



# Phylogenetic Diversity of Koala Retrovirus within a Wild Koala Population

K. J. Chappell,<sup>a</sup> J. C. Brealey,<sup>a</sup> A. A. Amarilla,<sup>a,b</sup> D. Watterson,<sup>a</sup> L. Hulse,<sup>c</sup> C. Palmieri,<sup>d</sup> S. D. Johnston,<sup>c</sup> E. C. Holmes,<sup>e</sup> J. Meers,<sup>d</sup> P. R. Young<sup>a</sup>

School of Chemistry and Molecular Biosciences, University of Queensland, St. Lucia, Queensland, Australia<sup>a</sup>; School of Pharmaceutical Sciences of Ribeirao Preto, University of Sao Paulo, Ribeirao Preto, Sao Paulo, Brazil<sup>b</sup>; School of Agriculture and Food Sciences, University of Queensland, Gatton, Queensland, Australia<sup>c</sup>; School of Veterinary Science, University of Queensland, Gatton, Queensland, Australia<sup>d</sup>; Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, New South Wales, Australia<sup>e</sup>

**ABSTRACT** Koala populations are in serious decline across many areas of mainland Australia, with infectious disease a contributing factor. Koala retrovirus (KoRV) is a gammaretrovirus present in most wild koala populations and captive colonies. Five subtypes of KoRV (A to E) have been identified based on amino acid sequence divergence in a hypervariable region of the receptor binding domain of the envelope protein. However, analysis of viral genetic diversity has been conducted primarily on KoRV in captive koalas housed in zoos in Japan, the United States, and Germany. Wild koalas within Australia have not been comparably assessed. Here we report a detailed analysis of KoRV genetic diversity in samples collected from 18 wild koalas from southeast Queensland. By employing deep sequencing we identified 108 novel KoRV envelope sequences and determined their phylogenetic diversity. Genetic diversity in KoRV was abundant and fell into three major groups; two comprised the previously identified subtypes A and B, while the third contained the remaining hypervariable region subtypes (C, D, and E) as well as four hypervariable region subtypes that we newly define here (F, G, H, and I). In addition to the ubiquitous presence of KoRV-A, which may represent an exclusively endogenous variant, subtypes B, D, and F were found to be at high prevalence, while subtypes G, H, and I were present in a smaller number of animals.

**IMPORTANCE** Koala retrovirus (KoRV) is thought to be a significant contributor to koala disease and population decline across mainland Australia. This study is the first to determine KoRV subtype prevalence among a wild koala population, and it significantly expands the total number of KoRV sequences available, providing a more precise picture of genetic diversity. This understanding of KoRV subtype prevalence and genetic diversity will be important for conservation efforts attempting to limit the spread of KoRV. Furthermore, KoRV is one of the only retroviruses shown to exist in both endogenous (transmitted vertically to offspring in the germ line DNA) and exogenous (horizontally transmitted between infected individuals) forms, a division of fundamental evolutionary importance.

**KEYWORDS** koala, koala retrovirus, envelope protein, endogenous, exogenous, evolution, phylogeny, genetic recombination

Over millions of years, retroviruses have infected all species of vertebrates that have been analyzed. A record of these ancient infections remains because occasionally retroviruses infect germ line cells and become incorporated into the host genome, a process known as endogenization (1). While genomic analysis has shown that retroviruses have frequently appeared during evolutionary history (approximately 8% of the human genome is comprised of endogenous retroviral sequences [2]), little is known

Received 12 September 2016 Accepted 14 November 2016

Accepted manuscript posted online 23 November 2016

**Citation** Chappell KJ, Brealey JC, Amarilla AA, Watterson D, Hulse L, Palmieri C, Johnston SD, Holmes EC, Meers J, Young PR. 2017. Phylogenetic diversity of koala retrovirus within a wild koala population. *J Virol* 91:e01820-16. <https://doi.org/10.1128/JVI.01820-16>.

**Editor** Frank Kirchhoff, Ulm University Medical Center

**Copyright** © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to K. J. Chappell, [k.chappell@uq.edu.au](mailto:k.chappell@uq.edu.au), or P. R. Young, [p.young@uq.edu.au](mailto:p.young@uq.edu.au).

about the processes by which retroviral sequences infiltrate and become permanent residents of the host genome. This is because almost all known endogenization events typically occurred millions of years ago; that of koala retrovirus (KoRV) is a recently identified exception (3).

KoRV is a gammaretrovirus related to gibbon ape leukemia virus (GALV) and woolly monkey virus (WMV). It is present in most wild koala populations and captive colonies and is thought to have been introduced relatively recently by interspecies transmission, although the exact source remains unknown (4, 5). While an accurate time scale has not been determined, evidence suggests that endogenization occurred between 120 and 50,000 years ago (5, 6). KoRV is one of the only retroviruses shown to exist in both endogenous (transmitted vertically to offspring in the germ line DNA) and exogenous (horizontally transmitted between infected individuals) forms (7). Other examples of retroviruses known to be both exogenous and endogenous are, by comparison, much more ancient; Jaagsiekte sheep retrovirus (JSRV) first integrated into the sheep genome approximately 5 to 7 million years ago (8), mouse mammary tumor virus (MMTV) 10 million years ago (9), feline leukemia virus (FeLV) 1 to 10 million years ago (10), and avian leukemia virus (ALV) up to 50 million years ago (11).

KoRV is ubiquitous in all koala populations sampled in the northern Australian state of Queensland (100% of koalas were infected), and although it is currently less prevalent in southern populations, its ongoing spread may eventually result in the infection of all koalas (12). KoRV has previously been demonstrated to be associated with high rates of neoplasia (leukemia and lymphoma) in both captive and wild koala populations (13, 14). Immunomodulation induced by KoRV has also been suggested to be a contributing factor in the high rates of chlamydiosis, a significant cause of koala mortality, morbidity, and infertility (4, 13). Koala populations are currently in serious decline across many areas of mainland Australia, and along with habitat fragmentation and destruction, domestic dog attacks, and road accidents, infectious disease is likely an important additional factor (15).

