Changes in Endogenous and Exogenous Koala Retrovirus Subtype Expression over Time Reflect Koala Health Outcomes

Bonnie L. Quigley,^a Samuel Phillips,^a Olusola Olagoke,^a Amy Robbins,^{a,b} Jonathan Hanger,^b Peter Timms^a

^aGenecology Research Centre, University of the Sunshine Coast, Sippy Downs, Queensland, Australia

^bEndeavour Veterinary Ecology, Toorbul, Queensland, Australia

ABSTRACT Koala retrovirus (KoRV) is unique in that it exists as both an exogenous and actively endogenizing gamma retrovirus of koalas. While nine subtypes of KoRV have been recognized, focused study of these subtypes in koalas over time and with different health outcomes has been lacking. Therefore, in this study, three wild koala cohorts were established and monitored to examine KoRV proviral and expression data from koalas that either remained healthy over time, began healthy before developing chlamydial cystitis, or presented with chlamydial cystitis and were treated with antibiotics. Deep sequencing of the proviral KoRV envelope gene revealed KoRV-A, -B, -D, and -F to be the major subtypes in this population and allowed for subtype-specific assays to be created. Quantification of KoRV transcripts revealed that KoRV-D expression mirrored the total KoRV expression levels (10^6 copies/ml of plasma), with KoRV-A and KoRV-F expression being ~ 10 -fold less and KoRV-B expression being ~ 100 -fold less, when detected. Strikingly, there was significantly higher expression of KoRV-D in healthy koalas than in koalas that developed chlamydial cystitis, with healthy koalas expressing a major KoRV-D/minor KoRV-A profile, whereas koalas that developed cystitis had variable KoRV expression profiles. Total anti-KoRV IgG antibody levels were found not to correlate with the expression of total KoRV or any individual KoRV subtype. Finally, KoRV expression was consistent between systemic and mucosal body sites and during antibiotic treatment. Collectively, this gives a comprehensive picture of KoRV dynamics during several important koala health states.

IMPORTANCE The long-term survival of the koala is under serious threat, with this iconic marsupial being declared “vulnerable” by the Australian Government and officially listed as a threatened species. KoRV is clearly contributing to the overall health status of koalas, and research into this virus has been lacking detailed study of the multiple subtypes at both the proviral and expressed viral levels over time. By designing new subtype-specific assays and following well-defined koala cohorts over time, this study has generated a new more complete picture of KoRV and its relationship to koala health outcomes in the wild. Only by building a comprehensive picture of KoRV during both koala health and disease can we bring meaningful koala health interventions into better focus.

KEYWORDS *Chlamydia*, KoRV, koala, retroviruses

Koala retrovirus (KoRV) is a gamma retrovirus of koalas (*Phascolarctos cinereus*) (1). It is unique in the fact that while still existing as an exogenous virus, it has also recently (within the last 50,000 years) endogenized into the koala genome (2, 3). However, not all koalas possess the endogenous retrovirus (ERV) form of KoRV (4), indicating that the process of endogenization is still under way. With all other known mammalian ERVs having endogenized millions of years ago, KoRV has presented a

Citation Quigley BL, Phillips S, Olagoke O, Robbins A, Hanger J, Timms P. 2019. Changes in endogenous and exogenous koala retrovirus subtype expression over time reflect koala health outcomes. *J Virol* 93:e00849-19. <https://doi.org/10.1128/JVI.00849-19>.

Editor Frank Kirchhoff, Ulm University Medical Center

Copyright © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Peter Timms, p.timms@usc.edu.au.

Received 20 May 2019

Accepted 13 June 2019

Accepted manuscript posted online 26 June 2019

Published 28 August 2019

unique opportunity to study how retroviruses interact with new hosts and evolve in these early stages of genome invasion.

Closely related to Gibbon ape leukemia virus (GALV), KoRV has been divided into three major clades and nine subtypes based on sequence differences in the receptor binding domain (RBD) of the envelope (*env*) protein gene (1, 5–9). KoRV-A defines the most prevalent clade and subtype, being found in both captive and wild koalas in Australia and zoos abroad, and is the endogenized form of this virus (1, 2, 4, 9, 10). KoRV-B comprises the second clade and subtype and has also been found in both wild Australian koala populations and national and international captive communities (5–7, 9, 11). The third clade of KoRV encompasses the remaining seven subtypes, with KoRV-C, KoRV-D, KoRV-E, and KoRV-F initially identified in captive koalas from around the world (5, 12, 13) and later in wild koalas within Australia (8, 9), while KoRV-G, KoRV-H, and KoRV-I have been identified in diseased wild Australian koalas (9).

Our understanding of the effect KoRV has on koalas is slowly developing. Chlamydial disease is the most prevalent and devastating infectious disease that affects wild koalas, leading to blindness and infertility (14). Conversely, captive koala populations suffer from staggering rates of leukemia and lymphoma, with more than 60% of deaths in some colonies attributed to these malignant neoplasms (6). Before different subtypes were recognized, analysis of captive koalas found that total KoRV RNA levels were significantly increased in animals suffering from leukemia or lymphoma and notably higher in animals with chlamydial disease than in healthy animals (15). In the wild, detection of KoRV-A or KoRV-B provirus in the koala genome has been linked to chlamydial disease in Southern or Central/Northern Australian koalas, respectively (10, 16), while KoRV-B infection in general has been associated with neoplasia, leukemia, and lymphoma worldwide (6, 10). With mounting evidence that KoRV appears to be exacerbating and/or promoting these serious health challenges, the time has come for focused research into KoRV subtype prevalence and expression in relation to koala health.

To that end, this study monitored wild koalas that remained healthy over time, koalas that were determined healthy for a time before developing chlamydial cystitis, and koalas that presented with chlamydial cystitis and were treated with antibiotics. From these carefully chosen groups, integrated KoRV provirus and expressed/circulating virus were investigated over time and by subtype to build a picture of KoRV diversity and expression as it related to koala health outcomes.

RESULTS

Amplification of the proviral KoRV envelope gene receptor binding domain region to identify subtypes present in Queensland koalas. Sixteen female koalas (generating 72 individual samples) were followed over time to represent two koala health outcomes. Group one was koalas that developed chlamydial cystitis (caused by *Chlamydia pecorum*) and consisted of five koalas that were determined healthy at two or more consecutive veterinary examinations followed by a diagnosis of chlamydial cystitis. Group two was koalas that remained healthy over time and consisted of 11 koalas that were determined healthy at three or more consecutive veterinary exams. Total nucleic acid was extracted from RNA later-stored plasma and separated into DNA and RNA fractions for KoRV characterization.

KoRV was initially characterized at the proviral level, using the DNA fraction, by established PCR (9) (universal KoRV *env* gene) (Fig. 1; see also Table S1 in the supplemental material) and Illumina sequencing targeting the RBD region within the *env* gene for all KoRV subtypes. Amplifiable DNA was confirmed in all 72 samples by the successful amplification of the koala β -actin gene target (using the established PCR from reference 5) (Table S1), and all samples generated the anticipated KoRV *env* amplicon for sequencing. The generated forward and reverse amplicon sequences were merged, trimmed for amplification primers, assembled into operational taxonomic units (OTUs) that represented sequences $\geq 99\%$ identical, and filtered to full-length sequence reads. After processing, an average of $83,042 \pm 23,585$ reads were produced

FIG 1 KoRV schematic diagramming the location of PCR primers used in this study. LTR, long terminal repeats; gag, group antigens gene; pol, polymerase gene; env, envelope gene; RBD, receptor bind domain. Primers are described in Table S1 in the supplemental material.

from each sample, representing a total of 110 unique OTUs. These 110 OTUs captured 95.2% of the full-length reads generated and were chosen to represent the KoRV proviral diversity of the data set. BLAST results comparing the OTUs to previously characterized KoRV *env* sequences placed 31 OTUs as KoRV-A, 27 OTUs as KoRV-B, 41 OTUs as KoRV-D, and 11 OTUs as KoRV-F.

