The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA
School of Veterinary Medicine and Science

MRes RESEARCH PROJECT DISSERTATION

NAME: Sukanya Dudeja

STUDENT ID: 20418933

PROJECT TITLE: The Detection of Polymorphism of RecKoRV Loci within the Victorian Koala Population

PROJECT SUPERVISOR: Dr. Rachael Tarlinton
Dr. Adam Blanchard

YEAR OF PROJECT: 2022

Contents

LIST OF ABBREVIATIONS

	Term
ALV	Avian Leukosis Virus
CA	Capsid
ds-DNA	Double stranded DNA
EAV	Endogenous Avian Retrovirus
E-MuLV	Ecotropic Murine Leukemia Virus
EIAV	Equine Infectious Anemia Virus
ERV	Endogenous Retrovirus
EVE	Endogenous Viral Elements
FeLV	Feline Leukemia Virus
GaLV	Gibbon Ape Leukemia Virus
GALV SF	GALV San Francisco
HIV	Human Immunodeficiency Virus
HPG	Hervey pteropid gammaretrovirus
IFNs	Interferons
IN	Integrase
IRS PCR	Interspersed Repetitive sequence Polymerase Chain Reaction
JSRV	Jaagsiekte sheep retrovirus
KI	Kangaroo Island
KoRV	Koala Retrovirus
L-Domain	Late-budding domain
LTR	Long Terminal Repeats
MA	Matrix
MERV-L	Murine Endogenous Retrovirus L
MLR	Mount Lofty Ranges
MuLV	Murine Leukemia Virus
MMTV	Mouse Mammary Tumor Virus
MPMV	Mason -Pfizer Monkey Virus
N-MuLV	N tropic Murine Leukemia Virus
NA	Northern Australia
NC	Nucleocapsid
NSW	New South Wales
OPA	Ovine Pulmonary Adenocarcinoma
P-MuLV	Polytropic Murine Leukemia Virus
PBS	Primer-Binding Site
PhER	Phascolarctos Endogenous Retroelement

PIC	Protein Integration Complex
PPT	Polypurine Tract
Psi	Packaging signal site
Pro	Viral Protease
R	Repeat region
RAV-0	Rous sarcoma associated virus 0
RBD	Receptor Binding Domain
RecKoRV	Recombinant Koala Retrovirus
RSV	Rous Sarcoma Virus
RT	Reverse Transcriptase
SEATO	Southeast Asia Treaty Organization
SIV	Simian Immunodeficiency Virus
SSAV	Simian Sarcoma Associated Virus
ss RNA	Single stranded RNA
SU	Surface protein
TM	Transmembrane protein
U5	Untranslated 5’ region
U3	Untranslated 3’ region
WMV	Wooly Monkey Virus
X-MuLV	Xenotropic Murine Leukemia Virus
XRV	Exogenous retrovirus
ZAP	Zinc Finger Antiviral Protein

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Koala Retrovirus (KoRV) is known to be the most recently acquired endogenous retrovirus and is still settling to its genomic parasitic lifestyle. KoRV has both infectious and endogenous (integrated into the host genome) forms. One hundred percentof the Koalas in northern Australia are positive for the virus. A higher proviral copy number per cell has been observed in Northern koalas due to endogenized KoRV compared to that in the south. The Koalas in southern Australia show a variability in the prevalence rate of the exogenous virus and a lower rate of KoRV induced disease. Southern Australian koalas earlier considered to be disease free or only having the exogenous counterparts of the virus, demonstrate a defective variant of KoRV known as Recombinant KoRV (RecKoRV). RecKoRV is formed due to the recombination between Phascolarctid Endogenous Retroelement (PhER) and KoRV. The presence of RecKoRV variants particularly in the founder population on the French island calls into question the existence of KoRV free animals. The difference in the KoRV profiles between the northern and the southern animals raises the possibility that these replication defective transcripts may be interfering with full length transcripts of the replication competent KoRV. The presence of RecKoRV variants particularly in the founder population and in the Victorian animals indicate that all southern animals have these variants. The aim of this study is to look at the polymorphism of the RecKoRV loci using integration site specific PCRs to explore whether these are fixed or variable in the Victorian koala population. RT-PCRs were performed on koala samples collected from the Cape Otway region in Victoria, Australia to check for the prevalence of KoRV. RecKoRV Insertion site specific PCRs were optimized and performed on a second set of samples which were collected from different regions in Victoria Australia and were negative for KoRV to detect the polymorphism of different RecKoRV loci. The design of these PCRs proved problematic with non-specific amplification, possibly due to the repetitive nature of the LTRs of retroviruses. Techniques such as inverse PCR may be necessary to analyze the insertion site variation of these RecKoRV loci.

1. LITERATURE REVIEW

Introduction

Retroviruses are found in all vertebrates and can lead to a range of different diseases including immunodeficiencies and malignancies (Rosenberg N, 1997). They convert viral RNA into DNA which can be incorporated into the genome of the host and become a stable provirus (VM Vogt, 1997). Since exogenous retroviruses (XRVs) infect somatic cells, the integrated proviruses are non-hereditary. At times the infection of germline cells can lead to the vertical transmission of the proviruses which then become a part of the host genome and are known as Endogenous Retroviruses (ERVs) (Gifford & Tristem, 2003).

A variety of outcomes can result from retroviral infections of a host species. Some retroviral insertions as ERVs can benefit their host's physiology, as seen in the case of syncytin an ERV envelope gene that plays an essential role during human placental morphogenesis, while others like endogenous Avian Leukosis Virus (en ALV) and endogenous Jaagsiekte sheep retrovirus are known to protect their hosts from infection by their exogenous counterparts. Recombination between enFeLV and exFeLV can lead to generation of new strains of the FeLV with altered properties and can cause immunosuppression and neoplasia in felids.

The outcomes of Koala retrovirus infection depend on the complexity and combination of various environmental, host and viral factors (Kinney & Pye, 2016). KoRV can lead to immunosuppression and induce neoplastic conditions in koalas (*Phascolarctos cinereus*). KoRV can be problematic for both wild koalas in Australia and captive koala populations around the world. KoRV associated malignancies pose a major threat to koala health and conservation (Enamul et al., 2020).

1. Retroviruses

The family *Retroviridae* are a family of enveloped RNA viruses known for their replicative strategy of retrotranscribing their RNA genome into DNA and thus integrating proviral sequences into the chromosomes of infected hosts. These sequences are 7-12 kb in size and are non-segmented, single stranded, positively polarized and linear. Specific cis or trans- active retroelements regulate various aspects of the retroviral replication (Kurth & Bannert, 2010).

Comprising of both pathogenic and non-pathogenic viruses, Retroviruses hold the capacity to infect mammals. The transmission of these viruses can be either intraspecies or trans-species based on the types of cells they infect, and the type of cell surface receptors found on those cells. All the retroviruses comprise of two full length transcripts that are linked to each other through H-bonds (hydrogen bonds) and possess a 5' cap and a 3' polyadenylated (A) tail. The genomic RNA lacks the terminal regions of the transcription driving 5' LTR and the 3' LTR which are found in the proviral form (Kurth & Bannert, 2010).

Screening of retroviruses from a wide variety of host taxa has suggested that the retroviruses emerged during the early paleozoic era along with their retroviral hosts approximately 460-550 million years ago (Hayward, 2017). Since infections with retroviruses are restricted to vertebrates, retroviral origin during the paleozoic era means that the retroviruses must have evolved from marine vertebrates as life on land during that time was extremely limited (Hayward, 2017).

It was after the discovery of the retroviruses, that they were classified and the difference in the structure of the simple and the complex retroviruses with reference to the presence of additional polyproteins in the latter was defined. Typically, all the exogenous retroviruses are spherical particles, about 100-150 nm in diameter (Kurth & Bannert, 2010).

1.1 Classification

At first, classified into 4 different groups, retroviruses were put into categories from type A to type D based on the shape of the core as observed using electron microscopy. Initially observed to be immature capsids, type A particles emerge as thick-shelled, hollow intracellular structures while type B particles possess round, non-central inner cores (Goff, 2007). Type A particles are no longer classified as a separate morphological class, while type B and D particles are typical of beta retroviruses (e.g., mouse mammary tumor virus (MMTV), Mason-Pfizer monkey virus (MPMV), Alpha, gamma, epsilon, and delta retroviruses have type C particles. An exception to this classification is the lentiviruses and spumaviruses having unique core types. Cylindrical or conical cores are observed in lentiviruses while spumaviruses have central, uncondensed cores and a spiked surface (Goff, 2007).

Taking into account the core structures, the retroviruses have two major morphogenetic assembly modes, the cores of type B and type-D viruses, assemble in the cytoplasm while the viruses of type C morphology (e.g., avian leukemia virus (ALV), murine leukemia virus (MuLV)) assemble at the cellular membrane concomitant with the budding process.

Alternatively, the *Retroviridae* family can be classified based on their genome into two classes, with simple (alpha, beta, gamma, and epsilon retroviruses) or complex genomes (lenti-, delta-, and spumaviruses). Simple retroviral genomes contain genes for the structural and functional polyproteins Gag, Pro, Pol, and Env, while the complex retroviral genomes produce additional proteins and RNAs with diverse virulence-enhancing functions. Complex retroviruses usually code for additional regulatory proteins derived from multiple spliced messages while simple retroviruses usually carry only elementary information (Escalera-Zamudio & Greenwood, 2016).

On the basis of evolutionary relatedness, the family *Retroviridae* comes under the order *Ortervirales* and comprises of two subfamilies namely the *Orthoretrovirinae* and *Spumaretrovirinae* with eleven genera and sixty-eight species. Out of all the genera, the spumaviruses (foamy viruses) are the most basal retroviruses with fewer interspecies transmission events (Hayward, 2017). Also, five out of the eleven genera with oncogenic potential are referred to as oncoviruses.

Retroviruses are also known to exist in endogenous and exogenous forms. The exogenous retroviral forms maintain their infectivity in the population and are transmitted horizontally. While the endogenous retroviruses (ERVs) may become fixed in the population i.e., due to integration of these ERVs into the germ line of their respective hosts, they have become an integral part of their genome. ERVs are different from exogenous retroviruses (XRVs) in the sense that they provide an extensive diversity of retroviral sequences that have accumulated over a range of millions of years due to interaction between the retroviruses and their vertebrate hosts (Escalera-Zamudio & Greenwood, 2016).

1.2 Structure

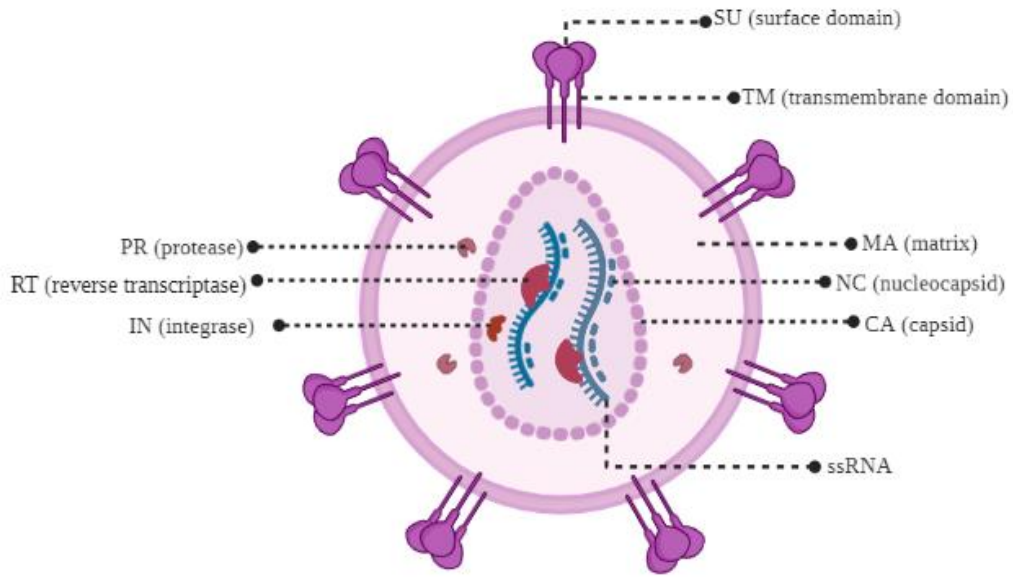


Figure 1 General Structure of Retrovirus comprising of four functional genes namely the *gag, pro, pol,* and *env*. The *gag* gene encodes the capsid (CA), matrix (MA) and nucleocapsid (NC). The *pro* gene encodes for the viral enzymes protease (PR) and integrase (IN). Env protein consists of the surface unit (SU) and transmembrane unit (SU). The env domain also has glycoprotein protrusions known as the envelope spike (created on Biorender).

A generalized DNA provirus has four functional genes namely *gag* (group antigen glycoprotein), *pro* (Protease), *pol* (polymerase, multifunction protein encoding reverse transcriptase), and *env* (Envelope) (Figure 1). The structural components of the viral core namely capsid (CA), matrix (MA), nucleocapsid (NC) are encoded by the *gag*. The capsid protein protects the core of the virus while the nucleocapsid protein protects the RNA genome from degradation. The length of the *gag* gene ranges from less than 1200 bp up to 2000 bp (Kurth & Bannert, 2010).

The *pro* gene, however, is approximately 700 bp in length and encodes the PR (protease) viral enzyme which is responsible for cleaving viral polyproteins into their separate subunits. The RT (reverse transcriptase) and IN (integrase) enzymes are encoded by the *pol* gene, the length of which is in the range of 2500 to 3500 bp. Of these two enzymes encoded by *pol* gene, the former catalyzes the transcription of viral RNA into DNA while the latter is responsible for the integration of the viral cDNA into host DNA. The length of the *env* gene varies from 1500 to 3000 bp (Goff, 2007). Furthermore, the Env protein encodes the

viral glycoprotein subunits SU (surface unit) and TM (transmembrane unit) which are crucial for the viral attachment and penetration (Escalera-Zamudio & Greenwood, 2016).

LTR (long terminal repeat) sequences flank the RNA form of the retroviral genome. The untranslated 5'(U5), repeat (R), and untranslated 3'(U3) are three distinct regions derived from the intronic end segments of the extracellular RNA genome of each LTR sequence. An unintegrated retrovirus has U5 and U3 regions at their respective 5' and 3' ends as the duplication of these regions takes place during reverse transcription (Gifford & Tristem, 2003). Quite a few elements involved in the transcription of integrated viruses are present in the LTRs. The same regulatory structures are present in the LTRs at 5'and 3' ends but most of the retroviruses use the 5' LTR for transcription initiation and 3' LTR for termination (Figures 2 and 3). Post integration the 5' LTR has an observed structure 5'-U3-R-U5-3'. The "att site" present at the far 5' end is used during integration as an attachment site. The key role of transcription initiation is played by the U3 region which is the promoter region of the retrovirus. A 5' cap is added to a particular site at the 3' end of U3 when the retroviral DNA is transcribed by the host while the remaining part of the 5' U3 is left untranscribed (Kurth & Bannert, 2010).

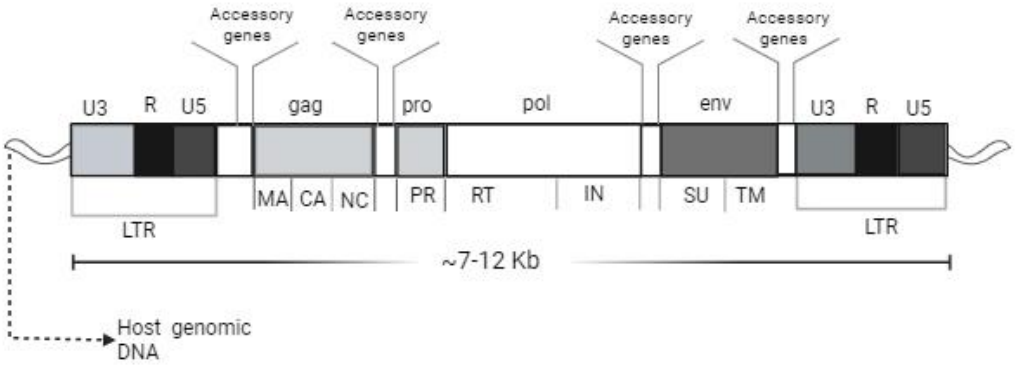


Figure 2 Integrated DNA provirus: A generalized replication competent DNA provirus comprises of four functional genes namely the *gag, pro, pol* and *env*. The structural components encoded by the *gag* gene are matrix (MA), capsid (CA) and nucleocapsid (NC). The *pro* gene encodes for protease (PR) while reverse transcriptase (RT) and integrase (IN) enzymes are encoded by the *pol* gene. The Env protein encodes for the surface (SU) and the transmembrane (TM) glycoproteins, essential for viral attachment and entry into the host cell. The complex retroviruses have certain additional genes known as accessory genes. The Long terminal repeats or LTRs flank the DNA provirus after the process of reverse transcription and are generated from the non-coding ends of extracellular RNA. The LTR is composed of U3, R and U5 regions (Adapted from Gifford & Tristem, 2003).

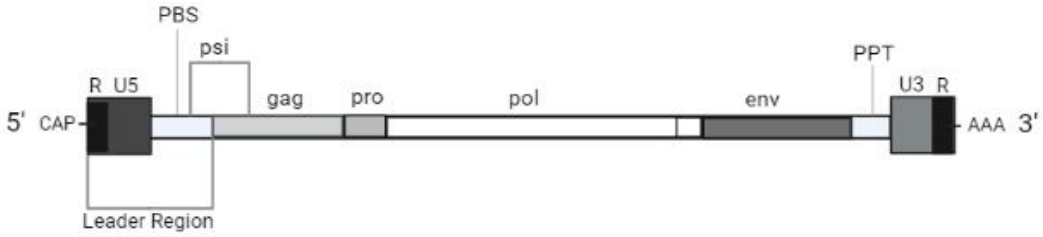


Figure 3 Viral genomic RNA: The tRNA attaches to the Primer binding site (PBS) towards the 5' end of RNA. The sequences necessary for the viral replication are PBS, (ψ) packaging sequences, polypurine tract (PPT), polyadenylation signal and poly(A)tail (AAA) (Adapted from Gifford & Tristem, 2003 and created on biorender).

The region U5 incorporates the Primer Binding Site (PBS) and the packaging signal (Psi) sites. The PBS site binds to a complementary cellular transfer RNA (tRNA) during reverse transcription while the Psi sites enhance the efficiency of RNA packaging into virions (Kurth & Bannert, 2010). The 3' LTR is primarily known to be involved in transcription termination and has the same 5'-U3-R-U5-3' structure as the 5' LTR after integration. The terminus of the 3' LTR has an att site, mirroring the 5' LTR (Kurth & Bannert, 2010).

1.3 Exogenous Viral Life Cycle

Retroviruses reproduce by manipulation of the host transcriptional and translational machinery. Retroviral replication has two phases namely the early and the late phase. The early phase starts when the virion attaches itself to the host's cell surface and it ends with the proviral DNA integrating itself into the host genome. The beginning of the late phase is marked by the initiation of viral transcription while the completion is when the progeny viruses are released (Figure.4) (Mothes & Uchil, 2010).

The initiation of the retroviral life cycle is by the interaction of the SU subunit of the Env protein interactions with a receptor on the cell surface (Goff, 2007). For a retrovirus to infect a cell, there needs to be a transference of the retroviral genome across its membrane and that of the cell (Blumenthal et al., 2012). This can be accomplished by either a pH independent pathway *i.e.*, by fusion of the viral and the cell surface membranes in response to receptor binding, or via a pH-dependent pathway *i.e.*, via receptor-mediated endocytosis (Goff, 2007) (Mothes & Uchil, 2010). Thus, the fusion peptide region of the TM protein is coupled with the host

cell membrane which generates changes in the host cell membrane and this peptide induces changes in the cell membrane of the host cell permitting the virus to enter the target cell. After the fusion of the virus with the host cell, uncoating of the viral core takes place and the retroviral genome is reverse transcribed into DNA.

At a point between the fusion of the viral membranes and the integration into the host chromosomal DNA, uncoating of the viral capsids and the reverse transcription of the retroviral RNA takes place. Although very little is known about the process of uncoating of the viral capsids, there are some studies which take into consideration the uncoating in the cytoplasm being linked to reverse transcription (Hulme et al., 2011; Mamede et al., 2017); while others consider the uncoating process occurs at the nuclear pore (Arhel et al., 2007).

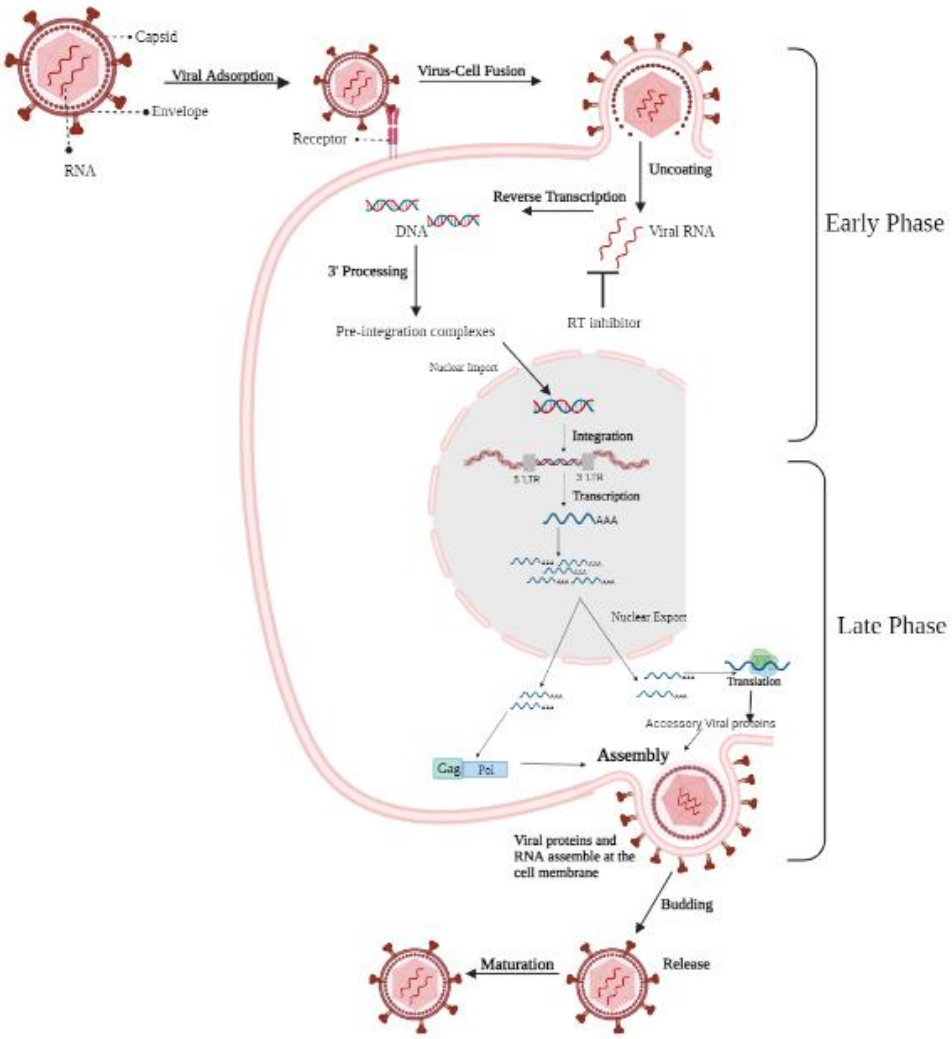
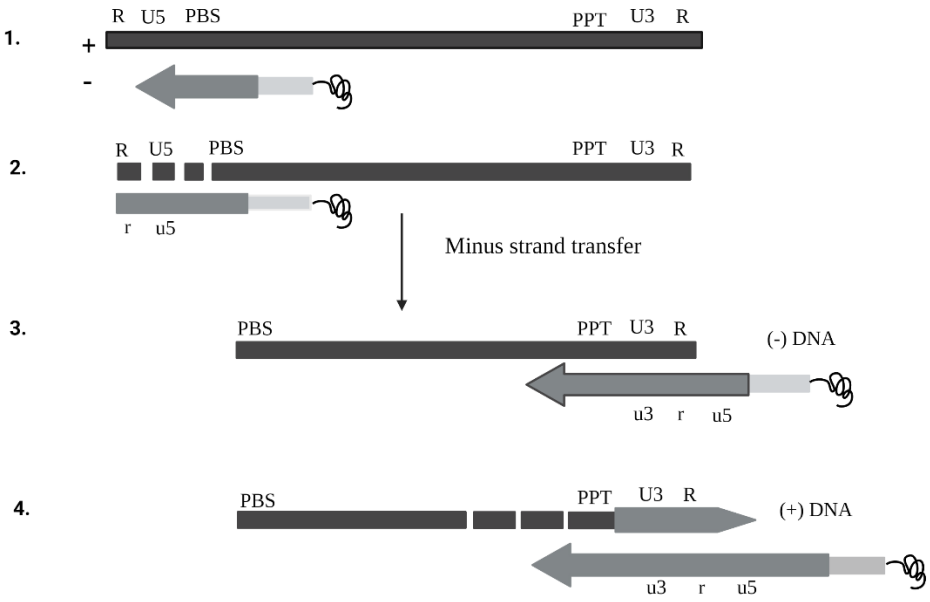


Figure 4 Retroviral replication cycle: After the interaction of the virus with the receptor, viral and host membrane fusion initiates retroviral entry in the host viral membrane, reverse transcription of RNA into ds DNA is followed by formation of the Pre integration complex (PIC), nuclear entry of PIC and viral DNA integration into the host genome is followed by transcription which is succeeded by nuclear export of mRNA, translation and viral assembly for production of progeny viruses (created on Biorender).

