Syst. Biol. 0(0):1–19, 2025

https://doi.org/10.1093/sysbio/syaf066
Advance Access publication September 27, 2025

Introgressed Variants Obscure Phylogenetic Relationships but Are Not Subject to Positive Selection in Australasian Long-Tailed Parrots

Brian Tilston Smith[1,*], Agusto Luzuriaga-Neira[1], David Alvarez-Ponce[2], Kaiya L. Provost[3], Gregory Thom[4], and Leo Joseph[5]

[1]*Department of Ornithology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA*
[2]*Biology Department, University of Nevada, Reno, Reno, NV 89557, USA*
[3]*Biology Department, Adelphi University, Garden City, NY 11530, USA*
[4]*Museum of Natural Science and Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA*
[5]*Australian National Wildlife Collection, National Research Collections Australia, CSIRO, GPO Box 1700, Canberra, ACT, 2601, Australia*
[*]*Correspondence to be sent to: Department of Ornithology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA; E-mail: bsmith1@amnh.org*

Received 24 October 2024; reviews returned 17 September 2025; accepted 22 September 2025
Associate Editor: Deren Eaton

Abstract.—Gene flow often obscures phylogenetic relationships, but the evolutionary significance of introgressed variants is unclear. Here, we examine the Australasian long-tailed parrots (Psittaculinae: Polytelini) in which an unexpected sister relationship between *Polytelis alexandrae* and the genus *Aprosmictus*, and not the other *Polytelis* species, has been observed. Using whole genomes, we tested whether this relationship was due to ancient introgression. We found that the majority of gene trees had *Ap. erythropterus* and *P. alexandrae* as sister taxa, whereas network analysis indicated monophyly of *Polytelis*, 48% of gene trees being in phylogenetic conflict due to introgression from *Ap. erythropterus* into *P. alexandrae*. Further analyses confidently confirmed that 4–8% of the genome of *P. alexandrae* was introgressed from *Ap. erythropterus* with signals of gene flow occurring throughout the genome. These findings indicate that topologies with *P. alexandrae* and the genus *Ap. erythropterus* as sister taxa were biased by gene flow and affirm that *Polytelis* is monophyletic. Next, we assessed the evolutionary outcomes for introgressed variants and found that, among introgressed protein-coding genes, only two (0.8%) were under positive selection, in comparison to 99 (1.7%) of non-introgressed genes. Our results indicate that, despite the ubiquity of detectable introgression in phylogenies, many genetic variants flowing between species may play a minor role in molecular adaptations. [Ancient introgression; Australia; birds; gene flow; phylogenetic network; *Polytelis*; positive selection..]

The phylogenomic era can be characterized by two waves of methodological advancement. The first was the incorporation of the multi-species coalescent methods that accounted for the independence of gene trees in phylogenetic inference to address incomplete lineage sorting (ILS) (Edwards 2009). Theoretical and empirical work demonstrated that gene tree discordance is expected and that coalescent-based approaches should produce more reliable parameter estimates in a phylogeny (Degnan and Salter 2005; Degnan and Rosenberg 2006; Maddison and Knowles 2006), an advance that has been instrumental in helping resolve challenging evolutionary relationships. The next wave has focused on addressing the impact of introgression in generating gene tree discordance by recognizing that the assumption of no horizontal gene flow does not hold for many systems (Leaché et al. 2013; Blair and Ané 2020). Recent empirical studies at varying phylogenetic scales routinely identify signatures of ancient introgression and gene flow among non-sister taxa (e.g., Thom et al. 2018; MacGuigan and Near 2019; Zhang et al. 2021; DeBaun et al. 2023), which, when not accounted for, can mislead phylogenetic relationships. Ultimately, the extent of introgression observed in phylogenies is dictated by the ecological and geographic factors governing the likelihood of contact, the degree of reproductive isolation between lineages, and the adaptive outcomes of introgressed alleles (Runemark et al. 2019).

Characterizing which portions of the genome are introgressed and whether they have been subject to natural selection is important for understanding both the sources of phylogenetic conflict and the contributions of gene flow to diversification. Events involving genetic introgression encompass a wide range of evolutionary scenarios and temporal scales. Gene flow could be ongoing between sister or non-sister lineages, the latter scenario having a large impact on tree topology heterogeneity across the genome (Leaché et al. 2013). Alternatively, genetic introgression could be ancient, meaning that gene flow involved ancestral and/or now extinct ghost lineages (Kuhlwilm et al. 2019; Zhang et al. 2019; Durvasula and Sankararaman 2020). The extent of gene flow can also vary across generations or according to how much of the genome was exchanged. In addition, the level of reproductive isolation between taxa, the frequency of genes that are resistant to gene flow, the amount of background selection, and the frequency of crossover events across the genome predict which re-

gions are more likely to introgress (Burri et al. 2015; Burri 2017; Wang et al. 2022). Collectively, the frequency, timing, and extent of gene flow will impact phylogenetic signal and the ability of methods to infer evolutionary relationships.

There has been progress in simultaneously estimating gene flow and the phylogeny (Solís-Lemus and Ané 2016; Wen et al. 2018; Allman et al. 2019), but the methodological work remains a patchwork of approaches. Methods include estimating a tree or network with hybrid edges by summarizing gene trees or allele counts (Hibbins and Hahn 2021). Alternative approaches include the frequently used D-statistic, which uses allele-count patterns leading to an excess of derived alleles under scenarios of gene flow (Patterson et al. 2012). However, the D-statistic does not account for variation in the genealogical history across the genome caused by stochastic fluctuations of coalescence and introgression (Ji et al. 2022). As a result, summary statistics might not accurately capture signatures of ancient introgression or detect gene flow between sister lineages and can have high error rates (Ji et al. 2022; Frankel and Ané 2023). Alternatively, full-likelihood models (e.g., Flouri et al. 2020) that presumably do not suffer from the same error rates, although computationally more demanding, can be implemented. Network-based approaches using either allele frequencies or pre-estimated gene trees are also commonly used to infer phylogenetic relationships with hybrid edges. A general limitation of introgression methods is that they provide a single estimate for the entire genome, making them inappropriate to study the patterns of more localized gene flow and the factors shaping introgression.

Although the ubiquity of introgression in phylogenomic data is increasingly apparent (Blair and Ané 2020), the phylogenetic implications of differential patterns arising from which regions of the genome are introgressed are a new frontier of inquiry. Introgression can occur early in the history of lineages when there are likely fewer reproductive barriers and genetic incompatibilities (Wang et al. 2020). In the absence of genetic incompatibilities, there would be fewer constraints to which part of the genome could introgress, producing greater topological homogeneity across the genome. In contrast, islands of speciation, which contain genes that mediate reproductive isolation, are less likely to introgress or be introgressed (Ravinet et al. 2017). That would in turn limit phylogenetic conflict to a narrow portion of the genome (Ottenburghs et al. 2023). In some cases, introgression has led to hybrid speciation, where the adaptive exchange of genes has directly led to reproductive isolation from parental lineages (Wang et al. 2021; Wu et al. 2023). Introgressed genetic variants could also play a role in adaptation, even if the genetic regions are not involved in genetic barriers between species. Standing genetic variation has been shown to have a larger role in adaptation than new mutations (Pritchard et al. 2010; Lai et al. 2019), and newly introgressed alleles could serve as new targets for natural selection. Understanding the role of introgressed genome regions in adaptation will help elucidate the processes that produce phylogenetic discordance.

The Australasian long-tailed parrots (Psittaculinae: Polytelini; Fig. 1a; Supplementary Fig. S1), which exhibit an unexpected pattern of phylogenetic discordance (Smith et al. 2023), offer an opportunity to test the evolutionary implications of differing levels of gene flow across the genome. Phylogenetic relationships inferred from concatenated alignments placed the three species of *Polytelis* of Australia as monophyletic, with *P. alexandrae* as sister to the sister pair of *P. swainsonii* and *P. anthopeplus* (Fig. 1b; hereafter Topology 1). In contrast, the species tree, estimated from gene trees while accounting for ILS, recovered *P. alexandrae* as sister to the two species of *Aprosmictus* (hereafter Topology 2) that occur in Australia, Papua New Guinea, and Indonesia (Fig. 1c), a relationship that is unexpected on phenotypic grounds (reviewed in Provost et al. 2018; Smith et al. 2024). The three species of *Polytelis* share a long-tailed, highly streamlined general morphology unique among all parrots, unique black and pink under-tail coloration, and very similar vocalizations (Forshaw 2002). They are otherwise highly dissimilar in plumage patterning and coloration. The genus *Aprosmictus*, consisting of *Ap. jonquillaceus* of Indonesia and *Ap. erythropterus* of Australia and New Guinea, are less slender, more compact birds characterized by shorter, broad tails and prominently red wing-coverts. *Aprosmictus* and *Polytelis* are the sister group to the king parrots, *Alisterus*. Similar pink under-tail coloration as in *Polytelis* is also found in *Aprosmictus* and *Alisterus* (details in Fig. 1a), and blue rump plumage is shared by *Polytelis alexandrae, Alisterus,* and *Aprosmictus* albeit much darker and more pronounced in the latter two genera. In captivity, hybridization among species within *Polytelis* and between *Aprosmictus* and *Polytelis* has been recorded (Forshaw 2002; Sindel and Gill 2003) although no backcrosses have been reported to our knowledge. *Polytelis alexandrae* is currently allopatric with respect to all other species in the complex, and any overlap with *Ap. erythropterus* is infrequent and geographically restricted to the peripheries of both species' ranges (see Forshaw 2002).

Here, we consider alternative hypotheses to explain these conflicting observed evolutionary histories, including one scenario not assessed in earlier studies, that is, ancient introgression between *Ap. erythropterus* and *P. alexandrae*. Gene flow among these lineages, which are placed in different genera and share a common ancestor some 5–10 Mya (Smith et al. 2023; Smith et al. 2024), would make these taxa appear to be more closely related than expected. If gene flow did occur, it may have been largely neutral or could have facilitated molecular adaptations over a broader timeframe. We set out to test alternative phylogenetic relationships and patterns of introgression in the long-tailed parrots using species-level whole genome data and verify our find-

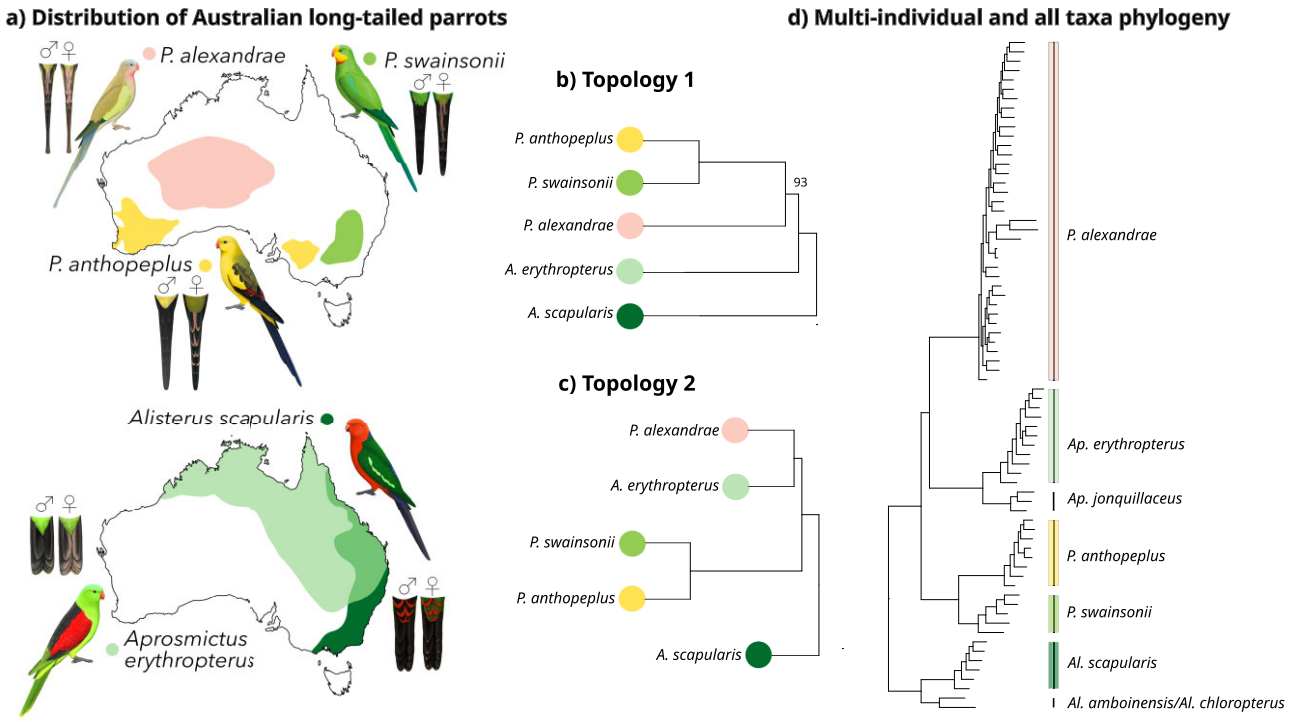

FIGURE 1. Geographic distributions of Australian species of long-tailed parrots and alternative phylogenetic hypotheses for the clade. (a) Shown are color-coded range maps for each species in Australia. The phylogenetic placement of *Polytelis alexandrae* varied across data sets and tree inference methods. Colored tips on the trees correspond to the occurrence of species in Australia. (b) A previously published concatenated tree using ultraconserved elements UCEs and a single individual per taxon found *P. alexandrae* as sister to *P. anthopeplus* and *P. swainsonii* (modified from Smith et al. 2024); we refer to this tree as Topology 1. (c) A single individual whole genome species tree, referred to as Topology 2, and (d) a multi-individual and concatenated UCE tree had *P. alexandrae* and *Ap. erythropterus* as sisters. All nodes have 100% ultrafast bootstrap support (b and d) or (c) 100% local posterior probabilities, unless noted. Bird illustrations and range maps by Julian Teh.

ings with expanded population-level sampling using ultraconserved elements. We test whether the unexpected species tree topology, where *P. alexandrae* and *Ap. erythropterus* are sister lineages, is due to high levels of ancient introgression or whether they are actual sister taxa. Next, we estimated positive selection along the *P. alexandrae* branch and characterized introgressed versus non-introgressed regions. By modeling the levels of gene flow, we are able to resolve phylogenetic relationships in the Australasian long-tailed parrots and show that the majority of gene flow was non-adaptive.

