Validating the use of non-invasively sourced DNA for population genetic studies using pedigree data

Faye Wedrowicz[1,2]**, Jennifer Mosse**[2]**, Wendy Wright**[2]**, and Fiona E. Hogan**[2]

[1]Faculty of Science, Monash University, Melbourne, Australia
[2]School of Applied and Biomedical Sciences, Federation University Australia,
Churchill, Victoria 3842, Australia

Correspondence to: Faye Wedrowicz (faye.wedrowicz@gmail.com)

Abstract. Non-invasive genetic sampling has provided valuable ecological data for many species – data which may have been unobtainable using invasive sampling methods. However, DNA obtained non-invasively may be prone to increased levels of amplification failure and genotyping error.

Utilizing genotype data from 32 pedigreed koalas, this study aimed to validate the reliability of final consensus genotypes obtained using DNA isolated from koala scats. Pedigree analysis, duplicate genotyping, analysis of mismatched loci and tests for null alleles were used to look for evidence of errors.

All genetically confirmed parent–offspring relationships were found to follow Mendelian rules of inheritance. Duplicate genotypes matched in all cases and there was no evidence of null alleles. Related individuals always had different 12-marker genotypes having a minimum of three unique loci (in one full sibling pair), a mode of seven unique loci and a maximum of 11 unique loci.

This study demonstrates the capacity of DNA recovered from koala scats to provide reliable genotypes that can unequivocally discriminate individuals and infer parentage, provided data are missing from no more than two loci. Validating data obtained using non-invasive sampling is an important step, allowing potential problems to be identified at an early stage.

1 Introduction

Koalas (*Phascolarctos cinereus*) are elusive and typically reside in tall eucalypt trees, which can make animal capture and the collection of samples (e.g. blood or biopsy) for DNA analysis costly and time consuming. Koala scats are a convenient DNA source that can be readily identified and collected from the forest floor. DNA isolated from koala scats can then be used to genotype individual koalas, providing a unique DNA profile including the gender of the individual sampled (Wedrowicz et al., 2013). The genotypic data obtained can also be used for a range of population genetic analyses. Previous genetic studies of koala populations have used DNA recovered from blood or biopsy samples to study koala populations using microsatellites (Houlden et al., 1996; Lee et al., 2011) and mitochondrial control region sequencing (Houlden et al., 1999). Sourcing DNA non-invasively from scats offers

a valuable tool that may be useful for genetic studies of wild populations (Morin et al., 2016; Piggott et al., 2006; Stenglein et al., 2011). The ability to identify individuals using molecular methods is critical for determining the number of unique individuals sampled. This information is also important for a range of other applications such as population monitoring and estimating population size using mark–recapture methods as well as investigating social structure, relatedness and dispersal (Taberlet et al., 1997). However, the presence of errors in consensus genotypes may incorrectly result in genotypes from the same individual appearing as though two different individuals have been sampled. This type of error may bias analyses, resulting in overestimation of the number of individuals sampled and thus population size (Waits et al., 2001), which may in turn reduce the reliability of inferences.

DNA samples obtained using non-invasive collection methods are often associated with greater rates of amplifica-

tion failure and genotyping error compared to those obtained from invasively collected samples such as from tissues (Pompanon et al., 2005; Taberlet et al., 1996). The major factors contributing to increased amplification failure and error rates in non-invasively collected samples are DNA degradation (reducing DNA quality) and lower quantities of DNA due to collection of fewer cells from non-invasive sources (Taberlet et al., 1996). Genotyping errors are noted when two or more genotypes appear to have originated independently from the same sample, the major error types being allelic dropout and false alleles (Beja-Pereira et al., 2009; Bonin et al., 2004). Allelic dropout (ADO) occurs when only one of the two alleles in a heterozygote is amplified, while false alleles (FA) are amplified PCR artefacts that may be mistaken for a true allele (Pompanon et al., 2005). Though there is always likely to be some degree of error in any given data set, the increased chance of errors in DNA collected from non-invasive sources necessitates the assessment and minimization of errors.

A method for the isolation and microsatellite genotyping of koala DNA from scats has been described by Wedrowicz et al. (2013) and levels of error were reported to be 1.8 % (ADO) and 0.3 % (FA) when DNA concentration was above $1 \, \text{ng} \, \mu\text{L}^{-1}$. Based on simulations, and depending on DNA concentration, three or four genotyping replicates were required to provide highly reliable genotype data. Consensus genotypes constructed from replicates were, however, not validated. The availability of empirical data permits the presence of errors in consensus genotypes to be examined and allows the likelihood that genotypes from two different individuals are the same or similar to be estimated (Paetkau, 2003).

Here we consider genotypic data obtained using DNA collected from a captive and pedigreed koala population to validate a non-invasive method for DNA collection and genotyping (Wedrowicz et al., 2013). The aim of this study is to confirm whether the number of replicate genotypes and scoring rules used by Wedrowicz et al. (2013) produce correct consensus genotypes for individual koalas.

2   Materials and methods

2.1   The captive koala population

The Koala Conservation Centre (KCC) located on Phillip Island, Victoria, Australia, began operation in 1992 in response to concerns regarding the long-term survival of Victorian koala populations, including those on Phillip Island. The KCC comprises approximately 7.6 ha of forest (enclosed and subdivided by koala proof fences) where three tree species browsed by koalas dominate: southern blue gum (*Eucalyptus globulus*), manna gum (*E. viminalis*) and swamp gum (*E. ovata*). The KCC currently sustains a free-ranging population of around 40 koalas, some of which are restricted to particular areas of the property within large internally fenced areas. The centre's koala population was initially established

using individuals from the South Gippsland region in Victoria, due to their presumed endemicity and low rates of symptomatic chlamydial disease despite a high prevalence of infection (Emmins, 1996).