The viral particles have duplicate copies of positive single stranded (+ss) RNA, the LTRs of which house sequence elements holding importance in the process of reverse transcription. Reverse transcription begins when the tRNA anneals to the complementary sequence in the viral RNA particularly the PBS at the 5’ end, thus synthesizing the minus strand DNA, also known as the minus strand strong stop DNA (Figure 5, step 1). This sequence consists of the R and U5 sequences. The minus strand strong stop DNA is transferred to the 3’ end of the viral genome by RNase H (Figure 5, step 2). Post strand transfer, the remaining RNA sequence is copied by minus strand extension (Figure 5, step 3). Afterwards, the RNase H degrades major components of the RNA genome except for the PPT (Figure 5, step 4). The PPT primes the plus strand synthesis by using the minus strand as a template. The extension of the plus strand continues to the 3’ end until terminated by adenosine residues (Figure 5, step 5). The degradation of the tRNA primer from the 5’ end of the minus strand is initiated by RNase H (Figure 5, step 6). This is followed by the transfer of the plus strand strong stop DNA to the PBS of the 3’ end of the minus strand DNA (Figure 5, step 7). Extension of the plus and minus strand thus leads to the production of the double stranded(ds)-cDNA with a copy of the LTR at each end (Figure 5, step 8) (Engelman, 2010).



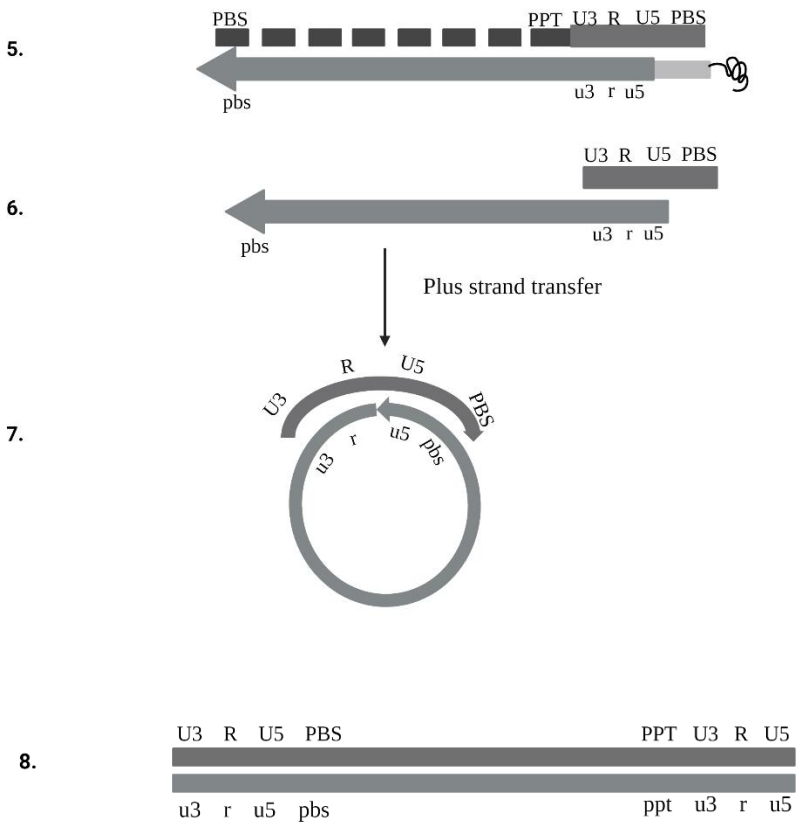


Figure 5 Mechanism of reverse transcription: The process of reverse transcription starts with:(1)the annealing of tRNA to the PBS at the 5'end (2)the minus strand strong stop DNA including R and U5 are synthesized by RT (3) the transfer of minus strand moves the tRNA to anneal to the 3' end of the viral genome (4) the elongation of minus strand and the continuation of degradation of plus RNA strand by RNase H (5) the copying of U3, U5 and R regions in minus-strand DNA by RT and reconstitution of PBS (6) the removal of tRNA and PPT primers from minus and plus strand DNAs by RNase H (7) Plus strand transfer is initiated by the annealing of complementary PBS sequence of the positive ssDNA and minus strand DNA (8) Elongation of the minus and plus strand DNA forming dsDNA with LTRs (Adapted from (Iwatani et al, 2007) and created on Biorender).

Prior to integration, the reverse transcribed retroviral copy forms a pre-integration complex (PIC), with naked viral core and discrete host proteins which are then imported into the nucleus (Kurth & Bannert, 2010). Integration of the retroviral genome into the host genome is a three-step process involving end processing, joining, and gap repair. Removal of a dinucleotide from the 3' ends of the ds-cDNA marks the end processing step (Hindmarsh & Leis, 1999). The hydroxyl groups are exposed due to the end processing are used by IN to cut the host DNA

and are joined to the viral DNA as a part of the joining step of the integration phase. The phosphodiester bonds in the host DNA are attacked by hydroxyl groups and a new phosphodiester bond formation between the 3' end of the virus and the host DNA occurs by displacement of one of the host DNA bonds (Goff, 2007). The ultimate step marking the end of the integration phase is the gap repair step which seals the gap between the att site at the 5' end of the retroviral and the host DNA (Goff, 2007).

In the late stage of the virus life cycle replication is mediated by the host instead of the retroviral enzymes (Goff, 2007). Factors in the promoter and enhancer region of the proviral U3 region of the LTR initiate transcription (Lenasi et al., 2010).The process of transcription uses the entire length of the retroviral genome for the generation of a single pre messenger RNA (mRNA) transcript (Goff, 2007). The transcript can be subjected to three fates. First, to serve as the genome for progeny viruses, there is a direct export of a fraction of transcripts from the nucleus. Some other transcripts when exported to the cytoplasm are translated to form Gag or Gag-Pol polyproteins. The remaining transcripts undergo splicing in the nucleus and are translated to form Env and in complex retroviruses they form accessory proteins (Goff, 2007). The Gag precursor protein is translated from full-length mRNA transcripts in the cell cytoplasm and later cleaved by Pro. Pro and Pol are also translated from an entire length of the mRNA transcripts in the cytoplasm and are later cleaved by Pro (Goff, 2007). The subunits of Gag polyprotein coordinate the assembly of new viruses and the Gag polyprotein precursor is adequate for the assembly of immature virus- like particles (Hinz et al., 2010) . The viral particles are then released by budding through the plasma membrane of the cell. The viruses of morphology type C assemble in the plasma membrane while the type B and type D viruses assemble in the cytoplasm and are transported to the plasma membrane, where the formation of envelope around the viral core is followed by their release (Goff, 2013). During packaging, the RNA genome is incorporated into the retroviral particle and the interaction between the psi elements the 5'UTR and NC subunit coordinates this process (Goff, 2007). The Gag domain coordinates the release of the retroviral particle from the host plasma membrane (Pincetic & Leis, 2009). After these retroviruses are released from the host cell, immature retroviruses undergo changes in their conformation to produce mature particles. The maturation of viruses makes them structurally distinct and confers stability. Thus, enhancing their ability to cause infection in the host cells (Goff, 2007).

1.4 Endogenous Retroviruses

Over the years it has become evident that occasionally, viruses mark their presence in the host genomes by leaving behind viral elements known as Endogenous Viral elements (EVE) which can prove to be advantageous for discerning their histories; the record for which is particularly rich for the retroviruses (Hayward, 2017).

As retroviruses do not usually lyse their target cell, integration allows a long-term association between cell and virus. If the infected cell is a germ cell, colonization of the germ line by the virus is possible. Numerous studies indicate that such events have occurred multiple times during evolution. Such inherited proviruses are called ERVs, and they provide a 'fossil record' of past retroviral infections dating back many millions of years, which suggests a never-ending stream of retroviral challenges for vertebrates. With ERVs (endogenous retroviruses), the vertical transmission of the retroviral sequences is seen to an extent that they become a part of the host germline within which they not just persist and replicate but as well evolve. The probable fate for most of the ERV lineages is extinction but some ERVs are capable of a certain level of expression and replication even after a period of a million years within their host genome (Gifford & Tristem, 2003). Although the evolution and the activity of ERVs in their hosts can continue through either adaptation to vertical transmission or by gradual changes in their genome though mutation, ERV insertions are unlikely to directly represent their retroviral ancestors (Greenwood et al., 2018).

Aggregation of mutations from one ERV generation to the next is irreversible. All the ERVs become defective eventually due to their inability to express any gene products which leads to the extinction of their lineage. Gradually, these defunct lineages are overridden by accrued mutations and degraded into the unidentifiable sequences of DNA known as junk DNA (Gifford & Tristem, 2003). However, in some cases ERVs are co-opted in their hosts as seen in case of the two ERV *env* genes syncytin-1 and syncytin-2, the fusogenic potential of which plays a critical role in the placental development, establishment, and regulation of feto-maternal immune tolerance in humans (Lokossou et al., 2014).

Degraded ERV insertions unable to code for the functional enzymes can remain in a replicative state if their regulatory sequences like PBS, PPT, psi, and LTR regions are undamaged and if the replicative proteins are provided by fully or partially functional ERVs in the same cells or infection causing exogenous viruses. Infection causing exogenous viruses may also be released periodically by replication competent ERV sequences (Gifford & Tristem, 2003).

1.5 Endogenous Retrovirus Lifecycle

ERVs proliferate in genomes by two potential mechanisms, namely the germline infection model and the retro-transposition model. These two models shed light on the mechanism behind the ERV proliferation in the genome. The former explains the direct reinsertion of the endogenized virus into the germline during an active reinfection phase while the latter employs the master gene theory to explain that a few active viral elements might be functional to an extent that they can yield independent copies in genomes (Magiorkinis et al., 2012). The recombination of older endogenous retroelements and the degradation of the invading retroviral genome occurs early at the time of the retroviral invasion with the disruption of the invading retrovirus. Alongside these events remobilization of the recombination retroelement within the host of the genome occurs (Löber et al., 2018). Retro-transposition or reinfection of germ cells leads to an increase in the ERV copy numbers. Repression by host mechanisms and mutations of ERV insertions during host cell replication results in a decline in the rate of amplification over time. Initial post entry endogenous retroviral integration is repressed by many mechanisms as seen in case of the Murine Leukemia Virus (Gifford & Tristem, 2003).

The ERV insertion frequency is influenced by the host selection pressures and, in this process, a part of their population is lost while others reach fixation. The change in the frequency of the ERV insertions is impacted by numerous factors including changes in the host population size, the genomic locations of the ERV inserts, the expression levels of the ERV inserts, and the fraction of coding DNA within the host genome (Gifford & Tristem, 2003). In the due course of time when the rate of amplification dwindles to zero, the ERV lineage is inactivated. Recombination events between the two full length retroviral elements in the host DNA leads to generation of solo LTRs (Hughes & Coffin, 2004). Even after the loss of their infectivity, these defunct retroviruses are still recognizable for millions of years (Brown et al., 2014).

There are also a large number of other non-retroviral transposable elements that are found in mammalian genomes both interspersed and as single copy endogenized genes. Some of these transposons also use reverse transcriptase to replicate. Long interspersed Element-1(LINE-1) or simply L1 is the most common self- replicating human transposon class. It replicates through transcription and reverse transcription, exhibits unfettered retrotransposition and generates new copies at different genomic locations. The overexpression of L1 RNA and protein causes apoptosis, tumor progression and DNA damage and repair. Mammalian cells use anti transposon restriction factors which are also anti-retroviral in nature in order to protect themselves from the deleterious effects of these transposable elements (Goodier, 2016; Han & Boeke, 2005).

1.6 Antiviral restriction Factors

Post endogenization, the functionality of ERVs is compromised and they are rendered defective due to progressive mutational degeneration. In certain cases however, the ERV regulatory sequences can be reselected for cellular functions by the hosts and some of them can inhibit XRV infections by acting as antiretroviral restriction factors (Kozak, 2015).

Antiretroviral restriction factors are cellular proteins which limit viral replication by targeting specific stages of their life cycle. Components of the innate immune response to viruses, restriction factors slow the viral replication in the respective host until the adaptive immune response starts to reduce or eliminate infections (Chemudupati et al., 2019).

1.6.1 APOBEC

APOBEC was the first evolutionarily conserved protein belonging to the family of cystidine aminases described. The APOBEC3 proteins are vertebrate specific and are synthesized in all placental mammals. Rodents have a single APOBEC3 member in the family of APOBEC proteins while there are seven APOBECs found in humans (Goodier, 2016). The ability of APOBEC3 to interfere with retroviral replication was discovered in 2002 when APOBEC3G was reported to block the replication of a class of HIV mutants with defects in the *vif* (virion infectivity factor) gene (an accessory gene of HIV-1) (Sheehy et al., 2002).

It was discovered that in the virions of Vif deficient HIV, the APOBEC3G proteins target the deamination of cytosines to uracils during the first strand synthesis of the HIV cDNA (Goodier, 2016). Cytosine transition mutations are observed as the antisense ssDNA intermediate formed is complementary to the viral RNA. As a result of transition mutations, the guanosines (G) are replaced with adenosines (A) in the positive strand of the proviral DNA (Nair & Rein, 2014). LINE-1 retrotransposition is inhibited by all APOBEC3 proteins, while the highest inhibition is shown by the A3A and A3B members of the human APOBEC3 family. Catalytically inactive APOBEC3s still cause the inhibition of non- LTR retrotransposons (Goodier, 2016).

1.6.2 SAMHD1

Human sterile alpha motif and HD- domain containing protein 1 or SAMHD1 was first identified as an antiretroviral restriction factor for HIV-1 in dendritic cells,

macrophages and in resting CD4+ T cells. SAMHD1 is the only enzyme in mammalian cells with dNTPase activity. Ubiquitously expressed in all cell types, it degrades dNTPs to deoxynucleoside and inorganic phosphates, SAMHD1 is essential for maintaining a balanced intracellular concentration of dNTPs (Chemudupati et al., 2019).

1.6.3 IFITM3

IFN-induced transmembrane protein 3 or IFITM3 is a protein that associates with membranes via transmembrane domains. The ability of IFITM3 to block viral entry is the primarily described mechanism of its antiretroviral activity. The IFITM3 protein is known to alter membranes in order to inhibit fusion of viruses. It has been suggested that IFITM3 can reduce the fluidity of membranes and the amphipathicity of the amphipathic helix in IFITM3 is required for obstruction of viral protein mediated membrane fusion. Stabilization or inhibition of the local membrane curvature by the amphipathic helix is one of the possible reasons for it to disfavour fusion. In the cells infected by viruses prior to IFITM3 induction, the upregulated IFITM3 is incorporated into the membrane envelope of budding virions which most likely reduces the infectivity of the virions. This mechanism of incorporation of IFITM3 into the virion to decrease infectivity was first documented in HIV-1 infections and was subsequently established in many more viruses which otherwise show resistance to IFITM3 inhibition in endosomes (Chemudupati et al., 2019).

1.6.4 Tetherin

Tetherin was named for its ability to inhibit HIV-1 release via a tethering mechanism. The N-terminal transmembrane domain and a C-terminal glycosylphosphotydilinositol (GPI) anchor are utilized by tetherin for attachment to the plasma membrane and the simultaneous incorporation into the membrane of HIV-1 particles by physical anchorage of the virions to the plasma membrane. Although tethrin dimerization is not required for restriction of all viruses that are inhibited by tethrin, dimerization is indispensable for its activity against HIV-1. The conserved coiled-coil domain within tetherin proteins is known to allow flexibility when the viruses are being tethered to the plasma membrane while the disulphide bonds between tethers is known to sustain the stability of tethers. Although lacking structural homology to the human homolog, murine tetherin impedes the release of both HIV-1 and Moloney MuLV. It has been demonstrated that tetherin shows little to no inhibitory response against HIV-1 in the presence of the viral accessory protein Vpu (viral protein U), as the co-localization of Vpu

with tetherin facilitates its removal by degradation through endosomal trafficking and its removal from plasma membranes (Chemudupati et al., 2019).

1.6.5 TRIM

TRIM, also known as Tripartite motif, is a conserved family of proteins characterized by a RING domain, one or two B-boxes and a predicted coiled- coil region. By inhibition of various stages of the retroviral life cycle, many TRIM proteins interfere with infectious retroviruses. The role of a few of them as antiviral restriction factors are described below.

The most well characterized cellular restriction factor, TRIM5α was first identified due to its potential to confer resistance and intrinsic immunity to HIV-1 infection in macaque cells. This cellular restriction factor blockades the early stages of infection post retroviral entry into the host cell. TRIM5α exists in the cytoplasm in diffused form or as aggregates known as cytoplasmic bodies. It inhibits the activity of wide range of retroviruses in a species specific manner (Rahm, 2016). TRIM5α obstructs the incoming virions by specific recognition of the CA motifs which have been assembled into the viral cores and stimulates the process of their partial uncoating. For TRIM5α to act as a restriction factor, all its protein domains have to be functional. TRIM5α orthologs from different species cause restriction of a wide range of retroviruses including HIV-1, HIV-2, SIV, MuLV, EIAV and FIV (Kurth & Bannert, 2010).

Another TRIM protein closely related to TRIM5 is TRIM22. A higher TRIM22 expression was observed in patients with lower viral load of HIV-1, which demonstrated its role as an antiviral restriction factor. TRIM22 suppresses retroviral transcription by interfering with Gag protein trafficking to the plasma membrane and binding of Sp1 to the LTRs. Multiple mechanisms are known by which TRIM22, an intracellular antibody receptor plays an important role in antiviral response, one of the mechanisms of its function is by neutralization of infection by targeting viral proteins for degradation by proteasomes and ATPase. Identified as a regulator of ZAP (Zinc finger antiviral protein), TRIM25 acts as a key regulator of ZAP’s antiviral activity (Van Gent et al., 2018).

1.6.6 ZAP

First identified in inhibition of MuLV, Zinc finger antiviral protein or ZAP is a broad-spectrum antiviral protein. There are alternatively spliced long, and short isoforms of ZAP known as ZAP-L and ZAP-S in humans. ZAP-L contains a poly-ADP-ribose polymerase-like domain or PARP-like domain which is not

enzymatically active but is necessary for restriction of alphaviruses, while no PARP like domain is found in ZAP-S. Both ZAP-S and L are induced by interferons (IFNs), however IFN treatment and virus infection increases ZAP-S more in comparison to ZAP-L. Despite their differences both ZAP-S and ZAP-L are known to inhibit the activity of retroviruses (Chemudupati et al., 2019).

1.7 Retroviral Pathogenesis and Interactions between the Exogenous and Endogenous Retroviruses

1.7.1 Avian leukosis virus (ALV)

The Avian leukosis viruses (ALV) are alpharetroviruses causing infections primarily in chickens, however they are capable of infecting pheasants, partridges, and quail. The induction of tumors by slow transforming ALVs is at a later stage after infection in the chicken. Erythroblastosis and lymphoid leukosis are tumors induced by insertional mutagenesis. The LTRs along with the ALV genome are incorporated either upstream or downstream or within host cellular proto-oncogenes. The activation of proto-oncogenes by the promotor or enhancer sequences of LTRs leads to an anomalous expression of the oncogene and neoplasia (Payne, 1998).

Lymphoid leukosis is the most common neoplasm induced by slow transforming ALVs. The *c-myc* gene in the B-cell is activated in the disease. Tumor progression and metastasis occurs by activation of other cellular oncogenes during B cell lymphomagenesis. In a few weeks post infection by ALV, the usual B cell components in the transformed lymphoid follicles (also known as the focal preneoplastic hyperplasia) are replaced with proliferating B lymphoblasts. The progression of a few transformed follicles in chickens over a time span of several months leads to neoplasia with metastasis followed by death. High doses of slow transforming ALVs with a latency period shorter than that of lymphoid leukosis leads to induction of erythroid leukosis. Although this form is less common, it leads to *c-erbB* gene activation in erythroid cells. Another sporadic disease manifestation in adult birds by slow transforming ALV is the disease called myeloblastosis which leads to transformation of myeloid cells and the occurrence of severe leukemia (Payne, 1998).

Acutely transforming ALV have at variable locations in their genomes one to two viral oncogenes. The oncogene carrying virions exhibit an inability to replicate due to genetic deletions in their structural genes. These acutely transforming and defective oncogenic virions need non-defective ALV helper viruses to replicate. The uncontrolled expression of viral oncogenes and their abnormal products result

in neoplasia by bringing about changes in the normal regulatory processes of cell growth and differentiation (Payne, 1998).

The ALV subgroups A-D and J are exogeneous while the subgroups E, F and I are endogenous. The endogenous ALV have defective LTRs with functionally deficient enhancer and promoter regions (Dudley, 2011). The ev loci are found within the genome of chickens (Boulliou et al., 1992). These ERVs are mostly defective due to the lack of complete retroviral genes required for the production of infectious virions (Payne, 1998). However, some of these ERVs can produce a mature infectious virion giving rise to ALV-E an archetype of which is the Rous sarcoma virus (RSV) – associated virus (0) (RAV-0). In an experiment, an increase in the identified endogenous ALV integration sites and the loci in domesticated chickens was observed compared to the red jungle fowl which suggested that ALV-E duplication corresponds with domestication (Chiu & VandeWoude, 2021).

At least three additional families of retrotransposons are found in chickens. The endogenous avian retrovirus (EAV) family lacks regions encoding polymerase and env genes important for provirus replication and transmission. The EAV family has large deletions in their retroviral genomes. The LTRs of EAVs are diverse apparently due to the degeneration of the sequences over a period of time due to their ancient integrations into the genome. Derived from recent integrations are two other ERV families called the ART-CH and EAV-HP closely related to ALV-J. These retroviral integrations are defective since they lack regions encoding Pol. Although the packaging signals are present in these defective transcripts, they need a helper virus to be packaged into a viral genome for replication (Hunt et al., 2008).

Host age and the ERV genome structure determine the interaction between the endogenous and the exogenous ALV. The endogenous ALVs are known to protect against the disease caused by exogenous ALV infection (Chiu & VandeWoude, 2021). The ALV-E includes a series of at least 23 loci consisting of both non-defective and defective retroviral inserts. The non-defective retroviral insertion loci can produce infectious virus particles and horizontally transmit without the use of a helper virus. The cellular receptors for ALV are encoded by a genetic complex TVB* which comprises of three alleles. The allele TVB*S1 encodes for the receptor of ALV-E. Some genotypes of the commercial chicken lines retain the receptor TVB*S1 but either fail to produce either replication competent ALV-E inserts or express envelope protein from defective ALV-E inserts. The glycoprotein encoded by the envelope gene of the defective inserts inhibit reinfection caused by the replication competent ALV-E by blocking the TVB*S1 receptor (Hunt et al., 2008).

