Genome-wide SNP loci reveal novel insights into koala (*Phascolarctos cinereus*) population variability across its range

Shannon R. Kjeldsen [1,*]

Email shannon.kjeldsen@jcu.edu.au

Kyall R. Zenger [1,2]

Kellie Leigh [3,4]

WilliamEllis [5]

Jenifer Tobey [6]

David Phalen [4]

Alistair Melzer [7]

Sean FitzGibbon [5]

Herman W. Raadsma [4]

[1] College of Marine and Environmental Sciences, James Cook University, Building 145, Townsville, QLD, 4811 Australia

[2] Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University, Townsville, 4811 Australia

[3] Science for Wildlife, PO Box 286, Cammeray, NSW, 2062 Australia

[4] Reprogen-Animal Bioscience Group, Faculty of Veterinary Science, University of Sydney, Private Mail Bag 4003, Narellan, NSW, 2570 Australia

[5] School of Agriculture and Food Science, The University of Queensland, Brisbane, 4072 Australia

[6] San Diego Zoo, San Diego, CA, 92101 USA

[7] School of Medical and Applied Science, CQ University, Rockhampton, 4702 Australia

Abstract

The koala (*Phascolarctos cinereus*) is an iconic Australian species that is currently undergoing a number of threatening processes, including disease and habitat loss. A thorough understanding of population genetic structuring and genomic variability of this species is essential to effectively manage populations across the species range. Using a reduced representation genome sequencing method known as double digest restriction-associated sequencing, this study has provided the first genome-wide SNP marker panel in the koala. In this study, 33,019 loci were identified in the koala and a filtered panel of 3060 high-utility SNP markers, including 95 sex-linked markers, were used to provide key insights into population variability and genomic variation in 171 koalas from eight populations across their geographic range. Broad-scale genetic differentiation between geographically separated populations (including sub-species) was assessed and revealed significant differentiation between all populations (F_{ST} range = 0.01–0.28), with the largest divergence observed between the three geographically distant subgroups (QLD, NSW and VIC) along the east coast of Australia (average F_{ST} range = 0.17–0.23). Sub-group divergence appears to be a reflection of an isolation by distance effect and sampling strategy rather than true evidence of sub-speciation. This is further supported by low proportions of AMOVA variation between sub-species groups (11.19 %). Fine-scale analysis using genome-wide SNP loci and the NETVIEW pipeline revealed cryptic genetic sub-structuring within localised geographic regions, which corresponded to the hierarchical mating system of the species. High levels of genome-wide SNP heterozygosity were observed amongst all populations (He = 0.25–0.35), and when evaluating across the species to other vertebrate taxa were amongst the highest values observed. This illustrates that the species as a whole still retains high levels of diversity which is comparable to other outbred vertebrate taxa for genome-wide SNPs. Insights into the potential for adaptive variation in the koala were also gained using

outlier analysis of genome-wide SNPs. A total of 10 putative outlier SNPs were identified indicating the high likelihood of local adaptations within populations and regions. This is the first use of genome-wide markers to assess population differentiation at a broad-scale in the koala and the first time that sex-linked SNPs have been identified in this species. The application of this novel genomic resource to populations across the species range will provide in-depth information allowing informed conservation priorities and management plans for in situ koalas across Australia and ex situ around the world.

Keywords

Koala
Phascolarctos cinereus
Genetic diversity
ddRAD
Selection
SNP

Shannon R. Kjeldsen and Kyall R. Zenger have contributed equally to this work.

Electronic supplementary material

The online version of this article (doi:10.1007/s10592-015-0784-3) contains supplementary material, which is available to authorized users.

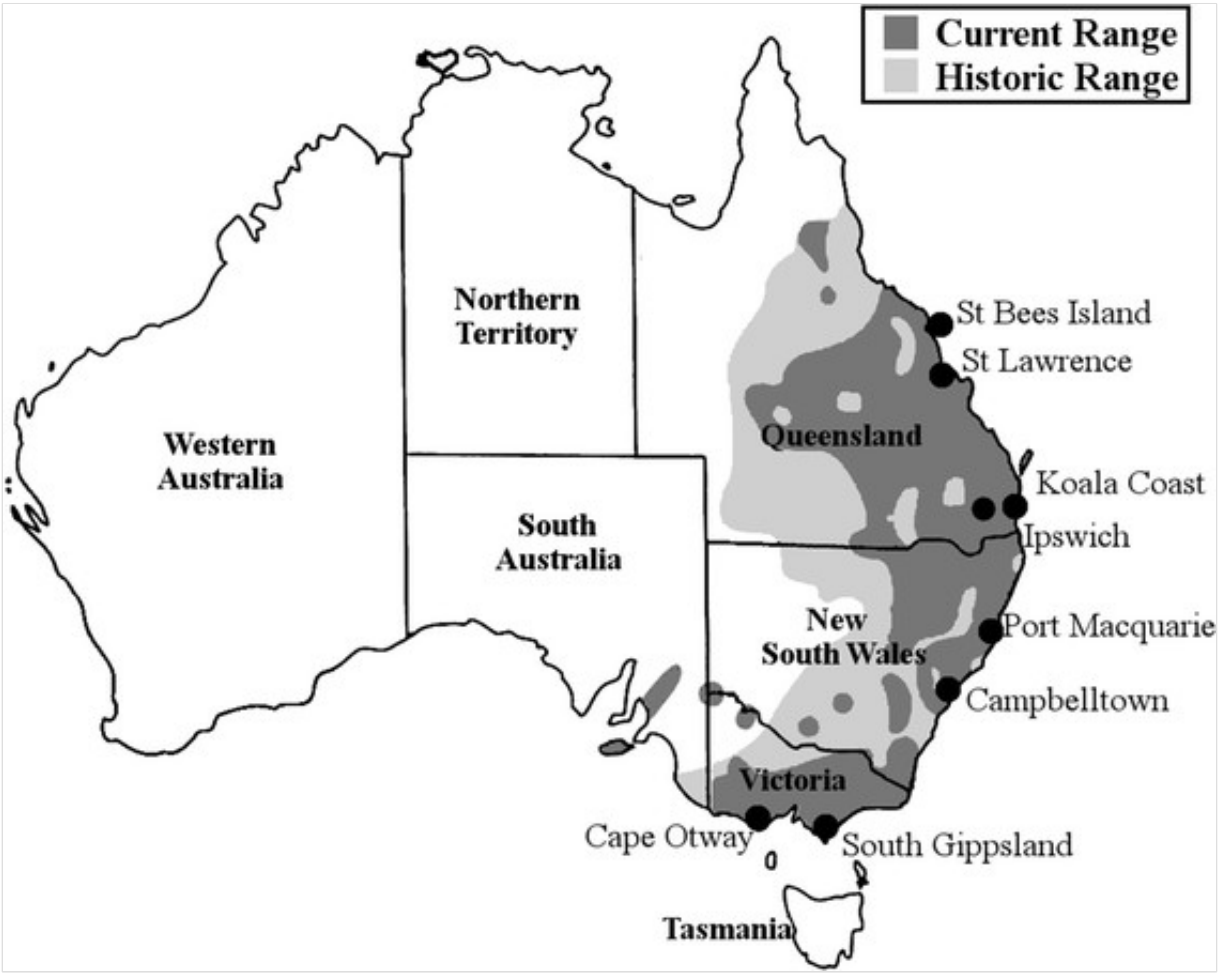
Introduction

Koalas (*Phascolarctos cinereus)* are one of Australia's most widely known and iconic marsupial species, found across the east coast of Australia from northern Queensland through to South Australia (Fig. 1). Koalas have faced a number of different conservation challenges throughout their history, from hunting for pelts, wide-spread disease including chlamydial infection and koala retrovirus and anthropogenic factors which led to habitat loss and predation by feral species (Avila-Arcos et al. 2013; Gordon et al. 2006; Melzer et al. 2000; Menkhorst 2008). There have been several documented events of local extinctions, range contractions, expansions and translocations throughout recent history. One of the greatest historical impact on the species was hunting in the early twentieth

century (Adams-Hosking et al. 2011; Gordon et al. 2006; Melzer et al. 2000; Menkhorst 2008). Despite numbers recovering well following extensive hunting, more recent loss of habitat, increasing urbanisation and disease have led to the koala being classified as vulnerable across most of its range (NSW, ACT and QLD) by the Federal Australian government, and listed as a threatened species under the US Endangered Species Act (Fish and Wildlife Service 1998; Natural Resource Management Ministerial Council 2009). Successful management is reliant on identifying the genetic structuring within and between populations, and identifying the genomic potential to adapt to different environments, particularly to inform captive breeding strategies (Moritz et al. 1996; Whisson et al. 2012). Information regarding fine-scale genetic relatedness within a population can also provide insights into social structuring of individuals within a region (Ross and Fletcher 1985). This information is particularly important in species where one or both sexes have a defined home range with a hierarchical social structure, as is the case with koalas (Mitchell 1990a; Thompson 2006; Ellis et al. 2009).

Fig. 1

Koala Distribution Map: Current national distribution of koalas and historical range with sampling sites indicated. Sample site references are outlined under ‘Map ID’ in Table 1. Adapted from distribution map created by Strahan et al. (1995)



Management of the koala has historically been a contentious issue as preliminary studies indicate high variability in population health and genetic structuring across Australia (Melzer et al. 2000). While in some regions koalas appear locally abundant, research indicates that numbers are generally declining across Australia (Gordon et al. 2006; Melzer et al. 2000; Natural Resource Management Ministerial Council 2009) and that local extinctions may currently be occurring (Lunney et al. 2002, 2014).

