Koala retrovirus genetic diversity and transmission dynamics within captive koala populations

Briony A. Joyce[a], Michaela D. J. Blyton[a], Stephen D. Johnston[b], Paul R. Young[a], and Keith J. Chappell[a,c,1]

[a]School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, QLD 4072, Australia; [b]School of Agriculture and Food Sciences, The University of Queensland, Gatton, QLD 4343, Australia; and [c]Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St. Lucia, QLD 4072, Australia

Edited by Maribeth V. Eiden, NIH, Bethesda, MD, and accepted by Editorial Board Member Stephen P. Goff May 18, 2021 (received for review November 19, 2020)

Koala populations are currently in rapid decline across Australia, with infectious diseases being a contributing cause. The koala retrovirus (KoRV) is a gammaretrovirus present in both captive and wild koala colonies that presents an additional challenge for koala conservation in addition to habitat loss, climate change, and other factors. Currently, nine different subtypes (A to I) have been identified; however, KoRV genetic diversity analyses have been limited. KoRV is thought to be exogenously transmitted between individuals, with KoRV-A also being endogenous and transmitted through the germline. The mechanisms of exogenous KoRV transmission are yet to be extensively investigated. Here, deep sequencing was employed on 109 captive koalas of known pedigree, housed in two institutions from Southeast Queensland, to provide a detailed analysis of KoRV transmission dynamics and genetic diversity. The final dataset included 421 unique KoRV sequences, along with the finding of an additional subtype (KoRV-K). Our analysis suggests that exogenous transmission of KoRV occurs primarily between dam and joey, with evidence provided for multiple subtypes, including nonendogenized KoRV-A. No evidence of sexual transmission was observed, with mating partners found to share a similar number of sequences as unrelated koala pairs. Importantly, both distinct captive colonies showed similar trends. These findings indicate that breeding strategies or antiretroviral treatment of females could be employed as effective management approaches in combating KoRV transmission.

koala | retrovirus | transmission | diversity | evolution

Retroviral sequences are an ancestral feature of all vertebrate genomes analyzed to date. By infiltrating and becoming a permanent fixture of the host genome, a process known as endogenization, these retroviral elements have helped shape vertebrate evolution. Most endogenization events occurred millions of years ago (1). An exception is the recently discovered koala retrovirus (KoRV), which appears to have endogenized less than 50,000 y ago (1, 2). KoRV is a gammaretrovirus, closely related to gibbon ape leukemia virus (GALV), present in both wild and captive Australian koala populations (3). It is found at 100% prevalence in the northern states of Queensland and New South Wales (3–10); however, a lower prevalence is observed in southern populations (in the states of Victoria and South Australia), with some studies suggesting certain populations are completely KoRV free (4, 11). KoRV has putatively been associated with the onset of neoplasia, including leukemia and lymphoma, blood and bone marrow disorders (myelodysplasia), and a wide range of opportunistic infections, including chlamydiosis, a bacterial pathogen causing koala morbidity, infertility, and mortality (8, 12–16). Uniquely, it is one of the only retroviruses that presents in both endogenous and exogenous forms (3, 17). Despite this, limited research has focused on the exogenous transmission of this virus, with KoRV transmission dynamics and/or mechanisms currently unconfirmed.

To date, there are nine identified KoRV subtypes (KoRV-A to I), each having a unique amino acid signature within the hypervariable receptor binding domain (RBD) of the envelope protein (3, 5, 17–21). KoRV-A was the first described subtype (3), which has since been detected in every KoRV-positive koala analyzed (4, 5, 17, 21, 22). It is the only subtype known to have endogenized and is incorporated at multiple sites in the germline DNA of northern koala populations, with consequent transmission through to progeny from both sire and dam (9). As a result, endogenous KoRV-A has been detected in all Queensland and New South Wales koalas analyzed to date at very high copy numbers (5, 6, 23). This endogenized variant is yet to be detected at high enough copy numbers in koala populations residing in southern parts of Australia to be considered endogenous, and evidence for the presence of replication competent virus in these southern populations is relatively weak (6, 9).

KoRV subtypes B to I have all been discovered within the last 9 years (5, 18, 19, 21). KoRV-B has been suggested to be more pathogenic compared to KoRV-A based on increased infectivity measured in vitro and some evidence supporting an association with disease (17, 23, 24). Whether the remaining KoRV subtypes are similarly associated with disease outcome remains to be extensively investigated. In comparison to KoRV-A, the prevalence of KoRV-B to I among koala populations is considerably reduced, with their identification in only a portion of the koalas analyzed to date (5–7, 11). Additionally, these subtypes are yet to be detected within the koala germline. Together, this suggests that KoRV subtypes B to I are exogenously acquired, meaning

Significance

A conserved retrovirus, koala retrovirus (KoRV)-A, is present within the genome of koalas in most Australian populations. Additional divergent sequences and subtypes are thought to be exogenously transmissible and more pathogenic. We present a comprehensive analysis of KoRV genetic diversity within two captive koala populations and statistically significant evidence of exogenous transmission occurring primarily through maternal transmission. These findings suggest strategies which may help limit the spread of pathogenic KoRV subtypes. We raise the possibility that captive breeding programs could target use of antiretroviral drugs to dams during the breeding season. While substantial research is needed to demonstrate safety and effectiveness, this presents a potentially important conservation strategy for koala populations severely affected by disease and currently in rapid decline.

Author contributions: M.D.J.B. and K.J.C. designed research; B.A.J. and M.D.J.B. performed research; B.A.J., M.D.J.B., and K.J.C. analyzed data; and B.A.J., M.D.J.B., S.D.J., P.R.Y., and K.J.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. M.V.E. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: k.chappell@uq.edu.au.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2024021118/-/DCSupplemental.

Published September 7, 2021.

they are not genetically inherited by progeny but instead are derived or transmitted via alternate routes (17, 19, 21, 23).

Vertebrates are able to acquire exogenous retroviral sequences in multiple ways. Some exogenous variants, such as those of feline leukemia virus (FeLV), are derived within the host through de novo recombination and are not directly transmissible between individuals (25–27). Alternatively, exogenous sequences can be actively transmitted between infected and uninfected individuals through contact with infectious fluids. Possible routes for exogenous retroviral transmission include sexual (28) and within milk (29, 30), feces (31), and saliva (32). For koalas in particular, there is also the potential for transmission to occur through ingestion of pap, a special form of maternal feces consumed by the joey at pouch emergence (33). A number of studies have highlighted an apparent correlation between the KoRV-B status of koala dams and joeys (17, 23, 34); however, due to the limited sample sizes and the absence of matched samples from fathers, such findings have been purely observational, and statistical significance could not be ascribed. Similarly, contact transmission between adult animals has been assumed based on transition to KoRV-B positive status over repeated sampling of individuals; however, again, these results are based on a limited sample size, and the possibility for an increase in KoRV load over time to confound qPCR analysis cannot be excluded (23).

