0000-0002-7664-3868>
 Ethan F. Gyllenhaal  <https://orcid.org/0000-0002-0835-8520>
 Xena M. Mapel  <https://orcid.org/0000-0002-7501-578X>
 Knud A. Jønsson  <https://orcid.org/0000-0002-1875-9504>
 Leo Joseph  <https://orcid.org/0000-0001-7564-1978>

REFERENCES

- Andersen, M. J., Hosner, P. A., Filardi, C. E., & Moyle, R. G. (2015). Phylogeny of the monarch flycatchers reveals extensive paraphyly and novel relationships within a major Australo-Pacific radiation. *Molecular Phylogenetics and Evolution*, 83(C), 118–136.
- Andersen, M. J., McCullough, J. M., Friedman, N. R., Peterson, A. T., Moyle, R. G., Joseph, L., & Nyári, Á. S. (2019). Ultraconserved elements resolve genus-level relationships in a major Australasian bird radiation (Aves: Meliphagidae). *Emu - Austral Ornithology*, 119(3), 218–232.
- Beehler, B. M., & Pratt, T. K. (2016). *Birds of New Guinea: Distribution, taxonomy, and systematics*. Princeton University Press.
- Benham, P. M., & Cheviron, Z. A. (2019). Divergent mitochondrial lineages arose within a large, panmictic population of the Savannah sparrow (*Passerculus sandwichensis*). *Molecular Ecology*, 28(7), 1765–1783.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.
- Bonnet, T., Leblois, R., Rousset, F., & Crochet, P.-A. (2017). A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *Evolution*, 71, 2140–2158. <https://doi.org/10.1111/evo.13296>.
- Borge, T., Lindroos, K., Nadvornik, P., Syvänen, A.-C., & Sætre, G.-P. (2005). Amount of introgression in flycatcher hybrid zones reflects regional differences in pre and post-zygotic barriers to gene exchange. *Journal of Evolutionary Biology*, 18(6), 1416–1424.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10(4), e1003537.
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, 29(8), 1917–1932.
- Burton, R. S., & Barreto, F. S. (2012). A disproportionate role for mt DNA in D obzhansky–Muller incompatibilities? *Molecular Ecology*, 21(20), 4942–4957.
- Burton, R. S., Pereira, R. J., & Barreto, F. S. (2013). Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematics*, 44(1), 281–302. <https://doi.org/10.1146/annurev-ecolsys-110512-135758>
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4), 540–552.
- Charlesworth, J., & Eyre-Walker, A. (2008). The McDonald–Kreitman test and slightly deleterious mutations. *Molecular Biology and Evolution*, 25(6), 1007–1015.
- Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30(23), 3317–3324.
- Claramunt, S., & Cracraft, J. (2015). A new time tree reveals Earth history's imprint on the evolution of modern birds. *Science Advances*, 1(11), e1501005.
- Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, 9(10), 1657–1659.
- Clements, J. F., Schulenberg, T. S., Iliff, M. J., Billerman, S. M., Fredericks, T. A., Sullivan, B. L., & Wood, C. L. (2019). *The eBird/Clements Checklist of Birds of the World: v2019*. Retrieved from <http://www.birds.cornell.edu/clementschecklist/download/>
- Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: Massive introgression by local genes. *Evolution; International Journal of Organic Evolution*, 62(8), 1908–1920.
- 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., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- del Hoyo, J., & Collar, N. J. (2016). *HBW and BirdLife International Illustrated Checklist of the Birds of the World. Volume 2: Passerines*. Lynx Edicions.
- Dickinson, E. C., & Christidis, L. (Eds.). (2014). *The Howard and Moore Complete Checklist of the Birds of the World, Volume 2: Passerines*. Aves Press.
- Drovetski, S. V., Semenov, G., Red'kin, Y. A., Sotnikov, V. N., Fadeev, I. V., & Koblik, E. A. (2015). Effects of asymmetric nuclear introgression, introgressive mitochondrial sweep, and purifying selection on phylogenetic reconstruction and divergence estimates in the Pacific clade of *Locustella* warblers. *PLoS One*, 10(4), e0122590.
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28(8), 2239–2252.
- Eaton, J. A., van Balen, S., Brickle, N. W., & Rheindt, F. E. (2016). *Birds of the Indonesian Archipelago: Greater Sundas and Wallacea*. Lynx.
- Edwards, S. V., Potter, S., Schmitt, C. J., Bragg, J. G., & Moritz, C. (2016). Reticulation, divergence, and the phylogeography–phylogenetics continuum. *Proceedings of the National Academy of Sciences*, 113(29), 8025–8032.
- Egea, R., Casillas, S., & Barbadilla, A. (2008). Standard and generalized McDonald–Kreitman test: A website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research*, 36(suppl_2), W157–W162.
- Ewing, G. B., & Jensen, J. D. (2016). The consequences of not accounting for background selection in demographic inference. *Molecular Ecology*, 25(1), 135–141.
- Faircloth, B. C. (2013). *illumiprocessor: A trimmomatic wrapper for parallel adapter and quality trimming*. <https://doi.org/10.1145/3132847.3132886>
- Faircloth, B. C. (2016). PHYLUCe is a software package for the analysis of conserved genomic loci. *Bioinformatics*, 32(5), 786–788.
- Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61(5), 717–726.
- Filardi, C. E., & Moyle, R. G. (2005). Single origin of a pan-Pacific bird group and upstream colonization of Australasia. *Nature*, 438(7065), 216–219.

