Immunogenetics
DOI 10.1007/s00251-017-1018-2

ORIGINAL ARTICLE

Characterisation of MHC class I genes in the koala

Yuanyuan Cheng[1] · **Adam Polkinghorne**[2] · **Amber Gillett**[3] · **Elizabeth A. Jones**[1] · **Denis O'Meally**[1,2] · **Peter Timms**[2] · **Katherine Belov**[1]

Received: 9 February 2017 / Accepted: 20 June 2017

Abstract Koala (*Phascolarctos cinereus*) populations are on the decline across the majority of Australia's mainland. Two major diseases threatening the long-term survival of affected koala populations are caused by obligate intracellular pathogens: *Chlamydia* and koala retrovirus (KoRV). To improve our understanding of the koala immune system, we characterised their major histocompatibility complex (MHC) class I genes, which are centrally involved in presenting foreign peptides derived from intracellular pathogens to cytotoxic T cells. A total of 11 class I genes were identified in the koala genome. Three genes, *Phci-UA*, *UB* and *UC*, showed relatively high genetic variability and were expressed in all 12 examined tissues, whereas the other eight genes had tissue-specific expression and limited polymorphism. Evidence of diversifying selection was detected in *Phci-UA* and *UC*, while gene conversion may have played a role in creating new alleles at *Phci-UB*. We propose that *Phci-UA*, *UB* and *UC* are likely classical MHC genes of koalas, and further research is needed to understand their role in koala chlamydial and KoRV infections.

Keywords Marsupial · Koala · MHC · Class I

Electronic supplementary material The online version of this article (doi:10.1007/s00251-017-1018-2) contains supplementary material, which is available to authorized users.

✉ Katherine Belov
kathy.belov@sydney.edu.au

[1] School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia

[2] Centre for Animal Health Innovation, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Sunshine Coast, QLD, Australia

[3] Australia Zoo Wildlife Hospital, Beerwah, QLD 4519, Australia

Background

The koala (*Phascolarctos cinereus*) is the last living member of the marsupial family Phascolarctidae and is one of Australia's most charismatic species. Unfortunately, however, koala numbers are on the decline across most mainland populations (Polkinghorne et al. 2013). Koalas in the states of Queensland, New South Wales, and Australian Capital Territory have been listed as vulnerable by the Australian Federal Government since 2012 (Kollipara et al. 2013), with recognised threats including habitat loss and fragmentation, road kills, bushfire, dog attacks, and a range of bacterial, viral and parasitic infections (Melzer et al. 2000; Dique et al. 2003; Lunney et al. 2007; McInnes et al. 2011; Rhodes et al. 2011). The two major diseases that threaten the long-term survival of koala populations are caused by obligate intracellular pathogens: *Chlamydia* (mainly *C. pecorum* and *C. pneumoniae*) (Polkinghorne et al. 2013) and koala retrovirus (KoRV) (Simmons et al. 2012). While chlamydial infection can cause genital tract and ocular disease leading to infertility and blindness, respectively (Polkinghorne et al. 2013), KoRV infection has been proposed to result in immunosuppression and potentially increased susceptibility to other diseases (e.g. chlamydia, lymphoid neoplasia) (Tarlinton et al. 2005; Maher and Higgins 2016).

The major histocompatibility complex (MHC) class I and class II genes play a crucial role in disease resistance, encoding polymorphic antigen-presenting molecules on the host cell surface. While class II molecules are primarily responsible for the recognition of extracellular and intravesicular pathogens, class I molecules present foreign peptides derived from intracellular pathogens, such as viruses or cytoplasmic bacteria, to cytotoxic T cells (Roy 2003; Cresswell et al. 2005). In koalas, MHC II genes have been previously characterised and extensively studied across

Published online: 01 July 2017

populations (Lau et al. 2013, 2014a, b; Abts et al. 2015); however, class I genes remained largely unexamined, with only two studies published in the past 20 years reporting 13 transcripts isolated from three individuals (Houlden et al. 1996; Abts et al. 2015). To bridge this gap of knowledge, we annotated class I genes in the PacBio koala genome assembly (Johnson et al. 2017) and classified the genes into classical or nonclassical by examining their tissue expression pattern, genetic variability and evolutionary features. While nonclassical class I (class Ib) genes can play diverse roles in immunological or nonimmunological processes (Braud et al. 1999), classical class I (class Ia) molecules are centrally involved in immune responses against intracellular pathogens and therefore are more likely to play a part in koala chlamydia and KoRV resistance. This study provides the foundation for future immunological research and disease association studies in koalas.

