The Koala and its Retroviruses:
Implications for Sustainability and Survival

edited by

Geoffrey W. Pye, Rebecca N. Johnson, and Alex D. Greenwood

nature culture **discover**

Australian **museum**

Koala Retrovirus (KoRV):
Molecular Biology and Evolution

Alex D. Greenwood*[1] and Alfred L. Roca[2]

[1] Department of Wildlife Diseases, Leibniz Institute for Zoo and Wildlife Research,
10315 Berlin, Germany

[2] Department of Animal Sciences, University of Illinois at Urbana-Champaign,
Urbana, IL 61801, United States of America
greenwood@izw-berlin.de · roca@illinois.edu

Abstract. The koala retrovirus (KoRV) is in transition between occurring as an exogenous retrovirus spread by infection and becoming an endogenous retrovirus spread primarily as part of the host germ line. While up to 10% of mammalian genomes are composed of such endogenous retroviruses (ERVs), KoRV is the only known example of a retrovirus in the process of making this transition. Thus, it presents a singular opportunity to study the host-pathogen interactions involved during retroviral invasion of a vertebrate germ line.

Overview of KoRV molecular biology

KoRV is most similar genetically to the gibbon ape leukemia virus (GALV). However, biologically, they are quite different. GALV is a highly aggressive oncogenic virus whereas KoRV, while associated with leukemia in koalas, is not as infectious (Hanger *et al*., 2000). Molecular studies of the genetic differences between KoRV and GALV have demonstrated that specific mutations likely account for the decreased pathogenicity of KoRV relative to GALV (Oliveira *et al*., 2006, Oliveira et al., 2007). Thus, replacing important GALV domains with their KoRV homologues decreases the infectivity of the GALV recombinants. Recently, however, an infectious KoRV clone has been developed that, despite its molecular differences with GALV, remains quite capable of infecting and replicating in tissue cultures from many mammalian species (Shojima *et al*., 2013). This KoRV clone provides a novel resource for comparative studies. Since this KoRV clone can infect a wide variety of mammalian cell types, it is not clear why KoRV has only been detected in koalas and is not more widely distributed.

KoRV titres are positively correlated with infection by the bacterial pathogen *Chlamydia*, which has a severe impact on koala health (Tarlinton *et al*., 2008). The interaction between KoRV and *Chlamydia* in terms of koala health needs to be clarified, in order to design appropriate interventions. Currently, research on KoRV and *Chlamydia* occur largely independently of one another, although both would benefit from coordination of efforts.

Genomically, KoRV integration sites vary across infected individuals, most likely KoRV inserts largely at random across the genome of koalas (Hanger *et al*., 2000; Tarlinton *et al*., 2006). In northern Australian koala populations, some copies of KoRV are found at the same locus across individuals, suggesting that the virus has been vertically transmitted as an endogenous retrovirus, i.e., has become part of the germ line. The KoRV genome is highly but variably expressed in tissues of infected individuals, as is common for exogenous and endogenous retroviruses alike (Seifarth *et al*., 2005). Among various other ERVs, expression of endogenous proviruses may evolve to benefit the host species, e.g., through development of novel gene

* author for correspondence

function or protection from infection by other retroviruses (Blikstad *et al*., 2008). However, in the case of KoRV, there is currently no evidence of a positive benefit to koalas from the ongoing endogenization of proviruses. Conversely, a detrimental role for KoRV has not been conclusively established either, and the mechanisms by which KoRV may be involved in enabling *Chlamydia* infections or in tumor formation are not well characterized.

Evolution of KoRV

Evolutionary analysis of KoRV suggests that it is most closely related to GALV, more distantly to pig endogenous retroviruses (PERVs), and embedded overall in a clade of retroviruses that include the murine leukemia viruses (MLVs) (Hanger *et al*., 2000). These comprise the gamma-retroviruses, a genus of the Retroviridae, with many members known to cause diseases, particularly cancers. The phylogeny is somewhat surprising as GALV has only been detected in captive or small introduced free-ranging populations of Southeast Asian gibbons, while KoRV is found in both captive and free-ranging Australian koalas (Reitz *et al*., 1979). The host species do not overlap in geographic distribution, which are separated by deep seas and were not connected by land bridges in the past. This would likely rule out the direct transfer of a virus from gibbon to koala, and may suggest transfer from a third species, rodent or perhaps bat. Bats are known to carry retroviruses similar to GALV and KoRV (Cui *et al*., 2012). Rodents, and particularly the diverse species of *Mus* from Southeast Asia, are also potential reservoirs: GALV-like sequences have been detected in *Mus caroli* and *Mus cervicolor* (Lieber *et al*., 1975; Benveniste *et al*., 1977). Modern sequencing efforts have not followed up on this rodent research, conducted nearly 40 years ago. Further progress in understanding the evolutionary origins, diversity and epidemiology of the KoRV/GALV family is of particular interest given the broad diversity of species infected by related viruses and considering that GALV is a pathogenic retrovirus that infects higher primates.

