Differential and defective expression of Koala Retrovirus reveal complexity of host and virus evolution

R.E Tarlinton[1], N. Sarker[2], J. Fabijan[3], T. Dottorini[1], L. Woolford[3], J. Meers[2], G.Simmons[2], H.Owen[2], J.M.Seddon[2], F. Hemmatzedah[3], D.Trott[3], N. Speight[3], R.D.Emes[1,4]

[1]School of Veterinary Medicine and Science, University of Nottingham, UK
[2]School of Veterinary Science, University of Queensland, Australia
[3]School of Animal and Veterinary Sciences, University of Adelaide, Australia
[4]Advanced Data Analysis Centre, University of Nottingham, UK

Abstract

Koala retrovirus (KoRV) is unique amongst endogenous (inherited) retroviruses in that its incorporation to the host genome is still active, providing an opportunity to study what drives this fundamental process in vertebrate genome evolution. RNA sequencing of KoRV from koala populations with high virus burden (Queensland) and low virus burden (South Australia) identified that South Australian animals, a population previously thought to have KoRV negative animals, harboured replication defective KoRV. This discovery provides the first evidence that a host population may maintain defective KoRV as protection from the infectious form of KoRV. This offers the intriguing prospect of being able to monitor and selectively breed for disease resistance to protect other wild koala populations from KoRV induced disease.

Introduction

Koalas (*Phascolarctos cinereus*) are an iconic marsupial species listed as vulnerable on the IUCN 'red list' of threatened species [1]. While a large part of their ongoing population decline is due to habitat loss, two major disease threats, chlamydial infection and Koala retrovirus (KoRV), are additionally limiting population viability [2]. These infections are particularly prevalent in the northern regions of Australia, namely Queensland and New South Wales, and less so in the south [3,4].

Following European settlement, large koala populations across Australia declined significantly due to hunting in the 1890's to 1920's, with southern populations nearing extinction. During this time, small refuge populations were established on offshore Victorian islands and these koalas have been subsequently used to restock most of their former southern range. This southern population is genetically distinct from the northern animals [5]. The mainland Mount Lofty Ranges koala population in South Australia originates from koalas from both the Kangaroo Island population [6] as well as koalas from Queensland and New South Wales [5,7].

Endogenous retroviruses (ERVs) are those that have become incorporated into their host's genome. They are ubiquitous in vertebrate genomes and in some cases constitute up to 10% of total genome content [8]. They are usually not functional as viruses due to the accumulation of mutations and deletions but are often expressed at an RNA level, where they are thought to play a role in genomic regulation [8] [9,10]. However, the reason for their spread and persistence in genomes, whether as parasites or commensals is still under debate [11]. KoRV is part of a small group of unusual "modern" endogenous retroviruses (including Murine leukaemia virus and Feline leukaemia virus). These modern ERVs are replication competent and display considerable overlap with their exogenous infectious counterparts, including swapping of gene segments [12].

KoRV is the most recent entrant into any known genome, with estimates of integration time somewhere between 200 and 49,000 years ago [13,14]. It is thought to have arisen from a recent species jump as its closest relatives are an endogenous virus in a *Melomys burtoni* sub-species (the grassland mosaic tailed rat) in northern Australia and Indonesia [15,16] and Gibbon ape leukaemia virus (GALV), the latter a pathogenic exogenous virus that most likely arose as a spill over event from south east Asian rodents in the late 1960s [17]. KoRV has been found in 100% of tested Queensland and

New South Wales koalas but appears to have a lower prevalence in southern populations [4]. The virus displays a high diversity in proviral copy number and integration sites between individuals and populations, with southern animals having lower copy numbers in their DNA [18,19] [4]. KoRV was originally identified during investigations into the high rates of lymphoid neoplasia (lymphoma and leukaemia) in Queensland koalas [13]. Koalas with lymphoid neoplasia have significantly higher KoRV viral loads in plasma [18] and some strains of KoRV also perturb the cytokine response profile of koala lymphocytes [20].

The originally identified virus is now known as KoRV A and appears to be present in all individuals that are KoRV-positive. A number of sequence variants of the *env* gene region, which encodes the surface unit (SU) of the envelope protein (Env), have also been identified (Figure 1). These vary between individuals and resemble the viral quasispecies common to infectious retroviruses [21]. There has been debate as to whether the KoRV B/J variant, which displays a different receptor usage to KoRV A, is an exogenous virus as it has been epidemiologically linked with clinical disease, however this is still unresolved [22] [23].
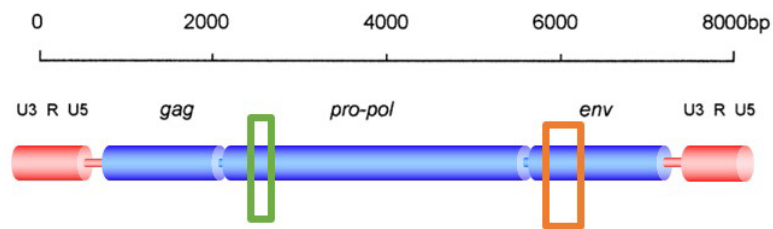

Figure 1: Schematic diagram of KoRV A complete genome (Genbank number AF151794), showing the location of the long terminal repeats LTR (red cylinders), genes (blue cylinders) and the location of the primers for the *pro-pol* gene used in most studies (position 2982-3002) green box and the hypervariable region of the *env* gene (position 5901-6476) orange box.