MATERIALS AND METHODS

Whole Genome Sequencing Protocol

We targeted three individuals from each of the following species: *Alisterus scapularis*, *Aprosmictus erythropterus*, *Polytelis alexandrae*, *Polytelis anthopeplus*, and *Polytelis swainsonii*. Because there are no tissue samples from wild individuals of *P. alexandrae*, we used samples from captivity in this part of the study. We did not have high-quality genetic samples of *Aprosmictus jonquillaceus* but it is included in a data set of ultraconserved elements discussed below. Sampling metadata is available in Supplementary Table S1. To extract and quantify DNA, we used a Qiagen high molecular weight DNA kit (MagAttract HMW DNA Kit; Qiagen) and a Qubit 2.0 Fluorometer (Thermo Fisher Scientific), respectively. First, we did a test run and sequenced the 15 individuals at low-coverage (< 5x) on an Illumina HiSeq X Ten PE150 machine. After data validation, we sequenced the same individuals on a lane of Illumina S4 NovaSeq 6000 to produce higher coverage genomes that ranged from 15 to 51x (Supplementary Table S2). Library preparation and sequencing were performed by Rapid Genomics (Gainesville, Florida).

Read Alignment, Variant Calling, and Filtering

We implemented an existing pipeline from Thom et al. (2024) to process raw reads and call variant sites. The pipeline follows a conventional workflow, using well accepted software and thresholds. To trim and filter raw reads, we used trimmomatic v0.36 (Bolger et al. 2014). We mapped raw reads to the chromosome-scale and annotated genome of *Melopsittacus undulatus*

(GCF_012275295.1) as a reference, and this version of the genome contains 30 autosomes, two sex chromosomes, and 832 unplaced scaffolds. This reference genome was the highest quality and most closely related taxon publicly available at the time in which we performed our bioinformatic analyses. Previous cytogenetic work indicated *M. undulatus* has 29 chromosome pairs, whereas *P. alexandrae* and *Alisterus scapularis* have 35 and 38 pairs, respectively (Van Dongen and De Boer 1984; De Lucca 1990; Christidis et al. 2008; Rus et al. 2016). The number of macrochromosomes varies from the reference *M. undulatus* (seven pairs) to *P. alexandrae* (nine) and *Alisterus scapularis* (eight). *Melopsittacus undulatus* and the species in our focal group shared a common ancestor ~28.64 (20.62–33.5) Mya (Smith et al. 2023; Smith et al. 2024), a level of divergence that reduces the efficacy of mapping. Processed reads were aligned to the *M. undulatus* reference using BWA v0.7.17 (Li and Durbin 2009) using the mem algorithm and default parameters. Using Picard v.2.0.1 (Broad Institute, Cambridge, MA; http://broadinstitute.github.io/picard/) and default parameters, we 1) sorted the SAM files produced by BWA with SortSam; 2) reassigned reads to groups with AddOrReplaceReadGroups; 3) identified duplicated reads with Markduplicates; 4) calculated summary statistics with CollectAlignmentSummaryMetrics, CollectInsertSizeMetrics, and CollectRawWgsMetrics; and 5) created indexes with BuildBamIndex. Next, using GATK v3.8 (McKenna et al. 2010), we 1) called SNPs and indels for each individual with HaplotypeCaller; 2) genotyped using GenotypeGVCFs (with heterozygosity set to 0.05); 3) flagged and filtered variants with VariantFiltration. Given the lack of a high confidence SNP panel, we implemented the hard filtering options recommended by the Broad Institute's Best Practices (https://gatk.broadinstitute.org/). We used the following settings to hard filter SNPs: 1) quality by depth below 2 (QD < 2.0); 2) SNPs where reads containing the alternative allele were considerably shorter than reads with the reference allele (ReadPosRankSum < −8); 3) SNPs with root mean square of the mapping quality lower than 40 (MQ < 40.0); 4) SNPs with evidence of strand bias (FS > 60.0 and SOR > 3.0); and 5) SNPs where the read with the alternative allele had a mapping quality that was lower than the reference allele (MQRankSumTest < −12.5). We used VCFTOOLS v0.1.15 (Danecek et al. 2011) to export variant sites from VCF files for bi-allelic sites with no missing data, a minimum read depth of 4x, a maximum read depth of 150x, and read quality score > Q20. We estimated the completeness of our mapped genomes with the Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.7.0 (Seppey et al. 2019) using the aves_odb10 data set.

Characterizing Genetic Variation

We estimated the fixation index (F_{ST}), the total divergence index between pairs of species (D_{xy}), and the nucleotide diversity index (π) for consecutive 50 SNP windows using the popgenWindow.py script (https://github.com/simonhmartin/genomics_general). To evaluate the heterozygosity in our mapped genomes, we estimated heterozygosity-rich regions, also known as runs of heterozygosity (ROHet), using the detectRUNS R package (Biscarini et al. 2018). For this analysis, we set the window size to 10 kb, and we only estimated the values in windows with a minimum of 20 SNPs. Because there are no modern genetic samples of *P. alexandrae* from the wild, we used genetic samples from captive birds in this phase of the study. If any of our captive samples had admixture due to hybridization in captivity over the last few generations, then we expected to find excess heterozygosity levels in comparison to wild individuals of the other species that do not have evidence for recent hybridization.

Population-level Sampling and Data Generation

To include broader population-level sampling, we used a sequence capture approach to target ultraconserved elements (UCEs; Faircloth et al. 2012), a widely used marker in avian phylogenomic studies, including our previous work on the focal system (Smith et al. 2023). We were particularly interested in including additional wild-collected 19th-century samples of *P. alexandrae* from well before it was a popular species in captivity. Because there are no modern tissue samples of wild *P. alexandrae*, we isolated DNA from wild-collected and captive-bred individuals from historical museum specimens. Our sample sizes for UCE sequencing are as follows (Supplementary Table S1): *Alisterus amboinensis* ($N = 1$); *Alisterus chloropterus* ($N = 1$); *Alisterus scapularis* ($N = 6$); *Aprosmictus erythropterus* ($N = 11$); *Aprosmictus jonquillaceus* ($N = 3$); *Polytelis alexandrae* ($N = 34$); *Polytelis anthopeplus* ($N = 8$); *Polytelis swainsonii* ($N = 5$); and an outgroup, *Melopsittacus undulatus* ($N = 1$). These UCE samples include one individual of each ingroup species from Smith et al. (2023). We followed the same wet-lab protocol and bioinformatic processing as Smith et al. (2023) unless noted otherwise.

DNA extractions from modern tissue samples were done using the standard protocol for the DNeasy tissue extraction kit (Qiagen, Valencia, CA). For historical samples from toe pads of museum specimens, we performed extractions on these samples in a clean molecular lab and modified the protocol to account for highly fragmented DNA. To size-select for shorter DNA fragments, we used the filter columns from the QIAquick PCR kit. DNA extracts were quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific). Library preparation of UCEs, enrichment, and Illumina (HiSeq X Ten) sequencing were performed by RAPiD Genomics (Gainesville, FL). The Tetrapod UCE 5K probe set was used to enrich 5,060 UCE loci (Faircloth et al. 2012).

De-multiplexed raw reads (FASTQ files) were trimmed and filtered for low-quality bases and adap-

TABLE 1. Genetic data sets and corresponding analyses.

Data set	*n*	Content	Analyses
Whole genomes	15	1076,869,200 sites	Quality control
Consecutive windows	5	295,503 windows	Summary statistics; IQTree2 Gene Trees
		295,503 gene trees	Astral; SNaQ; Aphid
		14,229 gene trees on Z-chromosome	Aphid
		2953 windows	BPP; D-statistics
Non-consecutive windows	5	38,662 windows	Summary statistics
		38,662 gene trees	Aphid; SNaQ; Aphid
		2347 gene trees on Z-chromosome	Aphid
		493 windows	BPP; D-statistics
Coding sequence		8440 genes	PAML
UCEs	72	2108,941 sites	IQTree2 concatenated topology
	5	3829 UCE loci	IQTree2 gene trees
		3829 gene trees	Aphid

Each window contained 50 SNPs. Consecutive windows: continuous sliding windows along the whole genome. Non-consecutive windows: non-continuous sliding windows with a sampling scheme of including one window and then skipping the next five. Coding sequence: single copy orthologous groups of protein coding sequence. UCEs: ultraconserved elements, and a reduced UCE data set with one individual per species was also analyzed. The complete whole genome data set consisted of 15 individuals and was partitioned into various subsets with one exemplar per the five focal species, and in some analyses the Z-chromosome was analyzed separately. The complete UCE data set contained 72 individuals and a subsampled data set with one individual per species.

tor sequences using Illumiprocessor v1 (Bolger et al. 2014). Next, in PHYLUCE (Faircloth 2015), we produced de novo assemblies of each individual using SPades (Bankevich et al. 2012) and created a chimera reference to maximize the number and length of UCE loci. Reads were aligned to the reference genome using BWA v0.7.13-r1126 (Li and Durbin 2009) with the *mem* algorithm. Resulting SAM files were converted to BAM format and sorted using SAMtools v1.10 (Li et al. 2009). Variant calling was performed using SAMtools *mpileup* (−C 30; −Q 20) and BCFtools v1.12 (Danecek et al. 2011). *Vcfutils* was used to retain only variant sites with a minimum coverage of 5× and a quality score of at least 20. FASTQ files were converted to FASTA format using seqtk (Li [n.d.]). Heterozygous sites were retained with IUPAC ambiguity codes. FASTA files were concatenated and processed using PHYLUCE (Faircloth 2015): loci with a minimum of 100 base pairs (bp) were aligned with MAFFT v.7.455 (Katoh and Standley 2013), and individual UCE alignments were retained if 75% of the samples were present.

Phylogenetic Inference

We inferred maximum likelihood phylogenies for the population-level UCEs and the whole genome data sets (Table 1) using IQ-TREE2 v.2.1.3 (Minh et al. 2020). For the UCEs, we used a concatenated alignment and estimated the best-fit substitution model for each gene partition using the ModelFinder option (Chernomor et al. 2016; Kalyaanamoorthy et al. 2017). Support values were assessed with 1000 ultrafast bootstraps for the concatenated tree (Hoang et al. 2018). For the whole genome data set, we randomly included one sample for each of the five focal taxa and estimated maximum likelihood gene trees for 50 SNP partitions in IQ-TREE2. We chose to estimate gene trees from 50 SNP windows instead of a fixed length to obtain enough signal to resolve phylogenetic relationships and to ensure that all windows have comparable genetic information and to capture signals of gene flow (Thom et al. 2024). To increase computational efficiency, HKY (Hasegawa, Kishino, and Yano; Hasegawa et al. 1985) was specified as the nucleotide substitution model for all windows. We estimated a species tree using the 295,503 gene trees in ASTRAL-III v.5.7.3 (Zhang et al. 2018), where node support was estimated using local posterior probabilities.

