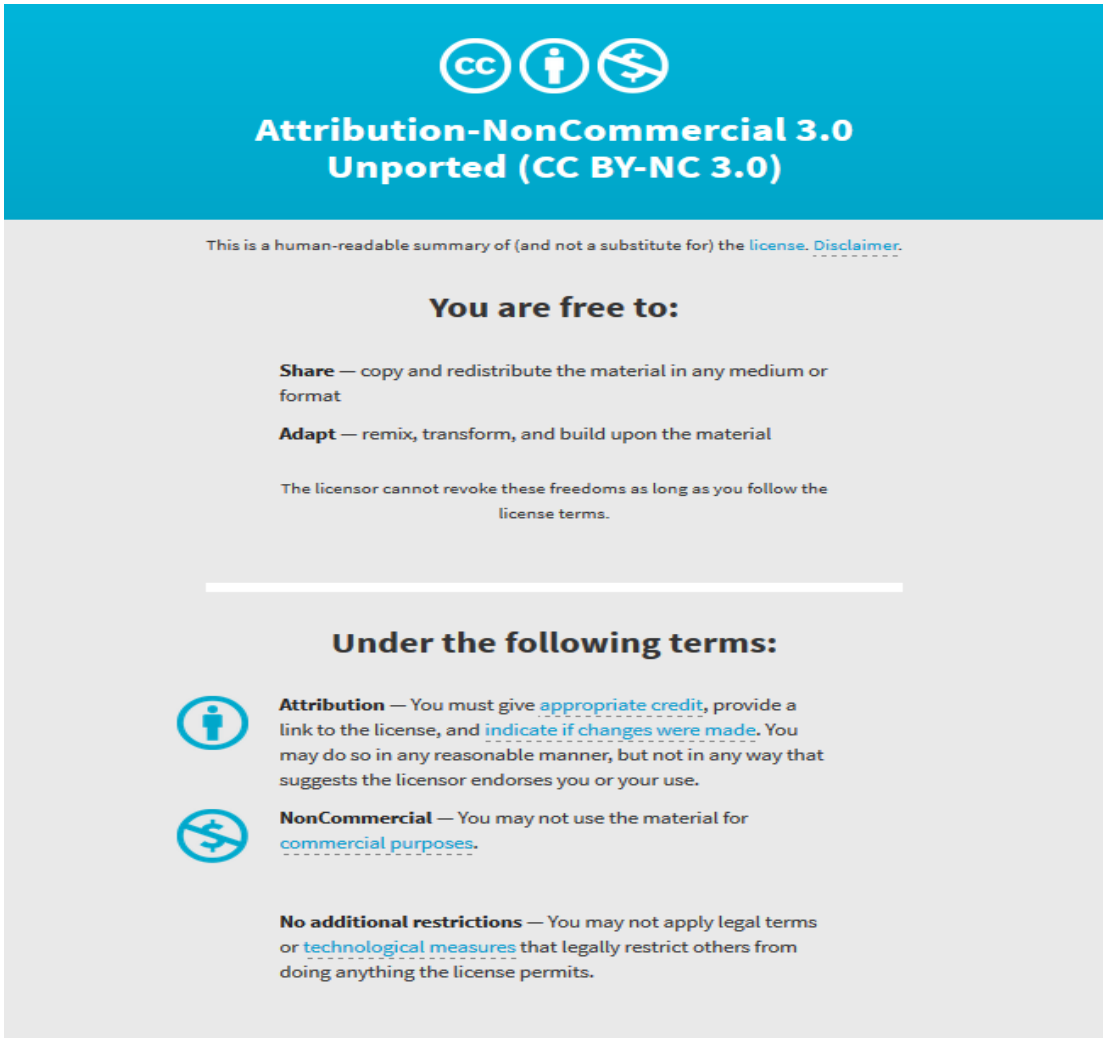
PUBLISHED VERSION

María C. Ávila-Arcos, Simon Y.W. Ho, Yasuko Ishida, Nikolas Nikolaidis, Kyriakos Tsangaras, Karin Hönig, Rebeca Medina, Morten Rasmussen Sarah L. Fordyce, Sébastien Calvignac-Spencer, Eske Willerslev, M. Thomas P. Gilbert, Kristofer M. Helgen, Alfred L. Roca, Alex D. Greenwood
One hundred twenty years of koala retrovirus evolution determined from museum skins
Molecular Biology and Evolution, 2013; 30(2):299-304