How and when did KoRV appear in the koala population? Until recently it was thought that the ancestor of KoRV may have entered the koala population as recently as ca. 200 years ago, and then spread from northern Australia southward (Tarlinton *et al*., 2008). The geographic spread of KoRV is incomplete since KoRV is ubiquitous among northern koalas but absent in some southern populations. In many northern locations 100% of koalas are positive for KoRV. Until recently it was thought that some parts of the south, and in particular islands off the southern Australian coast, were completely KoRV free. However, recent research suggests that, while low in prevalence, KoRV may be present in most if not all southern populations (Simmons *et al*., 2012). Whether this represents historical or recent introduction of KoRV to southern populations is unclear. In northern koalas, individuals may share the same KoRV integration sites (Hanger *et al*., 2000; Tarlinton *et al*., 2006). This suggests that KoRV, while persisting as an exogenous retrovirus, also exists as an endogenous retrovirus. Of course, since some koalas are completely KoRV-free, there are not yet any fixed ERVs, those found at the same chromosomal location in all members of a species. This lack of fixation is an indication that the process of retroviral endogenization by KoRV in koalas is a recent one (relative to many ERVs found in other taxa) that is still underway.

Recent analysis by our group suggests that while the overall evolutionary trajectory postulated for KoRV is largely correct, the time frame appears to have been underestimated (Ávila-Arcos *et al*., 2013). Using museum samples dating from the late 1800s to present, we demonstrated using next generation sequencing (NGS) that KoRV was widespread in northern Australian koalas by the 1800s. Since KoRV was already ubiquitous among koalas in northern Australia by the late 1800s (close to the previously postulated time at which the virus first infected koalas), it seems likely that the initial infection of koalas by KoRV occurred far earlier than 200 years ago.

We also found that evolution of the provirus across this time has been extremely slow (Ávila-Arcos *et al*., 2013). Only minor mutations that appeared to be individual-specific were found in the region of *env* coding for the receptor-binding domain. The receptor binding domain of the virus is exposed to the host immune system and is thus under the most pressure to make compensatory changes. The slow evolutionary rate of KoRV suggests that it has been under limited pressure to evolve, and also suggests that koalas have been affected for an extended period of time by harmful KoRV strains that reduce host fitness. We recently demonstrated by examining mitochondrial DNA variation from museum samples that genetic diversity in koalas has been low for more than a century, and it is possible that KoRV pathogenicity has persisted for this time due to the limited genetic diversity of the host species, which could preclude substantial increases in host fitness in response to KoRV (Tsangaras *et al*., 2012). Selective sweeps of resistance to KoRV would not be possible if koalas lack the genetic variation necessary to mediate differences in fitness. On a more positive note, the slow evolutionary rate of KoRV may also suggest that vaccines targeting viral proteins face limited diversity in the targeted proteins, increasing the chance that vaccinated koalas may be successfully protected from infection.

Looking forward

For retroviruses such as KoRV that can form quasi-species and that exist in the genome at high copy number, complex analytical tools may be required. Analytical problems are compounded for historical samples since DNA damage can introduce substantial amounts of artifactual variation. As described in the previous section, NGS can provide novel data and insights; however, new methods are showing even greater promise.