Previous population genetic studies have utilised neutral genetic markers, including minisatellites (Taylor et al. 1991), microsatellites (Houlden et al. 1996a, b; Lee et al. 2012a; Ruiz-Rodriguez et al. 2014), mtDNA (Houlden et al. 1999; Taylor et al. 1997; Tsangaras et al. 2012) and MHC genes (Houlden et al. 1996c; Lau et al. 2013), but have often had difficulty defining genetic structuring, especially on a national scale. To date, genetic studies on the koala have used at most 14 markers to estimate diversity and divergence within and between populations and this research has been limited to studies at a regional scale (Ellis et al. 2002; Lee et al. 2012a). A study by Houlden et al. (1996a) using six neutral markers, along with an assortment of other population genetic studies and anecdotal

evidence (Houlden et al. 1996b; Lee et al. 2010; Lee et al. 2012a; Martin 1985; Seabrook et al. 2011, 2002; Sherwin et al. 2000; Smith and Smith 1990; Wilmer et al. 1993) indicated a general decline in koala numbers and a reduction in genetic diversity. Other studies have suggested low diversity at a species level due to several significant bottlenecks and inbreeding, particularly in island populations (Fowler et al. 2000; Houlden et al. 1996b, c; Lee et al. 2012a; Tsangaras et al. 2012; Wilmer et al. 1993). Recent research by Lee et al. (2010) specifically noted a strong decline in genetic diversity in the south eastern Queensland region. Likewise, studies looking at MHC diversity in Victorian koalas have also noted significant founder effects due to translocations from bottlenecked populations (Lau et al. 2014). However, there is no baseline for genetic diversity in the koala as levels of genetic variation prior to European settlement are largely unknown (Houlden et al. 1996).

No studies have assessed the genomic health of koalas at a national scale (Lee et al. 2012a). Based on morphological and geographical information, koalas have been grouped into three sub-species, loosely corresponding to the state political borders along the east coast of Australia (Natural Resource Management Ministerial Council 2009). This classification was largely based on skull morphology and general phenotype (size and fur colour), which vary greatly across the species range. However, genetic studies carried out since the grouping was established did not detect sufficient levels of differentiation for a sub-species classification (Houlden et al. 1999). In other species, such as trout and wolf-like canids, genome-wide markers have been used successfully to assess both broad-scale population structuring and speciation and have consistently been shown to provide accurate assessments of species divergence (Pollinger et al. 2011; Stephens et al. 2009). In the koala however, broad-scale genetic structuring across the species range using genome-wide markers has not been investigated.

With an increase in the availability and affordability of next-generation sequencing technologies, genome-wide markers are becoming increasingly popular in evolutionary and ecological research. These genotyping by sequencing methods allow for unprecedented ease of research into non-model organisms. Many studies have indicated that the resolution provided by traditional markers such as allozymes, mini/microsatellites and mitochondrial DNA (mtDNA) is not comparable with their genome-wide

equivalents (DeFaveri et al. 2013; Miller et al. 2014; Rasic et al. 2014). Markers with a higher genomic density can provide more comprehensive genomic information enabling detailed studies of general diversity, divergence and adaptive variation (Angeloni et al. 2012; Davey et al. 2011). Bi-allelic single nucleotide polymorphism (SNP) markers are less informative per marker than microsatellites, but have the advantage of being highly abundant across the genome and are theoretically evenly spread across the genome (DeFaveri et al. 2013; Vignal et al. 2002) on coding and non-coding loci. Using complexity reduction methods such as double digest restriction-associated DNA sequencing (ddRAD) (Peterson et al. 2012), many thousands of SNPs can be identified for the same cost as developing only a few microsatellites, making them an attractive alternative to traditional marker sets (Peterson et al. 2012). The SNPs may also be present in both coding and non-coding regions of the genome and since many of these are likely to fall on the sex chromosomes (Carmichael et al. 2013) or in regions of the genome which are under selection, they can also be informative for research concerning sex-biased dispersal, sex-linked traits and adaptive variation. The identification of adaptive variation is also imperative to the effective management of vulnerable populations and to avoid possible outbreeding depression (Foll and Gaggiotti 2008; Novicic et al. 2011; Pilot et al. 2014). However, background selection, which may also occur within populations, must be taken into account (Comeron 2014; Tiffin and Ross-Ibarra 2014). Both directional and background selection can skew pairwise F_{ST} tests and can be difficult to untangle, but it is important to resolve these differences when we aim to use signatures of local adaptation as a proxy for determining the possibility of outbreeding in management plans (Edmands 2007; Gravuer et al. 2005). Genome-wide SNP genotyping is now considered the preferred marker for population based diversity studies in many species, and can be a versatile tool to provide insights into genetic structuring and the micro-evolutionary processes in the koala. These markers can be used to address critical questions for the species, including assessments of speciation (Jones et al. 2012; Leache et al. 2014), inter-regional and intra-regional population diversity and relatedness (Johnston et al. 2014; Larson et al. 2014), inbreeding, signatures of population reductions, effective populations sizes (Ne) (Johnston et al. 2014), parentage (Fernández et al. 2013), and evidence of adaptive variation (Nielsen et al. 2005). In this study, we report on the development of the first panel of genome-wide SNP markers in the koala that will serve as an important tool for the

conservation of the species and can be used to provide insights into population structuring and variability across the species range. This study assessed (i) the viability of ddRAD as a tool to rapidly discover SNP markers in this species (ii) levels of genome-wide variation within animals from eight geographically distinct regions from across the species range, and (iii) the potential utility of the SNP markers we identified to answer a broad range of ecological questions, including broad- and fine-scale population structuring.

Materials and methods

Sampling and DNA extraction

To evaluate genetic differentiation amongst geographically distant or potentially bottlenecked populations using genome-wide SNP loci, we obtained blood and tissues samples from koalas in representative regions of Queensland, New South Wales and Victoria, along with one introduced population on St Bees Island, QLD (see Fig. 1 for sampling locations). Tissue samples were opportunistically obtained from William Ellis, Sean FitzGibbon and Alistair Melzer and other researchers (see acknowledgements) and preserved in 70 % ethanol, while whole-blood samples were stored at −20 °C. To ensure the highest quality of DNA for ddRAD library preparation, all DNA samples were extracted using a modified CTAB/Cholorform-Isoamyl method (Adamkewicz and Harasewych 1996) and further purified using a Sephadex G-50 approach (GE Healthcare Life Sciences 2000) to ensure no inhibitors were carried through to ddRAD library preparation.

ddRAD library preparation

In silico simulations of EcoRI and MspI double digests were first performed on the Tamar wallaby (*Macropus eugenii*) and Opossum (*Monodelphis domestica*) genomes and extrapolated to the equivalent koala genome size, to evaluate the effectiveness of these restriction enzymes on reducing the complexity of the koala genome for ddRAD library preparation (Peterson et al. 2012). Simulations were carried out using the SimRAD restriction enzyme analysis package (Lepais and Weir 2014) with different fragment size selection windows in order to obtain 20,000–30,000 regions according to Peterson et al. (2012). A size selection window of 450 ± 44 bp was selected based on the number of theoretical

fragments (28,590) and size selection accuracy of the PippinPrep targeted size selection machine (Peterson et al. 2012).

Libraries were generated using a modified version of the Peterson et al. (2012) ddRAD protocol. Briefly, individual genomic DNA (~1 µg) was digested overnight with 10 units EcoRI-HF and 10 units MspI at 37 °C. All digests were checked on a 0.8 % agarose gel to ensure complete digestion. Digests were then cleaned using Sera-Mag SpeedBead Carboxylate-modified Microparticles (Thermo Scientific 2014) and quantified accurately using the Biotium ACCULBLUE High Sensitivity dsDNA quantification kit (Biotium 2013). Digested samples were standardised to 400 ng and sorted into groups of 48. A ligation reaction was carried out where unique in-line barcodes and a common biotinylated adaptor were added to each fragment using T4 Ligase and buffer (Peterson et al. 2012). Ligated samples were then pooled into their sets of 48 for barcoding, cleaned using Sera-Mag SpeedBeads Carboxylate-modified Microparticles (Thermo Scientific 2014) to remove excess adaptors and reduce the volume to 50 µl. These pools were quantified using a NanoDrop Spectrophotometer (Desjardins and Conklin 2010). No more than 5 µg of each pool of 48 samples was loaded onto the PippinPrep targeted size selection machine (Sage Science Inc 2013) and a size range of 450 ± 44 bp was selected. A 2 % Agarose gel cassette with Ethidium bromide and no overflow detection (Sage Science Inc 2013) was used to size select samples. Two elutions were taken along with a 0.1 % Tween-20 rinse in order to maximise recovery. Eluate was cleaned using Streptavidin magnetic beads (Thermo Scientific 2014) to remove any fragments lacking a biotinylated adaptor. Illumina flow cell adaptors (P1) and one of 12 unique indexes (P2) were attached to each fragment via PCR using a Biorad C1000 thermal cycler: initial denaturation at 98 °C for 30 s, followed by 14–17 cycles of 95 °C for 15 s each, 66 °C for 30 s and 72 °C for 45 s and a final extension step at 72 °C for 600 s. Multiplexed PCR products were pooled and cleaned again using Sera-Mag SpeedBeads Carboxylate-modified Microparticles (Thermo Scientific 2014) to reduce the volume again to 50 µl. Each pool of 48 was run on an Agilent 2100 Bioanalyzer (Agilent Technologies 2013) to ensure that size distribution was uniform between pools and to quantify for equimolar pooling prior to sequencing. The final multiplexed library was sent in a 20 nmol concentration to the Australian Genome Research Facility (AGRF) for paired-end sequencing (101 bp) on the Illumina HiSeq2000 platform at an

average read depth of ~1 million reads per individual.

Quality control and SNP filtering

The quality of raw sequences was screened using FastQC program (Andrews 2010) and reads with an average Q-score of <30 were discarded. To call individual genotypes, each library was processed through the *denovo_map.pl* pipeline in STACKs v1.20 (Catchen et al. 2011). This software de-multiplexes, quality checks, aligns and calls sequence variants across individuals and amongst populations. To avoid complications associated with mis-indexing, a combination of a unique 5 bp barcode coupled with a 6 bp index was used for each sample, in conjunction with the standard Illumina TruSeq indexes (Peterson et al. 2012). In order for a read to be retained, it needed to have all three barcodes completely intact, with no mismatching allowed. Any reads containing low quality or ambiguous barcodes were discarded. Following the extraction of individual trimmed reads, paired-end sequences were concatenated to form a single continuous read of 195 bp, using custom bash scripts. Sequence variant calling in STACKs was conducted with default program parameters with the exception of mismatches when aligning loci within individuals (*ustacks*—n = 6), and further mismatches allowed when creating a reference catalogue of loci (*cstacks*—m = 3). Furthermore, a minimum sequencing depth (*populations*—m = 10) and a minor allele frequency (MAF) (*populations*—a = 0.02) were selected (Zenger et al. 2007), to create the final filtered genotype file. In order to minimise negative effects of missing data when calculating frequency-based genetic distance parameters (see Willing et al. 2012), only a single SNP was retained in each locus and only SNPs which were genotyped in >10 individuals and common to >2 populations. Additionally, all common autosomal loci deviating from Hardy–Weinberg equilibrium (HWE) in all populations were identified using the program Genetix (Belkhir et al. 1996) and removed if significantly ($P < 0.01$) deviating from HWE. In order to identify contamination, all loci were searched against bacterial and viral databases (Johnson et al. 2008) and any matching regions were removed from the dataset. For frequency-based analyses, a sufficient level of individuals needed to be genotyped at a locus to make assumptions at a population level (Huang and Knowles 2014).