Originally published at:
http://doi.org/10.1093/molbev/mss223

PERMISSIONS

http://creativecommons.org/licenses/by-nc/3.0/



21 June 2017

One Hundred Twenty Years of Koala Retrovirus Evolution Determined from Museum Skins

María C. Ávila-Arcos,[†,1] Simon Y.W. Ho,[†,2] Yasuko Ishida,[†,3] Nikolas Nikolaidis,[†,4,5] Kyriakos Tsangaras,[6] Karin Hönig,[6] Rebeca Medina,[4,5] Morten Rasmussen,[1] Sarah L. Fordyce,[1] Sébastien Calvignac-Spencer,[7] Eske Willerslev,[1] M. Thomas P. Gilbert,[1] Kristofer M. Helgen,[8] Alfred L. Roca,[*,3] and Alex D. Greenwood[*,6]

[1]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

[2]School of Biological Sciences, University of Sydney, New South Wales, Australia

[3]Department of Animal Sciences, University of Illinois at Urbana-Champaign

[4]Department of Biological Science, California State University, Fullerton

[5]Center for Applied Biotechnology Studies, California State University, Fullerton

[6]Leibniz-Institute for Zoo and Wildlife Research, Berlin, Germany

[7]Robert-Koch-Institut, Emerging Zoonoses, Berlin, Germany

[8]National Museum of Natural History, Smithsonian Institution, Washington, DC

[†]These authors contributed equally to this work.

***Corresponding authors:** E-mail: greenwood@izw-berlin.de; roca@illinois.edu.

Associate editor: Beth Shapiro

Abstract

Although endogenous retroviruses are common across vertebrate genomes, the koala retrovirus (KoRV) is the only retrovirus known to be currently invading the germ line of its host. KoRV is believed to have first infected koalas in northern Australia less than two centuries ago. We examined KoRV in 28 koala museum skins collected in the late 19th and 20th centuries and deep sequenced the complete proviral *envelope* region from five northern Australian specimens. Strikingly, KoRV *env* sequences were conserved among koalas collected over the span of a century, and two functional motifs that affect viral infectivity were fixed across the museum koala specimens. We detected only 20 *env* polymorphisms among the koalas, likely representing derived mutations subject to purifying selection. Among northern Australian koalas, KoRV was already ubiquitous by the late 19th century, suggesting that KoRV evolved and spread among koala populations more slowly than previously believed. Given that museum and modern koalas share nearly identical KoRV sequences, it is likely that koala populations, for more than a century, have experienced increased susceptibility to diseases caused by viral pathogenesis.

Key words: KoRV, *Phascolarctos cinereus*, endogenous retroviruses, ancient DNA.

Letter

Although endogenous retroviruses (ERVs) are common components of vertebrate genomes (Weiss 2006), most are remnants of retroviral invasions of host germ lines that occurred in the distant past, making it difficult to reconstruct the events involved in their endogenization. The koala retrovirus (KoRV) is unique as the only known retrovirus currently in the process of invading the germ line of its host species, the koala (*Phascolarctos cinereus*), enabling prospective studies of the process of retroviral endogenization (Tarlinton et al. 2006). KoRV was recognized as an ERV in koalas from northern Australia based on shared restriction-digest patterns, indicative of integration at the same locus (Fiebig et al. 2006; Tarlinton et al. 2006). As endogenous components of the germ line, KoRV proviruses would be subject to Mendelian inheritance. Yet KoRV is also subject to horizontal transmission (Hanger et al. 2000). KoRV can reach high titers in infected koalas and likely causes immunosuppression (Fiebig et al. 2006), leading to susceptibility to secondary and often fatal pathologies such as *Chlamydia* infection or leukemias (Hanger et al. 2000; Tarlinton et al. 2005; Fiebig et al. 2006).