Inferring Species Relationships on a Network

We inferred a network using Species Networks applying Quartets (SNaQ; Solís-Lemus and Ané 2016) from the 295,503 gene trees estimated in IQ-TREE2 (Table 1) to determine whether the close sister relationship between *P. alexandrae* and *Ap. erythropterus* could be explained by introgression. Inheritance probabilities, describing the proportion of genomic regions inherited by a hybrid node from each of its parental lineages (γ), were estimated. SNaQ estimates a network, bifurcating or cyclic, by maximizing the pseudolikelihood for a set of gene trees. Using Topology 2 as the starting tree, we tested for the number of reticulation events ($h = 0$–5) and selected the value based on the pseudolikelihood score separately using two separate data sets of gene trees. One data set contained gene trees from consecutive 50 SNP windows across the genome. The second data set consisted of gene trees from non-consecutive windows

where one 50 SNP window was retained and next five were excluded.

To evaluate the confidence of the best network based on the pseudolikelihood, we performed a bootstrap analysis using 100 bootstrapped trees generated for each of the gene trees. Then, the bootsnaq function was used to select one of the bootstrapped trees to generate a data set from which this function estimates a network for each gene tree, for 10 runs and 100 replicates, resulting in a set of bootstrapped networks. Support values reflect how many times an introgression edge was in the bootstrapped network (Solís-Lemus et al. 2017). SNaQ analyses were conducted using PhyloNetworks (Solís-Lemus et al. 2017) in Julia (Bezanson et al. 2017).

Introgression Summary Statistics

The Dtrios program implemented in the D-suite package (Malinsky et al. 2021) was used to calculate D-statistics and the associated estimate of admixture fraction, *f*, referred to as the *f4*-ratio (Green et al. 2010; Durand et al. 2011; Martin et al. 2014). The D-statistic evaluates the frequency of two types of allele sites, ABBA and BABA, where A represents the ancestral allele and B the derived allele state in the context of four taxa analyses. Because our ingroup includes five species, we excluded two samples (*Al. scapularis* and *P. swainsonii*) that were redundant for the analysis, and we specified *P. anthopeplus* as P1, *P. alexandrae* as P2, *Ap. erythropterus* as P3, and the outgroup *M. undulatus* as O (P4). ABBA sites correspond to derived alleles shared between P1 and P3, whereas P2 and O share the ancestral allele (Patterson et al. 2012). If post-divergence gene flow between *P. alexandrae* and *Ap. erythropterus* occurred, we expect to find a higher proportion of BABA sites, where P2 and P3 share the derived allele and P1 and O share the ancestral allele, conflicting with the species tree. We also employed the *f-branch* metric, a metric that summarizes *f* scores (Malinsky et al. 2018). We estimated D-statistics and the *f*-branch metric for separate data sets (Table 1) as follows: 1) a global estimate across all windows; 2) estimates from 100 consecutive windows where each individual window contained 50 SNPs; and 3) estimates from 100 non-consecutive windows with a sampling scheme of "sample one window and exclude the next five," an arbitrary distance to reduce linkage among windows. The continuous and non-continuous sampling of windows allowed us to assess the impacts of linkage of sites. The 100 window partitions averaged 688,346 bp, corresponding with the genomic region used in the BPP analysis. This approach allowed us to examine gene flow patterns across the genome and to use comparable windows across the majority of analyses. We also examined whether using different individuals and non-consectutive windows impacted our inference of gene flow patterns using D-statistics.

Approximate Likelihood Estimates of Introgression from Gene Trees

We used the software Aphid v.0.11 to estimate gene flow from rooted triplets (Galtier 2024). Aphid distinguishes gene flow and ILS from gene trees using the topology and its branch lengths in an approximate likelihood framework. Gene tree conflict where the conflicting branch has a long branch is attributed to ILS, whereas a short branch is attributed to gene flow. The percentage of gene trees in conflict due to ILS and gene flow is reported, as well as the probability of phylogenetic conflict due to gene flow. We used ([*P. anthopeplus*, *P. alexandrae*], *Ap. erythropterus*) as the triplet rooted with *M. undulatus*. We ran Aphid on three separate data sets that comprised maximum likelihood gene trees of (Table 1): 1) 295,503 gene partitions across the autosomes and the Z chromosome; 2) 14,229 gene partitions across the Z chromosome; and 3) 4373 UCE loci that included a historical sample of a wild individual of *P. alexandrae.* Because Aphid analyses are conducted on the gene trees of single windows, we also present results similar to the noncontinuous window data sets by plotting one window and excluding the next five windows resulting in 49,200 across the whole genome and 2347 on only the Z-chromosome. The aim of conducting the analysis on only the UCE data set was to assess if we detected a different pattern using a wild caught individual. The autosomal and Z chromosome data sets were based on the same individuals using the single exemplar per taxon in the gene trees estimated in IQ-TREE2.

We used SnpEff (Cingolani et al. 2012) to summarize the *M. undulatus* genome annotation along each genomic window. Specifically, we calculated the proportion of each window assigned to introns, downstream or upstream of regulatory regions, exons, 3'-UTR, 5'-UTR, and intergenic regions. We classified the effect of each variant into four categories: 1) modifier, that is, variants with very low or unknown impact on protein function, often located in non-coding regions; 2) low impact, that is, variants in coding regions that do not change the protein function, such as synonymous mutations; 3) moderate impact, that is, variants that might change the amino acid sequence producing changes in protein effectiveness, but are unlikely to cause complete loss of protein functions; and 4) high impact, that is, variants that have a significant effect on protein function or gene structure, including a switch or loss of the function, generally located in coding or splicing regions. The aim of this approach was to determine whether gene trees in conflict with the species tree due to gene flow showed different frequencies in these categories compared with gene trees not in conflict. Under non-adaptive introgression, we do not expect to find differences between the mutational categories. In contrast, under adaptive introgression we expect to find a higher frequency of regions of the genome subject to positive selection (e.g., exons). We used a probability threshold of 0.70 from

the Aphid analyses for determining which regions were introgressed as per the recommendation of the software author. To assess whether there were significant differences in the number of SNPs located in different genomic regions determined by SnPeff, we performed χ^2 tests between introgressed versus non-introgressed windows for each category. Additionally, we calculated the effect size using Cohen's d to evaluate the potential effect of the large sample size of SNPs. These tests were performed using native libraries in R v4.2.1 (Team 2020).

Inferring Introgression Probabilities Using Full Likelihood

To further estimate patterns of gene flow, we used a full likelihood method in BPP v.4.7 (Flouri et al. 2018, 2020, 2023), allowing us to compare signatures of gene flow inferred from different methods (BPP; D-statistics; and Aphid). We estimated introgression probabilities between *P. alexandrae* and *Ap. erythropterus* using the A00 analysis. The A00 analysis performs a Bayesian analysis under the MSC with fixed species topology and fixed species assignment. Thus, the A00 analysis only estimates divergence times and population sizes (scaled by mutation rates). Using the MSC-M model, the tree topology was fixed to have a monophyletic *Polytelis* following Smith et al. (2024), where only one individual was used per taxon. We assigned gamma priors to $\theta \sim (2100)$ and $\tau 0 \sim \text{IG}(2, 10)$. We ran BPP independently on the previously described data sets (Table 1): 1) continuous windows over 100 consecutive windows, producing 2953 partitions; and 2) non-continuous windows where 100 windows were produced by sampling one window and excluding the next five windows producing 493 partitions. As with previous analysis, each individual window was delimited by having 50 SNPs The continuous 2953 and 493 non-contiguous partitions were run separately in BPP. Each partition was run for 500,000 generations, sampling every 2, and with a burn-in period of 250,000 generations. Partitions on the Z-chromosome were run separately with an inheritance scalar of 0.75.

We also used BPP to estimate the time to the most recent common ancestor (TMRCA) for *P. alexandrae* and *Ap. erythropterus* for each 50 SNP window along the genome. We were specifically interested in the depth of TMRCA values between introgressed and non-introgressed regions. Introgressed regions were determined using the Aphid gene flow probability threshold of 0.70. We used the A00 model and included only *P. alexandrae* and *Ap. erythropterus* in the analysis in which we were only interested in the TMRCA for these two taxa. We also wanted to avoid biasing ages due to gene tree heterogeneity among loci. We assigned gamma priors to $\tau \sim (21{,}000)$ and $\tau 0 \sim \text{IG}(22{,}000)$ and ran each window separately. Each partition was run for 10,000 generations, sampling every two, and with a burn-in period of 2000 generations.

Tests of Positive Selection

Introgressed loci can evolve neutrally or be subject to natural selection, with selection leaving different adaptive signatures on the genome at macro- versus micro-evolutionary scales. At micro-evolutionary scales, positive selection frequently operates via selective sweeps that produce regions of high differentiation and low diversity in linked sites (Pritchard et al. 2010), a pattern that is expected to erode over time due to new mutations and recombination. At macroevolutionary scales, positive selection is predicted to produce an excess of nonsynonymous substitutions in protein-coding genes (Yang and Bielawski 2000). Given the level of divergence between *P. alexandrae* and *Ap. erythropterus*, we expect signatures of molecular adaptations to be detectable in the ratio of nonsynonymous to synonymous substitutions in the coding sequences (CDSs). If introgression contributed to adaptive divergence at a macroevolutionary scale, we expect that introgressed loci will be subject to high rates of positive selection in *P. alexandrae* (e.g., Ma et al. 2019). We specifically test whether introgression led to adaptive versus non-adaptive divergence both cumulatively and at the individual locus level.

To predict the CDS regions for each species, we used Gene Model Mapper (GeMoMa) v1.8 (Keilwagen et al. 2019), a software package that uses a homology-based approach. As references for the CDS prediction, we used three bird genomes (*Melopsittacus undulatus*: GCF_012275295.1; *Strigops habroptilus*: GCF_004027225.2; and *Gallus gallus*: CF_016699485.2) that were annotated with RNAseq data.

We conducted a number of tests of positive selection for each *P. alexandrae* gene. To identify orthologous groups across Psittaculidae, we used the protein sequences of all five Polytelini species included in our study (*P. alexandrae*, *P. swainsonii*, *P. anthopeplus*, *Ap. erythropterus*, and *Al. scapularis*) plus *M. undulatus* as input for OrthoFinder 2.5.5 (Emms and Kelly 2019). We identified 8440 orthologous groups with a single copy in each species found on chromosomes 1–30, including Z-chromosome (Table 1). For each of these groups, we aligned the protein sequences using PRANK v.170427 (Löytynoja 2014). The resulting alignments were filtered in three ways: 1) we used Gblocks 0.91b (Castresana 2000) to remove poorly alignable regions; 2) we used a sliding-window approach to identify and remove regions of 15 amino acids for which one of the sequences presented 10 or more singleton amino acids (not present in any of the other five sequences); and 3) we used a similar approach to remove regions of five amino acids for which one of the sequences presented five singleton amino acids. The original CDS sequences, the protein alignments, and the results of the filtering were used to generate CDS alignments.

Each CDS alignment was subjected to a number of analyses using the codeML program in the PAML pack-

age 4.4d (Yang 2007). First, we used the M0 model to calculate a single d_N, d_S, and d_N/d_S ratio for each gene (in this analysis, d_N/d_S is assumed to be constant across all sites and branches). Second, we used the free-ratios model to calculate a separate d_N, d_S, and d_N/d_S ratio for each gene and branch in the phylogeny. Third, we tested for the presence of sites under positive selection in each gene by comparing the fit of models M2a and M1a using a likelihood ratio test—twice the difference in the likelihood of both models, $2\Delta\ell = 2 \times (\ell_{M8} - \ell_{M7})$ was assumed to follow a χ^2 distribution with two degrees of freedom (Whelan and Goldman 1999). Model 2a allows for the presence of a class of codons with $d_N/d_S > 1$ (indicative of the action of positive selection), whereas model 1a does not; thus, a significantly better fit of model 2a indicates that the gene has been subjected to positive selection. Fourth, for each gene and branch (8440 genes × 9 branches), we tested for the presence of sites under positive selection by comparing the fit of model A (MA) and null model A1 (MA1). Model A allows for the presence of a class of codons with $d_N/d_S > 1$ whereas null model A1 does not; thus, a significantly better fit of model A indicates that the gene has been subjected to positive selection. In these comparisons, $2\Delta\ell = 2 \times (\ell_{MA} - \ell_{MA1})$ was assumed to follow a 50%:50% mixture of a point of mass 0 and a χ^2 distribution with one degree of freedom (Yang et al. 2005; Zhang et al. 2005). For genes in regions predicted to be introgressed by Aphid, we used a tree in which *P. alexandrae* was the sister taxon to *Ap. erythropterus*. For all other genes, we used a tree in which *P. alexandrae* was the sister group to the *P. swainsonii*/*P. anthopeplus* clade.