- Filardi, C. E., & Smith, C. E. (2005). Molecular phylogenetics of monarch flycatchers (genus *Monarcha*) with emphasis on Solomon Island endemics. *Molecular Phylogenetics and Evolution*, 37(3), 776–788.
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6(8), 925–929.
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, 196(4), 973–983.
- Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, 133(3), 693–709.
- Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., & Ho, S. Y. W. (2019). EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecology and Evolution*, 9, 3891–3898. <https://doi.org/10.1002/ece3.5015>.
- Good, J. M., Hird, S., Reid, N., Demboski, J. R., Stepan, S. J., Martin-Nims, T. R., & Sullivan, J. (2008). Ancient hybridization and mitochondrial capture between two species of chipmunks. *Molecular Ecology*, 17(5), 1313–1327.
- Gopalakrishnan, S., Sinding, M.-H., Ramos-Madrugal, J., Niemann, J., Samaniego Castruita, J. A., Vieira, F. G., Carøe, C., Montero, M. D. M., Kuderna, L., Serres, A., González-Basallote, V. M., Liu, Y.-H., Wang, G.-D., Marques-Bonet, T., Mirarab, S., Fernandes, C., Gaubert, P., Koepfli, K.-P., Budd, J., ... Gilbert, M. T. P. (2018). Interspecific gene flow shaped the evolution of the genus *Canis*. *Current Biology: CB*, 28(21), 3441–3449.e5.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652.
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5(10), e1000695.
- Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology*, 65(5), 910–924.
- Heinsohn, T., & Hope, G. (2006). The Torresian connections: Zoogeography of New Guinea. In *Evolution and biogeography of Australian vertebrates*. Australian Scientific Publishing.
- Hill, G. E. (2019). Mitonuclear Ecology. *Molecular Biology and Evolution*, 32(8), 1917–1927.
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405.
- Joseph, L. (2018). Phylogeography and the role of hybridization in speciation. In D. T. Tietze (Ed.), *Bird species: How they arise, modify and vanish* (pp. 165–194). Springer International Publishing.
- Joseph, L., Bishop, K. D., Wilson, C. A., Edwards, S. V., Iova, B., Campbell, C. D., Mason, I., & Drew, A. (2019). A review of evolutionary research on birds of the New Guinean savannas and closely associated habitats of riparian rainforests, mangroves and grasslands. *Emu - Austral Ornithology*, 119(3), 317–330.
- Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: Simulations and implications for human population structure. *Heredity*, 106(4), 625–632.
- 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.
- Kearns, A. M., Joseph, L., Toon, A., & Cook, L. G. (2014). Australia's arid-adapted butcherbirds experienced range expansions during Pleistocene glacial maxima. *Nature Communications*, 5, 3994.
- Kim, B. Y., Huber, C. D., & Lohmueller, K. E. (2018). Deleterious variation shapes the genomic landscape of introgression. *PLoS Genetics*, 14(10), e1007741.
- Lambeck, K., & Chappell, J. (2001). Sea level change through the last glacial cycle. *Science*, 292(5517), 679–686.
- Lambeck, K., Esat, T. M., & Potter, E.-K. (2002). Links between climate and sea levels for the past three million years. *Nature*, 419(6903), 199–206.
- Lanfear, R., Calcott, B., Kainer, D., Mayer, C., & Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology*, 14, 82.
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34(3), 772–773.
- Leigh, J. W., & Bryant, D. (2015). popart: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110–1116.
- Lerner, H. R. L., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011). Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Current Biology*, 21(21), 1838–1844.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760.
- 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.
- Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, 19(3), 639–647.
- Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., & Usadel, B. (2012). RobiNA: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research*, 40, W622–W627.
- Mallet, J. (2005). Speciation in the 21st Century. *Heredity*, 95, 105–109.
- Mayr, E., & Diamond, J. M. (2001). *The birds of northern melanesia: Speciation, ecology, and biogeography*. Oxford University Press.
- McCormack, J. E., Tsai, W. L. E., & Faircloth, B. C. (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*, 16(5), 1189–1203.
- McCullough, J. M., Joseph, L., Moyle, R. G., & Andersen, M. J. (2019). Ultraconserved elements put the final nail in the coffin of traditional use of the genus *Meliphaga* (Aves: Meliphagidae). *Zoologica Scripta*, 48(4), 411–418.
- McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, 351(6328), 652–654.
- McElroy, K., Black, A., Dolman, G., Horton, P., Pedler, L., Campbell, C. D., & Joseph, L. (2020). Robbery in progress: Historical museum collections bring to light a mitochondrial capture within a bird species widespread across southern Australia, the Copperback Quail-thrush *Cinclosoma clarum*. *Ecology and Evolution*, 10(13), 6785–6793.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, 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.
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Conference*, 1–8.
- Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey, J. D., Travers, S. L., Brown, R. M., & Faircloth, B. C. (2016). Tectonic collision and uplift of Wallacea triggered the global songbird radiation. *Nature Communications*, 7, 12709.
- Mundy, N. I., Unitt, P., & Woodruff, D. S. (1997). Skin from feet of museum specimens as a non-destructive source of DNA for avian genotyping. *The Auk*, 114(1), 126–129.
- Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, 17(9), 2103–2106.

