Phylogeography of the iconic Australian pink cockatoo, Lophochroa leadbeateri

Phylogeography of the iconic Australian pink cockatoo, *Lophochroa leadbeateri*

Running title: Phylogeography of the pink cockatoo

Kyle M. Ewart[1,2], Rebecca N. Johnson[1,2], Leo Joseph[3], Rob Ogden[4], Greta J. Frankham[2,5], Nathan Lo[1]

[1] The University of Sydney, School of Life and Environmental Sciences, NSW, Australia
[2] Australian Centre for Wildlife Genomics, Australian Museum Research Institute, NSW, Australia
[3] Australian National Wildlife Collection, National Research Collections Australia, CSIRO, Canberra, Australia
[4] Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Edinburgh, UK
[5] Centre for Forensic Science, University of Technology Sydney, PO Box 123, Broadway, NSW, 2007 Australia

~

Corresponding author: Kyle M. Ewart; Kyle.Ewart@austmus.gov.au

Abstract

The pink cockatoo (*Lophochroa leadbeateri*; or Major Mitchell's cockatoo) is one of Australia's most iconic bird species. Two subspecies based on morphology are separated by a biogeographical divide, the Eyrean Barrier. Testing the genetic basis for this subspecies delineation, clarifying barriers to gene flow and identifying any cryptic genetic diversity will likely have important implications for conservation and management. Here, we used genome-wide SNPs and mitochondrial DNA data to conduct the first range-wide genetic assessment of the species. The aims were to investigate the pink cockatoo's phylogeography, characterise conservation units and reassess subspecies boundaries. We found consistent but weak genetic structure between the two subspecies based on nuclear SNPs. However, phylogenetic analysis of nuclear SNPs and mitochondrial DNA sequence data did not recover reciprocally monophyletic groups, indicating that the subspecies are not evolutionarily distinct. Consequently, we have proposed that the two currently recognized subspecies be treated as separate management units rather than evolutionarily significant units. Because poaching is suspected to be a threat to this species, we assessed the utility of our data for wildlife forensic applications. We demonstrated that a subspecies identification test could be designed using as few as twenty SNPs.

Key words

conservation genetics - *Lophochroa leadbeateri* - phylogeography - population genomics - wildlife forensics - wildlife trade

Introduction

The pink cockatoo (also known as the Major Mitchell's cockatoo), *Lophochroa leadbeateri* (Vigors, 1831), is an iconic bird species endemic to Australia. It is considered by many to be the most beautiful and spectacular of the cockatoos (Cacatuidae; Rowley & Chapman, 1991; Schodde, 1994), having pink-white plumage and an impressive bright red, yellow and white crest. The pink cockatoo is a hardy species that occurs in low densities throughout Australia's harsh arid and semi-arid regions.

Within the pink cockatoo's wide yet patchy distribution, four core breeding regions are apparent (Blakers *et al*., 1984; Fig. 1a). Although previous authors have recognised a variable number (0-4) of subspecies (ssp; e.g. Mathews, 1912 – 3 ssp; Peters, 1937 – 4 ssp; Condon, 1975 - 0 ssp; Hall, 1974 – 3 ssp; Schodde, 1997 – 2 ssp), two subspecies*, L. l. leadbeateri* and *L. l. mollis* (*cf* Forshaw & Cooper, 1981) have been generally accepted since the publication of Schodde's 1994 study (Fig. 1a) on the basis of body size and colour and pattern of the crest. These subspecies are separated by the Eyrean Barrier (Fig. 1a): a well-documented biogeographic barrier in southern Australia for a range of bird species (Ford, 1974; Schodde, 1982; Kearns *et al.*, 2009; Dolman & Joseph, 2012). *Lophochroa leadbeateri leadbeateri* is east of the Eyrean Barrier and has a more prominent yellow band in its crest and is larger in body size, while *L. l. mollis* is west and north of the Eyrean Barrier (Schodde 1994, 1997; Forshaw & Cooper, 2002).

Despite its wide distribution, the pink cockatoo is of conservation concern. In the eastern part of its distribution it is listed as Vulnerable (New South Wales and Queensland - Biodiversity Conservation Act, 2016; Nature Conservation Wildlife Regulation, 2006), or Threatened (Victoria - Flora and Fauna Guarantee Act, 1988; Walker *et al.*, 1999) (see Fig. 1a for State

49 localities). The species' abundance and range in north-western Victoria and western New
50 South Wales have been greatly reduced through the removal of habitat, in particular the loss
51 of hollow-bearing trees (Garnett *et al.*, 2011). Like other cockatoo species, the pink cockatoo
52 is unable to excavate its own hollows for nesting and so requires naturally occurring tree
53 hollows (Mackowski, 1984; Cameron, 2007). Further, increased agriculture and clearing of
54 feeding habitat have impacted the species, particularly in the southwest of its range in the
55 Western Australian wheat belt region (Rowley & Chapman, 1991). Another threat to this
56 species is poaching (Forshaw & Cooper, 1981; Higgins, 1999), which Rowley & Chapman
57 (1991) found to impact the most critical stage of the species' life cycle: recruitment of young.
58 Poaching is directly linked to demand for the species in the illegal pet trade. Together, these
59 factors indicate a need for improved understanding of phylogeographic patterns within the
60 species to aid in the conservation management of the species.
61
62 Genomic tools allow researchers to investigate how genetic diversity is distributed among
63 populations. They may help to identify and manage at-risk populations. Characterising
64 discrete units of genetic variation, termed conservation units (Ryder, 1986), and clarifying
65 barriers to gene flow within the pink cockatoo will facilitate conservation strategies that
66 maximize the evolutionary potential of the species (Frankham *et al*., 2010). The putative
67 subspecies barrier, the Eyrean Barrier, comprises the Flinders Ranges and Lake Eyre Basin
68 (Schodde, 1982). It is thought to have limited dispersal during the Plio-Pleistocene due to the
69 presence of vast lakes associated with the Lake Eyre Basin, and then in the Pleistocene due to
70 extreme aridity (Ford & Parker, 1973; Ford, 1974; Schodde, 1982; Joseph *et al.*, 2006).
71 However, the timing and strength with which the Eyrean Barrier has separated populations
72 within species is known to vary between avian taxa (Schodde, 1982; Dolman & Joseph,
73 2012; McElroy *et al*., 2018). Whether the morphological differences between pink cockatoo

74 subspecies at this barrier reflect underlying genetic divergence and potential conservation
75 units is unknown. Schodde (1994) suggested that there is currently no dispersal between
76 subspecies over this barrier, and that the two may even warrant recognition at species rank.
77 Further, it is unknown whether cryptic genetic structure exists across other well-characterised
78 southern Australian arid-zone biogeographic barriers within the pink cockatoo distribution,
79 such as the Nullarbor and Murravian Barriers (see Schodde & Mason, 1999). The impact of
80 these biogeographic barriers varies considerably between species (Neaves *et al*., 2012).
81
82 Clarifying the species' evolutionary history and intraspecific taxonomy have been
83 problematic due to a combination of poor sampling, relatively weak morphological
84 divergence across the species (e.g. see Forshaw, 2011), and the need to disentangle patterns
85 of geographical, sexual and age-related variation. Genomic analyses have the potential to
86 help to characterize conservation units, investigate connectivity among core breeding
87 populations, and resolve lingering taxonomic uncertainties about subspecies boundaries
88 (Baumsteiger *et al*., 2017; Marie *et al*., 2019; Tonzo *et al*., 2019; Ewart *et al*., 2020).
89 Furthermore, genetic data could facilitate the development of wildlife forensic tools, such as
90 geographical provenance and progeny testing, to increase the capacity for detection and
91 prosecute trafficking crimes involving this species (Walker *et al.*, 1999; Huffman & Wallace,
92 2011). The pink cockatoo is listed under CITES Appendix II, and trade in the species is
93 strictly regulated under Australian legislation.
94
95 Here we perform the first comprehensive phylogeographic assessment of the pink cockatoo
96 to address the topics we have raised above. This builds on two earlier genetic studies
97 involving this species based on allozymes (Adams *et al.*, 1984), and a multilocus nuclear and
98 mitochondrial DNA (mtDNA) data set (White *et al.*, 2011); both used only a few individuals

to address the species' systematic position with respect to other cockatoos. Pink cockatoo specimens from across the species' range have been collected over many decades and are stored in museums throughout Australia and elsewhere. Owing to developments in museum genomics, genome-wide data of use in population-level studies can be generated from old museum specimens (Rowe *et al.*, 2011; Ewart *et al.*, 2019). We generated thousands of genome-wide single nucleotide polymorphism markers (SNPs) and sequence data at three mtDNA markers from pink cockatoo frozen tissue and toe pad samples across their entire distribution. We performed comprehensive population genomic analyses to investigate potential barriers to gene flow for the purposes of clarifying taxonomy and informing conservation management. These data can be interpreted in light of the biogeography and palaeoenvironmental history of Australia's arid and semi-arid zones, and compared to the steadily increasing body of phylogeographic analyses of species having broadly similar distributions across southern Australia (Neaves *et al.*, 2009; Dolman & Joseph, 2012, 2015; Engelhard *et al.*, 2015; Ansari *et al.*, 2019).

Methods

Sample acquisition and DNA extractions

We acquired pink cockatoo frozen tissue (frozen liver/muscle) (n=45) and toe pad (n=51) samples from across their distribution (Fig. 1a & Table S1). Samples were obtained from: the Australian National Wildlife Collection, Canberra (ANWC); the Australian Museum, Sydney (AM); Museum Victoria, Melbourne (MV); and the Western Australian Museum, Perth (WAM). Collection dates for these samples ranged from 1883 to 2011 (Table S1).