Results

Whole Genome Characterization

The whole genome data set, including *Melopsittacus undulatus*, contained 14,785,561 SNPs with 10,505,270 transitions and 4280,291 transversions. Statistics for the genomes of each individual are reported in Supplementary Table S2. The average coverage and standard deviation and BUSCO scores for each species were *Alisterus scapularis* (41x; 9.4; 94.4%), *Aprosmictus erythropterus* (35.9x; 0.2; 94.4%), *Polytelis alexandrae* (26.6x; 15.9; 92%), *Polytelis anthopeplus* (20.2x; 13.3; 92.9%), and *Polytelis swainsonii* (29.8x; 7.6; 93.6%). We mapped reads up to chromosome 30, including the Z-chromosome. Reads mapped to microchromosomes above chr 21 and the W-chromosome, including unplaced scaffolds, were often short regions and were discarded in most downstream analyses.

The five focal taxa had similar patterns of genetic diversity as measured by F_{ST} (Supplementary Table S3), D_{XY} (Supplementary Table S4), and runs of heterozygosity except for one individual of *P. alexandrae* (Supplementary Tables S5–S6). F_{ST} values varied across the genome (Supplementary Fig. S2). The sample ale_ZMUC140801 had a higher number of runs of heterozygosity (35 vs. 2–9 for the other samples), possibly due to its lower coverage. This sample was removed, and re-calculated summary statistics were similar across species. For downstream analyses, which entailed using a single exemplar per taxon, we used ale_PRS4773 for *P. alexandrae*. *Polytelis alexandrae* had the highest π value (0.077), and *P. anthopeplus* the lowest (0.034; Supplementary Table S7).

UCE Phylogeny

The concatenated alignment contained 72 individuals, 2108,941 sites, 21,590 parsimony-informative sites, 11,743 singleton sites, and 2075,608 invariable sites. The concatenated tree had high support (ultrafast bootstraps = 100%) for relationships among *Polytelis* and *Aprosmictus* (Fig. 1d). *Polytelis alexandrae* and *Aprosmictus* were sisters, which in turn were sister to the sister taxa *P. anthopeplus* and *P. swainsonii* (Topology 2). All species were monophyletic. Evidence of phylogeographic structuring was observed between *Ap. erythropterus* samples from Australia and New Guinea. Within *P. anthopeplus*, however, and pending further sampling and closer analysis, we tentatively note that no phylogeographic structure was evident between its disjunct southeastern and southwestern Australian populations. Structuring among the *P. alexandrae* appeared to reflect mainly differences in information content of samples (Fig. 1b; Supplementary Fig. S3). Captive and wild-caught individuals did not exhibit a different phylogenetic pattern. The three individuals for which we sequenced genomes formed a well-supported clade.

Phylogenetic Tree and Network Estimation

The ASTRAL species tree (Fig. 1c) constructed using ML gene trees shows discordance with the concatenated tree estimated in previous studies (Smith et al. 2023; Smith et al. 2024), specifically for the placement of *P. alexandrae* (Fig. 1b). Different triplet topologies were found throughout and among chromosomes (Fig. 2a). Topology 2 was the most frequent (37.6% of the trees), followed by Topology 1 (25.4%), then an alternative minor topology where *P. anthopeplus* and *Ap. erythropterus* were sister (20.6%), and 16.4% of trees had an unresolved topology. The ASTRAL tree places *P. alexandrae* as sister to *Ap. erythropterus* (Topology 2) whereas the previously estimated concatenated tree places it as sister of the clade formed by the other two species of *Polytelis* (Topology 1).

When inferring relationships using a network approach, pseudolikelihood scores from SNaQ found that one hybridization event was optimal for the network (-log pseudolikelihood = 0.2039 (consecutive windows) and 0.0254 (non-consecutive windows); Supplementary Table S8). For simplicity, we will lead with results from the consecutive window data set and summarize at the

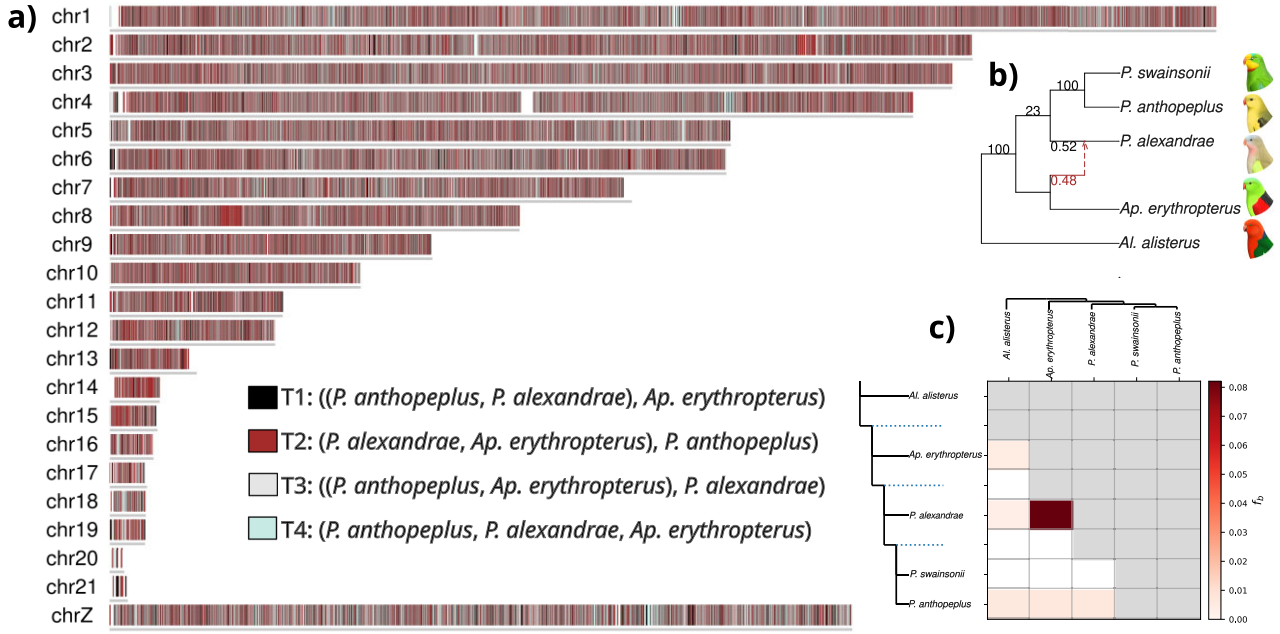

FIGURE 2. Topological variation across the genome, network relationships, and introgression proportions. (a) Each color corresponds with a triplet topology. Topology 1 (T1; 25.4% of trees) has a monophyletic *Polytelis*. Topology 2 (T2; 37.6%) *P. alexandrae* and *Ap. erythropterus* are sister. Topology 3 (T3; 20.6%) *P. anthopeplus* and *Ap. erythropterus* are sister. 16.4% of the topologies were unresolved, Topology 4 (T4). (b) The best-fit phylogenetic network shows a hybrid edge connecting *P. alexandrae* and *Ap. erythropterus*. The proportion of the genome contributed (γ) to *P. alexandrae* from *Ap. erythropterus* was 0.48. (c) The *f*-branch matrix shows pairwise introgression proportions from gene tree counts, with darker colors indicating higher introgression. The tree to the left of the matrix shows ancestral branches among species, and the tree on the top shows the expected relationships among species. Bird illustrations by Julian Teh.

end how the non-consecutive window data set differed. In the best-fit network, a proportion of the genome was contributed to *P. alexandrae* from *Ap. erythropterus* (γ = 0.48). After placing the introgression event and rooting the network using *M. undulatus*, the network topology matches Topology 1, that is, monophyly of *Polytelis* was restored, and *P. alexandrae* was sister to *P. anthopeplus* and *P. swainsonii* (Fig. 2b). The network had highly supported values for all nodes except for the placement of *P. alexandrae* as sister to *P. anthopeplus* and *P. swainsonii* (bootstrap = 0.23). The bootstrap values of the hybrid edge averaged 2% across the entire genome and ranged from 0% to 10% at the chromosome level (Supplementary Table S9; Supplementary Fig. S4). These bootstrap values reflect the number of times in which a network edge, including hybridization edges appears in the set bootstrapped trees. In the non-consecutive window data set, the best-fit network had the sample relationships as the consecutive window network except the hybridization edge was from a ghost lineage into *Ap. erythropterus* (Supplementary Table S9; Supplementary Fig. S5).

Introgression Statistics

Introgression tests using a genome-wide estimate of the D-statistic found significant evidence of gene flow between *Ap. erythropterus* and *P. alexandrae*. Results were similar when either *P. anthopeplus* (D-stat = 0.18; *z*-score = 80.47; *P*-value = 2.3×10^{-16}) or *P. swainsonii* (D-stat = 0.19; *z*-score = 88.32; *P*-value = 2.3×10^{-16}) was used as P1 in the triplet. Window-based estimates of D-statistics showed high variation in *z*-scores across the genome (Consecutive windows mean = 0.173; standard deviation = 0.094; range = 0.000–0.806; % of loci with *z*-score > 2.5 = 47%; Non-consecutive windows mean = 0.189; standard deviation = 0.090; range = 0.001–0.656; % of loci with *z*-score > 2.5 = 53%). Using the *f*-4 ratio, which was also similar across data sets, we estimated the introgression proportion between *Ap. erythropterus* and *P. alexandrae* to be ~8%, whereas all other comparisons were < 1% (Supplementary Table S10). Using Topology 1, the *f*-branch statistic indicated that the highest introgression in the clade was between *P. alexandrae* and *Ap. erythropterus*. D-statistic and *f*-branch statistic results were robust to which individual we used (Supplementary Fig. S6).

Approximate Likelihood Estimates of Introgression

The Aphid analyses show that the high frequency of Topology 2 across the genome was due to both ILS and introgression between *P. alexandrae* and *Ap. erythropterus* (Supplementary Table S11). This pattern was observed

in all data sets except for the Z-chromosome, which showed lower topological heterogeneity and reduced gene tree conflict due to gene flow.

For the consecutive window data set across the whole genome, the most frequent gene tree (87,100 gene trees; 37.6%) was ([*P. alexandrae*, *Ap. erythropterus*], *P. anthopeplus*), consistent with the species tree topology (Topology 2; Fig. 2a). The next frequent topology (58,876 gene trees; 25.4%) was ([*P. anthopeplus*, *P. alexandrae*], *Ap. erythropterus*), consistent with *Polytelis* being monophyletic (Topology 1; Fig. 2a). The other two topologies were ([*P. anthopeplus*, *Ap. erythropterus*], *P. alexandrae*) and (*P. anthopeplus*, *Ap. erythropterus*, *P. alexandrae*), representing 20.6% (47,703) and 16.4% (37,861) of gene trees, respectively. Posterior estimates of no-event = 0.001, no-conflict ILS = 0.257, and no-conflict gene flow = 0.04. The posterior estimate of conflict due to ILS was 0.484, and for gene flow it was 0.217. The posterior imbalance for the dominant topology for ILS was 0.506 and 0.971 for gene flow. Of the gene flow-conflicted loci with threshold values > 0.70, there were 8120 triplets consistent with Topology 2 ([*P. alexandrae*, *Ap. erythropterus*], *P. anthopeplus*), 560 with Topology 1 ([*P. anthopeplus*, *P. alexandrae*], *Ap. erythropterus*), and 120 with the alternative triplet ([*P. anthopeplus*, *Ap. erythropterus*], *P. alexandrae*).

Of the 14,229 gene trees on the Z-chromosome, 31.4% (3082 gene trees) supported Topology 2 ([*P. alexandrae*, *Ap. erythropterus*], *P. anthopeplus*) and 25.7% (2526 gene trees) supported Topology 1 ([*P. anthopeplus*, *P. alexandrae*], *Ap. erythropterus*). The two other triplets, ([*P. anthopeplus*, *P. alexandrae*], *Ap. erythropterus*) and (*P. anthopeplus*, *P. alexandrae*, *Ap. erythropterus*), were supported by 22.1% (2169 gene trees) and 20.8% (2044 gene trees) of the genes, respectively. Posterior estimates were of no-event = 0.003, no-conflict ILS = 0.317, and no-conflict gene flow = 0.002. The posterior estimate of conflict due to ILS was 0.656, and for gene flow it was 0.022. The posterior imbalance for the dominant topology for ILS was 0.506 and 0.971 for gene flow.

The non-consecutive and consecutive window data sets showed similar tree topology frequencies and posterior estimates (Supplementary Table S11). The Z-chromosome non-consecutive window data set showed higher gene tree conflict due to gene flow than the consecutive window (0.116 vs. 0.022), an estimate still lower than that of the whole genome.

For the 3829 UCE loci, where we used a single exemplar of the focal taxa, including a wild-collected *P. alexandrae* from a historical museum specimen, similar topological patterns were observed except that the most dominant topology was the unresolved triplet (*P. anthopeplus*, *P. alexandrae*, *Ap. erythropterus*), supported by 65.6% of the genes (1653 gene trees). Topology 2 ([*P. alexandrae*, *Ap. erythropterus*], *P. anthopeplus*) was the next most frequent triplet at 16.1% (405 gene trees), followed by the triplet matching Topology 1 ([*P. anthopeplus*, *P. alexandrae*], *Ap. erythropterus*) at 9.4% (237 gene trees). The remaining triplet ([*P. anthopeplus*, *P. alexandrae*], *Ap. erythropterus*) was observed in 224 gene trees (8.9%). Posterior estimates were of no-event = 0.001, no-conflict ILS = 0.241, and no-conflict gene flow = 0.040. The posterior estimate of conflict due to ILS was 0.486, and for gene flow it was 0.232. The posterior imbalance for the dominant topology for ILS was 0.561 and 0.842 for gene flow.