Methods

Gene identification

The koala reference genome was sequenced on a Pacific Biosciences RS II platform and estimated to have a 57.3-fold coverage (Johnson et al. 2017). An additional 30× of 150 bp paired-end Illumina X Ten reads was used for polishing the PacBio assembly, resulting in a high-quality genome with 95.1% of 4104 mammalian Benchmarking Universal Single-Copy Orthologs (BUSCOs) recovered, which is the greatest among all published marsupial genomes and comparable with the human genome (Johnson et al. 2017). BLAST searches were performed in the koala reference genome and a set of transcriptomes (Hobbs et al. 2014) using known marsupial MHC class I genes, including grey short-tailed opossum (*Monodelphis domestica*) *Modo-UA1* [GenBank accession number: KC560794], *Modo-UB* [NM_001079820], *Modo-UC* [NM_001079819], *Modo-UE* [EU886708], *Modo-UG* [DQ138606], *Modo-UI* [EU886707], *Modo-UJ* [NM_001171836], *Modo-UK* [EU886706], *Modo-UM* [NM_001171834], *Modo-MR1* [AB719956], *Modo-UA2*, *-UF*, *-UH* and *-UL* available from http://bioinf.wehi.edu.au/opossum/seq/Class_I.fa; tammar wallaby (*Macropus eugenii*) *Maeu-UA*, *-UB*, *-UC*, *-UD*, *-UE*, *-UF*, *-UH*, *-UI*, *-UJ*, *-UK*, *-UL*, *-UM*, *-UN*, *-UO* and *-UP* obtained from Siddle et al. (2009); Tasmanian devil (*Sarcophilus harrisii*) *Saha-UA*, *-UB*, *-UC* and *-UD* from Cheng et al. (2012) and *Saha-UK* and *-UM* from Cheng and Belov (2014). BLAST hit sequences, along with 13 previously reported koala class I sequences [U33807–U33815, KP792542, KP792544, KP792554 and KP792563], were used to perform further searches in the genome to ensure that all potential lineage-specific duplicate genes were found. Coding sequences of identified genes were predicted based on human and formerly characterised marsupial genes. Phylogenetic analysis was conducted in MEGA6 (Tamura et al. 2013) using the neighbour-joining method (Saitou and Nei 1987) with 1000 bootstrap replicates to infer the level of confidence on the phylogeny (Felsenstein 1985). Genes were named after their counterparts in other marsupials if clear homologous relationships were shown; otherwise, species-specific names were assigned. Alleles were named following the nomenclature system proposed by Ellis et al. (2006).

Classical and nonclassical gene categorisation

Tissue expression Gene expression was examined using non-quantitative RT-PCR in 12 types of tissues, including the bone marrow, lymph node, spleen, thymus, liver, stomach, duodenum, jejunum, colon, lung, kidney and testis. Opportunistic samples were collected from two male koalas that were euthanised at the Australia Zoo Wildlife Hospital due to advanced disease or injury. Tissue samples were stabilised in RNAlater (Sigma-Aldrich) during necropsy and total RNA was extracted using an RNeasy Mini Kit (QIAGEN) with a DNase digestion step to eliminate DNA contamination. The integrity and concentration of RNA was assessed on a Bioanalyzer (Agilent Technologies). All samples had an RNA Integrity Number (RIN) higher than 7.5 except for one testis sample that had RIN = 6.0. Approximately 200 ng of RNA was used for complementary DNA (cDNA) synthesis with SuperScript VILO Master Mix (Invitrogen), and a 1:10 dilution of the cDNA was made for use in PCR. Locus-specific primers were designed within exon 2 (forward) and exon 3 (reverse) of the genes (Online Resource 1), with primer specificity confirmed by sequencing PCR amplicons purified from gels. PCRs were performed in a total volume of 25 μL containing 1 U of *Taq* DNA Polymerase (Invitrogen), 2.0 mM Mg^{2+}, 0.2 mM each dNTP, 0.6 μM each primer and 1 μL cDNA. PCR conditions were as follows: 100 °C hot lid, 94 °C initial denaturation for 3 min, 33 cycles of 94 °C denaturation for 30 s, 62 °C annealing for 30 s, 72 °C extension for 45 s, and 72 °C final extension for 10 min.

Genetic variability Genetic variability within the peptide-binding α1 domain (exon 2) of identified class I genes was assessed via allele cloning and sequencing. Fifteen individuals belonging to two populations were used (Online Resource 2): five from a Queensland koala population at Lone Pine Koala Sanctuary (blood samples collected with approval by the Queensland University of Technology Animal Ethics Committee, Approval #0700000559), and 10 from a New South Wales population at Port Macquarie (DNA provided by the Australian Centre for Wildlife Genomics at the Australian Museum). Locus-specific primers were designed in intron 1 (towards 3′ end) and intron 2 (5′ end), with the

amplicon size ranging from 313 to 443 bp (Online Resource 1). PCRs contained 1× High Fidelity Buffer (Invitrogen) consisting of 60 mM Tris-HCl (pH 8.9) and 18 mM $(NH_4)_2SO_4$, 2.0 mM $MgSO_4$, 0.2 mM each dNTP, 0.5 μM each primer, 1.5 U Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen), and approximately 2 ng/μL genomic DNA. PCRs were carried out as follows: 100 °C hot lid; 94 °C for 2 min; 34 cycles of 94 °C 30 s, 62 °C 30 s, and 68 °C 45 s; and 68 °C 10 min. PCR amplicons were purified from agarose gels using QIAquick Gel Extraction Kit (Qiagen), ligated into pGEM-T Easy Vector (Promega), and cloned using JM109 High Efficiency Competent Cells (Promega). Six positive clones per reaction were picked and plasmids were extracted using DirectPrep 96 MiniPrep Kit (Qiagen) on a QIAvac Multiwell vacuum manifold (Qiagen). Plasmids were sequenced with T7 primer at the Australian Genome Research Facility (AGRF).

Sequence analysis Sequences were quality-checked in Sequencher 4.1.4 (Gene Codes) and aligned using CLUSTAL W (Thompson et al. 1994). To rule out PCR and sequencing errors, sequence variants that were found in multiple animals or amplified from a single animal in two separate PCRs were considered to indicate real alleles. Allele frequencies were calculated and tested for Hardy-Weinberg equilibrium using Cervus 3.0 (Kalinowski et al. 2007). Functionally significant residues, including those involved in peptide binding and/or T cell receptor or CD8 interaction, were predicted by aligning koala genes to human *HLA-A*, *B*, and *C* (Bjorkman and Parham 1990). Mean rates of synonymous and nonsynonymous nucleotide substitutions within and outside the peptide-binding region (PBR) were computed and tested for positive or negative selection in MEGA6 using the modified Nei-Gojobori method with Jukes-Cantor correction to account for multiple substitutions at a single site (Nei and Kumar 2000). Recombination break points of putative gene conversion events were inferred using the Genetic Algorithm Recombination Detection (GARD) method (Kosakovsky Pond et al. 2006) via the Datamonkey webserver (Delport et al. 2010). Individual sites under episodic diversifying selection were detected using the mixed effects model evolution (MEME) method (Murrell et al. 2012), and negatively selected sites were inferred using the fixed effects likelihood (FEL) method (Kosakovsky Pond and Frost 2005).