Thinly sliced toe pads (~2 mm thick) were sampled from traditional museum specimens, and DNA was extracted following Ewart *et al.* (2019). These DNA extractions were performed in a clean room facility dedicated to historical museum samples likely to have degraded DNA. Genomic DNA was extracted from frozen tissue samples following the manufacturer's protocols for the 'Bioline Isolate II Genomic DNA kit' Bioline (Australia). DNA concentration was measured using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific).

SNP genotyping

SNP data was generated using DArTseq™, a reduced representation sequencing method (methods described in Kilian *et al.*, 2012; Cruz *et al.*, 2013). This was performed by Diversity Arrays Technology (DArT) in Canberra, Australia. DArTseq has previously been used to generate SNP data for a range of phylogeographic, phylogenetic, and population genetic studies on vertebrate species (Melville *et al*., 2017). Briefly, different combinations of restriction enzymes were tested, and the *PstI-SphI* enzymes were selected for digestion of cockatoo DNA. DNA was then processed as per Kilian *et al.* (2012), using two different adaptors that correspond with the restriction site overhangs, both containing an Illumina flow cell attachment sequence, and one (the *PstI*-compatible adapter) also containing a sequencing primer sequence and varying length barcode region. The library was subject to PCR (using REDTaq DNA Polymerase, Sigma-Aldrich) as follows: initial denaturation at 94°C for 1 min, then 30 cycles of 94°C for 20 sec, 58°C for 30 sec and 72°C for 45 sec, and a final extension step at 72°C for 7 min. The library was then normalized and sequenced by first performing a c-Bot (Illumina) bridge PCR, followed by single end sequencing for 77 cycles on an Illumina Hiseq2500.

146 The resultant short-read sequences were processed using the DArT Pty Ltd analytical
147 pipelines. First, poor quality sequences were removed (using a Phred score ≥10), and
148 sequences were demultiplexed (using a barcode Phred score ≥30). Second, sequences were
149 trimmed to 69 bp and clustered with a Hamming distance threshold of 3. Low-quality regions
150 from singleton tags were corrected where possible. Third, SNPs were called using the
151 proprietary DArTsoft14 SNP calling pipeline. Real alleles were discriminated from
152 paralogous sequences by assessing a range of parameters including sequence depth, allele
153 count and call rate.
154
155 *SNP filtering*
156 We applied numerous SNP filtering criteria depending on the analysis (following Ewart *et*
157 *al*., 2019). First, we removed the duplicate/triplicate samples with the highest amount of
158 missing data. Second, we removed potentially erroneous SNPs, and SNPs with a high level of
159 missing data, based on reproducibility (100%) and call rate (>80%), using the R package
160 dartR version 1.0.5 (Gruber *et al.*, 2018). Third, to meet the population genetic assumptions
161 of some analyses, we removed linked SNPs, outlier SNPs that potentially represented loci
162 under selection, and SNPs out of Hardy-Weinberg equilibrium (HWE). To remove linked
163 SNPs to meet the assumption of linkage disequilibrium for some of the analyses, we retained
164 only one SNP per DArTseq locus using the R package dartR. To identify and remove outlier
165 SNPs that are potentially under directional or balancing selection to meet the assumption of
166 neutrality for some analyses, we used LOSITAN (Beaumont & Nichols, 1996; Antao *et al*.,
167 2008). For this analysis, samples were divided into subspecies, then performed 100,000
168 simulations, applying the 'infinite alleles' mutation model, a 0.95 confidence interval and a
169 0.1 false discovery rate. To identify departure from HWE, we used ARLEQUIN version 3.5
170 (Excoffier & Lischer, 2010), implementing 1,000,000 Markov Chain steps and a burnin of

100,000. We removed loci with a p-value <0.01 that potentially deviate from HWE. For this analysis, we considered all samples as one population, which is likely a conservative approach, as we would expect some false positives due to the Wahlund effect.

To investigate whether remnant poor-quality SNPs were skewing results, additional filters were applied to represent a 'stringently filtered' data set, and analyses were repeated. Here, we filtered SNPs for average locus coverage (>20 X) using the R package dartR, and minor allele frequency (MAF) (>0.05) using the R package poppr version 2.6.1 (Kamvar *et al*., 2014, 2015).

Additionally, to ensure that the inclusion of toe pad samples from old museum specimens did not skew results, SNPs were re-called using only the more contemporary tissue samples (using the SNP calling methods outlined in the previous section). SNPs were subsequently re-filtered. Additional details on SNP filtering methods and variants are provided in the Supplementary Material (Appendix I).

SNP quality control

To quantify genotyping error, we included 18 replicate and 4 triplicate samples among the 96 pink cockatoo samples analysed (indicated in Table S1). We used various replicate/triplicate types to investigate the factors that may influence error, including: frozen tissue replicates (from the same and different DArTseq plates), toe pad replicates (from the same and different DArTseq plates), frozen tissue / toe pad replicates (i.e. a frozen tissue and toe pad from the same individual), and tissue DNA replicates (from the same and different DArTseq plates).

We calculated SNP error rates (i.e. the number of SNP mismatches between replicate pairs over the total number of SNPs that were not missing in both replicates) using R functions from Mastretta-Yanes *et al*. (2015). Error rates were calculated pre- and post- SNP-filtering.

Generation of mitochondrial DNA sequence data

To generate mitochondrial reference genomes, we performed low-coverage whole genomic sequencing for four pink cockatoo samples (indicated in Table S1), following the NEBNext DNA library preparation protocol, with a pre-treatment of 500 bp shearing using Covaris M220. The libraries were then sequenced on an Illumina MiSeq using paired-end 251 bp sequencing. Library preparation and sequencing were performed at the Monash University Malaysia Genomics Facility (Selangor, Malaysia). The resultant paired sequence reads were trimmed using the BBDuk plugin in Geneious version 10.2.4 (Kearse *et al*., 2012), then assembled using Geneious and NOVOPlasty (Dierckxsens *et al.*, 2017). We then designed primers for the *ND4* and *ND5* genes and d-loop (for *ND2*, we used primers from Sorenson, 2003), and amplified and sequenced 15 samples from across the pink cockatoo range (indicated in Table S1). Thus, the mtDNA analyses were carried out using 19 samples (4 using low-coverage whole genomic sequencing, and 15 using Sanger sequencing). The d-loop marker was subsequently excluded as it was unable to be reliably sequenced (possibly due to the presence of control region duplications, which are often found in parrot species; Schirtzinger *et al*., 2012; Eberhard & Wright, 2016). Additional details on mitochondrial genome assemblies, primers, PCR conditions and sequencing can be found in the Supplementary Material (Appendix II).

220 *Identifying population structure*

221 We used five methods to investigate population structure present in the SNP genotype data.
222 Details of the different SNP filtering strategies and samples used in the different analyses are
223 provided in the Supplementary Material (Appendix I; Table S1). First, genetic variation was
224 summarized and visualized using a principal coordinates analysis (PCoA). This was
225 performed using the R packages dartR and ade4 version 1.7 (Chessel *et al.*, 2004). Second,
226 STRUCTURE version 2.3 (Pritchard *et al.*, 2000) was used to investigate genetic structure and
227 admixture. For this analysis, we modelled up to five ancestral populations (K=1-5),
228 implementing 10 replicates for each K, assuming admixture and correlated allele frequencies
229 (Porras-Hurtado *et al*., 2013). We ran the analysis for 2 million iterations with a burn-in of 1
230 million. This analysis was parallelized and automated using StrAuto version 1.0 (Chhatre &
231 Emerson, 2017). We considered six different estimators to determine the optimal value of K,
232 generated using StructureSelector (Li & Liu, 2018). Replicate runs were merged and bar plots
233 were generated using CLUMPAK (Kopelman *et al.*, 2015), implemented through
234 StructureSelector. We took a hierarchical approach, whereby the population clusters
235 identified using the full dataset were separated, re-filtered, then run independently. Third, to
236 investigate whether patterns of genetic differentiation derived from continuous (i.e. isolation
237 by distance; IBD) or discrete (e.g. biogeographic barriers) phylogeographic processes, we
238 performed a conStruct analysis (Bradburd *et al*., 2018), implementing the spatial model. A
239 conStruct (i.e. 'continuous structure') analysis is similar to the STRUCTURE analysis, but
240 controls for geographic distance between samples. Based on initial optimization, we ran two
241 independent conStruct analyses, with the 'adapt delta' parameter (the target average proposal
242 acceptance probability) set at 0.85, implementing two chains with 100,000 MCMC iterations
243 for each run. We checked for consistency between chains and independent runs, and visually
244 checked for convergence using the trace plots generated by conStruct. To determine an

appropriate level of parameterization, we ran five replicates of a cross-validation analysis comparing the spatial and non-spatial models for K = 1-5 for each replicate. We used a random 90% subsample as the training partition, and ran the analysis for 10,000 MCMC iterations.

Fourth, to measure genetic divergence between subspecies, we calculated pairwise F_{ST} values (Weir & Cockerham, 1984) using the R package hierfstat version 0.4.22 (Goudet & Jombart, 2015). F_{ST} values were considered significant if their associated confidence intervals (based on 0.025 and 0.975 quantiles, implementing 1000 bootstraps) did not encompass 0. To investigate differentiation within and between subspecies we performed an AMOVA using the R package poppr, and checked for significance using 10,000 permutations implemented in the R package ade4. To investigate whether any genetic structure patterns were driven by closely related individuals (e.g. cousins), we performed an inter-individual kinship analysis using the R package SNPRelate version 1.14 (Zheng *et al*., 2012).

We performed a haplotype network analysis to investigate population structure within the mtDNA sequence dataset. We performed this analysis using PopART (Leigh & Bryant, 2015), based on concatenated *ND2*, *ND4* and *ND5* sequences (a total of 2037 bp) and 19 samples, implementing the statistical parsimony TCS method (Clement *et al.*, 2000). Additionally, we calculated net nucleotide divergence (*Da*) between the two subspecies based on the mtDNA sequence dataset using the R package strataG version 2.4.905 (Archer *et al*., 2017).