Although all koalas in northern Australia are infected by KoRV, the virus is not ubiquitous in south-eastern Australia or in southern Australian islands (Tarlinton et al. 2006), suggesting that KoRV entered the koala population recently in northern Australia and has been spreading south (Tarlinton et al. 2006). A cross-species transfer of KoRV into northern Australian koalas was proposed to have occurred as recently as within the last two centuries (Tarlinton et al. 2008), placing the spread and endogenization of KoRV within the scope of ancient DNA analysis. We therefore sought to examine the evolution of KoRV using museum specimens of koalas.

DNA was extracted in dedicated ancient DNA facilities from 29 koala skins collected from the late 19th century to the 1980s (table 1). Mitochondrial DNA (mtDNA) was sequenced from 18 koalas. Three of the samples for which

Table 1. Koala Museum Samples.

Koala Sample Origins	Museum Number	Collection Date[a]	*pol*	*env3*[b]	*env4*[b]	mtDNA	Full *env* Sequence
Northern Australian							
Bohusläns Museum	um3435	1891	+[c]	+	+	+	Y
Goteborg Museum	collan18193	1870–1891	+	–	+	–	
	maex1738	1870–1891	+	+	+	+	Y
Kansas University Museum	159224	1980s	+	+	+	+	Y
Museum of Comparative Zoology	MCZ 12454[d]	1904	+	+	+	+	Y
	MCZ 8574[d]	1904	+	+	+	+	
Museum of Victoria	c2831	1923	+	+	+	+	
	c2832	1923	+	+	+	+[e]	
Queensland Museum	QM J2377	1915	–	–	–	–	
	QM J6480	1938	+	+	+	+	
	QM J7209	1945	+	+	+	+	
	QM J8353	1952	–	–	–	–	
	QM JM1875	1960s	+	+	+	+	
	QM JM64	1973	+	+	+	+	
	Number 424	1970–1980s	–	–	–	+	
	Number 6121	1970–1980s	+	+	–	+	
	Number 7463	1970–1980s	+	+	+	+	
	Number 7625	1970–1980s	+	+	+	+	
Royal Ontario Museum	9111010180	1891	–	–	–	–	
Stockholm Museum	582119	1911	+	+	+	+	Y
University of Michigan Museum of Zoology	122553	1966	–	–	–	–	
	124673	1977	–	–	–	–	
Southern Australian							
New South Wales							
Australian Museum	AM M17311	1883	–	–	–	–	
	AM M17299	1883	–	–	–	–	
	AM M17300	1883	–	–	–	–	
	AM B4593	1884	–	–	–	+[e]	
	AM M1461	1899	–	–	–	+	
	AM M12482	1971	+	–	+	+	
Victoria							
Australian Museum	AM M4841	1930	–	–	–	–	

NOTE.—For museum specimens, the authors thank F. Johansson and G. Nilson (Bohusläns Museum), R. Timm (Natural History Museum, University of Kansas), J. Chupasko and H. Hoekstra (Harvard Museum of Comparative Zoology), W. Longmore (Museum of Victoria), O. Grönwall and U. Johansson (Swedish Natural History Museum), J. Eger (Royal Ontario Museum), S. Hinshaw (University of Michigan Museum of Zoology), D. Stemmer and C. Kemper (South Australian Museum), S. Ingleby (Australian Museum), and S. Van Dyck and H. Janetzki (Queensland Museum).

[a]Collection dates were examined by K.H. The date ranges listed are as exact as possible based on museum records.

[b]See supplementary table S1, Supplementary Material online, for primer combinations.

[c]The "+" and "–" symbols represent successful or unsuccessful amplification (and sequencing). For full *env* sequence, Y ("yes") is indicated for the five individuals attempted (each was successful).

[d]MCZ 12454 and 8574 are the same koala individual.

[e]These mtDNAs could only be amplified once, despite multiple PCR attempts.

mtDNA amplified were not positive for KoRV; two of these were from southern Australia. Of 16 northern Australian koalas for which at least one PCR was successful, 15 were positive for KoRV. KoRV was ubiquitous among northern Australian koalas more than a century ago, suggesting that KoRV had already spread across northern Australia by the late 1800s and that it has subsequently spread quite slowly. The limited spread of KoRV may be consistent with the sedentary and solitary behavioral ecology of koalas, which have infrequent and short interactions between individuals even during the breeding season (Mitchell 1991; Ellis and Bercovitch 2011). It would also be consistent with the limited dispersal of koalas suggested by the geographic distribution of mtDNA haplotypes (Houlden et al. 1999) or with the disruption of contiguous koala range by human hunting or habitat destruction.