Full-Likelihood Introgression Probabilities

We found high heterogeneity in introgression probabilities among windows but in general, along each chromosome there were windows with high and low introgression probabilities (Supplementary Table S13). The average introgression probability for each chromosome ranged from 0.52 to 0.85, the small chromosomes (Chr20: 0.54; Chr21: 0.52) and Z-chromosome (0.66) having the lowest values, and other small chromosomes (Chr18: 0.85; Chr19: 0.82) having the highest. Larger chromosomes tended to have lower introgression probabilities than smaller chromosomes in both average values and minimum introgression probabilities. All chromosomes had regions with introgression probabilities > 0.95 except Chr18 and Chr19, which consisted of a single BPP run. For the BPP analyses on non-consecutive windows, we found similar introgression probabilities for chromosomes 1–9, 15, 18, and 20–21, and often higher values for the other chromosomes than those estimated for consecutive windows (Supplementary Table⟨?PMU ?⟩ S13).

We found that τ estimates of the TMRCA for *P. alexandrae* and *Ap. erythropterus* for each 50 SNP window were significantly different between introgressed and non-introgressed windows (Mann–Whitney *U* test: $W = 4.7 \times 10^8$; *P*-value $< 2.2 \times 10^{-16}$; Fig. 3a–c). Mean τ values were lower for introgressed (0.0011) than for non-introgressed (0.0016) regions, as were the ranges of τ values (introgressed: 0.0001–0.0060; non-introgressed: 0.0001–0.0073; Fig. 3a–c). The distribution of τ estimates had the lowest values for windows that introgressed and had Topology 2 (T2), indicating that the coalescent times were more younger than other sections of the genome.

Comparing Introgression Analyses

Comparisons across the introgression analysis (BPP, D-statistics, and Aphid) are limited because thresholds for delimiting strong support for introgression are not directly comparable (e.g., *P*-values vs. posterior probabilities). We present Manhattan plots of the results of the three methods used to estimate introgression to show how gene flow inferences vary across methods and the comparability of findings (Fig. 4a–c; Supplementary Fig. S7). Instead, we estimated the frequency at which re-

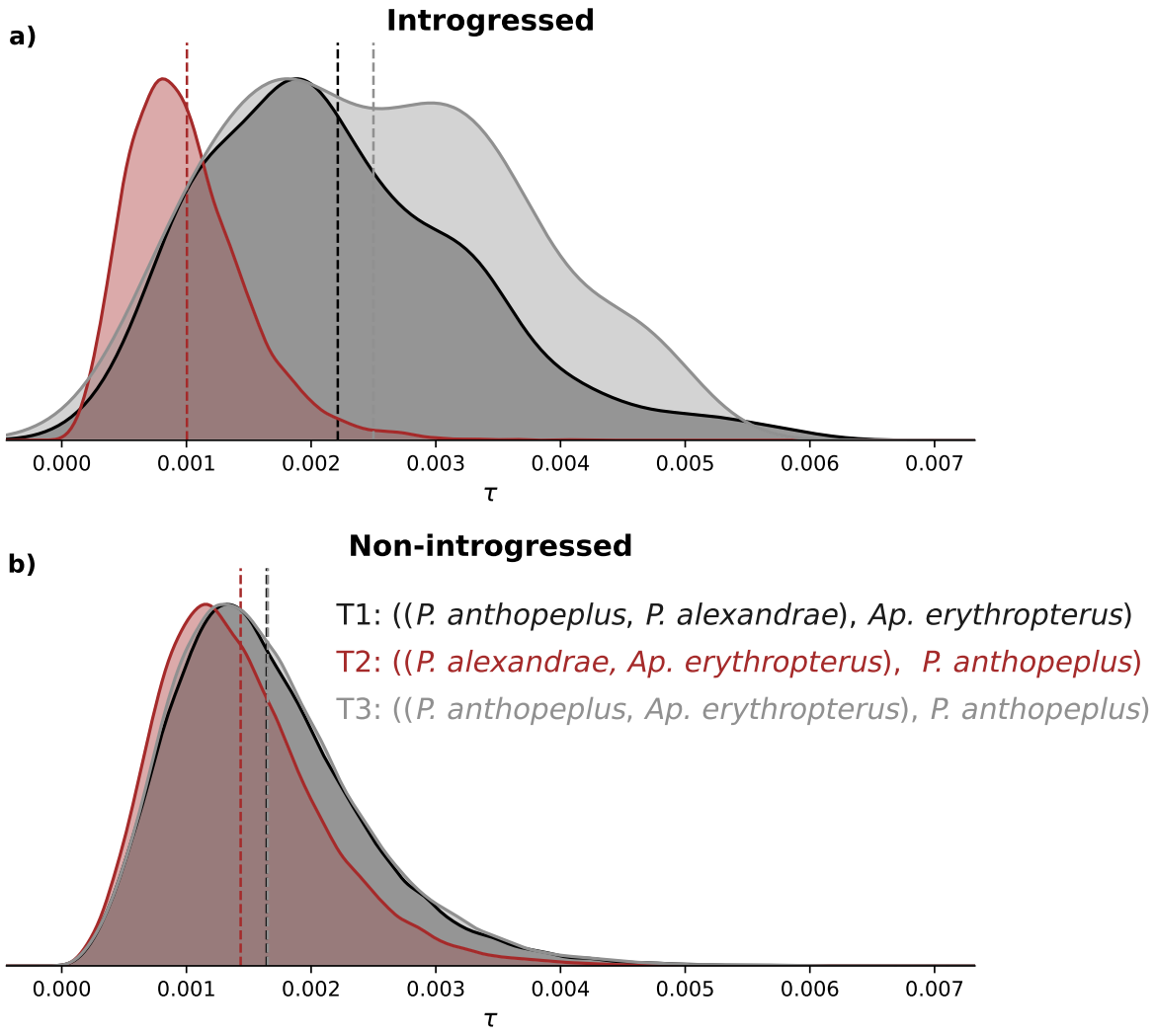

FIGURE 3. Histograms of τ estimates for *Ap. erythropterus* and *P. alexandrae from* introgressed (a) and non-introgressed (b) windows. The distribution of windows that were inferred to be introgressed and had Topology 2 (T2) had the lowest values. Each plot contains distributions of τ values for windows with Topology 1 (T1), where *Polytelis* is monophyletic, Topology 2 (T2), where *Ap. erythropterus* and *P. alexandrae* are sister, or an alternative topology, Topology 3 (T3), where *Ap. erythropterus* and *P. anthopeplus* are sister. Windows that had significant gene tree conflict due to introgression are plotted (left) separately from those not inferred to be introgressed (right). Each point from the distribution was estimated from a 50 SNP window. Introgressed windows had gene flow probabilities ≥ 0.70 from Aphid analyses. Dotted lines indicate mean values of each distribution.

gions significantly inferred to have gene flow were identified in multiple analyses. BPP, D-statistics, and Aphid similarly identified regions of the genome with a high probability of introgression. The D-statistics identified significant gene flow in the most windows and Aphid the least. From the D-statistic analysis, 1392/2953 (47%) with z-scores $\geq$ 2.5; Aphid, 8157/231541 (3.5%) windows probabilities ≥ 0.70; and BPP, 318/2953 (8%) windows had high posterior probabilities (≥ 0.95). Comparability among these analyses was limited because the BPP analysis was conducted on multiple windows where Aphid provides a global estimate and for individual gene trees. 178/318 (56%) windows estimated with posterior probabilities with ≥ 0.95 in BPP also had D-statistic z-scores $\geq$ 2.5. 15% (1224/2953) of windows estimated by Aphid to have gene flow (Probability ≥ 0.70) also had z-scores $\geq$ 2.5. The 8157 high probability windows inferred from Aphid were found on 270/318 (85%) of the high probability introgression windows in BPP and 1332/1392 (96%) on windows with z-scores $\geq$ 2.5.

Introgressed Windows and Positive Selection

Mutational categories of all windows were similar between introgressed and non-introgressed regions: high-impact mutations were the rarest (~1%), and low-impact mutations were the most frequent (Figs. 5a and

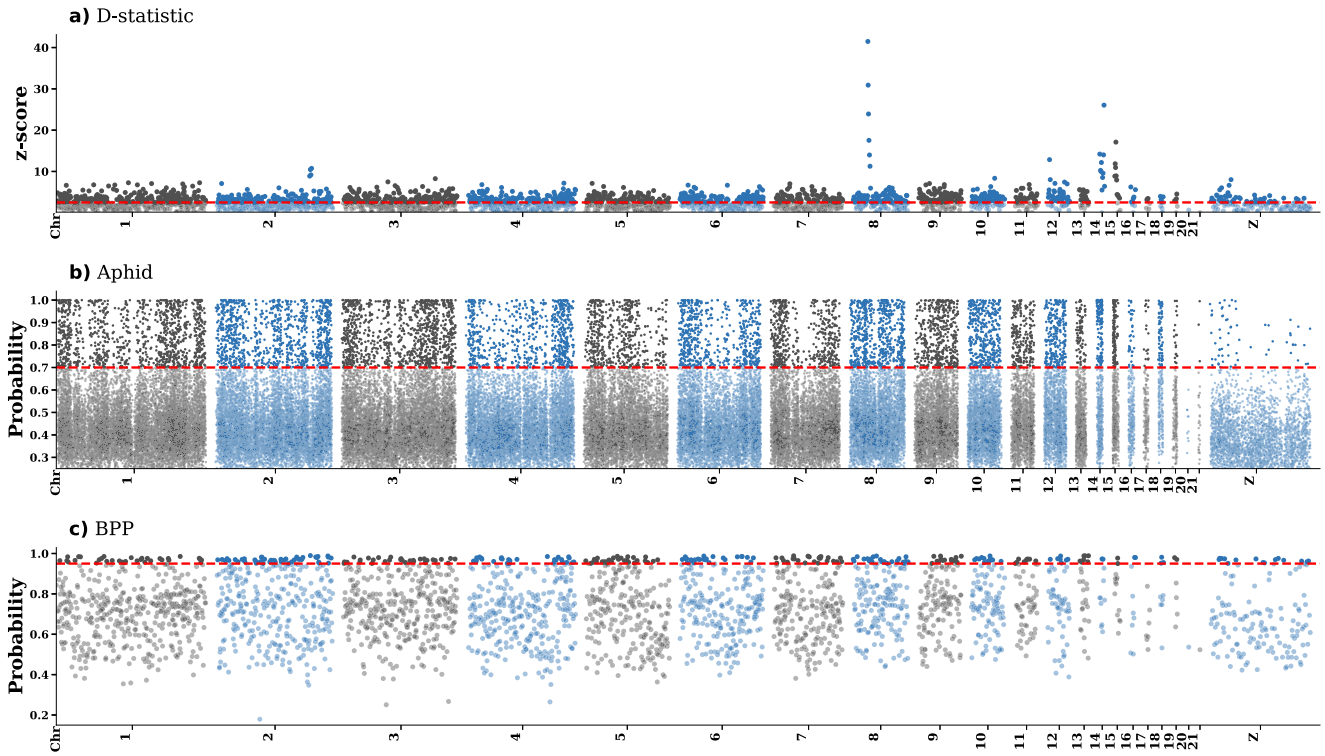

FIGURE 4. Manhattan plots showing different measures of introgression across the genome from consecutive windows. (a) z-scores for the D-statistic where values > 2.5 indicate significant departures from 0 indicating introgression between *Polytelis alexandrae* and *Aprosmictus erythropterus*. (b) Probabilities from Aphid of gene tree conflict where probabilities above 0.7 were considered evidence of gene flow. (c) Introgression probabilities from BPP for 100-loci windows. Dotted red lines denote thresholds of significance where points above the dotted line are deemed to have strong statistical support for being introgressed: (a) z-scores ≥ 2.5; (b) probabilities ≥ 0.70; and (c) posterior probabilities ≥ 0.95.

b). χ^2 tests showed that there were significant differences within each annotation category (P-value < 0.05) except for 3′-UTR (P-value $= 0.24$). The effect sizes were low for all comparisons (Supplementary Table S12). The largest effect sizes as determined by Cohen's d were within 5′-UTR (0.3913), intergenic regions (-0.3128), and exons (0.2660). Combined, regions located downstream and upstream of regulatory regions and exons represented less than 10% of regions in the gene trees. Although some differences were detected in mutational categories, there were clear trends in introgressed versus non-introgressed windows.