KoRV-A is the only subtype of KoRV that has become endogenous in koalas, and it is known that koalas from the study region contain this endogenized KoRV (10). Therefore, it was not surprising that KoRV-A was detectable in every sample (Table 1) and represented 97% to 100% of the KoRV proviral *env* sequences detected within any sample. OTU A1026 was the most prevalent OTU in the study, accounting for 86% of all KoRV-A proviral *env* sequences and 85.7% of all KoRV proviral *env* sequences detected. However, the other 30 KoRV-A OTUs were detectable in all 72 samples, indicating that diversity within this endogenous *env* region was still present and maintained in all 16 koalas tested.

KoRV-B, KoRV-D, and KoRV-F are exogenous virus subtypes in koalas, and their presence as detectable provirus was limited to a maximum of 0.8%, 2.4%, and 0.6% of reads detected in any sample, respectively. KoRV-B provirus was detectable in 4/16 koalas (25%) (Table 1), which is congruent with the 24% KoRV-B infection rate for koalas from this population (based on detectable KoRV-B *env* proviral DNA from blood samples [10]). Infection rates for the other KoRV subtypes have not been reported, as specific subtype assays have not been available. Interestingly, KoRV-D, which represented the most diverse proviral *env* sequence with 41 OTU representatives, was detectable in 14/16 koalas (88%), while KoRV-F provirus was only detectable in 4/16 koalas (25%) (Table 1). The other established KoRV subtypes (C, E, G, H, and I) were not detected as proviral elements in the samples tested.

KoRV-D/F clade contains more diversity than KoRV-B or KoRV-A clades in proviral *env* sequences. A minimum spanning tree (Fig. 2) and a maximum likelihood phylogenetic tree (see Fig. S1) were generated from all the KoRV *env* proviral OTUs, including a representative reference sequence from each established KoRV subtype (A to I) and a corresponding *env* sequence from GALV as an outgroup. Both methods revealed relationships seen in a previous study (9), with KoRV-A most closely related to GALV and three distinct KoRV clades formed by KoRV-A, KoRV-B, and KoRV-D to -I (Fig. 2). Also in agreement with previous proviral study, the KoRV-A sequence clade studied contained the least *env* diversity, while KoRV-D to -I sequence clade had the greatest *env* diversity (Fig. 2).

TABLE 1 Detectable KoRV in time course koalas (*n* = 16)^a

KoRV subtype	% (<i>n</i>) <i>env</i> provirus detectable by:		Avg <i>env</i> provirus when detected (copies/ml of plasma)	<i>env</i> gene RNA detectable by qPCR (% [<i>n</i>])	Avg expression <i>env</i> RNA when detected (copies/ml of plasma)
	Amplicon sequencing	qPCR			
KoRV-A	100 (16)	100 (16)	1.31 × 10 ⁶	100 (16)	8.08 × 10 ⁵
KoRV-B	25 (4)	0 (0)	BD ^b	94 (15)	2.45 × 10 ⁴
KoRV-D	88 (14)	0 (0)	BD	81 (13)	6.16 × 10 ⁶
KoRV-F	25 (4)	0 (0)	BD	31 (5)	8.41 × 10 ⁴

^aDetectable in at least 1 sample from koala.
^bBD, below detection limit (1 × 10³ copies/ml of plasma).

FIG 2 Minimum spanning tree diagramming the relatedness of the 110 KoRV proviral OTUs detected from deep amplicon sequencing analysis. GALV, gibbon ape leukemia virus. KoRV-A OTU is in green, KoRV-B OTU in red, KoRV-D OTU in blue, and KoRV-F OTU in yellow. Reference sequences are indicated with a bolded outline and include KoRV-A (Hanger [AF151794.2](#)), KoRV-B (Shojima [AB822554.1](#)), KoRV-C (Shojima [AB828005.1](#)), KoRV-D (Shojima [AB828004.1](#)), KoRV-E (Chappell [D15] [KX588043.1](#)), KoRV-F (Chappell [F12] [KX587994.1](#)), KoRV-G (Chappell [G3] [KX587961.1](#)), KoRV-H (Chappell [H1] [KX588036.1](#)), KoRV-I (Chappell [I1] [KX588021.1](#)), and GALV (Wilson [U20589.1](#):465-1019).

KoRV env CETTg motif attenuated in some, but not all, proviral env sequences.

An important motif in the RBD of retroviral env proteins is the CETTg motif (amino acids 167 to 171 in KoRV-A) (6, 17). In the proviral env sequences determined from the koalas in this study, different proportions of intact and attenuated CETTg motifs were detected in different KoRV subtypes (Table 2). All 16 koalas contained the same three variations in the KoRV-A proviral env motifs, in the same proportions: intact CETTg motif (7 OTU, 2.3% of KoRV-A proviral reads), attenuated CETAG motif (23 OTU, 97.4% of KoRV-A proviral reads), and a novel CETVG motif (1 OTU, 0.3% of KoRV-A proviral reads) (Table 2). From the four koalas with detectable KoRV-B env provirus, all possessed only the attenuated CETAG motif (27 OTU). These KoRV-B env proviruses contrast with the original KoRV-B isolate described with an intact CETTg motif (6). The 14 koalas with detectable KoRV-D env provirus contained one, two, or three variations of the motif: intact CETTg (8 OTU, 13% of KoRV-D proviral reads), attenuated CETAG motif (21 OTU, 60% of KoRV-D proviral reads), and a novel CETSG motif (12 OTU, 27% of KoRV-D proviral reads) (Table 2). Finally, four koalas had detectable KoRV-F env provirus containing one of two variations of the motif: intact CETTg motif (4 OTU, 34% of KoRV-F proviral reads) and attenuated CETAG motif (7 OTU, 66% of KoRV-F proviral reads) (Table 2).