Gene flow patterns

To investigate the influence of geographic distance in our genetic structure results, we investigated the correlation between genetic and geographic distance (i.e. IBD). As there are no discrete sampling sites (reflecting the pink cockatoo's continuous distribution; Fig. 1a), we analysed inter-individual distances. Individual-based genetic distances were based on PCA-based Euclidean distance, following Shirk *et al.* (2017), calculated using 45 principal components (35 when using only fresh tissue samples), and performed using the R package adegenet version 2.1.0 (Jombart, 2008). We then performed a Mantel test using these Euclidean genetic distances and geographic distance (in kilometres) using the R packages adegenet and dartR.

Due to the ongoing debate surrounding the use of Mantel tests to infer IBD patterns (e.g. Diniz-Filho *et al*., 2013), especially when considering inter-individual distances, we analysed interpopulation gene flow along a transect following methods in Ogden & Thorpe (2002). Indirect gene flow inferences were based on pairwise F_{ST} measurements (calculated as above, but scaled by pairwise geographic distance) between five 'sample clusters' (three individuals per cluster) across Australia, focusing on the putative subspecies barrier (Fig. 3b; Table S1). Willing *et al*. (2012) demonstrated that F_{ST} values can be estimated with relatively small sample sizes when using thousands of SNPs. To complement this analysis of gene flow across the Eyrean Barrier, we ran a conStruct analysis using the same 15 samples in the transect above. We used the same settings as the previous conStruct analysis, except the 'adapt delta' parameter was set to 0.7.

295 ## *Genetic diversity*

296 To measure the genetic diversity within each subspecies, we calculated allelic richness,
297 heterozygosity and private allele counts for each SNP marker. Allelic richness was calculated
298 using the R package PopGenReport version 3.0.4 (Adamack & Gruber, 2014), implementing
299 rarefaction to account for differences in sample size. Observed and expected heterozygosity
300 were calculated using GenAlEx (Peakall & Smouse, 2006, 2012). A count of private alleles
301 per population was calculated using the R package poppr. Mitochondrial DNA diversity was
302 measured in terms of nucleotide diversity, proportion of polymorphic sites, and number of
303 haplotypes using Geneious and the R package pegas version 0.1 (Paradis, 2010).
304

305 ## *Population growth*

306 To investigate factors that may have caused discordant mtDNA and nuclear DNA clustering
307 patterns (see the '*Results*' section) and to test for population growth, we computed Tajima's
308 D (Tajima, 1989), Fu's Fs (Fu, 1997) and Ramos-Onsin's R2 (Ramos-Onsins & Rozas, 2002)
309 statistics using DnaSP 6.12.03 (Rozas *et al*., 2017), based on mtDNA sequence data (2037 bp
310 of concatenated *ND2*, *ND4* and *ND5* sequences). The significance of the statistics was
311 inferred using coalescent simulations with 1000 replicates. Additionally, a mismatch
312 distribution plot was generated using the R package pegas.
313

314 ## *Phylogenetic methods*

315 We performed phylogenetic analyses to investigate whether genetic units identified in the
316 population genetic analyses were evolutionarily distinct within a phylogenetic framework.
317 Phylogenetic analyses based on SNPs were performed using SNAPP (Bryant *et al*., 2012),
318 implemented in BEAST version 2.4 (Bouckaert *et al.*, 2014), to compare 'species'
319 hypotheses (ESU hypotheses in this case). We used SNAPP to compare the relative support

for two models: one enforcing monophyly of each of the two subspecies (which corresponds to two genetic units in population genetic analyses; see Results section), and one without enforcing monophyly. As SNAPP is computationally intensive, we included four individuals per subspecies and 1000 randomly selected SNPs with no missing data from the putatively neutral SNP data set (see Supplementary Material, Appendix I, for more details) to improve computational tractability. We ran SNAPP for 4 million Markov chain Monte Carlo (MCMC) steps, sampling every 1000 steps after a burn-in of 400,000 steps. We used allele frequencies for the forward and backward mutation rates, and the default settings for priors. Model support was subsequently estimated using the AICM (Akaike information criterion through MCMC) method in Tracer version 1.6 (Rambaut *et al*., 2014). AICM was chosen over the preferred stepping-stone and path sampling analyses to improve computational tractability. As AICM has been shown to suffer from poor repeatability (Baele *et al*., 2012), we ran three replicate SNAPP analyses for each model (i.e. three enforcing monophyly of subspecies, and three not enforcing monophyly) and subsequently estimated AICM for each of the six runs.

To complement the SNAPP analysis, we performed a maximum likelihood phylogenetic analysis using RAxML (Stamatakis, 2014) based on concatenated SNP data (see Supplementary Material, Appendix I). We implemented the GTR substitution model with gamma-distributed rates among sites and the Lewis-type ascertainment bias correction to account for the exclusion of invariant sites, and performed 1000 bootstrap replicates to estimate node support. Trees were rooted using the midpoint method and visualised using Figtree 1.4.2 (Rambaut, 2009).

We performed a Bayesian phylogenetic analysis of mtDNA data (2,037 bp of concatenated *ND2*, *ND4* and *ND5*) using MrBayes 3.2 (Ronquist *et al*., 2012). This analysis was performed

using four independent Markov chains, each run for 100 million steps with a 25% burn-in, and sampled every 100 steps with convergence diagnostics calculated every 100 steps. We implemented the HKY substitution model with gamma-distributed rates among sites. Convergence diagnostics were assessed using Tracer (ESS values <200 were considered inadequate). This analysis was performed with and without an outgroup (*Cacatua pastinator*; GenBank accession: JF414240). Trees were rooted using either the midpoint method or an outgroup, and visualized using Figtree.

Testing SNPs for wildlife forensic applications

We filtered a subset of SNPs based on their utility in a geographic provenance assignment test by investigating SNP contributions in a discriminant analysis of principal components (DAPC). DAPC minimizes variation within groups, and maximizes variation between groups. First, we performed DAPC on the entire SNP dataset with no missing data (see extra filtering details in Supplementary data, Appendix I) using the R package adegenet. We considered two populations (K=2), corresponding to separation of the two subspecies, then repeated the analysis considering three populations (K=3) to investigate whether more fine-scale geographic assignment was possible. Second, SNPs were ranked based on their contribution to the clustering analysis. Third, we iterated through decreasing numbers of SNPs (increments of five SNPs) to investigate the minimum number of SNPs required to separate the two subspecies clusters. Finally, we tested the utility of a refined set of SNPs for geographic/subspecies assignment by assigning six randomly selected individuals (three individuals per subspecies) in separate tests using GeneClass2 (Piry *et al.*, 2004). For this analysis, we implemented the frequency-based assignment method (Paetkau *et al.*, 1995) and a 0.05 assignment threshold. The individual being tested was removed from the 'reference'

data before each analysis. Likelihood ratios were calculated from the assignment likelihood results, considering different prosecution and defence hypotheses.

Results

SNP genotyping

Seventy-eight samples were successfully genotyped using DArTseq (Table S1). DNA extracts from one frozen tissue sample (out of 45) and 20 toe pad samples (out of 51) were unsuitable for successful DArTseq library preparation. The oldest sample successfully genotyped was collected in 1912; all samples collected before this date failed. The DArTsoft14 pipeline called 20,324 SNPs from the successfully genotyped 78 samples (with 36.32% missing data). This SNP data set was reduced to 4,135 SNPs (with 12.26% missing data) after filtering for quality and missing data, 2,131 SNPs (with 11.78% missing data) after filtering for neutrality and linkage, and 1,279 SNPs (with 10.35% missing data) after applying more stringent filtering (see Supplementary Material, Appendix I for data filtering details, and Table S1 to view which individuals were used in each analyses). When using only the more contemporary tissue samples for SNP calling, the DArTsoft14 pipeline called 16,472 SNPs (with 16.79% missing data), which was reduced to 6,466 SNPs (with 3.07% missing data) after filtering for quality and missing data, and to 4,891 SNPs (with 1.95% missing data) after filtering for neutrality and linkage.

SNP quality control

Of the 18 replicate and 4 triplicate samples examined, some failed. We found two additional replicate samples based on their genetic signature (i.e. they had different sample numbers and were held in different Museums but they were parts from the same individual in two

393 collections). This was subsequently confirmed with the relevant Museums. Overall, a total of
394 13 replicates and 4 triplicates were used to quantify genotyping error (Table S2).
395
396 Filtering reduced the allele error rate in all samples except one (ANWC B38557; this sample
397 also had a very high proportion of missing data) (Table S1). After filtering, SNP error rates
398 for frozen tissue and DNA replicates/triplicates were all <3%. The SNP error rate and/or
399 shared missing data (missing in both replicates) was particularly high in eight 'toe pad/toe
400 pad' and 'tissue/toe pad' replicates (ranging from 12.10-23.08% and 0.63-97.17%
401 respectively after filtering). Although several problematic samples were removed from many
402 of the population genetic analyses (see Supplementary Material, Appendix I), error in toe pad
403 samples was variable, ranging from 2.87-23.08% in 'toe pad/toe pad' replicates after
404 filtering, hence toe pad samples with relatively high error rates are likely present in some
405 analyses.
406
407 *Genetic structure*
408 The PCoA revealed three distinct clusters: one *L. l. mollis* cluster and two *L. l. leadbeateri*
409 clusters (Fig. 1b). Within *L. l. leadbeateri*, five individuals from central Queensland formed a
410 cluster that was distant from the other samples. Kinship between these individuals was
411 relatively high (0.045-0.144; Table S3) compared to the average kinship of the entire dataset
412 (0.008; excluding self-kinship values), which may distort the level of genetic structure in this
413 region. When removing four out of the five central Queensland samples in a PCoA, the
414 remaining sample clusters with the other *L. l. leadbeateri* individuals (this result is consistent
415 when different central Queensland individuals are used; Figure S1). The only other
416 Queensland individual in the data set, from southern Queensland (see Fig. 1a), clustered with
417 the other *L. l. leadbeateri* samples. There were five other outlier samples. The four outliers