Results and discussion

Identification of koala MHC class I genes

Eleven transcribed class I genes were identified on six scaffolds in the koala genome assembly, namely *Phci-UA*, *UB*, *UC*, *UD*, *UE*, *UF*, *UG*, *UH*, *UI*, *UJ* and *UK* (Table 1). *Phci-UG* and *UH*, *Phci-UD* and *UI*, and *Phci-UC*, *UE*, *UF* and *UK* are located on three scaffolds, respectively, while *Phci-UA*, *UB* and *UJ* are found on separate scaffolds. The scaffold that contains *Phci-UC*, *UE*, *UF* and *UK*, i.e. scaffold#255 (3,701,240 bp), represents the core MHC region harbouring 138 MHC genes (Johnson et al. 2017).

Sequence similarities between genes ranged from 63.8 to 95.5% in coding sequences and from 52.8 to 94.1% in encoded proteins (Table 2). The highest similarity was found between *Phci-UC* and *UE* (95.5%), which formed a clade with 83% bootstrap support (Fig. 1). These two genes are located in close proximity in the genome (~72 kb apart on the same scaffold) and have likely resulted from a recent duplication. *Phci-UA* and *UB* also showed well-supported phylogenetic relationships (90%), though they are found on different scaffolds and their relative position cannot be determined based on available data.

Table 1 MHC class I genes in the koala genome assembly

Gene name	Class Ia or Ib	Scaffold	Scaffold size (bp)	Coordinates		Protein size (AA)
				Start	End	
Phci-UA	Ia	scaf00263	3,507,541	3,143,043	3,139,830	361
Phci-UB	Ia	scaf00381	1,011,865	15,571	11,695	357
Phci-UC	Ia	scaf00255	3,701,240	764,363	767,044	356
Phci-UD	Ib	scaf00347	1,650,515	1,588,259	1,590,606	343
Phci-UE	Ib	scaf00255	3,701,240	839,056	841,729	356
Phci-UF	Ib	scaf00255	3,701,240	986,711	989,369	357
Phci-UG	Ib	scaf00129	9,086,063	3,819,564	3,816,289	369
Phci-UH	Ib	scaf00129	9,086,063	3,858,138	3,854,952	361
Phci-UI	Ib	scaf00347	1,650,515	1,561,565	1,564,690	357
Phci-UJ	Ib	scaf00454	237,317	97,231	94,476	359
Phci-UK	Ib	scaf00255	3,701,240	403,212	407,221	361

Table 2 Pair-wise sequence similarity of the coding sequence of koala MHC class I genes

aa\nt[a]	UA	UB	UC	UD	UE	UF	UG	UH	UI	UJ	UK
Phci-UA	–	90.2%	70.1%	76.1%	69.6%	70.8%	86.3%	90.1%	85.3%	87.3%	67.9%
Phci-UB	84.7%	–	67.4%	76.5%	67.4%	68.5%	84.4%	86.3%	84.9%	85.6%	68.2%
Phci-UC	59.3%	58.5%	–	64.1%	95.5%	89.7%	66.7%	68.8%	70.6%	70.2%	64.2%
Phci-UD	66.5%	69.3%	57.4%	–	64.0%	64.6%	73.6%	75.7%	79.6%	76.5%	66.3%
Phci-UE	59.3%	58.8%	94.1%	58.2%	–	88.7%	66.6%	68.2%	70.2%	70.0%	63.8%
Phci-UF	60.2%	59.9%	85.5%	57.7%	86.0%	–	67.8%	69.1%	71.1%	72.4%	64.4%
Phci-UG	77.2%	76.6%	56.2%	64.0%	56.7%	57.8%	–	87.6%	83.1%	84.7%	69.4%
Phci-UH	81.7%	79.2%	59.6%	67.1%	59.9%	60.2%	79.4%	–	84.6%	86.8%	68.1%
Phci-UI	75.9%	76.3%	60.4%	72.7%	60.4%	61.9%	74.0%	74.8%	–	86.6%	70.2%
Phci-UJ	78.4%	78.4%	60.7%	68.2%	60.4%	62.1%	78.1%	79.2%	76.9%	–	71.0%
Phci-UK	55.6%	57.2%	52.8%	56.9%	53.6%	54.4%	57.5%	55.9%	58.1%	61.3%	–

[a] Above diagonal: nucleotide sequence identity; below diagonal: amino acid sequence identity

Most of the koala class I genes (all except *Phci-UK*) showed no clear orthologous relationship with previously reported marsupial genes (Fig. 1), which is common for MHC class I genes due to their rapid evolution involving lineage-specific gene duplication, recombination and deletion events (Nei et al. 1997).