Sex-linked markers

Sex-linked loci were identified by comparing expected Mendelian patterns of loci in individuals of known sex (males = 36, females = 28). To ascertain if the loci matched expected patterns of autosomal loci, a Fisher's exact test (Altham 1969) was carried out with a correction for false discovery rate (FDR) of 10 %. If the locus indicated sex-linked Mendelian inheritance (X or Y) in the test individuals, they were added to a short list. Meaning that to be short listed as X-linked, all males needed to appear homozygous (despite actually being hemizygous). To be shortlisted as Y-linked, all males needed to be homozygous and the locus was required to be missing in all females. As an additional test for Y-linked loci, deviation from HWE (in males) and linkage disequilibrium (LD) tests were conducted using Arlequin (Schneider et al. 2000) across all remaining individuals. Finally, shortlisted loci were evaluated against known marsupial sex chromosome sequence data (*M. eugenii*—Genbank accession ABQO000000000.2 and *M. domestica*—Genbank accession GCF000002295.2). In order for a locus to be confidently classified as X- or Y-linked, it needed to adhere to all of the aforementioned tests. All short listed loci were removed from the autosomal dataset.

Population genetic diversity

To evaluate the genetic diversity within populations, standard diversity indices including average expected heterozygosity (He), average observed heterozygosity (Ho), allelic diversity and inbreeding coefficient (F_{IS}) were calculated through Genetix (Belkhir et al. 1996). Additionally, effective population size (Ne) was calculated with NeEstimator (Peel et al. 2004) using the linkage disequilibrium option (Ne_{LD}). To assess individual genome-wide diversity and inbreeding measures, multi-locus heterozygosity (MLH) and internal relatedness (IR) were calculated for all individuals using the R package *Rhh* (Alho et al. 2010).

Broad-scale population divergence

To illustrate the usefulness and consistency of this dataset, a number of different analyses were performed to assess broad-scale diversity and divergence. Genetic differences between populations and their significance were evaluated using Weir and Cockerham's unbiased F-statistics (Weir and Cockerham 1984), and hierarchical analysis of molecular variance (AMOVA) between populations, states, and regions (north, middle and

south east coast) was calculated in Arlequin (Schneider et al. 2000). Broad-scale relationships between populations were visualised by constructing a neighbour-joining (NJ) genetic distance tree with Nei's standard distances (Nei et al. 1983). Genetic distances were calculated using Microsat2 (http://genetics.stanford.edu/hpgl/projects/microsat/) and tree reconstruction was performed using Mega6 (Tamura et al. 2013). Population assignments were confirmed using an assignment test carried out using GeneClass2 (Piry et al. 2004). A principal component analysis (PCA) using prior population clustering was conducted using the R package, *adegenet* (Jombart 2008) and was subsequently visualised through a DAPC scatterplot. The populations of Campbelltown and South Gippsland were excluded from the PCA as small sample size and lower genotyping rate were found to bias results.

Fine-scale population structure resolution

The utility of genome-wide loci to unravel fine-scale genetic structuring was assessed across all individuals in the Queensland populations (St Bees Island, St Lawrence, Koala Coast and Ipswich) as well as in the large, isolated Port Macquarie population. Family groups were identified through the calculation of maximum likelihood (ML) estimates of relatedness and their relationships within the program ML-Relate (Kalinowski et al. 2006). Individual relationships amongst all individuals within the QLD populations and across all populations were then calculated and visualised using the NETVIEW pipeline v0.5.1 (Steinig et al. 2015 accepted) at kNN values between 5 and 35.

Outlier loci detection

To identify any candidate loci under selection in three populations with low genetic differentiation ($F_{ST} < 0.10$) within QLD (Ipswich, Koala Coast and St Lawrence), outlier analyses were conducted using a frequency-based approach using Lositan (Antao et al. 2008). To identify outlier SNPs in Lositan, a total of 50,000 simulations were run at an FDR level of 0.1 at a 95 % confidence interval (CI), with a "Neutral" mean F_{ST} being used and a "Forced mean F_{ST}" being fitted under an infinite allele model. Biological replicates were used in the form of two environmentally similar areas; Ipswich and Koala Coast. Only markers which were highly differentiated from the simulated mean and which were present in >2 pairwise tests were called as outliers. To circumvent documented issues associated with purely

F_{ST}-based outlier detection methods (Narum and Hess 2011), a Bayesian method was also utilised, with 1:10 prior odds for a neutral model and all other parameters left as default (20 pilot runs of 5000 iterations followed by 100,000 iterations with an additional burn-in of 50,000) (Foll 2012) . Alpha levels and F_{ST} values were ordered from largest to smallest and outliers were then identified at FDR levels of 0.1, 0.15 and 0.2 using the Bayescan 2.01 function, *plot_R.r*. Both directional (diversifying) and balancing or purifying (background) selective markers were putatively identified, with outliers returning a positive alpha value being deemed diversifying and those returning a zero or negative alpha value being considered to be under balancing or purifying selection (Foll and Gaggiotti 2008; Novicic et al. 2011; Pilot et al. 2014). Outliers identified in both Lositan and Bayescan were then compared at their predefined FDR level. Candidate outlier SNPs were only confidently called if they appeared in both methods (Kovach et al. 2012; Larmuseau et al. 2010; Pujolar et al. 2014).

Results

Sequencing and SNP discovery

A total of 317,573,718 paired-end sequence reads were obtained across all 171 unique individuals. After quality filtering through the initial modules of the STACKs pipeline, 6.68 % of total reads were discarded due to low quality scores (Q-score < 30) and ambiguous barcode sequencing. These reads were clustered into a catalogue of 1,088,361 RAD loci across all individuals which were used to confirm genotyping calls. The median number of reads per individual was 750,299 and ranged between 148,615 and 3,037,452. An average of 33,019 stacks or ‘loci’ were recovered from each individual at an average read depth of 17.3 reads per stack. Based on a minimum read depth of 10 and a MAF of >0.02, 13,998 polymorphic SNPs were retained across all individuals. Higher levels of missing data were observed between geographically distant populations, which has also been noted in more recent studies involving the ddRAD protocol (Andrews et al. 2014). This is largely attributed to a higher likelihood of mutations in enzyme restriction sites in more divergent populations, leading to disproportionate sampling between individuals (Andrews et al. 2014). To overcome this problem and to test the robustness of the dataset, various levels of missing data were run through each analysis to ensure

minimisation of bias. A significant change in both heterozygosity and F_{IS} was observed when evaluating populations with read depths of below 10 at each locus. To rule out the effects of missing data, a subset of 500 of the most common SNPs were re-analysed for F_{IS}. Results of this re-testing did not reveal any significant differences in F_{IS} values ($P > 0.05$). Inbreeding coefficients were high in all populations and ranged from 0.11 to 0.32 with Koala Coast displaying the highest F_{IS} value (0.32). Missing data skewed F_{ST} and F_{IS} results when <10 individuals were genotyped at a locus within a population and so an extra filtering step to ensure >10 individuals were genotyped at each locus, in at least two populations was carried out on this dataset. Although there were some differences in SNP profiles between populations, the datasets were still large enough with sufficiently high density so that this SNP incompleteness did not affect the outcome in this case (Huang and Knowles 2014). Of the loci recovered, 311 were putatively identified as sex linked based on genomic database matches and these were removed from the 'autosomal' dataset, with 95 loci (X chromosome = 58, Y chromosome = 37) adhering to the two most informative identification criteria (Mendilian patterns and BLAST matches to marsupial sex chromosomal regions). Following selection of a single SNP per locus, conformity to HWE testing across populations, minimum number of individuals genotyped in each population and screening for sequence contamination, a final set of 3060 polymorphic autosomal SNPs were retained for further analysis (Supplementary Material 1).

Population genetic diversity

Average observed heterozygosity (Ho) ranged from 0.23 through to 0.29 and average expected heterozygosity (He) ranged from 0.26 to 0.35 (Table 1). Estimated effective population size (Ne_{LD}) across individual populations ranged from 2.7 (Campbelltown 95 % CI = 2.4–3.2) to infinity, with Campbelltown returning the smallest Ne_{LD}, likely due to small sample size and sub-sampling effect. Cape Otway and Port Macquarie returned Ne_{LD} = 46.7 [95 %CI 40.8, 54.4] and 116.8 [95 %CI 109.8, 124.6] respectively (Table 1). Individual MLH averaged over each population ranged from 0.18 to 0.29, standardised MLH (sMLH) being highest in Cape Otway (Table 1) and no private alleles were identified in the filtered dataset. Average IR, another measure of population diversity (Alho et al. 2010), ranged from 0.20 to 0.42.