418 near the origin of the PCoA plot (Fig. 1b) are likely explained by their high level of missing
419 data (>70%) (missing data are replaced by the mean allele frequency in the PCoA analysis).
420 The origin of the outlier from the Northern Territory (MV Z50083) is unclear. It may have
421 been either a migrant, an escaped aviary bird from the *L. l. leadbeateri* range, or the result of
422 a processing error (e.g. mislabelling, DNA contamination etc.).
423
424 Genetic variability in the STRUCTURE analysis was best explained using *K*=2-5, depending on
425 the estimator considered (Fig. S2). We present the major modes generated by CLUMPAK for
426 *K*=2 and *K*=3 (Figs 1c, d). The STRUCTURE analysis revealed a clear genetic break between
427 the two subspecies, with the exception of one outlier sample from the Northern Territory
428 (identified in the PCoA; Fig. 1). Individuals from central Queensland were distinct when
429 using *K*=3 (Fig. 1d) and in the analysis based on *L. l. leadbeateri* samples only (Fig. S3a).
430 Similar to the case for the PCoA, this result is likely driven by the relatively high relatedness
431 between these central Queensland individuals. In the STRUCTURE analysis based on *L. l.*
432 *mollis* samples only, subtle population differentiation, although not robustly supported,
433 coincided with samples from the south-western wheatbelt region (Fig. S3b-c).
434
435 Genetic variability in the conStruct analysis was best explained using *K*=2-3 (Fig. S4). Some
436 isolation by distance is evident as the spatial model is preferred over the non-spatial model.
437 In the conStruct analysis using *K*=2, there is clear population differentiation between the two
438 subspecies (excepting the Northern Territory outlier sample identified above; this sample was
439 removed from subsequent analyses; Fig. S5a-b), corroborating the STRUCTURE analysis (Fig.
440 1c-d). There was slight variability in the admixture plots between different chains and
441 independent analyses, however the main patterns were consistent (we present one chain from
442 each independent analysis; Fig. S5a-b). Inadequate convergence and consistency between

443 chains/analyses when using K=3 indicated that the results were unreliable at this level of
444 parameterization.
445
446 Relatively low but significant genetic differentiation was evident between the two subspecies
447 (F_{ST} = 0.039; confidence interval: 0.035, 0.042). In the AMOVA based on the full dataset
448 (i.e. 56 individuals and 2131 SNPs), the proportion of genetic variation within individuals
449 was 69.8%. This is significantly lower than expected based on random permutations ($p <$
450 0.001). The proportion of genetic variation within and between subspecies (25.8% and 4.4%,
451 respectively) were, however, both greater than expected ($p < 0.001$) (Table S4 & Fig. S6).
452 These patterns are indicative of population structure, and not a single panmictic population.
453 In the PCoA, STRUCTURE, F_{ST} and AMOVA analyses, use of different SNP datasets (i.e.
454 SNPs based on only tissues, and SNPs that underwent more stringent filtering) exhibited very
455 similar results (Figs S6, S7, S8 & Tables S4, S5).
456
457 Ten haplotypes were observed from the 19 mtDNA samples that were sequenced (i.e. 2,037
458 bp of concatenated *ND2*, *ND4* and *ND5* genes; Table S7). The haplotype network analysis
459 based on mtDNA exhibited a star-like pattern (Fig. 2a). A central haplotype predominated,
460 while other haplotypes were connected by the common haplotype. The common central
461 haplotype comprises individuals from both subspecies from across the species range. The
462 mtDNA *Da* between subspecies was 0.004%. Overall, mtDNA structure did not reflect
463 patterns found in SNP clustering analyses.
464
465 *Gene flow patterns*
466 The inter-individual Mantel tests revealed significant IBD when analysing the full dataset and
467 when analysing only more contemporary frozen tissue samples (all $p < 0.001$) (Fig. S9).

However, inter-individual genetic distances were found to be relatively invariable (note the near-horizontal relationship between genetic and physical distance in Fig. S9a). Relatively low genetic distances across Australia indicate that differentiation among geographic locations is weak. Further, in some cases, spatial patterns inferred from Mantel tests are problematic (Legendre & Fortin, 2010; Legendre *et al.*, 2015). We did not consider mtDNA in this analysis, as mtDNA is known to produce unreliable IBD results (Teske *et al.*, 2018).

There was a reduction in gene flow between the 'sample clusters' spanning the putative subspecies along the transect (Fig. 3). Although the level of differentiation was relatively low, all pairwise F_{ST} estimates along the transect were significant except for one (between 'cluster 1' and '2'; see Fig. 3b). The conStruct analysis based on these 15 transect samples corroborated the other population structure analyses. Clear genetic differentiation was evident between the two subspecies (Fig. S5c-d). Although there was slight variability between the independent analysis and separate chains, the main population structure patterns were consistent.

Genetic diversity

Lophochroa leadbeateri mollis had the highest genetic diversity for all metrics, although not considerably higher than *L. l. leadbeateri* (Table 1). Genetic diversity measurements varied when using different SNP datasets, but were qualitatively consistent (Table S6). As expected, when applying more stringent filtering (including a MAF filter), the number of private alleles and allelic richness decreased. Without subspecies divisions, mtDNA nucleotide diversity was 0.0012 (Table S7); *ND2* was considerably more diverse than *ND4* and *ND5*.

493 *Population growth*

494 Analyses of 'randomness', 'neutrality', Tajima's D (-1.851), Fu's Fs (-4.865) and Ramos-
495 Onsin's R2 (0.052), were all significant ($p < 0.05$ in each case). The unimodal mismatch
496 distribution (with a high value at zero mismatches) of the mtDNA data also indicates the
497 occurrence of an expansion event (Figure S10; Rogers & Harpending, 1992). These results
498 are consistent with a scenario of rapid growth in population size.

499

500 *Phylogenetics*

501 The SNAPP model for which monophyly was not enforced received the highest support
502 (Table S8). AICM was relatively consistent between replicates, ranging from 16838.7 to
503 16846.6 for model one (monophyly not enforced), and from 16874.8 to 16879.1 for model 2
504 (monophyly enforced). The two subspecies each exhibited monophyly in the RAxML
505 analysis (excepting the one aforementioned outlier sample from Northern Territory), although
506 bootstrap support was relatively low (i.e. 73%; Figure S11). These results indicate that the
507 existence of two ESUs corresponding each of the two subspecies are not unambiguously
508 supported.

509

510 Similar to the haplotype network analysis, phylogenetic analysis of mtDNA did not
511 correspond to the SNP population structure results and did not exhibit any discernible
512 geographic patterns (Fig. 2b, S12).

513

514 *Wildlife forensics*

515 The initial DAPC used for SNP selection clearly separated the two subspecies (Fig. 4a), in
516 line with the other genetic structure analyses. We retained 35 principal components for this

analysis. The minimum number of SNPs required to separate the subspecies via DAPC was twenty (Fig. 4b). We considered adequate separation when all samples were correctly sorted into their corresponding subspecies clusters. We retained five principal components when performing the DAPC using twenty SNPs. When considering three populations (K=3), the central Queensland individuals formed a separate cluster having no overlap but only when ≥75 SNPs were utilized (Fig. S13). It should be noted, however, that this clustering is likely driven by the high relatedness between these central Queensland samples.

The GeneClass2 analyses correctly assigned all six individuals with high support. When assigning an individual to the correct subspecies (e.g. claiming a *L. l. leadbeateri* individual was *L. l. leadbeateri*), all likelihood ratios were >28.71, and averaged 1.81×10^{7} (Table S9). The likelihood ratios were higher when assigning *L. l. mollis* than when assigning *L. l. leadbeateri*, averaging 3.54×10^{8} and 8.23×10^{6} respectively. When assigning an individual to the incorrect subspecies (e.g. claiming an *L. l. leadbeateri* individual was an *L. l. mollis* individual), all likelihood ratios were $<3.48 \times 10^{-2}$, and averaged 5.81×10^{-3}.

Discussion

We have performed the first comprehensive phylogeographic study of one of Australia's most charismatic but relatively understudied parrots, the pink cockatoo. Our extensive data set revealed two major genetic clusters corresponding to the currently recognized subspecies, and an additional, divergent cluster comprising closely related Central Queensland members of *L. l. leadbeateri* (importantly, this cluster disappeared when only one representative was used). We use these results to reassess the species' conservation priorities and taxonomy, which are currently based on morphology.

Population structure

Lophochroa leadbeateri is a widespread species that does not have defined geographically disjunct population isolates. Our SNP data show consistent but relatively weak levels of genetic structure between the two currently recognized subspecies at the Eyrean Barrier. It is important to determine whether this result is derived from historical biogeography (i.e. the Eyrean Barrier) or sampling gaps (i.e. IBD) as has been highlighted by several authors (Latch *et al*., 2014; Bradburd *et al*., 2018; Chambers & Hillis, 2020). We found that genetic structure between the two subspecies based on SNPs was apparent even when accounting for geographic distance (Figs. 3 & S5). Contrastingly, distinct subspecies clusters were not apparent in the mtDNA analyses. This is possibly due to incomplete lineage sorting and/or higher female dispersal, and is consistent with the weak and/or recent phylogeographic structure across the continent inferred by the SNP analyses. Large effective population sizes retaining ancestral variation even after long periods of isolation and/or recent divergence times could potentially preclude signals of population divergence in mtDNA (Hartl *et al.*, 1997; Maddison, 1997).