1.7.2 Jaagsiekte sheep retrovirus (JSRV)

The Jaagsiekte Sheep Retrovirus (JSRV) is the causative agent of Ovine pulmonary adenocarcinoma (OPA) a transmissible lung cancer prevalent in a high percentage of sheep, only 30% of which develop OPA lesions (Chiu & VandeWoude, 2021; Sharp & Demartini, 2003). JSRV is an acute transforming retrovirus, and its oncogene is found in the Env structural protein. The cytoplasmic tail of the TM domain contains a YXXM motif that is conserved between all transforming JSRV strains (Hofacre & Fan, 2010). JSRV is known to coexist with highly related endogenized JSRV which were co-opted by the host due to the protection offered against the XRVs (Armezzani et al., 2014). The ovine endogenous retroviruses lack the ScaI enzyme restriction site in the *gag* gene which is found in all oncogenic JSRVs and can be used to differentiate between the exogeneous (exJSRV) and the endogenous JSRV(enJSRV) (Shi et al., 2021). In a survey to check JSRV infection status in sheep from China, one strain of JSRV found had "LHMKYXXM" motif but no ScaI enzyme site, although it was known to be closely related phylogenetically to the strain identified in the USA, it is still suspected that it is an exJSRV strain due to the presence of the "LHMKYXXM" motif (Shi et al., 2021).

Approximately 15-20 copies of the ERVs related to the enJSRVs are present in the genome of the sheep. Multiple copies of enJSRV have been observed at multiple loci on chromosome 6 and chromosome 9 of both wild and domesticated sheep (Carlson et al., 2003). Of the three enJSRV loci sequenced at the amino acid level, the sequence similarity is 94 to 95 % in Gag, 95% to 99% in pro, 99% in Pol and 92% in env to that of exJSRV (Spencer et al., 2003). Almost 90-98% of the amino acids in both JSRV and enJSRV are similar (Palmarini et al., 2000) across all their genome except for the U3 region in enJSRV which shares only 56 % of sequence identity with its exogenous counterpart due to a difference in length.

The clara cells and the alveolar type II cells in the lung are readily accessible to exJSRV in early infection, due to the establishment of the infection through the oral/respiratory route. Thus exJSRV first infects and replicates in the lung epithelial cells (Archer et al., 2012). While the enJSRVs are abundantly expressed in the sheep reproductive organs including epithelia of the cervix, uterus and oviduct (Armezzani et al., 2014). The difference in the LTRs of JSRV and enJSRV determines their expression in the different tissues with enJSRV LTRs being most active in the reproductive tract particularly the uterus; the exJSRV LTRs are more active in the lung cells (Carlson et al., 2003).

Recombination events have not been documented between the enJSRV and the exJSRV and the interactions described between the enJSRV and the exJSRV are antagonistic in nature (Chiu & VandeWoude, 2021). The inhibition of the JSRV infection by enJSRV is through receptor interference. Both the exJSRVs and the

enJSRV use Hyaluronidase 2 as the cellular receptor. The saturation of the Hyaluronidase 2 receptor by the enJSRVs env decreases the availability of the receptors at the cell surface, thereby inhibiting the entry of the exJSRV into the target cells. Some enJSRVs, e.g., the enJS56A1 employ the JSRV late restriction mechanism (JLR) which blockades the post integration steps of the JSRV cycle by impairing the transport and the exit of the virus. The trans-dominant enJSRV Gag proteins form aggregates with the JSRV gag and are subsequently steered to the proteosome for degradation (Mura et al., 2004).

1.7.3 Murine leukemia virus (MuLV)

Murine Leukemia Viruses (MuLVs) are blood borne pathogens capable of causing disease via horizontal transfer in populations of both wild and laboratory mice. MuLV infected mouse species show a global distribution pattern. The MuLV host range is affected by polymorphism in the sequences of the host cell receptors and the receptor variants are responsible for different viral restriction patterns in different species of mice (Kozak, 2015).

The pool of cell surface receptor molecules during infection with murine leukemia virus (MuLV) are saturated by the continuously synthesized surface domain of env protein. The viral interference induced by inhibition of reinfection by the MuLVs recognizing the same receptor defines the different subgroups of MuLVs (Lavignon & Evans, 1996). MuLVs are typically found in inbred strains of mice in laboratory and the house mouse subspecies as exogenous MuLVs and endogenous MuLVs. These mouse gammaretroviruses are associated with leukemogenesis. There are three major subgroups of Murine Leukemia Virus (MuLV). The Ectotropic MuLVs (E-MuLVs) are present only in laboratory strains and wild mice species. The Polytropic MuLVs (P-MuLVs) or the mink cell focus inducing virus (MCF- MuLV) can infect a limited number of hosts including human and minks along with laboratory and wild mice. Xenotropic MuLV (X- MuLVs) and amphotropic MuLVs (A-MuLVs) carry the capacity to cause infections in a wide range of mammalian hosts (Kozak, 2015). However, the X-MuLVs can infect cells of other species but not mice. Neither classifiable as X-MuLV nor as P-MuLV, the XP-MuLVs. E-MuLVs and XP-MuLVs are present as ERVs in laboratory mice. A small number of E-MuLVs are found in common laboratory mice strains while all strains carry multiple germline copies of XP-MuLVs (Kozak, 2015).

A pseudotype virus is a replication defective virus produced when the structural and the enzymatic core of one virus is combined with the env protein of another. The P-MuLVs can be transformed to produce viruses alone by transmitting their genomes in the E-MuLV pseudotypes. Infectious P-MuLVs are generated by

recombination between the E-MuLVs and the homologous ERVs present in the genomes of the inbred mice (Lavignon & Evans, 1996).

Primarily, this recombination was believed to occur in the env and the LTR sequences. Recombination has been observed throughout the E-MuLV but is limited in *gag* due to antiretroviral factors of the host. Although the LTRs of pathogenic P-MuLV have gone through significant genetic changes, increased virulence cannot be attributed to a specific change. However, pathogenesis can be attributed to specific substitutions in the env gene. The env gene recombination thus, causes change in the receptor usage by replacing the E-MuLV receptor binding sites with that of the P-MuLV. The P-MuLV transmission in the E-MuLV pseudotypes may introduce leukemia through insertional mutagenesis (Lavignon & Evans, 1996).

Over 50 loci have been identified for all the three subgroups of endogenous MuLV in the C75BL mouse genome. Unlike the JSRV and the ASLV ERV, the endogenous MuLV can produce infectious virions. Many MuLV viral particles can coexist as both XRV and ERV (Kozak, 2015).

Most of the E-MuLV ERVs (Emvs) of laboratory mice are full length functional proviruses with minor defects. The mice carrying active Emvs can show a latency in the production of infectious viruses. Infectious MCF-MuLVs are generated in laboratory mice carrying replication competent or exogenous E-MuLVs. The signs of virus induced lymphomas are also shown by other mouse strains with multiple Emvs. Activation of germline Emvs or acquisition of infectious E-MuLVs by horizontal transfer marks the beginning of leukemogenesis (Kozak, 2015). The Emvs expressed poorly are restricted by certain antiretroviral restriction factors.

The first antiretroviral restriction factor to be discovered was Fv1 (Lilly, 1967).Originally identified as a gene controlling the susceptibility of mice to MuLV induced leukemia, it was also found in cultured mouse cells. Finally cloned in the 1996 the Fv1 gene exhibited a close relation with the gag gene of ERV family, murine endogenous retrovirus L (MERV-L) (Nair & Rein, 2014). The MuLV capsid protein is targeted by Fv1 for restriction. The Fv1 orthologs found in the genus Mus and in other rodent families suggest the same region of conserved synteny of Fv1 orthologs rodents other than *Mus musculus* (Boso et al., 2018).

The inhibition of retroviral infection occurs after reverse transcription but prior to integration of the viral genome. The activity of Fv1 is believed to occur post entry of the PIC into the nucleus. The viral determinant to the susceptibility to inhibition is the CA protein (Luban, 2010). The detection of CA protein of the incoming virus PIC and blockading of viruses of appropriate tropism is done by Fv1. In the case of higher viral loads, the saturation of Fv1 leads to a failure in blockade of infection caused by a second virus. (Goff, 2004).

1.7.4 Mouse Mammary Tumor Virus (MMTV)

MMTV is a type B beta retrovirus with a 9 kb RNA genome, flanked by 5' and 3' LTRs. The 3' LTR contains an open reading frame (ORF) that encodes viral accessory protein, superantigen (Sag) (Holt et al., 2013). Known by the name of mouse strain of origin Mouse Mammary Tumor Virus (MMTV) or MMTV(C3H) is transferred from mother to offspring through milk. Gut associated lymphoid tissue, precisely Peyer's patches is the initial site for MMTV infection. The expression of virally encoded *Sag* (super antigen) gene determines the ability of MMTV to disseminate from gut to mammary gland. Various cytokines and chemokines are activated by stimulation of Sag receptive $CD4^{+}$T cells. The activated T cells, further lead to the activation of B cells by upregulating the CD40 ligand and its binding to the CD40 receptor present on the B cells. This further leads to recruitment of additional B and T cells, activation of immune cells and multiplication in the reservoir of B cells infected with MMTV. The organ susceptible to tumor is infected by the virus when the MMTV infected B cells reach the developing mammary gland (Holt et al., 2013). The reinfections and proviral integrations in the mammary cells lead to development of carcinomas. Although functional Sag protein is essential for development of MMTV infection, both Sag dependent and Sag independent pathways exist. As observed in neonates, infection in the mammary epithelial cells seems to be independent of Sag function, however for disease progression, Sag dependent activation of T cells and further proliferation of B cell is an absolute requirement (Pederson & Sorenson, 2010). Endogenous MMTVs are designated as Mtv and are present in two or more copies in both inbred and wild mice (Dudley, 2011). Most Mtvs are deficient to produce an infectious virus, however Mtv2 is transmitted both through milk and according to mendelian inheritance patterns. A constant Sag expression is observed in almost all Mtvs (Holt et al., 2013). Germline infection is the common cause of oncogenesis caused by MMTV, Mtvs also participate actively in disease progression. Mtv1 and Mtv2 both cause mammary tumors in mice but the infectivity by the former is far less clear than the latter (Dudley, 2011).

Since both MMTV and Mtvs have viral oncogenes, insertional mutagenesis is suspected to be responsible for the development of most mammary tumors. Many Mtvs although incapable of producing infectious viral particles are known to induce murine breast cancers. The mouse B6 strain shows resistance to retrovirally induced mammary tumors and lack of exogenous MMTVs. Treatment of this strain with mammary hormones and chemical carcinogens like dimethylbenzathrazene (DMBA) leads to production of exMMTV. In the case of the B10 strain of mice, infectious proviruses are observed to arise due to

recombination events between Mtv8 and one of the two ERVs resident in the B6 and B10 strains of mice namely Mtv9 and Mtv17 (Dudley, 2011).

The endogenous Mtv6, Mtv8 and Mtv9 proviruses harbored by the BALB/cJ strain show a low prevalence of spontaneous mammary tumors. The infectious virus particles can be produced at a low frequency by activation of Mtv transcription post treatment with chemical carcinogens and hormones (Dudley, 2011). Some investigators suggests that mammary tumorigenesis is preceded by recombination events between the exogenous MMTV and Mtvs, as the mammary tumors obtained from MMTV(C3H) showed integrated copies of recombinants between C3H MMTV and Mtv1 (Dudley, 2011).

In GR mice the Mtv2 is responsible for MMTV expression and tumor development. It has also been observed that in the absence of Mtvs the mammary tumor incidence due to exogenous MMTVs is reduced from a 100% to 10 % in BALB mouse strains (Holt et al., 2013).

1.7.5 Feline leukemia virus (FeLV)

FeLV is a feline gammaretrovirus that can cause disease in felids. Exogenous FeLV (exFeLV) provirus is predominantly responsible for causing fulminant disease in both domesticated and wild felids (Chiu et al., 2018). Lymphosarcoma or leukemia are frequently occurring cancers in a small percentage of the felids affected by the virus (Ortega et al., 2020). The viral manifestation of infection varies considerably in positive animals and includes lethargy, pyrexia, weight loss and anorexia. FeLV is capable of infecting red bone marrow, either reducing the production of red blood cells (RBC) or leading to production of defective RBCs. In certain cases, FeLV can also lead to the deterioration of feline immune system.

The endogenous and exogenous FeLVs show a sequence similarity of 86%, with differences in the *gag* and *env*, and LTR. FeLV-A is considered to be the only transmissible XRV subgroup of FeLV, showing a 100 % transmissibility in clinical cases (E. S. Chiu & VandeWoude, 2021; Jarrett & Russell, 1978).

Oncogenesis, recombination or interference can be a result of interaction between the enFeLV and its exogenous counterparts (Polani et al., 2010). The interaction between the highly transmissible FeLV-A and the enFeLV after co-packaging of the transcripts into a single virion leads to the formation of recombinant FeLVB (E. Chiu et al., 2018). The process of recombination occurs during DNA synthesis directed by retroviral reverse transcriptase, post co-packaging of the ERV and XRV transcripts. Higher mortality and morbidity is observed when FeLV-B infection follows in FeLV-A positive animals (Cano-Ortiz et al., 2022). As FeLV-B is not replication competent it is co-transmitted with FeLV-A as a helper virus.

Very few cases of horizontal transmission of FeLV-B have been described. A replication competent FeLV-B sequence in an endangered non-domestic cat has been described. An absence of enFeLV from non-felis (domestic cats and closely related species) cat species makes them vulnerable to horizontally transmissible FeLV-B infection (E. S. Chiu et al., 2019). An acceleration in the progression of disease and the development of tumors are the results of FeLV-B infection (Bechtel et al., 1999). In an experiment performed on a domestic cat breeding colony, it was documented that the enFeLV copy numbers were higher in the males and the copy numbers were inversely related to the development of disease i.e., the higher the copy number of enFeLV, the lower was the prevalence of FeLV. The domesticated female felids showed the development of progressive FeLV disease and FeLV subtypes, even though the copy number of enFeLV was lower (Powers et al., 2018). Previous experiments had also indicated that recombination events between enFeLV and FeLV-A can enhance pathogenicity for the host while enFeLV can also neutralize the detrimental outcomes of FeLV-A infection (Tandon et al., 2008). De novo mutations in the *env* of FeLV-A gives rise to a rare subtype FeLV-C resulting in aplastic anaemia. Endogenous retroelements divergent from enFeLV recombine with XRVs in domestic felines gives rise to FeLV-D. The presence of FeLV-E subtype was seen in thymic lymphoma in cats (Cano-Ortiz et al., 2022). A T-lymphotropic subgroup of FeLV, namely FeLV-T is generated due to the mutation in the env gene of FeLV-A. The FeLV-T lacks the capacity to induce infections in the host due to a histidine substitution in the Env receptor binding domain. It requires an enFeLV env protein to function as a cofactor, for the maintenance of FeLV-T as an acutely lethal FeLV subtype (E. S. Chiu & VandeWoude, 2021).

1.7.6 Gibbon Ape Leukemia Virus (GALV)

Gibbon Ape Leukemia Virus is an exogenous and highly oncogenic gammaretrovirus that was first isolated in the late 1960s from the US Army- South East Asia Treaty Organization (SEATO) medical research facility in Bangkok, Thailand (McKee et al., 2017). It shows close antigenic relation to the Simian Sarcoma Associated Virus (SSAV) (Murphy & Switzer, 2008). Melomys burtoni Retrovirus (MbRV) and MelWMV proviral sequences isolated from the rodents grassland melomys (*Melomys burtoni*) subspecies respectively show a closest relation to GALV (McKee et al., 2017). GALV is capable of inducing a range of lymphoid tumors including myeloid leukemia in juvenile gibbons (Delassus et al., 1989).

Gibbons were used as models for studying human disease pathogens and their inoculation with blood and tissue samples from humans, rodents and other gibbons could have led to the occurrence of GALV in the captive gibbon population at

SEATO. GALV was identified in four gibbons being used for blood transfusion and transmission of mosquito born infections from humans and rodents respectively to gibbons, particularly during malaria and dengue virus studies (Brown & Tarlinton, 2017).

Other cases included one in 1967, a 1.5-year-old-white-cheeked-gibbon (*Nomascus leucogenys*) in the national Zoological Park in Washington D.C. that died six months after its arrival with the autopsy report stating the cause of death as acute lymphocytic leukemia. Another incidence was reported in the Southeast Asian Treaty organization (SEATO) in Bangkok, Thailand of lymphocytic leukemia in three percent of the white-handed gibbons (*Hylobates lar*) (Eiden & Taliaferro, 2010). Further investigations on the leukemogenesis were attributed to a gammaretrovirus named Gibbon ape leukemia virus (GALV). It was observed that not all GALV- infected gibbons were positive for leukemia. The occurrence of both the virus and anti-viral antibodies were seen in asymptomatic animals. Even after a long-term exposure of one of the gibbons to a highly viremic gibbon, it remained healthy possibly by exhibiting a viral interference mechanism, i.e. by downregulating the expression of viral receptor in the presence of GALV envelope (Eiden & Taliaferro, 2010). In the 1970s, a white-handed gibbon and a pet woolly monkey, housed together in San Francisco become sick within a year, and the woolly monkey was diagnosed with fibrosarcoma. The isolated viruses from both these animals were gammaretroviruses and they were termed as GALV San Francisco (GALV SF) and Woolly monkey virus (WMV) respectively, particularly due to a difference within the env gene of these viruses compared to that observed in captive gibbons. WMV is composed of two viruses, a replication competent SSAV and a replication defective simian sarcoma virus (Eiden & Taliaferro, 2010).

When gibbons connected to the SEATO colony in various locations were tested for antibodies to GALV, a widespread prevalence was found. GALV was found frozen in the brain samples of the gibbons in Gulf South Research Institute, Louisiana, and Hall's Island in Bermuda. During this period GALV was known to be an infectious agent in 11% of captive gibbon population (Brown & Tarlinton, 2017). The occurrence of GALV in the brain samples of gibbons could either be due to their interactions with SEATO gibbons or their contact with the other gibbons in Southeast Asia prior to their shipment. The other possibilities could be the inoculation of gibbons with human brain tissue affected with kuru disease or contamination by GALV-Br of the cell line cocultured with gibbon brain tissue (Brown & Tarlinton, 2017). SEATO facility reports highlight that the biological materials from human patients from Papua New Guinea (PNG) were injected into captive gibbons for up to a year before the GALV-SEATO strain was first identified. Another strain of GALV identified in the US research facility was GALV-Br which was isolated from gibbons injected with human brain samples from PNG (McKee et al., 2017).

There are a total of seven strains of GALV isolated. Out of these four strains namely GALV-SEATO, GALV-SF, GALV- Brain (GALV-Br), and GALV- H were isolated directly from gibbons. Since WMV developed in response to GALV and is related to GALV is classified as a strain of GALV. Two other strains namely GALV- X and GALV-Mr were isolated from HUT-78 and marmoset cells respectively particularly as contaminants of cell cultures (Alfano, Kolokotronis, et al., 2016). Absence of GALV in contemporary European and North American Gibbons has been suggested from the laboratory analysis. The lack of worldwide GALV case documentation and the absence of active virus circulation in gibbons since 1978 suggest the iatrogenesis of GALV during the 1960s-1970s (Brown & Tarlinton, 2017; McKee et al., 2017).

1.8 Koala Retroviruses (KoRV)

Koalas faced a danger of extinction in the early 20th century due to hunting pressures, habitat intrusion, fragmentation and disease. Parallelly, an additional stress prevailed on their health by the deforestation of their staple diet of eucalyptus (Eiden & Taliaferro, 2010). It was in 1987, that thirteen cases of lymphoid neoplasia were reported in kolas in South Wales and Queensland (Canfield et al., 1987). Subsequently 80 % of all fatality was attributed to lymphoma and leukemia in the captive koala populations of southeast Queensland (Eiden & Taliaferro, 2010; Gillett, 2014). Electron micrographs from leukemic tissue of an adult female koala demonstrated the presence of oncoviral particles with a similar morphology to retroviruses in proximity to neoplastic cells. This virus was subsequently sequenced and realized to be an endogenous retrovirus, though clearly a replication competent one (Hanger et al., 2000).

Known to be a gammaretrovirus, KoRV is a spherically shaped virus 80-100 nm in diameter, with a positive sense, ss-RNA genome and a size of approximately 8.4 kb (Enamul et al., 2020). KoRV is a recently endogenized retrovirus discovered in the koala genome thought to have entered anywhere from a few decades to approximately 22,200- 49,900 years (Xu & Eiden, 2015). The integration patterns of KoRV in the koala genomes suggest that the ERV form of KoRV is not possessed by all koalas and hence the process of endogenization is still in action in these animals (Hobbs et al., 2017; R.Tarlinton et al., 2006). Initially, identification of KoRV as an ERV was based on its occurrence in the germ line cells of all the Koalas under consideration. The generation of a full length replication competent virus in the peripheral blood mononuclear cells (PBMCs) of Koalas showed a close relationship between KoRV and GALV (Xu & Eiden, 2015).

A protein motif present in Env is linked to reduced infectivity of KoRV. This CETTG motif has been identified as invariant among highly infectious gammaretroviruses but is absent in endogenous KoRV. In the KoRV env sequences of 17 KoRV isolates this CETTG residue motif is substituted with CETAG in 15 isolates and CGTAG in two isolates (Oliveira et al., 2007). The absence of CETTG motif from the KoRV envelope and the mutation in the Late-budding domain (L-domain) lead to a failure in the efficient detachment of KoRV from the host-cell membrane, resulting in the reduction in the KoRV titer and an attenuation of KoRV infectivity. This reduction in KoRV titer is seen due to the accumulation of the KoRV viral components within the producer cells (Eiden & Taliaferro, 2010; Oliveira et al., 2007).

1.8.1 GALV and KoRV Origin

Koala Retroviral sequences show a close phylogenetic relationship to the exogenous gammaretrovirus GALV. Within gammaretroviruses, KoRV and GALV form a monophyletic clade and share a sequence similarity of 80% (Alfano, Michaux, et al., 2016). Since there is no evidence of coexistence of KoRV and GALV or recombination between the two, the suggestion of their transference to their respective hosts via host switching events probably holds true. A close genetic relationship of GALV to retroviral sequences in Australian rodent species and Indonesian subspecies of *Melomys burtoni* (Alfano, Michaux, et al., 2016; Simmons et al., 2014), named as MbRV and MelWMV respectively, exhibits the potentiality of rodents to act as vectors for interspecies retroviral transfer (McMichael et al., 2019). The geographic distribution of rodent species *Melomys burtoni* (MbRV) overlaps with that of koalas, and MbRV exhibits a sequence similarity of 93 and 83 % with GALV and KoRV respectively, although neither MbRV nor MelWMV viruses are replication competent (Alfano, Michaux, et al., 2016). Candidates acting as potential vectors for cross-species transmission of retroviruses from gibbons to koalas are most likely the ones being able to move between mainland Australia and Southeast Asia, thus establishing a close contact between the two. Since a novel gammaretroviruses has been identified in Australian *Pteropus* species of the order Chiroptera here is a possibility of bats acting as reservoir hosts (Hayward et al., 2020; McMichael et al., 2019). Hervey pteropid gammaretrovirus (HPG) shows the capability of causing infections in bat and human cells and similar levels of cell tropism are observed in KoRV-A and GALV, but not in mice. It is suggested that bats might have played a role in the gammaretrovirus transmission between the gibbons and koalas, due to their capability to traverse waterbodies separating Australia and Southeast Asia (Hayward et al., 2020). The habitat of black flying fox (*Pteropus alecto*) overlaps and connects the habitats of koalas and gibbons. The finding of flying fox

gammaretrovirus (FFTV) forming a clade with both KoRV and GALV, suggests bats as potential hosts during mammalian gammaretrovirus diversification, specifically in the origin of GALV and KoRV (McMichael et al., 2019). Population studies have identified HPG and KoRV-related sequences in several locations of northeast Australia, while phylogenetic analysis done has placed bat viruses in the KoRV-related retroviruses as the basal group (Hayward et al., 2020).