Development of subtype-specific PCR assays for KoRV-D and KoRV-F. Detection and quantification of individual KoRV subtypes directly from clinical samples requires

TABLE 2 CETTg motif in KoRV env proviral OTUs obtained from deep amplicon sequencing of Queensland koalas

KoRV subtype (no. of koalas with provirus)	CETTg motif observed	Hypothesized effect on syncytium formation	No. of OTUs with motif	Proportion of proviral reads (%)
KoRV-A (16)	CETTg	Wild-type capability	7	2.3
	CETAG	Attenuated	23	97.4
	CETVG	Attenuated	1	0.3
KoRV-B (4)	CETAG	Attenuated	27	100
KoRV-D (14)	CETTg	Wild-type capability	8	13
	CETAG	Attenuated	21	60
	CETSG	Attenuated	12	27
KoRV-F (4)	CETTg	Wild-type capability	4	34
	CETAG	Attenuated	7	66

subtype-specific PCR primers to the variable region of the *env* gene for quantitative PCR (qPCR). Subtype-specific primers for KoRV-A and KoRV-B are currently available (18), using a universal KoRV *env* forward primer that binds to all KoRV subtypes (located 46 bp from the start of the KoRV-A *env* gene) and specific reverse primers designed in the variable RBD region of each subtype. To evaluate whether the current KoRV-A and KoRV-B primer sets would capture the diversity of *env* gene OTUs detected in this study, as well as to design and test new KoRV-D and KoRV-F subtype-specific primers, all 110 *env* gene OTU sequences were aligned and inspected for PCR primer landing sites. Based on the alignment (summarized in Fig. S2), two degenerate positions were added to the established universal KoRV *env* forward primer to capture all the diversity observed (Table S1). The established KoRV-A reverse primer required no modification and was used as previously published (18), while the KoRV-B reverse primer was shifted several nucleotides, and a degenerate position was added to encompass all the KoRV-B *env* gene diversity encountered (Table S1). New KoRV-D and KoRV-F reverse primer sequences, which included degenerate positions, were chosen from the alignment to specifically amplify only those *env* subtypes (Table S1, Fig. S2). All four KoRV subtype-specific reverse primers were tested with the universal forward primer on a panel of DNA samples containing different combinations of known proviral subtypes (based on the provirus sequencing results discussed above), optimized for annealing temperature and primer concentration by qPCR, and sequenced to confirm amplicons generated were specific to the subtype targeted (example qPCR standard and melt curves for KoRV-B, KoRV-D, and KoRV-F assays are given in Fig. S3). All four subtype-specific PCR primer sets generated amplicons of the expected size only in samples known to contain that subtype, and all products sequenced matched the intended subtype targeted.

Only KoRV-A provirus was quantifiable by qPCR. Using the four subtype-specific KoRV qPCR assays (different from the universal *env* primer set used for deep amplicon sequencing), DNA samples were quantified for the amount of provirus present in each sample. For comparison to a koala gene, the koala β -actin gene was also quantified from each sample (5). KoRV-A provirus was detected in every sample and was present at an average level of 1.31×10^6 copies/ml of plasma (Table 1). KoRV-B, KoRV-D, and KoRV-F provirus were undetectable in all 72 samples, indicating that their presence was below the detection limit of the qPCR assay of 1×10^3 copies/ml of plasma (Table 1).

Average KoRV-D *env* expression was higher than KoRV-A, KoRV-B, and KoRV-F *env* expression. To determine the levels of expressed KoRV from each sample, the RNA fraction from each sample was DNase treated, reverse transcribed, and quantified by qPCR. In addition to the koala β -actin gene and KoRV *env* gene subtype-specific assays, total KoRV expression was determined using an established assay targeting a region of the KoRV polymerase (*pol*) gene conserved between all subtypes (15) (Table S1). Efficient removal of contaminating DNA from the RNA fraction was ensured by testing the RNA fraction samples with the KoRV *pol* assay after DNase treatment and confirming all samples were negative. Koala β -actin gene expression was determined to allow for comparison to previously published work and to serve as a marker for RNA quality. We found 71/72 (99%) of samples had detectable β -actin expression to an average level of 3.74×10^4 copies/ml of plasma.

Total KoRV *pol* expression was detectable at some point in all 16 koalas (64/72 [89%] samples), with an average expression of 1.36×10^7 copies/ml of plasma (minimum 5.21×10^3 copies/ml of plasma, maximum 1.71×10^8 copies/ml of plasma). When KoRV subtypes were specifically quantified, KoRV-A *env* gene expression was detectable at some point in all 16 koalas (59/72 [82%] samples) at an average level of 8.08×10^5 copies/ml of plasma. KoRV-B *env* gene expression was detected sporadically in 15/16 (94%) koalas (34/72 [47%] samples) at an average level of 2.45×10^4 copies/ml of plasma. KoRV-D *env* gene expression was regularly detectable in 13/15 (81%) koalas (55/72 [76%] samples) at an average level of 6.16×10^6 copies/ml of plasma. Finally, KoRV-F *env* gene expression was detectable in only 5/15 (31%) koalas (12/72 [17%] samples) at an average level of 8.41×10^4 copies/ml of plasma (Table 1). Quantities

FIG 3 Total KoRV and KoRV subtype expression by koala health outcome. Box and whisker plots are displayed for each category, and individual sample results are shown as black dots. Healthy koalas ($N = 11$) were sampled a total of 53 times, while koalas that developed cystitis ($N = 5$) were sampled a total of 19 times during this study. Using a mixed model ANOVA (to account for repeated measures) on the last 3 samples from each koala, total KoRV-pol expression levels ($F[1,14] = 10.480$, $P = 0.006$) and KoRV-D env expression levels ($F[1,14] = 13.760$, $P = 0.002$) were significantly different between healthy koalas and koalas that developed cystitis.

detected of each target in each sample are graphed by individual koala in Fig. S4. Overall, KoRV-D env expression was only 2-fold less than the total KoRV pol expression levels, with KoRV-A expression 17-fold less, KoRV-F expression 161-fold less, and KoRV-B expression 555-fold less when detected.

KoRV-D env expression is significantly higher in koalas that remained healthy than in koalas that developed chlamydial cystitis. KoRV-A and KoRV-B proviral detection has been linked to chlamydial disease in koalas in South Australia and Central/Northern Australia, respectively (10, 16). To determine whether the expression of different subtypes of KoRV could be linked to chlamydial disease, samples were grouped and analyzed based on whether the tested koala remained healthy over the monitoring period (11 koalas; 53 samples) or whether the koala developed chlamydial cystitis during monitoring (5 koalas; 19 samples) (Fig. 3). Strikingly, there was higher expression of total KoRV pol in healthy koala samples than in samples from koalas that developed cystitis. Taking the last three sampling points from all koalas in the study, we found that this expression difference was statistically significant (mixed models analysis of variance [ANOVA] $F[1,14] = 10.480$, $P = 0.006$). This difference was accounted for in the KoRV-D env expression (Fig. 3), with the last three sampling points from all koalas also showing significantly higher KoRV-D env expression in healthy koala samples than in koalas that developed cystitis (mixed models ANOVA $F[1,14] = 13.760$, $P = 0.002$). In contrast, there was no significant difference in the expression of KoRV-A, KoRV-B, or KoRV-F env between the two clinical koala groups.

Consistent KoRV env subtype expression, dominated by KoRV-D env, is representative of healthy koalas over time. To further characterize the KoRV env subtype expression patterns within koalas over time, pie charts were generated to reflect the proportion of the total KoRV env expression each subtype contributed to each sample (Fig. 4). This graphical timeline highlights how much each subtype contributed to the total KoRV env expression at any time point and whether this expression profile changed between samples over time. Overlaid on the graphical timelines are *C. pecorum* testing results (from both ocular [OC] and urogenital [UGT] sites) for each sample. Immediately apparent was the strikingly consistent pattern of KoRV env expression profiles within the healthy koalas, made up of a major component of KoRV-D and a minor component of KoRV-A (Fig. 4). This contrasted sharply with the variable KoRV env expression profiles in the koalas that developed cystitis, often lacking major KoRV-D expression and occasionally lacking any detectable KoRV env expression at all (Fig. 4). *C. pecorum* infection was not common in this sample group, being detected sporadically in both healthy koalas (some containing the ompA genotype A')

A. Developed cystitis

B. Healthy

FIG 4 KoRV subtype expression profile over time and in relation to clinical observations. Numbers on the arrows between pie charts indicate months between samples. *Chlamydia pecorum* was tested in each sample, and positive or missing samples are indicated in the bottom right of each pie chart. M, missing swab (not tested); OC, positive at ocular site; UGT, positive at urogenital site; low, <500 copies/ μ l swab extract; high, >500 copies/ μ l swab extract. When *C. pecorum* genotypes were determinable, the *ompA* genotype is indicated, as well.

and koalas that developed cystitis (some containing the *ompA* genotype G). However, limited numbers hinder the ability to draw any links between KoRV *env* expression and the presence of chlamydial organisms.