This study examines KoRV variant differences between genetically distinct populations from South East Queensland and South Australia. Sequence reads of RNA isolated from koala lymph nodes were mapped to the published KoRV sequences to examine differences in the KoRV profiles between the two populations.

Results

All 10 QLD animals were positive for KoRV *pol* nested PCR on DNA extracted from whole blood, 14/19 (74%) of the SA animals were also positive. Demographic data for individual animals are presented in Table 1.

Mapping of the Illumina reads directly to the KoRV type A and B reference genomes (Figure 2) demonstrated that when normalised for total mapped read depth, coverage was very similar for both the SA and QLD groups of koalas across the ends of the genomes (LTR-*gag*, and *env*-LTR). However, between positions 1389 and 7124 of the KoRV A sequence the SA group showed a mean coverage of < 10% of the QLD group suggesting that part of *gag*, all of *pro-pol* and part of the *env* genes are largely missing in the RNA transcripts, with six SA koalas not expressing this region at all. The target site of the standard KoRV *pol* qPCR used in most studies is contained within this missing region [27]. Some of SA animals were KoRV PCR positive for the proviral *pol* gene suggesting that at least partial proviruses for this region are present but are expressed at levels undetectable in the transcriptome.

The higher number of reads in the *env* and LTR regions can be explained by the presence of spliced *env* transcripts in addition to full length genomic transcripts as has been reported by other groups [28], although these are not detected as complete individual transcripts by the mapping methods used in this study.
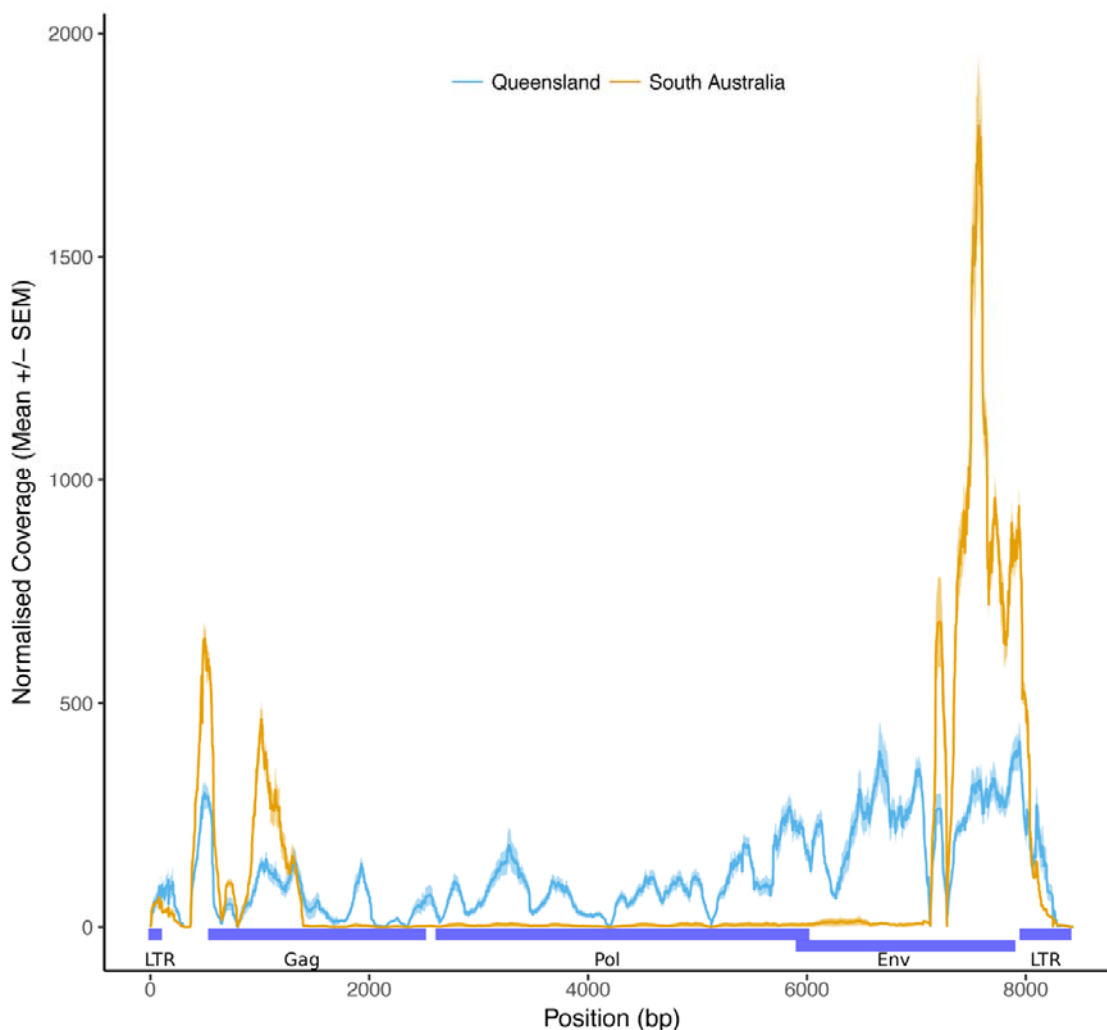

Figure 2: Coverage of reads mapped to the full length KoRV type A reference genome AF151794. For each group the mean normalised coverage ((per position coverage/total coverage) x 1×10^6) is represented by a line and +/- the standard error is shaded around the mean. QLD samples in blue, SA in orange. KoRV genomic regions are marked underneath the read maps with blue bars, these regions are: 5' LTR 1-505, gag 961-2526, pol 2641-6024, env 5906-7885, 3' LTR 7927-8431.