2.2   Scat collection and DNA isolation

Three fresh scat samples (showing a shiny outer surface) were collected from the ground directly beneath 32 individual koalas of known identity on two sampling occasions during 2007 ($n = 11$) and 2013 ($n = 21$). Samples collected in 2007 were independent from those used in the Wedrowicz et al. (2013) study. Individual koalas were identified by colour-coded ear tags. Scats were collected using wooden toothpicks and stored upright in open ended containers until surface washing. DNA was retrieved from a surface wash of each scat as described in Wedrowicz et al. (2013). To prevent sample contamination, DNA isolation, PCR setup and post-PCR analysis were carried out in separate areas of the laboratory and filtered pipette tips were used for pipetting DNA or DNA products. DNA was isolated from surface washes of two of the three scat samples for each individual animal using the QIAamp® DNA stool mini kit (Qiagen) as described by Wedrowicz et al. (2013); the surface wash of the third scat was stored in reserve at $-20\,°\text{C}$. DNA isolation from samples collected in 2007 used a slightly different protocol to that described in Wedrowicz et al. (2013) as outlined in the Supplement.

2.3   DNA screening and genotyping

DNA was quantified using the Qubit® dsDNA HS assay kit (Life Technologies). DNA quality was assessed using standard PCR and electrophoresis. Microsatellite Pcv31 (Cristescu et al., 2009) was amplified using reactions comprising $5 \, \mu\text{L}$ GoTaq® green master mix (Promega), $0.5 \, \mu\text{M}$ of each primer and $0.1 \, \mu\text{g} \, \mu\text{L}^{-1}$ of bovine serum albumin (BSA) made up to $10 \, \mu\text{L}$ with nuclease free water. A sexing PCR using primers GpdEx12, GpdEx13R (Loebel et al., 1995; Loebel and Johnston, 1997), IMY1 and IMY2 (Watson et al., 1998) and the PCR parameters described in Wedrowicz et al. (2013) was also carried out. For each individual, the DNA isolate producing the brightest bands on PCR gels, was chosen for genotyping.

Wedrowicz et al. (2013) describes a method for reliably genotyping koalas from DNA isolated from a single koala scat using specific sample collection and storage procedures and optimized DNA isolation protocols. Genotyping errors are accounted for by replicating genotypes three or four times per sample, according to DNA concentration, minimizing the chance that final genotypes contain errors. Replicate genotypes are used to produce a consensus genotype, with more than 99 % confidence in the resultant genotype (Wedrowicz et al., 2013). However reproducible errors, such as null alleles, may remain undetected using this method, requiring

other forms of error checking such as pedigree analysis for detection (Pompanon et al., 2005).

To obtain genotypes for each sample, DNA isolates were amplified using 12 microsatellite markers: K2.1, K10.1, Pcv2, Pcv6.1, Pcv6.3, Pcv24.2, Pcv25.2, Pcv30, Pcv31 (Cristescu et al., 2009), Phc2, Phc4 and Phc13 (Houlden et al., 1996). Amplification and product separation using capillary electrophoresis was conducted at the Australian Genome Research Facility (AGRF), Melbourne, Australia. Genotypes were replicated as recommended by Wedrowicz et al. (2013) according to the total DNA concentration (three replicates if DNA concentration was greater than $1.0 \, \text{ng} \, \mu\text{L}^{-1}$; four replicates if DNA concentration was 0.25–$1.0 \, \text{ng} \, \mu\text{L}^{-1}$).

2.4   Estimation of error rates

Rates and instances of genotyping error were calculated from replicate genotypes using GIMLET v 1.3.3 (Valière, 2002). Consensus genotypes were constructed based on Taberlet et al. (1996) and Valière et al. (2007) using the following rules: (1) alleles had to appear at least twice to be counted; (2) where four replicates were used, homozygous alleles had to appear at least three times; and (3) loci giving ambiguous results were omitted (scored as a failed reaction). In order to check for genotyping errors between consensus genotypes of independent duplicates (different scats from the same individual), DNA isolated from the second scat was also genotyped for nine randomly chosen individuals (28 % of samples). Micro-checker was used to test for the presence of null alleles (Van Oosterhout et al., 2004).

2.5   Mitochondrial DNA sequencing

To obtain maternally inherited haplotype data for the KCC koala population, primers KmtL1 and KmtH2 (Fowler et al., 2000) were used to amplify and sequence approximately 700 base pairs within the mitochondrial DNA (mtDNA) control region. Mitochondrial DNA was amplified using BIO-X-ACT™ short DNA polymerase (Bioline). Reactions consisted of $1 \times$ OptiBuffer, $1 \times$ Hi-Spec additive, 2 mM MgCl$_2$, 0.5 mM each dNTP, 1 unit of BIO-X-ACT™ short DNA polymerase and 0.5 μM of each primer, made up to 20 μL with water. DNA was initially denatured for 3 mins at 95 °C, followed by 35 cycles of 94, 50 and 72 °C for 1 min each and finishing with a single final extension of 5 min at 72 °C. PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega) and sequencing was carried out by AGRF. Resulting sequences were aligned and trimmed using MEGA 6 (Tamura et al., 2013).

2.6   KCC pedigree data

Pedigree information was obtained from KCC records. Five maternal founders and five generations were represented in the KCC pedigree (Fig. 1). At the Conservation Centre, koalas within the same maternal lineage are generally given names that begin with the same letter; names assigned by the KCC to each individual are used throughout this paper. Maternal relationships were known with high confidence as juveniles were caught and tagged after leaving the pouch but before becoming independent. Confident identification of paternal relationships can be more challenging as, although individual female–male pairs are housed together within the same enclosure for breeding, other KCC koalas may sometimes escape or enter the enclosures; opportunities for other males within the captive colony to breed with the intended female therefore exist. To account for potential errors in the pedigree, all parental relationships were considered putative until confirmed by the molecular data. The term "putative" is therefore used throughout when referring to the KCC pedigree data alone.