In this study, we sought to investigate KoRV transmission dynamics in captive koala colonies of known pedigree (n = 109). Patterns of KoRV diversity were highlighted within two Southeast Queensland captive populations, providing insight into the complexities of KoRV genetics and evolution. Furthermore, sequence sharing between dam-joey, sire-joey, and mating koala pairs were directly compared to infer the transmission of various KoRV subtypes. This research not only informs future conservation efforts in the management of both captive and wild populations but also provides greater insight into the mechanisms and role of exogenous transmission in shaping a recently endogenized retrovirus.

Results

KoRV Genetic Diversity within the RBD. Deep sequencing was carried out on an ~500 nucleotide (nt) region of the KoRV envelope (*env*) gene, encompassing the hypervariable RBD, for 109 captive koalas housed in two separate Southeast Queensland colonies (colony A, n = 45; colony B, n = 64). A representative library of integrated KoRV proviral sequences was produced from genomic DNA (gDNA) isolated from peripheral blood mononuclear cells (PBMCs). We chose to utilize gDNA from PBMCs, as this was shown to provide a good representation of viral RNA (vRNA), with around 80% of unique KoRV sequences detected in gDNA also present in vRNA (*SI Appendix*, Fig. S3). While there was only a weak correlation between relative copy numbers determined between datasets (*SI Appendix*, Fig. S3; r_s = 0.281; 95% credible interval [CI], −0.057 to 0.561), this is presumably due to vRNA being affected by transcriptional activity, transcriptional errors, and recombination. We therefore hypothesized that sequencing of proviral integrations may provide a cleaner library for assessment of transmission.

The read count was found to vary considerably between individuals with 21,625 KoRV reads present per koala on average, ranging from 1,247 to 203,804 after filtering and quality control. Clustering at a 97% identity threshold resulted in the identification of 421 unique KoRV sequences, following the exclusion of 391 nonfunctional sequences (deposited to GenBank). Protein alignment of these in silico translated sequences to those of known KoRV subtypes led to their classification as one of eight subtypes (A to D, F to I). The known KoRV subtypes have an average RBD nt sequence similarity of 60.4%, ranging between 40.6% and 76.1%. One divergent group of sequences was found to have a unique amino acid signature within the RBD (*SI Appendix*, Fig. S1), with nt sequences less than 75.7% similar to those of the known subtypes. Hence, following convention (5, 19, 21), this group was classified as a novel subtype (KoRV-K). KoRV-E was the only previously identified subtype not detected in this study.

The highest number of unique sequences were discovered for KoRV-A, B, and D, which were also found to be the most prevalent subtypes detected (Table 1). As expected, KoRV-A was identified in all analyzed koalas, making it the most prevalent subtype in these captive populations. This was followed closely by KoRV-D, which was detected in 104 out of 109 individuals. Having only been identified within one koala, KoRV-C was found to be the least prevalent subtype, followed closely by KoRV-F and G, which were detected in fewer than 10 individuals. KoRV-G was found to have the least genetic diversity, with only two sequences detected from all KoRV-G positive koalas (n = 10), both of which were present in the same koala 90% of the time.

The majority of the identified sequences were detected in multiple individuals; however, 91 (21.6%) sequences were only identified within a single koala (Table 1). The endogenous KoRV-A sequence (GenBank accession number AF151794) was detected in all koalas as the most abundant sequence, where, on average, it accounted for 93% of an individual's reads, ranging between 82% and 100%. In comparison, the next most abundant KoRV-A sequence, detected in 66 koalas, only accounted for 0.5% of the reads on average, ranging between 0.004% and 4.08%. The most frequently detected non-KoRV-A sequence was a KoRV-B sequence identified in 79 koalas. Despite the high prevalence, this sequence only accounted for, on average, 0.6% of the total KoRV reads, ranging between 0.003% and 3.56%, and 27.83% of total KoRV-B reads, ranging between 0.18% and 63.86%. The other subtypes were also found to have dominant sequences that were identified in 85.37%, 75.47%, and 62.86% of KoRV-H, I, and K positive koalas, respectively, each accounting for 41% and 69% of reads for their respective subtypes, on average. Despite the small sample sizes, koalas positive for KoRV-F and G were all found to possess the same dominant sequence that, on average, accounted for 89.54% and 74.69% of KoRV-F and G reads, respectively. No dominant sequence was detected for KoRV-D.

Subtype Variation between Animals and Captive Populations. KoRV-A was the most abundant subtype in all analyzed koalas, accounting for 94% of an individual's reads on average, ranging from 84% to 100%. In comparison, the other subtypes were much less abundant, individually accounting for less than 9.9% of a koala's total reads (Fig. 1). The percentage of reads attributed to each subtype for each koala is shown in Fig. 1 and *SI Appendix*, Table S1. The majority of subtypes were found to have a similar

Table 1. KoRV subtype diversity in captive koala populations

	No. individuals (%)			No. unique sequences	
Subtype	Colony A	Colony B	Combined	Identified	Shared (%)
KoRV-A	45 (100)	64 (100)	109 (100)	92	80 (87)
KoRV-B	39 (86.7)	53 (76.6)	92 (84.4)	79	62 (78.5)
KoRV-C	1 (2.2)	0 (0)	1 (0.9)	4	0 (0)
KoRV-D	43 (95.6)	61 (95.3)	104 (95.4)	104	81 (77.9)
KoRV-E	N.D	N.D	N.D	N.D	N.D
KoRV-F	5 (11.1)	3 (4.7)	8 (7.3)	5	2 (40)
KoRV-G	3 (6.7)	7 (10.9)	10 (9.2)	2	2 (100)
KoRV-H	18 (40)	23 (35.9)	41 (37.6)	25	21 (84)
KoRV-I	30 (66.7)	23 (35.9)	53 (48.6)	58	42 (72.4)
KoRV-K	5 (11.1)	30 (46.9)	35 (32.1)	52	40 (76.9)
Total:	45	64	109	421	330 (78.4)

Number of newly identified sequences for each subtype and number of koalas possessing that subtype are shown for both populations. Subtypes were identified in the genomic DNA of koalas from both colony A (n = 45) and B (n = 64). Shared sequences correspond to those present in two or more koalas. N.D, not detected.

prevalence in both populations (Table 1), with the proportion of reads attributed to each subtype also remaining consistent (Fig. 1). However, newly identified KoRV-K was significantly more prevalent within colony B, being detected in 46.9% of individuals compared to 11.1% in colony A (Table 1; $U = 925$; $P = 0.0001$), and also at a significantly greater abundance, accounting for 1.95% (0.002% to 5.582%) of reads on average compared to 0.08% (0.002% to 0.365%) in colony A (Fig. 1*B*; $U = 17$; $P = 0.004$). A similar observation was made for KoRV-I, which was found to be almost twice as prevalent in colony A compared to colony B (Table 1; $U = 997.5$; $P = 0.002$), although the average read abundance for this subtype was consistent across both populations (1.1% to 1.4%; Fig. 1*B*; $U = 285$; $P = 0.288$).