Pseudomapping of the sequence reads to the KoRV A genome (complete *gag, pro-pol* and *env* genes) and type sequences of the hypervariable region of the *env* gene (base pairs 6000-6575 of KoRV A) of each of the previously identified KoRV subtypes (KoRV A to I as per the classification scheme used in Chappell *et.al.* 2016 [21]) demonstrated that while QLD koalas had multiple subtypes within individuals, SA

animals had far lower KoRV subtype diversity. Significantly different expression was observed for KoRV A,B,D,E and G variants between QLD and SA samples (unpaired t-test with unequal variance, P<0.05) (Figure 3). It was observed that QLD animals were older (mean tooth wear class 4.22, 95% CI 3.88-4.56) than SA (mean tooth wear class 3.05, 95% CI 2.58-3.52) and so age may confound KoRV expression comparisons. When the same test was repeated for samples from koalas with tooth class 4 (7 QLD 8 SA samples), expression of A,B,E and G variants remained significantly different between locations (Extended data file 3), supporting the finding that KoRV expression is significantly different between the QLD and SA populations. Eleven out of nineteen SA animals (58%) had KoRV A. Six of these koalas had only KoRV A reads (Figure 3 and Table 1). Four animals had reads for KoRV A and one other variant only (D or E). Two animals had reads for KoRV E but no detectable reads for any other variant (including KoRV A). Only one SA koala (Z Table 1) had counts comparable to the QLD cohort with a similar range of variants (A, B, C, D, E, F, G, I), while the rest had counts that were <10% of the QLD koalas. *Pol* gene counts were also similarly considerably lower in the SA koalas than the QLD group. Relative expression as estimated count values for individual animals for each gene region and KoRV subtype are presented in Extended data file 2.
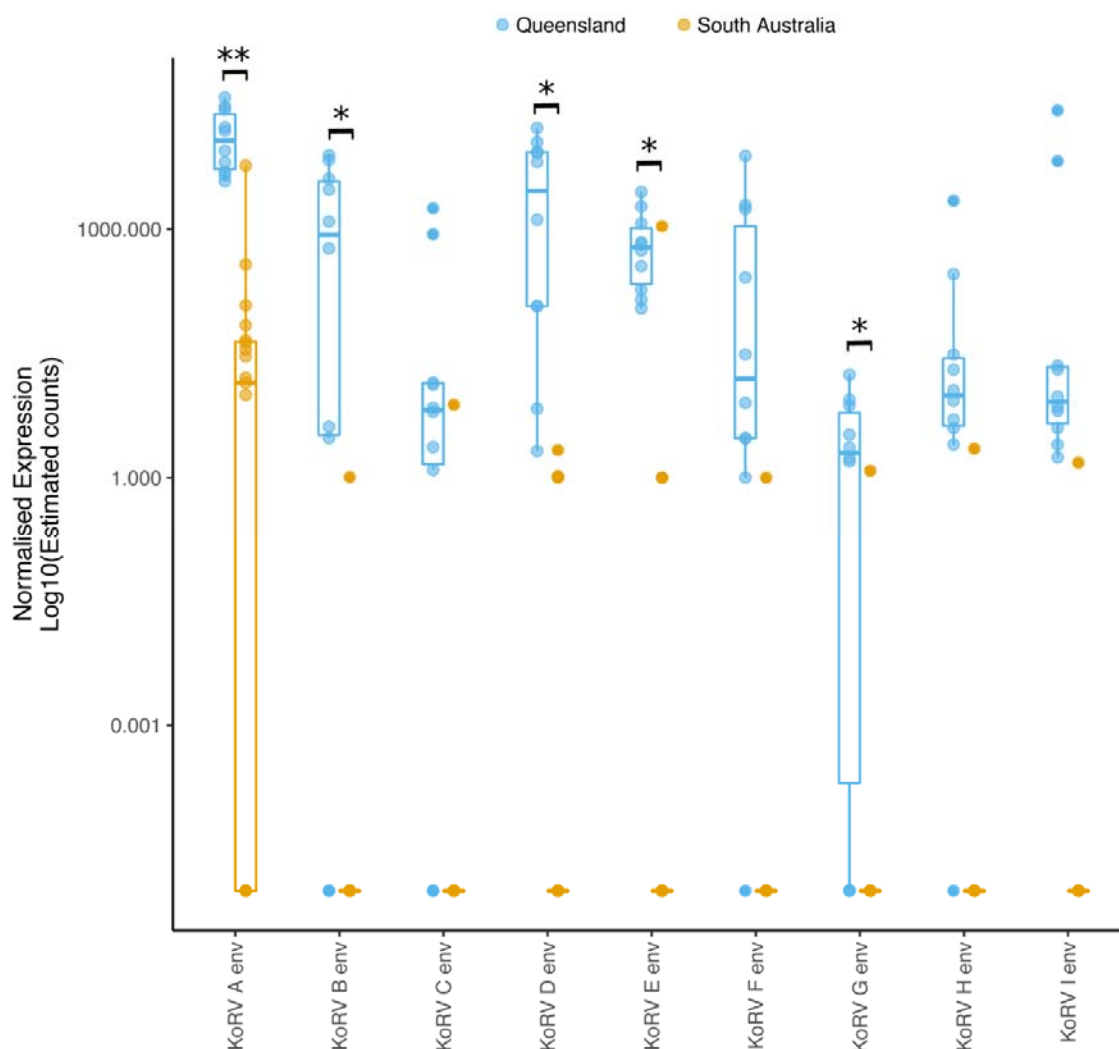
Figure 3: Normalised expression Log10(estimated counts) of KoRV A complete *env* gene and the 575 nucleotides of the hypervariable region of the subtypes (B-I). Box and whisker plots show the median and interquartile ranges (box) and minimum/maximum expression (whiskers) of groups. Data for individual animals within a group are shown by circles. QLD animals in blue and SA animals in orange. *Env* variants with significantly different expression between QLD and SA groups marked with black bars (** = P<0.01, *= P<0.05)