Five subtypes of KoRV, denoted KoRV-A, -B/J, -C, -D, and -E, have been reported to date; they differ primarily in sequences encoding the envelope protein, particularly within a hypervariable region of the receptor binding domain (RBD) (amino acids [aa] 91 to 135 [KoRV-A numbering]) (14, 16–21). KoRV-A was first identified in wild and captive Australian koala populations (16) and is the most commonly identified subtype (6, 14, 18, 20, 22–24). KoRV-B, -C, and -D (originally referred to as clones 11-4 [or KoRV-J], 11-2, and 11-1, respectively) were all initially identified in captive koalas housed in zoos in Japan (17, 20), while KoRV-E was more recently identified in koalas housed in a zoo in the United States (21). KoRV-B was subsequently identified in captive koalas housed in zoos in the United States and Germany (14, 18), as well as in wild Australian koalas (23). KoRV-C has also been identified in a captive koala housed in a zoo in the United States (19), while KoRV-D has been identified in wild Australian koalas (23). Another potentially new subtype was recently reported and tentatively designated KoRV-F (21), although this sequence does not differ substantially from that of KoRV-D.

Notably, differences in the envelope protein RBD between subtypes A and B are responsible for these subtypes using distinct host cell surface receptors (14, 20, 25). KoRV-A has been shown to use the sodium-dependent phosphate transporter encoded by the Pit-1 (SLC20A1) gene (25), while KoRV-B uses the thiamine transporter encoded by the THTR1 (SLC19A2) gene (14, 20). It has been suggested that KoRV-B is more infectious, transmissible, and associated with a higher risk of neoplastic disease (14). While subtypes C to E have not been investigated to the same extent as subtypes A and B, sequences encoded by the envelope RBD differ substantially, which is compatible with the possibility that they also display altered replication properties stemming from differential receptor usage.

In this study, we sought to determine the extent and pattern of KoRV phylogenetic diversity and subtype prevalence among wild koalas within southeast Queensland, Australia. This analysis expands upon previous analyses on captive koalas housed in zoos outside Australia and reveals the complexity of KoRV evolution in natural popu-

lations. Of note, our study reveals that the majority of koalas are simultaneously infected with multiple subtypes and that transmission is likely ongoing.

## RESULTS

**Genetic variation within the KoRV envelope gene.** We conducted deep sequencing of an approximately 500-nucleotide (nt) region of the KoRV *env* gene, incorporating a previously recognized hypervariable hot spot within the receptor binding domain, from 18 wild koalas. KoRV viral RNA within plasma was analyzed for 10 koalas (1 to 10), and the integrated proviral sequences present in genomic DNA within whole blood was analyzed for eight koalas (11 to 18). After quality control and filtering, an average of ~30,000 reads was produced for each sample, ranging between 20,000 and 60,000 (the number of reads attributed to each sequence type for each animal is presented in Fig. S3, S4, and S5 in the supplemental material). Sequence validation resulted in the exclusion of approximately 3% of reads due to the presence of missense mutations, premature stop codons, or large deletions.

The validated data set included a total of 108 unique sequences and a sequence identical to that of KoRV-A. A phylogenetic tree was estimated using the entire data set, as well as the equivalent region from nine previously published KoRV sequences and GALV and WMV sequences utilized as outgroups (Fig. 1). All of the identified sequences grouped with previously known subtypes A, B, and D, although some sequences related to subtype D exhibited relatively long branches with high bootstrap values (see below) (although the subtype A sequences were not strictly monophyletic in the tree presented in Fig. 1, they are when the outgroup sequences are excluded [not shown]). The previously identified subtypes C and E were not present in any of the samples sequenced in this study.

Our analysis revealed significant genetic diversity in the *env* gene, including numerous point mutations, deletions, insertions, and potential recombination events, which resulted in a complex pattern of phylogenetic diversity. In particular, although subtypes C and D have been described previously based upon differences in the hypervariable region (17, 19, 23), our more detailed analysis revealed that these subtypes in reality belong to a highly diverse paraphyletic group (i.e., subtype D) that contains a number of other subtypes (E, F, G, H, and I) within its component genetic diversity (Fig. 1). In contrast, subtype A harbors markedly lower levels of genetic diversity and was less distant from the GALV and WMV outgroups. This may indicate that subtype A represents a collection of endogenous rather than exogenous viruses, as the former are expected to exhibit lower rates of evolutionary change.

As noted above, a number of *env* sequences exhibited long branches within the paraphyletic subtype D. More importantly, these divergent sequences exhibited such distinct amino acid signatures in the hypervariable region of the *env* RBD that, following convention, we propose that they constitute new virus subtypes: KoRV-F, -G, -H, and -I (Fig. 2). Interestingly, the *env* hypervariable region of KoRV-F appeared to show homology to regions of KoRV-C (aa 118 to 123, PWPGF<sub>T</sub>), as well as KoRV-D (aa 100 to 117, SxQARPP<sub>LYD</sub>xPxGTPGA).