Hybridization capture is a method whereby PCR products or synthesized oligonucleotides (similar to PCR primers) can be used as "baits" to "fish" out or capture related sequences from genomic DNA or reverse transcribed RNA libraries (Maricic *et al*., 2010). Such methods have several advantages over PCR amplification, especially as applied to retroviruses or to ancient DNA (Fig. 1). First, while PCR is very sensitive, it is susceptible to false negatives due to primer-target mismatch. Second, for PCR to work, DNA molecules must be present that are the size of the target amplicon, typically at least 50 bp and often much larger. By contrast, the PCR products or oligonucleotides used for enrichment can bind their targets even if the target DNA is much shorter or contains mismatches. This is important in particular for retroviruses since different virus copies may be quite divergent. Also, short target DNA fragments typical of historical samples (often no larger than 100 bp and sometimes under 20 bp) can be readily retrieved by hybridization. The number of targets that can be "baited" in a given experiment is effectively unlimited. Coupled with NGS, many kilobases or tens of kilobases of target DNA

Library production

Adaptor ligation

Koala DNA- Museum or modern

Bait production

Generate PCR products from genomic DNA, fragment to 200-800 bp, ligate biotin adaptors

Bind biotinylated baits to magnetic beads

Capture: Mix bait and target and wash

Purify/wash

Non-target sequences lost

Desired sequences captured

Non-KoRV sequences lost, Novel sequences enriched

Amplify enriched library

Hybridize library and baited beads

Illumina sequence and assemble sequences to generate consensus
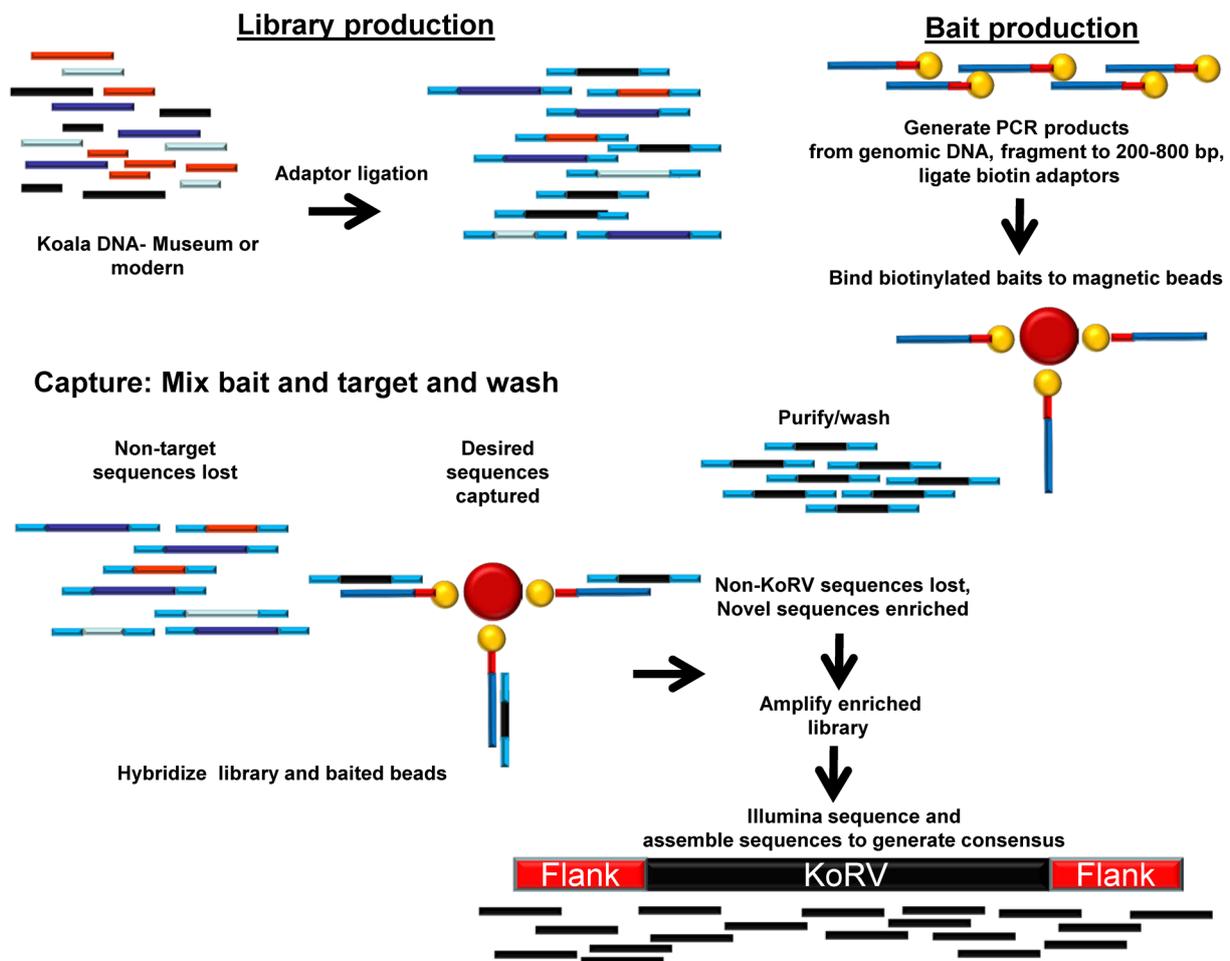
| Flank | KoRV | Flank |