Table 1

Genetic diversity indices for each koala population sampled

State	Location	n	Ho	He	Fis (P < 0.01)	sMLH (±SD)		IR (±SD)
QLD	St Bees Island	19	0.29	0.35	0.23	0.29 (±0.06)	1.05 (±0.24)	0.29 (±0.15
QLD	St Lawrence	19	0.26	0.30	0.20	0.24 (±0.04)	1.03 (±0.15)	0.21 (±0.11
QLD	Koala Coast	24	0.22	0.30	0.32	0.18 (±0.09)	0.79 (±0.39)	0.42 (±0.29
QLD	Ipswich	23	0.27	0.31	0.19	0.25 (±0.06)	0.98 (±0.21)	0.26 (±0.16
NSW	Port Macquarie	45	0.23	0.28	0.21	0.23 (±0.05)	0.94 (±0.20)	0.25 (±0.15
NSW	Campbelltown	9	0.27	0.33	0.27	0.25 (±0.11)	0.97 (±0.41)	0.34 (±0.27
VIC	South Gippsland	19	0.24	0.30	0.27	0.24 (±0.11)	1.01 (±0.48)	0.31 (±0.34
VIC	Cape Otway	13	0.24	0.25	0.11	0.24 (±0.04)	1.07 (±0.16)	0.20 (±0.11

Broad-scale population divergence

Pairwise F_{ST} values between populations displayed varying levels of genetic differentiation, ranging from 0.28 between St Lawrence and South Gippsland, to 0.01 between Koala Coast and Ipswich—the two closest sampled populations (Table 2) . Interestingly, despite being geographically further apart, Port Macquarie revealed less differentiation from the population in Ipswich, QLD (F_{ST} 0.11) than from Campbelltown, NSW (F_{ST} 0.13). Differentiation between mainland populations within QLD was low (F_{ST} 0.01–0.08) when compared to the two Victorian populations (F_{ST} 0.10). Assignment tests correctly assigned 100 % of individuals to their source population, with only 15.85 % of individuals being assigned to a second, geographically close population. Broad-scale genetic structuring using NETVIEW revealed three major genetic clusters

across the sampled populations, with the QLD populations in one large group, Port Macquarie clustering out on its own, and Cape Otway clustering apart from Port Macquarie (Fig. 2a). Principal component analysis through DAPC (Jombart 2008) indicated three genetic groups (Fig. 3), again with Port Macquarie and Cape Otway forming their own defined clusters, similar to the NETVIEW clustering. South Gippsland and Campbelltown populations were excluded from these analyses due to low sample size and less complete genotyping. The NJ distance tree also indicated three broad genetic clusters across the sampled populations, however here the NSW populations grouped more closely with the QLD populations, than with the southern groups (Fig. 4). To further assess this demarcation of genetic groups, a hierarchical AMOVA was carried out within state boundaries, as per current sub-species classifications and with clustering groups revealed in DAPC and NETVIEW analyses. Only 11.19 % of variation could be accounted for between sub-species groups, with a larger portion of variation being observed among populations within groups (28.29 %) and among individuals within populations (46.94 %). These results are not in keeping with current sub-species classifications (Goldfuss 1817 in Iredale and Troughton 1934; Thomas 1923; Troughton 1935).

Table 2

F_{ST} values based on Weir and Cockerham's unbiased genetic distances (1984)

	St Lawrence	**St Bees Island**	**Ipswich**	**Koala Coast**	**Port Macquarie**	**Campbel**
St Bees Island	0.10	+	+	+	+	+
Ipswich	0.06	0.15	+	+	+	+
Koala Coast	0.08	0.14	0.01	+	+	+
Port Macquarie	0.15	0.20	0.11	0.13	+	+
Campbelltown	0.19	0.21	0.15	0.17	0.13	+
South Gippsland	0.28	0.22	0.23	0.25	0.25	0.20
Cape Otway	0.24	0.21	0.23	0.24	0.21	0.21
+ denotes value significant to $P < 0.05$						

Fig. 2

a Population clustering of all populations using an isolation by state (IBS) matrix constructed using the NETVIEW v5.0 pipeline visualised at kNN = 10. **b** and **c** Population clustering of all QLD populations using an isolation by state (IBS) matrix constructed using the NETVIEW v5.0 pipeline visualised at kNN = 5 (**b**) and kNN = 25 (**c**)

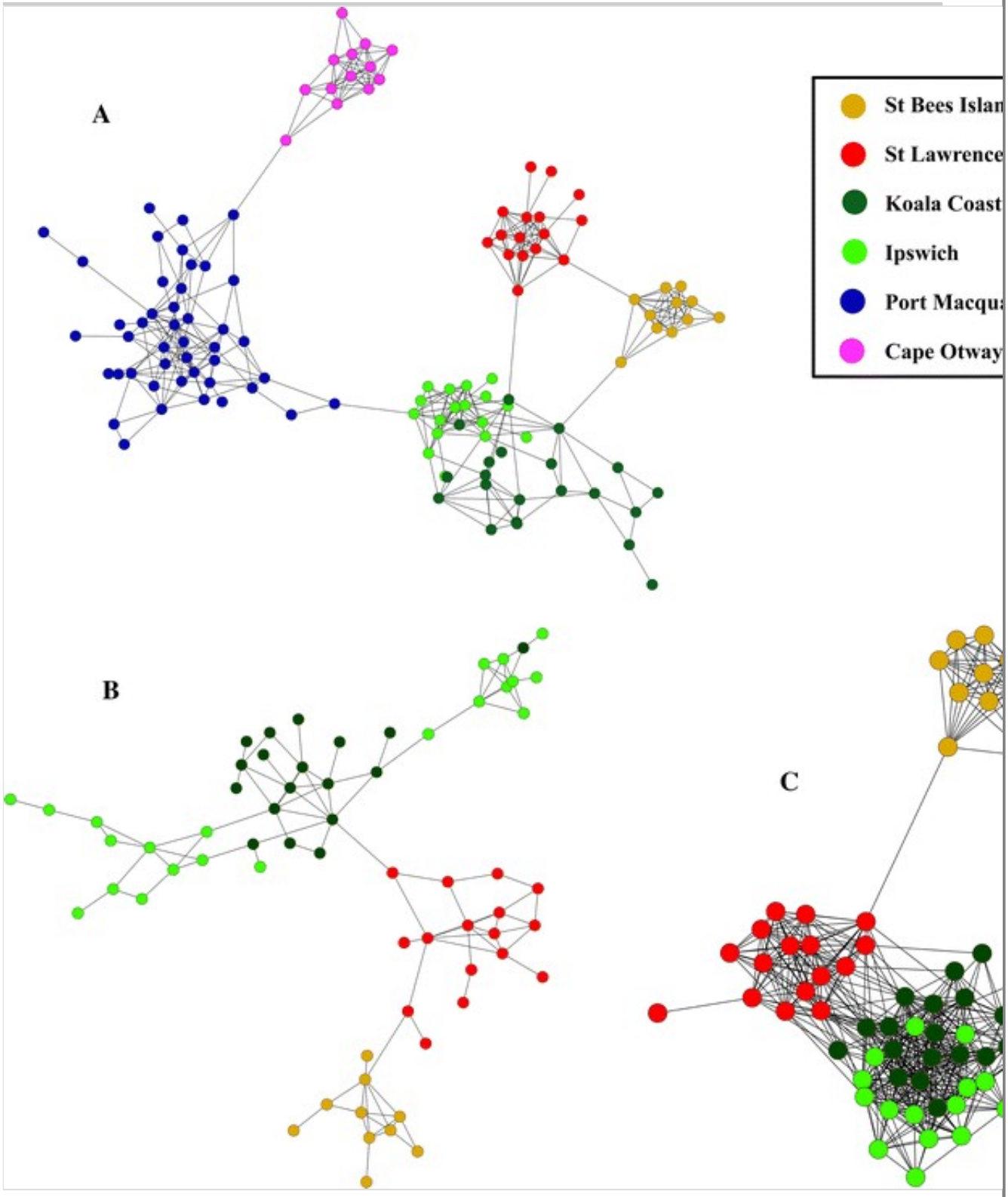


Fig. 3

DAPC scatterplot created through the R package *adegenet* with prior population clustering. Sample populations are outlined under ‘Map ID’ in Table 1

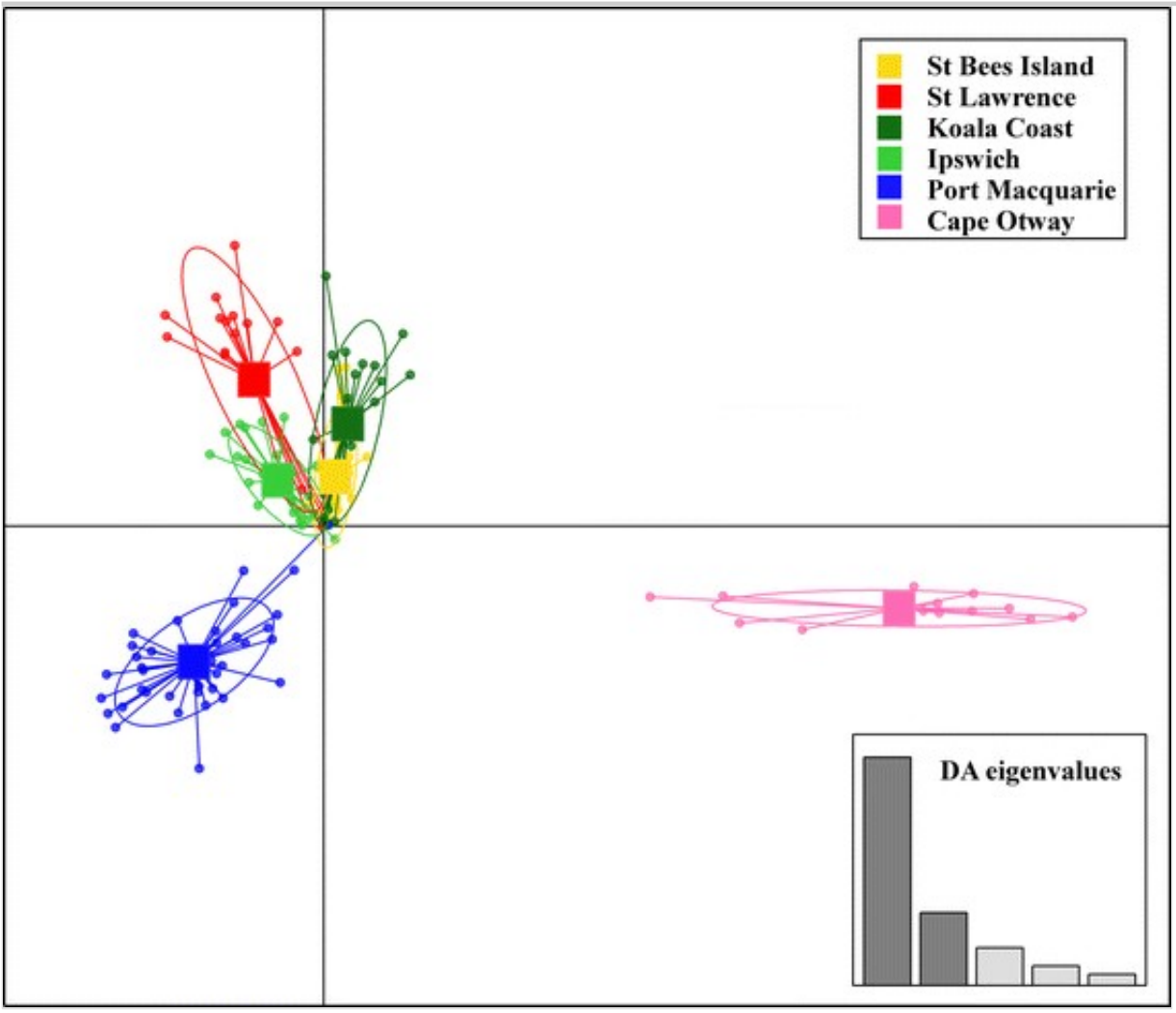
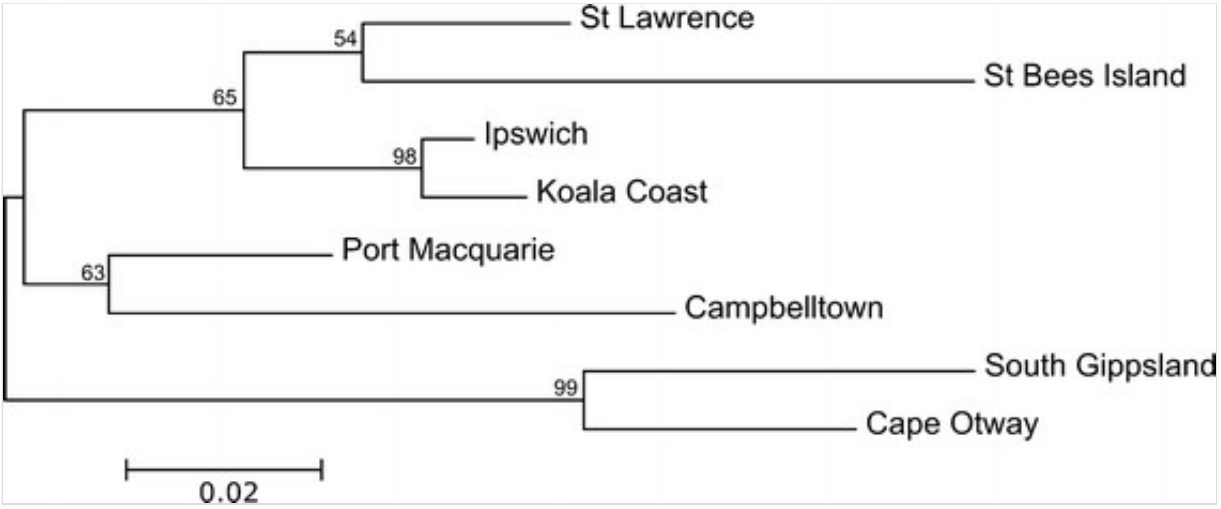