Discussion

The findings of the current study suggest that KoRV infection involves a more complex host-viral relationship than previously recognised, particularly in SA koalas. Other studies have shown differences between northern and southern koala populations in the prevalence of KoRV infection, levels of KoRV proviral and viral loads and disease burden. Our study has revealed additional host and viral factors that indicate these population differences are more complicated than merely presence or absence of virus and virus load.The results of this study were unexpected. The central portion of the virus genome was significantly lower in transcript coverage in the SA koalas. There are two possible explanations for this: either this portion of the DNA is missing from proviral loci in the SA animals, or it is present but not being transcribed.

Five SA koalas didn't have any reads at all for the *pol* gene and these transcripts are unlikely to function as an infectious virus. It is possible that these koalas have truncated KoRV proviral loci in their genomes directly transcribing these variants. We cannot tell from this study whether these transcripts arise from a single genome locus that is identical in all SA animals or multiple loci. Recombination of ERV loci in the genome into new RNA variants in the transcriptome is a well described phenomenon in other species [29,30] and careful comparison of paired DNA and RNA samples from the same animals would be required to untangle this.

Koalas with genuinely truncated proviral loci would have been identified as KoRV negative in previous studies as the standard tests for the virus are conventional PCR or qPCR assays targeting the portion of the *pol* gene that is missing in these transcripts [4,27,31]. However 14 (74%) of the SA animals did have this gene region in their DNA (Table 1) and many of them had detectable (if low) reads across the full KoRV genome so it is reasonable to assume that at least some of them had full length proviral DNA that was not being expressed. Other studies using KoRV *pol* PCR tests for proviral loci in DNA have also indicated that at least some southern animals have this gene but at much lower copy numbers than in QLD animals [4]. The pattern of deletion for more ancient retroviral loci is one of loss of the *env* genes with maintenance of the *gag-pol* genes to facilitate spread within individual cells [11]. The replication defective variants missing their *pro-pol* genes in the current study indicate that the drivers of retroviral endogenisation in the face of an infectious virus challenge are very different to the long term ones in well adapted virus/host systems.

These SA koalas may be infected with an exogenous virus variant that is currently under immune control, with proviral loci present but not expressed as RNA. This would be similar to the situation with cats infected with Feline leukaemia virus (FeLV) which manage to control but not clear virus infection [32]. QLD animals with full length endogenous loci are likely tolerized to the virus by in-utero expression of viral genes and their immune systems are unable to recognise and respond to any exogenous virus. This is the probable explanation for many QLD animals having no KoRV antibodies and not developing them on vaccination [33] [34]. SA animals expressing only partial viral RNA for the *gag* (capsid gene), if this is not being translated as protein, may not be hampered in virus control this way. This does not explain however why these SA animals are expressing partial viral RNA. Further work will need to establish whether KoRV positive SA animals routinely have detectable antibody to the virus (which would imply control of infectious virus as for FeLV cats).

These replication defective variants may have originally arisen by being "carried" along with replication competent viruses as occurs for other retroviruses such as Rous Sarcoma Virus [35]. It seems likely that these variants along with full length ones were present before the southern animals were genetically isolated in the 1920's and that the other alleles were lost due to the genetic bottlenecks in the Mount Lofty population. Due to the admixture of northern genotype animals in the Mount Lofty population it is still possible that other more isolated southern genotype populations may have genuinely KoRV free animals. Retesting these southern populations with *gag* or LTR primers should be a priority.

This host genetic restriction in the SA population may also have resulted in animals with viral receptor alleles that are unable to bind infectious KoRV, restricting infectious virus replication and transmission. This situation occurs in several mouse strains resistant to certain murine leukaemia virus strains [36], though to date there are no known variations between southern and northern koalas for the KoRV A and B receptors, Pit1, and THTR1 [23,37]. It is also possible that mutations in other genes important in retroviral replication (such as retroviral restriction factors) differ between the two populations resulting in restricted replication in the SA animals but this remains to be explored.