1.8.2 KoRV subtypes

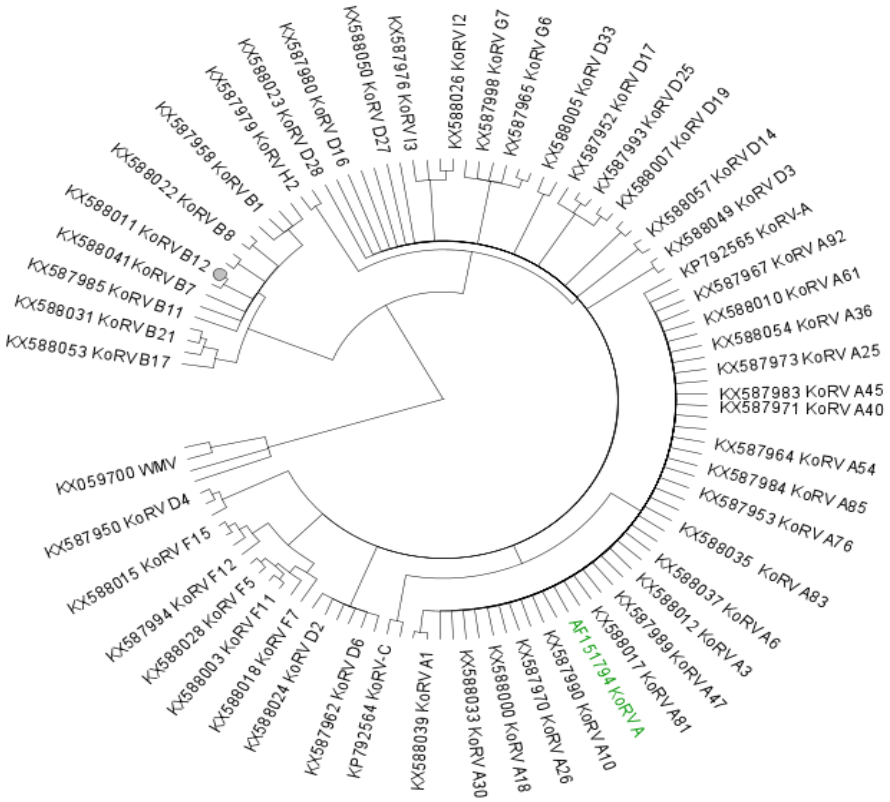


Figure 6 Phylogenetic Tree generated using PhyML on Geneious Prime 2022.2.2 –the Phylogenetic tree represents the differences in the KoRV *env* sequences. The sequence AF151794 KoRV A (highlighted green) is the KoRV-A reference sequence which is going through a process of endogenization while WMV (Wooly Monkey Virus) forms the outgroup. The sequences were aligned using MAFFT, boot strap value-500, the numbers in the phylogenetic tree represent the Genbank Accession number, the sequences for which have been taken from (Quigley et al., 2019).

Based on the sequence differences in the receptor binding domain (RBD) of the *env* gene, KoRV has been classified into three major clades and nine subtypes

(Figure 6). Found in both captive and free ranging koalas the most dominant subtype and clade of KoRV going through the endogenization process is KoRV-A (Hanger et al., 2000; Quigley et al., 2018; Simmons et al., 2012; R. Tarlinton et al., 2006). The second major clade found in both the Australian wild and captive koala communities globally is KoRV-B. A third clade of KoRV comprises of the seven remaining KoRV subtypes namely KoRV-C, D, E, F found in captive koala populations across the world (Shojima et al., 2013; R. Tarlinton et al., 2005) and KoRV-G, H, and I found in the diseased Australian wild Koala population (Quigley et al., 2019).

The KoRV-A subtype was identified in both captive and wild koala populations in Australia and in international zoos. KoRV-A is the only subtype endogenized and shows a 100 % prevalence in northern Australian koalas (R. Tarlinton et al., 2006). A KoRV-A isolate derived from the DNA extracted from PBMCs of a healthy male koala at the Duisburg zoo, Berlin showed two mutations in the envelope gene to that of the original Australian isolate and was named as $KoRV_{D-B}$ (Fiebig et al., 2006). Other KoRV-A isolates from the partial pol and envelope genes sequenced PCR amplicons of five animals in Japanese zoo named as OJ-1-OJ-3, OJ-5 showed >99 % homology with the original Australian KoRV- A isolate (accession no. AF151794), while the OJ-4 isolate was identified to form a new KoRV subgroup called KoRV-J (Miyazawa et al., 2011). Initial isolation of the KoRV-J variant was from a koala held at the Kobe municipal Oji zoo in Japan (Miyazawa et al., 2011). Although identified as different subtypes initially, KoRV J is just another name given to the subtype KoRV-B.

The KoRV-A and KoRV-B sequences show similarity in the *gag/ pol* regions, leading to a failure of KoRV-A *gag/pol* derived probes to differentiate between the two. The KoRV-A and KoRV-B sequences show 94 % residue identity in the envelope genes (Xu et al., 2013; Xu & Eiden, 2015). Both KoRV-A and B exhibit sufficient divergence in their abilities to employ and bind to the receptors. KoRV-A uses the sodium-dependent phosphate transporter, Pit 1 (Xu et al., 2013). The envelope gene of KoRV-B and J has an altered RBD resulting in the use of an alternative receptor namely the thiamine transport protein 1 (THTR1) (Shojima et al., 2013; Xu et al., 2013). All northern Australian koalas have an endogenized KoRV-A which may produce infectious virus particles and is known to spread vertically and horizontally (McEwen et al., 2021).

KoRV can exist in the Koala genome in several forms, including full length provirus (with LTRs, complete gag, pol and env gene), defective provirus (with LTRs, complete env gene, partially or completely missing gag and pol genes) or as recombinant KoRV (RecKoRV) (with a retroelement between the LTRs) (Hobbs et al., 2017; Quigley et al., 2019). One variant of KoRV that is thought to transmit as a defective virus in koalas is KoRV-D (Hobbs et al., 2017; Quigley et al., 2019). The expression of KoRV- D *env* in the plasma was suggestive of its

transmissibility in some form. The explanation that the presence of KoRV-D at higher levels in healthy koalas is not detrimental to their health fits well if KoRV-D is transmitted as a defective virus (Quigley et al., 2019).

Another recently identified variant KoRV-E, was found in captive koalas in zoos in the United states (Xu et al., 2015). On alignment of KoRV-F to the KoRV-C envelope (previously identified in Japan) KoRV-F showed a residue identity of 95% and a resemblance in the variable region A of the envelope region of KoRV-C when aligned with the KoRV-C isolated from Japan (Xu et al., 2015). KoRV-F is a variant of KoRV-D, however it is rarely found in the Koala population. KoRV-G was identified in two koalas with severe chlamydial disease, KoRV-I in a koala with leukemia and KoRV-H was identified in a koala of unknown health status (Chappell et al., 2017; Quigley et al., 2019). With the division of KoRV into subtypes, the KoRV-B subtype has been most related to lymphoma, leukemia, and neoplasia in koalas (Quigley et al., 2018; Waugh et al., 2017; Xu et al., 2013).

1.8.3 Maternal Transmission of KoRV

The KoRV subtypes that are thought to transmit exogenously through dam-joey interactions include KoRV- A, B, D, H, I and K. Possible routes for the exogenous retroviral transmission include the sexual route and transmission through milk, feces and saliva. Transmission of exogenous KoRV has the higher likelihood of occurrence between the dam and offspring particularly due to the close proximity between them (Joyce et al., 2021). Transmission can be due to the sharing of potentially infectious fluids, including milk and pap. Although, the active virus has not been recovered from the milk, KoRV peptide sequences have been discovered in both early and late lactation milk in koalas (Joyce et al., 2021).

In a study, KoRV proviral loads for KoRV-A, B and C were recorded for both the parents and the joey in the Japanese zoo. It was observed that the KoRV-A proviral loads were higher for the sire, followed by the joey and were lowest in the dam. The joey showed higher KoRV-C proviral loads as compared to its parents. This possibility of a KoRV-B negative joey being born to KoRV-B positive parents and it testing positive for KoRV-A indicates endogenous vertical transmission of KoRV-A from dam to joey. This aligns with the 100 % vertical transmission of KoRV-A transmission in the northern koala population (Hashem et al., 2020). In another study, higher levels of KoRV-A sharing were observed in maternally related koalas than any other group. However, close interactions between the dam and joey lead to exogenous transmission of other variants of KoRV (Joyce et al., 2021). Although KoRV-B to I are exogenously transmitted and endogenous KoRV-A is ubiquitous, sharing of KoRV-A in maternally related

koalas suggests the occurrence of an active viral transmission of this subtype (Joyce et al., 2021).

1.8.4 KoRV pathogenesis

Variable receptor usage by KoRV subgroups might be indicative of a wide range of diseases observed in koalas. In some captive koalas in Japan, it has been observed that the disruption of thiamine transport function by KoRV-J caused anemia (Shojima et al., 2013). KoRV as an infectious agent has attracted attention for its role in periodontal disease, as exogenous KoRV may initiate an immunomodulatory response in koalas on contraction of bacterial infections within the oral cavity (Butcher et al., 2020; Enamul et al., 2020). Increased KoRV-A proviral and expressed loads lead to more readily detectable levels may also underly the likelihood of malignant neoplasms and chlamydial disease (Quigley et al., 2019). Koalas infected with KoRV-B, E and F showed significantly higher proportions of cancers than those infected by KoRV-A, including a higher occurrence of neoplasia's in offspring infected with non-A subtypes of KoRV (Zheng et al., 2020). Association of KoRV-B infection in wild koalas has been linked to other neoplasia's like mesothelioma, osteochondroma and an unspecified proliferative bone condition in koalas (Quigley et al., 2018; Zheng et al., 2020).

Although KoRV subtyping was not reported, it was seen that koalas with neoplasia had significantly higher proviral and plasma viral loads when compared to other disease categories (R. Tarlinton et al., 2005; Zheng et al., 2020). Higher viral loads were observed in the plasma and proviral DNA in koalas with cancers compared to those that were alive or died of other causes (Zheng et al., 2020). Based on the detectable measurement of viral loads, higher median plasma viral loads were observed in all subtypes (except KoRV-J) than KoRV-A, even though KoRV-A is expressed as an ERV (Zheng et al., 2020). The viral loads in the blood of some KoRV positive animals can be as high as $> 10^9$ genome equivalents/ ml. There exists a correlation between the viral loads and the progression of disease. A higher viral load increases the chance of insertional mutagenesis and hence heightens the chances of tumor induction (Fabijan et al., 2020; McEwen et al., 2021; Quigley et al., 2019; Sarker et al., 2020; R. Tarlinton et al., 2005).

All the retroviral TM proteins possess a highly conserved immunosuppressive domain which is identical in all the KoRV subtypes and the high viral loads in KoRV positive animals is likely involved in the induction of immunodeficiencies in Koalas (Denner & Young, 2013). Immunomodulation by KoRV can lead to lymphoma, leukemia, anemia, mesothelioma, craniofacial tumors, stomatitis, rhinitis, gingivitis, and opportunistic infections like cryptococcosis, toxoplasmosis and chlamydiosis in koalas (Denner & Young, 2013). However,

leukemia and lymphoma are the common neoplastic diseases observed , leading to a death rate of 3-5 % in wild and 80 % of captive koala population in Australia (Canfield et al., 1987; Hanger et al., 2000; Shojima et al., 2013; R. Tarlinton et al., 2005, 2008).

1.8.5 Association between Chlamydia and KoRV in Koalas

Chlamydia infections in Koalas are linked to ocular and urogenital pathology. The prevalence of overt chlamydial disease and *Chlamydia pecorum* infections are higher in northern koalas. *Chlamydia* is well studied in the New South Wales (NSW) and Queensland (QLD) populations due to the occurrence of severe clinical disease in the northern Australian koalas (Speight et al., 2016).

In southern Australia, the Victorian koala population has been restocked from the French Island population which has been healthy and free of *C. pecorum* but with signs of chlamydia disease particularly "wet bottom" (Legione et al., 2017). Although ocular and urinary tract infections are reported in southern koalas, they may be due other pathogens with the capability of showing similar clinical signs (Polkinghorne et al., 2013).

A higher rate of chlamydiosis is observed in KoRV positive koalas and it is a significant cause of infertility, morbidity and mortality in them (Polkinghorne et al., 2013). Although the link between KoRV and chlamydiosis is unclear, it is believed that the exogenous KoRV-A infection may lead to an alteration in the immune system of koalas making them susceptible to opportunistic infections like Chlamydia (R. Tarlinton et al., 2005). The prevalence of chlamydiosis in Southern Australian koalas is unknown with only fewer reported cases of ocular chlamydiosis (Fabijan et al., 2017).

1.8.6 Retroviral Germline Invasion, endogenization by KoRV-A

Mostly the cross-species transmission of retroviral pathogens leads to somatic cell infections without germ line invasions. The establishment of ERVs within the host species allows the study of these cross-species transmission. Although, the obscurity of the XRV germ line invasion and the ERV establishment in the host species over a period of million years makes it harder to understand the host-virus evolutionary dynamics, the earliest stages of colonization of the host germ line can be studied relatively easily in koalas. The invasion of the koala genome by KoRV-A within the last 50,000 years represents very recent endogenization by KoRV (Löber et al., 2018; Quigley et al., 2019).

Multiple studies have demonstrated that KoRV integrations are polymorphic across the koala population. Variation in the integration patterns of KoRV-A amongst unrelated koalas was suggested by southern blot analysis (Hanger et al., 2000). However, the integration patterns of KoRV amongst various tissues and germ line cells in the same animal were observed to be identical KoRV integrants but are not fixed across koalas, hence they are present at non-identical chromosomal locations in different individual animals (Hanger et al., 2000). In support of the above a pedigree study of the dam-sire-joey triad with 39 KoRV-A integration sites were examined (Greenwood et al., 2018) and it was observed that the KoRV integrants carried by joeys were present in either sire or dam. This showed that each KoRV-A was transmitted by inheritance rather than infection. In another study it was observed that an unrelated dam and sire had only one common integration site, highlighting a high degree of insertional polymorphism across unrelated Koalas (Greenwood et al., 2018). Additionally an experiment in which seven unrelated koalas were tested and 429 5' and 331 3'distinct sequences flanking KoRV-A integration sites across koalas were identified, 93 % of these sequences were unique to individuals (Greenwood et al., 2018; Tsangaras et al., 2014). When the KoRV-A flanking sequences were studied in ten museum koala specimens, hundreds of integration sites were identified (J. Cui et al., 2015), more than 90 % of which were unshared by other sampled members of the koala species (P. Cui et al., 2016; Tsangaras et al., 2014). A difference in the prevalence of copy numbers of KoRV amongst koalas in different geographic locations also acts as an evidence for recent germ line integration by KoRV-A (Greenwood et al., 2018; Xu & Eiden, 2015).

The correlation between disease and KoRV proviral copy numbers when observed highlighted that a higher proviral copies/ cell equivalent in the QLD population corresponds to the endogenously transmitted virus. However, the estimates of about 10^{-3}-10^{-4} proviral copies/ cell in the Victorian population explains the restriction of provirus to fewer infected cells, indicating exogenous transmission in the southern koala population (Simmons et al., 2012). The southern koala population tested for KoRV pol gene PCR tested negative suggesting a lack of KoRV in the Southern Australian koalas (Tarlinton et al., 2006).

1.8.7 Difference in Viral Prevalence in the Northern and Southern Australian Koala Population

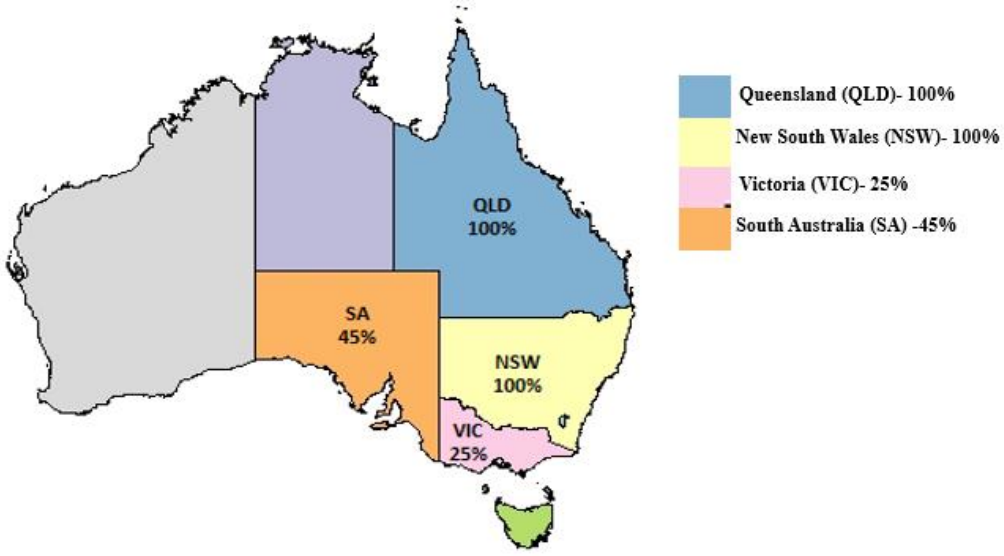


Figure 7 KoRV prevalence map- The map highlights the difference in KoRV prevalence between the northern and southern Australian koala population, with a 100 % northern koalas testing positive for KoRV; the KoRV prevalence is observed to be significantly lower in South Australia and lowest in Victoria (Adapted from Enamul et al., 2020).

Koalas have been divided broadly into two populations based on the epidemiological research being carried on KoRV: northern population (Queensland and New South Wales) and southern population (Victoria and Southern Australia)(Enamul et al., 2020). The differences in diversity between the two populations are attributed to the establishment of a genetic bottleneck in the southern animals caused by a decline in their population from 1890s to 1920s due to hunting. Small refuge populations established on the Victorian offshore islands have since been used for restocking the koala population in the southern range (Tarlinton et al., 2021). The introduction of northern koalas into the southern Australian population and the NSW koalas to Mount Lofty Ranges (MLR) in South Australia has diluted the gene pools and greater diversity of haplotypes was found in MLR compared to the Kangaroo Island (KI) population (Fabijan et al., 2019). A higher prevalence of kidney disease is observed as a result of oxalate nephrosis in the MLR koalas while KI koalas show a higher susceptibility to testicular aplasia. The presence of oxalate nephrosis solely in the southern

Australian koalas points to genetic segregation between the northern and southern population (Sarker et al., 2020).

A 100 % prevalence of endogenous KoRV-A is observed in Queensland (QLD) and New South Wales (NSW) koalas while in southern Australian koalas the occurrence of KoRV is quite low, with only 25 % of the Victorian wild koalas testing positive for KoRV (Figure 7) (Legione et al., 2017; Sarker et al., 2020; Simmons et al., 2012). KoRV-A is endogenized in northern koalas but not in the southern Australian koala population. The presence of KoRV-A (endogenized in NA koalas) in southern Australian animals is at lower than one copy/ genome equivalent, and indicative of somatic rather than germ line insertions (P. Cui et al., 2016; Hobbs et al., 2017; McEwen et al., 2021; Tarlinton et al., 2022).

In another study, when the southern Australian koala populations from Kangaroo Island (KI) and MLR were tested for KoRV, 42.4 % of the KI population and 65.3 % of the MLR populations positive for KoRV provirus, most of them tested positive for KoRV-A (Fabijan et al., 2019). A higher than normal rate of infection recorded in both KI and MLR populations could be attributed to the infections caused by exogenous KoRV-A, due to the animals testing positive for proviral KoRV-pol gene but negative for proviral env KoRV-A and KoRV-B genes (Fabijan et al., 2019). Consistently higher provial copy numbers of KoRV in the northern Australian koalas reflect endogenous infections. Despite endogenous infections, a higher load of KoRV viral RNA in the plasma of QLD koalas indicate that the northern koalas have lesser control over viral replication and KoRV transcription. Southern koalas show relatively lower KoRV RNA in plasma and sometimes no detectable levels of viral RNA in plasma (Meers et al., 2014; Sarker et al., 2020).

There is a possibility that the virus is spreading into Victoria and southern Australian islands post recent endogenization in the north. The southward spread of KoRV could initially be a consequence of horizontal transmission, followed by an endogenization wave after a period of time after significant number of animals have been infected (Legione et al., 2017; Simmons et al., 2012). Another possibility is that the Southern Australian koala population is more resistant to KoRV infection due to subtle genetic differences, and are only susceptible to infections by XRV and so are able to mount immune response to the virus (Sarker et al., 2019; Simmons et al., 2012). A third possibility is that the southern animals were originally colonized by endogenous KoRV-A but that these were lost due to genetic drift and founder effects during the very sharp genetic contraction in the 1920s (and re-founding of the population from a tiny number of individuals). These animals, not tolerated to an endogenous virus would likely mount a more effective immune response to the infectious version and have a lower viral load and prevalence rate than their northern counterparts. Conversely they might carry

replication defective endogenous KoRV, that confer the southern Australian koalas protection against XRV induced pathogenesis (Sarker et al., 2019).

1.8.8 RecKoRVs

Initially when the PacBio sequencing of a northern Australian koala named 'Bilbo' was sequenced and assembled, a proviral integrant called RecKoRV1 was identified. The RecKoRV1 sequence had the 5' KoRV LTR , gag leader region, env region and the 3' KoRV LTR, and the KoRV pol gene region had been replaced with another retroelement known as PhER (Löber et al., 2018). This RecKoRV insertion was observed 15 times in the genome of this single animal. It was observed that the interspersed non-KoRV sequence between the two fragments is abundant in the koala genome. Several KoRV-negative individuals in the Adelaide hills of South Australia, were positive for RecKoRV1. The koalas positive for RecKoRV1 and negative for KoRV likely reflect the mendelian segregation of the integrants in a population which at low frequencies have low copies of both KoRV and RecKoRV, so that a limited proportion of the individuals carry either or both. The regional difference in the expression of PhER may affect the distribution of RecKoRVs by altering the type of PhER template available for recombination. Although, the KoRV and PhER recombination can occur in any koala population that have both, the distinct set of recombination breakpoints suggests independent formation of recombinants between KoRV and PhER. Thus, providing a possible explanation of the difference in the RecKoRV variants found in the northern and southern koala population (Löber et al., 2018).

PhER is 8 kb in size and exhibits some characteristics of an ERV like 478bp LTR, limited sequence similarity to *env* gene sequences of retroviruses and lacks the capacity to code proteins. Recombination events between KoRV and PhER lead to generation of a recombinant KoRV (RecKoRV) (Figure 8) which consists of (1) KoRV 5'LTR, gag leader sequence and truncated 5' end of *gag* from KoRV (2) 4.9 kb 3' end of PhER with 3' LTR (3) KoRV truncated 3' end of *env* and 5'LTR. The variants RecKoRV1, 2 and 3 were found in the genome animal (Bilbo) animal and the variants RecKoRV2 and 3 had different LTR arrangements compared to the frequently occurring 6.9 kb RecKoRV1. Retrotransposition and exogenous infection were believed to be the two possibilities by which RecKoRV could have replicated in the genome of the animal. Although RecKoRV1 lacks protein coding capacity, retrotransposition could be responsible for putative somatic integration exhibited by RecKoRV1 in Bilbo (Hobbs et al., 2017).

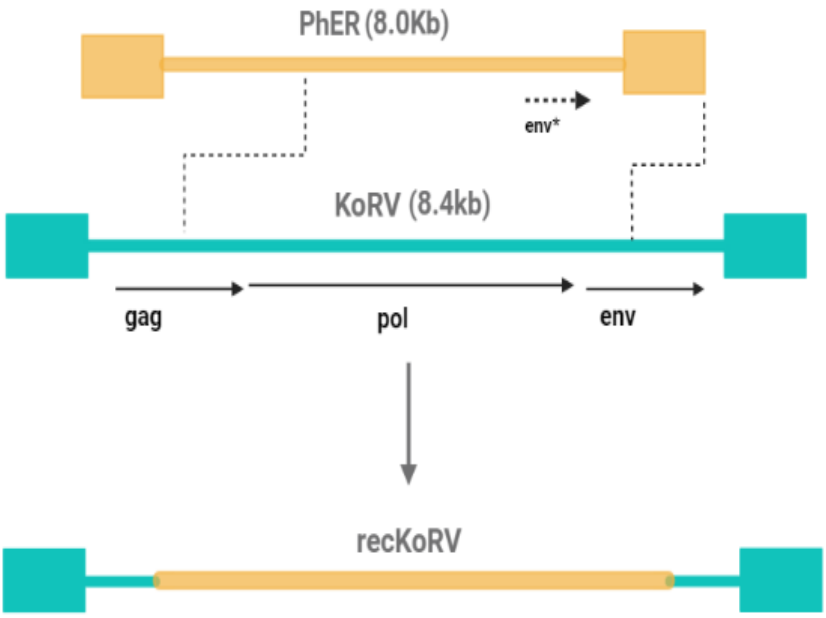


Figure 8 RecKoRV formed by recombination between PhER and KoRV- When PhER (orange) recombines with KoRV (blue). It leads to the formation of a defective KoRV sequence known as RecKoRV. The RecKoRV sequence has the LTRs of KoRV (blue) interspersed by the PhER (orange) retroelement.