2.7   Probability of identity

The probability of identity and probability of identity between siblings were calculated using GenAlEx (Peakall and Smouse, 2012). The similarity between genotypes of pairs of individuals was also considered by examining the number of loci with different genotypes as described by Paetkau (2003). We use the same system as Paetkau (2003) to describe the number of mismatched loci between pairs of individuals. Two individuals with identical genotypes at all 12 loci (no mismatches) are referred to as a 0MM pair; a 1MM pair describes a pair of individuals with a single mismatching locus, i.e. a unique genotype at one locus and identical genotypes at all other loci; and a 12MM pair has a unique genotype at every locus (Paetkau, 2003). To count the number of mismatched loci between each pair of individuals in the KCC population, the R package *allelematch* (Galpern et al., 2012; R Core Team, 2014) was used.

In order for the degree of relatedness to be compared to the frequency of mismatched loci, parental information from the KCC pedigree confirmed by the genetic data was used to calculate pedigree relationship coefficients (R) between all pairs of individuals using the *pedantics* package (Morrissey and Wilson, 2010; R Core Team, 2014). Parentage information was omitted where the parents of an individual koala were unknown or uncertain, meaning that a small number of pairwise relationships may have been classified as unrelated when the true degree of relationship may have been higher. Values calculated by the *pedantics* package (derived from the genetically confirmed pedigree data) were used to assign each pairwise relationship as first order (I: parent–offspring (PO) or full siblings (FS), $R \geq 0.50$), second order (II: half-siblings (HS), avuncular or grandparent–grandoffspring, $0.25 \leq R < 0.50$), third order (III: cousins, $0.125 \leq R < 0.25$) or unrelated (UR: $R < 0.125$).
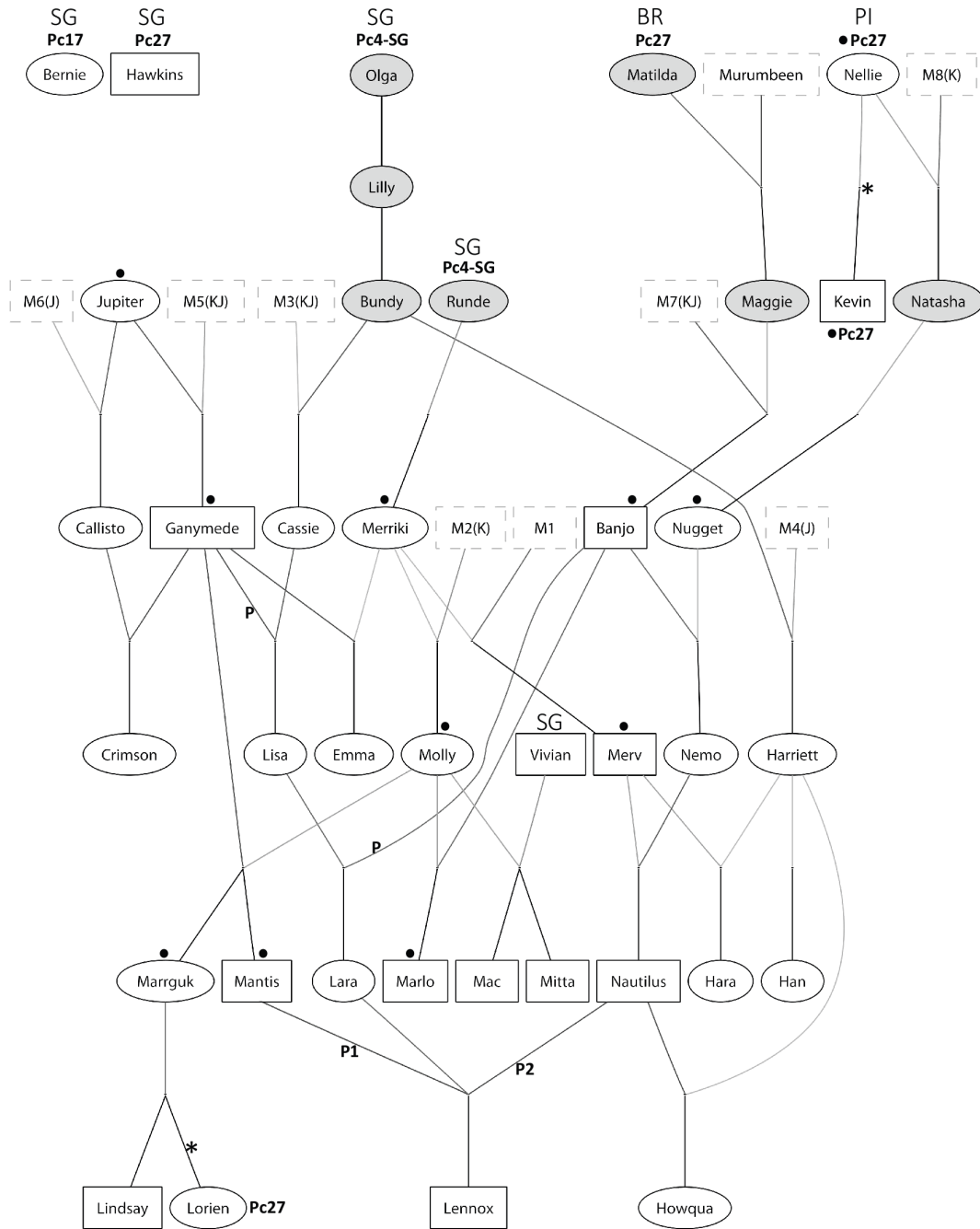
Figure 1. Pedigree diagram for sampled koalas. Ovals indicate females and rectangles males. Individuals marked with a circle (●) were sampled in 2007, while all others were sampled in 2013. Molly and Marlo were sampled in both 2007 and 2013. Individuals that are shaded grey were not sampled. Presumed maternal relationships not supported by the molecular data are marked with an asterisk (*). Putative paternal relationships not confirmed by the genetic data are not shown. Possible paternal relationships identified from the molecular data are marked with a "P" or "P1" and "P2" where more than one candidate father was identified. The origin of founders is also indicated by SG: South Gippsland; PI: Phillip Island; or BR: Brisbane ranges (founded by French and Phillip Island stock). Unknown fathers are numbered M1–M8. Many of these fathers are thought to be Karlos (K), Jack (J) or one of the two (KJ), though these individuals were not sampled.