Replication defective ERV sequences play an intriguing role in disease pathogenesis in other 'modern' endogenous retroviruses with infectious counterparts. The best studied of these are the Avian leucosis group E viruses (ALVE) in chickens, also known as ev loci. These are genetically very similar to the circulating exogenous (infectious) avian leukosis/sarcoma viruses [38]. Selective breeding of birds with low viral excretion has led to strains of chickens with undetectable levels of infectious viral particles. The genotype of these strains varies, from those that completely lack ALVE loci, to those that contain defective ALVE loci that do not produce infectious virus particles, to those that have active ALVE loci but lack the receptor for the virus (meaning that it cannot re-infect somatic tissue). Some strains of birds that produce ALVE envelope proteins (but not full virus particles) are actually resistant to infectious ALVE as the defective envelope proteins "blockade" the receptor for the virus preventing infection [39].

A similar receptor blockade by defective Env proteins occurs in Jaagsietke sheep retrovirus (JSRV)[40], in part explaining the tissue tropism of the exogenous virus for tissues where the endogenous variants are not expressed. Endogenous JSRV loci also exert a further block on exogenous viral replication at the viral assembly stage, where defective Gag proteins from the ERV loci are packaged along with infectious variants preventing the viral particles from being packaged and transported correctly for viral release from the cell. Receptor blockade by endogenous Env proteins has also been reported in Murine Leukaemia virus variants in mice, along with a Gag mediated block at the pre-integration step of viral replication [41].

None of these scenarios fully explain the situation with these KoRV transcripts as it would appear impossible for complete Env and Gag proteins to be produced in many of these animals. They do however raise the intriguing possibility that these replication defective transcripts are interfering in some way with the full length virus variants completing their replication cycle. Future work will need to include in vivo studies of receptor usage by the truncated variants identified here and whether these variants do (and at what stage) blockade infectious virus replication.

This study does not resolve the issue of which (if any) of the identified KoRV env subtypes is the transmissible version of the virus but, it does offer some possibilities. As has been reported in many other studies [21,22,37,42] our northern animals display considerable quasispecies variation in their KoRV isolates as would be expected for a infectious replicating retrovirus. Comprehensive phylogenetics from Chapell *et.al.* (2017) [21] demonstrate that KoRV A is basal to both KoRV B and a large number of sequence variants under the paraphyletic group KoRV D. KoRV variants C and E (found here) have previously only been identified in koalas in zoos outside Australia [37] and not previously identified in wild koalas; the KoRV E variant uses a different receptor to KoRV A and is speculated to be a potential exogenous virus. No animals prior to this study that have

tested positive for KoRV have lacked the KoRV A variant [21,31,37]. The two SA animals in this study with only KoRV E reads without detectable KoRV A are therefore of interest, although the very low read counts for these animals mean these results must be treated with some caution. Eleven of our 19 SA koalas expressed KoRV A but at <10% of the counts of QLD koalas. Only 5 of these animals expressed other variants along with KoRV A, 4 of these expressed D or E only. Only one animal (which may represent an escape from suppression of viral replication, a well described phenomenon in retroviruses) had more than one other KoRV *env* subtype, including KoRV B which has not been reported previously in southern Australian koalas [31], despite over 160 animals being examined in a study in Victoria.

We consider that KoRV A, rather than representing the endogenous version of the virus, could be the primarily transmitted virus with other variants arising within individual animals. This situation would be analogous to that seen in FeLV where the FeLV A strain is the only one that spreads from cat to cat, with the B, C, D, T and feline fibrosarcoma strains arising independently in individual animals via *env* mutations or recombination with endogenous loci or the acquisition of cellular oncogenes. The derived strains of FeLV use different receptors, and display altered pathogenicity, to the FeLV A strain [43,44] again analogous to KoRV where the KoRV B and E variants use a different receptor to KoRV A [23,42]. Should it prove the case that only one strain of KoRV is transmissible, this bears promise for the success of a vaccination effort [34]; the commercial FeLV vaccines that target only the A variant have been highly successful in reducing the incidence of clinical FeLV disease [45].

The discovery of these replication defective KoRV sequences in SA animals has opened up a number of intriguing implications for both controlling disease in koala populations and the drivers of retroviral endogenisation in their hosts. The hypothesis that the replication defective variants may blockade infectious KoRV replication, if substantiated, opens up the option to use selective breeding to re-introduce this trait into the KoRV susceptible northern population.

Tables

Table 1: KoRV variant expression of individual animals

ID	Location[a]	Sex	Age[b]	Provirus PCR[c]	KoRV variants[d]
A	QLD	M	4	+	ALL
B	QLD	M	4	+	ALL
C	QLD	F	4	+	ALL
D	QLD	F	4	+	ALL
E	QLD	F	>3	+	A, C, D, E, F, G, H, I
F	QLD	M	4	+	A, D, E, I
G	QLD	M	5	+	A, B, C, D, E, F, H, I
H	QLD	M	4	+	A, B, D, E, F, G, H, I
I	QLD	F	4	+	ALL
J	QLD	M	5	+	A, B, C, D, E, F, H, I
K	SA	F	4	+	A
L	SA	M	3	+	A
M	SA	M	2	+	N[e]
N	SA	M	4	-	N
O	SA	M	3	+	N
P	SA	F	3	+	N
Q	SA	M	4	+	A
R	SA	M	2	+	A
S	SA	M	2	+	N
T	SA	M	2	+	A, D
U	SA	F	4	+	E
V	SA	M	4	+	E
W	SA	F	4	+	N
X	SA	F	3	-	A, D
Y	SA	F	4	-	A
Z	SA	M	3	+	A, B, C, D, E, F, G, I
A1	SA	F	1	+	A, E
A2	SA	M	2	-	A, D
A3	SA	M	4	-	A

[a] Population location: QLD – Queensland; SA – South Australia
[b] Age determined my dentition and the degree of wear of the upper pre-molar (Martin *et al*. 1999)
[c] based on nested PCR for *pol* provirus on whole blood DNA
[d] KoRV variants determined by KoRV transcripts; ALL = all published variants (A to I)
[e] N= no env hypervariable region detected.