A test of purifying selection was done on each of the 8440 protein-coding genes for which we could identify a single ortholog in *P. alexandrae*, *P. swainsonii*, *P. anthopeplus*, *Ap. erythropterus*, *Al. scapularis*, and *M. undulatus*. Out of these genes, 237 were inferred to be introgressed based on our Aphid analysis (probability threshold ≥ 0.70), 5755 were inferred to be non-introgressed. A further 2448 were categorized as of unknown introgression state because Aphid failed to calculate their gene flow probabilities due to unexpected branch lengths. For each gene, we first computed a single d_N/d_S ratio for all codons and branches using PAML's M0 model. The median d_N/d_S was 0.141, indicating that strong purifying selection acted on most genes. The d_N/d_S ratios of introgressed genes were significantly lower than the d_N/d_S ratios of non-introgressed genes (median d_N/d_S for introgressed genes $= 0.137$; median d_N/d_S for non-introgressed genes $= 0.151$; Mann-Whitney's U test, P-value $= 0.041$; Supplementary Table S14), indicating that introgressed genes evolved under stronger purifying selection. We next used the free-ratios model to compute a separate d_N/d_S ratio for each gene and branch. Results for the *P. alexandrae* branch also indicated a predominance of purifying selection, and that introgressed genes (median $d_N/d_S = 0.137$) were under stronger purifying selection than non-introgressed genes (median $d_N/d_S = 0.151$; Mann-Whitney's U test, P-value $= 0.034$; Supplementary Table S14).

We then conducted tests of positive selection on each gene. Comparison of models M2a and M1a, which allows testing for the presence of a fraction of codons under positive selection without focusing on any specific branch, identified signatures of positive selection in 583 genes (P-value < 0.05). Our branch-and-site tests of positive selection in the *P. alexandrae* branch

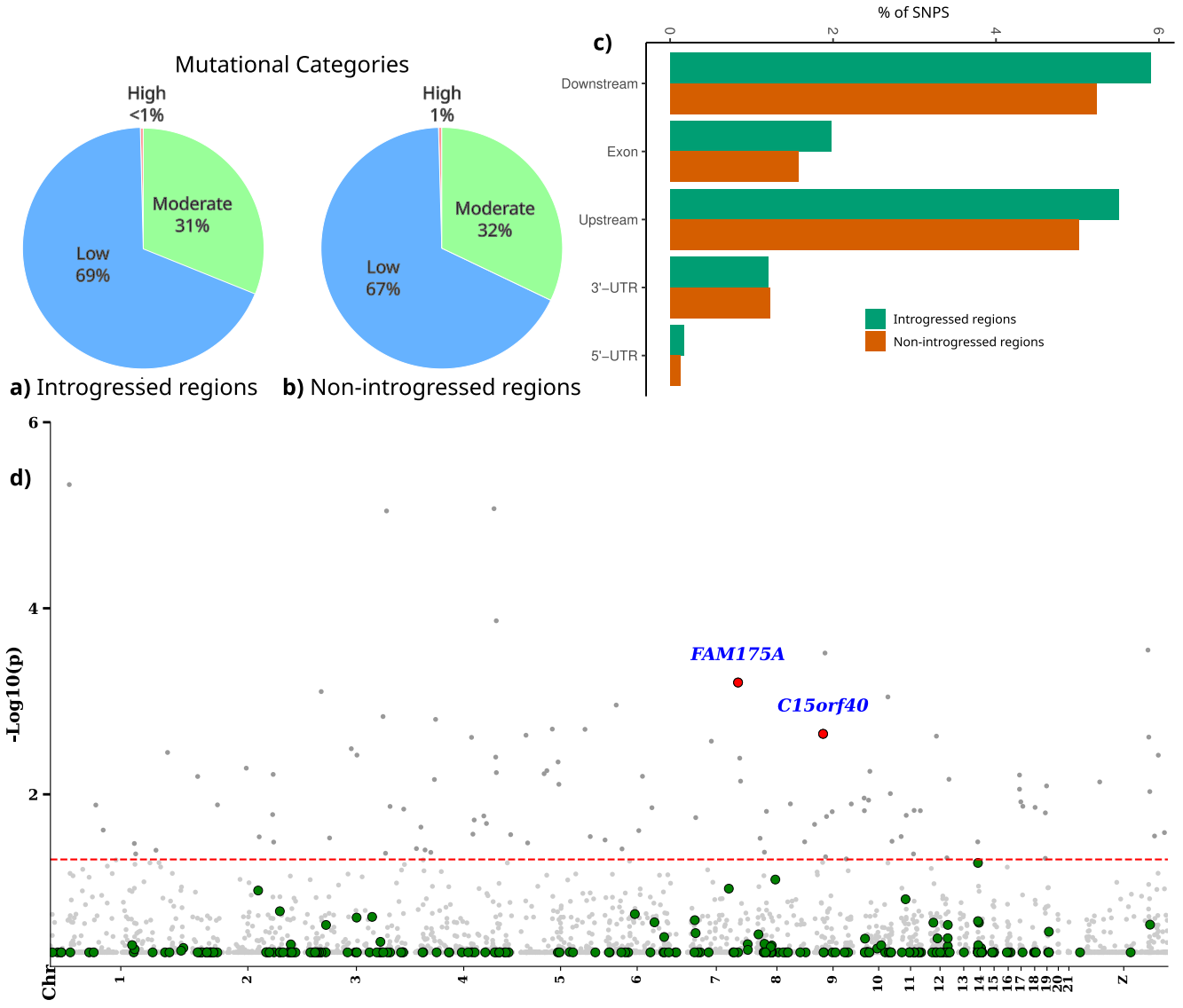

FIGURE 5. Categorization of introgressed regions. Patterns of mutational categories for introgressed (a) and non-introgressed regions (b) are similar. Mutational categories: (1) low impact variants are in coding regions that do not change the protein function; (2) moderate impact variants are those that might change the amino acid sequence producing changes in protein effectiveness but are unlikely to cause complete loss of protein functions; and (3) high impact variants have a significant effect on protein function or gene structure, including the switch or loss of the function. High impact mutations were the rarest, and low impact ones were the most common. (c) The percentage of introgressed versus non-introgressed regions differed within and among variant functional categories. (d) Manhattan plot showing log *P*-values for tests of positive selection on the *P. alexandrae* lineage. The dotted line represents a *P*-value of 0.05, and all points above the line are considered as having statistical support for selection. Introgressed regions are shown with larger points and non-introgressed regions are smaller. The majority of introgressed regions were not under positive selection. Only two genes were found to be both under positive selection and introgressed (*FAM175A* and *C15orf40*). A non-introgressed region with a log *P*-value of 8.7 on chromosome 8 was removed from the plot to reduce whitespace.

identified signatures of positive selection in 146 genes (*P*-value < 0.05), including two introgressed genes (*FAM175A* and *C15orf40*; Fig. 5d, Table 2, Supplementary Table S14) and 99 non-introgressed genes. A Fisher's exact test showed that introgressed and non-introgressed genes did not significantly differ in the fraction of genes under positive selection (*P*-value = 0.440).

Discussion

We found that the non-monophyly of *Polytelis* was caused by recovery of a close sister-group relationship between the highly divergent species *Aprosmictus erythropterus* and *Polytelis alexandrae* and that this was due to ancient introgression. Thus, *Polytelis alexandrae* should remain in *Polytelis*. Phylogenetic placement of *P. alexan-*

Table 2. Introgressed regions under positive selection.

Gene symbol	Gene name	Gene function	Chr	BPP Post Prob	Aphid Prob	*z*-score *P*-value
FAM175A	Family with sequence similarity 175 Member A	Cell cycle regulation	7	0.80	0.81	0.031
C15orf40	Chromosome 15 open reading frame 40	Cell signaling	9	0.68	0.76	0.002

Shown are the loci in windows that were both under positive selection in *P. alexandrae* and introgressed from *Ap. erythropterus*. Selection tests were conducted using the species tree topology (Topology 2). Included is the gene symbol, gene name, function based on the chicken genome, chromosome (Chr), introgression posterior probability from BPP (BPP Post Prob), the probability of gene tree conflict due to gene flow from Aphid (Aphid Prob), and the *P*-value of *z*-scores from D-statistics (*z*-score *P*-value).

drae outside its nominal genus *Polytelis* has been unstable across data sets and inference methods (Schweizer et al. 2010; Schweizer et al. 2011; Provost et al. 2018; Smith et al. 2023, 2024), and our results show that this was due to the predominance of gene trees containing, unexpectedly, *Ap. erythropterus* and *P. alexandrae* as sister taxa. We found that up to 20% of the gene trees were in conflict due to gene flow (Aphid analyses, Supplementary Table S11), with allele frequency patterns indicating that 8–20% of *P. alexandrae*'s genome was inherited from *Ap. erythropterus*. These findings favor the retention of *P. alexandrae* in the genus *Polytelis* and highlight how paraphyletic groups may be due to introgression and not taxonomic misclassification. Much of the genetic exchange did not lead to detectable molecular adaptations over time because 1% of protein-coding genes had significant evidence of positive selection in *P. alexandrae*. Further, of those 146 genes (Supplementary Table S14), only two had a high probability of being introgressed from *Ap. erythropterus* (Table 2). For ancient introgression scenarios, such as the one between *Ap. erythropterus* and *P. alexandrae*, where there is a high degree of phenotypic and ecological divergence and absence of introgressed traits, genetic exchange may produce high levels of topological heterogeneity while not being subject to strong levels of selection. This suggests that these introgressed regions are likely not contributing to an adaptation nor reproductive isolation.

Genome-wide Introgression

We employed a range of methods that varied in input data, resolution, and statistical complexity, consistently finding evidence for ancient introgression. Recent work has shown that methods using allele count patterns suffer from power issues when there is substitution rate variation (Frankel and Ané 2023), and full likelihood approaches for estimating introgression have been shown to improve accuracy (Ji et al. 2022). However, we found that our results were largely stable across methods, samples, and the degree of linkage in data sets, but the extent of inferred gene flow varied among software packages. It is unclear how our results may have been biased by the finding that the major topology across the genome did not match the "true" phylogeny. Even though the most frequent gene tree topology placed *P. alexandrae* as sister to *Ap. erythropterus*, the majority of discordance observed in individual gene trees was attributed to ILS for the majority of genomic windows, whereas a smaller proportion of discordant gene trees could be attributed to ancient introgression. The stability in detecting ancient introgression across methods may reflect that we examined a single introgression event among highly divergent lineages. Significant *z*-scores (*P*-values ≤ 0.05) from allele frequency counts and gene tree conflict due to gene flow (Aphid; Prob. ≥ 0.70) were found in the majority of windows with a high probability of introgression in the full likelihood method, BPP (posterior probability ≥ 0.95). A key distinction between the software packages was the computation time for running our data sets. BPP (~5 days on 30 CPUs) was more computationally intensive than D-suite (1–2 days 7 CPUs), followed by Aphid (~12 hours on 40 CPUs, excluding the estimation of input gene trees). SNaQ, the network based method that we used, was the most computationally intensive, particularly during bootstrapping (~10 days on 40 cores) and detected the highest proportion of introgression. The network had a hybrid edge between *P. erythropterus* and *P. alexandrae*, and the former contributed 48% to the latter (Fig. 2), a value we found varied across runs and whose support for this relationship was low. It is unclear how generalizable our results are to other empirical systems. We could assume, however, that taxa with diversification histories similar to the long-tailed parrots will show stable introgression signals across methods even if the percentage of the genome found to be introgressed varies.

Gene flow was not localized to particular regions of the genome, being widespread across chromosomes, a pattern that has emerged in other systems. For example, *Anopheles* mosquitoes (Fontaine et al. 2015), *Lonchura* munias (Stryjewski and Sorenson 2017), and *Phylloscopus* leaf warblers (Zhang et al. 2021) all show pervasive gene flow across the genome. In Amazonian birds, introgression was shown to be genome-wide but more frequent in high recombination regions and on microchromosomes (Thom et al. 2024). These examples speciated sometime in the last three million years and so differ from our results in the degree of divergence be-

tween taxa. Gene flow in these examples is presumably well before major genetic incompatibilities evolved. In contrast, previous molecular dating has placed the split between *Ap. erythropterus* and *P. alexandrae* at 4.7–10.3 Mya (Smith et al. 2024). Our estimates of introgressed regions indicate that gene flow occurred ∼1–2 million years ago (Fig. 3; τ of 0.001 corresponds to 1 million years, assuming a generic substitution rate of 1×10^{-9} substitutions/site/year), which is well after the estimated divergence between *Ap. erythropterus* and *P. alexandrae.* It is worth pointing out that the Aphid software distinguishes gene tree conflict due to ILS versus gene flow due to the depth of the node, where younger nodes are considered to be due to gene flow. This approach necessarily biases toward younger ages. Examples of speciation-with-gene flow often report genetic introgression during the early stages of speciation (Wang et al. 2020). The literature also has numerous cases of ancient introgression detected deeper in phylogenies (e.g., Fontaine et al. 2015; Liu et al. 2022; Suvorov et al. 2022; Zhou et al. 2022; DeBaun et al. 2023), indicating that phylogenetic signal due to introgression persists over long time scales, even if occurring early in the speciation process. In some cases, signals of introgression appear not to persist (Gardner et al. 2023). The relationship between the origin, extent, and persistence of introgressed signals across phylogenetic scales warrants further inquiry.