Table 1. Genotyping success rates according to sampling year and DNA concentration, grouped into categories A–D.

Total DNA isolated (ng μL^{-1})	Sampling year	Total PCRs (no. of individuals)	PCR success rate (±95 % CI %)	Average peak height for successful PCRs (±95 % CI)	Average number successful loci per consensus genotype	Allelic dropout: average rate across loci % (frequency)	False alleles: average rate across loci % (frequency)
A: 0.10–0.25	2007	96 (2)	90 ± 4.5	7591 ± 1098	11.0/12	1.0 (1)	0.0 (0)
	2013	48 (1)	71 ± 6.6	4236 ± 610	8.0/12	0.0 (0)	0.0 (0)
	Overall	144 (3)	80 ± 4.1	6110 ± 693	10.0/12	1.0 (1)	0.0 (0)
B: 0.25–0.50	2007	192 (4)	69 ± 3.3	4665 ± 437	8.0/12	5.6 (5)	0.0 (0)
	2013	192 (4)	89 ± 2.3[A]	5641 ± 405	11.3/12	5.1 (5)	0.0 (0)
	Overall	384 (8)	79 ± 2.1[A]	5214 ± 298	9.6/12	6.2 (10)	0.0 (0)
C: 0.50–1.0	2007	0 (0)	–	–	–	–	–
	2013	288 (6)	92 ± 1.6	5420 ± 328	11.8/12	0.8 (2)	0.8 (2)
	Overall	288 (6)	92 ± 1.6[A,B]	5420 ± 328	11.8/12	0.8 (2)	0.8 (2)
D: > 1.0	2007	252 (7)	88 ± 1.9[A,B]	5040 ± 317	10.6/12	2.7 (5)	0.8 (1)
	2013	648 (18)	97 ± 0.65[C]	7539 ± 265[B,C]	11.8/12	0.4 (2)	0.0 (0)
	Overall	900 (25)	94 ± 0.76[A,B]	6812 ± 212[B,C]	11.4/12	1.1 (7)	0.2 (1)

[A, B, C] Superscripts indicate categories that differed significantly ($p < 0.05$).

2.8 Parentage analysis

The full exclusion method of paternal analysis was carried out manually to confirm parent–offspring relationships. Offspring genotypes were considered compatible with parental genotypes if one allele was shared at each of the 12 loci genotyped. PARENTE (Cercueil et al., 2002) uses the same method to identify potential mothers, fathers and parent pairs and was also used to confirm parentage. Given known mothers, tests for paternity were carried out using CERVUS (Kalinowski et al., 2007). KCC records and the genotyping results were used to visualize the KCC pedigree sampled using Pedigraph V1 (Garbe and Da, 2008; Fig. 1). Parentage was also determined from the genotypic data using the pedigree reconstruction software FRANz 2.0.0 (Riester et al., 2009). FRANz was utilized as it is able to infer multigenerational pedigrees without the need for prior information, such as maternal relationship, that is often required by other software.

3 Results

3.1 Genotypic data

Consensus DNA profiles were obtained for 32 koalas at the KCC. Genotypic data were obtained for a total of 42 DNA isolates (32 individuals: 23 single samples, 9 duplicate samples and 1 triplicate sample); of these, 17 were genotyped four times (DNA concentration below 1 ng μL^{-1}) and 25 were genotyped three times (DNA concentration over 1 ng μL^{-1}). Of the 32 individuals genotyped, 21 (66 %) had complete genotypes, 3 had missing data at one locus (9 %), 4 had missing data at two loci (12 %) and 4 individuals (9 %)

were not successfully amplified or scored at more than two loci (three, four, four and seven missing loci). Missing data were mostly present in samples collected in 2007 (Table 1), which could be due to the slightly different method used for DNA isolation. All consensus genotypes were identical at every available locus for the scat samples from the nine individuals genotyped in replicate (Supplement, Table S1). Scats from two individuals (Molly and Marlo) were obtained in both 2007 and 2013; for each individual, genotypes at all available loci were identical for the 2007 and 2013 samples (Supplement, Table S1), confirming that the intended individual was sampled on both occasions.

The rates of allelic dropout and false alleles calculated from replicate data were 1.9 and 0.2 % (averaged across loci) respectively, which is similar to the overall error rates of 1.8 and 0.3 % reported by Wedrowicz et al. (2013) for samples containing 1–20 ng μL^{-1} of total DNA. No evidence of errors in consensus genotypes obtained from the KCC population were noted. Five putative parent–offspring relationships (two maternal and three paternal) were found to be incompatible. Incompatible genotypes identified in putative parent–offspring relationships displayed (1) mismatched genotypes at multiple loci and (2) unanimity with additional known relationships and so were refuted as errors. Four of the five incompatible parent–offspring pairs had more than three mismatching loci. One putative parent pair (Lara–Merv) mismatched at only one locus (K2.1), where both the parent and offspring were heterozygous. In the parent (Merv), the discrepant allele of 166 base pairs was refuted as an error due to its presence in his mother and another of his offspring (Supplement, Fig. S1). The mismatching 164 base pair allele

Table 2. Variability in 640 bp mtDNA control region sequence for individuals sampled at the KCC.