Extended data

Extended data File 1: Details of the animals included in this study: N/A = sample not available for testing, region of origin, QLD= Queensland, SA= South Australia, sex, M= Male, F= female, tooth wear class (age classification on a 7 point scale [46]), KoRV proviral status (pol gene PCR on DNA from whole blood), lymph node RNA quality and NGS read details: concentration, A260/280 and A260/230 ratios, RIN value, number of paired raw reads and number of trimmed reads from each sample.

Extended data File 2: Estimated counts of KoRV subtypes for individual animals, Column A = Sequence read archive (SRA) identifier,Column B = Koala ID as per Supplementary file 1, Column C, D, E: TPM for KoRV A, *gag, pro/pol* and *env* (Genbank number AF151794), Column F-M = KoRV *env* subtype based on the first 575 nucleotides of the *env* gene subtypes (B-I). Column N= State of origin, Column O= KoRV pol gene PCR on whole blood DNA.

Extended data File 3: Normalised expression Log10(estimated counts) of KoRV A complete *env* gene and the 575 nucleotides of the hypervariable region of the subtypes (B-I) for animals from tooth wear (age) class 4. Box and whisker plots show the median and interquartile ranges (box) and minimum/maximum expression (whiskers) of groups. Data for individual animals within a group are shown by circles. QLD animals in blue and SA animals in orange. *Env* variants with significantly different expression between QLD and SA groups marked with black bars (** = P<0.001, *= P<0.005)

Methods
Ethics
 Ethical approval for this study was granted by the University of Queensland Animal Ethics Committee, permit number ANFRA/SVS/461/12 and ANRFA/SVS/445/15, the Queensland Government Department of Environment and Heritage Protection permit number WISP11989112, the University of Adelaide Animal Ethics Committee permit number S-2013-198 and the South Australian Government Department of Environment, Water and Natural Resources Scientific Research Permit Y26054.
Samples
 Samples were collected from wild-rescued koalas euthanised for clinical reasons and submitted for post-mortem examinations from South East Queensland (Greater Brisbane) (n=10) and South Australia (Mount Lofty Ranges) (n=19). Age was determined by dentition and the amount of wear on the upper premolar [24]. Submandibular lymph nodes were collected within 2-6 hours of death into RNALater® and stored at -80ºC. Where possible, blood was collected into EDTA prior to euthanasia (BD vacutainer).
 Total RNA was extracted using an RNeasy Mini kit with on column DNAase1 digestion (Qiagen). RNA quantity and quality were assessed via anXpose spectrophotometer (Bioke) and Agilent 2100 Bioanalyzer. mRNA was prepared for sequencing using the Illumina TruSeq stranded mRNA library prep kit and 100 base pair, paired end sequencing was performed on an Illumina HiSeq. Details of the koalas, sample quality and read quantity are provided in Extended data file 1.
 DNA was extracted from 100 µl of EDTA blood using a DNeasy blood and tissue kit (Qiagen). KoRV DNA proviral presence in whole blood was established using a published nested PCR for the KoRV polymerase gene [4]. Of the ten koalas from South East Queensland (QLD), six were male and four female and all were adults, with a tooth

wear class (TWC) 4 or 5. Nineteen koalas were sampled from the Mount Lofty Ranges, South Australia (SA); seven female and 12 male. Six were juvenile (TWC 1 or 2) and 13 were adults (TWC 3 or 4).

KoRV genome coverage

To reduce mis-mapping due to the abundance of highly repetitive long terminal repeat sequences, the adapter-trimmed fastq files were first mapped using Hisat2 [25] to the isolated Long Terminal Repeat (LTR) region of the koala KoRV type sequence (accession AF151794). LTR depleted reads were then mapped to the complete genomes of Koala retrovirus KoRV A and KoRV B (AF151794.2 and KC779547.1 respectively) using Hisat2 [25]. Per-base coverage was determined from bam files for each isolate using samtools version 1.3.1 depth (with parameters –aa –q 10 –d 20000).

KoRV subtype gene expression

To quantitate the expression of KoRV subtype genes, LTR depleted reads for individual koalas were pseudoaligned to the *gag*, *pol* and *env* genes of KoRV subtype A (accessions AAF15097.1_1, AAF15097.1_2 and AAF15097.1_3 respectively) and the first 575 nucleotides of the *env* genes of KoRV subtypes B-I (accessions AB822553.1, AB828005.1, AB828004.1, KX588043.1, KX587994.1, KX587961.1, KX588036.1 and KX588021.1 respectively) using Kallisto [26]. These nucleotides correspond to the hypervariable region of the *env* gene that is used in KoRV subtype classification.