Anti-KoRV *env* IgG levels do not correlate with KoRV subtype *env* expression levels. To further investigate the link between KoRV and koala health, the total circulating anti-KoRV *env* IgG antibody levels were determined for each sample from matched serum samples. Our assay utilized a near full-length KoRV-A *env* protein, and so to gain confidence that antibodies to multiple KoRV subtypes would bind this KoRV-A-based substrate, multiple full-length *env* sequences were obtained from

FIG 5 KoRV expression profiles before, during, and after antibiotic treatment for *Chlamydia* cystitis at systemic and mucosal body sites for koala Ed (A) and koala Rora (B). PBMC, peripheral blood mononuclear cells; UGT, urogenital swabs.

GenBank and aligned. Comparison of KoRV-A, -B, -C, -D, and -F protein sequences revealed no less than 86% identity between KoRV-B and KoRV-A env proteins and no less than 91% identity between KoRV-C, -D, and -F env to KoRV-A env. Given that the majority of differences observed were in the RBD, the high degree of sequence conservation throughout the rest of the env protein lead us to continue with this assay to measure total anti-KoRV env IgG. End point titers of anti-KoRV env IgG antibody averaged $3,911 \pm 2,013$ for 70/72 sample time points (two time points did not have matched serum samples). Interestingly, there was no correlation between anti-KoRV env IgG levels and the expression of total KoRV (KoRV-pol, $P = 0.935$) or any individual KoRV env subtype (KoRV-A, $P = 0.912$; KoRV-B, $P = 0.118$; KoRV-D, $P = 0.234$; KoRV-F, $P = 0.066$).

KoRV env expression is consistent between mucosal and systemic body sites and is unaffected by antibiotic treatment for chlamydial cystitis. Given the apparent importance of stable dominant KoRV-D env expression to koala health, we sought to investigate KoRV env expression in more detail during and after antibiotic treatment for chlamydial cystitis. To that end, we changed methodologies and employed total RNA sequencing to a depth of ~ 30 million reads per sample to follow two koalas with chlamydial cystitis during treatment, recovery, and release. Both koalas displayed stable KoRV env subtype expression over the 8 weeks they were admitted, at both their systemic (peripheral blood mononuclear cells [PBMC]) and mucosal (urogenital) sites, and both sites shared the same KoRV expression patterns (Fig. 5; Fig. S5). Koala Ed expressed KoRV-A, KoRV-D, and KoRV-F env transcripts at a level of $\sim 100,000$ transcripts/million each in every sample, with lower levels ($< 5,000$ transcripts/million) of the other KoRV subtypes (B, C, E, G, H, and I) detected sporadically (Fig. S5). Alternatively, koala Rora expressed KoRV-A and KoRV-B env transcripts at a level of $\sim 100,000$ transcripts/million each in every sample, with KoRV-C, KoRV-D, KoRV-E, KoRV-F, and KoRV-I detected at expressions levels of $\sim 10,000$ transcripts/million, while KoRV-G and KoRV-H were virtually undetectable (Fig. S5). The KoRV env expression pattern did not change considerably between samples collected before antibiotic treatment (during

cystitis), during antibiotic treatment, or up to 1 month postantibiotic treatment for either koala (Fig. 5).

DISCUSSION

KoRV was initially documented in koalas in 2000 (1), and since then, it has become recognized as a common infection of koalas, both in captivity and in the wild. Studies have linked general KoRV infection and infection with the specific subtypes KoRV-A and KoRV-B to serious koala conditions, including chlamydial disease and malignant neoplasms (6, 10, 16). However, focused study of KoRV subtypes beyond KoRV-A and KoRV-B has been limited to proviral investigation (9), and there has been no study of expressed/circulating KoRV by subtype or over time. To redress these gaps, this study investigated several carefully chosen koala cohorts over time to determine KoRV proviral and expressed envelope gene properties in relation to koala health outcomes.

In this study, we investigated the RBD region of the KoRV *env* in both its proviral and expressed forms from koala cell-free plasma. Plasma is the established sample of choice for KoRV RNA expression studies (15) and does contain DNA from several sources, including apoptosis, spontaneous cellular release, breakdown of blood cells, bacteria, viruses, leukocyte surface DNA, cell and tissue necrosis, cellular release of exosomes, transposons, and retrotransposons, with apoptotic and spontaneous release contributing the largest fractions (19–21). Acknowledging that DNA recovered from cell-free plasma will be less than from cellular sources, we sought to verify that our *env* proviral detection results (by deep amplicon sequencing from the DNA fraction) accurately reflected our expressed *env* transcript results (by qPCR detection from the RNA fraction). Since all the koalas in this study had matched results from the same sample, we were able to compare the two methods for subtype detected correspondence. For the vast majority of samples and subtypes, detection of proviral sequences corresponded with the detection of expressed *env* transcripts. KoRV-A proviral sequence and expressed *env* reads were detectable in every koala. Of the four koalas with detectable KoRV-B proviral sequence, all four had detectable levels of KoRV-B *env* expression. For KoRV-D, only two koalas had no KoRV-D *env* provirus detected, with a third koala having very few proviral sequences; these three koalas were the only ones with no expressed KoRV-D *env* detected. Finally, KoRV-F provirus was detected in four koalas, and three of them also had detectable KoRV-F *env* expression. Overall, the limited discrepancies between methods occurred where KoRV-B or KoRV-F was detected as expressed *env* RNA by qPCR and the corresponding proviral sequences fell below the deep sequencing detection limit. Given the majority of concordance between proviral deep sequencing and *env* transcript expression, we found that plasma was an acceptable source of proviral DNA for investigation. Additionally, we found that proviral deep sequencing was more sensitive than direct proviral subtype-specific qPCR quantification, and qPCR quantification of provirus from plasma-derived DNA is not recommended for future studies.

KoRV-A is the dominant subtype of KoRV present in koalas and has become endogenized in the koala populations other than southern koalas. In the KoRV-A proviral *env* sequences detected from the 16 koalas studied, all animals shared the same 37 OTU sequences in the same relative proportions. This is consistent with these sequences being endogenized as permanently integrated elements in the koala genome that are now shared in the population through vertical transmission. After endogenization, retroviruses are generally subjected to recombination and rearrangement events that eventually result in complete or partial inactivation of the viral genes (22). Interestingly, the KoRV-A *env* transcript was still detected at almost 10^6 copies/ml of plasma, indicating that KoRV-A *env* provirus copies are still being transcribed. Although we could not distinguish what proportion of these KoRV-A transcripts were from packaged RNA genomes and which were from released mRNA transcripts from host cells, the detection of any KoRV-A *env* in RNA form indicates that the genome inactivation of KoRV-A is not complete.