Haplotype	n	Percent of samples	Nucleotide position			
			6	21	112	247
Pc27	13	40.6 %	G	T	G	G
Pc4–SG	18	56.3 %	–	–	–	A
Pc17	1	3.1 %	A	C	A	A

was present in two independent samples from the offspring (Lara), and was present in seven other individuals within the population; CERVUS identified another candidate male (Banjo) as the most likely father. Mutation of the K2.1 allele from 166 bases in the parent (Merv) to 164 bases in the offspring (Lara) is also a possibility though the mutation rate of microsatellites, averages around 5×10^{-4} mutations per locus per generation (Selkoe and Toonen, 2006). Further genotyping using additional markers for Lara (offspring), Lisa (mother) and Merv and Banjo (potential fathers) would therefore be required to definitely assign parentage in this case. All genotypes were found to follow Mendelian modes of inheritance. Null alleles were not detected using Micro-Checker. Considering all individuals sampled at the KCC (including Bernie and Hawkins; two recent immigrants introduced from South Gippsland), Hardy–Weinberg proportions (HWP) were followed for all but one locus (Pcv6.3); all loci were in HWP after the two immigrants, which had no sampled offspring, were removed.

3.2 Sequencing performance and mtDNA variability

Following alignment and trimming, 642 bases of mtDNA control region sequence were obtained. Three mtDNA haplotypes with four variable sites were identified in the KCC individuals sampled in this study (Table 2). Haplotype names have recently been standardized by Neaves et al. (2016) and are used here. Two of the haplotypes, Pc17 and Pc27, were previously described by Houlden et al. (1999). Pc27 is the most common haplotype in South Gippsland and the only haplotype found in individuals of French and Phillip Island origin (Houlden et al., 1999); haplotype Pc27 was present in 13 (41 %) of the 32 KCC individuals. One (3 %) immigrant koala (Bernie) possessed the Pc17 haplotype. The third haplotype, designated Pcv4–SG, has not been previously reported in Victoria. Pcv4–SG was the most common haplotype in the KCC data set, being present in 18 (56 %) of the 32 individuals sampled. Pcv4–SG was 100 % identical to Pc4 reported by Houlden et al. (1999) found in northern NSW koalas (Coonabarabran and Port Stephens) but only covered 74 % of the full length of Pc4. Sequencing a larger section of the control region would be required in order to more accurately define this haplotype.

Table 3. The number of loci displaying different genotypes between pairs of individuals of known relationship in the KCC koala population. FOR: first-order relationship ($R \geq 0.50$); SOR: second-order relationships ($0.25 \geq R > 0.50$); TOR: third-order relationship ($0.125 \geq R > 0.25$); UR: unrelated ($R < 0.125$).

Mis-matched loci	FOR	SOR	TOR	UR	Total (percent related)
0MM	0	0	0	0	0 (–)
1MM	0	0	0	0	0 (–)
2MM	0	0	0	0	0 (–)
3MM	1	0	0	0	1 (100 %)
4MM	1	2	0	3	6 (50 %)
5MM	4	2	2	8	16 (50 %)
6MM	0	8	2	19	29 (34 %)
7MM	8	6	8	51	73 (30 %)
8MM	5	6	8	68	87 (22 %)
9MM	2	7	6	70	85 (18 %)
10MM	1	3	4	37	45 (18 %)
11MM	0	0	1	26	27 (4 %)
12MM	0	0	0	9	9 (0 %)
Totals	22	34	31	291	378 (23 %)

3.3 Individual identification

All positive sexing PCRs correctly identified the gender of the individual sampled. Using the complete 12-marker set of microsatellites , the probability of identity (P_{ID}) and probability of identity between siblings (P_{IDsibs}) in the KCC population was $1.3 \times 10^{-8} (\approx 1$ in $75\,000\,000)$ and 3.9×10^{-4} (≈ 1 in 2500 or 0.04 %), respectively. To retain a high level of power of the marker set to discriminate individuals, a limit of missing data at two loci was imposed (i.e. consensus genotypes with missing data at more than two loci were discarded). The P_{ID} and P_{IDsibs} in the KCC population using the ten least informative loci (i.e. excluding the two most informative loci: K2.1 and Pcv2) was $2.5 \times 10^{-6} (\approx 1$ in 400 000) and 2.8×10^{-3} (≈ 1 in 350 or 0.28 %). Excluding consensus genotypes with missing data at more than two loci, there were no instances of 0MM, 1MM or 2MM pairs. There was one occurrence of a 3MM pair between full siblings (Mitta and Mac) and six 4MM pairs (three related and three unrelated: Table 3), suggesting that the likelihood of any two individuals having the same 12-marker genotype is likely to be very low. Using this microsatellite suite for the individual identification of koalas at the KCC, we therefore recommend that at least ten loci are successfully typed and scored to have a high level of confidence in individual assignment.

3.4 Parentage analysis

A pedigree diagram was constructed for the individuals sampled in this study using the KCC pedigree data, supplemented by information obtained from the molecular data (Fig. 1).

The pedigree information supplied by the KCC consisted of 22 putative maternal relationships and 14 putative parent pair–offspring relationships. Of the mother–offspring relationships 20 of the 22 (91 %) were confirmed using the full exclusion method. Two individuals (Kevin and Lorien) were not tagged when juveniles; hence, there was some initial uncertainty regarding their parentage. The putative maternal relationships for Kevin and Lorien were not confirmed by the genetic data (Kevin and Lorien had four and three mismatched loci with their presumed mothers respectively). Genotypic data identified Jupiter as a potential mother of Kevin, with a matching allele at each of the 10 loci available between the pair. None of the sampled individuals were found to have a genotype compatible with maternity of Lorien; however, not all potential parents at the KCC were sampled. As expected, all mtDNA haplotypes were maternally inherited. Haplotype data also contradicted the parent–offspring relationship between Marrguk (Pc4–SG) and Lorien (Pc27), showing that Lorien originates from a maternal lineage with the Pc27 haplotype (Fig. 1).