- Omland, K. E., Baker, J. M., & Peters, J. L. (2006). Genetic signatures of intermediate divergence: Population history of Old and New World Holarctic ravens (*Corvus corax*). *Molecular Ecology*, *15*(3), 795–808.
- Oswald, J. A., Overcast, I., Mauck, W. M. 3rd, Andersen, M. J., & Smith, B. T. (2017). Isolation with asymmetric gene flow during the nonsynchronous divergence of dry forest birds. *Molecular Ecology*, *26*(5), 1386–1400.
- Ottenburghs, J., Kraus, R. H. S., van Hooff, P., van Wieren, S. E., Ydenberg, R. C., & Prins, H. H. T. (2017). Avian introgression in the genomic era. *Avian Research*, *8*(1), 30.
- Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), 2337–2360.
- Peñalba, J. V., Joseph, L., & Moritz, C. (2019). Current geography masks dynamic history of gene flow during speciation in northern Australian birds. *Molecular Ecology*, *28*(3), 630–643.
- Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology & Evolution*, *24*(7), 386–393.
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, *8*(11), e1002967.
- Pons, J.-M., Sonsthagen, S., Dove, C., & Crochet, P.-A. (2014). Extensive mitochondrial introgression in North American Great Black-backed Gulls (*Larus marinus*) from the American Herring Gull (*Larus smithsonianus*) with little nuclear DNA impact. *Heredity*, *112*, 226–239. <https://doi.org/10.1038/hdy.2013.98>.
- R Development Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904.
- Rand, D. M., & Kann, L. M. (1996). Excess amino acid polymorphism in mitochondrial DNA: Contrasts among genes from *Drosophila*, mice, and humans. *Molecular Biology and Evolution*, *13*(6), 735–748.
- Rheindt, F. E., & Edwards, S. V. (2011). Genetic introgression: An integral but neglected component of speciation in birds. *The Auk*, *128*(4), 620–632.
- Rougemont, Q., Moore, J.-S., Leroy, T., Normandeau, E., Rondeau, E. B., Withler, R. E., Van Doornik, D. M., Crane, P. A., Naish, K. A., Garza, J. C., Beacham, T. D., Koop, B. F., & Bernatchez, L. (2020). Demographic history shaped geographical patterns of deleterious mutation load in a broadly distributed Pacific Salmon. *PLoS Genetics*, *16*(8), e1008348.
- Rougeux, C., Gagnaire, P., & Bernatchez, L. (2019). Model-based demographic inference of introgression history in European whitefish species pairs. *Journal of Evolutionary Biology*, *32*(8), 806–817.
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302.
- Schodde, R., & Mason, I. J. (1999). *The directory of Australian birds: Passerines*. CSIRO Publishing.
- Shipham, A., Schmidt, D. J., Joseph, L., & Hughes, J. M. (2016). A genomic approach reinforces a hypothesis of mitochondrial capture in eastern Australian rosellas. *The Auk: Ornithological Advances*, *134*(1), 181–192.