The striking KoRV-B finding was that current estimates of KoRV-B infection rates may

be grossly underestimated in koalas. In this study of mostly healthy koalas, KoRV-B *env* expression was detected within at least one sample from a koala in 15/16 animals (94%). This mirrors previous proviral deep sequencing results, with 14/18 (78%) severely ill koalas having detectable KoRV-B provirus (9). Currently, screening for KoRV-B from clinical samples targets the proviral KoRV-B *env* gene from genomic DNA preparations at a single time point, with Queensland and New South Wales populations currently believed to have ~25% KoRV-B prevalence and southern koala populations (Victoria and South Australia) believed to be KoRV-B free (10, 16). This survey found that KoRV-B *env* expression was often detected in only one or two samples from a koala and at levels that were orders of magnitude lower than other KoRV subtypes. Generally, low levels of both KoRV-B *env* provirus and expressed transcripts in otherwise healthy koalas could be falling below the detection limit of current methods and leading to underestimations of the KoRV-B infection rate in koala populations. In addition, increases in both KoRV-B proviral and expressed loads to levels that are more readily detectable may be a contributor to the pathologies that have become associated with KoRV-B infection (increased chlamydial disease and malignant neoplasms [6, 10]).

The high level of KoRV-D *env* expression found in healthy koalas in this study was an unexpected finding. KoRV-D *env* provirus was expected to be common among koalas but not abundant in terms of copy number, having been previously reported in 17/18 (94%) Queensland koalas but representing only a maximum of 16.1% of the deep sequencing KoRV *env* reads for any koala (9). The diversity of KoRV-D *env* OTUs found in this study was also anticipated, as the same study found similar diversity in the KoRV-D sequences detected (9). However, it was very surprising that KoRV-D *env* transcripts were detected at levels higher than KoRV-A expression levels. These levels of KoRV expression were independently confirmed by assaying an alternative KoRV gene target, the conserved KoRV *pol* gene, to ensure total KoRV expression levels were reaching levels that could not be explained by KoRV-A expression alone. KoRV-D *env* expression was 8-fold higher than KoRV-A *env* expression. One possible reason for this substantial expression difference may be found in the long terminal repeats (LTRs) of these subtypes. The U3 untranslated region of KoRV LTRs are known to contain different numbers and types of direct repeat enhancer element sequences, leading to different numbers of transcription factor binding sites (13, 23). While the number of these enhancer elements has been shown to vary within subtypes, KoRV-A isolates are generally reported with one copy of the enhancer, while a characterized KoRV-B isolate has been found to contain four repeats and a characterized KoRV-F isolate has been found to possess five repeats (13). Given the variation known to exist, this region of the LTR is worthy of future study to determine if increases in enhancer region duplication could contribute to expression differences detected between KoRV subtypes.

A complicating factor in interpreting what KoRV-D *env* expression could mean to the koala is the fact that, to date, KoRV-D provirus has only been detected in the koala genome as a defective variant. It is known that KoRV can exist in the koala genome in several forms, including full-length provirus (containing LTRs and complete *gag*, *pol*, and *env* genes), defective provirus (containing LTRs and a complete *env* gene, but missing part or all of *gag* and *pol* genes), and recombinant (rec) KoRV (an alternative retrovirus sequence between LTRs) (23). Where KoRV-D provirus has been examined, it has only been found in the defective provirus state (13, 23). The conserved structure of the defective KoRV-D provirus found within the completely sequenced koala genome suggested that it is transmitted as a defective virus (23). Our study focused on detection of the *env* gene and transcript within KoRV, preventing us from commenting on the intact or defective status of subtypes detected. However, this study shows that KoRV-D *env* is expressed in plasma and adds evidence that KoRV-D, in some form, could be transmissible. If the KoRV-D detected in this study is defective, that may contribute to the explanation of why high levels of expressed KoRV-D are not detrimental to the health of koalas.

The other named subtypes of KoRV appear to be less common or limited to specific circumstances in the koala population. KoRV-F appears to be a variant of KoRV-D that

is less established in the koala population. A previous study found KoRV-F *env* proviral sequence in 8/18 (44%) koalas tested (9), while our present study found KoRV-F *env* proviral sequence in 4/16 (25%) animals. KoRV-C, KoRV-E, and KoRV-I were only detectable in this study from the two koalas that were deep sequenced to the level of millions of transcripts per sample. Even at this level, only one koala had notable levels of these three subtypes. Thus far, KoRV-C detection has been limited to captive colonies in Japan and the United States (5, 12), while KoRV-E detection has been limited to a captive colony in the United States (13). KoRV-G, KoRV-H, and KoRV-I were only previously reported in two, one, and one koala, respectively, with severe chlamydial disease (KoRV-G), leukemia (KoRV-I), and unknown health status (KoRV-H) (9). In our present study, the only koala found to have KoRV-I had chlamydial cystitis, and KoRV-G and KoRV-H were not detected at consistent levels. Looking at the close relationships between KoRV-D to KoRV-I subtypes by minimum spanning tree analysis, further focused study will be needed to determine if these subtypes use distinct receptors for entry or whether they represent minor variations of more established KoRV-D or KoRV-F subtypes.

While overall KoRV and KoRV-D expression was highest in healthy koalas in this study, this result does not directly conflict with previous reports that indicated that koalas with chlamydial disease had higher levels of KoRV than their healthy counterparts. In a previous study, healthy koalas were found to have between 1.7×10^5 and 3.6×10^9 (mean 7.7×10^7) KoRV *pol* copies/ml of plasma (15). In this study, samples from koalas healthy at the time of sampling (all the healthy koala samples and samples from koalas prior to developing cystitis) showed a similar range of 5.2×10^3 to 1.7×10^8 (mean 1.3×10^7) copies/ml of plasma when KoRV *pol* was detected. The difference between the studies becomes apparent in the KoRV *pol* levels in koalas with chlamydial disease, with a previous study indicating koalas with chlamydial disease had KoRV *pol* levels from $10^{6.5}$ to 10^{10} (mean 6.9×10^8) copies/ml of plasma (15), while this study found levels from 10^3 to 10^6 (mean 7.7×10^5) copies/ml of plasma when KoRV *pol* was detected. Differences between these studies include the numbers of koalas tested with chlamydial disease (20 koalas in the previous study, only five koalas in this study) and the fact that koalas in this study were closely monitored for health and recaptured quickly when disease presented, while koalas from the previous study were brought to wildlife hospitals after being ill for an unknown period of time. These differences in KoRV viral load results indicate that longitudinal testing is needed in koalas with chlamydial disease to understand when during disease presentation virus loads may increase and what role different subtype expression patterns may be contributing to overall KoRV viral loads.

A previous study found that KoRV-A *env* containing mutations in the CETTG motif had reduced cytopathology (47% to 64% reduction of syncytium formation) compared to that from *env* constructs with intact motifs (17). A suggested mechanism for this reduced cytopathology was the proposed loss of an O-linked glycosylation site (and glycan) at the second T (CETTG) position (17). In our proviral analysis, we found this second T position to be the only position in this motif with alternative amino acids, occurring in 66% to 100% of *env* sequences detected by subtype. This suggests that large proportions of expressed KoRV may have reduced cytopathology compared to that for related gamma retroviruses with completely intact CETTG motifs. A trend has been observed that known mammalian gamma retroviral *env* sequences derived from infectious murine leukemia viruses (MLVs), feline leukemia viruses (FLVs), GALV, and woolly monkey virus (WMV) have intact CETTG motifs, while gamma retroviral *env* sequences from endogenous viruses such as porcine endogenous retrovirus (PERV) B and C and baboon endogenous virus (BAEV) have mutations to this site (17). The observation that KoRV appears to have both intact and mutated CETTG motifs supports our current understanding of this unique virus as having both infectious/exogenous and endogenous characteristics.