Since PhER cannot code for an intact integrase, both KoRV and RecKoRV1 rely on KoRV integrase for insertion into the koala genome (Löber et al., 2018). Transcripts of PhER detected in the koala transcriptome, suggests that KoRV and PhER can be co-packaged in the same virion (Hobbs et al., 2014; Löber et al., 2018) and can lead to recombinant transcripts. Seventeen recombination breakpoints have been detected between KoRV and RecKoRV1. Among the different recombination breakpoints between KoRV and PhER, a high degree of population structuring was detected. In some populations of NSW, RecKoRV1 3' breakpoint was completely absent. A drastic difference was seen in the complement of recombination between KoRV and PhER in the genomes of two koalas (Pacific Chocolate and Birkie), from NSW and QLD respectively, suggested independent formation of recombinants between the two populations. The RecKoRV genesis and distribution can be affected by regional differences in PhER expression, this can affect the type of PhER available for recombination (Löber et al., 2018).

Also, observed in St. Bees Island of Queensland, it was seen that 4 out of 15 koalas were positive for RecKoRV1, however 15/15 were positive for KoRV in contrast to the Koala population of mainland QLD where all animals tested

positive for KoRV and RecKoRV1 (Löber et al., 2018). It was in the 1930s that the population of koalas at St. Bees island was established by translocation of 12-17 animals from QLD, insertional polymorphism amongst RecKoRV1 loci amongst this population or loss of RecKoRV1 loci during genetic drift could be the possible reasons for difference in the recombinant breakpoints between the St. Bees and QLD population (Löber et al., 2018).

As demonstrated in recent work (Tarlinton et al., 2022) southern Australian koalas that tested negative on the *pol* gene PCR and were thought to be free of KoRV, do in fact carry RecKoRV variants in their genomes. Nanopore sequencing results of 5 animals (2 south Australian and 3 Victorian) from CRISPR pull down enrichment of KoRV sequences demonstrated the presence of multiple RecKoRV sequences in each animal, however no full length KoRV was detected. These RecKoRV sequences show differences to the RecKoRVs identified in the northern animals (Löber et al., 2018) and appear to be polymorphic across the population with no RecKoRV insertions present across all southern Australian animals. The presence of RecKoRV sequences in the founder population from French Island implies that the southern Australian animals likely were historically positive for either exogenous or endogenous KoRV, however the genetic bottlenecks of these animals resulted in the loss of KoRV but not RecKoRV out of this population.

1.9 Summary

KoRV exists in both endogenous and infectious forms, however KoRV-A is the only endogenized subtype. Endogenous KoRV- A and exogenous KoRV with multiple subtypes with higher viral loads and disease prevalence have been observed in the northern koalas, while a lower disease rate observed in southern koalas is due to exogenous KoRV. This difference in KoRV prevalence in the two populations could be attributed to founder effects and the establishment of a genetic bottle neck in the southern population due to historical hunting pressures (Tarlinton et al., 2021)**.** Some southern Australian koala populations believed to be KoRV free, now demonstrate the presence of defective KoRV transcripts, lacking the KoRV polymerase gene at least in some of the animals (Tarlinton et al., 2022). The interaction of KoRV with the ERVs in the Koala genome, and its recombination with PhER leading to the generation of RecKoRV, is detected in koalas from across Australia, with RecKoRV1 being the most prevalent variant in the northern koala population (Hobbs et al., 2017; Löber et al., 2018; Quigley & Timms, 2020) The RecKoRV sequences found the southern population vary from those found in the northern Australian koala population. (Hobbs et al., 2014; Löber et al., 2018). The presence of RecKoRV in the germ line cells can lead to its vertical spread across the koala populations. In any population where both PhER and KoRV are present, there is a possibility of recombination. The distinct sets of

recombination breakpoints between the genomes of koalas from QLD and NSW suggest the independent formation of recombinants between the two populations. Thus, providing a probable explanation for the difference in the RecKoRV sequences found in the southern and northern Australian koala populations. The difference in the RecKoRV1 insertion sites between the two QLD koala named Bilyarra and Bilbo suggest that there has not been enough time for broad distribution of RecKoRV1 integrations, and that the RecKoRV1 has been able to retrotranspose to different loci in the koala genome (Löber et al., 2018). It is likely that these RecKoRV variants along with the infectious KoRV were present before the genetic isolation of the southern koalas, and that these infectious KoRV alleles either were lost or never integrated into the genome of the southern Australian koalas due to establishment of genetic bottlenecks (Tarlinton et al., 2021, 2022).

This study sought to confirm the polymorphism of KoRV and RecKoRV sequences across the southern koala population using samples previously collected from Victorian and South Australian animals. The study was planned to inform further work on the potential function and impact of RecKoRV on KoRV replication.

1.10 Objectives

- Examination of KoRV load in the samples from southern animals from a Victorian island population.
- Screening southern animals for polymorphism of RecKoRV loci using integration site specific PCRs in order to explore whether these are fixed or variable in the population.

2. MATERIALS AND METHODS

2.1 Sample Collection and Preparation

2.1.1 Koala Plasma Samples

Fifty koala plasma samples collected into RNA later were obtained from Dr. Michael Lynch (Veterinary Services at Zoos Victoria, Melbourne) from Victorian Koalas of the Cape Otway region were collected as per the method in (R. Tarlinton et al., 2005) in September to December 2018 as part of population management. Tooth score class or tooth wear scales were used to divide the koala samples into separate age classes for giving the approximation of age of the animals (Gordon, 1991). The samples were received and stored in 1.5 ml cryovials in -80 ºC.

	Number
Male	25
Female	25
Tooth Score Class	
1 2 3 4 >4	3 7 7 21 12

Table 1 Overview of number of koala plasma samples and classification based on Tooth Score Class

2.1.2 Koala Spleen Samples

Fifty four koala spleen samples were obtained from Dr. Alistair Legione (University of Melbourne, Melbourne) and were part of those summarized in (Legione et al., 2017). These samples are from Victorian koalas in different regions of Victoria, Australia (mentioned in Table 2 and Figure 9). These samples were received in 2 ml microtubes and were primarily stored in -20 ºC.

Victorian Region	Number of Koalas
Raymond Island	10
Mount Eccles	1
Cape Otway (South Coast)	10
Flinders (Mornington Peninsula)	3
Somers (Mornington Peninsula)	6
French Island	1
Trafalgar (Gippsland)	1
Rawson (Gippsland)	1
Traralgon South (Gippsland)	1
Sandy Point (Gippsland)	2
Mirboo North (Gippsland)	1
Strzelecki (Gippsland)	1
Leongatha (Gippsland)	2
Neerim South (Gippsland)	1
Bolwarra (Far West)	2
Casterton (Far West)	1
Mywee (Far North)	4
Koonoomoo (Far North)	6

Table 2 List of the Victorian koala Samples used for RecKoRV insertion site PCRs (Legione et al., 2017).

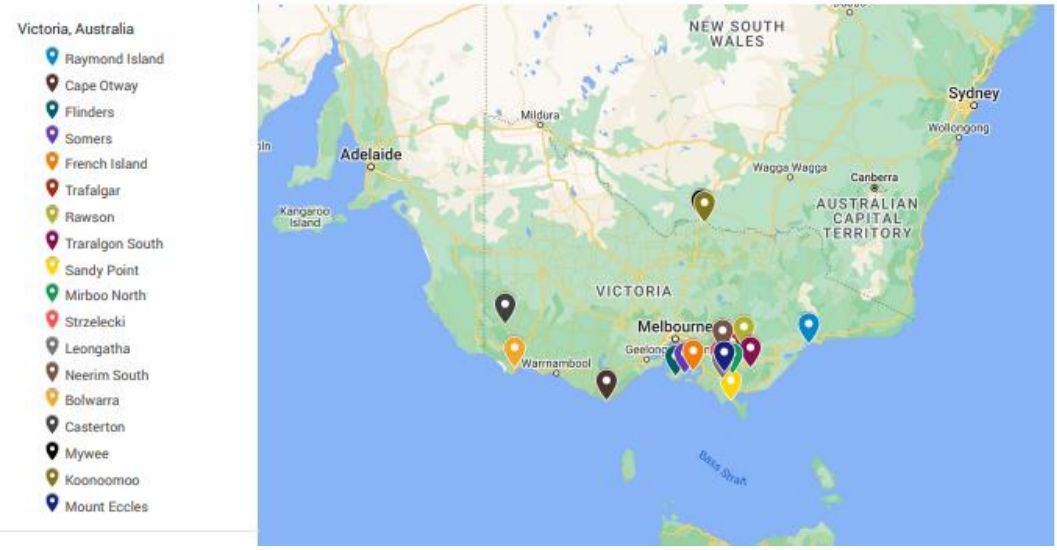


Figure 9 Map of Victoria, Australia (obtained from Google Maps): With a list of all the places from where the Victorian koala plasma samples were collected as referenced in Table 2.

2.1.3 Koala Tissue Samples

Spleen samples were also collected from a cadaver of a koala housed at Longleat safari park (originally from southern Australia) which died of oxalate nephrosis were stored in RNA later in universals at -20 ºC. This animal tested negative for KoRV before they were imported from MLR to Longleat Safari Park.

Koala Name	**Genetic population**	**KoRV-A status**	**Cause of death**
Wilpena	South Australia (Mt Lofty ranges)	Negative	Oxalate nephrosis

Table 3 Wilpena koala samples

2.1.4 Koala Blood Sample

One blood sample of a northern koala named Ben from the captive koala population of Dreamworld theme park, Queensland who died at the age of 11 due to Lymphosarcoma (Tarlinton et al 2006) was used as a positive control for KoRV. The sample was stored at -80 ºC and then was used for genomic DNA isolation.

2.2 Nucleic Acid Extraction

2.2.1 RNA

RNA extraction was performed on the plasma in RNA later samples using the Macherey Nagel NucleoSpin RNA, mini kit for RNA purification as per manufacturer's instructions. Extracted RNA concentration was measured at 260 nm using nanodrop (nanodrop 8000, Thermofisher). The samples were then stored at -80 ºC until further use.

2.2.2 DNA

DNA extraction was performed on the tissue samples obtained from Longleat Safari Park and the koala spleen samples using the Macherey Nagel, NucleoSpin tissue column for DNA from cells and tissue according to manufacturer's instructions. Extracted DNA samples were measured at 260 nm using nanodrop. The samples were stored at -20 ºC until further use.

2.3 Preparation of positive control DNA

2.3.1 Preparation of positive control DNA for KoRV qPCR

The pGEM-T easy vector containing the complete KoRV proviral gene ("pcindy") was obtained from Dr. Jon Hanger (as described in Hanger et al., 2000). Stocks of the plasmids were stored at -80 ºC.

The competent cells JM109 (NEB) were thawed on ice and the cells were mixed by gently flicking the tube. To 50 µl of the JM109 cells 4 µl of the pcindy plasmid was added. The mixture was incubated on ice for 20 minutes. Post incubation on

ice, the cells were subjected to the heat shock treatment on a heat block maintained at 42 ºC for 45-50 seconds. The tube with the cells and plasmid mix was then returned to ice for 2 minutes. This mixture was then added to 250 µl of Luria Bertani (LB) broth and was incubated in a shaking incubator maintained at 35 ºC and 150 rpm for 1 hour and 30 mins.

A 100 µl of the transformation culture was plated on the transformants were allowed to grow on 2-Yeast Tryptone (YT) agar with 12.5 µg/ml ampicillin. The transformed bacteria were then allowed to grow, and the colonies were then picked from the plate and inoculated into 2YT broth with 12.5 µg/ml ampicillin and incubated at 37 ºC for 2 hours and incubated in a shaking incubator maintained at 150 rpm. Plasmid extraction was performed using NucleoSpin plasmid mini-prep DNA columns (Machery Nagel) as per manufacturer's instructions and the concentration of the extracted plasmids were measured using nanodrop at 260 nm. This concentration along with the molecular weight of the extracted plasmid were used to determine the plasmid copy number. Glycerol stocks were prepared by mixing 800 µl of transformed bacterial culture with 200 µl of glycerol and stored at -80 ºC.

One of the Queensland animal blood samples was used for DNA extraction. DNA was extracted using the Macherey Nagel NucleoSpin tissue columns for DNA from cells and tissues. The concentration of extracted DNA was measured at 260 nm on nanodrop.

2.3.2 Preparation of positive control for β-actin gene qPCR

DNA was extracted from the Queensland animal and was subjected to β-actin PCR with the PCR product then purified using the Macherey Nagel PCR purification kit. Alternatively, DNA extracted from the spleen sample from the Longleat animal was used as a positive control.

2.4 cDNA synthesis from RNA

RT was performed in two steps, using M-MLV-RT (Promega) as per manufacturer's instructions and random hexamer primers. The first primer annealing step involved incubation of 15 µl reaction volume consisting of 0.5 µg/µl of random hexamers, up to 2 µg (Promega: C1181) of RNA samples and RNase free water at 70 ºC for 5 minutes. The incubated reaction was immediately cooled on ice. This was followed by the addition of 10 µl of Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT) reaction volume, incubation at 37 ºC for 60 minutes. The M-MLV RT reaction consisted of M-MLV-5X

reaction buffer, 10 mM dNTP mix (Promega cat no: U1511), Recombinant RNAsin® Ribonuclease inhibitor (Promega cat no: N2111), 200 units of M-MLV RT (Promega cat no: M170B) and Rnase free water. All cDNA products were stored at -20 ^{o}C for conventional PCR.

2.5 PCR primers and probes

KoRV pol primers and probe sequences for PCR and RT-PCR were as those in (R. Tarlinton et al., 2005). Beta actin primer sequences designed for koala β-actin sequence were taken from (Sarker et al., 2018). The primers were synthesized by Sigma-Aldrich and probes by Eurofins.

The RecKoRV primers were designed using Primer 3 software on Geneious prime 2022.2.2 (by Dotmatics) and were synthesized by Sigma-Aldrich.

2.5.1 RT-PCR primers and probes

The primer sequences for KoRV pol gene were:

5’-TTGGAGGAGGAATACCGATT-ACAC-3’ (sense)

5’-GCCAGTCCCATACCTGCCTT-3’ (antisense)

The TaqMan probe sequence for KoRV was:

5′-FAM-TCGACCCGTCATGGC-MGBNFQ-3′

The TaqMan probe sequence was labelled at the 5’ end with FAM (6-Carboxyfluorescein) dye and at 3’end was labelled with MGBNFQ (Minor Groove Binder Non-Fluorescent quencher) molecule.

The primer sequences for β-actin were:

5’-TTGCTGACAGGATGCAGAAG-3’ (sense)

5’-ATCCACATCTGCTGGAAGGT-3’ (antisense)

2.5.2 Endpoint PCR primers

The same β-actin and KoRV primers as used for RT-PCR, were used for endpoint PCR as well. RecKoRV primers were designed on Geneious Prime (Figure 10) using primer 3 software.

The primer sequences for RecKoRV3 primers were:

RecKoRV3a: 5'-CCAACAAGGTTTGACTAGCCTCC-3' (sense)

RecKoRV3b: 5'-TGTTGTTGAGTCTTTGTGACCCC-3' (antisense)

The primer sequences for RecKoRV5 primers were:

RecKoRV5a: 5'-TCACCCCCTGCCTCTCTACA-3' (sense)

RecKoRV5b: 5'-GAACAAAGGGCTGGCAGGTG-3'(antisense)

The primer sequences for RecKoRV7 primers were:

RecKoRV7a: 5'-ACCAGACCCTAGACAACGAGG-3'(sense)

RecKoRV7b: 5'-AAAGGAAGAAGGGGTGCAAAGA-3'(antisense)

2.6 RecKoRV PCR Primer Design using Primer3 software in Geneious Prime

The assembly of KoRV containing nanopore reads for the koala "Wilpena" that were mapped to the Koala reference genome assembly (Tarlinton et al., 2022) were received from Dr. Alistair Legione in geneious file format. De novo assembly of these reads was done using the Geneious Assembler to obtain contigs. The contigs obtained were then used to generate a consensus sequence. The consensus sequence obtained was then aligned to the KoRV LTRs for identification of the RecKoRV insertion sites. After the location of KoRV LTRs were identified on the sequence, Primer 3 package on Geneious Prime 2022.2.2 (by Dotmatics) was used at standard settings to design these insertion site specific primers (Figure 10).

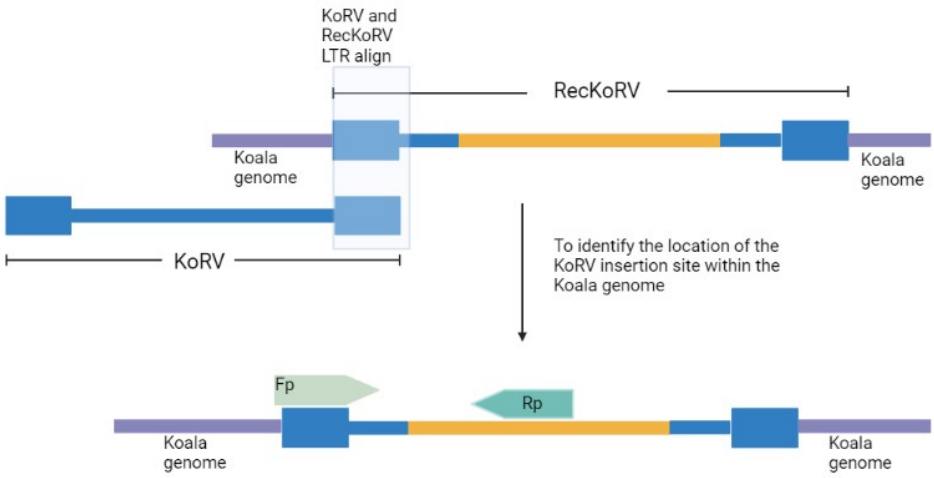


Figure 10 Primer Design for RecKoRV insertion sites*:* The above picture represents the alignment of the KoRV and RecKoRV LTRs done for the identification of the RecKoRV insertion site. The primers (Fp- Forward, Rp- Reverse) were then designed for the RecKoRV insertion site identified using Primer3 primer design software on Geneious Prime 2022.2.2.

2.6.1 Endpoint PCR

The composition of reaction mixtures (all reagents from Promega) for the KoRV pol gene were as follows:

2.5 µl of 10X standard taq buffer, 1.5 µl of 25 mM $MgCl_2$, 0.5 µl of 10 µM KoRV sense primer, 0.5 µl of 10 µM KoRV antisense primer, 0.5 µl of taq DNA polymerase, 0.5 µl of 10 mM dNTPs ,18.5 µl nuclease free H_2O, 0.5 µl template DNA.

For β-actin (housekeeping) gene PCR, the DNA isolated from Queensland animal was diluted 10 times to be used as template.

The composition of reaction mixtures for β-actin gene Endpoint PCR were as follows:

5 µl of 10X standard taq buffer, 3 µl of 25 mM $MgCl_2$, 1 µl of 10 µM β-actin sense primers, 1 µl of 10 µM β-actin antisense primers, 1 µl of taq DNA polymerase, 1 µl of 10 mM dNTPs, 37.5 µl nuclease free H_2O, 1 µl template DNA.

Cycling conditions for the KoRV pol gene and the β-actin gene on PCR blocks were as follows: 95 ºC for 2 mins, followed by 45 cycles of 95 ºC for 30 sec, 60 ºC for 30 sec, 72 ºC for 30 sec, then a final elongation step at 72 ºC for 2 min and hold at 4 ºC.

2.6.2 Quantitative Real-time PCR:

Endpoint PCR was performed on the isolated plasmid for the amplification of a conserved sequence of KoRV of about 111 bp and the PCR product was purified. The dilution series of PCR product from 10^7 to 10^3 was prepared just prior to use as a standard curve for qPCR .

Endpoint PCR for the β-actin gene was performed on the DNA isolated for amplification of the β-actin housekeeping gene sequence of about 145 bp to verify the quality of nucleic acid samples and to check the quality and suitability of the samples before proceeding the detection of polymorphism using RecKoRV site specific PCR. For convenience the same control was applied to the qPCR samples. The PCR product was purified, and the dilutions were prepared from 10^{10} to 10^6. The dilution series of PCR products was prepared just prior to use as a standard curve for qPCR.

The composition of reaction mixtures for the β-actin gene qPCR were as follows:

12.5 µl 1X Power Up SYBR green master mix (applied biosystems), 1 µl of 10 µM β-actin sense primer, 1 µl of 10 µM β-actin antisense primer, 5.5 µl of nuclease Free H_2O, 5 µl of template DNA. The composition of the reaction mixture for the KoRV pol qPCR gene was as follows:

6.25 µl 2X Promega probe master mix, 1 µl of 10 µM KoRV sense primer, 1 µl of 10 µM KoRV antisense primer, 1 µl of 10 µM KoRV probe, 10.75 µl of nuclease Free H_2O, 5 µl of Template DNA.

Cycling conditions for KoRV pol gene qPCR was as follows: Initial Denaturation for 95 ºC for 2 mins, followed by 95 ºC for 15 sec and annealing at 60 ºC for 30 sec. A standard curve was generated for each run using the tenfold dilution series for KoRV pol and β-actin gene standards from 10^6 to 10^2 and the absolute quantification of samples was done on a Biorad CFX Connect qPCR machine.

2.6.3 Annealing Temperature Gradient PCRs:

Annealing Temperature Gradient PCRs were performed for all the three RecKoRV insertion sites identified, and the annealing temperatures were optimized based on the gel electrophoresis results.

The temperature gradient range for RecKoRV7 and RecKoRV3 was set between 56 ºC to 66 ºC.

For RecKoRV5 the annealing temperature gradient range was set between 60 ºC to 67 ºC.

2.6.4 Endpoint PCR for RecKoRV7:

The composition of reaction mixtures for RecKoRV7 were as follows:

2.5 µl of 10X standard taq buffer with $MgCl_2$ (NEB), 1 µl of 10 µM RecKoRV7a, 1 µl of 10 µM RecKoRV7b, 1 µl of 10mM dNTPs, 0.5 µl of taq DNA polymerase, 18 µl nuclease free H_2O, 1 µl template DNA.

Cycling Conditions:

95 ºC for 5 mins, followed by 30 cycles of 95 ºC for 30 sec, 58.6 ºC for 30 sec, 72 ºC for 30 sec, then a final elongation step at 72 ºC for 10 mins and hold at 4 ºC.

The band size expected was 755 bp.

2.6.5 Endpoint PCR for RecKoRV3:

The composition of reaction mixtures for RecKoRV3 were as follows:

12.5 µl of 2X Promega Go Taq Long Amp PCR master mix, 1 µl of 10 µM RecKoRV3a, 1 µl of 10 µM RecKoRV3b, 9.5 µl nuclease free H_2O, 1 µl template DNA.

Cycling Conditions:

95 ºC for 5 mins, followed by 30 cycles of 95 ºC for 30 sec, 60 ºC for 30 sec, 72 ºC for 1 min 30sec, then a final elongation step at 72 ºC for 5 mins and hold at 4 ºC set on the Bioer thermocycler.

The band size expected was 1213 bp.

2.6.6 Endpoint PCR for RecKoRV5:

The composition of reaction mixtures for RecKoRV3 were as follows:

12.5 µl of 2X Promega Go Taq Long Amp PCR master mix, 1 µl of 10 µM RecKoRV5a, 1 µl of 10 µM RecKoRV5b, 9.5 µl nuclease free H_2O, 1 µl template DNA.

Cycling Conditions:

95 ºC for 5 mins, followed by 30 cycles of 95 ºC for 30 sec, 62.8 ºC for 30 sec, 72 ºC for 2 mins, then a final elongation step at 72 ºC for 5 mins and hold at 4 ºC.

The band size expected was 2885 bp.