Fig. 4

Evolutionary relationships between populations and regions, inferred using the Neighbor-Joining method with Nei’s standard genetic distance. The optimal tree with the sum of branch length = 0.37 is shown



AQ1

AQ2

Fine-scale population structure resolution and adaptive variation

Within the QLD group of populations, NETVIEW clustering at kNN = 5 (Fig. 2b) was able to reveal fine-scale family structuring, showing small groups of 3–5 individuals clustering closely together. St Bees Island was the most distantly related to the mainland groups, with most genetic links being observed between St Lawrence and St Bees Island. A total of 31 half-sibling relationships (ML-relatedness > 0.15) were identified in all QLD populations, along with 14 full-sibling and six parent-offspring pairings. Possible family groups were also identified within Ipswich, Koala Coast and St Lawrence, although at a NETVIEW clustering of kNN > 10. These three populations appeared to be highly related (Fig. 2b, c), despite St Lawrence being >600 km away from either Ipswich or Koala Coast.

This relationship is also shown in low F_{ST} values (0.01–0.08) within this region (Table 2.). Due to its large sample size and distribution, the Port Macquarie population was also assessed for fine-scale structuring. In Port Macquarie (as within the QLD region), trio and small family groups were identified using NETVIEW analysis at kNN = 5 and low level relatedness (0.01–0.10) was observed across most individuals sampled. A total of eight half-sib pairs, three full-sib pairs and one parent-offspring pair were identified using ML-relatedness measures. At a clustering level of kNN > 10 in NETVIEW two sub-groupings begin to become apparent which mirror the sampling location of individuals (i.e. northern Port Macquarie and southern Port Macquarie region). The NETVIEW analyses of both QLD and Port Macquarie indicate strong localised sub-structuring

within these regions, which is supported by inflated F_{IS} values (i.e., Wahlund effect).

Outlier loci were identified in three populations in QLD with low levels of genetic differentiation (FST < 0.10) at FDRs of 0.1, 0.15 and 0.2 using a frequency-based approach in Lositan (Antao et al. 2008) and a Bayesian approach in Bayescan 2.01 (Foll and Gaggiotti 2008). In Lositan, the overall simulated F_{ST} mean was low between the three tested populations (F_{ST} = 0.058). In total, 48 candidate directional outlier SNPs were identified as having a significantly higher F_{ST} than the simulated average between all QLD populations (F_{ST} 0.359–0.786). Furthermore, 35 of these were common to at least two pairwise population tests. A total of 16 outlier SNPs were identified in the Bayescan analysis with 11 being common to at least two pairwise population tests and appearing as common to the biological replicates. These markers also revealed alpha levels of significantly higher than 0 (alpha level range 0.66 – 1.29), which is indicative of directional selection, rather than balancing or background selection (Foll and Gaggiotti 2008; Novicic et al. 2011; Pilot et al. 2014). Of these, 10 markers were called in both analysis methods, indicating that these markers are likely to be biologically relevant. Out of the 10 common outlier SNPs, 7 were classified as balancing or purifying loci in pairwise comparisons between Ipswich and Koala Coast (biological replicates) and then as directional outliers when both populations were compared against St Lawrence. All candidate outlier SNP loci were searched against available marsupial genomes (*M. eugenii* – Genbank accession ABQO000000000.2 and *M. domestica* – Genbank accession GCF000002295.2), however no significant matches were observed.

Discussion

Population diversity and broad-scale divergence

Refining transition zones and management boundaries is crucial to any future conservation efforts for the koala. Based on our results and the high level of diversity in morphology that has been observed in the koala (Houlden et al. 1999), a successful conservation plan for the koala must strike a balance between preserving a maximum level of diversity and avoiding potential complications from admixture, such as outbreeding depression (Frantz et al. 2006; Schwartz et al. 2007; Whisson et al. 2012).

The application of high throughput next generation sequencing of ddRAD loci allowed the development of the first large genome-wide SNP dataset in a diverse panel of koalas. From a practical perspective, the markers developed using this method will provide the tools for a much more comprehensive assessment of the conservation status and genetic diversity of both wild and captive populations of the koala across much of its range. It will also allow for the development of more effective genetic management strategies, defining management boundaries for the species across all geographical cohorts and in captive breeding programmes.

The results outlined in this study have provided the first insights into genome-wide diversity in koala populations across their range. Overall diversity within populations was revealed to be equivalent to other vertebrate species, including canid, felid and ungulate species (Table 3). When compared to other vertebrate species genotyped using SNP markers and with a similar ecology and anthropogenic history, the populations sampled in this study show equivalent or greater values of both observed and expected heterozygosity (Table 3). Compared to large, constant populations such as wolves in North America (Gray et al. 2009) and feral pig populations in Europe (Goedbloed et al. 2013), the koala showed notably higher levels of diversity when assessed using genome-wide markers. Furthermore, multi-locus heterozygosity (MLH) was also revealed to be high in koalas in this study (MLH = 0.23). In contrast, an inbred population of harbor seals and a strain of outbred mice (Hoffman et al. 2014) demonstrated a lower level of MLH (0.06 and 0.16 respectively). Despite a general view that koalas have reduced genetic diversity across their range, based on limited studies using neutral markers (Lee et al. 2010; Melzer et al. 2000; Phillips 2000), the results of this SNP study, which included samples from five of the sites used in previous studies (although not the same animals), indicate that in fact they may have equivalent genetic diversity to other stable outbred wild taxa (Table 3). Pre-emptive management through detailed monitoring of genetic diversity in national koala populations is important to ensure populations remain genetically diverse, especially given the current threats to the species such as disease and habitat loss (Melzer et al. 2000), however this study indicates that more passive management and monitoring may be preferable to active intervention at this point (at least for the populations sampled in this study). It is worth noting that while species-wide diversity was

generally high compared to the aforementioned taxa, there was considerable variation in levels of diversity within the koala populations studied and sampling is proposed to identify populations that have limited genetic diversity and may require more targeted management and monitoring. The Koala Coast population showed a reduced level of diversity when compared to the others in this study, specifically showing lower heterozygosity and higher average IR across all individuals (Table 1). This population has at least one recorded historical bottleneck and has shown a similar trend in other studies (Lee et al. 2010). Heterozygosity was generally high in St Bees Island, in contrast to other island populations, such as Kangaroo Island, which have been reported to be highly inbred (Lee et al. 2012a). Our results concur with another study that suggested St Bees Island may be one of the most genetically diverse of all island populations (Lee et al. 2012a). However, as no other studies have been conducted using genome-wide SNPs, it is difficult to make direct comparisons with other island groups at this stage. The lower levels of heterozygosity observed in South Gippsland are supported by extensive documentation of hunting and local extinction in Victorian populations in the early twentieth century (Houlden et al. 1999; Wilmer et al. 1993). It would be useful to further sample animals in the Strzelecki region for comparison, as this population is said to have escaped much of the historical hunting and so may retain more historical diversity (Lee et al. 2012b).