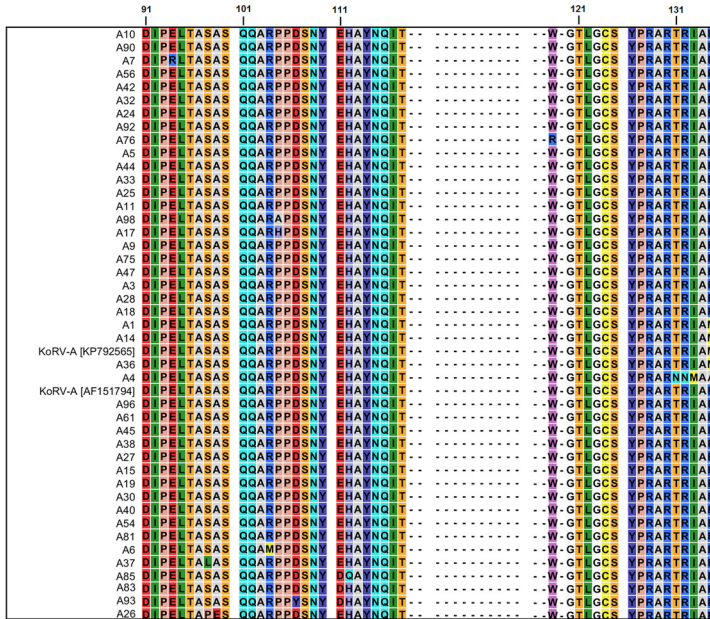
Notably, these currently and newly identified subtypes were present in multiple koalas. Overall, KoRV-A was detected in all 18 animals, while the other subtypes were present in a subset of koalas: KoRV-B/J, 14/18; KoRV-D, 17/18; KoRV-F, 8/18; KoRV-G, 2/18; KoRV-H, 1/18; and KoRV-I, 1/18. Hence, all individual animals tested were positive for at least two distinct hypervariable region subtypes, with the majority positive for three or four, suggesting that multiple subtypes cocirculate in a single population (Table 1; Fig. 3).

Small sequence deletions were frequently detected around the region encoding amino acids 36 to 39. These deletions were predominantly of three amino acids at positions 36 to 38, 37 to 39, or, in one instance, 35 to 37. In one case a deletion of two amino acids at positions 35 and 36 was identified, while in another a deletion of 11 amino acids at positions 35 to 42 was identified.





**KoRV-A**



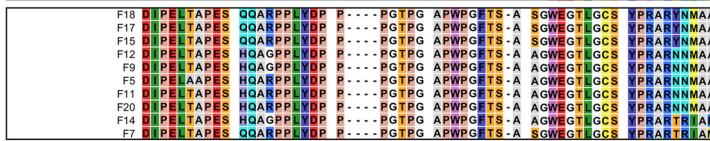
**KoRV-C**



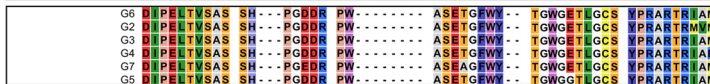
**KoRV-H**



**KoRV-F**



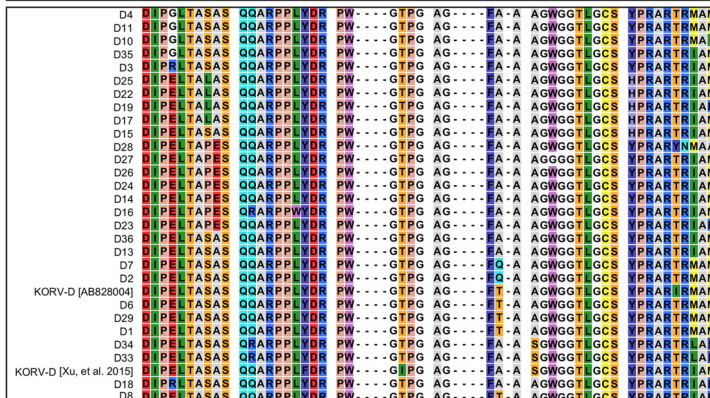
**KoRV-G**



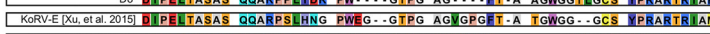
**KoRV-I**



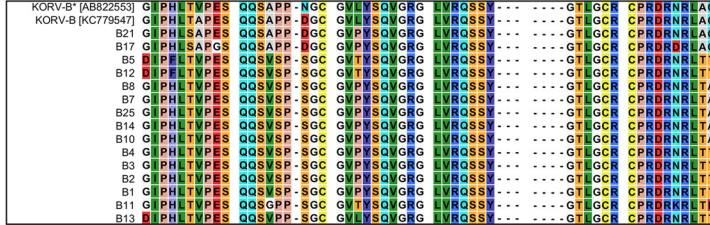
**KoRV-D**



**KoRV-E**



**KoRV-B**



**FIG 2** Envelope hypervariable region diversity. An alignment of amino acid sequences comprising the hypervariable region within the receptor binding domain (RBD) of the envelope protein (amino acids 91 to 135 [KoRV-A numbering]) is shown.

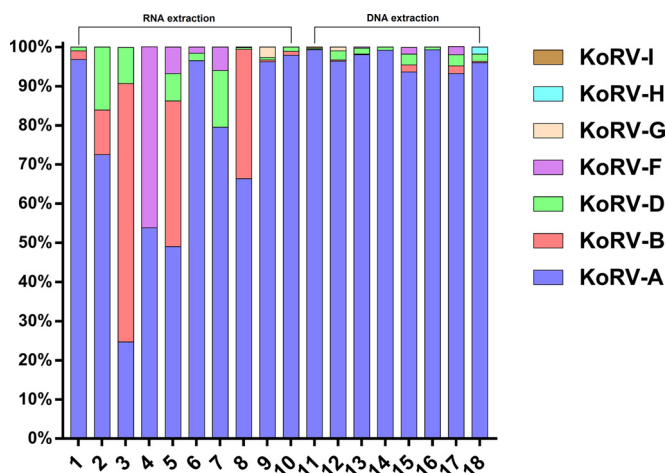
PROF OC

**TABLE 1** Percentages of reads grouping with groups A, B, and D and with hypervariable region subtypes

Koala	% of reads grouping with group:									
	A	B	D	D hypervariable region subtype:						
				C	D	E	F	G	H	I
1	96.8	2.2	1		1					
2	72.5	11.4	16.1		16.1					
3	24.7	66	9.2		9.2					
4	53.8	0.04	46.2				46.2			
5	49	37.2	13.8		7		6.8			
6	96.5		3.5		1.9		1.6			
7	79.5		20.5		14.5		6			
8	66.4	33	0.6		0.5		0.1			
9	96.2	0.5	3.3		0.6			2.7		
10	97.9	1	1.1		1.1					
11	99.3	0.1	0.6		0.3					0.3
12	96.4	0.3	3.3		2.3			1		
13	98	0.2	1.8		1.5		0.3			
14	99.2		0.8		0.8					
15	93.6	1.8	3.6		2.8		1.7			
16	99.3		0.7		0.7					
17	93.2	2	4.8		2.8		2.1			
18	96	0.3	3.7		1.9				1.8	