- Sloan, D. B., Havird, J. C., & Sharbrough, J. (2017). The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology*, *26*(8), 2212–2236.
- Smadja, C. M., & Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, *20*(24), 5123–5140.
- Smeds, L., Qvarnström, A., & Ellegren, H. (2016). Direct estimate of the rate of germline mutation in a bird. *Genome Research*, *26*(9), 1211–1218.
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313.
- Swanson, W. J., Nielsen, R., & Yang, Q. (2003). Pervasive adaptive evolution in mammalian fertilization proteins. *Molecular Biology and Evolution*, *20*(1), 18–20.
- Swofford, D. (2003). *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595.
- Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, *3*(2), 170–177.
- Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*(16), 3907–3930.
- Toews, D. P. L., Mandic, M., Richards, J. G., & Irwin, D. E. (2014). Migration, mitochondria, and the yellow-rumped warbler. *Evolution; International Journal of Organic Evolution*, *68*(1), 241–255.
- Tsai, W. L. E., Schedl, M. E., Maley, J. M., & McCormack, J. E. (2019). More than skin and bones: Comparing extraction methods and alternative sources of DNA from avian museum specimens. *Molecular Ecology Resources*, *20*, 1220–1227.
- 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.
- Weir, J. T., & Schluter, D. (2007). The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, *315*(5818), 1574–1576.
- Winker, K., Glenn, T. C., & Faircloth, B. C. (2018). Ultraconserved elements (UCEs) illuminate the population genomics of a recent, high-latitude avian speciation event. *PeerJ*, *6*, e5735.
- Wong, W. S. W., Yang, Z., Goldman, N., & Nielsen, R. (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*, *168*(2), 1041–1051.
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591.
- Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C. G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O. G., Janacek, S. H., Juettemann, T., To, J. K., ... Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Research*, *46*(D1), D754–D761.
- Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821–829.
- Zhang, D., Tang, L., Cheng, Y., Hao, Y., Xiong, Y., Song, G., Qu, Y., Rheindt, F. E., Alström, P., Jia, C., & Lei, F. (2019). “Ghost introgression” as a cause of deep mitochondrial divergence in a bird species complex. *Molecular Biology and Evolution*, *36*(11), 2375–2386.
- Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, *7*(1–2), 203–214.

SUPPORTING INFORMATION

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

How to cite this article: Andersen MJ, McCullough JM, Gyllenhaal EF, et al. Complex histories of gene flow and a mitochondrial capture event in a nonsister pair of birds. *Mol Ecol*. 2021;30:2087–2103. <https://doi.org/10.1111/mec.15856>