2.6.7 Agarose Gel Electrophoresis:

PCR products were visualized by gel electrophoresis, 1 % TAE gel with Nancy red dye at 80 V for 60 min with imaging on Invitrogen ibright imaging system.

2.6.8 Sanger sequencing of PCR products

After the visualization of PCR products by gel electrophoresis, the bands were extracted from the gels and purified using the Nucleospin gel and PCR clean-up kit (Macherey Nagel). The concentration of the PCR products was checked on the Thermofisher scientific nanodrop 8000 and were then sent for Sanger sequencing to eurofins genomics to confirm the identity of the sequences.

3. RESULTS

3.1 Copy numbers in Victorian koalas

Fifty Koala plasma samples were run in duplicates and tested for β-actin housekeeping gene, out of which only 26 samples had amplifiable beta-actin RNA. Out of these 26 samples only 4 were positive for the KoRV pol gene, giving an overall KoRV prevalence of 15.3 % in the Victorian Koala population tested. In all the cases, the CT values were quite high and the copy numbers per ml were very low.

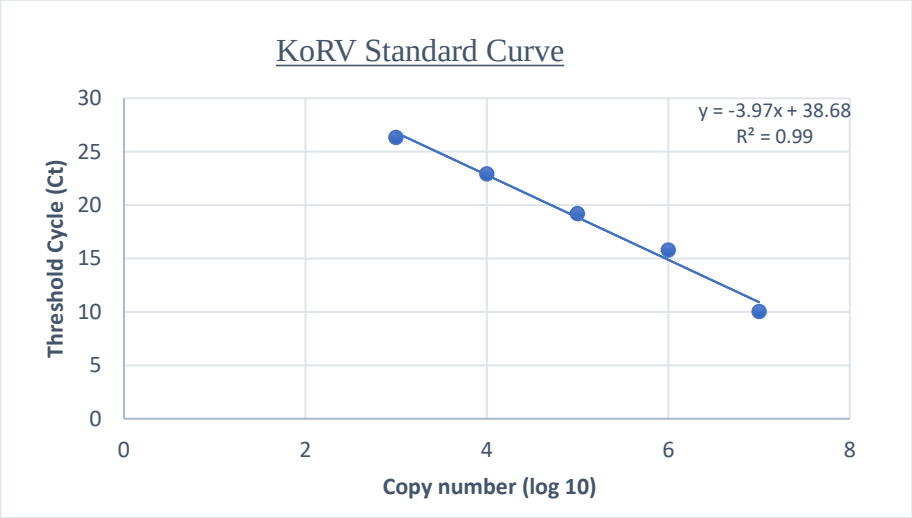


Figure 11 Figure 11 KoRV qPCR standard curve: x -axis represent the copy number/ ml after serial dilutions of the samples and y – axis represents the Ct values. The standard curve equation obtained was used to calculate the KoRV copy number/ml in the samples positive for KoRV (mentioned in table 4)

Sample Numbers	Ct mean	Copy number (copies/ml)
Sample 3	41.04	21-26
Sample 4	42.91	29-55
Sample 6	40.42	15-20
Sample 43	36.33	21-26

Table 4 Ct values and Copy number calculations for KoRV positive samples

3.2 Optimization of Insertion – site PCRs

3.2.1 β-actin

All 54 samples that were tested for the β-actin gene, tested positive at the expected band size of 145 bp.

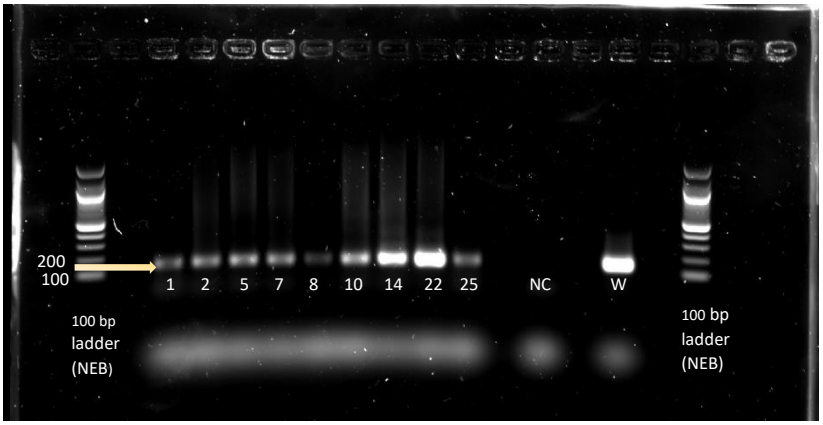


Figure 12 Example of Beta actin Endpoint PCR results for DNA from Victorian koala spleen Lane1= 100 bp ladder (NEB), W= Wilpena (Positive Control), NC= Negative Control. Band observed at the right size at about 145bp. The numbers mentioned below each lane represent the sample numbers of the koalas positive for the housekeeping gene and were further tested for each RecKoRV insertion sites.

3.2.2 RecKoRV7:

The temperature range set for the RecKoRV7 annealing temperature gradient were as follows:

NC/1	2	3	4	5	6	7	8	9	10	11	12
56.0 ºC	56.2 ºC	56.6 ºC	57.5 ºC	58.6 ºC	60.0 ºC	61.7 ºC	63.1 ºC	64.2 ºC	65.2 ºC	65.7 ºC	66.0 ºC

Table 5 Annealing Temperature Gradient Range for RecKoRV7 maintained on the Bioer Thermocycler

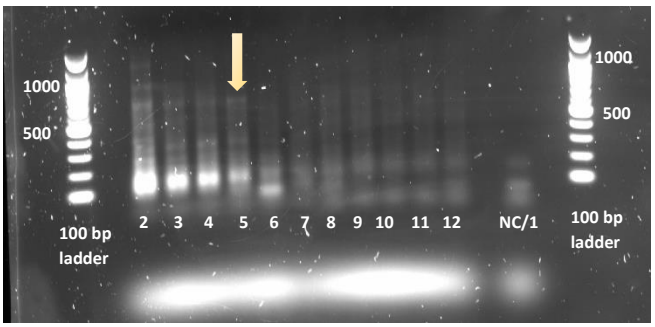


Figure 13 RecKoRV7 Gradient shows the annealing temperature gradient with a range between 56 ºC to 66ºC for RecKoRV7 using Wilpena koala sample. With NC/1 at 56 ºC and the number 12 at 66 ºC. Well labelled as 5= 58.6 ºC was chosen as the annealing temperature as the right size band was observed at 58.6 ºC.

All 54 samples were tested for RecKoRV7. A band was observed around the expected size of 755 bp as well as smaller non-specific bands in 16 samples (Figure 13).

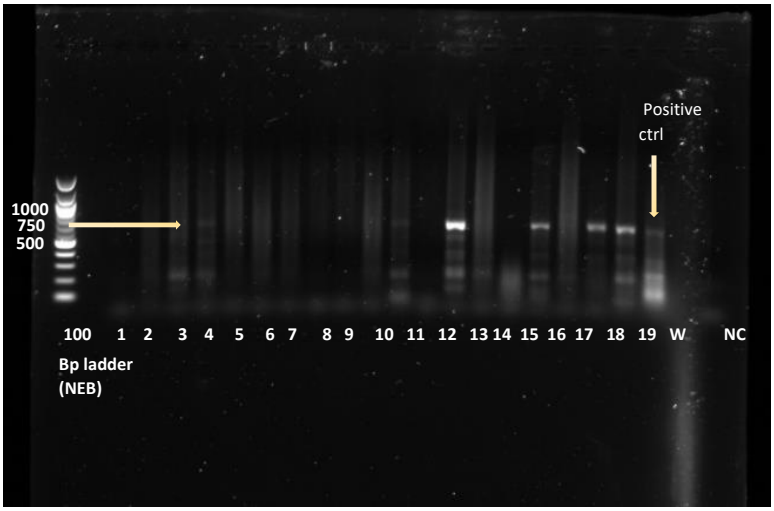


Figure 14 Example of RecKoRV7 site-specific PCR results for DNA from Victorian koala spleen -Samples positive for RecKoRV7 showed bands at the same band size as the W= Positive control (Wilpena) at about 755 bp. The numbers mentioned below each lane represent the koala sample numbers. NC= Negative Control.

3.2.3 RecKoRV3

The temperature range set for the RecKoRV3 annealing temperature gradient PCRwere as follows:

NC/1	2	3	4	5	6	7	8	9	10	11	12
56.0 ºC	56.2 ºC	56.6 ºC	57.5 ºC	58.6 ºC	60.0 ºC	61.7 ºC	63.1 ºC	64.2 ºC	65.2 ºC	65.7 ºC	66.0 ºC

Table 6 Annealing Temperature Gradient Range for RecKoRV3 maintained on the Bioer Thermocycler

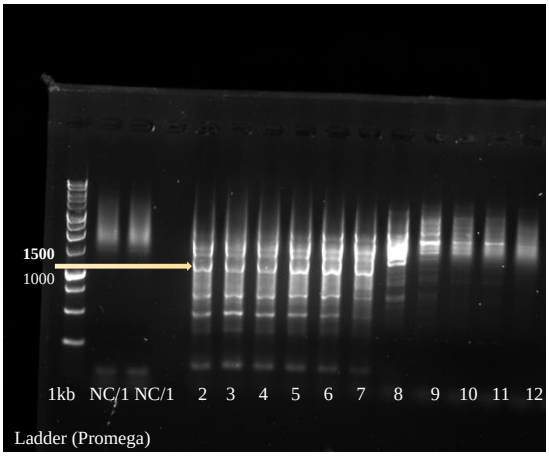


Figure 15 RecKoRV3 Gradient shows the annealing temperature gradient for RecKoRV3 using Wilpena koala sample. The temperature range was set between 56 ºC to 66 ºC, with NC/1 at 56 ºC while the well labelled 12 with sample at annealing temperature 66 ºC. Well labelled as 6= 60 ºC annealing temperature chosen. NC/1 = Negative Control

Clear bands of the correct size (1213 bp) were not achieved for the RecKoRV3 primer set, either with the positive control or the test samples. Multiple non-specific bands were seen in all samples (figure 16).

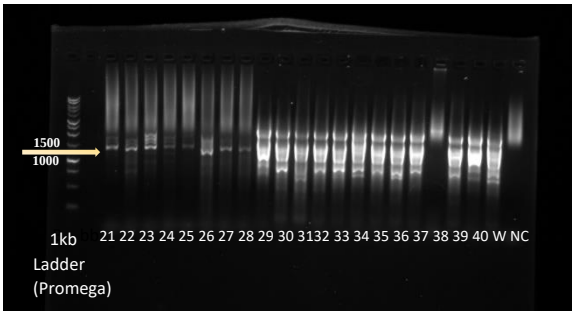


Figure 16 Example of RecKoRV3 site-specific PCR results for DNA from Victorian koala spleen- the number mentioned below each well represent the koala spleen sample numbers. While a few samples showed the right sized bands, but it couldn't be compared with W= Wilpena (Positive control) due to multiple bands, NC= Negative Control.

3.2.4 RecKoRV5:

The temperature range set for the RecKoRV3 annealing temperature gradient PCR were as follows:

NC/1	2	3	4	5	6	7	8	9	10	11
60.1 ºC	60.4 ºC	61.0 ºC	61.8 ºC	62.8 ºC	63.9 ºC	65.0 ºC	65.8 ºC	66.4 ºC	66.8 ºC	67.0 ºC

Table 7 Annealing Temperature Gradient Range for RecKoRV5 maintained on the Bioer Thermocycler

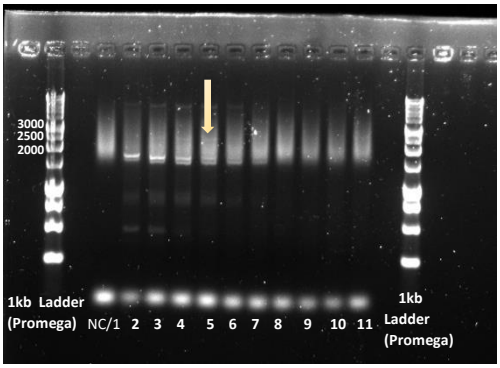


Figure 17 RecKoRV5 Gradient shows the annealing temperature gradient with the temperature range between 60 ºC to 66 ºC for RecKoRV5 using Wilpena koala sample. For the well labelled NC/1, the annealing temperature was maintained at 60 ºC and the sample in 11 the annealing temperature was 67 ºC. The optimal annealing temperature for RecKoRV5 was recorded as 62.5 ºC, as shown in well labelled as 5= 62.8 ºC

Multiple bands were observed for RecKoRV5 allele specific PCR (figure 18) which could not confirm the results. Although the expected band size was 2885 bp.

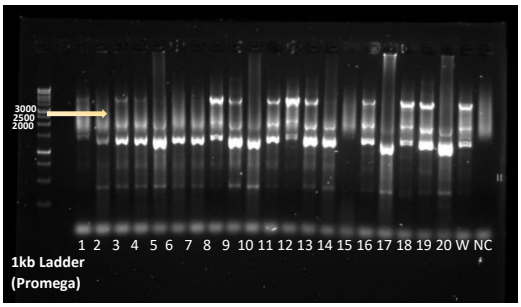


Figure 18 Example of RecKoRV5 site-specific PCR results for DNA from Victorian koala spleen: The wells labelled from number 1 to 20 represent the koala spleen sample numbers while W= Wilpena (Positive control) and NC= Negative Control. Nonspecific bands were observed which could not confirm the results.

3.2.5 Sangar sequencing of PCR products

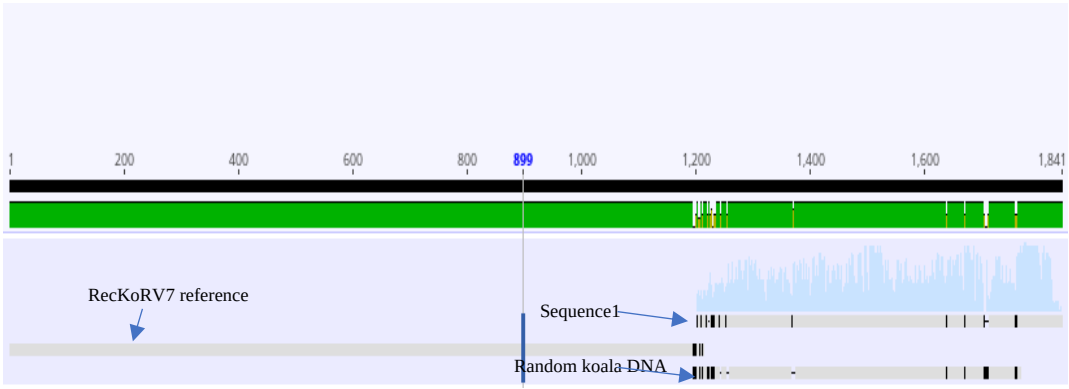


Figure 19 Sanger Sequencing Results and Alignment- Sequence 1 is the RecKoRV7 reverse sequence (Sanger sequencing results and on blast N, it shows sequence similarity with Koala DNA and not RecKoRV7 reference sequence indicating random amplification of Koala DNA.

The PCR products for RecKoRV7 and RecKoRV3 were sent for sequencing to Eurofins genomics. The sequencing reaction for RecKoRV3 failed. The sequencing results obtained for RecKoRV7 (Figure 19) were obtained as RecKoRV forward and RecKoRV reverse sequences. Using BLASTN for checking the sequence similarity of the sequence to the Koala genome, the only koala transcript that it aligned with was a transcript associated with lymphoma and not with any of the KoRV sequences indicating the amplification of random koala genome instead of the KoRV locus.

4. DISCUSSION

4.1 Overview of Research Aims

The research work presented here aimed to screen the Victorian koala population for the presence of KoRV using copy number calculations based on the results of qPCR experiments and to detect polymorphism of RecKoRV loci using integration site specific PCR. The integration site specific PCR was designed to explore whether the RecKoRV insertion sites are fixed or variable in the southern koala population. To study polymorphism we designed primers specific to several RecKoRV insertion sites identified in the animal named Wilpena used as part of the (Tarlinton et al., 2022) study.

4.2 KoRV prevalence:

The prevalence of KoRV in the southern Australian koala population is quite variable, reported as between 15-72 % in other studies (Legione et al., 2017). In the Victorian koala plasma samples from Cape Otway obtained from Dr. Michael Lynch, the percentage prevalence of KoRV was observed to be towards the lower end of the reported range. Out of the 50 samples that were tested, only 26 passed the β-actin quality control. Out of these 26 samples, only 4 were positive for KoRV and had much lower copies of the virus than those in the Northern Australian Koalas (Simmons et al., 2012; R. Tarlinton et al., 2005) giving an overall KoRV prevalence in that population of 15.3 %. The copy number values recorded were also quite low, in line with previous reports of lower copy numbers in southern animals (Sarker et al., 2020; Simmons et al., 2012).

It has been proposed that KoRV is going through a process of endogenization in southward direction which is based on the investigations of variations in the KoRV prevalence between the northern and southern koala populations. Victorian koalas show lower numbers of proviral copies of KoRV when compared to the Northern koalas (Simmons et al., 2012).

The common origin of the Victorian and South Australian koala population may explain the difference in the KoRV pathogenicity between the northern and southern Australian koala populations. Koalas that predominantly originated from mainland Victoria were used to populate South Australia. Additionally, some koalas introduced in the South Australian Koala population were from the northern populations which included the descendants of Queensland and Victorian koalas bred in captivity (Fabijan et al., 2019). Reports also mention the possibility of the introduction of koalas of NSW origin in the MLR. The koalas found in the

MLR region exhibit a dilution in the gene pool and highest level of diversity amongst the southern koalas which is congruous with the observations of founding animals being obtained from QLD and NSW. The maternal mitochondrial control region haplotype spread across SA and Victoria suggest recent history of translocations that have been documented. Koalas at the Eyre Peninsula, Kangaroo Island and Bessiebelle are similar to those in French Island which is believed to be the founder population (Neaves et al., 2016). The koala population found in the French Island is a closed population that was established around 1900 with a low number of koalas. Since the initial translocations for restocking koala populations the island has had no new individuals added. The finding of RecKoRV in this population (Tarlinton et al., 2022) hence suggests that KoRV might have been present in the founder populations of the southern animals.

It has generally been seen that there is an exogenous transmission of KoRV-A in southern koalas and the functional endogenization of KoRV-A has not been observed in the southern Australian koala population.

4.3 RecKoRV

It is predicted that KoRV on recombination with PhER would suffer loss of virulence, since none of the recombinants are expected to code for an intact virus (Löber et al., 2018). While there is a possibility of RecKoRV potentially exerting deleterious effects on the host by retro transposition into new genomic locations, there is likely still a reduction in the deleterious effects by the provirus when compared to intact KoRV. There is a higher probability that the presence of replication competent KoRV in all tissues may caus new somatic integrations and reintegrations with the potential to affect oncogenes and the likeliness of causing higher disruption than RecKoRV (McEwen et al., 2021).

Northern Australian koalas being born with KoRV-A probably fail to recognize it as foreign and therefore fail to mount an immune response. The establishment of a genetic bottleneck in the southern Australian koala population may have led to the loss of endogenized KoRV but not RecKoRV (Tarlinton et al., 2022). There is a possibility that RecKoRV variants might protect the southern koala population from diseases linked to the exogenous KoRV, however that may depend on the location of RecKoRV sequence and may depend on if nearby genes are interrupted or silenced (Tarlinton et al., 2022). This interruption and silencing of genes may also be involved in the reduction of the spread of exogenous KoRV in southern Australian koalas. As seen in the case of other ERVs like FeLV and MuLV, endogenous retroelements can lead to an inhibition in the entry or release of retroviruses in cells, thus increasing resistance to future retroviral infection (Greenwood et al., 2018; McDougall et al., 1994). Receptor interference seen in

case of enJSRV and ALV-E infections can lead to decreased availability of the viral receptors inhibiting the entry and infection caused by their replication competent counterparts (Hunt et al., 2008; Mura et al., 2004).

Initial results from nanopore sequencing (Tarlinton et al., 2022) indicated that RecKoRV loci are polymorphic across South Australian and Victorian populations and this work sought to confirm that across a larger number of animals. Endpoint PCR using three different RecKoRV primer pairs was performed for three different RecKoRV insertion sites that were identified in the Wilpena koala sample (used as a control). All the 54 Victorian spleen samples were screened to check whether these insertion sites are polymorphic within the southern Australian koalas or not. The presence of multiple bands on the agarose gel which could not be resolved with annealing temperature gradient PCRs and the repetitive nature of the LTR in the RecKoRV3 and RecKoRV5 sequences led to problems with analyzing the results. Although some samples showed the right sized bands for the RecKoRV3 sequences at 1213 bp, the failure of anger sequencing reaction led to problems with confirmation of results. The results for RecKoRV5 were also inconclusive.

For RecKoRV7 after setting the annealing temperature gradient PCR, the temperature was optimized at 58.6 ºC and the right sized bands were observed for 16 samples at 755 bp. However, the sanger sequencing results did not match with the RecKoRV7 insertion site sequence and showed amplification of random koala DNA. There is a possibility that the multiple copies of KoRV LTRs for different insertion sites increases the potential of mis priming of the primers, thus making it difficult to design primers for a specific insertion site. Thus, there is a need to look at other techniques to screen for sequences which are polymorphic and repetitive. Other techniques that could be used are as follows:

- Using different primers or insertion sites- although the limitation of this technique being the limited number of insertion sites with enough sequencing quality and depth leading to problems with accurately designing primers as seen in with the above three RecKoRV sequences.
- Long read sequencing again either whole genome or the CRISPR pull down used to originally isolate the Wilpena sequences. However, the high cost of sequencing is the major limitation of this technique.
- Using Inverse PCR methods by digesting the samples with restriction enzymes and self-ligating to circularize them. Further using nested PCR with two primers pairs to amplify the unknown sequence followed by sequencing (Greenwood et al., 2018). Though this method has the disadvantage of only sequencing the insertion sites and not the complete RecKoRV locus.

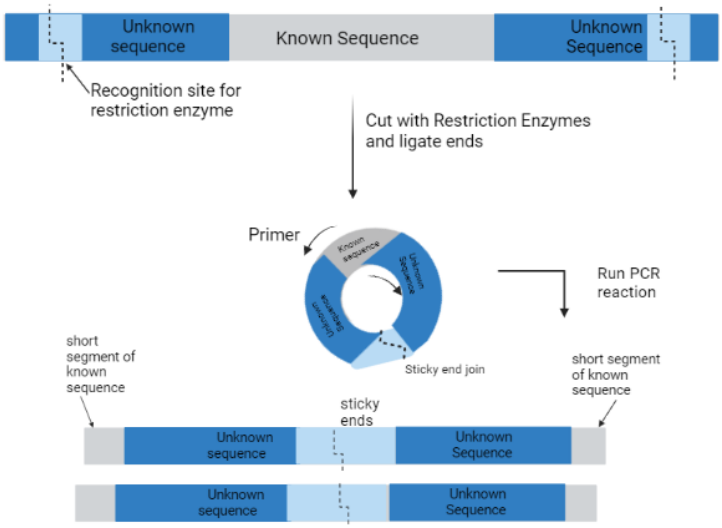


Figure 20 Diagram explaining the working of Inverse PCR reaction: The figure shows the restriction digestion of the target DNA at specific restriction sites. The linear DNA is circularized and then amplified with primers that anneal to the known region. The PCR products have the unknown DNA which can be cloned and sequenced (created on Biorender).

In order to look at the RecKoRV insertion sites, one of the above-mentioned techniques could be optimized and used to amplify the sequences and confirm the polymorphic nature of the RecKoRV sequences within the Victorian koala population. The shortage of time and money made it challenging to use these alternative techniques to look at the RecKoRV insertion sites in this project.

It will be worthwhile using an affordable technique like Inverse PCR for screening different RecKoRV loci and for the identification of the RecKoRV variants in the southern Australian koala populations. Since these RecKoRV variants in southern Australian koalas are not fixed and might be protecting them from the diseases linked with exogenous KoRV. The identification of the polymorphism of these RecKoRV loci within the Victorian koala population would then help explore the links between infectious KoRV prevalence and clinical disease incidence. If these RecKoRV variants are polymorphic and reduce the susceptibility of the Victorian koalas to KoRV, then selective breeding for the reintroduction of this trait into the KoRV positive northern koalas will help control KoRV associated malignancies in koalas and take us a step further towards achieving the aim of koala health and conservation.