The significant population expansion result, further evidenced by the star-like haplotype network (Fig. 2a), may have proliferated the frequency of a common haplotype and explain the absence of distinct geographically disjunct haplotype clusters. The common haplotype (see Fig. 2a) comprised individuals from across the species' range, including an individual from central Queensland (B28102) and individuals from south-west Western Australia (A35378, Z23813 and B53847), indicating that the species has the capacity to disperse over long-distances. However, the weak differentiation detected by SNPs indicates that the Eyrean barrier may have limited dispersal, similar to other vertebrate species found in this region (Neaves *et al*., 2012; McElroy *et al*., 2018).

Overall, these data suggest that the Eyrean barrier has been either a somewhat effective, although relatively recent biogeographic barrier to gene flow in this species, or a more long-term but porous barrier. The subtle morphological divergence between subspecies reported by Schodde (1994) is consistent with a relatively recent divergence time. Morphological differences can accumulate rapidly in bird taxa, often before mtDNA genetic divergence (Zink & Barrowclough, 2008; Safran *et al.*, 2016).

The weak substructure evident within each of the two subspecies is consistent with relatively regular gene flow between members of the four core breeding populations (Fig. 1a). In *L. l. leadbeateri*, the genetic differentiation we identified between individuals from central Queensland individuals and all other populations is likely an artefact of analysing related individuals. Although the relatively high relatedness between these individuals may be due to actual genetic structure in this region (i.e. higher levels of inbreeding in a genetically isolated population), it is more likely that individuals from a family unit were sampled. All five central Queensland individuals were collected in the same region, four of which were collected three days apart (while the other was collected ~3 years later), and the kinship analysis suggests these individuals could be second- and/or third-order relatives (Table S3). In *L. l. mollis*, there is limited genetic differentiation between the population in the south-westernmost 'wheatbelt' area and other populations (Fig. S3b). This population inhabits mulga shrubland, and was previously considered a separate subspecies (Peters, 1937). However, the genetic structure in this region is subtle and inconsistent; notably, some of the associated samples do have high levels of missing data. Analysing additional geographically intermediate samples may help clarify the presence of potential cryptic genetic diversity

591 within the two subspecies, and hence elucidate management strategies to conserve their
592 genetic variation.
593
594 The pink cockatoo's shallow phylogeographic structure across its range corresponds to that
595 seen in some other Australian arid zone bird species (Joseph & Wilke, 2006; Dolman &
596 Joseph, 2015). Engelhard *et al.* (2015), for example, found mtDNA genetic structure, albeit
597 weak, in another cockatoo in the same subfamily (Cacatuinae), the galah (*Eolophus*
598 *roseicapilla*). However, there are numerous examples of similarly distributed bird species
599 that do exhibit more marked genetic differentiation across much the same range, such as the
600 copper-backed and chestnut quail-thrush (*Cinclosoma clarum* and *C. castanotum*,
601 respectively), the white-eared honeyeater (*Nesoptilotis leucotis*), the splendid fairy-wren
602 (*Malurus splendens*), and the Australian ringneck (*Barnardius zonarius*) (Joseph & Wilke,
603 2006; Kearns *et al.*, 2009; Dolman & Joseph, 2015, 2016). We recently found evolutionarily
604 distinct isolates within arid zone populations of another inland cockatoo species, the red-
605 tailed black-cockatoo *Calyptorhynchus banksii*. In that case the southwestern 'wheatbelt'
606 population was found to be genetically and taxonomically distinct (Ewart *et al.*, 2020).
607 Varying responses to biogeographical barriers among the pink cockatoo and these other arid
608 bird taxa are likely due to differences in habitat specificity and vagility (Toon *et al.*, 2007).
609

610 ## *Conservation implications*

611 Robust delineation of conservation units is vital for effective conservation prioritization.
612 Conservation units can be apportioned as either management units (i.e. a demographically
613 independent unit of genetic variation; Moritz, 1994; Palsbøll *et al.*, 2007) or evolutionarily
614 significant units (i.e. independently evolving units of genetic variation; Ryder, 1986; Moritz,
615 1994). Based on the genetic structure results presented above, the two subspecies should be

616 considered separate management units. Given the lack of support for two evolutionarily
617 distinct clades (i.e. they do not exhibit reciprocal monophyly) in the phylogenetic analysis
618 based on nuclear SNPs, the low F_{ST} values, and the lack of mtDNA support, these
619 conservation units do not appear to constitute separate ESUs.
620
621 Assessing population fragmentation within each of the two subspecies is critical, as small
622 isolated populations often suffer from genetic erosion (Frankham *et al*., 2017). The additional
623 substructure we identified in central Queensland could indicate that this population is at risk
624 of genetic isolation, although it is likely that the genetic differentiation detected in this region
625 is likely driven by high relatedness among the samples examined (see above). Denser
626 sampling of unrelated individuals, and geographically wider sampling to fill gaps in this
627 study should be implemented to clarify the genetic structure in this region and determine
628 whether or not it should be regarded as separate management unit.
629
630 *Taxonomic reassessment*
631 Incorrect delineation of subspecies can misguide subsequent studies and conservation
632 strategies (Zink, 2004; Braby *et al.*, 2012; Huang & Knowles, 2016). Typically, different
633 subspecies exhibit at least some mtDNA phylogenetic resolution (e.g. Kearns *et al.*, 2015,
634 2016). Net divergence, *Da*, at the mtDNA *ND2* gene between the two nominal pink cockatoo
635 subspecies was only 0.009%. In several other avian species that exhibit ND2 differentiation
636 at the Eyrean Barrier, the value is much higher. Examples include the white-eared honeyeater
637 (2.23%; Dolman & Joseph, 2015), the mulga parrot subspecies (1.92%; McElroy *et al*.,
638 2018), and the Australian ringneck (1.72%; Joseph & Wilke, 2006). Accordingly, the
639 minimal mtDNA differentiation may be taken to suggest that the species is monotypic (i.e.,
640 no subspecies). Conversely, a lack of mtDNA-based subspecies divergence does not

641 necessarily justify/dictate taxonomic modifications (Ball & Avise, 1992; Funk & Omland,
642 2003; Omland *et al.*, 2006). Traits other than genetics and morphology, including
643 vocalizations, ecological characteristics, and frequency of subspecies hybrids, can be taken
644 into account (Remsen, 2005; also see Ford & Parker, 1973). Therefore, although they may
645 not be evolutionary distinct genetically (i.e. they may not represent separate ESUs), we
646 advocate continued recognition of two subspecies within the pink cockatoo.
647
648 *Wildlife forensics implications*
649 The generation of SNP data and the population genetic inferences presented in this study
650 could facilitate the development of wildlife forensic techniques for the pink cockatoo (Ogden,
651 2011). Typically, a species or subspecies identification test is based on analysis of mtDNA
652 due its high mutation rate, lack of recombination, availability of homologous reference data,
653 and the ease with which it is amplified and sequenced (Linacre & Tobe, 2011; Johnson *et al.*,
654 2014). However, the lack of reciprocal monophyly of subspecies/populations in our analyses
655 of mtDNA loci means they may not be suitable for performing a subspecies identification or
656 geographic provenance tests. Any forensic testing of pink cockatoo subspecies should
657 therefore rely on nuclear DNA markers. We have provided proof of concept that reliable
658 population identification testing can be performed in this species using as few as 20 SNPs (all
659 likelihood ratios were >28 when the prosecution hypothesis was correct). Including more
660 SNPs and samples would intuitively yield greater assignment power and confidence.
661 Furthermore, different SNPs could be selected that are more informative to identify
662 individuals in certain subregions. Being able to identify source populations will help direct
663 enforcement and compliance resources to areas most vulnerable to illegal collection and
664 allow repatriation of seized animals to their subspecies/population of origin (Alacs &
665 Georges, 2008). This study serves as an example on how to construct subspecies/population

identification or geographic provenance tests for species with relatively shallow phylogenetic structure.

Additionally, the SNPs generated in this paper could be utilized in the development of parentage testing. A parentage test of offspring along with their putative mother and/or father could determine in an investigation whether a pink cockatoo is wild or captive-bred. SNPs with a high MAF are particularly useful for parentage analysis (Andrews *et al.*, 2018); even after filtering for missing data, locus quality, outliers, HWE, and linkage, the dataset contained 176 SNPs with a MAF >0.4 (data not shown).

Benefits and caveats of this genome-wide SNP data set

This study serves as another example of genome-wide SNP data being able to resolve populations where mtDNA and/or relatively few nuclear markers lacked resolution or were misleading (Leslie & Morin, 2016; Rodríguez-Ezpeleta *et al.*, 2016; Younger *et al.*, 2017). Utilizing many genetic markers alleviates issues faced when basing important conservation decisions and/or taxonomy on a small number of markers. Further, we have harnessed advancements in museum genomics to successfully genotype numerous old museum specimens (toe pads) collected over decades with reasonably high success (61% of toe pads were successfully genotyped), which provided critical representation of the species' distribution (Fig. 1a).

Utilizing SNPs generated from old museum specimens, however, presents a number of issues. Even after filtering, some toe pad genotypes had high SNP error rates and considerable missing data (Table S2). Although some troublesome genotypes were removed from most analyses (Appendix I), given the variability of the error rates of toe pad samples,

some samples with relatively high error rates may not have been excluded during the filtering stages. The error and missing data in these old museum specimens are expected to be biased towards low diversity SNPs, and random errors are expected to homogenize genetic structure (Ewart *et al.*, 2019). Further, even though the pink cockatoo is a very long-lived species (Brouwer *et al.*, 2000), it is possible that the genetic structure may have changed over the sampling period (i.e. genotyped individuals were sampled between 1912 and 2011). To ensure these biases did not lead to false population genetic inferences, we demonstrated that comparable results were obtained when analysing a SNP dataset called exclusively from more contemporary tissue samples. These results corroborate the value of using toe pad samples genotyped with this platform to support spatial conclusions, but may present problems for temporal inferences (Ewart *et al.*, 2019).