We determined the 1980 bp full-length *env* region sequence from five koalas positive for KoRV (table 1), with a bias toward older samples that might reveal the earliest KoRV variants present in koalas. A multiplex strategy coupled with FLX deep sequencing was employed (Krause et al. 2006), as thoroughly described in supplementary fig. S1, Supplementary Material online. We compared sequences of KoRV in the museum samples to those previously obtained

from modern koalas. Across all individuals, at all but one position along the *env* gene, we detected a character state that matched that previously reported for modern KoRV (accession number DQ174772). Despite the presence of polymorphisms, for all museum koalas, one of the two character states present at each *env* polymorphism matched the sequence present in modern KoRV (Hanger et al. 2000), except at a single nucleotide site. This indicated that proviral *env* sequences had been conserved for more than a century. The slow evolutionary rate for KoRV *env* may be due to our methods primarily identifying endogenous KoRV proviruses, which would have been subject to slower (host) evolutionary rates than exogenous viruses.

The full proviral *env* region could not be amplified as a single PCR product, so it was not possible to determine the phase of the polymorphisms to perform phylogenetic analysis. Nonetheless, actual polymorphisms were distinguished from DNA damage using a conservative approach (described in supplementary fig. S1 and supplementary methods, Supplementary Material online). A total of 20 *env* polymorphic sites were found among the KoRV sequences from museum specimens (fig. 1). Yet only one of the polymorphisms detected was present across all museum koalas, while two others were shared by more than one koala; the remaining 17 were unique to individual koala samples. Thus, there was a dearth of sequence diversity for KoRV *env* among the museum specimens, contrasting with many ERVs in other species that exhibit multiple subtypes of highly divergent viral strains. Although numerous exogenous transmissions and multiple invasions of the germ line appear to have occurred in koalas (Tarlinton et al. 2006), our results suggest that the various transmissions or multiple germ-line invasions likely involved KoRV retroviruses that had similar sequences.

A comparison with previously reported sequences found that at only three of the sites (positions 243, 560, and 733, fig. 1) were the same polymorphisms detected in both museum and modern koala KoRV sequences among regions of sequence overlap (multiple modern KoRV sequences were available for ~900 bp that overlapped 8 of the 20 polymorphic sites identified in museum samples) (Tarlinton 2006). The putative ancestral state (shared by the common ancestor of gibbon ape leukemia virus [GALV] and KoRV) at each of the polymorphic sites detected in museum koala samples was determined by comparing the KoRV sequences to those of the three strains of GALV. Only at positions 243, 560, 966, and 1365 (fig. 1) did the character state across GALV strains differ from that of modern KoRV, which likely retained the common ancestral character state at the other 16 polymorphic sites. Of the 20 polymorphisms, ten represented nonsynonymous mutations. The GA-Branch method was used to estimate that the d*N*/d*S* ratio was 0.15 for KoRV ENV, whereas the free-ratio model estimated the ratio as 0.05. In contrast, the estimated d*N*/d*S* across GALV strains was 0.44 and 0.53, respectively, using the same methods. The low d*N*/d*S* estimate for KoRV suggested the action of purifying selection.