To investigate whether the expression of different KoRV *env* subtypes had any relationship to anti-KoRV antibodies detectable within an animal, we compared anti-

KoRV IgG levels to *env* expression of each subtype. A KoRV-A-derived *env* protein was used in the antibody assay (with greater than 86% identity to the other subtype protein sequences), and no correlation was found between any KoRV subtype *env* expression and anti-KoRV *env* IgG levels. While it is possible that antibodies to non-KoRV-A subtypes may have bound poorly in our assay and been missed, the lack of correlation between anti-KoRV *env* IgG levels and KoRV-A expression suggests that transcription levels of the *env* protein are not a strong predictor of KoRV *env* antibodies in koalas.

By monitoring two koalas at multiple body sites and during antibiotic treatment using total RNA sequencing, we were able to answer some basic questions. The first was that paired samples tested from PBMCs and urogenital swabs gave the same KoRV profile, indicating that KoRV *env* expression does not appear to be different between systemic and mucosal body sites in the koala. In a clinical setting, swabs are routinely collected for chlamydia testing and, with the minor addition of an RNA preservative, could also be used for KoRV expression testing. This could simplify sample collection processes (removing the need to collect blood) and increase the opportunity to test and monitor KoRV subtype expression. As well, monitoring of KoRV expression during chlamydial disease, during antibiotic treatment and after recovery, showed no impact or alteration of KoRV expression, indicating that these factors do not have a major influence on this retrovirus's expression.

The long-term survival of the koala is under serious threat, with this iconic marsupial being declared vulnerable by the Australian Government in 2012 and officially listed as a threatened species. Health challenges from *Chlamydia* and neoplasms are eminent pressures on koala populations, both in the wild and in captivity. KoRV is clearly contributing to the overall health status of koalas, and continued research into KoRV biology only expands the intricacies and complexities involved. As future study incorporates both proviral and expressed viral elements from all KoRV subtypes, a more complete picture of both KoRV and koala health outcomes will come into better focus.

MATERIALS AND METHODS

Animals and sample collection. (i) Time course koalas. Sixteen female koalas between the ages of 2 and 12 years were included in this study. They were selected from a 4-year population-wide management program by the Queensland Government Department of Transport and Main Roads for the Moreton Bay Rail (MBR) project, in the Moreton Bay Region, Queensland, Australia (project center point: 27.25°S, 153.02°E). Koalas in this population were captured, clinically examined by experienced wildlife veterinarians (both visually and by ultrasound), radio collared, and released back into the wild. Koalas were monitored by remote and field telemetry, recaptured, and subjected to veterinary examination and sampling at regular intervals (approximately 6-monthly). Blood, urogenital swabs, and ocular swabs were collected from koalas under general anesthesia during veterinary examinations. From blood, 200 μ l of plasma was separated and mixed with 300 μ l RNeasy lysis solution (Thermo Fisher Scientific). All samples were stored at -20°C until transport to the laboratory, where they were stored at -80°C until processing. All procedures were approved by the University of the Sunshine Coast (USC) Animal Ethics Committee (animal ethics number AN/A/13/80) and by the Queensland Government (scientific purposes permit WISP11532912). All experiments were performed in accordance with relevant guidelines and regulations.

(ii) Koalas with cystitis treated with antibiotics. Two koalas (one male, Ed, and one female, Rora) presented with signs of chlamydial cystitis (both visually and by ultrasound) at either Currumbin Wildlife Hospital (CWH) or Australia Zoo Wildlife Hospital (AZWH) and were transported (within 3 days) to AZWH for treatment. Before treatment, urogenital swabs and blood were collected under anesthesia. Rora then received chloramphenicol as a 60 mg/kg subcutaneous injection once a day for 28 days, while Ed received doxycycline (diluted 50:50 in sterile saline) at 5 mg/kg subcutaneously once a week for 4 weeks. Every 2 weeks, urogenital swabs and blood were collected, for up to 8 weeks. Urogenital swabs were placed directly in 500 μ l of RNeasy lysis solution, while blood samples were processed within 2 h of collection to isolate PBMCs. For PBMC isolation, 4 ml of blood was added to 4 ml phosphate-buffered saline (PBS) containing 2% fetal calf serum (FCS) and mixed by inversion. The mixture was then added to a 15-ml SepMate (*in vitro* diagnostic [IVD]) tube (STEMCELL Technologies) containing 4 ml of Ficoll-Paque (GE Healthcare Bio-Sciences AB) and centrifuged at $1,200 \times g$ for 10 min at room temperature (RT). PBMCs were harvested after centrifugation by pouring the top tube contents into a new 15-ml tube. Harvested PBMCs were then washed twice in PBS containing 2% FCS and centrifuged at $200 \times g$ for 10 min at RT. The resulting cells were resuspended in 200 μ l of RNeasy lysis solution and stored at -20°C until further use.

Preparation of DNA and RNA (cDNA) from plasma samples. Total nucleic acid was extracted from 72 plasma samples. The entire 500- μ l mixture of plasma/RNeasy lysis solution was extracted using the QIAamp Viral RNA minikit (Qiagen) according to the manufacturer's instructions. Nucleic acid was eluted into a final volume of 60 μ l, divided into a 20- μ l aliquot for DNA proviral analysis and a 40- μ l aliquot for RNA analysis. The RNA aliquots were treated with TURBO DNA-free (Thermo Fisher Scientific) as per the

manufacturer's instructions and further divided into two pools: the first for direct qPCR analysis to ensure efficient removal of all DNA and the second for cDNA synthesis using the iScript cDNA synthesis kit (Bio-Rad) as per the manufacturer's instructions. Purified DNA, RNA, and cDNA were stored at -20°C during processing and -80°C for long-term storage.

KoRV env proviral sequencing and OTU generation. KoRV *env* amplicons spanning the RBD of the *env* gene (positions 22 to 514 of the KoRV-A *env* gene) were generated with the universal *env* primers env22.F and env514.R (see Table S1 in the supplemental material) according to reference 9 with minor modifications. Primers were combined with a HotStarTaq Plus Master Mix kit (Qiagen) and purified DNA for 35 rounds of amplification at 55°C annealing to generate the KoRV *env* amplicons for sequencing. Amplicons were sent to the Ramaciotti Centre for Genomics (Sydney, Australia), where they were Illumina barcoded using eight rounds of amplification, followed by sequencing on a MiSeq sequencing system (Illumina) using V3 300-bp paired-end chemistry.