99%). The difference in the proportion of KoRV-A sequences within samples 1 to 10 compared to samples 11 to 18 was found to be statistically significant ( $P = 0.018$ ). For subtypes B, D, and F, no particular sequence was dominant, with different sequences producing a high number of reads in different samples. For both RNA- and DNA-derived samples, the major KoRV-A sequence was identical to that which has been previously identified (16), and this sequence constituted ~96% of all KoRV-A reads.

**Genetic variation between animals and geographical distribution.** Roughly half of the 108 unique sequences were identified in multiple animals, potentially indicating either relatively recent transmission or endogenization within an ancestor, while half were identified in only one individual, consistent with ongoing viral evolution. A breakdown of the sequences identified in each sample and the number of reads is presented in Fig. S3, S4, and S5. Koalas 1, 3, 8, and 16 shared a number of KoRV-D sequences, as did koalas 5, 8, 9, 13, and 14, while koalas 4, 5, 6, 13, 15, and 17 shared a number of KoRV-F sequences (Fig. S5). These animals tended to be geographically colocated (see Fig. S2 in the supplemental material), consistent with either horizontal transmission or potentially unique reintegrations into the germ line DNA within a



**FIG 3** Percentage of reads grouping with hypervariable region subtypes.

shared common ancestor. Koalas 1, 3, and 16 were all found within a 5-km radius southeast of Brisbane, while koalas 4, 5, 6, 8, 9, 14, 15, and 17 were all found within a 10-km radius to the north of Brisbane.

Koala 7 possessed a large number of unique sequences and was distant from the other koalas in this study. Similarly, koala 11 was the only animal in which KoRV-I sequences were identified, while koala 12 was one of two animals in which KoRV-G sequences were identified (the other being koala 9). Despite being located distant from the other koalas included in the study, koalas 11 and 12 shared some KoRV-D sequences that were identical or similar to those found in other koalas. Conversely, koala 2 was found in the same area as koalas 1, 3, and 16 but did not share similar sequences with those animals, instead possessing a number of unique KoRV-D sequences (Fig. S5).

**Presence of the CETTG motif.** Another feature of interest in the RBD is a CETTG motif located at amino acids 167 to 171 (KoRV-A numbering), which has been previously identified in GALV and KoRV-B (14) and has been reported to be involved in replicative potential and/or pathogenicity (14, 26). This motif is conserved among all exogenous murine gammaretroviruses but is mutated in endogenous elements. In addition to KoRV-B, we identified sequences containing an intact CETTG motif within subtypes A, D, and F in samples from two of 18 koalas (koalas 2 and 7). Within these samples, sequences including an intact CETTG motif were a minority comprising fewer than 1% of the total sequences for each subtype.

**Phylogenetic evidence for recombination.** The short amplicon length together with the presence of the hypervariable region made a full resolution of the history of recombination within KoRV impossible. We therefore estimated separate phylogenetic trees based on alignment regions 1 to 212 and 212 to 536 (equivalent to nt 23 to 232 and nt 233 to 491 [KoRV-A numbering]) to identify cases of strong phylogenetic incongruence compatible with the occurrence of recombination. Eight cases were identified in which sequences displayed widespread phylogenetic movement between the two trees (clearly visible as the mix of colors in Fig. S6 in the supplemental material). Each case showed strong bootstrap support (>70%) in the tree based on the region 1 to 212 (Fig. S6a) but separated into different and well-supported hypervariable region subtypes in the tree based on alignment 213 to 536 (Fig. S6b). Notably, in six of the eight cases, these putative recombinant sequences were confined to individual animals, while in the other two cases, the sequences were present in two and three animals, respectively (see Fig. S3, S4, S5, and S7 in the supplemental material). Confirmation of recombination in these viruses will require dedicated analyses with longer sequences.

## DISCUSSION

Although KoRV is highly prevalent in wild koala populations, with infection rates as high as 100% in many regions (12), little is known about the genetic diversity of this virus, including the presence and prevalence of divergent subtypes. This in part reflects the fact that our current understanding of KoRV diversity comes almost exclusively from captive colonies located outside Australia. To expand upon these limited studies, we analyzed viruses sampled from the wild koala population of southeast Queensland. Specifically, we collected samples from 18 diseased individuals and analyzed KoRV genetic diversity using deep sequencing of an approximately 500-nucleotide region of the *env* gene that encodes a previously identified hypervariable region within the receptor binding domain. This region is of particular significance as it has been implicated in alternate receptor usage and has been used to define unique KoRV subtypes (14, 16–21). Our study represents the first analysis of KoRV genetic diversity in a wild koala population and the first application of deep sequencing to profile the viral population within infected animals.