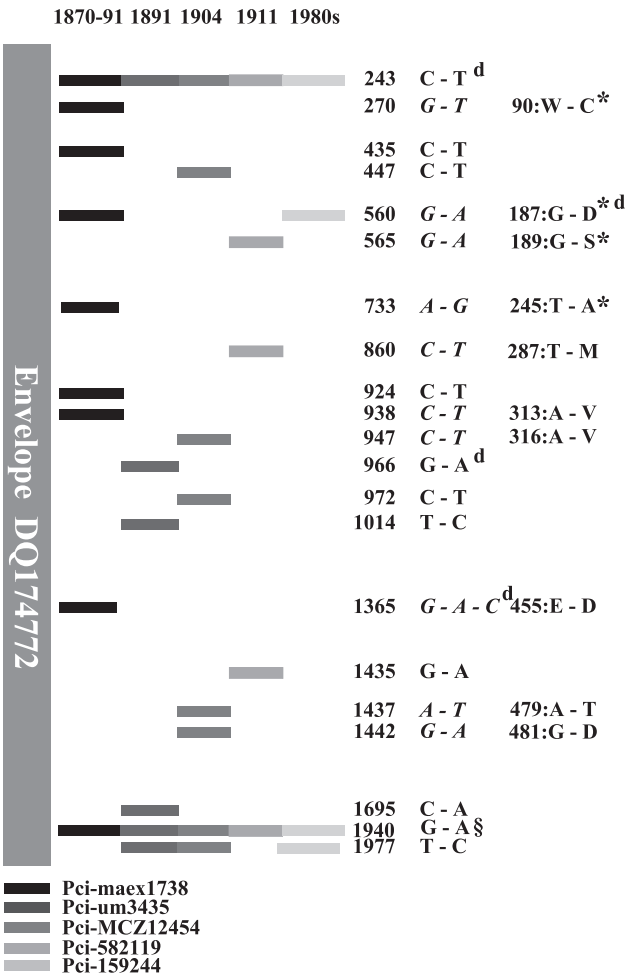


Fig. 1. Variation in the *env* region of KoRV in five koala museum specimens collected across a century. Nucleotide positions are listed for polymorphisms (generally two character states) detected in each KoRV. Both character states listed were present at each polymorphism listed for each individual. The modern sequence (GenBank DQ174772) matched KoRV sequences from all koala museum specimens, including the first of the listed character states at each of the polymorphic sites shown. The only exception was a fixed nucleotide difference at position 1940, indicated by the symbol "§" in which the modern reference sequence had guanine but all museum sequences had adenine. Nucleotide positions are numbered from the start codon of *env*. A "d" denotes polymorphisms in which the reference sequence did not match the character state present in GALV. Nonsynonmous nucleotide variation is italicized, with codon number and single letter amino acid codes next to the corresponding nucleotide polymorphism, with the amino acid matching the KoRV reference listed first. An asterisk (*) indicates amino acid changes that could be structurally modeled.

The exogenous GALV is the closest known relative of KoRV (Hanger et al. 2000). We modeled the three-dimensional structure of KoRV, GALV, and a related porcine ERV (PERV) (Hanger et al. 2000) using as a template the murine leukemia virus (Fass et al. 1997). KoRV, GALV, and PERV ENV proteins were predicted to have a similar folding pattern and overall conserved structural conformations (fig. 2 and supplementary fig. S2, Supplementary Material online). However, the putative receptor-binding site displayed major differences