Out of the total 109 koalas analyzed, 96.3% had detection of two or more KoRV subtypes, with some individuals having up to eight (A26, B60, and B64). Interestingly, four koalas were positive for only KoRV-A. One of these koalas was from colony A (A32, male), and three were housed in colony B (B10, female; B51, male and B52, male) (Fig. 1*A*). Koala B10 and B51 are both siblings initially born in Perth from parents with colony A and New South Wales heritage. Koala A32 was also born in Perth to the same sire as B10 and B51 but to a different dam, who is the half-sister of B10 and B51. Conversely, koala B52 was wild bred, with no known relation to the other three koalas. Three of these KoRV-A positive only koalas were resampled at ~18 mo post initial sampling, which revealed none had acquired any additional subtypes within this period. Furthermore, there was no detection of any KoRV-A sequences that were not the originally described endogenous sequence within these koalas at either sampling point. Koala B52 was unable to be resampled as he died from declining welfare during this time.

Subtype Transmission. The sharing of identical KoRV sequences was analyzed between unrelated ($n = 5{,}687$), dam-joey ($n = 50$), sire-joey ($n = 29$), maternally related (m-related, $n = 49$), paternally related (p-related, $n = 49$), and mating partner ($n = 22$) koala pairs under null hypothesis selection by fitting generalized linear mixed models using the MCMCglmm 2.29 package in R (35). The models were fitted in a Bayesian framework using a Poisson model with the number of sequences shared between each koala pair fitted as the response variable and pair classification as the fixed explanatory variable. A maternal lineage was defined by those related through a strictly female line. This analysis was carried out separately for each subtype excluding KoRV-C, F, and G, which were detected in too few koalas to be statistically tested. Due to the overall low sharing observed

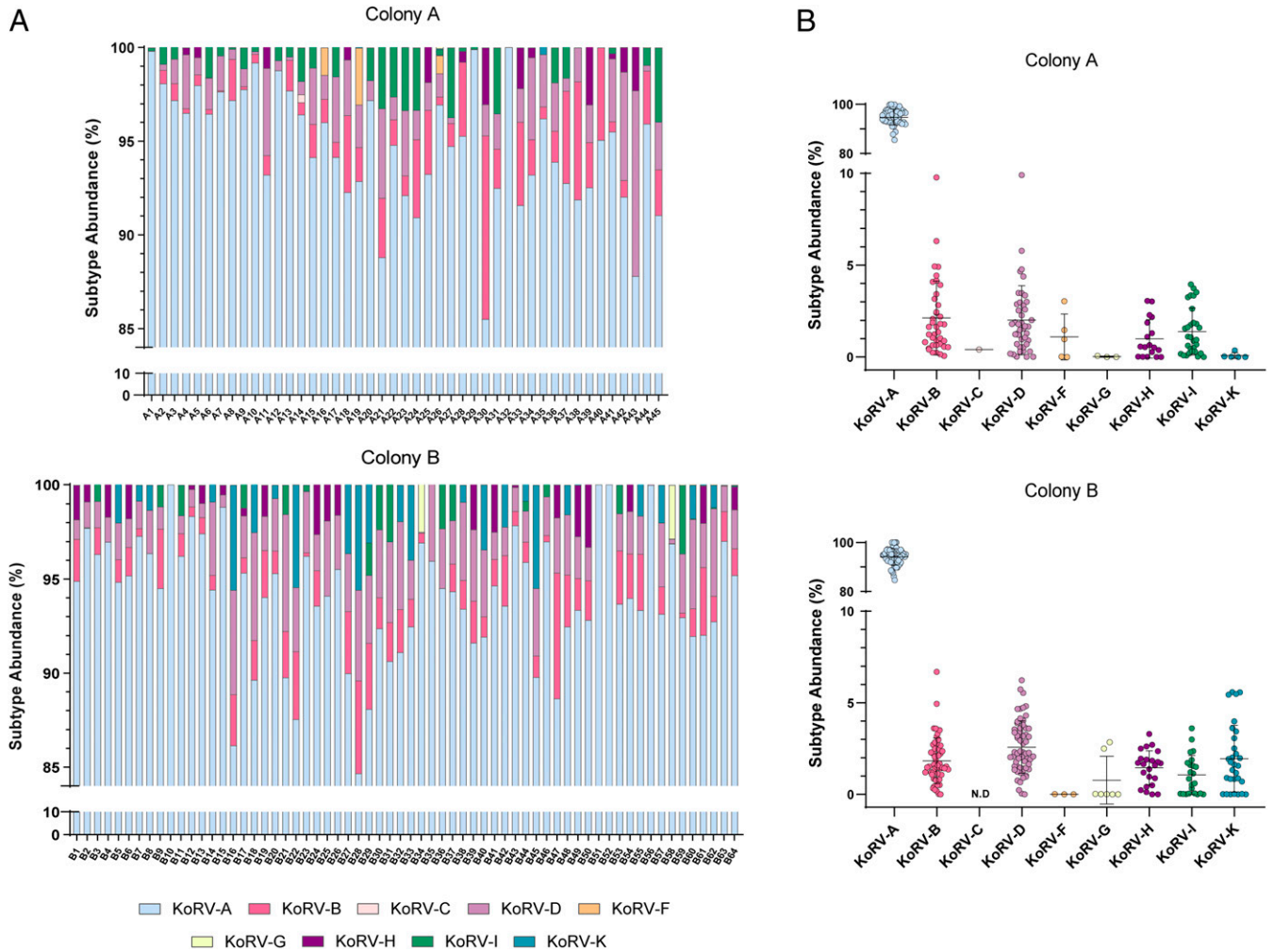

Fig. 1. Percentage of KoRV reads grouped by subtypes. Prevalence of KoRV subtypes in genomic DNA from koalas housed at colony A ($n = 45$) and colony B ($n = 64$). (*A*) Subtype abundance for each animal is shown for both populations. Colors indicate the different subtypes detected. Of note, koala B56 had low detection of KoRV-D (0.006%, *SI Appendix*, Table S1). (*B*) Percentage relative abundance for each subtype is summarized for both populations. Each point represents an individual koala with the mean ± SD shown. N.D, not detected.

between koala pairs for KoRV-H, I, and K, these subtypes were combined for this analysis to allow for a more accurate fit of the model. The original, endogenous KoRV-A sequence was also removed as it was shared among all koalas.