The detection of 108 unique sequences significantly extends the total number of KoRV sequences from the nine previously reported, thereby providing a more precise picture of genetic diversity. Furthermore, the application of deep sequencing to whole-cell DNA as well as plasma-derived RNA, presumably packaged viral transcripts and potentially infectious, has provided insight into the relative levels of these subtypes

that are either embedded within the host genome or actively transcribed. The data revealed extensive genetic diversity that is seemingly a product of point mutations, deletions, and insertions. The presence of significant phylogenetic incongruence in the data is also indicative of recombination, as observed in other retroviruses (27), although we were unable to unambiguously resolve recombinants, their parents, and the break-points due to the short amplicon length.

Notably, the phylogenetic diversity of KoRV fell into three major groups, two comprising previously identified subtypes A and B and a third containing hypervariable region subtypes C to I. The identification of four previously unreported hypervariable region subtypes, along with the high level of sequence diversity from the analysis of only 18 individual animals from a relatively small geographic region, suggests that more extensive genetic diversity is likely to be identified for KoRV across its extensive geographic range.

While some of the subtypes previously identified outside Australia (C and E) were not detected in any individuals in our study, others (A, B, and D) were frequently detected. In addition to previously identified subtypes, four potentially new subtypes that depict the high levels of genetic diversity in the hypervariable region (F, G, H, and I) were identified, although it is noteworthy that all these fell within the diverse (and paraphyletic) group D. Apart from KoRV-A, which was ubiquitous and perhaps endogenous (see below), subtypes B, D, and F were highly prevalent, with at least one of these three subtypes found in all animals analyzed.

Subtypes B, D, and F exhibited rich intrasubtype genetic diversity, with mean genetic distances of 4 to 5% in *env*. By comparison, ~96% of sequence reads within KoRV-A were identical to the previously identified prototype sequence. Based on prevalence and the relative lack of sequence diversity, KoRV-A has previously been proposed as the endogenous subtype that is vertically transmitted to offspring (14). Given that these will acquire mutations at the same (low) rate as host genomic DNA, endogenous virus copies are expected to evolve markedly more slowly than their exogenous counterparts, in which genetic diversity is continually produced through replication with a highly error-prone reverse transcriptase. In addition, KoRV-A reads were detected at a significantly higher rate when DNA was used as a starting template (koalas 11 to 18) than when RNA was used (koalas 1 to 10). Conversely, KoRV-B has been proposed as an exogenous subtype that is horizontally transmitted between individuals but not vertically transmitted, and it exhibits higher levels of genetic diversity. This hypothesis is based on a limited pedigree analysis in which a single KoRV-B-positive sire that did not transmit KoRV-B to two progeny was identified (14). The paraphyletic group containing KoRV-C, -D, -E, -F, -G, -H, and -I identified in the current study similarly exhibits a high level of genetic diversity consistent with ongoing exogenous viral evolution.

Interestingly, identical or highly similar sequences for subtypes B, D, and F were present in multiple animals, particularly in those animals from a single area, indicating either recent horizontal transmission or further unique endogenization events in common ancestors. In contrast, animals located at greater distance from the others within the study were more likely to have unique sequences. However, this trend was not absolute, with one animal (koala 2) identified in a region similar to that for three other animals that did not possess any of the *env* gene sequences present in those animals. This geographical distribution is consistent with either continued horizontal transmission or unique endogenization events within ancestors leading to vertical transmission.

It also needs to be noted that as we did not analyze isolated infectious virus, we cannot exclude the possibility that sequencing reads could originate from transcribed RNA and not from replication-competent virus. Previous studies have derived infectious viral isolates from subtypes A, B, C, and D (14, 17, 25). In our study, the high percentage of reads attributable to KoRV-A, -B, -D, and -F within plasma from individuals suggests that these viruses are likely to be actively replicating. However, sequences defined as potential subtypes G, H, and I constituted fewer than 3% of reads and were identified



in only one or two individual animals. While these sequences appear to encode intact functional envelope proteins, the possibility that they are inactive cannot be excluded. Future studies aimed at deriving viral isolates for these novel subtypes will need to be performed to clarify the nature of these variations with respect to KoRV infectivity.

One key remaining question is whether any of the individual subtypes are more likely to be associated with disease/pathology. While KoRV infection has been shown to be associated with high rates of neoplasia and has been implicated in immunomodulation that may predispose animals to chlamydiosis (4, 13, 14), little is known about the contribution of genetically diverse subtypes to koala morbidity. The KoRV-B subtype has been suggested to be associated with a higher risk of neoplasia and to be more infectious and transmissible. However, these conclusions were based on only a small number of captive koalas (14). While our study was not designed to evaluate the contributions of respective subtypes to disease severity, it is clear from the veterinary and histological assessment that the animals tested were diseased regardless of the presence of KoRV-B. These observations suggest that association of KoRV-B with disease may not be as definitive as previously suggested (14).

We also examined the CETTG motif at amino acid positions 167 to 171 (KoRV-A numbering), which has also been suggested to contribute to increased viral infectivity and virulence (14, 26). We identified an intact CETTG motif in a minority of sequences belonging to subtypes A, B, D, and F in samples from two of 18 animals (koalas 2 and 7). These animals did not appear to be experiencing any increased morbidity compared to animals in which the motif was disrupted in all sequences present (see Fig. S1 in the supplemental material). Additionally, sequences containing the intact CETTG constituted fewer than 1% of the valid reads in these animals. Our preliminary findings, therefore, indicate that the presence of this motif may contribute little to viral transcription and packaging, as suggested previously (14).

In summary, by employing deep sequencing, we have been able to determine the diversity of KoRV sequences both within individual animals and within a subset of the wild koala populations around southeast Queensland. This study significantly expands the total number of KoRV sequences available across the hypervariable region of the *env* RBD and provides a more precise picture of KoRV evolution and genetic diversity. All individual animals tested were positive for two or more divergent KoRV sequences, with numerous point mutations, deletions, insertions, and putative recombination events found to contribute to sequence diversity. Although our study was relatively limited in its scope, we identified four potentially new envelope hypervariable region subtypes, suggesting that we have only just begun to scratch the surface of KoRV genetic diversity. The greater understanding of KoRV subtype prevalence and genetic diversity provided by this study, and future endeavors, will be important both for conservation efforts and for the investigation of the relationship between exogenous and endogenous retroviruses.