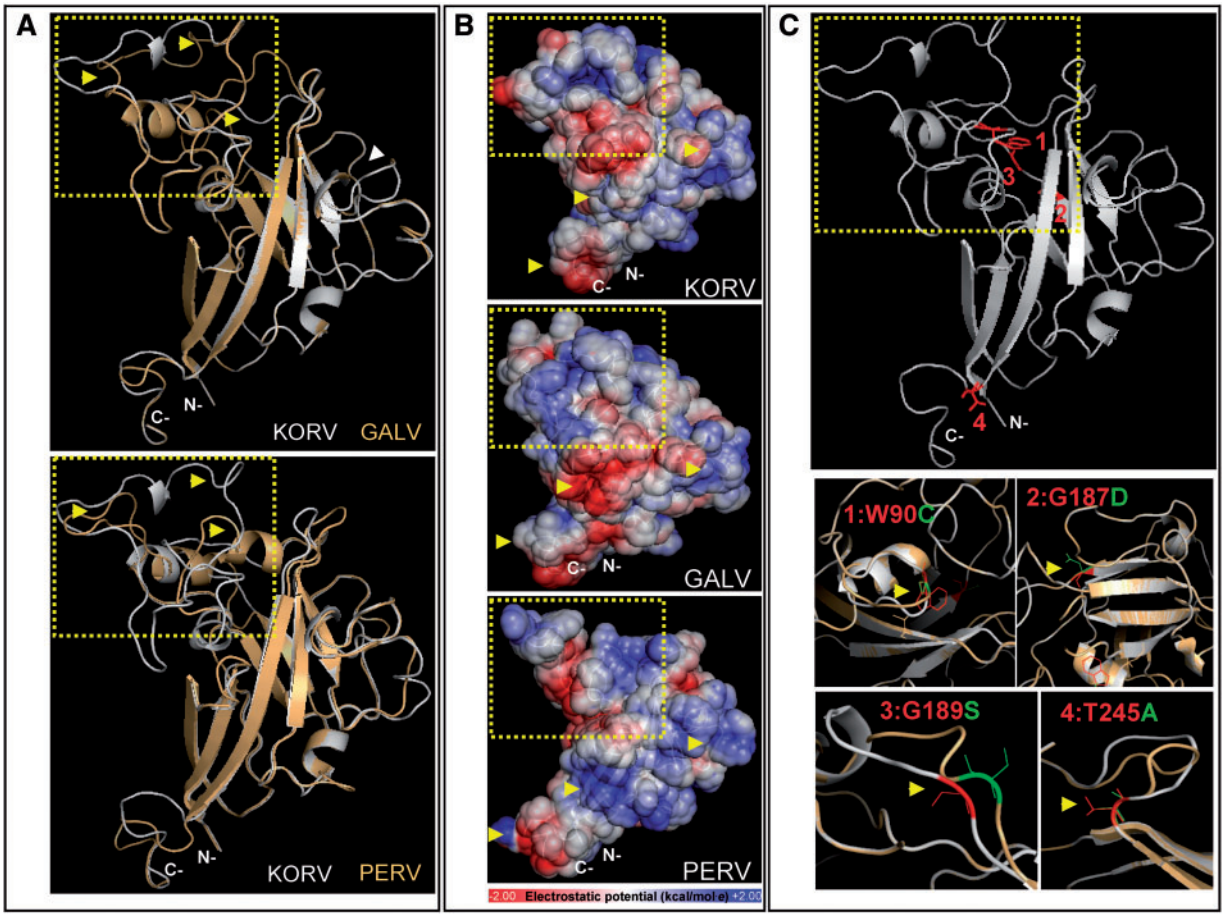


Fig. 2. Museum KoRV ENV protein structural variation. (*A*) Structural superimpositions between KoRV and GALV (top panel), or KoRV and PERV (lower panel), show in cartoon representations the overall similarity of the structures, with major differences located at the receptor-binding sites (yellow boxes). These differences (arrows) include the presence or absence of α helices, different angles between the loops, and the presence of additional loops. (*B*) Distribution of the electrostatic potential on solvent-accessible surfaces of KoRV (top panel), GALV (middle panel), and PERV (lower panel) structures, showing that the major differences in protein charge are located at the receptor-binding sites, indicated by yellow boxes. Additional differences are indicated by arrows. Positively charged groups are blue, negative charges are red, and white gray areas are neutral. (*C*) The structural model of KoRV ENV proteins showing the positions of four nonsynonymous mutations (upper panel; mutations are also listed in fig. 1). Lower panels show structural alignments at sites with amino acid variants (arrows), focusing on mutations and resultant structural changes. The G189S mutation was predicted to cause a major structural alteration that changed the angle and conformation of a conserved loop of unknown function. W90C and T245A both exchanged a surface amino acid with a partially buried one. Both G187D and G189S were predicted to alter the electrostatic profile of the protein surface (not shown).

in both the folding pattern and the number and position of specific secondary elements (fig. 2 and supplementary figs. S2 and S3, Supplementary Material online). This suggested that, across KoRV, GALV, and PERV, positive selection may have diversified the receptor-binding domains, while purifying selection maintained overall folding patterns.

Electrostatic properties greatly differed in parts of the KoRV, GALV, and PERV ENV (Weiner et al. 1982; Matthew 1985) (fig. 2*B* and supplementary fig. S3, Supplementary Material online). Although the physiological importance of these changes would need to be experimentally characterized, these predictions could have functional consequences should interactions between these viruses and their cellular receptors be affected by electrostatic forces, as has been suggested for other viruses (Tse et al. 2011; Xie et al. 2011). The small region of the ENV protein sequence that could be reliably modeled included only four of the amino acid variations detected across KoRV (fig. 2*C*), each of which would produce changes in the topology and surface exposure of amino acids (fig. 2*C* and supplementary fig. S4, Supplementary Material online).