Forward and reverse amplicon reads were filtered to remove sequences less than 150 bp and primer trimmed with cutadapt (24), followed by merging with FLASH (25). Operational taxonomic units (OTUs) were determined within QIIME 1.9.1 (26) using UCLUST and USEARCH (27) to cluster reads with at least 99% identity and confirm clusters by mapping FLASH-merged reads back to the OTUs generated. OTUs generated from single reads (singletons) were removed from the data set, and the remaining OTUs were BLAST searched (28) against a library of known KoRV *env* sequences (generated from the data and reference sequences from reference 9). OTUs with less than a near full-length match (480 bp, usually representing non-KoRV artefacts) and in-frame stop codons and frameshifts were removed. Postprocessing, a final data set of 110 KoRV *env* proviral OTUs remained, which represented 95.2% of the FLASH-merged sequences generated from sequencing.

KoRV env proviral clustering and phylogenetic analysis. KoRV *env* proviral OTUs were aligned with reference sequences from the nine established KoRV subtypes (KoRV-A, Hanger [AF151794.2](#); KoRV-B, Shojima [11-4/OJ-4] [AB822553.1](#); KoRV-C, Shojima [11-2/OJ-4] [AB828005.1](#); KoRV-D, Shojima [11-1/OJ-4] [AB828004.1](#); KoRV-E, Chappell [D15] [KX588043.1](#); KoRV-F, Chappell [F12] [KX587994.1](#); KoRV-G, Chappell [G3] [KX587961.1](#); KoRV-H, Chappell [H1] [KX588036.1](#); KoRV-I, Chappell [I1] [KX588021.1](#)), as well as GALV (Wilson [U20589.1](#):465-1019) as an outgroup, using default parameters in ClustalW (29, 30).

For the minimum spanning tree, aligned DNA OTU sequences were inputted in PHYLOViZ Online (<http://online.phyloviz.net/index>) and used with the programs default goeBURST algorithm (31) to diagram the relatedness of KoRV *env* proviral OTU. For the phylogenetic tree, the evolutionary history of the aligned DNA OTU sequences was inferred by using the maximum likelihood method based on the Tamura-Nei model in MEGA7 (32, 33). The tree with the highest log likelihood ($-5,129.30$) was kept, with branch lengths measured in the number of substitutions per site. Nodes with $\geq 70\%$ bootstrap support (from 1,000 replicates) were labeled.

Quantitative PCR. Quantification of all targets in this study was achieved using iTaq Universal SYBR green Supermix (Bio-Rad) on a CFX 96 Touch System (Bio-Rad) by comparison to a standard curve generated from a dilution series of PCR products of a known concentration (10^6 to 10^1). Koala β -actin gene quantification (both DNA and RNA/cDNA) was carried out as described in reference 5, while total KoRV *pol* quantification (both DNA and RNA/cDNA) was carried out as described in reference 15. Subtype-specific KoRV *env* quantification (both provirus and expressed) was carried out with the primer sets listed in Table S1 (universal forward primer KoRV_UF [TCYTGGGAAGTGGRAAAGAC] with subtype-specific reverse primers KoRV_AR [GGGTTCCTCCCAAGTGATCTG], KoRV_BR [GACTAACCCCTGCCKACCT], KoRV_DR [GRTTCCCAAGKCGR] or KoRV_FR [GAYGTAAARCCAGGCCAAGG]), with KoRV-A primers used at a concentration of $0.3 \mu\text{M}$ while KoRV-B, KoRV-D, and KoRV-F primers were used at a concentration of $0.9 \mu\text{M}$. Reactions were performed with an initial denaturing of 95°C for 5 min, followed by 40 cycles of denaturation at 94°C for 15 s, annealing at 54°C for 15 s, extension at 72°C for 15 s, and data acquisition at 82°C for 10 s, followed by melt curve analysis from 65°C to 95°C in 0.5°C increments. Results were processed using Bio-Rad CFX Manager software.

Sample processing and testing for *Chlamydia pecorum*. Urogenital and ocular swab samples were DNA extracted as described in reference 34. The extracted samples were screened for the presence of *C. pecorum* using a diagnostic qPCR assay targeting a 208-bp fragment of the *C. pecorum* HP gene using the primers listed in Table S1 (35) and the qPCR conditions described in reference 36. When *C. pecorum*-positive samples were identified, *ompA* genotyping was performed as described in reference 11.

Anti-KoRV IgG endpoint titers. Total anti-KoRV IgG antibody levels were determined from serum as detailed in reference 37. Briefly, a near full-length recombinant KoRV-A *env* protein, minus the first 100 amino acids (inclusion of the first 100 amino acids prevented protein expression, therefore the recombinant protein begins at the RBD), was generated with a glutathione S-transferase (GST) tag for affinity purification (37). To compare the *env* protein conservation between KoRV subtypes, full-length *env* protein sequences were compared from KoRV-A ([AF151794](#), [BAN63359.1](#)), KoRV-B ([NC_021704.1](#), [BAN63358.1](#)), KoRV-C ([ALX81658.1](#), [BAN63361.1](#)), KoRV-D ([BAN63360.1](#)), and KoRV-F ([ANT96689.1](#)). Protein sequences were aligned using ClustalW (29, 30), and percent identities were calculated between sequences trimmed to the recombinant protein length using GenDoc (38). For the KoRV-specific IgG enzyme-linked immunosorbent assay (ELISA), serum from koalas was first preincubated with GST to remove GST antibodies. Assay plates were coated with $1 \mu\text{g}$ recombinant KoRV-A *env* protein overnight before washing and incubating with serially diluted serum samples. After additional washing, incubation with a sheep anti-koala IgG, further washing, incubation with a donkey anti-sheep horseradish peroxidase (HRP) IgG, final washing, and development with tetramethylbenzidine (TMB) liquid substrate (Sigma-Aldrich), optical density was read at 450 nm and endpoint titers (EPT) calculated (37).

Statistics. Mixed model ANOVAs (using a general linear model) were performed using SPSS software (released 2013, IBM SPSS Statistics for Windows, version 22.0; IBM Corp., Armonk, NY). To account for the non-normal distribution of RNA expression results, the actual copies per milliliter of plasma values were divided by the detection limit of the assay (1,000 copies/ml of plasma) and then natural log transformed before being evaluated. Because differing numbers of samples were evaluated for each koala, the last three sampling points were taken from each animal for statistical comparison.

RNA isolation from urogenital swabs and PBMCs. For total RNA extraction, 500 μ l of RNAlater swab homogenate or 200 μ l of harvested PBMCs was processed using the RNeasy Mini kit (Qiagen) according to the manufacturer's instructions for RNA purification from animal cells, spin protocol. After column isolation, 14 μ l of RNA solution was further treated with TURBO DNA-free (Life technologies) as per the manufacturer's instructions, followed by ethanol precipitation to remove any contaminating salts. All samples were then quantified for RNA determination using a Qubit RNA determination kit (Thermo Fisher Scientific) as per the manufacturer's instructions. All RNA aliquots were stored at -20°C until further use.

Total RNA sequencing and KoRV *env* transcript processing. Isolated RNA from PBMCs and urogenital swabs was sent to the Ramaciotti Centre for Genomics (Sydney, Australia) for library preparation and sequencing. Library preparation was performed in 12-sample batches using the SMARTer stranded total RNA-Seq kit v2-Pico input mammalian-strand-specific preparation kit (Clontech) and sequenced on a paired-end NextSeq 75-bp high-output run.

Sequence files were assessed for quality using FastQC (39) before KoRV *env* reads were pseudo-aligned to the nine reference KoRV subtypes (listed in "KoRV *env* proviral clustering and phylogenetic analysis" above) and quantified using Kallisto (40). Identified reads were normalized to transcripts per million and tabulated for analysis.