Putative paternal relationships (and therefore parent pairs) were confirmed for 11 out of 14 (78 %) offspring, while three putative paternal relationships (the putative fathers of Lisa, Lara and Lennox) were refuted by the genotypic data. For the three individuals whose presumed fathers were not validated by the molecular data (Lisa, Lara and Lennox, Fig. 1), alternative potential fathers were identified from the males sampled. Since all males at the KCC were not sampled these relationships could not be unequivocally confirmed; however, paternity assignment using CERVUS specified significant odds ratios for the Ganymede–Lisa, Banjo–Lara and Mantis–Lennox father–offspring relationships.

3.5 Pedigree reconstruction

The pedigree reconstruction software, FRANz 2.0.0 (Riester et al., 2009) was used in order to evaluate whether a multigenerational pedigree could be constructed based on the genetic data (12-marker genotypes and sex markers) alone. Individuals with missing data at more than two loci were excluded, leaving nine maternal relationships and seven parent pairs with genetically confirmed relationships. All nine maternal relationships and seven parent pairs were correctly identified by FRANz. Additionally, the known mothers/potential fathers for two of the three individuals with uncertain paternity were also identified by FRANz. All identified parent pairs had a log of the odds (LOD) value greater than 4.0, meaning that the odds of the identified relationships detected having occurred by chance are greater than 10^4 to 1. One individual (Lindsay; 7.7 % of identifications above LOD = 4) was incorrectly assigned to a parent pair by FRANz. Since Lindsay's true mother amplified at only five loci and was excluded, the true mother's full sibling was assigned instead; the true father was unknown. A LOD of 4 will therefore provide a good cutoff above which parents identi-

fied using FRANz can be considered highly likely, provided genotypes do not have missing data at more than two loci.

4 Discussion

Declining wildlife populations and threatened or vulnerable species may benefit from the development of non-invasive genetic sampling regimes allowing for the rapid acquisition of large amounts of population data. Such data have the potential to facilitate a greater understanding of the processes that may be implicated in population declines and thus allow for the development and action of appropriate management strategies in an attempt to prevent further declines and potential extinction. By sampling scats from a captive population of koalas, with known pedigree, this study has shown that the method for isolating and genotyping DNA from koala scats used here provides accurate consensus genotypes. This study also demonstrated the ability of the 12-microsatellite suite to unequivocally identify the individual from which the sample was obtained and to infer parent–offspring relationships. The applicability of mtDNA sequencing using DNA isolated from koala scats was also established. Overall, this study has confirmed that sampling DNA from koala scats is a robust and reliable alternative to traditional DNA sources that may be beneficial to future conservation studies for the koala.

4.1 Error assessment

Similar to results reported by Wedrowicz et al. (2013), the error rate found within replicate genotypes was determined to have an average of 1.9 % (ADO) and 0.2 % (FA) across loci. Using pedigree analysis, duplicate genotyping, analysis of mismatched loci and tests for null alleles, the consensus genotypes of the 32 individuals used in this study were free of detectable error. This provides a strong indication that the methods and scoring rules used to obtain genotypic data using DNA isolated from koala scats are associated with acceptably low error rates.

Examination of pairs of individuals with a small number of mismatched alleles (Paetkau, 2003) can be a very useful method for checking errors in samples collected from wild populations. Within the captive KCC population, no 0MM (same 12-marker genotype), 1MM or 2MM pairs were observed, while a 3MM pair was observed once between a pair of full siblings. Given that the KCC population is likely to contain a higher proportion of related individuals than wild populations, it would seem unlikely that any two individuals would have less than four mismatched loci between them. Identifying genotypes that mismatch at three or less loci in wild collected data sets is an ideal step for identification of potential errors before carrying out further analyses.

4.2 Individual identification

The suite of markers considered here can be used to un-equivocally identify individuals and infer parentage, which are both important capabilities for ecological investigations of koala populations. As discussed above, the distribution of mismatched loci found between pairs of individuals in the KCC population, containing numerous first degree relatives, suggests that the chance of observing a 0MM pair is extremely small. P_{ID} and P_{IDsibs} for the koala population at the KCC was 1 in 75 000 000 and 1 in 2500 when all 12 markers were successfully amplified. P_{IDsibs} is suggested as a conservative limit from which to gauge the probability of two individuals sharing the same genotype by chance (Waits et al., 2001). The frequency of full siblings in wild koala populations is unknown, but likely to be negligible as most siblings born to the same mother are found to have differing paternities (Ellis et al., 2002). P_{IDsibs} is therefore a very conservative measure for defining genotype matches for koalas, which would be advantageous as incorrectly identifying individuals could lead to misinformed management. Previously utilized cut-offs for confident assignment of full genotype matches as one specific individual have included 0.05 (1 in 20; Woods et al., 1999) for P_{IDsibs} and < 0.001–0.0001 (less than 1 in 1000–10 000) P_{ID} in wildlife forensics (Waits et al., 2001). The P_{ID} and P_{IDsibs} reported in this study are therefore well within acceptable limits.

4.3 Inference of parentage

Genetic data can be used to assist captive management in a range of ways, including assessments of founder relationships, filling gaps within pedigrees, quantifying and monitoring genetic diversity, classifying the region of origin and identifying genetically valuable individuals (Ivy et al., 2009). Parent–offspring relationships were confirmed in almost all cases in this study (31/36). Two uncertain maternal relationships not confirmed by the genetic data highlights the potential for errors to be present in pedigree information. Pedigree data are often used in captive breeding programmes to select breeding pairs that are sufficiently unrelated while genetic estimates of relatedness can be useful to minimize inbreeding between wild founders (Bergner et al., 2014). When using pedigree data, the presence of pedigree errors may undermine management schemes, resulting in unintended breeding between related individuals that could have a negative effect on fitness (Hammerly et al., 2016). Supplementing pedigree data with genetic data may be useful for captive management programmes in order to ensure management decisions are based on the most accurate information possible by using genetic data to validate pedigree data and chosen breeding pairs. As an example, a study of captive Attwater's prairie chickens carried out by Hammerly et al. (2016) firstly assigned breeding pairs based on pedigree estimates of relatedness and then reassigned the chosen pair if the genetic estimate of pairwise relatedness was greater than 0.125. This method of mate selection was found to significantly increase the survival rate of chicks compared to choosing breeding pairs based on the pedigree alone (Hammerly et al., 2016). Inferring parentage within wild populations with no prior pedigree knowledge is also possible, though likely to present a greater challenge, as sampling of all candidate parents (especially fathers) may not always be achievable.