Conclusion

This is the first species-wide genetic study on the pink cockatoo. The extensive dataset provides a basis for effective conservation management for this species. Well-informed management strategies based on genetics can now be implemented with the aim of maximising the species' genetic diversity and its potential to adapt to changing environments (Frankham, 2005; Huffman & Wallace, 2011). Further, preliminary analyses with these data indicate they could be successfully used in the development of a wildlife forensic toolbox to detect and prosecute trafficking crimes associated with this species.

The phylogeographic analyses we have performed represent a robust approach for investigating species that are widespread, yet have shallow phylogeographic structure. SNPs revealed subtle patterns of genetic differentiation that were not detected through analysis of mtDNA and morphology.

Acknowledgements

We are grateful to the University of Sydney Industry and Community Engagement Seed Fund, Birdlife Australia (Stuart Leslie Bird Research Award), Australian Museum Foundation, Australian Government Research Training Program (RTP) Scholarship and University of Sydney Merit Award, and Michael Sverns, Glenn Sharp, and Stephen Lavelle from the Department of Environment, Land, Water and Planning (State Government of Victoria) for funding. We thank the following people/organizations for their help in obtaining samples: Robert Palmer (Australian National Wildlife Collection), Scott Ginn, Leah Tsang and Richard Major (Australian Museum), Ron Johnstone and Rebecca Bray (Western Australian Museum), Claire Keely (Museum Victoria), and John Martin (Royal Botanic Gardens, Sydney). We also thank Frankie Sitam and Jeffrine Rovie-Ryan for providing access and training at the National Wildlife Forensic Laboratory (Kuala Lumpur, Malaysia), Linda Neaves for her assistance with the mtDNA PCRs and sequencing, Kate FitzGerald for her assistance with the graphics, and Simon Ho and Mark Eldridge for their expert advice.

Collection of pink cockatoos for research purposes of this project was approved by the Australian Museum Animal Care and Ethics Committee (approval number 16-02). All earlier samples had been collected under all appropriate ethics and scientific collecting permits. Sulphur-crested cockatoo samples were obtained as part of another project, under the ethics permit AEC 151020-0. Transfer of pink cockatoo DNA from Australia to Malaysia was authorized under CITES permit number PWS2018-AU-000019.

Conflict of interest: none

Data accessibility

SNP data and mtDNA sequence data will be made available on Dryad Digital Repository and GenBank respectively upon acceptance.

References

Adamack AT, Gruber B. 2014. PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution* 5: 384-387.

Adams M, Baverstock PR, Saunders DR, Schodde R, Smith GT. 1984. Biochemical systematics of the Australian cockatoos (Psittaciformes: Cacatuinae). *Australian Journal of Zoology* 32: 363-377.

Alacs E, Georges A. 2008. Wildlife across our borders: a review of the illegal trade in Australia. *Australian Journal of Forensic Sciences* 40: 147-160.

Andrews KR, Adams JR, Cassirer EF, Plowright RK, Gardner C, Dwire M, Hohenlohe PA, Waits LP. 2018. A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RAD seq data. *Molecular Ecology Resources* 18: 1263-1281.

Ansari MH, Cooper SJB, Schwarz MP, Ebrahimi M, Dolman G, Saint KM, Donnellan SC, Bull CM, Gardner MG. 2019. Plio-Pleistocene diversification and biogeographic barriers in southern Australia reflected in the phylogeography of a widespread and common lizard species. *Molecular Phylogenetics and Evolution* 133: 107-119.

Antao T, Lopes A, Lopes RJ., Beja-Pereira A, Luikart G. 2008. LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics* 9: 323.

Archer FI, Adams PE, Schneiders BB. 2017. stratag: an r package for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources* 17: 5-11.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution* 29: 2157-2167.

772 Ball Jr RM, Avise JC. 1992. Mitochondrial DNA phylogeographic differentiation among
773 avian populations and the evolutionary significance of subspecies. *The Auk* 109: 626-
774 636.
775 Baumsteiger J, Moyle PB, Aguilar A, O'Rourke SM, Miller MR. 2017. Genomics clarifies
776 taxonomic boundaries in a difficult species complex. *PloS One* 12: e0189417.
777 Beaumont MA, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of
778 population structure. *Proceedings of the Royal Society of London B* 263: 1619-1626.
779 Biodiversity Conservation Act, 2016. NSW State Government, Office of Environment and
780 Heritage. Available at:
781 https://www.environment.nsw.gov.au/threatenedspeciesapp/profile.aspx?id=10116
782 Blakers M, Davies SJJF, Reilly PN. 1984. *The atlas of Australian birds*. Melbourne:
783 Melbourne University Press.
784 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A,
785 Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary
786 analysis. *PLoS Computational Biology* 10: e1003537.
787 Braby MF, Eastwood R, Murray N. 2012. The subspecies concept in butterflies: has its
788 application in taxonomy and conservation biology outlived its usefulness?. *Biological*
789 *Journal of the Linnean Society* 106: 699-716.
790 Bradburd GS, Coop GM, Ralph PL. 2018. Inferring continuous and discrete population
791 genetic structure across space. *Genetics* 210: 33-52.
792 Brouwer K, Jones ML, King CE, Schifter H. 2000. Longevity records for Psittaciformes in
793 captivity. *International Zoo Yearbook* 37: 299-316.
794 Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. 2012. Inferring
795 species trees directly from biallelic genetic markers: bypassing gene trees in a full
796 coalescent analysis. *Molecular Biology and Evolution* 29: 1917-1932.
797 Cameron M. 2007. *Cockatoos*. Melbourne: CSIRO Publishing.
798 Chambers EA, Hillis DM. 2020. The multispecies coalescent over-splits species in the case of
799 geographically widespread taxa. *Systematic Biology* 69: 184-193.
800 Chessel D, Dufour AB, Thioulouse J. 2004. The ade4 package-I-One-table methods. *R News*
801 4: 5-10.
802 Chhatre VE, Emerson KJ. 2017. StrAuto: Automation and parallelization of STRUCTURE
803 analysis. *BMC Bioinformatics* 18: 192.
804 Clement M, Posada DCKA, Crandall KA. 2000. TCS: a computer program to estimate gene
805 genealogies. *Molecular Ecology* 9: 1657-1659.

806 Condon HT. 1975. *Checklist of the Birds of Australia: Non-passerines (Vol. 1*). Melbourne:
807 Royal Australasian Ornithologists Union.
808 Cruz VM, Kilian A, Dierig DA. 2013. Development of DArT marker platforms and genetic
809 diversity assessment of the U.S. collection of the new oilseed crop lesquerella and
810 related species. *PLoS ONE* 8: e64062.
811 Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle
812 genomes from whole genome data. *Nucleic Acids Research* 45: e18.
813 Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, Telles MPDC, Rangel
814 TF, Bini LM. 2013. Mantel test in population genetics. *Genetics and Molecular*
815 *Biology* 36: 475-485.
816 Dolman G, Joseph L. 2012. A species assemblage approach to comparative phylogeography of
817 birds in southern Australia. *Ecology and Evolution* 2: 354-369.
818 Dolman G, Joseph L. 2015. Evolutionary history of birds across southern Australia: structure,
819 history and taxonomic implications of mitochondrial DNA diversity in an ecologically
820 diverse suite of species. *Emu* 115: 35-48.
821 Dolman G, Joseph L. 2016. Multi-locus sequence data reveal Pleistocene speciation in semi-
822 arid southern Australian birds (Cinclosoma spp.) was associated with increased genetic
823 drift. *BMC Evolutionary Biology* 16: 226.
824 Eberhard JR, Wright TF. 2016. Rearrangement and evolution of mitochondrial genomes in
825 parrots. *Molecular Phylogenetics and Evolution* 94 (Part A): 34-46.
826 Engelhard D, Joseph L, Toon A, Pedler L, Wilke T. 2015. Rise (and demise?) of subspecies in
827 the Galah (*Eolophus roseicapilla*), a widespread and abundant Australian cockatoo.
828 *Emu-Austral Ornithology* 115: 289-301.
829 Ewart KM, Johnson RN, Ogden R, Joseph L, Frankham GJ, Lo N. 2019. Museum specimens
830 provide reliable SNP data for population genomic analysis of a widely distributed but
831 threatened cockatoo species. *Molecular Ecology Resources* 00: 1-15.
832 Ewart KM, Johnson RN, Ogden R, Joseph L, Ho SYW, Frankham G, Eldridge MDB, Lo N.
833 2020. Genome-wide SNP analyses identify novel evolutionarily significant units in an
834 iconic Australian bird species, the red-tailed black-cockatoo (*Calyptorhynchus*
835 *banksii*). *Heredity* 1-16.
836 Excoffier L, Lischer HEL. 2010. Arlequin suite version 3.5: A new series of programs to
837 perform population genetics analyses under Linux and Windows. *Molecular Ecology*
838 *Resources* 10: 564-567.