Overall, dam-joey koala pairs were found to share significantly more KoRV sequences compared to the other analyzed pair types. Taking all subtypes into account, dam-joey pairs shared 9.3 sequences (95% CI, 7.9 to 11.3) on average, which was threefold more than the number shared by unrelated pairs (averaged 3.0 sequences shared; 95% CI, 2.4 to 3.7; Fig. 2; $P < 0.0001$) and sire-joey pairs (averaged 3.2 sequences shared: 95% CI, 2.6 to 4.0; Figs. 2 and 3; $P < 0.001$). This finding reached statistical significance for each individual subtype with the exception of KoRV-B, where this was just above the defined cutoff of 0.05 (averaged 1.6 dam-joey and 1.1 sire-joey sequences shared; $P = 0.07$). Similar findings were evident when comparing m-related pairs (averaged 15.9 sequences shared; 95% CI, 12.8 to 21.0) with unrelated pairs ($P < 0.0001$) and p-related pairs (averaged 3.9 sequences shared; 95% CI, 3.5 to 4.6; Figs. 2 and 3; $P < 0.0001$). Again, this trend was observed for KoRV-B, though not significant (averaged 1.8 m-related and 1.4 p-related sequences shared; $P = 0.149$). P-related koala pairs did show some increase in sequence sharing relative to unrelated pairs, but only for subtypes A and B and to a much lower extent (Fig. 2; $P < 0.05$).

Mating partners were found to be no more likely to share sequences than unrelated pairs for any individual subtype. On average, mating partner pairs shared 2.5 sequences (95% CI, 1.9 to 3.6) compared to 3.0 for unrelated pairs (95% CI, 2.4 to 3.7; $P > 0.32$).

Discussion

Despite KoRV having a high prevalence and potential impact on koala health, limited research has been conducted on the transmission of this virus and its subtypes. Most studies to date have only concentrated on the endogenous KoRV-A (1, 3, 9, 23) and the exogenous KoRV-B (17, 23, 34, 36). While these small-scale analyses have highlighted instances of maternal and adult to adult transmission of exogenous KoRV, these studies have been conducted using partial koala pedigrees with low sample numbers and with no statistical comparisons to other pair types to definitively ascribe transmission routes (17, 23, 34). To expand

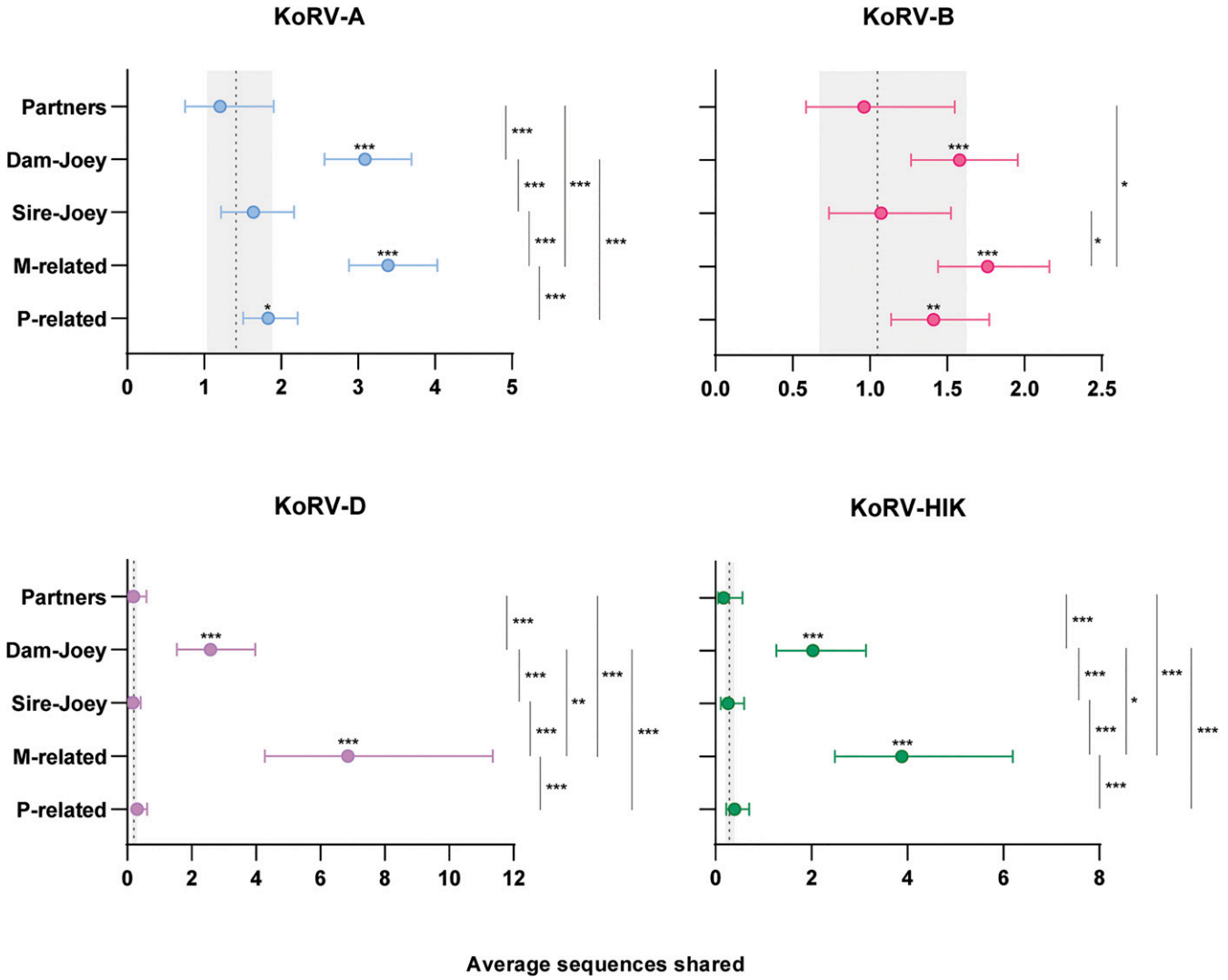

Fig. 2. Average sequences shared with 95% CIs from generalized linear mixed models of subtype sharing. The expected number of shared sequences for each subtype is shown for p-related ($n = 49$), m-related ($n = 49$), sire-joey ($n = 29$), dam-joey ($n = 50$), and mating partner ($n = 22$) koala pairs. Average sharing between unrelated ($n = 5,687$) koalas is represented by a dashed line with 95% CIs highlighted in gray. Maternal relatives were defined by those related through a strictly female lineage. The original, endogenous KoRV-A sequence was omitted from this analysis. Asterisks above nodes indicate significance to the unrelated reference group. Significance between different pair groupings are shown on the right. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