Table 3

Heterozygosity values of various species based on SNP genotyping methods

Common name	**Species**	**He**	**Ho**	**References**
Outbred or large, wild populations				
Deer	*Odocoileus spp.*	0.31	0.30	Haynes and Latch (2012)
Koala	*Phascolarctos cinereus*	0.29	0.27	This study
Feral Pig	*Sus scrofa domesticus*	0.28	0.27	Goedbloed et al. (2013)
Chinook Salmon	*Oncorhynchus tshawytscha*	0.26	0.25	Narum et al. (2008)
Coyote	*Canis latrans*	0.25	0.20	Kyle et al. (2006); Koblmuller et al. (2009)

Wolf (North American)	*Canis lupus*	0.24	0.22	Gray et al. (2009); Cronin et al. (2015)
Populations with a recent known bottleneck or domesticated species				
Domestic Cat (Russian Blue)	*Felis catus*	–	0.19	Kurushima et al. (2013)
Eurasian Beaver	*Castor fiber*	0.19	0.17	Senn et al. (2014)
Angus	*Bos taurus*	–	0.17	MacEachern et al. (2009)
Brown Bear	*Ursus arctos*	0.17	0.16	Cronin et al. (2014), Miller et al. (2012)
Wolf (Italian)	*Canis lupus*	0.17	0.15	Gray et al. (2009); Fabbri et al. (2007)
Arctic Ringed Seal	*Pusa hispida hispida*	0.14	0.13	Olsen et al. (2011)
Holstein	*Bos taurus*	–	0.12	MacEachern et al. (2009)
Polar Bear	*Ursus maritimus*	0.04	0.05	Cronin et al. (2014)
Black Bear	*Ursus americanus*	0.02	0.02	Cronin et al. (2014)

While there is no known reproductive divide between groups of koala, this study indicates a significant level of genetic differentiation between geographic regions. All analysis performed indicated that northern and southern subgroups were genetically distant from one another, with Port Macquarie separating out on its own. Interestingly, Port Macquarie clustered more closely to some of the southern QLD populations than when compared with corresponding NSW groups despite geographic proximity to NSW groups. This trend has also been observed in MHC diversity studies across the species range (Lau et al. 2014). While it is clear that a general isolation by distance effect is being observed, further sub-sampling of populations within Northern NSW is needed to complete our understanding of the relationships between these groups. The AMOVA results showed very little differentiation between states overall (11.19 %) and affirmed the lack of support for any sub-species assignment. The NJ tree (Fig. 4) shows three genetic clusters, but also reveals that the NSW populations cluster more closely with populations in QLD, rather than their

Victorian counterparts. This is supported by smaller F_{ST} values between these groups, and it appears that across Australia only two large genetic groups may be present, although a denser sampling strategy should be observed in future studies to confirm this. The largest variation between groups was observed between two of the most geographically distant mainland populations assessed in this study (St Lawrence, QLD and South Gippsland, VIC) and genetic distances between populations were generally high. However, F_{ST} values of 0.3–0.4 and a greater variation between sub-species groups would have been needed to provide support to any sub-species classification (Frankham et al. 2002), and this was not observed in this dataset. Genome-wide marker sets have been used to resolve sub-speciation and ecological management questions in a number of aquatic animal and plant species, including trout and rice (Feltus et al. 2004; Stephens et al. 2009). Similarly, the genome-wide SNPs developed in this study will help to resolve these issues in the koala, provided an additional and denser sampling strategy is employed to clarify potential 'transition zones'.

Effective population size can be an indicator of population health within a species. Given that this study only had access to a single temporal sample, a linkage disequilibrium method was used. Several population sample sizes were not sufficient to attain an accurate estimate of Ne_{LD}, with estimates for St Bees Island, St Lawrence, Ipswich and South Gippsland reporting infinite Ne values. The fact that most populations still returned an infinite Ne_{LD} indicates that the true values may be quite large, since if the small number of individuals sampled were highly related, a small Ne_{LD} would have been expected. The two populations that returned robust values, Cape Otway (Ne_{LD} = 46.7 [95 %CI 40.8, 54.4]) and Port Macquarie (Ne_{LD} = 116.8 [95 %CI 109.8, 124.6]), both had the highest genotyping rate and Port Macquarie also had the largest sample size. Further sampling across multiple time periods and generations would be required for more accurate estimations, as it is possible that social sub-structuring may have skewed results in this case (Luikart et al. 2010). It is generally accepted that an Ne of 50 to 100 is sufficient for maintenance of short-term fitness (Shaffer 1981), and an Ne of roughly 10 % of the total census size allows a species to avoid an 'extinction spiral' (Frankel 1981). An accurate national census size is difficult to attain in the koala as most estimates rely on public sightings, which can skew results, or small scale transect distance-sampling techniques, which are expensive and time-consuming so

are rarely carried out on a large scale. Given that koalas have been regarded as having a hierarchical social system (White 1999; Mitchell 1990a, b) and general observed heterozygosity was high, this study concluded that the small Ne and the inflated F_{IS} values were due to social structuring and Wahlund effect (Sinnock 1975) rather than any excessive inbreeding within populations.

Fine-scale population structure resolution

The ability of genome-wide markers to investigate fine-scale genetic structuring and relatedness (Cánovas et al. 2014; Consortium 2009; Kijas et al. 2009; Miller et al. 2012; Pollinger et al. 2011; Stephens et al. 2009) has been demonstrated successfully in this study, teasing apart relatedness and structuring in both the QLD region and Port Macquarie koalas. Within the QLD region, a historic link between the St Lawrence population and two south eastern QLD populations has been identified using genetic distance clustering (identity by state) through the NETVIEW pipeline (Steinig et al. 2015 accepted), indicating historical links given the extremely low probability of translocations between these areas. Lower levels of differentiation ($F_{ST} \leq 0.10$) and links defined within NETVIEW clustering (Fig. 2b, c) are supported by historical records of translocations between central QLD and the introduced population of St Bees Island (Berck 1995). Our study also revealed there may have historically been active gene flow between koalas in the Ipswich region and the Koala Coast (Fig. 2b, c), which are in close proximity but largely divided by significant barriers to dispersal, including large motorways and urbanised expanses. The genetic connectivity between these regions may be partly due to road mitigation measures implemented by the local governments to aid koala dispersal, such as raised wildlife crossings (Brisbane City Council 2014). It is also likely that haphazard translocations of rehabilitated animals by the public could have contributed to gene flow in this region. Further collaring and tracking studies are needed to confirm this.

A total of 95 sex-linked genetic markers were identified in this dataset. These markers could potentially be used in future studies to help investigate sex-biased dispersal and connectedness within and between closely linked populations. They may also allow for genetic tracking of paternity in the koala. Genetic analysis of sex-biased dispersal in the koala has until now largely been reliant on the use of mtDNA (Fowler et al.

2000; Taylor et al. 1997; Tsangaras et al. 2012; Wilmer et al. 1993), which can be inherently limiting due to its female biased inheritance pattern. A strong subset of sex-linked SNPs is an effective method to both confidently identify individuals and to test paternity (Heaton et al. 2002).

Using two methods to detect outlier loci, this study identified 11 putative outlier loci that were common to all QLD populations sampled, with 10 of these returning positive alpha levels. It is possible that there is a base level of background selection occurring in these populations, however given that the 10 common outliers identified had positive alpha levels and they were common in two independent biological replicate comparisons, it is likely that these loci are associated with regions of the genome under directional selection. While it is most likely that this study has not detected all regions of the genome under selection, it highlights that differential selective pressures are present amongst the populations tested. This is further supported by the distinctive population clustering observed in the NEVTIEW analysis (Fig. 2). A more comprehensive national analysis of koala populations is warranted to fully understand the differentiation and stratification of the national population in distinct sub-populations. A strong understanding of adaptive potential and the biology of a species are vital to a successful translocation and management plan. Koalas are already actively managed across their range in the face of growing human development, yet there is little genetic information available to guide this process. The identification of genetic sub-groups and management boundaries is critical for all genetically differentiated terrestrial species (Pollinger et al. 2010). While neutral markers such as microsatellites have been used extensively in wild populations to assess diversity within and between regions, more recent evidence suggests that a combination of adaptive and neutral markers such as those offered by a ddRAD dataset should be utilised when developing management plans in order for them to be most effective (Féral 2002; Fraser and Bernatchez 2001; Funk et al. 2012). Management plans should consider current and historical gene flow, and environmental differences which have shaped local koala populations. If populations are under different environmental pressures and have been separated for >20 generations (as is likely to be the case with the koala at the extremes of its range and even within state borders), care should be taken to assess adaptive variation as well as neutral diversity between managed populations (see Frankham et al. 2011). If a population declines to the point of requiring intervention or translocations, closely

related populations under similar environmental pressures, genetic makeup and selective forces should be considered, rather than geographic distance. The genetic similarity based on genome wide relationships from the SNP set developed here and as visualised through NETVIEW will strongly assist in the management of koala populations.

We recommend that future studies employ a more rigorous sampling design (higher sample numbers) and couple this with environmental and phenotypic/morphological data in order to pinpoint regions of the koala genome under selection. Additionally, the utility of this method to identify markers potentially under selection could be used to inform future translocation planning for this species (Funk et al. 2012).

Usefulness of ddRADseq and limitations of the dataset

Our findings support that ddRAD has proven to be a cost effective (~$20 per sample) and efficient method for rapid discovery of large genome-wide datasets in a number of species (Peterson et al. 2012). Recent studies have indicated that the power of a marker set increases with a greater number of markers utilised (Miller et al. 2014) and SNPs have been demonstrated to be far more efficient and practical tool for population genetic studies than traditional neutral markers. In order to obtain the best quality SNP datasets, some forethought and planning needs to be conducted prior to library preparation. The *in silico* analysis performed prior to library preparation indicated that a size range of 350–450 bp would recover between 20,000 and 35,000 regions in the koala genome. A range of 400 ± 44 bp was chosen as this was easily visible on an agarose gel image. The ddRAD protocol is dependent on high quality, genomic DNA and success relies on selecting enzyme combinations appropriate to genome content and structure. Thus, it is strongly recommended that an *in silico* digest is performed prior to library preparation and sequencing. The modified protocol used in this study maybe optimised to not only be applied to different species and taxa, but also gives flexibility in the number of regions of the genome recovered (Peterson et al. 2012). This allows for the fine-tuning of sequencing depth and the number of markers recovered. DNA preparation and standardisation between samples was also critical, as was evident when comparing sequencing runs 1 and 2, with the second run yielding significantly more reads per individual (84.9 ± 0.03 %). This was achieved

through increasing consistency between samples at each step in the protocol and by decreasing the number of individuals pooled per sequencing run (~300 samples per lane). While the overall datasets can be improved by re-sequencing of libraries, this significantly increases cost and resource requirements.