Data Availability

KoRV sequence data (as fasta formatted data) are available from adac figshare [https://figshare.com/articles/KoRV_Genome_Alignment_resulting_Fasta_files_/]. Raw sequence reads fastq format are available at ENA with the accession number [PRJEB21505].

References

1   Woinarski, J. & Burbidge, A. A. Phascolarctos cinereus. *The IUCN Red List of Threatened Species 2016* **e.T16892A21960344** (2016).
2   Committee, E. a. C. R.   (ed Senate Committee) (Senate Printing Unit, Parliament House Canberra, 2011).
3   Polkinghorne, A., Hanger, J. & Timms, P. Recent advances in understanding the biology, epidemiology and control of chlamydial infections in koalas. *Vet Microbiol* **165**, 214-223, doi:10.1016/j.vetmic.2013.02.026 (2013).
4   Simmons, G. S. *et al.* Prevalence of koala retrovirus in geographically diverse populations in Australia. *Aust Vet J* **90**, 404-409, doi:10.1111/j.1751-0813.2012.00964.x (2012).
5   Neaves, L. E. *et al.* Phylogeography of the Koala, (Phascolarctos cinereus), and Harmonising Data to Inform Conservation. *PLoS One* **11**, e0162207, doi:10.1371/journal.pone.0162207 (2016).
6   Robinson, A. C. in *The koala: proceedings of the Taronga symposium on koala biology, management and medicine.*   132-143 (Zoological Parks Board, Sydney).
7   lindsay, H. A. Re-establishing the koala in SA. *Wild Life* **12**, 257-262 (1950).
8   Hayward, A., Grabherr, M. & Jern, P. Broad-scale phylogenomics provides insights into retrovirus-host evolution. *Proc Natl Acad Sci U S A* **110**, 20146-20151, doi:10.1073/pnas.1315419110 (2013).
9   Manghera, M. & Douville, R. N. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? *Retrovirology* **10**, 16, doi:10.1186/1742-4690-10-16 (2013).
10   Chuong, E. B., Rumi, M. A., Soares, M. J. & Baker, J. C. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nature genetics* **45**, 325-329, doi:10.1038/ng.2553 (2013).
11   Magiorkinis, G., Gifford, R. J., Katzourakis, A., De Ranter, J. & Belshaw, R. Env-less endogenous retroviruses are genomic superspreaders. *Proc Natl Acad Sci U S A* **109**, 7385-7390, doi:10.1073/pnas.1200913109 (2012).
12   Eiden, M. V., Radke, K., Rovnak, J. & Quackenbush, S. L. in *Retroviruses*   (eds R Kurth & N. Bannert) 371-394 (Caister Academic Press, 2010).

13   Hanger, J. J., Bromham, L. D., McKee, J. J., O'Brien, T. M. & Robinson, W. F. The Nucleotide Sequence of Koala (*Phascolarctos cinereus*) Retrovirus: a Novel Type C Endogenous Virus Related to Gibbon Ape Leukemia Virus. *Journal of Virology* **74**, 4264-4272 (2000).

14   Avila-Arcos, M. C. *et al.* One hundred twenty years of koala retrovirus evolution determined from museum skins. *Mol Biol Evol* **30**, 299-304, doi:10.1093/molbev/mss223 (2013).

15   Simmons, G., Clarke, D., McKee, J., Young, P. & Meers, J. Discovery of a novel retrovirus sequence in an Australian native rodent (Melomys burtoni): a putative link between gibbon ape leukemia virus and koala retrovirus. *PLoS One* **9**, e106954, doi:10.1371/journal.pone.0106954 (2014).

16   Alfano, N. *et al.* Endogenous Gibbon Ape Leukemia Virus Identified in a Rodent (Melomys burtoni subsp.) from Wallacea (Indonesia). *J Virol* **90**, 8169-8180, doi:10.1128/jvi.00723-16 (2016).

17   Brown, K. & Tarlinton, R. Is Gibbon Ape Leukaemia Virus still a threat? . *Mammal Review* **47**, 53-61 (2017).

18   Tarlinton, R. E., Meers, J. & Young, P. R. Retroviral invasion of the koala genome. *Nature* **442**, 79-81, doi:10.1038/nature04841 (2006).

19   Cui, P. *et al.* Comprehensive profiling of retroviral integration sites using target enrichment methods from historical koala samples without an assembled reference genome. *PeerJ* **4**, e1847, doi:10.7717/peerj.1847 (2016).

20   Maher, I. E. & Higgins, D. P. Altered Immune Cytokine Expression Associated with KoRV B Infection and Season in Captive Koalas. *PLoS One* **11**, e0163780, doi:10.1371/journal.pone.0163780 (2016).

21   Chappell, K. J. *et al.* Phylogenetic diversity of Koala Retrovirus within a Wild Koala Population. *J Virol*, doi:10.1128/jvi.01820-16 (2016).

22   Xu, W. & Eiden, M. V. Koala Retroviruses: Evolution and Disease Dynamics. *Annual review of virology* **2**, 119-134, doi:10.1146/annurev-virology-100114-055056 (2015).

23   Shojima, T. *et al.* Identification of a novel subgroup of Koala retrovirus from Koalas in Japanese zoos. *J Virol* **87**, 9943-9948, doi:10.1128/jvi.01385-13 (2013).