ACKNOWLEDGEMENTS

I would like to thank:

- Dr. Rachael Tarlinton (SVMS) for her supervision, constant support and guidance throughout the project.
- Dr. Adam Blanchard (SVMS) for his support with the bioinformatic analysis of the data
- Dr. Ternenge Apaa and Cristobal Castillo for their constant help and motivation
- Dr. Alistair Legione (University of Melbourne) for provision of Victorian koala spleen samples and RecKoRV sequence information
- Dr. Micheal Lynch (Zoos Victoria) for provision of Victorian koala plasma samples
- The lab technicians (SVMS) for their advice on laboratory techniques and equipment.

BIBLIOGRAPHY

Alfano, N., Kolokotronis, S.-O., Tsangaras, K., Roca, A. L., Xu, W., Eiden, M. V., & Greenwood, A. D. (2016). Episodic Diversifying Selection Shaped the Genomes of Gibbon Ape Leukemia Virus and Related Gammaretroviruses. *Journal of Virology*, *90*(4), 1757–1772. https://doi.org/10.1128/JVI.02745-15

Alfano, N., Michaux, J., Morand, S., Aplin, K., Tsangaras, K., Löber, U., Fabre, P.-H., Fitriana, Y., Semiadi, G., Ishida, Y., Helgen, K. M., Roca, A. L., Eiden, M. V., & Greenwood, A. D. (2016). Endogenous Gibbon Ape Leukemia Virus Identified in a Rodent (Melomys burtoni subsp.) from Wallacea (Indonesia). *Journal of Virology*, *90*(18), 8169–8180. https://doi.org/10.1128/JVI.00723-16

Archer, F., Mornex, J.-F., & Leroux, C. (2012). Interplay Between JSRV, an Oncogenic Retrovirus, and the Pulmonary Epithelium. *Current Immunology Reviews*, *8*, 209–215.

Arhel, N. J., Souquere-Besse, S., Munier, S., Souque, P., Guadagnini, S., Rutherford, S., Prévost, M.-C., Allen, T. D., & Charneau, P. (2007). HIV-1 DNA Flap formation promotes uncoating of the pre-integration complex at the nuclear pore. *The EMBO Journal*, *26*(12), 3025–3037. https://doi.org/10.1038/sj.emboj.7601740

Armezzani, A., Varela, M., Spencer, T. E., Palmarini, M., Arnaud, F., & Dé Rick Arnaud, F. (2014). "Ménage à Trois": The Evolutionary Interplay between JSRV, enJSRVs and Domestic Sheep. *Viruses*, *6*(12), 4926–4945. https://doi.org/10.3390/v6124926

Bechtel, M. K., Hayes, K. A., Mathes, L. E., Pandey, R., Stromberg, P. C., & Roy-Burman, P. (1999). Recombinant feline leukemia virus (FeLV) variants establish a limited infection with altered cell tropism in specific-pathogen-free cats in the absence of FeLV subgroup A helper virus. *Veterinary Pathology*, *36*(2), 91–99. https://doi.org/10.1354/vp.36-2-91

Blumenthal, R., Durell, S., & Viard, M. (2012). HIV Entry and Envelope Glycoprotein-mediated Fusion. *Journal of Biological Chemistry*, *287*(49), 40841–40849. https://doi.org/10.1074/jbc.R112.406272

Boso, G., Buckler-White, A., & Kozak, C. A. (2018). Ancient Evolutionary Origin and Positive Selection of the Retroviral Restriction Factor Fv1 in Muroid Rodents . *Journal of Virology*, *92*(18), 850–868. https://doi.org/10.1128/JVI.00850-18/SUPPL_FILE/ZJV018183850S1.PDF

Boulliou, A., Le Pennec, J. P., Hubert, G., Donal, R., & Smiley, M. (1992). The Endogenous Retroviral ev 21 Locus in Commercial Chicken Lines and its Relationship with the Slow-Feathering Phenotype (K). *Poultry Science*, *71*(1), 38–46. https://doi.org/10.3382/PS.0710038

Brown, K., Emes, R. D., & Tarlinton, R. E. (2014). Multiple Groups of

Endogenous Epsilon-Like Retroviruses Conserved across Primates. *Journal of Virology*, *88*(21), 12464–12471. https://doi.org/10.1128/JVI.00966-14

Brown, K., & Tarlinton, R. E. (2017). Is gibbon ape leukaemia virus still a threat? *Mammal Review*, *47*(1), 53–61. https://doi.org/10.1111/mam.12079

Butcher, R., Pettett, L., Fabijan, J., Ebrahimie, E., Mohammadi-Dehcheshmeh, M., Speight, K., Boardman, W., Bird, P., & Trott, D. (2020). Periodontal disease in free-ranging koalas (*Phascolarctos cinereus*) from the Mount Lofty Ranges, South Australia, and its association with koala retrovirus infection. *Australian Veterinary Journal*, *98*(5), 200–206. https://doi.org/10.1111/avj.12919

Canfield, P. J., Brown, A. S., Kelly, W. R., & Sutton, R. H. (1987). Spontaneous lymphoid neoplasia in the koala (Phascolarctos cinereus). *Journal of Comparative Pathology*, *97*(2), 171–178. https://doi.org/10.1016/0021-9975(87)90037-5

CANFIELD, P., SABINE, J., & LOVE, D. (1988). Virus particles associated wth leukaemia in a koala. *Australian Veterinary Journal*, *65*(10), 327–328. https://doi.org/10.1111/j.1751-0813.1988.tb14518.x

Cano-Ortiz, L., Tochetto, C., Roehe, P. M., Franco, A. C., & Junqueira, D. M. (2022). Could Phylogenetic Analysis Be Used for Feline Leukemia Virus (FeLV) Classification? *Viruses*, *14*(2), 249. https://doi.org/10.3390/v14020249

Carlson, J., Lyon, M., Bishop, J., Vaiman, A., Cribiu, E., Mornex, J.-F., Brown, S., Knudson, D., DeMartini, J., & Leroux, C. (2003). Chromosomal Distribution of Endogenous Jaagsiekte Sheep Retrovirus Proviral Sequences in the Sheep Genome. *Journal of Virology*, *77*(17), 9662–9668. https://doi.org/10.1128/JVI.77.17.9662-9668.2003

Chappell, K. J., Brealey, J. C., Amarilla, A. A., Watterson, D., Hulse, L., Palmieri, C., Johnston, S. D., Holmes, E. C., Meers, J., & Young, P. R. (2017). Phylogenetic Diversity of Koala Retrovirus within a Wild Koala Population. *Journal of Virology*, *91*(3). https://doi.org/10.1128/JVI.01820-16.

Chemudupati, M., Kenney, A. D., Bonifati, S., Zani, A., McMichael, T. M., Wu, L., & Yount, J. S. (2019). From APOBEC to ZAP: Diverse mechanisms used by cellular restriction factors to inhibit virus infections. In *Biochimica et Biophysica Acta - Molecular Cell Research* (Vol. 1866, Issue 3, pp. 382–394). Elsevier B.V. https://doi.org/10.1016/j.bbamcr.2018.09.012

Chiu, E., Hoover, E., & VandeWoude, S. (2018). A Retrospective Examination of Feline Leukemia Subgroup Characterization: Viral Interference Assays to Deep Sequencing. *Viruses*, *10*(1), 29. https://doi.org/10.3390/v10010029

Chiu, E. S., Kraberger, S., Cunningham, M., Cusack, L., Roelke, M., & VandeWoude, S. (2019). Multiple Introductions of Domestic Cat Feline Leukemia Virus in Endangered Florida Panthers1. *Emerging Infectious*

Diseases, *25*(1), 92–101. https://doi.org/10.3201/eid2501.181347

Chiu, E. S., & VandeWoude, S. (2021). Endogenous Retroviruses Drive Resistance and Promotion of Exogenous Retroviral Homologs. *Annual Review of Animal Biosciences*, *9*(1), 225–248. https://doi.org/10.1146/annurev-animal-050620-101416

Cui, J., Tachedjian, G., & Wang, L. F. (2015). Bats and Rodents Shape Mammalian Retroviral Phylogeny. *Scientific Reports*, *5*. https://doi.org/10.1038/SREP16561

Cui, P., Löber, U., Alquezar-Planas, D. E., Ishida, Y., Courtiol, A., Timms, P., Johnson, R. N., Lenz, D., Helgen, K. M., Roca, A. L., Hartman, S., & Greenwood, A. D. (2016). Comprehensive profiling of retroviral integration sites using target enrichment methods from historical koala samples without an assembled reference genome. *PeerJ*, *2016*(3), e1847. https://doi.org/10.7717/PEERJ.1847/SUPP-12

Delassus, S., Sonigo, P., & Wain-Hobson, S. (1989). Genetic organization of gibbon ape leukemia virus. *Virology*, *173*(1), 205–213. https://doi.org/10.1016/0042-6822(89)90236-5

Denner, J., & Young, P. R. (2013). Koala retroviruses: characterization and impact on the life of koalas. *Retrovirology*, *10*. https://doi.org/10.1186/1742-4690-10-108

Dudley, J. (2011). Retroviruses and Insights into Cancer. In *Retroviruses and Insights into Cancer* (Issue January). https://doi.org/10.1007/978-0-387-09581-3

Eiden, M. V., & Taliaferro, D. L. (2010). Emerging Retroviruses and Cancer. In *Retroviruses and Insights into Cancer* (pp. 307–333). Springer New York. https://doi.org/10.1007/978-0-387-09581-3_11

Enamul, M., Kayesh, H., Md, ·, Hashem, A., & Tsukiyama-Kohara, K. (2020). Koala retrovirus epidemiology, transmission mode, pathogenesis, and host immune response in koalas (Phascolarctos cinereus): a review. *Archives of Virology*, *165*, 2409–2417. https://doi.org/10.1007/s00705-020-04770-9

Engelman, A. N. (2010). Reverse Transcription and Integration. In R. Kurth & N. Bannert (Eds.), *Retroviruses* (First Edit, pp. 129–160). Caister Academic Press.

Escalera-Zamudio, M., & Greenwood, A. D. (2016). On the classification and evolution of endogenous retrovirus: human endogenous retroviruses may not be ‘human’ after all. *APMIS*, *124*(1–2), 44–51. https://doi.org/10.1111/apm.12489

Fabijan, J., Miller, D., Olagoke, O., Woolford, L., Boardman, W., Timms, P., Polkinghorne, A., Simmons, G., Hemmatzadeh, F., Trott, D. J., & Speight, K. N. (2019). Prevalence and clinical significance of koala retrovirus in two

South Australian koala (Phascolarctos cinereus) populations. *Journal of Medical Microbiology*, *68*, 1072–1080. https://doi.org/10.1099/jmm.0.001009

Fabijan, J., Sarker, N., Speight, N., Owen, H., Meers, J., Simmons, G., Seddon, J., Emes, R. D., Tarlinton, R., Hemmatzadeh, F., Woolford, L., & Trott, D. J. (2020). Pathological Findings in Koala Retrovirus-positive Koalas (Phascolarctos cinereus) from Northern and Southern Australia. *Journal of Comparative Pathology*, *176*, 50–66. https://doi.org/10.1016/j.jcpa.2020.02.003

Fabijan, J., Woolford, L., Lathe, S., Simmons, G., Hemmatzadeh, F., Trott, D. J., & Speight, N. (2017). Lymphoma, Koala Retrovirus Infection and Reproductive Chlamydiosis in a Koala (Phascolarctos cinereus). *Journal of Comparative Pathology*, *157*(2–3), 188–192. https://doi.org/10.1016/J.JCPA.2017.07.011

Fiebig, U., Hartmann, M. G., Bannert, N., Kurth, R., & Denner, J. (2006). Transspecies Transmission of the Endogenous Koala Retrovirus. *Journal of Virology*, *80*(11), 5651–5654. https://doi.org/10.1128/JVI.02597-05

Gifford, R., & Tristem, M. (2003). Invited Review The Evolution, Distribution and Diversity of Endogenous Retroviruses *. In *Virus Genes* (Vol. 26, Issue 3).

Gillett, A. K. (2014). An examination of disease in captive Australian koalas (Phascolarctos cinereus) and potential links to koala retrovirus (KoRV). *Technical Reports of the Australian Museum, Online*, *24*, 39–45. https://doi.org/10.3853/j.1835-4211.24.2014.1612

Goff, S. P. (2004). Retrovirus Restriction Factors. *Molecular Cell*, *16*(6), 849–859. https://doi.org/10.1016/J.MOLCEL.2004.12.001

Goff, S. P. (2007). Host factors exploited by retroviruses. *Nature Reviews Microbiology*, *5*(4), 253–263. https://doi.org/10.1038/nrmicro1541

Goff, S. P. (2013). Retroviridae. In D. M. Knipe & P. M. Howley (Eds.), *Fields Virology* (Sixth Edit, pp. 1424–1473). Lippincott Williams & Wilkins.

Goodier, J. L. (2016). Restricting retrotransposons: A review. In *Mobile DNA* (Vol. 7, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s13100-016-0070-z

Gordon, G. (1991). Estimation of the age of the Koala, Phascolarctos cinereus (Marsupialia: Phascolarctidae), from tooth wear and growth. *Australian Mammalogy*, *14*(1), 5. https://doi.org/10.1071/AM91001

Greenwood, A. D., Ishida, Y., O'Brien, S. P., Roca, A. L., & Eiden, M. V. (2018). Transmission, Evolution, and Endogenization: Lessons Learned from Recent Retroviral Invasions. *Microbiology and Molecular Biology Reviews*, *82*(1). https://doi.org/10.1128/MMBR.00044-17

Han, J. S., & Boeke, J. D. (2005). LINE-1 retrotransposons: Modulators of quantity and quality of mammalian gene expression? In *BioEssays* (Vol. 27, Issue 8, pp. 775–784). https://doi.org/10.1002/bies.20257

Hanger, J. J., Bromham, L. D., McKee, J. J., O'Brien, T. M., & Robinson, W. F. (2000). The Nucleotide Sequence of Koala (Phascolarctos cinereus) Retrovirus: a Novel Type C Endogenous Virus Related to Gibbon Ape Leukemia Virus . *Journal of Virology*, *74*(9), 4264–4272. https://doi.org/10.1128/JVI.74.9.4264-4272.2000

Hashem, A., Maetani, F., Enamul, M., Kayesh, H., Eiei, T., Mochizuki, K., Ito, A., Sakurai, H., Asai, T., & Tsukiyama-Kohara, K. (2020). *Transmission of Koala Retrovirus from Parent Koalas to a Joey in a Japanese Zoo*. *94*, 19–20. https://doi.org/10.1128/JVI.00019-20

Hayward, A. (2017). Origin of the retroviruses: when, where, and how? *Current Opinion in Virology*, *25*, 23–27. https://doi.org/10.1016/J.COVIRO.2017.06.006

Hayward, J. A., Tachedjian, M., Kohl, C., Johnson, A., Dearnley, M., Jesaveluk, B., Langer, C., Solymosi, P. D., Hille, G., Nitsche, A., Sánchez, C. A., Werner, A., Kontos, D., Crameri, G., Marsh, G. A., Baker, M. L., Poumbourios, P., Drummer, H. E., Holmes, E. C., … Tachedjian, G. (2020). Infectious KoRV-related retroviruses circulating in Australian bats. *Proceedings of the National Academy of Sciences*, *117*(17), 9529–9536. https://doi.org/10.1073/pnas.1915400117

Hindmarsh, P., & Leis, J. (1999). Retroviral DNA Integration. *Microbiology and Molecular Biology Reviews*, *63*(4), 836–843. https://doi.org/10.1128/MMBR.63.4.836-843.1999

Hinz, A., Miguet, N., Natrajan, G., Usami, Y., Yamanaka, H., Renesto, P., Hartlieb, B., McCarthy, A. A., Simorre, J. P., Göttlinger, H., & Weissenhorn, W. (2010). Structural Basis of HIV-1 Tethering to Membranes by the BST-2/Tetherin Ectodomain. *Cell Host & Microbe*, *7*(4), 314–323. https://doi.org/10.1016/J.CHOM.2010.03.005

Hobbs, M., King, A., Salinas, R., Chen, Z., Tsangaras, K., Greenwood, A. D., Johnson, R. N., Belov, K., Wilkins, M. R., & Timms, P. (2017). Long-read genome sequence assembly provides insight into ongoing retroviral invasion of the koala germline. *Scientific Reports*, *7*(1), 15838. https://doi.org/10.1038/s41598-017-16171-1

Hobbs, M., Pavasovic, A., King, A. G., Prentis, P. J., Eldridge, M. D. B., Chen, Z., Colgan, D. J., Polkinghorne, A., Wilkins, M. R., Flanagan, C., Gillett, A., Hanger, J., Johnson, R. N., & Timms, P. (2014). A transcriptome resource for the koala (Phascolarctos cinereus): Insights into koala retrovirus transcription and sequence diversity. *BMC Genomics*, *15*(1), 1–16. https://doi.org/10.1186/1471-2164-15-786/FIGURES/4

Hofacre, A., & Fan, H. (2010). Jaagsiekte sheep retrovirus biology and oncogenesis. *Viruses*, *2*(12), 2618–2648. https://doi.org/10.3390/v2122618

Holt, M. P., Shevach, E. M., & Punkosdy, G. A. (2013). Endogenous mouse mammary tumor viruses (Mtv): New roles for an old virus in cancer, infection, and immunity. *Frontiers in Oncology*, *3 NOV*(November), 1–10. https://doi.org/10.3389/fonc.2013.00287

Hughes, J. F., & Coffin, J. M. (2004). Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: Implications for human and viral evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(6), 1668–1672. https://doi.org/10.1073/PNAS.0307885100/ASSET/C19B7DF0-0C6E-4471-A810-FF918CEA41E6/ASSETS/GRAPHIC/ZPQ0050437120003.JPEG

Hulme, A. E., Perez, O., & Hope, T. J. (2011). Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(24), 9975–9980. https://doi.org/10.1073/pnas.1014522108

Hunt, H., Fadly, A., Silva, R., & Zhang, H. (2008). Survey of endogenous virus and TVB* receptor status of commercial chicken stocks supplying specific-pathogen-free eggs. *Avian Diseases*, *52*(3), 433–440. https://doi.org/10.1637/8183-112907-Reg.1

Iwatani, Y., Chan, D. S. B., Wang, F., Maynard, K. S., Sugiura, W., Gronenborn, A. M., Rouzina, I., Williams, M. C., Musier-Forsyth, K., & Levin, J. G. (2007). Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic Acids Research*, *35*(21), 7096–7108. https://doi.org/10.1093/NAR/GKM750

Jarrett, O., & Russell, P. H. (1978). Differential growth and transmission in cats of feline leukaemia viruses of subgroups A and B. *International Journal of Cancer*, *21*(4), 466–472. https://doi.org/10.1002/ijc.2910210411

Joyce, B. A., Blyton, M. D. J., Johnston, S. D., Young, P. R., & Chappell, K. J. (2021). Koala retrovirus genetic diversity and transmission dynamics within captive koala populations. *Proceedings of the National Academy of Sciences*, *118*(38). https://doi.org/10.1073/pnas.2024021118

Kinney, M. E., & Pye, G. W. (2016). KOALA RETROVIRUS: A REVIEW. *Journal of Zoo and Wildlife Medicine*, *47*(2), 387–396. https://doi.org/10.1638/2015-0185.1

Kozak, C. A. (2015). Origins of the endogenous and infectious laboratory mouse gammaretroviruses. *Viruses*, *7*(1), 1–26. https://doi.org/10.3390/v7010001

Kurth, R., & Bannert, N. (2010). *Retroviruses*.

Lavignon, M., & Evans, L. (1996). A multistep process of leukemogenesis in Moloney murine leukemia virus-infected mice that is modulated by retroviral

pseudotyping and interference. *Journal of Virology*, *70*(6), 3852–3862. https://doi.org/10.1128/jvi.70.6.3852-3862.1996

Legione, A. R., Patterson, J. L. S., Whiteley, P., Firestone, S. M., Curnick, M., Bodley, K., Lynch, M., Gilkerson, J. R., Sansom, F. M., & Devlin, J. M. (2017). Koala retrovirus genotyping analyses reveal a low prevalence of KoRV-A in Victorian koalas and an association with clinical disease. *Journal of Medical Microbiology*, *66*(2), 236–244. https://doi.org/10.1099/jmm.0.000416

Lenasi, T., Contreras, X., & Peterlin, B. M. (2010). Transcription, Splicing and Transport of Retroviral RNA. In R. Kurth & N. Bannert (Eds.), *Retroviruses* (First Edit, pp. 161–186). Caister Academic Press.

Lilly, F. (1967). Susceptibility to Two Strains of Friend Leukemia Virus in Mice. *Science*, *155*(3761), 461–462. https://doi.org/10.1126/science.155.3761.461

Löber, U., Hobbs, M., Dayaram, A., Tsangaras, K., Jones, K., Alquezar-Planas, D. E., Ishida, Y., Meers, J., Mayer, J., Quedenau, C., Chen, W., Johnson, R. N., Timms, P., Young, P. R., Roca, A. L., & Greenwood, A. D. (2018). Degradation and remobilization of endogenous retroviruses by recombination during the earliest stages of a germ-line invasion. *Proceedings of the National Academy of Sciences*, *115*(34), 8609–8614. https://doi.org/10.1073/pnas.1807598115

Lokossou, A. G., Toudic, C., & Barbeau, B. (2014). Implication of Human Endogenous Retrovirus Envelope Proteins in Placental Functions. *Viruses*, *6*, 4609–4627. https://doi.org/10.3390/v6114609

Luban, J. (2010). Retroviral Restriction Factors. In R. Kurth & N. Bannert (Eds.), *Retroviruses: Molecular Biology, Genomics and Pathogenesis* (January 2010, pp. 285–308).

Magiorkinis, G., Gifford, R. J., Katzourakis, A., De Ranter, J., & Belshaw, R. (2012). Env-less endogenous retroviruses are genomic superspreaders. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(19), 7385–7390. https://doi.org/10.1073/PNAS.1200913109

Mamede, J. I., Cianci, G. C., Anderson, M. R., & Hope, T. J. (2017). Early cytoplasmic uncoating is associated with infectivity of HIV-1. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(34), E7169–E7178. https://doi.org/10.1073/pnas.1706245114

McDougall, A. S., Terry, A., Tzavaras, T., Cheney, C., Rojko, J., & Neil, J. C. (1994). Defective endogenous proviruses are expressed in feline lymphoid cells: evidence for a role in natural resistance to subgroup B feline leukemia viruses. *Journal of Virology*, *68*(4), 2151–2160. https://doi.org/10.1128/jvi.68.4.2151-2160.1994

McEwen, G. K., Alquezar-Planas, D. E., Dayaram, A., Gillett, A., Tarlinton, R., Mongan, N., Chappell, K. J., Henning, J., Tan, M., Timms, P., Young, P. R.,

Roca, A. L., & Greenwood, A. D. (2021). Retroviral integrations contribute to elevated host cancer rates during germline invasion. *Nature Communications*, *12*(1), 1316. https://doi.org/10.1038/s41467-021-21612-7

McKee, J., Clark, N., Shapter, F., & Simmons, G. (2017). A new look at the origins of gibbon ape leukemia virus. *Virus Genes*, *53*(2), 165–172. https://doi.org/10.1007/s11262-017-1436-0

McMichael, L., Smith, C., Gordon, A., Agnihotri, K., Meers, J., & Oakey, J. (2019). A novel Australian flying-fox retrovirus shares an evolutionary ancestor with Koala, Gibbon and Melomys gamma-retroviruses. *Virus Genes*, *55*(3), 421–424. https://doi.org/10.1007/s11262-019-01653-3

Meers, J., Simmons, G., Jones, K., Clarke, D. T. W., & Young, P. R. (2014). Koala retrovirus in free-ranging populations—prevalence. *Technical Reports of the Australian Museum, Online*, *24*, 15–17. https://doi.org/10.3853/j.1835-4211.24.2014.1608

Miyazawa, T., Shojima, T., Yoshikawa, R., & Ohata, T. (2011). Isolation of koala retroviruses from koalas in Japan. *The Journal of Veterinary Medical Science*, *73*(1), 65–70. https://doi.org/10.1292/JVMS.10-0250

Mothes, W., & Uchil, P. D. (2010). Retroviral Entry and Uncoating. In R. Kurth & N. Bannert (Eds.), *Retroviruses* (pp. 107–128).