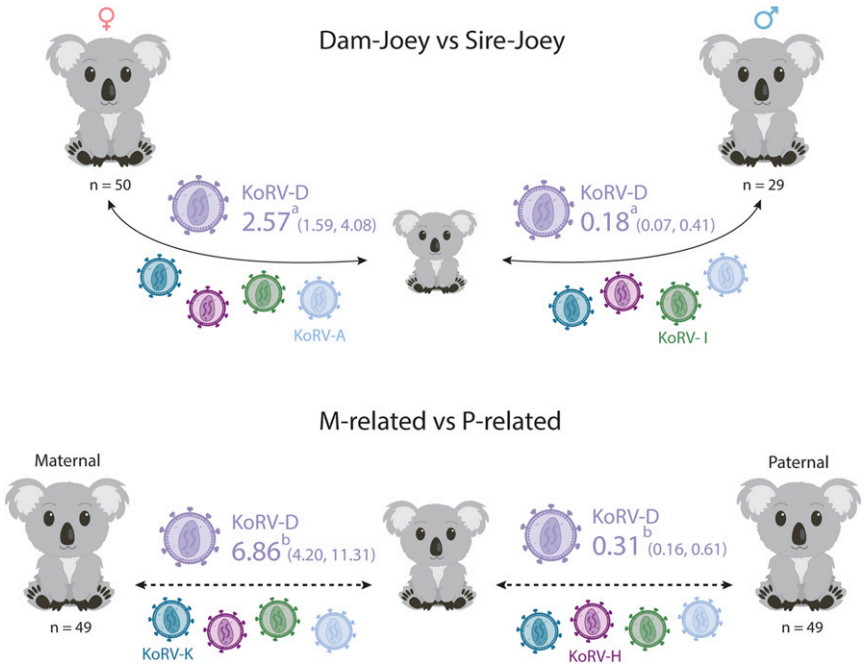

Fig. 3. Key sequence sharing dynamics. Average number of sequences shared between dam-joey, sire-joey, and extended m-related and p-related pairs is shown for KoRV-D. Similar trends were observed for KoRV-A (blue), H (dark purple), I (green), and K (cyan). Direct and extended relations are represented by full and dashed lines, respectively. Maternal relation was defined by those related through a strictly female lineage. Analysis includes koalas housed in both colony A and B. Expected mean is shown with lower and upper CIs displayed in brackets. *n* denotes the number of koala pairs. $^{a,b}P < 0.001$.

upon these limited studies, we analyzed KoRV genetic diversity in 109 captive koalas housed in two distinct populations in Southeast Queensland. Using known koala pedigrees, sequences shared between koalas of varying relation were compared to infer subtype-specific transmission. This study represents a large-scale transmission study conducted on KoRV and a detailed KoRV genetic diversity analysis conducted on captive Australian koalas.

Both captive colonies included in this study were found to contain distinct KoRV subtype diversity. A total of 421 unique KoRV sequences were identified across both captive koala populations based on a clustering identity of 97%. Notably, while a 99% cut off is usually observed within KoRV diversity studies (5, 6), this threshold was chosen based on the frequency of PCR error and sequencing artifacts. In accordance with other studies (5–7), KoRV-A, B, and D were found to be the most prevalent subtypes in both populations. Detection of KoRV-C in one animal presents the first case of a KoRV-C positive koala within Australia, with this subtype only previously detected in koalas held in overseas zoos (20, 21). Of note, this animal was born to wild-bred parents. KoRV-H was found to be more prevalent within captive koalas in Southeast Queensland in comparison to wild koalas from the same region, where it was only detected in 1 animal out of 18 (5). By contrast, these 18 wild koalas had a greater prevalence of KoRV-F in comparison to the captive populations, where it was only detected at minor levels. These patterns of specific subtype prevalence within localized breeding populations reflect localized transmission dynamics. In addition to the detection of known subtypes, the animals from this study also possessed sequences of a potentially novel subtype, KoRV-K, which was found to be much more prevalent and abundant within colony B compared to colony A. Previous studies have shown KoRV subtypes C to I to fall within the paraphyletic subgroup D (5, 6); notably, this subtype was also found to cluster within this group (*SI Appendix*, Fig. S2). KoRV-E failed to be detected within any of the koalas included in this study, making it the only subtype yet to be identified within Australian populations.

While almost one quarter of the detected KoRV sequences were only identified within a single koala, highlighting the continual within-host evolution of this virus, the observation that the majority of sequences for all subtypes were present in multiple animals indicates that active transmission of the different subtypes is ongoing. Captive populations around Australia often exchange their koalas to increase genetic diversity within populations, thereby instigating possible KoRV transmission between institutions. Transmission between wild and captive koalas is also possible, with captive institutions commonly incorporating rehabilitated wild koalas into their breeding programs. These could both be contributing factors to the similar KoRV genetic profiles (subtype prevalence and abundance) we observed between the captive populations in this study, particularly due to them both residing in Southeast Queensland.

The data generated through this study provide evidence to suggest that all subtypes tested, including KoRV-A, B, D, H, I, and K, transmit exogenously, primarily through dam-joey interactions. This complements the results from small-scale pedigree studies, which found the primary route of transmission for KoRV-B to occur between dam and joey, although no statistical analyses or comparisons to other pair types were conducted (17, 23, 34). Interestingly, both KoRV-A and B sequences did display some evidence of sire-joey transmission; however, this was to a much lower extent than dam-joey transmission. As cohousing of sires and offspring is not standard practice in zoos, it is an unlikely explanation for this transmission. Additionally, if this were the case, we would expect to see more sequence sharing among mating partner and unrelated koala pairs. Homologous recombination is known to occur at high frequencies for all analyzed retroviruses, including HIV, which undergoes approximately two to three recombination events per genome per replication cycle (37). KoRV-B has never been identified as an endogenous provirus, and while we cannot rule out this possibility as an explanation for the sire-joey transmission, a more likely explanation is the sire-joey transmission of endogenous KoRV-A sequences followed by their recombination in progeny with the exogenous KoRV-B.