Conclusions

The use of ddRAD sequencing and genotyping in conjunction with high throughput next-generation sequencing allows for a simple and effective method of genome-wide marker discovery. With careful planning and experimental design, this method has allowed for robust estimations of diversity and divergence in the koala; a species currently without an available reference genome (Andrews et al. 2014). This genomic resource is the first of its kind in the koala and will provide a basis for other genome-wide population studies in the future. The markers identified in this study have indicated that species-wide diversity in the koala is equivalent to, if not higher than other wild, outbred vertebrate species, but it is unclear how the koala compares to other marsupials, due to the lack of genome-wide research in the area. This finding is contradictory to a general view of the species as having low genetic diversity (Houlden et al. 1999; Tsangaras et al. 2012). This study has further resolved genetic groupings in the koala, showing three broad genomic clusters across Australia and a high level of variation between populations, indicating an isolation by distance model and rejection of a distinct sub-species classification, although a more comprehensive sampling strategy to pinpoint possible transition zones is needed to confirm this. The application of this method to a wider range of samples and populations across the species range will provide in-depth information that can inform conservation priorities, management and possible translocation plans for koalas across Australia.

Acknowledgments

We would like to express our gratitude to Eike Joachim Steinig for his help and support with regard to NETVIEW and various other analyses, Damien Higgins for supplying our Port Macquarie samples, Diedre de Villiers for supplying Koala Coast and Ipswich samples and David Phalen for providing samples from NSW and VIC. Monal Lal for his comments and advice on draft manuscripts and Maria Nayfa for her helpful advice

regarding outlier analyses. Funding was provided via an Australian Research Council (ARC) Linkage Project grant (LP120200630), with industry funding and in-kind support from partners San Diego Zoo Koala Education & Conservation Program and James Cook University, Sydney University, and Science for Wildlife.

Electronic supplementary material

Below is the link to the electronic supplementary material.

Supplementary Material 1

Genotype file in the STRUCTURE format containing 3060 SNP markers across all individuals. A value of '−9' indicates missing data. Supplementary material 1 (XLSX 3331 kb)

References

Adamkewicz SL, Harasewych MG (1996) Systematics and biogeography of the genus Donax (Bivalvia: Donacidae) in eastern North America. Am Malacol Bull 13:97–103

Adams-Hosking C, Moss P, Rhodes J, Grantham H, McAlpine C (2011) Modelling the potential range of the koala at the Last Glacial Maximum: future conservation implications. Aust Zool 35:983–990

Alho JS, Välimäki K, Merilä J (2010) Rhh: an R extension for estimating multilocus heterozygosity and heterozygosity–heterozygosity correlation. Mol Ecol Resour 10:720–722

Altham PME (1969) Exact Bayesian analysis of a 2 × 2 contingency table, and fisher's "Exact" significance test. J R Stat Soc 31:261–269

Andrews S (2010) FastQC A quality control tool for high throughput

sequence data.
http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Andrews KR, Hohenlohe PA, Miller MR et al (2014) Trade-offs and utility of alternative RADseq methods: reply to Puritz et al. Mol Ecol 23:5943–5946

Angeloni F, Wagemaker N, Vergeer P, Ouborg J (2012) Genomic toolboxes for conservation biologists. Evol Appl 5:130–143

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 Bioinf 9:323

Avila-Arcos MC, Ho SYW, Ishida Y et al (2013) One 120 years of Koala retrovirus evolution determined from museum. Mol Biol Evol 30(2):299–304

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Lab Génome, Popul Interact CNRS UMR 5000:1996–2004

Berck L (1995) St. Bees Island : its history, life-styles & tales/by Lionel Berck. Boolarong Press, Brisbane

Biotium (2013) AccuBlue™ High Sensitivity dsDNA Quantitation Kits.

Brisbane City Council (2014) Protecting biodiversity. QLD Government, Brisbane. http://www.brisbane.qld.gov.au/environment-waste/natural-environment/bushland-parklands-wetlands/protecting-biodiversity

Cánovas A, Bannasch DL, Oberbauer AM, Medrano JF (2014) Single nucleotide polymorphism (SNP) variation of wolves (*Canis lupus*) in southeast Alaska and comparison with wolves, dogs, and coyotes in North America. J Hered 106(1):26–36

Carmichael SN, Bekaert M, Taggart JB et al (2013) Identification of a sex-linked SNP marker in the salmon louse (*Lepeophtheirus salmonis*)

using RAD sequencing. PLoS One 8:8

Catchen J, Amores A, Hohenlohe P, Cresko W, Postlethwait J (2011) Stacks: building and genotyping loci de novo from short-read sequences. G3 1(3):171–182

Comeron JM (2014) Background selection as baseline for nucleotide variation across the *Drosophila* Genome. PLoS Genet 10:e1004434

Consortium BH (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324:528–532

Cronin MA, Rincon G, Meredith RW et al (2014) Molecular phylogeny and SNP variation of polar bears (*Ursus maritimus*), brown bears (*U. arctos*), and black bears (*U. americanus*) derived from genome sequences. J Hered 105(3):312–323

Cronin MA, Cánovas A, Bannasch DL, Oberbauer AM, Medrano JF (2015) Single Nucleotide Polymorphism (SNP) Variation of Wolves (Canis lupus) in Southeast Alaska and Comparison with Wolves, Dogs, and Coyotes in North America. J Hered 106: 26-36.

Davey JW, Hohenlohe PA, Etter PD et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510

DeFaveri J, Viitaniemi H, Leder E, Merila J (2013) Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. Mol Ecol Resour 13:377–392

Desjardins P, Conklin D (2010) NanoDrop microvolume quantitation of nucleic acids. *J Vis Exp*

Edmands S (2007) Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and

management. Mol Ecol 16:463–475

Ellis W, Hale PT, Carrick FN (2002) Breeding dynamics of koalas in open woodlands. Wildl Res 29:19–25

Ellis WA, Melzer A, Bercovitch FB (2009) Spatiotemporal dynamics of habitat use by koalas: the checkerboard model. Behav Ecol Sociobiol 63:1181–1188

Fabbri E, Miquel C, Lucchini V et al (2007) From the Apennines to the Alps: colonization genetics of the naturally expanding Italian wolf (*Canis lupus*) population. Mol Ecol 16:1661–1671

Feltus FA, Wan J, Schulze SR et al (2004) An SNP resource for rice genetics and breeding based on sub-species indica and japonica genome alignments. Genome Res 14:1812–1819

Féral J-P (2002) How useful are the genetic markers in attempts to understand and manage marine biodiversity? J Exp Mar Biol Ecol 268:121–145

Fernández ME, Goszczynski DE, Lirón JP et al (2013) Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. Genet Mol Biol 36:185–191

Fish and Wildlife Service (1998) Endangered and threatened wildlife and plants; proposed determination of threatened status for the Koala (ed. INTERIOR DOT)

Foll M (2012) BayeScan v2. 1 user manual. Ecology 20:1450–1462

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. Genetics 180:977–993

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

Frankel OH (1981) Conservation and evolution. CUP Archive, Cambridge

Frankham R, Briscoe DA, Ballou JD (2002) Introduction to conservation genetics. Cambridge University Press, Cambridge

Frankham R, Ballou JD, Eldridge MDB et al (2011) Predicting the probability of outbreeding depression predicción de la depresión por exogámica. Conserv Biol 25:465–475

Frantz A, Pourtois JT, Heuertz M et al (2006) Genetic structure and assignment tests demonstrate illegal translocation of red deer (*Cervus elaphus*) into a continuous population. Mol Ecol 15:3191–3203

Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Mol Ecol 10:2741–2752

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends Ecol Evol 27:489–496

GE Healthcare Life Sciences (2000) Sephadex G-50 fine. GE Healthcare Life Sciences

Goedbloed D, Megens H, Van Hooft P et al (2013) Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. Mol Ecol 22:856–866

Goldingay RL, Dobner B (2014) Home range areas of koalas in an urban area of north-east New South Wales. Aust Mammal 36:74–80

Gordon G, Hrdina F, Patterson R (2006) Decline in the distribution of the Koala *Phascolarctos cinereus* in Queensland. Aust Zool 33:345–358

Gravuer K, von Wettberg E, Schmitt J (2005) Population differentiation and genetic variation inform translocation decisions for *Liatris scariosa var. novae-angliae*, a rare New England grassland perennial. Biol Conserv 124:155–167

Gray MM, Granka JM, Bustamante CD et al (2009) Linkage disequilibrium and demographic history of wild and domestic canids. Genetics 181:1493–1505

Haynes GD, Latch EK (2012) Identification of novel single nucleotide polymorphisms (SNPs) in deer (*Odocoileus spp*.) using the bovineSNP50 BeadChip. PLoS One 7:e36536

Heaton MP, Harhay GP, Bennett GL et al (2002) Selection and use of SNP markers for animal identification and paternity analysis in US beef cattle. Mamm Genome 13:272–281

Hoffman JI, Simpson F, David P, et al. (2014) High-throughput sequencing reveals inbreeding depression in a natural population. Proceedings of the National Academy of Sciences of the United States of America 111: 3775-3780.