One distinct feature of the koala MHC as compared to other marsupials is the lack of gene *UM*. *UM* and *UK* are two marsupial-specific class Ib genes that have been found in all three formerly sequenced marsupial genomes i.e. the opossum, tammar wallaby and Tasmanian devil (Fig.1). These species belong to distinct taxonomic lineages which last shared an ancestor

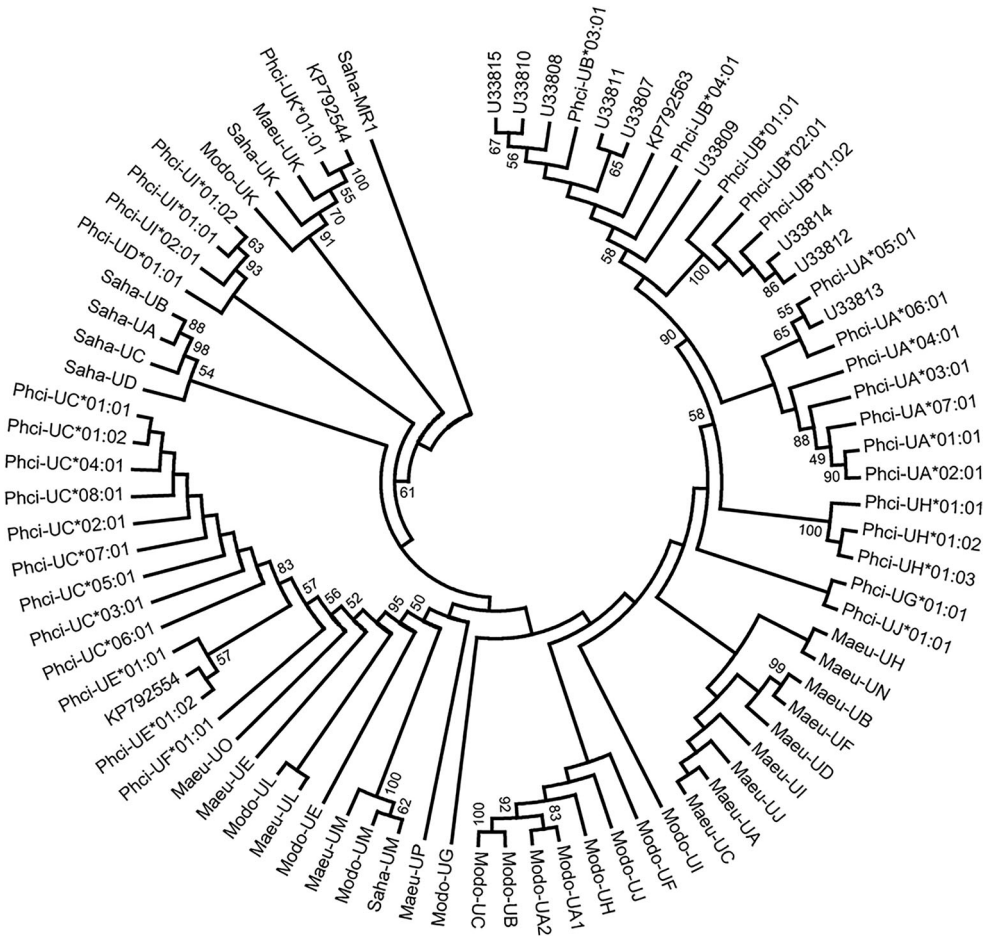

Fig. 1 Phylogenetic analysis of marsupial MHC class I genes using coding sequences. Prefix in gene names indicates species: *Phci* koala, *Maeu* tammar wallaby, *Saha* Tasmanian devil, *Modo* grey short-tailed opossum. Sequences U33807–U33815, KP792542, KP792544, KP792554 and KP792563 are previously reported koala class I transcripts. Low statistical support values (<50%) are not shown. Koala sequences identified in this study can be found in GenBank with accession numbers KY612373–KY612406

over 68 million years ago (Kirsch et al. 1997). As it is unusual to observe such clear orthology between class Ib genes of distantly related species (Hughes and Nei 1989), it has been speculated that *UM* and *UK* may play certain important marsupial-specific functions, resulting in the long-term conservation of these genes through marsupial evolution by strong purifying selection (Siddle et al. 2009). Formerly reported marsupial *UM* and *UK* genes in the opossum (Baker et al. 2009), tammar wallaby (Siddle et al. 2009) and Tasmanian devil (Cheng and Belov 2014) were used to search the koala reference genome and a set of transcriptomes (Hobbs et al. 2014). While *UK* was identified in the core MHC region (scaffold#255), similar to what was seen in other marsupials, no koala sequence was found to show orthology to other marsupial *UM* genes. This suggests that the loss of *UM* gene in the koala was likely due to a sudden mutational event, such as unequal recombination during meiosis or genomic segment removal by a transposable element, rather than gradual pseudogenisation by point mutations. Whether the loss of *UM* in the koala had functional consequence or selective significance is unclear.

Compared to human class Ia genes, five koala genes, *Phci-UA*, *UB*, *UD*, *UG* and *UH*, have a longer α1 domain due to three additional amino acid residues at positions 81–83 (Fig. 2), whereas *Phci-UC* and *UE* contain a deletion around this region (positions 78–83) resulting in a smaller α1 domain than the other genes. Codon sites that are essential for class I protein structures are well conserved in all koala genes, including the glycosylation site in the α1 domain (position 113 in Fig. 2) and four cysteines in the α2 and α3 domains which form disulphide bonds (Bjorkman and Parham 1990). Six residues in the α3 domain that are implicated in CD8 binding are also mostly conserved, except that the two aspartic acid residues at position 251 and 255 in *HLA-A* are substituted by similarly negatively charged glutamic acid residues in some koala genes.