Transmission is likely to occur more readily between dam and offspring due to their close proximity and sharing of potentially

infectious fluids, including milk and pap. Notably, while no active virus has been recovered from milk, KoRV sequences and peptides have previously been discovered in both early and late lactation koala milk (38). These modes of exogenous transmission have been observed with other related gammaretroviruses, including GALV and FeLV, where transmission is known to occur prenatally and postnatally through contact of infectious secretions, including milk and feces (31, 39–41). We therefore propose that transmission of exogenous KoRV subtypes A, B, D, H, I, and K most likely occurs through ingestion of milk, pap, and/or infected fluids during the perinatal period or parturition. Future investigations by our group will involve identifying the possible modes of this exogenous transmission between koalas.

Our findings suggest that KoRV transmission among adult koalas is infrequent, with average sequence sharing among unrelated pairs less than one for KoRV-D and close to one for KoRV-B despite the high prevalence of these subtypes in the populations. Additionally, four koalas were identified as positive for KoRV-A only within this study and did not acquire any other subtypes or sequences within a ∼18-mo period. These koalas were housed in both populations and were recently transferred from Perth Zoo ($n = 3$) or brought in for care from the wild ($n = 1$) less than 3 years prior to initial sampling. The greatest KoRV diversity and viral loads recorded to date have been observed in Southeast Queensland (4, 6, 12, 42), where the sample populations are located. As such, the KoRV-A only koalas may not have experienced the same level of KoRV exposure while joeys as compared to the other captive koalas sampled. Together, this suggests that a more intimate interaction, such as that between dam and joey, may be required for exogenous transmission to occur.

It is well established within the literature that KoRV-A has endogenized within Queensland populations, with transmission occurring vertically through the koala germline (1, 3, 9). This was supported by our results, which found all koalas to possess the same dominant endogenized KoRV-A sequence. However, our analysis also revealed a higher level of KoRV-A sharing among m-related koalas compared to all other groups, suggesting ongoing exogenous transmission. Vast sequence diversity was also observed within this subtype, with 92 sequences detected, 87% of which were identified in more than one koala. Exogenous transmission of KoRV-A has generally been assumed in southern koala populations where functional KoRV-A does not appear to be endogenized (4, 9); however, our analysis indicates active viral transmission of this subtype is also still occurring where endogenous KoRV-A is ubiquitous.

This study provides evidence that sexual transmission of KoRV is not extensive. While sexual contact is known to be a primary route of transmission for other known retroviruses, such as HIV (43), this has yet to be conclusively documented for any gammaretrovirus. The possibility of sexual contact during mating being a cause of KoRV transmission has been previously proposed (23); however, our analysis found no evidence of this, with no significant difference in the extent of sharing among mating partner and unrelated pairs for any of the subtypes. Of the mating partner pairs analyzed, there were three in which only the male was KoRV-B positive and five in which only the female was KoRV-B positive (*SI Appendix*, Table S2), supporting the suggestion that exogenous KoRV is not readily sexually transmitted.

While the association of KoRV with disease and cancer onset is becoming increasingly apparent (34, 44), research has already begun focusing on suitable KoRV prevention and management strategies. Various groups have proposed vaccination as a means of protection and conservation for koalas through inducing neutralizing antibodies and reducing viral loads (45–48). However, the results from our study suggest vaccination will be ineffective, as such strategies would be unlikely to hinder transmission between KoRV positive dams and their joeys at a very early age. Alternative methods targeted toward breeding strategies or perhaps antiretroviral treatment of mothers during the breeding season may be more effective and deserve further investigation.

In summary, this study analyzed KoRV genetic diversity and transmission dynamics within two healthy captive koala populations residing in Southeast Queensland. The number of uniquely identified sequences and detection of a subtype significantly expands known KoRV diversity and provides insight into viral evolution. Key transmission dynamics were highlighted, revealing close interactions, as seen between dam and joey, to be the primary mode of exogenous transmission within these populations. Possible routes for this transmission include contact with infected fluids, including milk and pap. These findings will be important for future koala conservation and captive management by inferring suitable breeding and management strategies.

Materials and Methods

Sample Collection and Processing. A total of 109 (53 female, 56 male) clinically healthy captive koalas were included in this study, sampled from two different populations in Southeast Queensland, Australia: colony A ($n = 45$) and colony B ($n = 64$). From each conscious koala, ∼2 mL of blood was drawn by a trained veterinarian during annual veterinary examinations and stored on ice during transport to the laboratory. Blood samples were centrifuged for 2 min at 11,200 × *g* to separate components, after which the blood plasma (∼500 μL), stored in 500 μL of RNA*later* stabilization solution (Invitrogen), and buffy coat (∼100 μL) fractions were separated. Proviral genomic DNA was then extracted from the buffy coat using the FavorPrep blood genomic DNA mini kit (Favorgen Biotech Corp), as per the manufacturer's instructions.

Illumina Sequencing. An amplicon library was prepared for all samples following the Illumina 16S Metagenomic Sequencing Library Preparation guide (15044223-B). Previously published oligonucleotide primers that flank the hypervariable region of the *env* gene (5) were used to amplify the ∼500 bp target sequence via PCR. These primers included both *env* complementary regions and Illumina adapter sequences. Phusion high-fidelity DNA polymerase (NEB) was used per the manufacturer's instructions with an annealing temperature of 55 °C and 25 rounds of amplification. Sequencing of the PCR amplicons was carried out at the Australian Centre for Ecogenomics (University of Queensland). Amplicons were purified using Agencourt AMPure XP beads (Beckman Coulter) and 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. Indexed amplicons were pooled in equimolar concentrations and sequenced on a MiSeq sequencing system (Illumina) using paired-end sequencing with V3 300 bp chemistry as per the manufacturer's protocol.