## MATERIALS AND METHODS

**Sample collection and processing.** Eighteen wild adult koalas were included in this study, all reported by the public as being sick and admitted to koala hospitals around southeast Queensland. The date and location where each of these koalas was found as well as details from the veterinary and histology reports are included in Fig. S1 and S2 in the supplemental material. Eleven adult males were admitted between January 2014 and January 2015 (koalas 1 to 10 and 12). A further six koalas (3 females and 3 males) were admitted during 2010 (koalas 13 to 18), and an adult female was admitted during 2013 (koala 11). For koalas 1 to 10, viral RNA was isolated from blood plasma using the High Pure viral nucleic acid kit (Roche, USA). For koalas 11 to 18, genomic DNA was isolated from whole blood using the PureLink genomic DNA minikit (Invitrogen, USA).

**Illumina sequencing.** Preparation of the amplicon library was performed as described by Illumina, using the workflow outlined in the manufacturer's guidelines (16S Metagenomic Sequencing Library Preparation guide 15044223-B). Oligonucleotide primers flanking the hypervariable region of the envelope gene (*env*) were used to amplify the target sequence by PCR. Primers included Illumina adaptor sequences (italics) and complementary regions: *env*22.F (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG GCTTCTACTCTCAAACCCGCGCC) and *env*514.R (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGTTG CCAAGTAGCGGTTC). Phusion high-fidelity DNA polymerase (NEB, UK) was used for PCR according to the manufacturer's instructions, with 25 rounds of amplification and an annealing temperature of 55°C. Attachment and barcode sequences were incorporated in an additional 8 rounds of amplification.

Sequencing of the PCR amplicons was performed at the Australian Centre for Ecogenomics, University of Queensland. The PCR products were purified using Agencourt AMPure XP beads (Beckman Coulter, USA). Purified DNA was indexed with unique 8-bp barcodes using the Illumina Nextera XT 384 sample index kit A-D (Illumina FC-131-1002) under standard PCR conditions with Q5 Hot Start high-fidelity 2× master mix (NEB, UK). Indexed amplicons were pooled in equimolar concentrations and sequenced on a MiSeq sequencing system (Illumina, USA) using paired-end sequencing with V3 300-bp chemistry according to the manufacturer's protocol.

Forward and reverse reads were merged based on approximately 100 nucleotides of overlapping sequencing using the PANDAseq Assembler 2.7 (28). The *de novo* operational taxonomic unit (OUT) picking method from QIIME 1.8 (29) was used with UCLUST (30) to cluster reads into OTUs with a similarity of 99% and then select representative sequences. Low-abundance OTUs with a minimum count of <0.01% of the total number of reads were removed from the analysis. Representative sequences were aligned with the *env* nucleotide sequence from KoRV-A (GenBank accession number AF151794) to confirm sequence validity using CLC-biological workbench 6 (CLCbio, Denmark). Sequences lacking homology to the *env* gene of KoRV-A and those containing missense mutations or large deletions were excluded from further analysis. Sequence termini were trimmed to nucleotides 23 to 513 (KoRV-A numbering). The validated data set included 108 unique *env* sequences of KoRV and a sequence identical to that of KoRV-A.

**Phylogenetic analysis.** The 108 *env* sequences obtained here were combined with nine previously reported unique KoRV *env* sequences, seven retrieved from GenBank (accession numbers AF151794, KP792564, KP792565, KC779547, AB822553, AB828004, and AB828005) and two unique sequences supplied by Maribeth Eiden (21). In addition, four sequences were utilized as outgroups to root the KoRV phylogeny: two of GALV (GenBank accession numbers KT24047 and KT24048) and two of WMV (KT724051 and KX059700, with the latter sampled from a rodent, *Melomys burtoni* [31]). These sequences were then aligned using MUSCLE (32) with default parameters ensuring that the nucleotide sequence alignment properly matched triplet codon structure. This resulted in a final data set of 121 *env* sequences, 536 nucleotides in length.

A phylogenetic tree of these data was estimated using the maximum likelihood method available in the PhyML package (33), assuming a GTR+I+Γ model of nucleotide substitution and employing SPR+NNI branch swapping. The value for the invariant-site parameter (I) was 0.264, while that for among-site rate variation (Γ) was 1.026. To determine the robustness of individual nodes on the phylogeny, a bootstrap resampling exercise was undertaken, utilizing 1,000 bootstrap replications under the same substitution model defined above but employing NNI branch swapping. Each representative sequence was assigned to a KoRV subtype (A to G) based on clustering within the phylogenetic tree and analysis of patterns of amino acid sequence variation within the hypervariable region, as in previous studies (14, 19, 20, 23).

A provisional analysis using recombination detection methods within the RDP4 package (34) provided evidence of a complex history of recombination (data not shown; available on request). However, due to the short amplicon length and the presence of a hypervariable region, it was impossible to fully characterize recombination events by unambiguously identifying breakpoints, recombinant sequences, and their parents. We therefore adopted a simpler, although equally robust, approach and looked for significant incongruence among phylogenetic trees inferred for (arbitrarily chosen) regions 1 to 212 and 213 to 536 of the sequence alignment. Accordingly, bootstrap maximum-likelihood phylogenies were estimated for both regions as described above; putative recombinant sequences were then identified as those with incongruent groupings supported by bootstrap values of >70%.

**Statistical analysis.** The proportions of KoRV-A sequences identified within extracted DNA and RNA samples were compared by the unpaired *t* test using GraphPad PRISM-7 software.

**Accession number(s).** All unique sequences have been deposited in GenBank and assigned accession numbers KX587950 to KX588057.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JVI.01820-16>.