Categorisation of classical and nonclassical genes

Three koala genes, *Phci-UA*, *UB* and *UC*, showed characteristics of class Ia genes based on the pattern of tissue expression and genetic variability. Class Ia genes are known to be

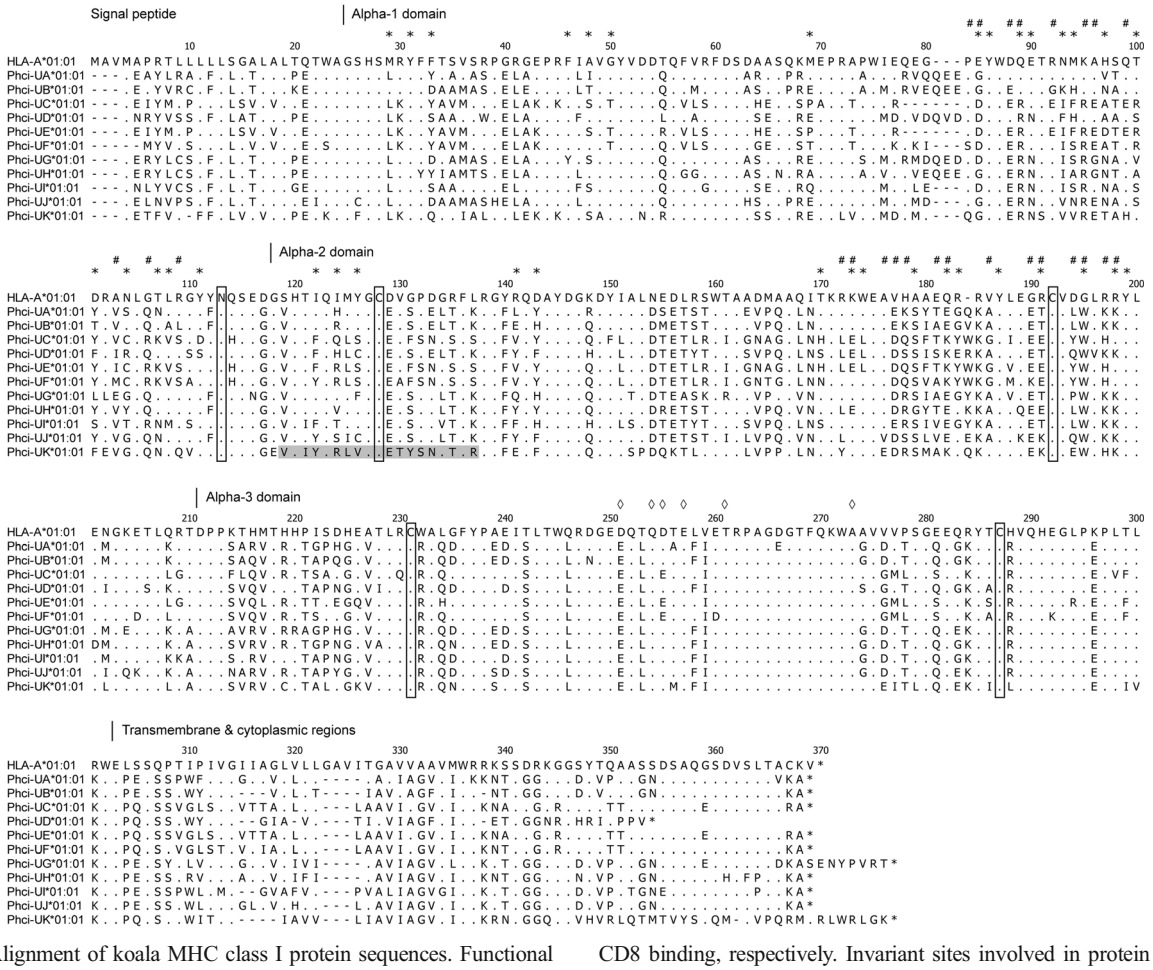

Fig. 2 Alignment of koala MHC class I protein sequences. Functional domains and features were inferred based on human class Ia genes (Bjorkman and Parham 1990). *Asterisks*, *hashes* and *diamonds* indicate residues implicated in peptide binding, T cell receptor interaction and CD8 binding, respectively. Invariant sites involved in protein structure are *boxed*. *Boxed* (*grey*) sequence segment in α2 domain of *Phci-UK* is absent in the alternatively spliced transcript variant

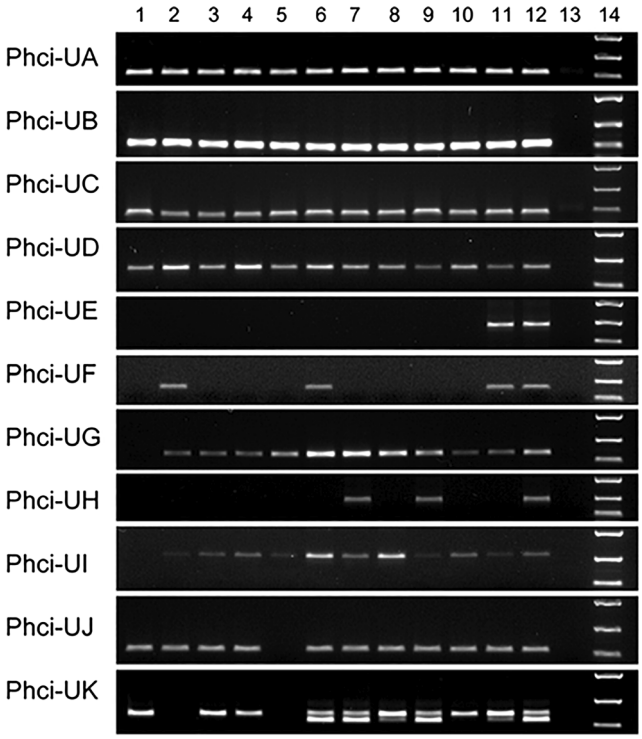

Fig. 3 Tissue transcription of koala class I genes. *Gel lanes*: (*1*) bone marrow, (*2*) duodenum, (*3*) jejunum, (*4*) kidney, (*5*) large intestine, (*6*) liver, (*7*) lung, (*8*) lymph node, (*9*) spleen, (*10*) stomach, (*11*) testis, (*12*) thymus, (*13*) no-template control and (*14*) DNA ladder of 250 bp, 500 bp and 1 kb