Bioinformatic Processing. Forward and reverse reads were merged based on roughly 20 nt of overlap, with up to 5% mismatch, using the Galaxy web platform on the public server at https://usegalaxy.org (49). Reads were then filtered by size (450 to 600 bp) and quality (90% of sequence with a cutoff value >20). The de novo operational taxonomic unit picking method in QIIME 2 (50) was used to cluster reads with a similarity of 97%. Representative sequences of each QIIME cluster were then blasted against the NCBI "nt" database to identify non-KoRV sequences. Those only containing a single read were also omitted. The putative KoRV sequences were then aligned to the *env* nt sequence of KoRV-A (GenBank accession number AF151794) to confirm sequence homology using CLC Workbench 8 (CLCBio). Sequences containing missense mutations or large deletions or those that lacked *env* gene homology to KoRV-A were excluded from subsequent analysis. Illumina adapter sequences were trimmed from sequence termini to nt 23 to 513 (KoRV-A numbering). To determine subtype, a protein alignment was carried out on the in silico translated sequences along with representative sequences for each subtype, with subtypes denoted based on homology within the hypervariable region, in accordance with previous studies (5, 17, 20, 21, 51). Sequences found to have less than 80% homology within this region to any known subtype were deemed novel subtypes. The validated dataset included 421 unique *env* sequences. Subtype prevalence and read abundance between colonies was compared by a Mann–Whitney *U* test using GraphPad Prism 8.0.1 software.

Sequence Sharing Analysis. The sample by sequence read count table generated from QIIME was converted into a sample by sequence presence/absence matrix. The number of identical sequences shared between koala pairs for each subtype was then calculated using custom code in RStudio 3.5.1 (52). Using supplied koala pedigrees from both captive populations, each koala pair was classified as unrelated ($n = 5{,}687$), dam-joey ($n = 50$), sire-joey ($n = 29$), mating

partners ($n = 22$), m-related (up to second cousins, $n = 49$), or p-related (up to second cousins, $n = 49$). Maternal relatives were defined by those related through a strictly female lineage. Subtypes H, I, and K were combined for this analysis as the extent of sharing for these subtypes among unrelated koalas was insufficient to allow model fitting. Due to the low incidence of KoRV-C, F, and G, these subtypes were excluded.

Generalized linear mixed models following a Poisson distribution were then fitted using the MCMCglmm 2.29 package in R (35) to determine if the extent of KoRV sequence sharing differed between koala pairs with different familial relationships. The models were fitted in a Bayesian framework with uninformative priors using a Poisson model that incorporated additive over dispersion. The number of sequences shared between each koala pair was fitted as the response variable with the fixed explanatory variable being the pair classification. Koala identifications were fitted as a random effect using a multiple membership model to account for the nonindependence between the pairwise comparisons. Specified model parameters were as follows: number of iterations (nitt) = 1,003,000 (KoRV-A and B), 1,503,000 (KoRV-D), or 2,003,000 (KoRV-HIK); number of initial iterations removed (burnin) = 3,000; and thinning interval (thin) = 100 (KoRV-A and B), 150 (KoRV-D), 200 (KoRV-HIK). The endogenous KoRV-A sequence (GenBank accession number AF151794) was omitted from this analysis as it was shared among all koalas.

Data Analysis. The expected number of sequences shared for each pair type within each subtype was calculated using the following formula: $e^{\text{post mean + post mean of intercept}}$, with CIs also determined using $e^{\text{credible interval value + post mean of intercept}}$. The expected number of sequences shared for each pair type across multiple subtypes was calculated by summing the post means for each subtype. CIs were then determined by taking the square root of both the upper and lower variance ($\text{CI}_{\text{upper or lower}} = \sqrt{a^2 + b^2 + d^2 + hik^2}$, where a, b, d, and hik are the variance of KoRV-A, B, D, and HIK, respectively).

Ethical Statement. All sampling procedures were approved by the University of Queensland Animal Ethics Committee (animal ethics number SCMB/094/18/DREAMWORLD).

Data Availability. All sequences reported in this paper have been deposited in GenBank and assigned accession numbers MW283966–MW284386.

ACKNOWLEDGMENTS. We thank the veterinarians and veterinary nurses from both captive koala housing institutions for access to their animals and for their assistance with sample collection. Further acknowledgment goes to the Australian Centre for Ecogenomics at the University of Queensland for the supply of unpublished Illumina sequencing data.