One unique finding among our introgression tests was the lower introgression probabilities found on the Z-chromosome (Fig. 4; Supplementary Table S11). Gene flow probabilities were much lower on the sex chromosome, and the frequency of triplet topologies was more uniform than on the autosomes. The Z-chromosome also had one of the lowest introgression probabilities according to BPP (Fig. 4c). This pattern is consistent with Haldane's Rule (Haldane 1922; Muller 1940, 1942; Dobzhansky 1982), which indicates that the lower fitness of hybrids of the heterogametic sex will result in lower introgression of sex chromosomes. Reduced introgression of the Z-chromosome is consistently found in avian hybridization studies (Ottenburghs 2022), and our results provide evidence that this also occurred in *P. alexandrae*. Interestingly, the species tree for the Z-chromosome (not presented) still supported the *P. alexandrae* and *Ap. erythropterus* sister relationship. This is presumably because the gene trees with this relationship were informative, and these gene trees reflected an unexpected topology due to rampant ILS. On autosomal regions the most predominant topology was the *P. alexandrae* and *Ap. erythropterus* sister relationship, but the majority of this discordance was due to ILS. Thus, in the face of high levels of ILS and reduced gene flow, the Z-chromosome could be biased towards harboring an inaccurate topology.

Differential selection on introgressed loci.— We also asked whether introgressed loci, should they be detected, were under positive selection. We identified only two such genes with known functions in the chicken genome, *FAM175A* and *C15orf40,* which involve cell cycle regulation and signaling, respectively. We were unable, however, to directly link these loci to known phenotypic changes that may have contributed to speciation in these birds. Adaptive introgression studies often involve both a phenotype and a candidate gene (Pardo-Diaz et al. 2012; Jones et al. 2018; Garg et al. 2019) and often focus on systems that occur across strong environmental gradients (e.g., Ma et al. 2019). The most significant ecological distinction between *P. alexandrae* and *Ap. erythropterus* is that the former occurs in the arid interior of central and Western Australia, whereas the latter inhabits more tropical and subtropical open woodlands across northern and eastern Australia and New Guinea. *Polytelis anthopeplus* and *P. swainsonii* also occur in more mesic open woodlands in temperate and semi-arid habitats. However, we had no a priori target genes to focus on in *Polytelis/Aprosmictus*, nor are there any apparently introgressed external phenotypic traits. We know of no characters that are uniquely shared between *P. alexandrae* and *Ap. erythropterus* and that do not occur in the other members of *Polytelis* or *Alisterus* (e.g., blue rump color albeit lighter in *P. alexandrae* and female *Ap. erythropterus*).

The degree of genetic exchange between *Ap. erythropterus* and *P. alexandrae* raises the question of whether this is consistent with a scenario of homoploid hybrid speciation. Recent research has highlighted a number of cases of hybrid speciation (e.g., Wang et al. 2021; Wu et al. 2023), where an emphasis has been placed on the role of introgression in reproductive isolation (Schumer et al. 2014; Long and Rieseberg 2024). Speciation genes in birds are often associated with plumage genes in recently diverged taxa that are morphologically similar (Toews et al. 2016; Turbek et al. 2021). In contrast, *Ap. erythropterus* and *P. alexandrae* share a common ancestor some 4.7–10.3 Mya (Smith et al. 2024), exhibit highly differentiated morphologies and plumage colors, and their present distributions essentially do not overlap. Although we cannot discard the possibility that the introgressed genes under selection may have contributed to speciation in *P. alexandrae*, we have no spatial, behavioral, or phenotypic context in which to interpret the results. Hybrid speciation is not compelling in this case.

We also explored whether there was asymmetry within variant functional categories between introgressed versus non-introgressed regions. In all functional categories, there were significant differences between introgressed and non-introgressed loci in those regions but the effect sizes were typically small (Cohen's $d < 0.2$). Collectively, this suggests that there were plenty of potential targets for positive selection in *P. alexandrae* from *Ap. erythropterus*, but few regions were actually subject to selection, a pattern mirrored in other studies (e.g., Ma et al. 2019). Interestingly, in-

trogressed genes were subject to significantly stronger purifying selection. Proportionately, positive selection acted at similar frequencies on non-introgressed regions within *P. alexandrae* versus regions introgressed from *Ap. erythropterus*; however, a larger number of existing variants were subject to selection. This parallels a finding, albeit inferred from allele frequencies at the population level, where selection was more likely to act on standing variation than on new mutations (Lai et al. 2019). Our selection tests were conservative in that they focused on selection on protein-coding sequences, which are relevant given the temporal divergences between *P. alexandrae* and *Ap. erythropterus*. However, there are numerous tests of selection on allele frequencies that can include non-coding regions and that require population-level sampling (Vitti et al. 2013). These non-tested signatures of selection may capture patterns of adaptive introgression that could have increased the fitness of *P. alexandrae*.

Conclusion

The phylogenomic toolkit is now well-suited to distinguish the causes of gene tree heterogeneity across the genome by accounting for ILS and gene flow during phylogenetic inference, as we have done here for the long-tailed parrots of Australia. Previous studies on the complex found a close relationship between two species (*Polytelis alexandrae* and *Aprosmictus erythropterus*) placed in disparate genera due to their morphological and behavioral distinctiveness. We found robust support that this unexpected relationship was due to ancient introgression and that taxonomically *P. alexandrae* should remain in *Polytelis*. Although genomic data have shown that signals of introgression are ubiquitous in phylogenetic studies, what is less clear is the evolutionary outcomes of the gene flow beyond causing discordance in phylogenetic signals. Despite high levels of ancient introgression between *P. alexandrae* and *Ap. erythropterus*, we found that the majority of introgressed loci played a limited role in molecular adaptations. A small proportion of protein-coding markers were introgressed, and only two were under positive selection, a pattern consistent with the idea that new variants are not frequent targets of directional selection. The expansion of the phylogenomic toolkit to more broadly incorporate additional approaches, such as assessing neutral versus adaptive evolution, now provides the opportunity to infer both the evolutionary relationships and the process that gave rise to the phylogenetic pattern.

Supplementary data

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.t4b8gtjbc

Funding

This study was funded in part by National Science Foundation awards to BTS (DEB-1655736 and DBI-2029955).

Acknowledgments

We thank the following people and institutions for providing critically important samples: American Museum of Natural History, New York (P. Sweet, T. Trombone, V. Chua), Australian National Wildlife Collection, CSIRO, Canberra (C. Wilson), South Australian Museum, Adelaide (M. Penck), Queensland Museum, Brisbane (H. Janetzki), Museum and Art Gallery of the Northern Territory, Darwin (G. Dally, S. Horner), National Museum of Natural History, Smithsonian Institution (B. Schmidt, H. James, G. Graves), University of Kansas Biodiversity Institute (M. Robbins, R. Moyle), Western Australian Museum, Perth (R. Bray), Zoological Museum, Natural History Museum of Denmark, Copenhagen (P. Hosner), Carnegie Museum of Natural History (S. Rogers), and Natural History Museum, Tring (A. Bond, M. Adams). We thank Julian Teh for providing beautiful bird and range map illustrations. Gordon Gullock helped with photography. We thank N. Galtier for gracious help with the Aphid analyses, D. Eaton, and anonymous reviewers.

Data Availability

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.t4b8gtjbc and NCBI under BioProject ID PRJNA1157411.

Conflict of Interest statement

None declared.

References

Allman E.S., Baños H., Rhodes J.A. 2019. NANUQ: a method for inferring species networks from gene trees under the coalescent model. Algorithms Mol. Biol. 14(1):24.

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., Nikolenko S.I., Pham S., Prjibelski A.D., Pyshkin A.V., Sirotkin A.V., Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19(5):455–477.

Bezanson J., Edelman A., Karpinski S., Shah V.B. 2017. Julia: a fresh approach to numerical computing. SIAM Rev. 59(1):65–98.

Biscarini F., Cozzi P., Gaspa G., Marras G. 2018. detectRUNS: an R Package to detect runs of homozygosity and heterozygosity in diploid genomes. CRAN (The Compr. R Arch. Network). https://cran.r-project.org/web/packages/detectRUNS (Accessed January 2024)

Blair C., Ané C. 2020. Phylogenetic trees and networks can serve as powerful and complementary approaches for analysis of genomic data. Syst. Biol. 69(3):593–601.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120.

Burri R. 2017. Interpreting differentiation landscapes in the light of long-term linked selection: differentiation and long-term linked selection. Evol. Lett. 1(3):118–131.
Burri R., Nater A., Kawakami T., Mugal C.F., Olason P.I., Smeds L., Suh A., Dutoit L., Bureš S., Garamszegi L.Z., Hogner S., Moreno J., Qvarnström A., Ružić M., Sæther S.-A., Sætre G.-P., Török J., Ellegren H. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Res. 25(11):1656–1665.
Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17(4):540–552.
Chernomor O., von Haeseler A., Minh B.Q. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Syst. Biol. 65(6):997–1008.
Christidis L., Shaw D.D., Schodde R. 2008. Chromosomal evolution in parrots, lorikeets and cockatoos (Aves: Psittaciformes). Hereditas. 114(1):47–56.
Cingolani P., Platts A., Wang L.L., Coon M., Nguyen T., Wang L., Land S.J., Lu X., Ruden D.M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: sNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. fly. 6(2):80–92.
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.
De Lucca E.J. 1990. Karyotype studies in twenty-two species of parrots (Psittaciformes: Aves). Genet. Mol. Biol. 13(4):599–612.
DeBaun D., Rabibisoa N., Raselimanana A.P., Raxworthy C.J., Burbrink F.T. 2023. Widespread reticulate evolution in an adaptive radiation. Evolution. 77(4):931–945.
Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2(5):e68.
Degnan J.H., Salter L.A. 2005. Gene tree distributions under the coalescent process. Evolution 59:24–37.
Dobzhansky T. 1982. Genetics and the origin of species. New York: Columbia University Press.
Durand E.Y., Patterson N., Reich D., Slatkin M. 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28(8):2239–2252.
Durvasula A., Sankararaman S. 2020. Recovering signals of ghost archaic introgression in African populations. Sci. Adv. 6(7):eaax5097.
Edwards S.V. 2009. Is a new and general theory of molecular systematics emerging? Evolution. 63(1):1–19.
Emms D.M., Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238.
Faircloth B.C. 2015. 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.
Flouri T., Jiao X., Huang J., Rannala B., Yang Z. 2023. Efficient Bayesian inference under the multispecies coalescent with migration. Proc. Natl. Acad. Sci. USA. 120(44):e2310708120.
Flouri T., Jiao X., Rannala B., Yang Z. 2018. Species tree inference with BPP using genomic sequences and the multispecies coalescent. Mol. Biol. Evol. 35(10):2585–2593.
Flouri T., Jiao X., Rannala B., Yang Z. 2020. A Bayesian implementation of the multispecies coalescent model with introgression for phylogenomic analysis. Mol. Biol. Evol. 37(4):1211–1223.
Fontaine M.C., Pease J.B., Steele A., Waterhouse R.M., Neafsey D.E., Sharakhov I.V., Jiang X., Hall A.B., Catteruccia F., Kakani E., Mitchell S.N., Wu Y.-C., Smith H.A., Love R.R., Lawniczak M.K., Slotman M.A., Emrich S.J., Hahn M.W., Besansky N.J. 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science. 347(6217):1258524.
Forshaw J.M. 2002. Australian parrots, (3rd rev. ed., p. 640). Robina Town Centre: Alexander Editions.
Frankel L.E., Ané C. 2023. Summary tests of introgression are highly sensitive to rate variation across lineages. Syst. Biol. 72(6):1357–1369.
Galtier N. 2024. An approximate likelihood method reveals ancient gene flow between human, chimpanzee and gorilla. Peer Commun. J. 4;e3
Gardner E.M., Bruun-Lund S., Niissalo M., Chantarasuwan B., Clement W.L., Geri C., Harrison R.D., Hipp A.L., Holvoet M., Khew G., Kjellberg F., Liao S., Pederneiras L.C., Peng Y.-Q., Pereira J.T., Phillipps Q., Ahmad Puad A.S., Rasplus J.-Y., Sang J., Schou S.J., Velautham E., Weiblen G.D., Zerega N.J.C., Zhang Q., Zhang Z., Baraloto C., Rønsted N. 2023. Echoes of ancient introgression punctuate stable genomic lineages in the evolution of figs. Proc. Natl. Acad. Sci. USA. 120(28):e2222035120.
Garg K.M., Sam K., Chattopadhyay B., Sadanandan K.R., Koane B., Ericson P.G.P., Rheindt F.E. 2019. Gene flow in the Müllerian mimicry ring of a poisonous Papuan songbird clade (Pitohui; Aves). Genome Biol. Evol. 11(8):2332–2343.
Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H.-Y., Hansen N.F., Durand E.Y., Malaspinas A.-S., Jensen J.D., Marques-Bonet T., Alkan C., Prüfer K., Meyer M., Burbano H.A., Good J.M., Schultz R., Aximu-Petri A., Butthof A., Höber B., Höffner B., Siegemund M., Weihmann A., Nusbaum C., Lander E.S., Russ C., Novod N., Affourtit J., Egholm M., Verna C., Rudan P., Brajkovic D., Kucan Ž., Gušic I., Doronichev V.B., Golovanova L.V., Lalueza-Fox C., de la Rasilla M., Fortea J., Rosas A., Schmitz R.W., Johnson P.L.F., Eichler E.E., Falush D., Birney E., Mullikin J.C., Slatkin M., Nielsen R., Kelso J., Lachmann M., Reich D., Pääbo S. 2010. A draft sequence of the Neandertal genome. Science. 328(5979):710–722.
Haldane J.B.S. 1922. Sex ratio and unisexual sterility in hybrid animals. J. Genet. 12(2):101–109.
Hasegawa M., Kishino H., Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22(2):160–174.
Hibbins M.S., Hahn M.W. 2021. Phylogenomic approaches to detecting and characterizing introgression. Genetics. 220(2):iyab173.
Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35(2):518–522.
Ji J., Jackson D.J., Leaché A.D., Yang Z.. 2022. Power of Bayesian and heuristic tests to detect cross-species introgression with reference to gene flow in the *Tamias quadrivittatus* group of North American chipmunks. Syst. Biol. 72(2):446–465.
Jones M.R., Mills L.S., Alves P.C., Callahan C.M., Alves J.M., Lafferty D.J.R., Jiggins F.M., Jensen J.D., Melo-Ferreira J., Good J.M. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science 360(6395):1355–1358.
Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14(6):587–589.
Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software Version 7: improvements in performance and usability. Mol. Biol. Evol. 30(4):772–780.
Keilwagen J., Hartung F., Grau J. 2019. GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. Methods Mol. Biol. 1962: 161–177.
Kuhlwilm M., Han S., Sousa V.C., Excoffier L., Marques-Bonet T. 2019. Ancient admixture from an extinct ape lineage into bonobos. Nat. Ecol. Evol. 3(6):957–965.
Lai Y.-T., Yeung C.K.L., Omland K.E., Pang E.-L., Hao Y., Liao B.-Y., Cao H.-F., Zhang B.-W., Yeh C.-F., Hung C.-M., Hung H.-Y., Yang M.-Y., Liang W., Hsu Y.-C., Yao C.-T., Dong L., Lin K., Li S.-H. 2019. Standing genetic variation as the predominant source for adaptation of a songbird. Proc. Natl. Acad. Sci. USA. 116(6):2152–2157.
Leaché A.D., Harris R.B., Rannala B., Yang Z. 2013. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63:17–30.
Li H. Seqtk: a toolkit for FASTA/Q Files. https://github.com/lh3/seqtk (Accessed April 2020)