ubiquitously expressed and usually have high levels of polymorphism, whereas class Ib genes tend to have restricted tissue expression and low sequence variability (Braud et al. 1999). Transcription of *Phci-UA*, *UB* and *UC* was detected in all 12 examined tissues belonging to the immune, digestive, respiratory, reproductive and urinary systems (Fig. 3). Seven, five and nine alleles were identified for the three genes, respectively (Fig. 4). *Phci-UC* contained a relatively lower level of polymorphisms, showing an average of 96.7% codon site identity in the α1 domain among nine alleles (Online Resource 3). More variant sites were found in *Phci-UA* and *UB*, with the average α1 domain identity being 87.6 and 86.1%, respectively. Allele frequencies observed in the 10 examined New South Wales koalas and five Queensland koalas are shown in Online Resource 4. No significant deviation from Hardy-Weinberg equilibrium was detected in either sample group. The two populations displayed distinct frequency distributions, though further population surveys using larger sample sizes will be necessary to produce more accurate estimations of allele frequencies for the genes.

Another indication of the possible role of *Phci-UA*, *UB* and *UC* as class Ia genes is evidenced by positive selection in the α1 domain, with significantly higher rates of nonsynonymous (d_N) than synonymous (d_S) nucleotide substitutions observed at certain codon sites in the peptide-biding region (PBR) (Fig. 4). Due to direct interactions with antigens, class Ia molecules are commonly under positive selection driven by host-pathogen co-evolution (Hughes 2002). As a result, class Ia genes usually contain a significant excess of nonsynonymous substitutions over synonymous substitutions, especially in the PBR (Hughes and Nei 1988). Conformingly, five sites were detected to be subject to episodic diversifying selection in the koala class Ia genes, all of which were putatively involved in peptide binding (Fig. 4). Comparisons of the mean rates of d_N

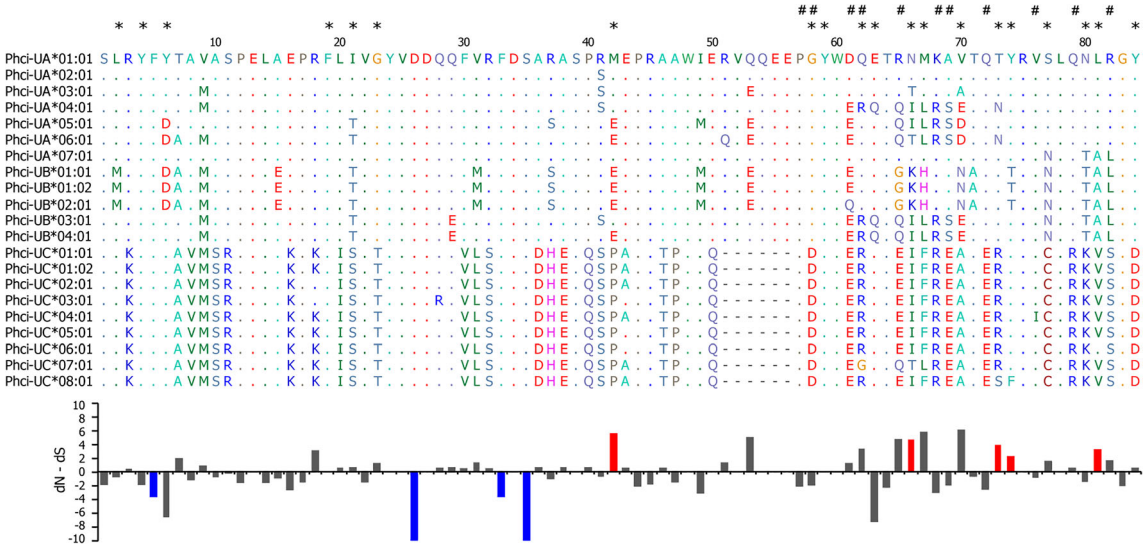

Fig. 4 Sequence analysis of α1 domain of koala class Ia alleles. *Asterisks* and *hashes above the alignment* indicate residues implicated in peptide binding and T cell receptor interaction, respectively. Site-wise rates of nonsynonymous and synonymous substitutions (d_N–d_S) calculated using FEL are shown below the sequences; *red columns* indicate codon sites under episodic diversifying selection (MEME $p < 0.05$), while *blue columns* represent negatively selected residues (FEL $p < 0.05$) (Colour figure online)

Table 3 Test for positive selection in koala MHC class Ia genes

Gene		d_N	d_S	d_N–d_S	p value[a]
Phci-UA	α1 domain	0.071 ± 0.014	0.041 ± 0.017	0.030 ± 0.019	0.049
	Non-PBR	0.040 ± 0.011	0.039 ± 0.020	0.001 ± 0.020	0.479
	PBR	0.173 ± 0.043	0.065 ± 0.056	0.108 ± 0.058	0.036
Phci-UB	α1 domain	0.096 ± 0.021	0.052 ± 0.021	0.043 ± 0.026	0.043
	Non-PBR	0.060 ± 0.018	0.012 ± 0.013	0.047 ± 0.018	0.005
	PBR	0.237 ± 0.085	0.201 ± 0.118	0.037 ± 0.126	0.387
Phci-UC	α1 domain	0.017 ± 0.005	0.015 ± 0.007	0.002 ± 0.010	0.410
	Non-PBR	0.012 ± 0.006	0.021 ± 0.011	−0.008 ± 0.013	0.254[b]
	PBR	0.032 ± 0.012	0	0.032 ± 0.011	0.005

[a] Tested for positive selection ($d_N > d_S$); p values smaller than 0.05 were considered significant