Recombinant GALV strains in which either of two GALV viral motifs is replaced by the homologous KoRV motif have shown reduced infectivity in vitro (Eiden et al. 2007). This has led to the suggestion that endogenization of KoRV into the koala germ line may have been enabled by changes in the motifs known to affect viral infectivity (Eiden et al. 2007). Yet protein modeling of a motif known to affect viral infectivity between GALV (CETTG motif) and KoRV (CETAG) found that the motif in both viruses had almost identical folding patterns, predicted to be largely buried rather than exposed to the surface (not shown). If these predictions prove to be accurate, the CETTG/CETAG motif might not function by binding to the receptor but might instead affect viral infectivity (Eiden et al. 2007) by another mechanism such

as conformational changes. In the five koala museum specimens examined at high coverage, the CETAG sequence of modern KoRV was found in all museum koala samples, with no polymorphisms detected.

In another motif, the *gag-pol* region L domain, GALV has the motif PRPPIY, whereas the homologous region in modern KoRV contains SRLPIY. Hybrid vectors based on GALV but containing the KoRV SRLPIY motif exhibit a several-fold reduction in viral titer in vitro (Eiden et al. 2007). This domain was sequenced in koala samples Pci-um3435, Pci-maex1738, and Pci-159224, spanning more than a century. All three sequences were identical to that of modern KoRV at this motif, with no polymorphisms detected. Thus, both of the functionally important amino acid motifs that have been identified as rendering KoRV less infectious than GALV were found to be present in KoRV since at least the late 19th century.

In summary, the dearth of substantial changes in KoRV *env* or functional motifs for more than a century and the widespread distribution of KoRV in the late 1800s were surprising results. However, these findings would be consistent with a historical account that an epidemic with symptoms that may have been similar to those caused by *Chlamydia* killed large numbers of koalas during 1887–1889 (Lee and Martin 1988). Our results suggest that KoRV has subjected koala populations to greatly increased susceptibility to disease for a period spanning more than a century.

Materials and Methods

Details are included as supplementary methods, Supplementary Material online. Briefly, DNA extractions of 29 museum specimens of 28 individual koalas (two samples were available for one of the koalas) (table 1) were performed in dedicated ancient DNA facilities in Berlin and Copenhagen, following all standards for working with ancient DNA (Cooper and Poinar 2000). PCR primer sequences and combinations are listed in supplementary table S1, Supplementary Material online. PCR strategy, protocols, and high-throughput sequencing using GS FLX (Roche Life Sciences) are detailed in supplementary figure S1, Supplementary Material online. A conservative approach was used to distinguish true variants from sequencing errors or damage (supplementary methods, Supplementary Material online, and supplementary fig. S1, Supplementary Material online). Novel sequences were deposited in GenBank (accession numbers JQ244835–JQ244839). The d*N*/d*S* ratio was estimated using GA-Branch in HyPhy (Pond and Frost 2005) and the free-ratio model in codeml in PAML (Yang 2007) (see supplementary methods, Supplementary Material online). The structural characteristics of proteins were predicted using SWISS-MODEL (Arnold et al. 2006), DaliLite (Holm and Park 2000) and PBEQ-Solver (Im et al. 2008). Further details are in supplementary methods, Supplementary Material online.

Supplementary Material

Supplementary methods, references, figs. S1–S4, and table S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors are grateful to the institutions and individuals listed in table 1 for museum specimens of koalas. They also thank G. Gordon, K. Aplin, and S. Jackson for helpful discussions; M. Farag for technical help; M. Malasky, R. Hanson, S. O'Brien, M. Bush, J. Graves, W. Sherwin, D. Wildt; and the Columbus, San Diego, and San Francisco zoos for modern koala samples. This work was supported by The Australian Research Council, Lundbeck Foundation "Pathogen Palaeogenomes," and the Danish National Research Foundation "GeoGenetics" grants to M.A.A., S.H., M.R., S.F., E.W., and M.T.P.G.; by start-up funds and a state-mini grant from California State University, Fullerton (CSUF) and a Faculty-Student research grant from CSUF to N.N. and R.M.; by an Equity Scholarship from CSUF to R.M.; and by the Smithsonian Institution to K.H. The project described was supported by grant number R01GM092706 from the National Institute of General Medical Sciences (NIGMS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIGMS or the National Institutes of Health.