Mura, M., Murcia, P., Caporale, M., Spencer, T. E., Nagashima, K., Rein, A., & Palmarini, M. (2004). Late viral interference induced by transdominant Gag of an endogenous retrovirus. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(30), 11117. https://doi.org/10.1073/PNAS.0402877101

Murphy, H. W., & Switzer, W. M. (2008). Occupational Exposure to Zoonotic Simian Retroviruses: Health and Safety Implications for Persons Working with Nonhuman Primates. *Zoo and Wild Animal Medicine*, 251–264. https://doi.org/10.1016/B978-141604047-7.50034-8

Nair, S., & Rein, A. (2014). Antiretroviral restriction factors in mice. *Virus Research*, *193*, 130–134. https://doi.org/10.1016/J.VIRUSRES.2014.07.002

Neaves, L. E., Frankham, G. J., Dennison, S., FitzGibbon, S., Flannagan, C., Gillett, A., Hynes, E., Handasyde, K., Helgen, K. M., Tsangaras, K., Greenwood, A. D., Eldridge, M. D. B., & Johnson, R. N. (2016). Phylogeography of the Koala, (Phascolarctos cinereus), and Harmonising Data to Inform Conservation. *PLOS ONE*, *11*(9), e0162207. https://doi.org/10.1371/journal.pone.0162207

Oliveira, N. M., Satija, H., Kouwenhoven, I. A., & Eiden, M. V. (2007). Changes in viral protein function that accompany retroviral endogenization. *Proceedings of the National Academy of Sciences*, *104*(44), 17506–17511. https://doi.org/10.1073/pnas.0704313104

Ortega, C., Valencia, A. C., Duque-Valencia, J., & Ruiz-Saenz, J. (2020). Prevalence and genomic diversity of feline leukemia virus in privately owned and shelter cats in Aburrá Valley, Colombia. *Viruses*, *12*(4). https://doi.org/10.3390/V12040464

Payne, L. N. (1998). Retrovirus-induced disease in poultry. *Poultry Science*, *77*(8), 1204–1212. https://doi.org/10.1093/ps/77.8.1204

Pederson, F., & Sorenson, A. (2010). Pathogensis of Oncoviral Infections. In R. Kurth & N. Bannert (Eds.), *Retroviruses: Molecular Biology, Genomics and Pathogenesis* (January 2010, pp. 237–268).

Pincetic, A., & Leis, J. (2009). The Mechanism of Budding of Retroviruses from Cell Membranes. *Advances in Virology*, *9*. https://doi.org/10.1155/2009/623969

Polani, S., Roca, A. L., Rosensteel, B. B., Kolokotronis, S. O., & Bar-Gal, G. K. (2010). Evolutionary dynamics of endogenous feline leukemia virus proliferation among species of the domestic cat lineage. *Virology*, *405*(2), 397–407. https://doi.org/10.1016/J.VIROL.2010.06.010

Polkinghorne, A., Hanger, J., & Timms, P. (2013). Recent advances in understanding the biology, epidemiology and control of chlamydial infections in koalas. *Veterinary Microbiology*, *165*(3–4), 214–223. https://doi.org/10.1016/j.vetmic.2013.02.026

Powers, J. A., Chiu, E. S., Kraberger, S. J., Roelke-Parker, M., Lowery, I., Erbeck, K., Troyer, R., Carver, S., & VandeWoude, S. (2018). Feline Leukemia Virus (FeLV) Disease Outcomes in a Domestic Cat Breeding Colony: Relationship to Endogenous FeLV and Other Chronic Viral Infections. *Journal of Virology*, *92*(18). https://doi.org/10.1128/JVI.00649-18

Quigley, B. L., Ong, V. A., Hanger, J., Timms, P., & Simon, V. (2018). Molecular Dynamics and Mode of Transmission of Koala Retrovirus as It Invades and Spreads through a Wild Queensland Koala Population. *Journal of Virology*, *92*(5), 1871–1888. https://doi.org/10.1128/JVI.01871-17

Quigley, B. L., Phillips, S., Olagoke, O., Robbins, A., Hanger, J., & Timms, P. (2019). Changes in Endogenous and Exogenous Koala Retrovirus Subtype Expression over Time Reflect Koala Health Outcomes. *Journal of Virology*, *93*(18). https://doi.org/10.1128/JVI.00849-19

Quigley, B. L., & Timms, P. (2020). Helping koalas battle disease-Recent advances in Chlamydia and koala retrovirus (KoRV) disease understanding and treatment in koalas. *FEMS Microbiology Reviews*, *024*, 583–605. https://doi.org/10.1093/femsre/fuaa024

Rahm, N. (2016). TRIM5alpha. *Encyclopedia of AIDS*, 1–9. https://doi.org/10.1007/978-1-4614-9610-6_370-1

Rosenberg N, J. P. (1997). Retroviral Pathogenesis. *Retroviruses. Cold Spring*

Harbor (NY): Cold Spring Harbor Laboratory Press; 1997.

Sarker, N., Fabijan, J., Emes, R. D., Hemmatzadeh, F., Meers, J., Moreton, J., Owen, H., Seddon, J. M., Simmons, G., Speight, N., Trott, D., Woolford, L., & Tarlinton, & R. E. (2018). Identification of stable reference genes for quantitative PCR in koalas OPEN. *SciEnTific RepORts |*, *8*, 3364. https://doi.org/10.1038/s41598-018-21723-0

Sarker, N., Fabijan, J., Owen, H., Seddon, J., Simmons, G., Speight, N., Kaler, J., Woolford, L., Emes, R. D., Hemmatzadeh, F., Trott, D. J., Meers, J., & Tarlinton, R. E. (2020). Koala retrovirus viral load and disease burden in distinct northern and southern koala populations. *Scientific Reports 2020 10:1*, *10*(1), 1–11. https://doi.org/10.1038/s41598-019-56546-0

Sarker, N., Fabijan, J., Seddon, J., Tarlinton, R., Owen, H., Simmons, G., Thia, J., Blanchard, A. M., Speight, N., Kaler, J., Emes, R. D., Woolford, L., Trott, D., Hemmatzadeh, F., & Meers, J. (2019). Genetic diversity of Koala retrovirus env gene subtypes: insights into northern and southern koala populations. *Journal of General Virology*, *100*(9), 1328–1339. https://doi.org/10.1099/jgv.0.001304

Sharp, J. M., & Demartini, J. C. (2003). Natural history of JSRV in sheep. *Current Topics in Microbiology and Immunology*, *275*, 55–79. https://doi.org/10.1007/978-3-642-55638-8_3

Sheehy, A. M., Gaddis, N. C., Choi, J. D., & Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, *418*(6898), 646–650. https://doi.org/10.1038/nature00939

Shi, W., Jia, S., Guan, X., Yao, X., Pan, R., Huang, X., Ma, Y., Wei, J., & Xu, Y. (2021). A survey of jaagsiekte sheep retrovirus (JSRV) infection in sheep in the three northeastern provinces of China. *Archives of Virology*, *166*(3), 831–840. https://doi.org/10.1007/S00705-020-04919-6/FIGURES/6

Shojima, T., Yoshikawa, R., Hoshino, S., Shimode, S., Nakagawa, S., Ohata, T., Nakaoka, R., & Miyazawa, T. (2013). Identification of a Novel Subgroup of Koala Retrovirus from Koalas in Japanese Zoos. *Journal of Virology*, *87*(17), 9943–9948. https://doi.org/10.1128/JVI.01385-13

Simmons, G., Clarke, D., Mckee, J., Young, P., & Meers, J. (2014). *Discovery of a Novel Retrovirus Sequence in an Australian Native Rodent (Melomys burtoni): A Putative Link between Gibbon Ape Leukemia Virus and Koala Retrovirus*. https://doi.org/10.1371/journal.pone.0106954

Simmons, G., Young, P., Hanger, J., Jones, K., Clarke, D., McKee, J., & Meers, J. (2012). Prevalence of koala retrovirus in geographically diverse populations in Australia. *Australian Veterinary Journal*, *90*(10), 404–409. https://doi.org/10.1111/j.1751-0813.2012.00964.x

Speight, K. N., Polkinghorne, A., Penn, R., Boardman, W., Timms, P., Fraser, T., Johnson, K., Faull, R., Bate, S., & Woolford, L. (2016). PREVALENCE

AND PATHOLOGIC FEATURES OF *CHLAMYDIA PECORUM* INFECTIONS IN SOUTH AUSTRALIAN KOALAS (*PHASCOLARCTOS CINEREUS*). *Journal of Wildlife Diseases*, *52*(2), 301–306. https://doi.org/10.7589/2015-05-120

Tandon, R., Cattori, V., Pepin, A. C., Riond, B., Meli, M. L., McDonald, M., Doherr, M. G., Lutz, H., & Hofmann-Lehmann, R. (2008). Association between endogenous feline leukemia virus loads and exogenous feline leukemia virus infection in domestic cats. *Virus Research*, *135*(1), 136–143. https://doi.org/10.1016/J.VIRUSRES.2008.02.016

Tarlinton, R. E., Fabijan, J., Hemmatzadeh, F., Meers, J., Owen, H., Sarker, N., Seddon, J. M., Simmons, G., Speight, N., Trott, D. J., Woolford, L., & Emes, R. D. (2021). Transcriptomic and genomic variants between koala populations reveals underlying genetic components to disorders in a bottlenecked population. *Conservation Genetics*, *22*(3), 329–340. https://doi.org/10.1007/s10592-021-01340-7

Tarlinton, R. E., Legione, A. R., Sarker, N., Fabijan, J., Meers, J., Mcmichael, L., Simmons, G., Owen, H., Seddon, J. M., Dick, G., Ryder, J. S., Hemmatzedah, F., Trott, D. J., Speight, N., Holmes, N., Loose, M., & Emes, R. D. (2022). Differential and defective transcription of koala retrovirus indicates the complexity of host and virus evolution. *Journal of General Virology*, *103*, 1749. https://doi.org/10.1099/jgv.0.001749

Tarlinton, R. E., Meers, J., & Young, P. R. (2006). Retroviral invasion of the koala genome. *Nature*, *442*(7098). https://doi.org/10.1038/nature04841

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. *Journal of General Virology*, *86*(3), 783–787. https://doi.org/10.1099/vir.0.80547-0

Tarlinton, R., Meers, J., & Young, P. (2008). *Biology and evolution of the endogenous koala retrovirus*. https://doi.org/10.1007/s00018-008-8499-y

Tsangaras, K., Siracusa, M. C., Nikolaidis, N., Ishida, Y., Cui, P., Vielgrader, H., Helgen, K. M., Roca, A. L., & Greenwood, A. D. (2014). Hybridization Capture Reveals Evolution and Conservation across the Entire Koala Retrovirus Genome. *PLOS ONE*, *9*(4), e95633. https://doi.org/10.1371/JOURNAL.PONE.0095633

Van Gent, M., Sparrer, K. M. J., & Gack, M. U. (2018). *Annual Review of Virology TRIM Proteins and Their Roles in Antiviral Host Defenses*. https://doi.org/10.1146/annurev-virology

VM Vogt. (1997). Retroviral Virions and Genomes. *Retroviruses. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997*.

Waugh, C. A., Hanger, J., Loader, J., King, A., Hobbs, M., Johnson, R., & Timms, P. (2017). Infection with koala retrovirus subgroup B (KoRV-B), but not

KoRV-A, is associated with chlamydial disease in free-ranging koalas (Phascolarctos cinereus). *Scientific Reports*, *7*(1), 134. https://doi.org/10.1038/s41598-017-00137-4

Xu, W., & Eiden, M. V. (2015). *Koala Retroviruses: Evolution and Disease Dynamics* *. https://doi.org/10.1146/annurev-virology-100114-055056

Xu, W., Gorman, K., Santiago, J., Kluska, K., & Eiden, M. (2015). Genetic Diversity of Koala Retroviral Envelopes. *Viruses*, *7*(3), 1258–1270. https://doi.org/10.3390/v7031258

Xu, W., Stadler, C. K., Gorman, K., Jensen, N., Kim, D., Zheng, H., Tang, S., Switzer, W. M., Pye, G. W., & Eiden, M. V. (2013). An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo. *Proceedings of the National Academy of Sciences*, *110*(28), 11547–11552. https://doi.org/10.1073/pnas.1304704110

Zheng, H., Pan, Y., Tang, S., Pye, G. W., Stadler, C. K., Vogelnest, L., Vinette Herrin, K., Rideout, B. A., & Switzer, W. M. (2020). *Koala retrovirus diversity, transmissibility, and disease associations*. *17*, 34. https://doi.org/10.1186/s12977-020-00541-1

APPENDICES

APPENDIX I

Detailed list of samples obtained from Dr. Micheal Lynch and used for RT-PCR to check for KoRV copy number calculations.

S.No.	DATE	Ear Tag No.	Sample Number	Sex	Weight	Teeth Score	Sample Type
1	10/12/2018	122	A4/122	F	5.1	2	Plasma in RNA later
2	10/11/2018	132	A15/132	F	8.7	5	Plasma in RNA later
3	10/11/2018	139	B10A/139	F	10.1	4B	Plasma in RNA later
4	10/11/2018	140	B12/140	F	7.9	4B	Plasma in RNA later
5	10/11/2018	142	B8A/142	F	9.6	4A	Plasma in RNA later
6	10/11/2018	143	B3/143	F	10.1	7B	Plasma in RNA later
7	10/11/2018	144	A3/144	F	8.2	3	Plasma in RNA later
8	10/11/2018	145	B15/145	F	9.2	4C	Plasma in RNA later
9	10/10/2018	160	B8/160	F	6	4C	Plasma in RNA later
10	10/10/2018	161	B11/161	F	5.1	2	Plasma in RNA later
11	10/09/2018	163	B28/163	F	3.4	1	Plasma in RNA later

12	10/09/2018	164	B27/164	F	4.2	2	Plasma in RNA later
13	10/09/2018	167	B20/167	F	7.6	3	Plasma in RNA later
14	10/09/2018	170	B22/170	F	9.3	4A	Plasma in RNA later
15	10/09/2018	172	B10/172	F	8.2	4C	Plasma in RNA later
16	10/09/2018	176	B5/176	F	9.1	6A	Plasma in RNA later
17	10/09/2018	178	B3/178	F	8	4A	Plasma in RNA later
18	10/08/2018	183	B9/183	F	7.4	4A	Plasma in RNA later
19	10/09/2018	185	B1/185	F	9.1	5	Plasma in RNA later
20	10/08/2018	186	A12/186	F	8.7	5	Plasma in RNA later
21	10/08/2018	189	A11/189	F	8.7	6A	Plasma in RNA later
22	10/08/2018	191	A9/191	F	9.1	4B	Plasma in RNA later
23	10/09/2018	194	B1/194	F	7.2	3	Plasma in RNA later
24	10/08/2018	197	A4/197	F	7.5	4B	Plasma in RNA later
25	10/08/2018	199	10:A2/199R	F	7.2	3	Plasma in RNA later
26	10/08/2018	229	A6/229	M	7.3	4A	Plasma in RNA later

27	10/08/2018	230	A7/230	M	11.1	5	Plasma in RNA later
28	10/08/2018	231	A15/231	M	4.5	1	Plasma in RNA later
29	10/08/2018	232	A8/232	M	12	5	Plasma in RNA later
30	10/08/2018	233	A10/233	M	4.1	2	Plasma in RNA later
32	10/08/2018	235	B8/235	M	10.3	4B	Plasma in RNA later
32	10/08/2018	236	B10/236	M	7	3	Plasma in RNA later
33	10/08/2018	237	A17/237	M	11.1	5	Plasma in RNA later
34	10/08/2018	238	B11/238	M	4.7	2	Plasma in RNA later
35	10/08/2018	239	A19/239	M	6.1	2	Plasma in RNA later
36	10/09/2018	244	B13/244	M	11.5	4A	Plasma in RNA later
37	10/09/2018	245	B12/245	M	9.9	4B	Plasma in RNA later
38	10/09/2018	246	B14/246	M	10.3	4C	Plasma in RNA later
39	10/09/2018	247	B16/247	M	11.6	3	Plasma in RNA later
40	10/09/2018	248	B15/248	M	11.3	5	Plasma in RNA later
41	10/09/2018	249	B17/249	M	12.9	4A	Plasma in RNA later

42	10/09/2018	250	A27/250	M	7.9	2	Plasma in RNA later
43	10/09/2018	252	A28/252	M	11.1	4B	Plasma in RNA later
44	10/09/2018	256	A30/256	M	3.5	1	Plasma in RNA later
45	10/10/2018	260	A3/260	M	11.2	4C	Plasma in RNA later
46	10/10/2018	261	B1/261	M	12.2	6A	Plasma in RNA later
47	10/11/2018	586	B5A/586	M	11.5	4C	Plasma in RNA later
48	10/11/2018	588	B4/588	M	13	3	Plasma in RNA later
49	10/11/2018	589	B6/589	M	11.9	6B	Plasma in RNA later
50	10/10/2018	591	B16/591	M	6.6	4A	Plasma in RNA later

APPENDIX II

Detailed list of the samples obtained from Dr. Alistair Legione and used as a part of his study in (Legione et al., 2017). In this study these were the samples that were used to check for polymorphism of RecKoRV.

Tube Number	Pathology Accession Number	Species	Origin	Sample Type	Sent By
1	W1048-13	Phascolarctos cinereus (koala)	Raymond Island	Spleen	Dr. Alistair Legione
2	W1050-13	Phascolarctos cinereus (koala)	Raymond Island	Spleen	Dr. Alistair Legione
3	W1097-13	Phascolarctos cinereus (koala)	Raymond Island	Spleen	Dr. Alistair Legione
4	W1096-13	Phascolarctos cinereus (koala)	Raymond Island	Spleen	Dr. Alistair Legione
5	W1052-13	Phascolarctos cinereus (koala)	Raymond Island	Spleen	Dr. Alistair Legione
6	W1051-13	Phascolarctos cinereus (koala)	Raymond Island	Spleen	Dr. Alistair Legione
7	W1129-13	Phascolarctos cinereus (koala)	Mount Eccles	Spleen	Dr. Alistair Legione
8	W1141-13	Phascolarctos cinereus (koala)	Cape Otway (South Coast)	Spleen	Dr. Alistair Legione
9	W1143-13	Phascolarctos cinereus (koala)	Cape Otway (South Coast)	Spleen	Dr. Alistair Legione
10	W1144-13	Phascolarctos cinereus (koala)	Cape Otway (South Coast)	Spleen	Dr. Alistair Legione
11	W1147-13	Phascolarctos cinereus (koala)	Cape Otway (South Coast)	Spleen	Dr. Alistair Legione
12	W1149-13	Phascolarctos cinereus (koala)	Cape Otway (South Coast)	Spleen	Dr. Alistair Legione
13	W1150-13	Phascolarctos cinereus (koala)	Cape Otway (South Coast)	Spleen	Dr. Alistair Legione
14	W63-14	Phascolarctos cinereus (koala)	Raymond Island	Spleen	Dr. Alistair Legione
15	W152-14	Phascolarctos cinereus (koala)	Raymond Island	Spleen	Dr. Alistair Legione
16	W1190A-13	Phascolarctos cinereus (koala)	Cape Otway (South Coast)	Spleen	Dr. Alistair Legione
17	W1190B-13	Phascolarctos cinereus (koala)	Cape Otway (South Coast)	Spleen	Dr. Alistair Legione
18	1W1184-13	Phascolarctos cinereus (koala)	Cape Otway (South Coast)	Spleen	Dr. Alistair Legione

19	W383-13	Phascolarctos cinereus (koala)	Flinders (Mornington Peninsula)	Spleen	Dr. Alistair Legione
20	W384-13	Phascolarctos cinereus (koala)	Flinders (Mornington Peninsula)	Spleen	Dr. Alistair Legione
21	W236-12	Phascolarctos cinereus (koala)	Somers (Mornington Peninsula)	Spleen	Dr. Alistair Legione
22	W235-16	Phascolarctos cinereus (koala)	Somers (Mornington Peninsula)	Spleen	Dr. Alistair Legione
23	W19-12	Phascolarctos cinereus (koala)	French Island	Spleen	Dr. Alistair Legione
24	W551-13	Phascolarctos cinereus (koala)	Trafalgar (Gippsland)	Spleen	Dr. Alistair Legione
25	W434-13	Phascolarctos cinereus (koala)	Flinders (Mornington Peninsula)	Spleen	Dr. Alistair Legione
26	W678-13	Phascolarctos cinereus (koala)	Rawson (Gippsland)	Spleen	Dr. Alistair Legione
27	W1078-13	Phascolarctos cinereus (koala)	Traralgon South (Gippsland)	Spleen	Dr. Alistair Legione
28	W505-12	Phascolarctos cinereus (koala)	Sandy Point (Gippsland)	Spleen	Dr. Alistair Legione
29	W305-12	Phascolarctos cinereus (koala)	Bolwarra (Far West)	Spleen	Dr. Alistair Legione
30	W183-13	Phascolarctos cinereus (koala)	Mirboo North (Gippsland)	Spleen	Dr. Alistair Legione
31	W228-18	Phascolarctos cinereus (koala)	Strzelecki (Gippsland)	Spleen	Dr. Alistair Legione
32	W281-11	Phascolarctos cinereus (koala)	Bolwarra (Far West)	Spleen	Dr. Alistair Legione
33	W512B-11	Phascolarctos cinereus (koala)	Somers (Mornington Peninsula)	Spleen	Dr. Alistair Legione
34	W513-11	Phascolarctos cinereus (koala)	Somers (Mornington Peninsula)	Spleen	Dr. Alistair Legione
35	W512A-11	Phascolarctos cinereus (koala)	Somers (Mornington Peninsula)	Spleen	Dr. Alistair Legione
36	W448-11	Phascolarctos cinereus (koala)	Leongatha (Gippsland)	Spleen	Dr. Alistair Legione
37	W245-11	Phascolarctos cinereus (koala)	Somers (Mornington Peninsula)	Spleen	Dr. Alistair Legione

38	W452-14	Phascolarctos cinereus (koala)	Mywee (Far North)	Spleen	Dr. Alistair Legione
39	W603-14	Phascolarctos cinereus (koala)	Mywee (Far North)	Spleen	Dr. Alistair Legione
40	W669-14	Phascolarctos cinereus (koala)	Koonoomoo (Far North)	Spleen	Dr. Alistair Legione
41	W343-15	Phascolarctos cinereus (koala)	Koonoomoo (Far North)	Spleen	Dr. Alistair Legione
42	W452-14	Phascolarctos cinereus (koala)	Koonoomoo (Far North)	Spleen	Dr. Alistair Legione
43	W453-15	Phascolarctos cinereus (koala)	Mywee (Far North)	Spleen	Dr. Alistair Legione
44	W984-15	Phascolarctos cinereus (koala)	Sandy Point (Gippsland)	Spleen	Dr. Alistair Legione
45	W250-16	Phascolarctos cinereus (koala)	Koonoomoo (Far North)	Spleen	Dr. Alistair Legione
46	W1045-15	Phascolarctos cinereus (koala)	Koonoomoo (Far North)	Spleen	Dr. Alistair Legione
47	W171-16	Phascolarctos cinereus (koala)	Casterton (Far West)	Spleen	Dr. Alistair Legione
48	W523-15	Phascolarctos cinereus (koala)	Neerim South (Gippsland)	Spleen	Dr. Alistair Legione
49	W28-15	Phascolarctos cinereus (koala)	Leongatha (Gippsland)	Spleen	Dr. Alistair Legione
50	W11-15	Phascolarctos cinereus (koala)	Mywee (Far North)	Spleen	Dr. Alistair Legione
51	W930-15	Phascolarctos cinereus (koala)	Raymond Island	Spleen	Dr. Alistair Legione
52	W946-15	Phascolarctos cinereus (koala)	Raymond Island	Spleen	Dr. Alistair Legione
53	W280-16	Phascolarctos cinereus (koala)	Koonoomoo (Far North)	Spleen	Dr. Alistair Legione
54	W21-16	Phascolarctos cinereus (koala)	Cape Otway (South Coast)	Spleen	Dr. Alistair Legione