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.
Liu X.-Q., Xia X.-M., Chen L., Wang X.-Q. 2022. Phylogeny and evolution of Cupressaceae: updates on intergeneric relationships and new insights on ancient intergeneric hybridization. Mol. Phylogenet. Evol. 177:107606.
Long Z., Rieseberg L.H. 2024. Documenting homoploid hybrid speciation. Mol. Ecol. e17412.
Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. Methods Mol. Biol. 1079:155–170.
Ma Y., Wang J., Hu Q., Li J., Sun Y., Zhang L., Abbott R.J., Liu J., Mao K. 2019. Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. Commun. Biol. 2(1):1–12.
MacGuigan D.J., Near T.J. 2019. Phylogenomic signatures of ancient introgression in a rogue lineage of darters (Teleostei: Percidae). Syst. Biol. 68(2):329–346.
Maddison W.P., Knowles L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. Syst. Biol. 55(1):21–30.
Malinsky M., Matschiner M., Svardal H. 2021. Dsuite—fast D-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. 21(2):584–595.
Malinsky M., Svardal H., Tyers A.M., Miska E.A., Genner M.J., Turner G.F., Durbin R. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat. Ecol. Evol. 2(12):1940–1955.
Martin S.H., Davey J.W., Jiggins C.D. 2014. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. Mol. Biol. Evol. 32(1):244–257.
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 Res. 20(9):1297–1303.
Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37(5):1530–1534.
Muller H.J. 1940. Bearing of the *Drosoph i la* work on systematics. In: Huxley J.S., editors. The new systematics. Oxford: Clarendon Press. p. 185–268.
Muller H.J. 1942. Isolating mechanisms, evolution, and temperature. Biol. Symp. 6:71.
Ottenburghs J. 2022. Avian introgression patterns are consistent with Haldane's rule. J. Hered. 113(4):363–370.
Ottenburghs J., Honka J., Heikkinen M.E., Madsen J., Müskens G.J.D.M., Ellegren H. 2023. Highly differentiated loci resolve phylogenetic relationships in the Bean Goose complex. BMC Ecol Evo 23(1):2.
Pardo-Diaz C., Salazar C., Baxter S.W., Merot C., Figueiredo-Ready W., Joron M., McMillan W.O., Jiggins C.D. 2012. Adaptive introgression across species boundaries in *Heliconius* butterflies. PLoS Genet. 8(6):e1002752.
Patterson N., Moorjani P., Luo Y., Mallick S., Rohland N., Zhan Y., Genschoreck T., Webster T., Reich D. 2012. Ancient admixture in human history. Genetics. 192(3):1065–1093.
Pritchard J.K., Pickrell J.K., Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr. Biol. 20(4):R208–R215.
Provost K.L., Joseph L., Smith B.T. 2018. Resolving a phylogenetic hypothesis for parrots: implications from systematics to conservation. Emu—Austral Ornithol. 118(1):7–21,
Ravinet M., Faria R., Butlin R.K., Galindo J., Bierne N., Rafajlović M., Noor M.A.F., Mehlig B., Westram A.M. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. J. Evol. Biol. 30(8):1450–1477.
Runemark A., Vallejo-Marin M., Meier J.I. 2019. Eukaryote hybrid genomes. PLoS Genet. 15(11):e1008404.
Rus A., Cigudosa J., Juan J.L.C., Gómez A.O., Almeida T.A., Joshua S., Miranda J. 2016. Chromosomal evolution in Psittaciformes. Revisited. Int. J. Biol. 8: 34.
Schumer M., Rosenthal G.G., Andolfatto P. 2014. How common is homoploid hybrid speciation? Evolution. 68(6):1553–1560.
Schweizer M., Seehausen O., Güntert M., Hertwig S.T. 2010. The evolutionary diversification of parrots supports a taxon pulse model with multiple trans-oceanic dispersal events and local radiations. Mol. Phylogenet. Evol. 54(3):984–994.
Schweizer M., Seehausen O., Hertwig S.T. 2011. Macroevolutionary patterns in the diversification of parrots: effects of climate change, geological events and key innovations. J. Biogeogr. 38(11):2176–2194.
Seppey M., Manni M., Zdobnov E.M. 2019. BUSCO: assessing genome assembly and annotation completeness. Methods Mol. Biol. 1962: 227–245.
Sindel S., Gill J.H. 2003. Australian coral-billed parrots (the *Alisterus*, *Aprosmictus* and *Polytelis* genera). New South Wales, Australia: Singil Press.
Smith B.T., Merwin J., Provost K.L., Thom G., Brumfield R.T., Ferreira M., Mauck III W.M., Moyle R.G., Wright T., Joseph L.. 2023. Phylogenomic analysis of the parrots of the world distinguishes artifactual from biological sources of gene tree discordance. Syst. Biol. 72(1):228–241.
Smith B.T., Thom G., Joseph L. 2024. Revised evolutionary and taxonomic synthesis for parrots (Order: Psittaciformes) guided by phylogenomic analysis. Bull. Am. Mus. Nat. Hist. 2024(468):1–87.
Solís-Lemus C., Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS Genet. 12(3):e1005896.
Solís-Lemus C., Bastide P., Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. Mol. Biol. Evol. 34(12):3292–3298.
Stryjewski K.F., Sorenson M.D. 2017. Mosaic genome evolution in a recent and rapid avian radiation. Nat. Ecol. Evol. 1(12):1912–1922.
Suvorov A., Kim B.Y., Wang J., Armstrong E.E., Peede D., D'Agostino E.R.R., Price D.K., Waddell P., Lang M., Courtier-Orgogozo V., David J.R., Petrov D., Matute D.R., Schrider D.R., Comeault A.A. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. Curr. Biol. 32(1):111–123.e5.
R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/ (Accessed 2024)
Thom G., Amaral F.R.D., Hickerson M.J., Aleixo A., Araujo-Silva L.E., Ribas C.C., Choueri E., Miyaki C.Y. 2018. Phenotypic and genetic structure support gene flow generating gene tree discordances in an Amazonian floodplain endemic species. Syst. Biol. 67(4):700–718.
Thom G., Moreira L.R., Batista R., Gehara M., Aleixo A., Smith B.T. 2024. Genomic architecture predicts tree topology, population structuring, and demographic history in Amazonian birds. Genome Biol. Evol. 16.(1):evae002.
Toews D.P.L., Taylor S.A., Vallender R., Brelsford A., Butcher B.G., Messer P.W., Lovette I.J. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. Curr. Biol. 26(17):2313–2318.
Turbek S.P., Browne M., Di Giacomo A.S., Kopuchian C., Hochachka W.M., Estalles C., Lijtmaer D.A., Tubaro P.L., Silveira L.F., Lovette I.J., Safran R.J., Taylor S.A., Campagna L. 2021. Rapid speciation via the evolution of pre-mating isolation in the Iberá Seedeater. Science. 371(6536):eabc0256.
Van Dongen M.W.M., De Boer L.E.M. 1984. Chromosome studies of 8 species of parrots of the families Cacatuidae and Psittacidae (Aves: Psittaciformes). Genetica. 65(1):109–117.
Vitti J.J., Grossman S.R., Sabeti P.C. 2013. Detecting natural selection in genomic data. Annu. Rev. Genet. 47(1):97–120.
Wang L., Liu S., Yang Y., Meng Z., Zhuang Z. 2022. Linked selection, differential introgression and recombination rate variation promote heterogeneous divergence in a pair of yellow croakers. Mol. Ecol. 31(22):5729–5744.
Wang X., He Z., Shi S., Wu C.-I. 2020. Genes and speciation: is it time to abandon the biological species concept? Natl. Sci. Rev. 7(8):1387–1397.

Wang Z., Jiang Y., Bi H., Lu Z., Ma Y., Yang X., Chen N., Tian B., Liu B., Mao X., Ma T., DiFazio S.P., Hu Q., Abbott R.J., Liu J. 2021. Hybrid speciation via inheritance of alternate alleles of parental isolating genes. Mol. Plant 14(2):208–222.
Wen D., Yu Y., Zhu J., Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. Syst. Biol. 67(4):735–740.
Whelan S., Goldman N. 1999. Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics. Mol. Biol. Evol. 16(9):1292–1299.
Wu H., Wang Z., Zhang Y., Frantz L., Roos C., Irwin D.M., Zhang C., Liu X., Wu D., Huang S., Gu T., Liu J., Yu L. 2023. Hybrid origin of a primate, the gray snub-nosed monkey. Science. 380(6648): eabl4997.
Yang Z. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22(4):1107–1118.
Yang Z., Bielawski J.P. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15(12):496–503.
Yang Z., Wong W.S.W., Nielsen R. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24(8):1586–1591.
Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19(S6):153.
Zhang D., Rheindt F.E., She H., Cheng Y., Song G., Jia C., Qu Y., Alström P., Lei F. 2021. Most genomic loci misrepresent the phylogeny of an avian radiation because of ancient gene flow. Syst. Biol. 70(5):961–975.
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. Mol. Biol. Evol. 36(11):2375–2386.
Zhang J., Nielsen R., Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22(12):2472–2479.
Zhou B.-F., Yuan S., Crowl A.A., Liang Y.-Y., Shi Y., Chen X.-Y., An Q.-Q., Kang M., Manos P.S., Wang B. 2022. Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. Nat. Commun. 13:1–14.

Syst. Biol. 0(0):1–19, 2025

https://doi.org/10.1093/sysbio/syaf066
Advance Access publication September 27, 2025