References

Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22:195–201.

Cooper A, Poinar HN. 2000. Ancient DNA: do it right or not at all. *Science* 289:1139.

Eiden MV, Oliveira NM, Satija H, Kouwenhoven IA. 2007. Changes in viral protein function that accompany retroviral endogenization. *Proc Natl Acad Sci U S A.* 104:17506–17511.

Ellis WAH, Bercovitch FB. 2011. Body size and sexual selection in the koala. *Behav Ecol Sociobiol.* 65:1229–1235.

Fass D, Davey RA, Hamson CA, Kim PS, Cunningham JM, Berger JM. 1997. Structure of a murine leukemia virus receptor-binding glycoprotein at 2.0 angstrom resolution. *Science* 277:1662–1666.

Fiebig U, Hartmann MG, Bannert N, Kurth R, Denner J. 2006. Transspecies transmission of the endogenous koala retrovirus. *J Virol.* 80:5651–5654.

Hanger JJ, Bromham LD, McKee JJ, O'Brien TM, Robinson WF. 2000. The nucleotide sequence of koala (*Phascolarctos cinereus*) retrovirus: a novel type C endogenous virus related to gibbon ape leukemia virus. *J Virol.* 74:4264–4272.

Holm L, Park J. 2000. DaliLite workbench for protein structure comparison. *Bioinformatics* 16:566–567.

Houlden BA, Costello BH, Sharkey D, Fowler EV, Melzer A, Ellis W, Carrick F, Baverstock PR, Elphinstone MS. 1999. Phylogeographic differentiation in the mitochondrial control region in the koala, *Phascolarctos cinereus* (Goldfuss 1817). *Mol Ecol.* 8:999–1011.

Im W, Jo S, Vargyas M, Vasko-Szedlar J, Roux B. 2008. PBEQ—solver for online visualization of electrostatic potential of biomolecules. *Nucl Acids Res.* 36:W270–W275.

Krause J, Dear PH, Pollack JL, Slatkin M, Spriggs H, Barnes I, Lister AM, Ebersberger I, Paabo S, Hofreiter M. 2006. Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* 439:724–727.

Lee AK, Martin R. 1988. The koala: a natural history. Kensington (Australia): New South Wales University Press.

Matthew JB. 1985. Electrostatic effects in proteins. *Annu Rev Biophys Biophys Chem.* 14:387–417.

Mitchell PJ. 1991. The home ranges and social activity of koalas—a quantitative analysis. In: Lee AK, Handasyde KA, Sanson GD, editors. Biology of the koala. Sydney (Australia): Surrey Beatty. p. 171–187.

Pond SL, Frost SD. 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol Biol Evol.* 22:1157.

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

Tarlinton R, Meers J, Young P. 2008. Biology and evolution of the endogenous koala retrovirus. *Cell Mol Life Sci.* 65:3413–3421.

Tarlinton RE. 2006. Characterisation of the epidemiology and molecular biology of koala retrovirus. Brisbane (Australia): University of Queensland.

Tarlinton RE, Meers J, Young PR. 2006. Retroviral invasion of the koala genome. *Nature* 442:79–81.

Tse H, Kao RY, Wu WL, Lim WW, Chen H, Yeung MY, Woo PC, Sze KH, Yuen KY. 2011. Structural basis and sequence co-evolution analysis of the hemagglutinin protein of pandemic influenza A/H1N1 (2009) virus. *Exp Biol Med.* 236:915–925.

Weiner PK, Langridge R, Blaney JM, Schaefer R, Kollman PA. 1982. Electrostatic potential molecular surfaces. *Proc Natl Acad Sci U S A.* 79:3754–3758.

Weiss RA. 2006. The discovery of endogenous retroviruses. *Retrovirology* 3:67.

Xie Q, Lerch TF, Meyer NL, Chapman MS. 2011. Structure–function analysis of receptor binding in adeno-associated virus serotype 6 (AAV-6). *Virology* 420:10–19.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.