24   Martin, R. & Handasyde, K. *The koala: Natural history, conservation and management*. (University of New South Wales Press, Sydney, 1999).

25   Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).

26   Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology* **34**, 525-527, doi:10.1038/nbt.3519 (2016).

27   Tarlinton, R., Meers, J., Hanger, J. & Young, P. Real-time reverse transcriptase PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas. *Journal of General Virology* **86**, 783-787 (2005).

28   Hobbs, M. *et al.* A transcriptome resource for the koala (Phascolarctos cinereus): insights into koala retrovirus transcription and sequence diversity. *BMC genomics* **15**, 786, doi:10.1186/1471-2164-15-786 (2014).

29   Young, G. R. *et al.* Resurrection of endogenous retroviruses in antibody-deficient mice. *Nature* **491**, 774-778, doi:10.1038/nature11599 (2012).

30   Grandi, N., Cadeddu, M., Blomberg, J. & Tramontano, E. Contribution of type W human endogenous retroviruses to the human genome: characterization of HERV-W proviral insertions and processed pseudogenes. *Retrovirology* **13**, 67, doi:10.1186/s12977-016-0301-x (2016).

31   Legione, A. R. *et al.* Koala retrovirus (KoRV) genotyping analyses reveal a low prevalence of KoRV-A in Victorian koalas and an association with clinical disease. *Journal of medical microbiology*, doi:10.1099/jmm.0.000416 (2016).

32   Lutz, H. *et al.* Feline leukaemia. ABCD guidelines on prevention and management. *J Feline Med Surg* **11**, 565-574 (2009).
33   Fiebig, U., Keller, M., Moller, A., Timms, P. & Denner, J. Lack of antiviral antibody response in koalas infected with koala retroviruses (KoRV). *Virus research* **198**, 30-34, doi:10.1016/j.virusres.2015.01.002 (2015).
34   Waugh, C., Gillett, A., Polkinghorne, A. & Timms, P. Serum Antibody Response to Koala Retrovirus Antigens Varies in Free-Ranging Koalas (Phascolarctos cinereus) in Australia: Implications for Vaccine Design. *J Wildl Dis* **52**, 422-425, doi:10.7589/2015-09-257 (2016).
35   Rubin, H. The early history of tumor virology: Rous, RIF, and RAV. *Proc Natl Acad Sci U S A* **108**, 14389-14396, doi:10.1073/pnas.1108655108 (2011).
36   Nethe, M., Berkhout, B. & van der Kuyl, A. C. Retroviral superinfection resistance. *Retrovirology* **2**, 52, doi:10.1186/1742-4690-2-52 (2005).
37   Xu, W., Gorman, K., Santiago, J. C., Kluska, K. & Eiden, M. V. Genetic diversity of koala retroviral envelopes. *Viruses* **7**, 1258-1270, doi:10.3390/v7031258 (2015).
38   Borisenko, L. & Rynditch, A. V. Complete nucleotide sequences of ALV-related endogenous retroviruses available from the draft chicken genome sequence. *Folia Biol (Praha)* **50**, 136-141 (2004).
39   Hunt, H., Fadly, A., Silva, R. & Zhang, H. Survey of endogenous virus and TVB* receptor status of commercial chicken stocks supplying specific-pathogen-free eggs. *Avian Dis* **52**, 433-440 (2008).
40   Spencer, T. E., Mura, M., Gray, C. A., Griebel, P. J. & Palmarini, M. Receptor usage and fetal expression of ovine endogenous betaretroviruses: implications for coevolution of endogenous and exogenous retroviruses. *J Virol* **77**, 749-753 (2003).
41   van der Kuyl, A. C. HIV infection and HERV expression: a review. *Retrovirology* **9**, 6, doi:10.1186/1742-4690-9-6 (2012).
42   Xu, W. *et al.* An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo. *Proc Natl Acad Sci U S A* **110**, 11547-11552, doi:10.1073/pnas.1304704110 (2013).
43   Miyake, A. *et al.* Novel Feline Leukemia Virus Interference Group Based on the env Gene. *J Virol* **90**, 4832-4837, doi:10.1128/jvi.03229-15 (2016).
44   Kawamura, M. *et al.* AKT capture by feline leukemia virus. *Archives of virology* **162**, 1031-1036, doi:10.1007/s00705-016-3192-1 (2017).
45   Hofmann-Lehmann, R. *et al.* Vaccination against the feline leukaemia virus: outcome and response categories and long-term follow-up. *Vaccine* **25**, 5531-5539, doi:10.1016/j.vaccine.2006.12.022 (2007).
46   Martin, R. W. Age-specific fertility in three populations of the koala, Phascolarctos cinereus Goldfuss, in Victoria. . *Australian Wildlife Research* **8**, 275-283 (1981).

Author Contributions

RET, JM, GS, JMS, HO, NSp, FH, DT, RDE conceived and designed the study, RT wrote the paper, all authors provided critical commentary, RDE generated the figures, NSa, JF, LW, HO, NSp TD and RDE performed parts of the sample collection, sample processing and or data analysis

Author Information

The authors declare no competing financial interests

Correspondence and requests for materials should be addressed to
rachael.tarlinton@nottingham.ac.uk