5 Conclusions

DNA isolated from koala scats, and investigated using a suite of DNA markers, provided data that can be confidently used to study both captive pedigrees and wild koala populations. Analyses of parentage and relatedness aid the selection of breeding pairs and can verify the parents of juveniles when parental information is unavailable or uncertain, so can assist the management of captive populations. The use of non-invasive genetic sampling for the study of wild populations has the potential to provide numerous advantages over invasive sampling methods. Compared to obtaining DNA from biopsies or blood from wild individuals, sourcing DNA from scats confers the advantage of permitting wide geographic studies across densely forested and inaccessible terrain. DNA sourced from scats also permits the collection of large sample sizes and, when compared to opportunistic sampling involving road kill or shelter animals, minimizes sampling bias.

Both microsatellite genotyping and sequencing of DNA isolated from koala scats have been shown to produce reliable results. Analysis of non-invasively isolated DNA using genomic methods, such as single nucleotide polymorphisms (SNPs), have been demonstrated (Fabbri et al., 2012; Snyder-Mackler et al., 2016) and may widen the applicability of DNA isolated from koala scats. Another potential benefit of DNA isolated from scats is that DNA from gut microbes or ingested plant material may also be present, which could be used to gather additional information, such as measures of health and dietary preferences (Bradley et al., 2007; Ley et al., 2008). Future advancements in technologies and methods will provide additional, and more efficient uses for non-invasive genetic sampling in conservation genetics. In the case of the koala, this has the potential to lead to standardized methods that can be analysed and compared across the koalas' entire range providing a better understanding of population genetic diversity and promoting the conservation of this iconic species.

The Supplement related to this article is available online at doi:10.5194/we-17-9-2017-supplement.

References

Beja-Pereira, A., Oliveira, R., Alves, P. C., Schwartz, M. K., and Luikart, G.: Advancing ecological understandings through technological transformations in noninvasive genetics, Mol. Ecol. Resour., 9, 1279–1301, 2009.

Bergner, L. M., Jamieson, I. G., and Robertson, B. C.: Combining genetic data to identify relatedness among founders in a genetically depauperate parrot, the Kakapo (*Strigops habroptilus*), Conserv. Genet., 15, 1013–1020, 2014.

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., and Taberlet, P.: How to track and assess genotyping errors in population genetics studies, Mol. Ecol., 13, 3261–3273, 2004.

Bradley, B. J., Stiller, M., Doran-Sheehy, D. M., Harris, T., Chapman, C. A., Vigilant, L., and Poinar, H.: Plant DNA sequences from feces: potential means for assessing diets of wild primates, Am. J. Primatol., 69, 699–705, 2007.

Cercueil, A., Bellemain, E., and Manel, S.: PARENTE: computer program for parentage analysis, J. Hered., 93, 458–459, 2002.

Cristescu, R., Cahill, V., Sherwin, W. B., Handasyde, K., Carlyon, K., Whisson, D., Herbert, C. A., Carlsson, B. L. J., Wilton, A. N., and Cooper, D. W.: Inbreeding and testicular abnormalities in a bottlenecked population of koalas (*Phascolarctos cinereus*), Wildl. Res., 36, 299–308, 2009.

Ellis, W. A., Hale, P. T., and Carrick, F.: Breeding dynamics of koalas in open woodlands, Wildl. Res., 29, 19–25, 2002.

Emmins, J. J.: The Victorian koala: Genetic heterogeneity, immune reponsiveness and epizootiology of Chlamydiosis, Doctor of Philosophy, Depatment of Pathology and Immunology, Monash University, Melbourne, Victoria, 346 pp., 1996.

Fabbri, E., Caniglia, R., Mucci, N., Thomsen, H. P., Krag, K., Pertoldi, C., Loeschcke, V., and Randi, E.: Comparison of single nucleotide polymorphisms and microsatellites in non-invasive genetic monitoring of a wolf population, Arch. Biol. Sci., 64, 321–335, 2012.

Fowler, E. V., Houlden, B. A., Hoeben, P., and Timms, P.: Genetic diversity and gene flow among southeastern Queensland koalas (*Phascolarctos cinereus*), Mol. Ecol., 9, 155–164, 2000.

Galpern, P., Manseau, M., Hettinga, P., Smith, K., and Wilson, P.: Allelematch: an R package for identifying unique multilocus genotypes where genotyping error and missing data may be present, Mol. Ecol. Resour., 12, 771–778, 2012.

Garbe, J. and Da, Y.: Pedigraph: a software tool for the graphing and analysis of large complex pedigree: User manual Version 2.4, Department of Animal Science, University of Minnesota, 2008.

Hammerly, S., Cerda, D., Bailey, H., and Johnson, J.: A pedigree gone bad: increased offspring survival after using DNA-based relatedness to minimize inbreeding in a captive population, Anim. Conserv., 19, 296–303, 2016.

Houlden, B. A., England, P. R., Taylor, A. C., Greville, W. D., and Sherwin, W. B.: Low genetic variability of the koala *Phascolarctos cinereus* in south-eastern Australia following a severe population bottleneck, Mol. Ecol., 5, 269–281, 1996.