Data availability. The 110 unique KoRV OTUs generated during this study have been deposited in GenBank and assigned the accession numbers [MK390503](#) to [MK390612](#).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JVI.00849-19>.

SUPPLEMENTAL FILE 1, PDF file, 1.5 MB.

ACKNOWLEDGMENTS

This project was significantly supported by the Queensland Government (Department of Transport and Main Roads) and specifically the Moreton Bay Rail project team. Funding for this project was provided by the Australian ARC Linkage Scheme (to P.T.).

We thank the many groups that have supported our overall koala disease work, including the Queensland Department of Environment and Heritage Protection, Moreton Bay Regional Council, Gold Coast City Council, Friends of the Koala, Lismore, Koala Action Inc., Endeavor Veterinary Ecology, Australia Zoo Wildlife Hospital, Lone Pine Koala Sanctuary, Currumbin Wildlife Sanctuary, Redland City Council, Zoos South Australia, and VIDO, Canada. Specifically, we thank the Department of Transport & Main Roads—Moreton Bay Rail Project Team for collaborating with this project enabling access to the work undertaken as part of the Moreton Bay Rail-Koala Tagging and Monitoring Program and the dedicated staff at Endeavor Veterinary Ecology, particularly, Jo Loader, for their help in capturing, radio collaring, and tracking the koalas, and undertaking the health assessments as well as collecting samples. We also thank Justin Scott and Farah Zahir from QFAB Bioinformatics, Institute for Molecular Biosciences, University of Queensland, Australia, for statistical support and advice.

REFERENCES

- 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>.
- Tarlinton RE, Meers J, Young PR. 2006. Retroviral invasion of the koala genome. *Nature* 442:79–81. <https://doi.org/10.1038/nature04841>.
- 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>.
- 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>.
- 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, 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 U.S. zoo. *Proc Natl Acad Sci U S A* 110:11547–11552. <https://doi.org/10.1073/pnas.1304704110>.
- 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>.
- 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>.
9. 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>.
 10. Quigley BL, Ong VA, Hanger J, Timms P. 2017. Molecular dynamics and mode of transmission of Koala Retrovirus as it invades and spreads through a wild Queensland koala population. *J Virol* 92:e01871-17. <https://doi.org/10.1128/JVI.01871-17>.
 11. Nyari S, Waugh CA, Dong J, Quigley BL, Hanger J, Loader J, Polkinghorne A, Timms P. 2017. Epidemiology of chlamydial infection and disease in a free-ranging koala (*Phascolarctos cinereus*) population. *PLoS One* 12: e0190114. <https://doi.org/10.1371/journal.pone.0190114>.
 12. 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>.
 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>.
 14. Polkinghorne A, Hanger J, Timms P. 2013. Recent advances in understanding the biology, epidemiology and control of chlamydial infections in koalas. *Vet Microbiol* 165:214–223. <https://doi.org/10.1016/j.vetmic.2013.02.026>.
 15. 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. <https://doi.org/10.1099/vir.0.80547-0>.
 16. Legione AR, Patterson JL, Whiteley P, Firestone SM, Curnick M, Bodley K, Lynch M, Gilkerson JR, Sansom FM, Devlin JM. 2017. Koala retrovirus genotyping analyses reveal a low prevalence of KoRV-A in Victorian koalas and an association with clinical disease. *J Med Microbiol* 66: 236–244. <https://doi.org/10.1099/jmm.0.000416>.
 17. 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>.
 18. Waugh CA, Hanger J, Loader J, King A, Hobbs M, Johnson R, Timms P. 2017. Infection with koala retrovirus subgroup B (KoRV-B), but not KoRV-A, is associated with chlamydial disease in free-ranging koalas (*Phascolarctos cinereus*). *Sci Rep* 7:134. <https://doi.org/10.1038/s41598-017-00137-4>.
 19. Gahan PB. 2010. Circulating nucleic acids in plasma and serum: diagnosis and prognosis in cancer. *EPMA J* 1:503–512. <https://doi.org/10.1007/s13167-010-0021-6>.
 20. Abolhassani M, Tillotson J, Chiao J. 1994. Characterization of the release of DNA by a human leukemia-cell line hl-60. *Int J Oncol* 4:417–421. <https://doi.org/10.3892/ijo.4.2.417>.
 21. Stroun M, Lyautey J, Lederrey C, Olson-Sand A, Anker P. 2001. About the possible origin and mechanism of circulating DNA apoptosis and active DNA release. *Clin Chim Acta* 313:139–142. [https://doi.org/10.1016/S0009-8981\(01\)00665-9](https://doi.org/10.1016/S0009-8981(01)00665-9).
 22. Xu W, Eiden MV. 2015. Koala retroviruses: evolution and disease dynamics. *Annu Rev Virol* 2:119–134. <https://doi.org/10.1146/annurev-virology-100114-055056>.
 23. Hobbs M, King A, Salinas R, Chen Z, Tsangaras K, Greenwood AD, Johnson RN, Belov K, Wilkins MR, Timms P. 2017. Long-read genome sequence assembly provides insight into ongoing retroviral invasion of the koala germline. *Sci Rep* 7:15838. <https://doi.org/10.1038/s41598-017-16171-1>.
 24. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
 25. Magoc T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>.
 26. 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>.
 27. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
 28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
 29. Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680. <https://doi.org/10.1093/nar/22.22.4673>.
 30. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>.
 31. Ribeiro-Gonçalves B, Francisco AP, Vaz C, Ramirez M, Carriço JA. 2016. PHYLOVIZ Online: web-based tool for visualization, phylogenetic inference, analysis and sharing of minimum spanning trees. *Nucleic Acids Res* 44:W246–51. <https://doi.org/10.1093/nar/gkw359>.
 32. Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>.
 33. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054>.
 34. Devereaux LN, Polkinghorne A, Meijer A, Timms P. 2003. Molecular evidence for novel chlamydial infections in the koala (*Phascolarctos cinereus*). *Syst Appl Microbiol* 26:245–253. <https://doi.org/10.1078/07232020322346092>.
 35. Jelocnik M, Islam MM, Madden D, Jenkins C, Branley J, Carver S, Polkinghorne A. 2017. Development and evaluation of rapid novel isothermal amplification assays for important veterinary pathogens: *Chlamydia psittaci* and *Chlamydia pecorum*. *PeerJ* 5:e3799. <https://doi.org/10.7717/peerj.3799>.
 36. Wan C, Loader J, Hanger J, Beagley K, Timms P, Polkinghorne A. 2011. Using quantitative polymerase chain reaction to correlate *Chlamydia pecorum* infectious load with ocular, urinary and reproductive tract disease in the koala (*Phascolarctos cinereus*). *Aust Vet J* 89:409–412. <https://doi.org/10.1111/j.1751-0813.2011.00827.x>.
 37. Olagoke O, Miller D, Hemmatzadeh F, Stephenson T, Fabijan J, Hutt P, Finch S, Speight N, Timms P. 2018. Induction of neutralizing antibody response against koala retrovirus (KoRV) and reduction in viral load in koalas following vaccination with recombinant KoRV envelope protein. *NPJ Vaccines* 3:30. <https://doi.org/10.1038/s41541-018-0066-4>.
 38. Nicholas K. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. *EMBnet news* 4:14.
 39. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
 40. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34:525–527. <https://doi.org/10.1038/nbt.3519>.