839 Fenster CB, Ballou JD, Dudash MR, Eldridge MD, Frankham R, Lacy RC, Ralls K, Sunnucks
840 P. 2018. Focus: ecology and evolution: conservation and genetics. *The Yale Journal of*
841 *Biology and Medicine* 91: 491.
842 Flora and Fauna Guarantee Act, 1988. Victorian State Government, Department of
843 Environment, Land, Water & Planning. Available at:
844 https://www.environment.vic.gov.au/__data/assets/pdf_file/0024/115827/FFG-
845 Threatened-List.doc.pdf
846 Ford J. 1974. Speciation in Australian birds adapted to arid habitats. *Emu* 74: 161-168.
847 Ford J. 1987. Minor isolates and minor geographical barriers in avian speciation in continental
848 Australia. *Emu* 87: 90-102.
849 Ford J, Parker SA. 1973. A second species of wedgebill? *Emu* 73: 113-118.
850 Forshaw JM. 2011. *Parrots of the world*. Melbourne: CSIRO Publishing.
851 Forshaw JM, Cooper WT. 1981. *Australian Parrots. Second edition*. Melbourne: Lansdowne
852 Press.
853 Forshaw JM, Cooper WT. 2002. *Australian Parrots. Third edition*. Melbourne: Lansdowne
854 Press.
855 Frankham R. 2005. Genetics and extinction. *Biological Conservation* 126: 131-140.
856 Frankham R, Ballou JD, Briscoe DA. 2010. *Introduction to conservation genetics*. *Second*
857 *edition*. Cambridge: Cambridge University Press.
858 Frankham R, Ballou JD, Ralls K, Eldridge MDB, Dudash MR, Fenster CB, Lacy RC,
859 Sunnucks P. 2017. *Genetic management of fragmented animal and plant populations*.
860 Oxford: Oxford University Press.
861 Fu Y-X. 1997. Statistical tests of neutrality of mutations against population growth,
862 hitchhiking and background selection. *Genetics* 147: 915-925.
863 Funk DJ, Omland KE. 2003. Species-level paraphyly and polyphyly: frequency, causes, and
864 consequences, with insights from animal mitochondrial DNA. *Annual Review of*
865 *Ecology, Evolution, and Systematics* 34: 397-423.
866 Garnett ST, Szabo JK, Dutson G. 2011. *The action plan for Australian birds 2010*. Melbourne:
867 CSIRO Publishing.
868 Goudet J, Jombart T. 2015. hierfstat: Estimation and Tests of Hierarchical F-Statistics. R
869 package version 0.04-22.
870 Gruber B, Unmack PJ, Berry OF, Georges A. 2018. dartr: An r package to facilitate analysis
871 of SNP data generated from reduced representation genome sequencing. *Molecular*
872 *Ecology Resources* 18: 691-699.

873 Hall BP. 1974. *Birds of the Harold Hall Australian Expeditions: 1962-1970*. Trustees of the
874 British Museum of Natural History, London.
875 Hartl DL, Clark AG, Clark AG. 1997. *Principles of population genetics, Vol. 116*. Sunderland:
876 Sinauer Associates.
877 Higgins PJ. 1999. *Handbook of Australian, New Zealand & Antarctic Birds. Vol. 4, Parrots to*
878 *Dollarbird*. Oxford: Oxford University Press.
879 Huang JP, Knowles LL. 2016. The species versus subspecies conundrum: Quantitative
880 delimitation from integrating multiple data types within a single Bayesian approach in
881 hercules beetles. *Systematic Biology* 65: 685-699.
882 Huffman JE, Wallace JR. 2011. *Wildlife forensics: methods and applications*. Hoboken: John
883 Wiley & Sons.
884 Johnson RN, Wilson-Wilde L, Linacre A. 2014. Current and future directions of DNA in
885 wildlife forensic science. *Forensic Science International: Genetics* 10: 1-11.
886 Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers.
887 *Bioinformatics* 24: 1403-1405.
888 Joseph L, Wilke T. 2006. Molecular resolution of population history, systematics and
889 historical biogeography of the Australian ringneck parrots *Barnardius*: are we there
890 yet?. *Emu-Austral Ornithology* 106: 49-62.
891 Joseph L, Wilke T, Ten Have J, Terry Chesser R. 2006. Implications of mitochondrial DNA
892 polyphyly in two ecologically undifferentiated but morphologically distinct migratory
893 birds, the masked and white-browed woodswallows Artamus spp. of inland Australia.
894 *Journal of Avian Biology* 37: 625-636.
895 Kamvar ZN, Brooks JC, Grünwald NJ. 2015. Novel R tools for analysis of genome-wide
896 population genetic data with emphasis on clonality. *Frontiers in Genetics* 6: 208.
897 Kamvar ZN, Tabima JF, Grünwald NJ. 2014. poppr: an R pack- age for genetic analysis of
898 populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2: e281.
899 Kearns A, Joseph L, Double M, Edwards S. 2009. Inferring the phylogeography and
900 evolutionary history of the Splendid Fairy-wren (*Malurus splendens*) from
901 mitochondrial DNA and spectrophotometry. *Journal of Avian Biology* 40: 7-17.
902 Kearns AM, Joseph L, White LC, Austin JJ, Baker C, Driskell AC, Malloy JF, Omland KE.
903 2016. Norfolk Island Robins are a distinct endangered species: ancient DNA unlocks
904 surprising relationships and phenotypic discordance within the Australo-Pacific
905 Robins. *Conservation Genetics* 17: 321-335.

Kearns AM, White LC, Austin JJ, Omland KE. 2015. Distinctiveness of Pacific Robin subspecies in Vanuatu revealed from disparate patterns of sexual dichromatism, plumage colouration, morphometrics and ancient DNA. *Emu-Austral Ornithology* 115: 89-98.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647-1649.

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, ... Aschenbrenner-Kilian M. 2012. Diversity arrays technology: a generic genome profiling technology on open platforms. In Data production and analysis in population genomics. Totowa: Humana Press, 67-89.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* 15: 1179-1191.

Latch EK, Reding DM, Heffelfinger JR, Alcalá-Galván CH, Rhodes Jr OE. 2014. Range-wide analysis of genetic structure in a widespread, highly mobile species (*Odocoileus hemionus*) reveals the importance of historical biogeography. *Molecular Ecology* 23: 3171-3190.

Legendre P, Fortin MJ. 2010. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources* 10: 831-844.

Legendre P, Fortin MJ, Borcard D. 2015. Should the Mantel test be used in spatial analysis?. *Methods in Ecology and Evolution* 6: 1239-1247.

Leigh JW, Bryant D. 2015. popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* 6: 1110-1116.

Leslie MS, Morin PA. 2016. Using genome-wide SNPs to detect structure in high-diversity and low-divergence populations of severely impacted eastern tropical Pacific spinner (*Stenella longirostris*) and pantropical spotted dolphins (*S. attenuata*). *Frontiers in Marine Science* 3: 253.

Li YL, Liu JX. 2018. StructureSelector: A web based software to select and visualize the optimal number of clusters using multiple methods. *Molecular Ecology Resources* 18: 176–177.

939 Linacre A, Tobe SS. 2011. An overview to the investigative approach to species testing in
940 wildlife forensic science. *Investigative Genetics* 2: 2.
941 Mackowski CM. 1984. The ontogeny of hollows in blackbutt (*Eucalyptus pilularis*) and its
942 relevance to the management of forests for possums, gliders and timber. In: Smith AP,
943 Hume ID, eds. *Possums and Gliders*. Sydney: Australian Mammal Society and Surrey
944 Beatty and Sons, 553-67.
945 Maddison WP. 1997. Gene trees in species trees. *Systematic Biology* 46: 523-536.
946 Marie AD, Stockwell BL, Rico C. 2019. DNA analysis of juvenile scalloped hammerhead
947 sharks *Sphyrna lewini* (Griffith & Smith 1834) reveals multiple breeding populations
948 and signs of adaptive divergence in the South Pacific. *Frontiers in Marine Science* 6:
949 718.
950 Mastrantonio V, Porretta D, Urbanelli S, Crasta G, Nascetti G. 2016. Dynamics of mtDNA
951 introgression during species range expansion: insights from an experimental
952 longitudinal study. *Scientific Reports* 6: 30355.
953 Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Piñero D, Emerson BC. 2015.
954 Restriction site-associated DNA sequencing, genotyping error estimation and de novo
955 assembly optimization for population genetic inference. *Molecular Ecology Resources*
956 15: 28–41.
957 Mathews GM. 1912. *The Austral Avian Record: A Scientific Journal Devoted Primarily to the*
958 *Study of the Australian Avifauna (Volume 1)*. London: Witherby & Company.
959 McElroy K, Beattie K, Symonds MR, Joseph L. 2018. Mitogenomic and nuclear diversity in
960 the Mulga Parrot of the Australian arid zone: cryptic subspecies and tests for selection.
961 *Emu-Austral Ornithology* 118: 22-35.
962 Melville J, Haines ML, Boysen K, Hodkinson L, Kilian A, Smith Date KL, … Parris KM.
963 2017. Identifying hybridization and admixture using SNPs: application of the DArTseq
964 platform in phylogeographic research on vertebrates. *Royal Society Open Science* 4:
965 161061.
966 Menkhorst P, Rogers D, Clarke R, Davies J, Marsack P, Franklin K. 2017. *The Australian bird*
967 *guide*. Melbourne: CSIRO Publishing, 260.
968 Moritz C. 1994. Defining "evolutionarily significant units" for conservation. Trends in
969 *Ecology and Evolution* 9: 373–375.
970 Nature Conservation (Wildlife) Regulation, 2006. Queensland State Government. Available
971 at: https://www.legislation.qld.gov.au/view/pdf/inforce/2010-05-21/sl-2006-0206.

Neaves LE, Zenger KR, Prince RIT, Eldridge MDB, Cooper DW. 2009. Landscape discontinuities influence gene flow and genetic structure in a large, vagile Australian mammal, *Macropus fuliginosus*. *Molecular Ecology* 18: 3363–78.

Neaves LE, Zenger KR, Prince RI, Eldridge MD. 2012. Impact of Pleistocene aridity oscillations on the population history of a widespread, vagile Australian mammal, *Macropus fuliginosus*. *Journal of Biogeography* 39: 1545-1563.

Ogden R. 2011. Unlocking the potential of genomic technologies for wildlife forensics. *Molecular Ecology Resources* 11: 109-116.

Ogden R, Thorpe RS. 2002. Molecular evidence for ecological speciation in tropical habitats. *Proceedings of the National Academy of Sciences* 99: 13612-13615.

Omland KE, Baker JM, Peters JL. 2006. Genetic signatures of intermediate divergence: population history of Old and New World Holarctic ravens (*Corvus corax*). *Molecular Ecology* 15: 795-808.