Houlden, B. A., Costello, B. H., Sharkey, D., Fowler, E. V., Melzer, A., Ellis, W., Carrick, F., Baverstock, P. R., and Elphinstone, M. S.: Phylogeographic differentiation in the mitochondrial control region in the koala, *Phascolarctos cinereus* (Goldfuss 1817), Mol. Ecol., 8, 999–1011, 1999.

Ivy, J. A., Miller, A., Lacy, R. C., and DeWoody, J. A.: Methods and prospects for using molecular data in captive breeding programs: an empirical example using parma wallabies (*Macropus parma*), J. Hered., 100, 441–454, 2009.

Kalinowski, S. T., Taper, M. L., and Marshall, T. C.: Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment, Mol. Ecol., 16, 1099–1106, 2007.

Lee, T., Zenger, K. R., Close, R. L., and Phalen, D. N.: Genetic analysis reveals a distinct and highly diverse koala (*Phascolarctos cinereus*) population in South Gippsland, Victoria, Australia, Aust. Mammal., 34, 68–74, 2011.

Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R., and Gordon, J. I.: Worlds within worlds: evolution of the vertebrate gut microbiota, Nat. Rev. Microbiol., 6, 776–788, 2008.

Loebel, D., Longhurst, T., and Johnston, P.: Full-length cDNA sequence of X-linked G6PD of an Australian marsupial, the wallaroo, Mamm. Genome, 6, 198–201, 1995.

Loebel, D. A. and Johnston, P. G.: Analysis of the intron-exon structure of the G6PD gene of the wallaroo (*Macropus robustus*) by polymerase chain reaction, Mamm. Genome, 8, 146–147, 1997.

Morin, D. J., Kelly, M. J., and Waits, L. P.: Monitoring coyote population dynamics with fecal DNA and spatial capture–recapture, J. Wildl. Manage., 80, 824–836, 2016.

Morrissey, M. B. and Wilson, A. J.: PEDANTICS: an R package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing, Mol. Ecol. Resour., 10, 711–719, 2010.

Neaves, L. E., Frankham, G. J., Dennison, S., FitzGibbon, S., Flannagan, C., Gillett, A., Hynes, E., Handasyde, K., Helgen, K. M., and Tsangaras, K.: Phylogeography of the Koala (*Phascolarctos*

cinereus) and Harmonising Data to Inform Conservation, PloS one, 11, e0162207, doi:10.1371/journal.pone.0162207, 2016.

Paetkau, D.: An empirical exploration of data quality in DNA-based population inventories, Mol. Ecol., 12, 1375–1387, 2003.

Peakall, R. and Smouse, P. E.: GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update, Bioinformatics, 28, 2537–2539, 2012.

Piggott, M., Banks, S., and Taylor, A.: Population structure of brush-tailed rock-wallaby (*Petrogale penicillata*) colonies inferred from analysis of faecal DNA, Mol. Ecol., 15, 93–105, 2006.

Pompanon, F., Bonin, A., Bellemain, E., and Taberlet, P.: Genotyping errors: causes, consequences and solutions, Nat. Rev. Genet., 6, 847–846, 2005.

R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2014.

Riester, M., Stadler, P. F., and Klemm, K.: FRANz: reconstruction of wild multi-generation pedigrees, Bioinformatics, 25, 2134–2139, 2009.

Selkoe, K. A. and Toonen, R. J.: Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers, Ecol. Lett., 9, 615–629, 2006.

Snyder-Mackler, N., Majoros, W. H., Yuan, M. L., Shaver, A. O., Gordon, J. B., Kopp, G. H., Schlebusch, S. A., Wall, J. D., Alberts, S. C., Mukherjee, S., Zhou, X., and Tung, J.: Efficient genome-wide sequencing and low-coverage pedigree analysis from noninvasively collected samples, Genetics, 203, 699–714, 2016.

Stenglein, J. L., Waits, L. P., Ausband, D. E., Zager, P., and Mack, C. M.: Estimating gray wolf pack size and family relationships using noninvasive genetic sampling at rendezvous sites, J. Mammal., 92, 784–795, 2011.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L. P., and Bouvet, J.: Reliable genotyping of samples with very low DNA quantities using PCR, Nucleic Acids Res., 24, 3189–3194, 1996.

Taberlet, P., Camarra, J. J., Griffin, S., Uhrès, E., Hanotte, O., Waits, L. P., Dubois-Paganon, C., Burke, T., and Bouvet, J.: Noninvasive genetic tracking of the endangered Pyrenean brown bear population, Mol. Ecol., 6, 869–876, 1997.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S.: MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0, Mol. Biol. Evol., 30, 2725–2729, 2013.

Valière, N.: GIMLET: a computer program for analysing genetic individual identification data, Mol. Ecol. Notes, 2, 377–379, 2002.

Valière, N., Bonenfant, C., Toïgo, C., Luikart, G., Gaillard, J.-M., and Klein, F.: Importance of a pilot study for non-invasive genetic sampling: genotyping errors and population size estimation in red deer, Conserv. Genet., 8, 69–78, 2007.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P.: MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data, Mol. Ecol. Notes, 4, 535–538, 2004.

Waits, L. P., Luikart, G., and Taberlet, P.: Estimating the probability of identity among genotypes in natural populations: cautions and guidelines, Mol. Ecol., 10, 249–256, 2001.

Watson, C., Margan, S., and Johnston, P.: Sex-chromosome elimination in the bandicoot *Isoodon macrourus* using Y-linked markers, Cytogenet. Genome Res., 81, 54–59, 1998.

Wedrowicz, F., Karsa, M., Mosse, J., and Hogan, F. E.: Reliable genotyping of the koala (*Phascolarctos cinereus*) using DNA isolated from a single faecal pellet, Mol. Ecol. Resour., 13, 634–641, 2013.

Woods, J. G., Paetkau, D., Lewis, D., McLellan, B. N., Proctor, M., and Strobeck, C.: Genetic tagging of free-ranging black and brown bears, Wildl. Soc. Bull., 27, 616–627, 1999.