**TEXT S1**, PDF file, 0.6 MB.

## ACKNOWLEDGMENTS

Illumina deep sequencing was conducted by the Australian Centre for Ecogenomics at the University of Queensland. We thank Maribeth Eiden for the supply of unpublished sequence data.

This work was supported by the Australian Research Council via Linkage Project LP0989701 (2010 to 2013) (Retroviral Invasion of the Koala Genome: Prevalence, Transmission and Role in Immunosuppressive Disease), which was awarded to J. Meers and P. R. Young, and by Queensland State Government Koala Research Grant KRG005 awarded to S. D. Johnston, C. Palmieri, and others. E. C. Holmes is supported by an NHMRC Australia Fellowship.

## REFERENCES

- Coffin JM. 2004. Evolution of retroviruses: fossils in our DNA. *Proc Am Philos Soc* 148:264–280.
- Lander ES, Linton LM, Birren B, Nussbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921. <https://doi.org/10.1038/35057062>.
- Denner J. 2007. Transspecies transmissions of retroviruses: new cases. *Virology* 369:229–233. <https://doi.org/10.1016/j.virol.2007.07.026>.
- Denner J, Young PR. 2013. Koala retroviruses: characterization and impact on the life of koalas. *Retrovirology* 10:108. <https://doi.org/10.1186/1742-4690-10-108>.
- Ishida Y, Zhao K, Greenwood AD, Roca AL. 2015. Proliferation of endogenous retroviruses in the early stages of a host germ line invasion. *Mol Biol Evol* 32:109–120. <https://doi.org/10.1093/molbev/msu275>.
- Avila-Arcos MC, Ho SY, Ishida Y, Nikolaidis N, Tsangaras K, Honig K, Medina R, Rasmussen M, Fordyce SL, Calvignac-Spencer S, Willerslev E, Gilbert MT, Helgen KM, Roca AL, Greenwood AD. 2013. One hundred twenty years of koala retrovirus evolution determined from museum skins. *Mol Biol Evol* 30:299–304. <https://doi.org/10.1093/molbev/mss223>.
- Tarlington RE, Meers J, Young PR. 2006. Retroviral invasion of the koala genome. *Nature* 442:79–81. <https://doi.org/10.1038/nature04841>.
- Spencer TE, Palmarini M. 2012. Application of next generation sequencing in mammalian embryogenomics: lessons learned from endogenous betaretroviruses of sheep. *Anim Reprod Sci* 134:95–103. <https://doi.org/10.1016/j.anireprosci.2012.08.016>.
- Baillie GJ, van de Lagemaat LN, Baust C, Mager DL. 2004. Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals. *J Virol* 78:5784–5798. <https://doi.org/10.1128/JVI.78.11.5784-5798.2004>.
- Benveniste RE, Sherr CJ, Todaro GJ. 1975. Evolution of type C viral genes: origin of feline leukemia virus. *Science* 190:886–888. <https://doi.org/10.1126/science.52892>.
- Dimcheff DE, Drovetski SV, Krishnan M, Mindell DP. 2000. Cospeciation and horizontal transmission of avian sarcoma and leukosis virus gag genes in galliform birds. *J Virol* 74:3984–3995. <https://doi.org/10.1128/JVI.74.9.3984-3995.2000>.
- Simmons GS, Young PR, Hanger JJ, Jones K, Clarke D, McKee JJ, Meers J. 2012. Prevalence of koala retrovirus in geographically diverse populations in Australia. *Aust Vet J* 90:404–409. <https://doi.org/10.1111/j.1751-0813.2012.00964.x>.
- Tarlington 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. <https://doi.org/10.1099/vir.0.80547-0>.
- Xu W, Stadler CK, Gorman K, Jensen N, Kim D, Zheng H, Tang S, Switzer WM, Pye GW, Eiden MV. 2013. An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo. *Proc Natl Acad Sci U S A* 110:11547–11552. <https://doi.org/10.1073/pnas.1304704110>.
- Department of Environment and Resource Management. 2009. Report on the decline of the Koala Coast koala population: population status in 2008. Department of Environment and Resource Management, Queensland, Australia.
- 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. <https://doi.org/10.1128/JVI.74.9.4264-4272.2000>.
- Miyazawa T, Shojima T, Yoshikawa R, Ohata T. 2011. Isolation of koala retroviruses from koalas in Japan. *J Vet Med Sci* 73:65–70. <https://doi.org/10.1292/jvms.10-0250>.
- Fiebig U, Keller M, Moller A, Timms P, Denner J. 2015. Lack of antiviral antibody response in koalas infected with koala retroviruses (KoRV). *Virus Res* 198:30–34. <https://doi.org/10.1016/j.virusres.2015.01.002>.
- Abts KC, Ivy JA, DeWoody JA. 2015. Immunomics of the koala (*Phascolarctos cinereus*). *Immunogenetics* 67:305–321. <https://doi.org/10.1007/s00251-015-0833-6>.
- 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. *J Virol* 87:9943–9948. <https://doi.org/10.1128/JVI.01385-13>.
- Xu W, Gorman K, Santiago JC, Kluska K, Eiden MV. 2015. Genetic diversity of koala retroviral envelopes. *Viruses* 7:1258–1270. <https://doi.org/10.3390/v7031258>.
- Fiebig U, Hartmann MG, Bannert N, Kurth R, Denner J. 2006. Transspecies transmission of the endogenous koala retrovirus. *J Virol* 80:5651–5654. <https://doi.org/10.1128/JVI.02597-05>.
- Hobbs M, Pavasovic A, King AG, Prentis PJ, Eldridge MD, Chen Z, Colgan DJ, Polkinghorne A, Wilkins MR, Flanagan C, Gillett A, Hanger J, Johnson RN, Timms P. 2014. A transcriptome resource for the koala (*Phascolarctos cinereus*): insights into koala retrovirus transcription and sequence diversity. *BMC Genomics* 15:786. <https://doi.org/10.1186/1471-2164-15-786>.
- Tsangaras K, Siracusa MC, Nikolaidis N, Ishida Y, Cui P, Vielgrader H, Helgen KM, Roca AL, Greenwood AD. 2014. Hybridization capture reveals endogenous and conservation across the entire Koala retrovirus genome. *PLoS One* 9:e95633. <https://doi.org/10.1371/journal.pone.0095633>.
- Oliveira NM, Farrell KB, Eiden MV. 2006. In vitro characterization of a koala retrovirus. *J Virol* 80:3104–3107. <https://doi.org/10.1128/JVI.80.6.3104-3107.2006>.
- Oliveira NM, Satija H, Kouwenhoven IA, Eiden MV. 2007. Changes in viral protein function that accompany retroviral endogenization. *Proc Natl Acad Sci U S A* 104:17506–17511. <https://doi.org/10.1073/pnas.0704313104>.
- Vuilleumier S, Bonhoeffer S. 2015. Contribution of recombination to the evolutionary history of HIV. *Curr Opin HIV AIDS* 10:84–89. <https://doi.org/10.1097/COH.0000000000000137>.
- Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. 2012. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 13:31. <https://doi.org/10.1186/1471-2105-13-31>.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>.
- Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
- Alfano N, Michaux J, Morand S, Aplin K, Tsangaras K, Lober U, Fabre PH, Fitriana Y, Semiadi G, Ishida Y, Helgen KM, Roca AL, Eiden MV, Greenwood AD. 2016. Endogenous gibbon ape leukemia virus identified in a rodent (*Melomys burtoni* subsp.) from Wallacea (Indonesia). *J Virol* 90:8169–8180. <https://doi.org/10.1128/JVI.00723-16>.
- Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. <https://doi.org/10.1186/1471-2105-5-113>.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst biology* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>.
- Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462–2463. <https://doi.org/10.1093/bioinformatics/btq467>.