Figure 1. An illustration of the steps involved in hybridization capture. Library preparation is standardized and requires DNA fragments of limited size. For modern samples, DNA must be fragmented to fit the minimum and maximum requirements of the NGS library insert size recommended. For ancient or historical DNA, fragmentation is often not necessary because the DNA is often highly degraded. Adaptors (shown as light blue) are ligated on both sides of each DNA strand. These will be used to amplify the libraries in later steps and, when multiple samples are used, to also identify each library, so that sequences can be separated by sample using the identifiers in a post-sequencing bioinformatics routine. "Bait" preparation involves generating PCR products for the desired targets. It can involve amplifying a long fragment and then shearing to ca. 200 bp, or generating multiple short amplicons. Regardless of how the bait is generated, the bait sequences will be ligated to a biotinylated adaptor (yellow circles) attached to an adaptor sequence (red). The biotin linked baits are then attached to streptavidin-coated magnetic beads. Baits and target libraries are then hybridized together. DNA in the library with sequences complementary to the baits will bind to them; non-matching sequences will not. The target-bait hybridized beads are then isolated using a magnet while the unhybridized DNA is washed off. Using heat or NaOH, the captured strands are separated from the baits resulting in a highly purified library representing the desired target, which is then sequenced using next-generation methods. Sequence reads are aligned to a reference sequence; in this case, KoRV.

can be retrieved during a single experiment. By appending a short fragment of DNA containing a specimen identifier to each set of captured DNA, multiple samples can be processed in one NGS run. Illumina and 454 GS FLX technologies provide ample read length and high throughput, suitable for a broad variety of experimental needs. Such techniques have been extremely successful for ancient DNA and virological studies, including the sequencing of *Yersina pestis* genomes from 500 year old plague pits (Schuenemann *et al.*, 2011), and identifying variation and integration sites for the Merkel cell polyomavirus in formalin fixed tissues (Duncavage *et al.*, 2011).

We are applying these methods to investigate the evolution of KoRV. These are proving superior to our previous analysis by PCR and NGS and have yielded full KoRV genomes from modern and museum specimens and have determined integration site variation (Tsangaras *et al.*, 2014). These methods can reveal the full palette of variation at any given site of the KoRV genome in any individual. KoRV present across individuals at the same locus, which likely represent endogenous proviruses, can be examined for their presence or absence in museum samples and further investigated. We have also applied the technique to rodents that could harbor relatives of KoRV and GALV with promising leads regarding the reservoirs for some known GALV strains. As these methods are in their infancy, they will develop further and will enable studies of KoRV molecular evolution to advance rapidly.

References

Ávila-Arcos, M. C., S. Y. Ho, Y. Ishida, N. Nikolaidis, K. Tsangaras, K. Honig, R. Medina, M. Rasmussen, S. L. Fordyce, S. Calvignac-Spencer, E. Willerslev, M. T. Gilbert, K. M. Helgen, A. L. Roca, and A. D. Greenwood. 2013. One hundred twenty years of koala retrovirus evolution determined from museum skins. *Molecular Biology and Evolution* 30(2): 299–304.
http://dx.doi.org/10.1093/molbev/mss223

Benveniste, R. E., R. Callahan, C. J. Sherr, V. Chapman, and G. J. Todaro. 1977. Two distinct endogenous type C viruses isolated from the Asian rodent *Mus cervicolor*: conservation of virogene sequences in related rodent species. *Journal of Virology* 21(3): 849–862.