Houlden BA, England P, Sherwin WB (1996a) Paternity exclusion in koalas using hypervariable microsatellites. J Hered 87:149–152

Houlden BA, England PR, Taylor AC, Greville WD, Sherwin WB (1996b) Low genetic variability of the koala *Phascolarctos cinereus* in south eastern Australia following a severe population bottleneck. Mol Ecol 5:269–281

Houlden BA, Greville WD, Sherwin WB (1996c) Evolution of MHC class I loci in marsupials: characterization of sequences from koala (*Phascolarctos cinereus*). Mol Biol Evol 13:1119–1127

Houlden BA, Costello BH, Sharkey D et al (1999) Phylogeographic differentiation in the mitochondrial control region in the koala, *Phascolarctos cinereus* (Goldfuss 1817). Mol Ecol 8:999–1011

Huang H, Knowles LL (2014) Unforeseen consequences of excluding

missing data from next-generation sequences: simulation study of RAD sequences. Syst Biol 57:719–731

Iredale T, Troughton ELG (1934) A check-list of the mammals recorded from Australia. Mem Aust Mus 6:1–122

Johnson M, Zaretskaya I, Raytselis Y et al (2008) NCBI BLAST: a better web interface. Nucleic Acids Res 36:W5–W9

Johnston SE, Orell P, Pritchard VL et al (2014) Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). Mol Ecol 23:3452–3468

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

Jones Felicity C, Chan Yingguang F, Schmutz J et al (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. Curr Biol 22:83–90

Kalinowski ST, Wagner AP, Taper ML (2006) ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. Mol Ecol Notes 6:576–579

Kijas JW, Townley D, Dalrymple BP et al (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One 4:e4668

Koblmüller S, Nord M, Wayne RK, Leonard JA (2009) Origin and status of the Great Lakes wolf. Mol Ecol 18:2313–2326

Kovach RP, Gharrett AJ, Tallmon DA (2012) Genetic change for earlier migration timing in a pink salmon population. Proc R Soc Lond B 279:3870–3878

Kurushima J, Lipinski M, Gandolfi B et al (2013) Variation of cats under domestication: genetic assignment of domestic cats to breeds and worldwide random-bred populations. Anim Genet 44:311–324

Kyle C, Johnson A, Patterson B et al (2006) Genetic nature of eastern wolves: past, present and future. Conserv Genet 7:273–287

Larmuseau M, Raeymaekers J, Hellemans B, Van Houdt J, Volckaert F (2010) Mito-nuclear discordance in the degree of population differentiation in a marine goby. Heredity 105:532–542

Larson WA, Seeb LW, Everett MV et al (2014) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). Evol Appl 7:355–369

Lau Q, Jobbins SE, Belov K, Higgins DP (2013) Characterisation of four major histocompatibility complex class II genes of the koala (*Phascolarctos cinereus*). Immunogenetics 65:37–46

Lau Q, Jaratlerdsiri W, Griffith J, Gongora J, Higgins D (2014) MHC class II diversity of koala (*Phascolarctos cinereus*) populations across their range. Heredity 113: 287-296

Leache AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. Syst Biol 63:534–542

Lee KE, Seddon JM, Corley SW et al (2010) Genetic variation and structuring in the threatened koala populations of Southeast Queensland. Conserv Genet 11:2091–2103

Lee KE, Seddon JM, Johnston S et al (2012a) Genetic diversity in natural and introduced island populations of koalas in Queensland. Aust J Zool 60:303–310

Lee T, Zenger KR, Close RL, Phalen DN (2012b) Genetic analysis reveals a distinct and highly diverse koala (*Phascolarctos cinereus*) population in South Gippsland, Victoria, Australia. Aust Mammal 34:68–74

Lepais O, Weir JT (2014) SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol Ecol Resour 14:1314–1321

Luikart G, Ryman N, Tallmon DA, Schwartz MK, Allendorf FW (2010) Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. Conserv Genet 11(2):355–373

Lunney D, O'Neill L, Matthews A, Sherwin WB (2002) Modelling mammalian extinction and forecasting recovery: koalas at Iluka (NSW, Australia). Biol Conserv 106:101–113

Lunney D, Stalenberg E, Santika T, Rhodes JR (2014) Extinction in Eden: identifying the role of climate change in the decline of the koala in south eastern NSW. Wildl Res 41:22–34

MacEachern S, Hayes B, McEwan J, Goddard M (2009) An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. BMC Genom 10:181

Martin R (1985) Overbrowsing, and decline of a population of the koala, *Phascolarctos cinereus*, in Victoria. I. Food preference and food tree defoliation. Wildl Res 12:355–365

Melzer A, Carrick F, Menkhorst P, Lunney D, John BS (2000) Overview, critical assessment, and conservation implications of koala distribution and abundance. Conserv Biol 14:619–628

Menkhorst P (2008) Hunted, marooned, re-introduced, contracepted: a history of Koala management in Victoria. Royal Zoological Society New South Wales, Mosman

Miller W, Schuster SC, Welch AJ et al (2012) Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proc Natl Acad Sci 109:E2382–E2390

Miller JM, Malenfant RM, David P et al (2014) Estimating genome-wide heterozygosity: effects of demographic history and marker type. Heredity 112:240–247

Mitchell P (1990a) The home ranges and social activity of koalas-a quantitative analysis. Biol Koala 35(3):171–187

Mitchell P (1990b) Social behaviour and communication of koalas. Biol Koala Chipp Nort (Aust) 2:151–170

Moritz C, Worthington Wilmer J, Pope L et al (1996) Applications of genetics to the conservation and management of Australian fauna: four case studies from Queensland. Mol Genet Approaches Conserv 26:442–456

Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. Mol Ecol Resour 11:184–194

Narum S, Banks M, Beacham T et al (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. Mol Ecol 17:3464–3477

Natural Resource Management Ministerial Council (2009) National Koala Conservation and Management Strategy 2009–2014. (Department of the Environment W, Heritage and the Arts, Canberra.). Natural Resource Management Ministerial Council, Canberra

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19: 153-170

Nielsen R, Williamson S, Kim Y et al (2005) Genomic scans for

selective sweeps using SNP data. Genome Res 15:1566–1575

Novicic ZK, Jelic M, Jovanovic M et al (2011) Microsatellite variability of *Drosophila subobscura* populations from the central Balkans. Evol Ecol Res 13:479–494

Olsen MT, Volny VH, Berube M et al (2011) A simple route to single-nucleotide polymorphisms in a nonmodel species: identification and characterization of SNPs in the Artic ringed seal (Pusa hispida hispida). Mol Ecol Resour 11:9–19

Peel D, Ovenden J, Peel S (2004) NeEstimator: software for estimating effective population size, version 1.3. Queensland Government, Department of Primary Industries and Fisheries

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7:11

Phillips SS (2000) Population trends and the koala conservation debate. Conserv Biol 14:650–659

Pilot M, Greco C, vonHoldt BM et al (2014) Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. Heredity 112:428–442

Piry S, Alapetite A, Cornuet J-M, et al. (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. J Hered 95: 536-539

Pollinger JP, Lohmueller KE, Han E et al (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. Nature 464:898–902

Pollinger JP, Earl DA, Knowles JC et al (2011) A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. Genome Res 21:1294–1305

Pujolar JM, Jacobsen MW, Als TD et al (2014) Genome-wide single-generation signatures of local selection in the panmictic European eel. Mol Ecol 23:2514–2528

Rasic G, Filipovic I, Weeks AR, Hoffmann AA (2014) Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, Aedes aegypti. BMC Genomics 15:12

Ross KG, Fletcher DJ (1985) Comparative study of genetic and social structure in two forms of the fire ant *Solenopsis invicta* (Hymenoptera: Formicidae). Behav Ecol Sociobiol 17:349–356

Ruiz-Rodriguez CT, Ishida Y, Greenwood AD, Roca AL (2014) Development of 14 microsatellite markers in the Queensland koala (*Phascolarctos cinereus* adustus) using next generation sequencing technology. Conserv Genet Resour 6:429–431

Schneider S, Roessli D, Excoffier L (2000) Arlequin: a software for population genetics data analysis. User Man Ver 2:2496–2497

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. Trends Ecol Evol 22:25–33

Seabrook L, McAlpine C, Phinn S, Callaghan J, Mitchell D (2002) Landscape legacies: koala habitat change in Noosa Shire, south-east Queensland. Aust Zool 32:446–461

Seabrook L, McAlpine C, Baxter G et al (2011) Drought-driven change in wildlife distribution and numbers: a case study of koalas in south west Queensland. Wildl Res 38:509–524

Senn H, Ogden R, Frosch C et al (2014) Nuclear and mitochondrial genetic structure in the Eurasian beaver (*Castor fiber*)–implications for future reintroductions. Evol Appl 7:645–662

Shaffer ML (1981) Minimum population sizes for species conservation. Bioscience 31:131–134

Sherwin WB, Timms P, Wilcken J, Houlden B (2000) Analysis and conservation implications of koala genetics. Conserv Biol 14:639–649

Sinnock P (1975) The Wahlund effect for the two-locus model. Am Nat 109:565–570

Smith P, Smith J (1990) Decline of the urban koala (*Phascolarctos cinereus*) population in Warringah Shire, Sydney. Aust Zool 26:109–129

Steinig EJ, Neuditschko M, Khatkar MS, Raadsma HW, Zenger KR (2015) netview p: a network visualization tool to unravel complex population structure using genome-wide SNPs. Mol Ecol Res

Stephens MR, Clipperton NW, May B (2009) Sub-species-informative SNP assays for evaluating introgression between native golden trout and introduced rainbow trout. Mol Ecol Resour 9:339–343

Strahan R (1995) The Mammals of Australia, 2nd edn. Reed Books, Chatswood, Australia.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Taylor AC, Graves JAM, Murray ND, Sherwin WB (1991) Conservation genetics of the koala (*Phascolarctos cinereus*) 2. Limited variability in minisatellite dna-sequences. Biochem Genet 29:355–363

Taylor AC, Graves JM, Murray ND et al (1997) Conservation genetics of the koala (*Phascolarctos cinereus*), low mitochondrial DNA variation amongst southern Australian populations. Genet Res 69:25–33

Thermo Scientific (2014) Sera-mag speedbeads carboxylate-modified magnetic particles

Thomas O (1923) On some queensland phalangeridae. Ann Mag Nat Hist 11:246–250

Thompson J (2006) The comparative ecology and population dynamics of koalas in the Koala Coast region of south-east Queensland. University of Queensland, St Lucia

Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. Trends Ecol Evol 29:673–680

Troughton ELG (1935) The southern race of the koala. Aust Nat 9:137–140

Tsangaras K, Avila-Arcos MC, Ishida Y et al (2012) Historically low mitochondrial DNA diversity in koalas (*Phascolarctos cinereus*). BMC Genet 13:11

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. Genet Select Evol 34:275–305

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

Whisson DA, Holland GJ, Carlyon K (2012) Translocation of overabundant species: implications for translocated individuals. J Wildl Manag 76:1661–1669

White NA (1999) Ecology of the koala (*Phascolarctos cinereus*) in rural south-east Queensland, Australia. Wildl Res 26:731–744

Willing E-M, 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

Wilmer JMW, Melzer A, Carrick F, Moritz C (1993) Low genetic diversity and inbreeding depression in Queensland koalas. Wildl Res 20:177–188

Zenger K, Khatkar M, Tier B et al (2007) QC analyses of SNP array

data: experiences from a large population of dairy sires with 23.8 million data points. Proc Seve Conf Genetic Improv Mak Happen 17:123–126