## AUTHOR QUERIES

### AUTHOR PLEASE ANSWER ALL QUERIES

1

AQau—Please confirm the given-names and surnames are identified properly by the colors.

■ = Given-Name, ■ = Surname

AQaff—Please confirm the following full affiliations or correct here as necessary. This is what will appear in the online HTML version:

<sup>a</sup>School of Chemistry and Molecular Biosciences, University of Queensland, St. Lucia, Queensland, Australia

<sup>b</sup>School of Pharmaceutical Sciences of Ribeirao Preto, University of Sao Paulo, Ribeirao Preto, Sao Paulo, Brazil

<sup>c</sup>School of Agriculture and Food Sciences, University of Queensland, Gatton, Queensland, Australia

<sup>d</sup>School of Veterinary Science, University of Queensland, Gatton, Queensland Australia

<sup>e</sup>Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, New South Wales, Australia

AQaff—This affiliation line will appear in the PDF version of the article and matches that on page 1 of the proof; corrections to this affiliation line may be made here **or** on page 1 of the proof:

School of Chemistry and Molecular Biosciences, University of Queensland, St. Lucia, Queensland, Australia<sup>a</sup>; School of Pharmaceutical Sciences of Ribeirao Preto, University of Sao Paulo, Ribeirao Preto, Sao Paulo, Brazil<sup>b</sup>; School of Agriculture and Food Sciences, University of Queensland, Gatton, Queensland, Australia<sup>c</sup>; School of Veterinary Science, University of Queensland, Gatton, Queensland, Australia<sup>d</sup>; Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, New South Wales, Australia<sup>e</sup>

AQfund—The Funding Information below includes information that you provided on the submission form when you submitted the manuscript. This funding data will not appear in the manuscript, but it will be provided to CrossRef in order to make the data publicly available. Therefore, please check it carefully for accuracy and mark any necessary corrections. Statements acknowledging financial support may also appear within the manuscript itself (in Acknowledgments); any such statements should also be checked for accuracy, but will have no bearing on funding data deposited with CrossRef.

Funder	Grant(s)	Author(s)	Funder ID
Queensland State Council: Koala Research Grant	KRG005	Steve D. Johnston	



## AUTHOR QUERIES

### AUTHOR PLEASE ANSWER ALL QUERIES

2

Department of Health   National Health and Medical Research Council (NHMRC)	Fellowship	Edward C. Holmes	<a href="https://doi.org/10.13039/501100000925">https://doi.org/10.13039/501100000925</a>
Australian Research Council (ARC)	LP0989701	Joanne Meers, Paul R. Young	<a href="https://doi.org/10.13039/501100000923">https://doi.org/10.13039/501100000923</a>

AQA—To ensure sequential order, references 26-34 have been renumbered in text and in References. Please check and correct the renumbering if necessary. If any reference should be deleted from the References list, please mark “reference deleted” in the margin next to that entry; do not renumber subsequent references.

AQB—Figure 3 was not cited in the original; OK to cite here? (All figures must be cited, in numerical order.) Also, the dashes in Table 1 have been deleted because dashes may not be used as placeholders for “zero” or “no data” per ASM style; if they are necessary, they must be defined in a footnote.

AQC—ASM policy requires that new nt/protein/microarray data be available to the public upon online posting of the article, so please verify all links to records (particularly for new sequences) and that each number retrieves the full record of the data (not just the home page). If the link takes you to an empty record, instruct the production staff to remove the link. If a new accession number is not linked in the proof or a link is broken, provide production staff with the specific URL for the record. (Note that the accession numbers for sequences determined in this study have been moved to the end of the text per ASM style.) If the accession numbers for new data are not publicly accessible by the proof stage, publication of your article may be delayed; please contact the ASM production staff immediately with the expected release date.

---