[b] Tested for negative selection ($d_S > d_N$)

and d_S within the α1 domain revealed that the PBR of *Phci-UA* and *UC* was under an overall effect of positive selection ($d_N > d_S$), whereas their non-PBR sites appeared to have evolved more neutrally (Table 3). Interestingly, a higher d_N than d_S was found outside the PBR in *Phci-UB*. Further sequence analysis using the GARD method for recombination detection (Kosakovsky Pond et al. 2006) revealed that this was likely caused by gene conversion, with two break points (position 49 and 70 in Fig. 4) with significant topological incongruence (p value <0.01) detected between *Phci-UB* and *UA* sequences. Gene conversion has been suggested as an important mechanism in creating new MHC alleles in many species, though whether such events are driven by adaptive selection is debated (Klein and Figueroa 1986; Zangenberg et al. 1995; Martinsohn et al. 1999; Hosomichi et al. 2008). As conversion introduces nucleotide polymorphisms to a gene by transferring sequence fragments from another allele or locus, it can cause bias in analyses of point substitutions (Kosakovsky Pond et al. 2006), which may explain the results at *Phci-UB*.

Unlike the three genes discussed above, the other eight koala class I genes all showed limited genetic variability and the majority had tissue-specific expression. *Phci-UD* was found to be expressed in all examined tissues; however, no polymorphism was detected in this gene, suggesting that it is likely nonclassical. At *Phci-UE*, *UH* and *UI*, two, three and three alleles were identified, containing one, two and two SNPs, respectively; most of these SNPs are synonymous except for one in *Phci-UI*. No sequence variants were found at *Phci-UF*, *UG*, *UJ* and *UK*. Two transcript variants were detected at *Phci-UK*, with the shorter variant lacking partial exon 3 at the 5′ end (Figs. 2 and 3).

Based on these observations, we propose that out of 11 MHC class I genes, three are likely classical genes involved

Table 4 MHC class I genes in four marsupial species

Species	Gene	Description	References
Koala	Phci-UA, UB and UC	Classical	This study; GenBank: KY612373–KY612406
	Phci-UD, UE, UF, UG, UH, UI, UJ and UK	Nonclassical	
Tammar wallaby	Maeu-UA, UB and UC	Un-linked to MHC, classical	(Siddle et al. 2009)
	Maeu-UD, UF, UH and UN	Un-linked to MHC, nonclassical	
	Maeu-UE, UK, UM and UO	Nonclassical	
	Maeu-UI, UJ, UL and UP	Expression not detected	
Tasmanian devil	Saha-UA, UB and UC	Classical	(Cheng et al. 2012)
	Saha-UD, UK and UM	Nonclassical	(Cheng and Belov 2014)
Grey short-tailed opossum	Modo-UA1	Classical	(Miska and Miller 1999)
	Modo-UA3 and UA4	Classical	(Krasnec et al. 2015)
	Modo-UG	Nonclassical	(Gouin et al. 2006)
	Modo-UE, UI, UJ, UK and UM	Nonclassical	(Baker et al. 2009)
	Modo-UA2, UF, UH and UL	Expression not detected	(Belov et al. 2006)
	Modo-UB and UC	Un-linked to MHC	(Miska et al. 2004)

in antigen presentation in the koala. Having multiple copies of class Ia loci is typical of the mammalian MHC, though the number can be highly variable between species (Amadou et al. 1999; Chardon et al. 1999; Beck and Trowsdale 2000). The same number of class Ia loci has been reported for four marsupials to date, including the koala, tammar wallaby, Tasmanian devil and opossum (Table 4), suggesting that having three class Ia genes may be a common feature of marsupial MHC.

Conclusions

Eleven putatively functional MHC class I genes were identified in the koala genome. Eight genes showed restricted tissue expression and limited sequence variability, which are characteristics of nonclassical loci. The marsupial-specific class Ib gene *UM*, which is well conserved across multiple marsupial lineages, is absent in the koala. Three class I genes, *Phci-UA*, *UB* and *UC*, are likely class Ia genes that play classical MHC functions. Further studies are underway to fully reveal the genetic diversity of these genes in koala populations using larger sample sizes and to elucidate whether polymorphisms within these genes, especially those found in the peptide-binding regions in α1 and α2 domains, are associated with disease resistance/susceptibility to chlamydia and KoRV in koalas.

Acknowledgements This work was funded by the Australian Research Council Discovery Project DP150101485 awarded to AP, PT and KB. We thank the Australian Centre for Wildlife Genomics at the Australian Museum for providing the NSW koala DNA samples. Special thanks to Galit Rawlinson and the staff at Lone Pine Koala Sanctuary, Brisbane, for assistance with sample collection and on-going support for our research studies.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abts KC, Ivy JA, DeWoody JA (2015) Immunomics of the koala (*Phascolarctos cinereus*). Immunogenetics 67:305–321

Amadou C, Kumánovics A, Jones EP, Lambracht-Washington D, Yoshino M, Lindahl KF (1999) The mouse major histocompatibility complex: some assembly required. Immunol Rev 167:211–221

Baker ML, Melman SD, Huntley J, Miller RD (2009) Evolution of the opossum major histocompatibility complex: evidence for diverse alternative splice patterns and low polymorphism among class I genes. Immunology 128:e418–e431

Beck S, Trowsdale J (2000) The human major histocompatibility complex: lessons from the DNA sequence. Annu Rev Genomics Hum Genet 1:117–137

Belov K, Deakin JE, Papenfuss AT et al (2006) Reconstructing an ancestral mammalian immune supercomplex from a marsupial major histocompatibility complex. PLoS Biol 4:e46

Bjorkman PJ, Parham P (1990) Structure, function, and diversity of class I major histocompatibility complex molecules. Annu Rev Biochem 59:253–288