Blikstad, V., F. Benachenhou, G. O. Sperber, and J. Blomberg. 2008. Evolution of human endogenous retroviral sequences: a conceptual account. *Cellular and Molecular Life Sciences* 65(21): 3348–3365.
http://dx.doi.org/10.1007/s00018-008-8495-2

Cui, J., M. Tachedjian, L. Wang, G. Tachedjian, L. F. Wang, and S. Zhang. 2012. Discovery of retroviral homologs in bats: implications for the origin of mammalian gammaretroviruses. *Journal of Virology* 86(8): 4288–4293.
http://dx.doi.org/10.1128/JVI.06624-11

Duncavage, E. J., V. Magrini, N. Becker, J. R. Armstrong, R. T. Demeter, T. Wylie, H. J. Abel, and J. D. Pfeifer. 2011. Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. *Journal of Molecular Diagnostics* 13(3): 325–333.
http://dx.doi.org/10.1016/j.jmoldx.2011.01.006

Hanger, J. J., L. D. Bromham, J. J. McKee, T. M. O'Brien, and W. F. Robinson. 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.
http://dx.doi.org/10.1128/JVI.74.9.4264-4272.2000

Lieber, M. M., C. J. Sherr, G. J. Todaro, R. E. Benveniste, R. Callahan, and H. G. Coon. 1975. Isolation from the Asian mouse *Mus caroli* of an endogenous type C virus related to infectious primate type C viruses. *Proceedings of the National Academy of Sciences, USA* 72(6): 2315–2319.
http://dx.doi.org/10.1073/pnas.72.6.2315

Maricic, T., M. Whitten, and S. Pääbo. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5(11): e14004.
http://dx.doi.org/10.1371/journal.pone.0014004

Oliveira, N. M., K. B. Farrell, and M. V. Eiden. 2006. In vitro characterization of a koala retrovirus. *Journal of Virology* 80(6): 3104–3107.
http://dx.doi.org/10.1128/JVI.80.6.3104-3107.2006

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

Reitz, M. S., Jr., F. wong-Staal, W. A. Haseltine, D. G. Kleid, C. D. Trainor, R. E. Gallagher, and R. C. Gallo. 1979. Gibbon ape leukemia virus-Hall's Island: new strain of gibbon ape leukemia virus. *Journal of Virology* 29(1): 395–400.

Schuenemann, V. J., K. Bos, S. DeWitte, S. Schmedes, J. Jamieson, A. Mittnik, S. Forrest, B. K. Coombes, J. W. Wood, D. J. Earn, W. White, J. Krause, and H. N. Poinar. 2011. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proceedings of the National Academy of Sciences, USA* 108(38): E746–752.
http://dx.doi.org/10.1073/pnas.1105107108

Seifarth, W., O. Frank, U. Zeilfelder, B. Spiess, A. D. Greenwood, R. Hehlmann, and C. Leib-Mosch. 2005. Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. *Journal of Virology* 79(1): 341–352.
http://dx.doi.org/10.1128/JVI.79.1.341-352.2005

Shojima, T., S. Hoshino, M. Abe, J. Yasuda, H. Shogen, T. Kobayashi, and T. Miyazawa. 2013. Construction and characterization of an infectious molecular clone of koala retrovirus. *Journal of Virology* 87(9): 5081–5088.
http://dx.doi.org/10.1128/JVI.01584-12

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

Tarlinton, R. E., J. Meers, and P. R. Young. 2006. Retroviral invasion of the koala genome. *Nature* 442(7098): 79–81.
http://dx.doi.org/10.1038/nature04841

Tarlinton, R. E., J. Meers, and P. R. Young. 2008. Biology and evolution of the endogenous koala retrovirus. *Cellular and Molecular Life Sciences* 65: 3413–3421.
http://dx.doi.org/10.1007/s00018-008-8499-y

Tsangaras, K., M. C. Ávila-Arcos, Y. Ishida, K. M. Helgen, A. L. Roca, and A. D. Greenwood. 2012. Historically low mitochondrial DNA diversity in koalas *(Phascolarctos cinereus)*. *BMC Genet* 13: 92.
http://dx.doi.org/10.1186/1471-2156-13-92

Tsangaras, K., M. C. Siracusa, N. Nikolaidis, Y. Ishida, P. Cui, H. Vielgrader, K. M. Helgen, A. L. Roca, and A. D. Greenwood. 2014. Hybridization capture reveals evolution and conservation across the entire koala retrovirus genome. *PLoS One* 9(4): e95633.
http://dx.doi.org/10.1371/journal.pone.0095633