1. Y. Ishida, K. Zhao, A. D. Greenwood, A. L. Roca, Proliferation of endogenous retroviruses in the early stages of a host germ line invasion. *Mol. Biol. Evol.* **32**, 109–120 (2015).
2. M. C. Ávila-Arcos *et al*., One hundred twenty years of koala retrovirus evolution determined from museum skins. *Mol. Biol. Evol.* **30**, 299–304 (2013).
3. J. J. Hanger, L. D. Bromham, J. J. McKee, T. M. O'Brien, W. F. Robinson, 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 (2000).
4. G. S. Simmons *et al*., Prevalence of koala retrovirus in geographically diverse populations in Australia. *Aust. Vet. J.* **90**, 404–409 (2012).
5. K. J. Chappell *et al*., Phylogenetic diversity of koala retrovirus within a wild koala population. *J. Virol.* **91**, e01820-16 (2017).
6. N. Sarker *et al*., Genetic diversity of koala retrovirus *env* gene subtypes: Insights into northern and southern koala populations. *J. Gen. Virol.* **100**, 1328–1339 (2019).
7. B. L. Quigley *et al*., Changes in endogenous and exogenous koala retrovirus (KoRV) subtype expression over time reflects koala health outcomes. *J. Virol.* **93**, 849 (2019).
8. R. Tarlinton, J. Meers, J. Hanger, P. Young, 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 (2005).
9. R. E. Tarlinton, J. Meers, P. R. Young, Retroviral invasion of the koala genome. *Nature* **442**, 79–81 (2006).
10. R. E. Tarlinton *et al*., Differential and defective expression of koala retrovirus reveal complexity of host and virus evolution. *bioRxiv* [Preprint] (2017). https://doi.org/10.1101/211466 (Accessed 17 August 2017).
11. A. R. Legione *et al*., 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 (2017).
12. J. Fabijan *et al*., Pathological findings in koala retrovirus-positive koalas (*Phascolarctos cinereus*) from northern and southern Australia. *J. Comp. Pathol.* **176**, 50–66 (2020).
13. A. S. Brown, A. A. Girjes, M. F. Lavin, P. Timms, J. B. Woolcock, Chlamydial disease in koalas. *Aust. Vet. J.* **64**, 346–350 (1987).
14. F. A. Cockram, A. R. Jackson, Keratoconjunctivitis of the koala, *Phascolarctos cinereus*, caused by *Chlamydia psittaci*. *J. Wildl. Dis.* **17**, 497–504 (1981).
15. F. A. Cockram, A. R. Jackson, Letter: Isolation of a Chlamydia from cases of keratoconjunctivitis in koalas. *Aust. Vet. J.* **50**, 82–83 (1974).
16. K. A. McColl, R. W. Martin, L. J. Gleeson, K. A. Handasyde, A. K. Lee, Chlamydia infection and infertility in the female koala (*Phascolarctos cinereus*). *Vet. Rec.* **115**, 655 (1984).
17. W. Xu *et al*., An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11547–11552 (2013).
18. T. Miyazawa, T. Shojima, R. Yoshikawa, T. Ohata, Isolation of koala retroviruses from koalas in Japan. *J. Vet. Med. Sci.* **73**, 65–70 (2011).
19. W. Xu, K. Gorman, J. C. Santiago, K. Kluska, M. V. Eiden, Genetic diversity of koala retroviral envelopes. *Viruses* **7**, 1258–1270 (2015).
20. K. C. Abts, J. A. Ivy, J. A. DeWoody, Immunomics of the koala (*Phascolarctos cinereus*). *Immunogenetics* **67**, 305–321 (2015).
21. T. Shojima *et al*., Identification of a novel subgroup of Koala retrovirus from Koalas in Japanese zoos. *J. Virol.* **87**, 9943–9948 (2013).
22. U. Fiebig, M. Keller, A. Möller, P. Timms, J. Denner, Lack of antiviral antibody response in koalas infected with koala retroviruses (KoRV). *Virus Res.* **198**, 30–34 (2015).
23. B. L. Quigley, V. A. Ong, J. Hanger, P. Timms, Molecular dynamics and mode of transmission of koala retrovirus as it invades and spreads through a wild Queensland koala population. *J. Virol.* **92**, 13 (2018).
24. C. A. Waugh *et al*., 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 (2017).
25. O. Jarrett, W. D. Hardy Jr, M. C. Golder, D. Hay, The frequency of occurrence of feline leukaemia virus subgroups in cats. *Int. J. Cancer* **21**, 334–337 (1978).
26. S. P. Dunham, E. Graham, Retroviral infections of small animals. *Vet. Clin. North Am. Small Anim. Pract.* **38**, 879–901, ix (2008).
27. J. C. Neil, R. Fulton, M. Rigby, M. Stewart, "Feline leukaemia virus: Generation of pathogenic and oncogenic variants" in *Retroviral Insertion and Oncogene Activation*, H.-J. Kung, P. K. Vogt, Eds. (Springer-Verlag Berlin Heidelberg, Berlin, Germany, 1991), pp. 67–93.
28. Centers for Disease Control, *Recommendations for Prevention of HIV Transmission in Health-care Settings* (US Department of Health and Human Services, Public Health Service, Centers for Disease Control, Atlanta, GA, 1987), vol. 37.
29. A. M. Pacitti, O. Jarrett, D. Hay, Transmission of feline leukaemia virus in the milk of a non-viraemic cat. *Vet. Rec.* **118**, 381–384 (1986).
30. S. Hino *et al*., Mother-to-child transmission of human T-cell leukemia virus type-I. *Jpn. J. Cancer Res.* **76**, 474–480 (1985).
31. M. A. Gomes-Keller *et al*., Fecal shedding of infectious feline leukemia virus and its nucleic acids: A transmission potential. *Vet. Microbiol.* **134**, 208–217 (2009).
32. M. A. Gomes-Keller *et al*., Detection of feline leukemia virus RNA in saliva from naturally infected cats and correlation of PCR results with those of current diagnostic methods. *J. Clin. Microbiol.* **44**, 916–922 (2006).
33. B. Kalman, H. Levigne, *The Life Cycle of a Koala* (Crabtree Publishing Company, New York, 2002).
34. H. Zheng *et al*., Koala retrovirus diversity, transmissibility, and disease associations. *Retrovirology* **17**, 34 (2020).
35. J. D. Hadfield, MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *J. Stat. Softw.* **33**, 1–22 (2010).
36. M. A. Hashem *et al*., Transmission of koala retrovirus from parent koalas to a joey in a Japanese zoo. *J. Virol.* **94**, e00019-20 (2020).
37. A. E. Jetzt *et al*., High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol.* **74**, 1234–1240 (2000).
38. K. M. Morris *et al*., Characterisation of the immune compounds in koala milk using a combined transcriptomic and proteomic approach. *Sci. Rep.* **6**, 35011 (2016).
39. T. G. Kawakami, L. Sun, T. S. McDowell, Natural transmission of gibbon leukemia virus. *J. Natl. Cancer Inst.* **61**, 1113–1115 (1978).
40. T. G. Kawakami, L. Sun, T. S. McDowell, Infectious primate type-C virus shed by healthy gibbons. *Nature* **268**, 448–450 (1977).
41. W. D. Hardy Jr *et al*., Biology of feline leukemia virus in the natural environment. *Cancer Res.* **36**, 582–588 (1976).
42. N. Sarker *et al*., Koala retrovirus viral load and disease burden in distinct northern and southern koala populations. *Sci. Rep.* **10**, 263 (2020).
43. M. W. Cloyd, "Human retroviruses" in *Medical Microbiology*, S. Baron, Ed. (University of Texas Medical Branch at Galveston, Galveston, TX, 1996), ed. 4, chap. 62.
44. G. K. McEwen *et al*., Retroviral integrations contribute to elevated host cancer rates during germline invasion. *Nat. Commun.* **12**, 1316 (2021).
45. O. Olagoke *et al*., 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 (Basel)* **3**, 8 (2018).
46. O. Olagoke, B. L. Quigley, F. Hemmatzadeh, G. Tzipori, P. Timms, Therapeutic vaccination of koalas harbouring endogenous koala retrovirus (KoRV) improves antibody responses and reduces circulating viral load. *NPJ Vaccines (Basel)* **5**, 60 (2020).
47. U. Fiebig *et al*., Induction of neutralizing antibodies specific for the envelope proteins of the koala retrovirus by immunization with recombinant proteins or with DNA. *Virol. J.* **12**, 68 (2015).
48. U. Fiebig, M. G. Hartmann, N. Bannert, R. Kurth, J. Denner, Transspecies transmission of the endogenous koala retrovirus. *J. Virol.* **80**, 5651–5654 (2006).
49. E. Afgan *et al*., The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44** (W1), W3–W10 (2016).
50. J. G. Caporaso *et al*., QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
51. M. Hobbs *et al*., A transcriptome resource for the koala *(Phascolarctos cinereus)*: Insights into koala retrovirus transcription and sequence diversity. *BMC Genomics* **15**, 786 (2014).
52. RStudio Team, *RStudio: Integrated Development for R* (RStudio, Inc., Boston, MA, 2016).