Paetkau D, Calvert W, Stirling I, Strobeck C. 1995. Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* 4: 347-354.

Palsbøll PJ, Berube M, Allendorf FW. 2007. Identification of management units using population genetic data. *Trends in Ecology and Evolution* 22: 11-16.

Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26: 419-420.

Peakall R, Smouse PE. 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Peakall R, Smouse PE. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537-2539.

Peters JL. 1937. *Check-list of birds of the world, volume III*. Cambridge: Harvard University Press, 176.

Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A. 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity* 95: 536-539.

Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu M. 2013. An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics* 4: 98.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

1005 Ralls K, Ballou JD, Dudash MR, Eldridge MD, Fenster CB, Lacy RC, Sunnucks P, Frankham
1006 R. 2018. Call for a paradigm shift in the genetic management of fragmented
1007 populations. *Conservation Letters* 11: e12412.
1008 Rambaut A. 2009. FigTree: tree figure drawing tool version 1.4.2. Available at:
1009 http://tree.bio.ed.ac.uk/software/figtree.
1010 Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6. Available at:
1011 http://beast.bio.ed.ac.uk/Tracer.
1012 Ramos-Onsins SE, Rozas J. 2002. Statistical properties of new neutrality tests against
1013 population growth. *Molecular Biology and Evolution* 19: 2092-2100.
1014 Remsen Jr JV. 2005. Pattern, process, and rigor meet classification. *The Auk* 122: 403-413.
1015 Rodríguez-Ezpeleta N, Bradbury IR, Mendibil I, Álvarez P, Cotano U, Irigoien X. 2016.
1016 Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP
1017 markers: effects of sequence clustering parameters and hierarchical SNP selection.
1018 *Molecular Ecology Resources* 16: 991-1001.
1019 Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of
1020 pairwise genetic-differences. *Molecular Biology and Evolution* 9: 552-569.
1021 Rowe KC, Singhal S, Macmanes MD, Ayroles JF, Morelli TL, Rubidge EM, Bi KE, Moritz
1022 CC. 2011. Museum genomics: low-cost and high-accuracy genetic data from historical
1023 specimens. *Molecular Ecology Resources* 11: 1082-1092.
1024 Rowley I, Chapman G. 1991. The breeding biology, food, social-organization, demography
1025 and conservation of the Major Mitchell or pink cockatoo, *Cacatua-Leadbeateri*, on the
1026 margin of the Western Australian wheat-belt. *Australian Journal of Zoology 39*: 211-
1027 261.
1028 Rozas J, Ferrer-Mata A, Sanchez DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE,
1029 Sánchez-Gracia A. 2017. DnaSP v6: DNA sequence polymorphism analysis of large
1030 datasets. *Molecular Biology and Evolution* 34: 3299-3302.
1031 Ryder OA. 1986. Species conservation and systematics: The dilemma of subspecies. *Trends in*
1032 *Ecology and Evolution* 1: 9-10.
1033 Safran RJ, Scordato ESC, Wilkins MR, Hubbard JK, Jenkins BR, Albrecht T, ... Nosil P.
1034 2016. Genome-wide differentiation in closely related populations: the roles of selection
1035 and geographic isolation. *Molecular Ecology* 25: 3865-3883.

1036 Schirtzinger EE, Tavares ES, Gonzales LA, Eberhard JR, Miyaki CY, Sanchez JJ, … Wright
1037 TF. 2012. Multiple independent origins of mitochondrial control region duplications in
1038 the order Psittaciformes. *Molecular Phylogenetics and Evolution* 64: 342-356.
1039 Schodde R. 1982. Origin, adaptation and evolution of birds in arid Australia. In: Barker WR,
1040 Greenslade PJM, eds. *Evolution of flora and fauna of arid Australia*. Adelaide:
1041 Peacock Publications, 191-224.
1042 Schodde R. 1994. The bird fauna of western New South Wales: geography and status. In:
1043 Lunney D, Hand D, Reed P, Butcher D, eds. *Future of the fauna of western New South*
1044 *Wales*. Mosman: Royal Zoological Society of NSW, 107-121.
1045 Schodde R. 1997. Cacatuidae. In: Houston WWK, Wells A, eds. *Zoological Catalogue of*
1046 *Australia. Vol. 37.2: Aves (Columbidae to Coraciidae).* Melbourne: CSIRO
1047 Publishing, 64-108.
1048 Schodde R, Mason IJ. 1999. *Directory of Australian birds: passerines: Passerines*.
1049 Melbourne: CSIRO publishing.
1050 Shirk AJ, Landguth EL, Cushman SA. 2017. A comparison of individual-based genetic
1051 distance metrics for landscape genetics. *Molecular Ecology Resources* 17: 1308-1317.
1052 Sorenson MD. 2003. Avian mtDNA Primers. Available at: http://people.bu.edu/msoren/
1053 primers.html.
1054 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1055 large phylogenies. *Bioinformatics* 30: 1312-1313.
1056 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA
1057 polymorphism. *Genetics* 123: 585-595.
1058 Teske PR, Golla TR, Sandoval-Castillo J, Emami-Khoyi A, van der Lingen CD, von der
1059 Heyden S, … Beheregaray LB. 2018. Mitochondrial DNA is unsuitable to test for
1060 isolation by distance. *Scientific Reports* 8: 8448.
1061 Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance
1062 in animals. *Molecular Ecology* 21: 3907-3930.
1063 Tonzo V, Papadopoulou A, Ortego J. 2019. Genomic data reveal deep genetic structure but no
1064 support for current taxonomic designation in a grasshopper species complex.
1065 *Molecular Ecology* 28: 3869-3886.
1066 Toon A, Mather PB, Baker AM, Durrant KL, Hughes JM. 2007. Pleistocene refugia in an arid
1067 landscape: analysis of a widely distributed Australian passerine. *Molecular Ecology*
1068 16: 2525-2541.

Walker IS, Sluiter IRK, Hawker P. 1999. Flora and fauna guarantee action statement, Major Mitchell's Cockatoo *Cacatua leadbeateri*. Department of Sustainability and Environment. Available at: https://www.environment.vic.gov.au/__data/assets/pdf_file/0025/32875/Major_Mitchells_Cockato_Cacatua_leadbeateri.pdf

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.

White NE, Phillips MJ, Gilbert MTP, Alfaro-Núñez A, Willerslev E, Mawson PR, Spencer PB, Bunce M. 2011. The evolutionary history of cockatoos (Aves: Psittaciformes: Cacatuidae*). Molecular Phylogenetics and Evolution* 59: 615-622.

Willing EM, Dreyer C, van Oosterhout C. 2012. Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLOS ONE* 7: e42649.

Younger JL, Clucas GV, Kao D, Rogers AD, Gharbi K, Hart T, Miller KJ. 2017. The challenges of detecting subtle population structure and its importance for the conservation of emperor penguins. *Molecular Ecology* 26: 3883-3897.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28: 3326-3328.

Zink RM. 2004. The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271: 561-564.

Zink RM, Barrowclough GF. 2008. Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* 17: 2107-2121.

Figure and table legends

Figure 1. (a) The distribution of *Lophochroa leabeateri leadbeateri* (blue) and *L. l. mollis* (orange) in Australia, adapted from Schodde (1994) and Menkhorst et al. (2017), and localities of the frozen tissue samples (stars) and toe pads samples (circles) genotyped in this study. The thick grey line represents the Eyrean Barrier, the darker shading represents core breeding zones, and the lighter shading and blurred fringes represent areas of potentially sparser distribution and/or non-breeding based on records from the Atlas of Living Australia database (https://www.ala.org.au; accessed 4 November 2020). (b) A PCoA plot for 57 pink cockatoo individuals using 4,135 SNPs. (c) and (d) STRUCTURE plots for 57 pink cockatoo individuals based on 2,131 SNPs when K=2 and K=3, respectively. The bottom-left photo is of *L. l. leadbeateri*, Mt. Hope, NSW. Photo: Corey Callaghan.

Figure 2. (a) TCS-based haplotype network analysis based on 19 pink cockatoo individuals using 2,037 bp of concatenated *ND2*, *ND4* and *ND5* genes. (b) Phylogeny of the pink cockatoo based on mtDNA data (see Table S1 for samples details), isolated from Fig. S12a (outgroup removed for clarity). Bayesian posterior probabilities are given above relevant branches. The 'CQ' and 'SW' labels next to the haplotypes (a) and taxon names (b) represent samples from central Queensland and south-west Western Australia respectively (see Fig. S3 for additional details). NB: the common haplotype in the haplotype network (a) contains haplotypes from both *L. l. leadbeateri* populations and the south-western Western Australia *L. l. mollis* population, but not the more north-easterly *L. l. mollis* population.

Figure 3. (a) Genetic divergence of populations along a transect based on inter-population pairwise $F_{ST}/(1 - F_{ST})$ calculated using 2,131 SNPs, divided by pairwise geographic distance, and plotted against the midpair distance of adjacent localities. (b) For this analysis, 15 pink cockatoo individuals were divided into five 'sample clusters' (3 individuals per cluster) along a transect. The vertical dotted red line in (a) indicates the pairwise comparison across the putative subspecies barrier.

Table 1. Genetic diversity measurements based on 2,131 SNPs in 56 pink cockatoo individuals, and 2,037 bp of concatenated *ND2*, *ND4* and *ND5* genes in 19 pink cockatoo individuals. Genetic diversity was measured within subspecies. Note, the haplotype common to both subspecies (see Fig. 2a) was counted twice in the 'number of haplotypes.'

Figure 4. DAPC analyses showing separation between *Lophochroa leadbeateri mollis* (orange) and *L. l. leadbeateri* (blue). The analyses were based on 49 pink cockatoo individuals using (a) 1,307 SNPs, and (b) 20 informative SNPs.