Braud VM, Allan DS, McMichael AJ (1999) Functions of nonclassical MHC and non-MHC-encoded class I molecules. Curr Opin Immunol 11:100–108

Chardon P, Renard C, Vaiman M (1999) The major histocompatibility complex in swine. Immunol Rev 167:179–192

Cheng Y, Stuart A, Morris K et al (2012) Antigen-presenting genes and genomic copy number variations in the Tasmanian devil (*Sarcophilus harrisii*). BMC Genomics 13:87

Cheng Y, Belov K (2014) Characterisation of non-classical MHC class I genes in the Tasmanian devil (*Sarcophilus harrisii*). Immunogenetics 66:727–735

Cresswell P, Ackerman AL, Giodini A, Peaper DR, Wearsch PA (2005) Mechanisms of MHC class I-restricted antigen processing and cross-presentation. Immunol Rev 207:145–157

Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455–2457

Dique DS, Thompson J, Preece HJ, Penfold GC, De Villiers DL, Leslie RS (2003) Koala mortality on roads in south-east Queensland: the koala speed-zone trial. Wildl Res 30:419–426

Ellis SA, Bontrop RE, Antczak DF et al (2006) ISAG/IUIS-VIC comparative MHC nomenclature committee report, 2005. Immunogenetics 57:953–958

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Gouin N, Wright AM, Miska KB et al (2006) *Modo-UG*, a marsupial nonclassical MHC class I locus. Immunogenetics 58:396–406

Hobbs M, Pavasovic A, King AG et al (2014) A transcriptome resource for the koala (*Phascolarctos cinereus*): insights into koala retrovirus transcription and sequence diversity. BMC Genomics 15:786

Hosomichi K, Miller MM, Goto RM et al (2008) Contribution of mutation, recombination, and gene conversion to chicken Mhc-B haplotype diversity. J Immunol (Baltimore, Md.: 1950) 181:3393–3399

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

Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170

Hughes AL, Nei M (1989) Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. Mol Biol Evol 6:559–579

Hughes AL (2002) Natural selection and the diversification of vertebrate immune effectors. Immunol Rev 190:161–168

Johnson RN, O'Meally D, Chen Z, et al. (2017) Out on a limb - koala genome reveals adaptations to a toxic diet and a prehistorical population crash. Under review.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Kirsch JAW, Lapointe F-J, Springer MS (1997) DNA-hybridisation studies of marsupials and their implications for metatherian classification. Aust J Zool 45:211–280

Klein J, Figueroa F (1986) Evolution of the major histocompatibility complex. Crit Rev Immunol 6:295–386

Kollipara A, Polkinghorne A, Wan C et al (2013) Genetic diversity of *Chlamydia pecorum* strains in wild koala locations across Australia and the implications for a recombinant *C. pecorum* major outer membrane protein based vaccine. Vet Microbiol 167:513–522

Kosakovsky Pond SL, Frost SDW (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22:1208–1222

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891–1901

Krasnec KV, Sharp AR, Williams TL, Miller RD (2015) The opossum MHC genomic region revisited. Immunogenetics 67:259–264

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, Griffith JE, Higgins DP (2014a) Identification of MHCII variants associated with chlamydial disease in the koala (*Phascolarctos cinereus*). PeerJ 2:e443

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

Lunney D, Gresser S, O'Neill LE, Matthews A, Rhodes J (2007) The impact of fire and dogs on koalas at Port Stephens, New South Wales, using population viability analysis. Pac Conserv Biol 13: 189–201

Maher IE, Higgins DP (2016) Altered immune cytokine expression associated with KoRV B infection and season in captive koalas. PLoS One 11:e0163780

Martinsohn JT, Sousa AB, Guethlein LA, Howard JC (1999) The gene conversion hypothesis of MHC evolution: a review. Immunogenetics 50:168–200

McInnes LM, Gillett A, Hanger J, Reid SA, Ryan UM (2011) The potential impact of native Australian trypanosome infections on the health of koalas (*Phascolarctos cinereus*). Parasitology 138:873–883

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

Miska KB, Miller RD (1999) Marsupial Mhc class I: classical sequences from the opossum, *Monodelphis domestica*. Immunogenetics 50: 89–93

Miska KB, Wright AM, Lundgren R et al (2004) Analysis of a marsupial MHC region containing two recently duplicated class I loci. Mamm Genome 15:851–864

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8:e1002764

Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci 94:7799–7806

Nei M, Kumar S (2000) Synonymous and nonsynonymous nucleotide substitutions. In: Molecular Evolution and Phylogenetics Oxford University Press, New York, USA

Polkinghorne A, Hanger J, Timms P (2013) Recent advances in understanding the biology, epidemiology and control of chlamydial infections in koalas. Vet Microbiol 165:214–223

Rhodes JR, Ng CF, de Villiers DL, Preece HJ, McAlpine CA, Possingham HP (2011) Using integrated population modelling to quantify the implications of multiple threatening processes for a rapidly declining population. Biol Conserv 144:1081–1088

Roy CR (2003) Immunology: professional secrets. Nature 425:351–352

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Siddle HV, Deakin JE, Coggill P et al (2009) MHC-linked and un-linked class I genes in the wallaby. BMC Genomics 10:310

Simmons GS, Young PR, Hanger JJ et al (2012) Prevalence of koala retrovirus in geographically diverse populations in Australia. Aust Vet J 90:404–409

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

Tarlinton R, Meers J, Hanger J, Young P (2005) Real-time reverse transcriptase PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas. J Gen Virol 86:783–787

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22

Zangenberg G, Huang M-M, Arnheim N, Erlich H (1995) New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. Nat Genet 10:407–414