PNAS RESEARCH ARTICLE | MICROBIOLOGY

Geographic patterns of koala retrovirus genetic diversity, endogenization, and subtype distributions

Michaela D. J. Blyton[a,b,c,1], Paul R. Young[a,c], Ben D. Moore[b], and Keith J. Chappell[a,c,1]

Edited by Stephen Goff, Columbia University Irving Medical Center, New York, NY; received December 15, 2021; accepted June 6, 2022

Koala retrovirus (KoRV) subtype A (KoRV-A) is currently in transition from exogenous virus to endogenous viral element, providing an ideal system to elucidate retroviral–host coevolution. We characterized KoRV geography using fecal DNA from 192 samples across 20 populations throughout the koala's range. We reveal an abrupt change in KoRV genetics and incidence at the Victoria/New South Wales state border. In northern koalas, *pol* gene copies were ubiquitously present at above five per cell, consistent with endogenous KoRV. In southern koalas, *pol* copies were detected in only 25.8% of koalas and always at copy numbers below one, while the *env* gene was detected in all animals and in a majority at copy numbers above one per cell. These results suggest that southern koalas carry partial endogenous KoRV-like sequences. Deep sequencing of the *env* hypervariable region revealed three putatively endogenous KoRV-A sequences in northern koalas and a single, distinct sequence present in all southern koalas. Among northern populations, *env* sequence diversity decreased with distance from the equator, suggesting infectious KoRV-A invaded the koala genome in northern Australia and then spread south. The exogenous KoRV subtypes (B to K), two novel subtypes, and intermediate subtypes were detected in all northern koala populations but were strikingly absent from all southern animals tested. Apart from KoRV subtype D, these exogenous subtypes were generally locally prevalent but geographically restricted, producing KoRV genetic differentiation among northern populations. This suggests that sporadic evolution and local transmission of the exogenous subtypes have occurred within northern Australia, but this has not extended into animals within southern Australia.

koala retrovirus | geography | genetic diversity | endogenization | evolution

Significance

Retrovirus infection is synonymous with disease; however, retroviruses can also become endogenous (incorporated into the germline) and directly contribute to a species' genetic makeup. This process has occurred multiple times in all vertebrates, yet little is known about the endogenization process. Koala retrovirus (KoRV) provides an ideal system to study this process as it is in transition between exogenous disease-causing virus and endogenous viral element. This study reveals a comprehensive picture of KoRV biogeography that informs understanding of how host population history, host suppression, and transmission dynamics can influence retroviral evolution. KoRV is also associated with chlamydiosis and neoplasia in the endangered koala. Improved understanding of how KoRV variants are distributed should guide conservation management to help limit disease.

Retroviruses play a unique role in host evolution and health. Generally, retroviruses are transmitted between individuals within a species and cause severe disease through exogenous infection of somatic cells, such as those of the immune system. Some retroviruses, however, can also infect the germline and become a fundamental component of the host's genome that is thereafter passed onto progeny, a process known as endogenization. Consequently, retroviral genetic sequences have accumulated in the genomes of all vertebrates throughout their evolution and in humans now compose 8% of the genome (1). However, little is known about the early processes by which retroviral sequences colonize a host genome and how this impacts the host and virus. This is because the initial endogenization of most retroviruses occurred many millions of years ago, with the koala retrovirus (KoRV) being an exception (2).

KoRV is a gammaretrovirus that has been associated with chlamydiosis and may cause neoplasia in its solitary arboreal host, the koala (*Phascolarctos cinerus*, Marsupialia) (3–6). It is thought that KoRV subtype A entered koala populations from an unknown source (2, 7), with endogenization subsequently occurring between 50,000, and 120 y ago (2, 8). This initial endogenization event is the youngest known, and KoRV is still found in both endogenous and exogenous forms today (7, 9, 10). As such, KoRV provides a valuable system to elucidate retroviral–host coevolution. In this study we investigate the biogeography of KoRV across the koala's natural range to reveal patterns of genetic diversification and differentiation that shed light on exogenous and endogenous adaptation and transmission dynamics.

KoRV infects koalas throughout their natural range in eastern Australia, yet the endogenous and exogenous forms show distinct distributions. In the northern part of the koala's range (in the states of Queensland [QLD] and New South Wales [NSW]), endogenous KoRV (enKoRV) loci are present in 100% of northern koalas, while in southern Australia (states of Victoria [Vic] and South Australia [SA]), quantitative PCR of the *pol* proviral gene suggests that KoRV only occurs as an exogenous virus at a prevalence of ∼15 to 50% (6, 11, 12). It has been suggested that this geographic pattern reflects the spread of endogenous KoRV from the north, where an original endogenization event occurred, to the south (11). Such a spread would be expected to produce an

Author contributions: M.D.J.B. and K.C. designed research; M.D.J.B. performed research; B.D.M. contributed new reagents/analytic tools; M.D.J.B. and K.C. analyzed data; and M.D.J.B., P.Y., B.D.M., and K.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

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

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2122680119/-/DCSupplemental.

Published August 9, 2022.

endogenization gradient. However, the incidence of koalas carrying endogenous KoRV appears to change abruptly between Victoria and NSW (11–13). No populations containing koalas both with and without endogenous KoRV have been reported. An alternative explanation proposed is that koalas in southern Australia are resistant to KoRV and/or have eliminated full-length, replication-competent KoRV from their genomes (14). Some koalas in SA express the terminal regions of the KoRV genome (14), which could indicate they carry either degraded KoRV variants and/or an apparently replication-incompetent, recombination-derived KoRV-like element known as recKoRV (15). A key factor is that koalas throughout Victoria and SA are generally genetically similar to one another as their populations were mostly reestablished from southern offshore islands at the beginning of the 20th century, after they were earlier decimated by hunting (16–22). This may have facilitated the dissemination of degraded KoRV and/or KoRV resistance in the south but could also directly explain the abrupt change in KoRV distributions between NSW and Victoria. Therefore, further investigation of KoRV's genetic and geographic patterns is required to elucidate historic and contemporary KoRV–koala dynamics.

Retroviral subtypes are one important component of genetic diversity and can reflect host–virus evolutionary history and transmission dynamics. New variants/subtypes may be selected to allow superinfection in populations where virus incidence is high (23, 24). KoRV can be delineated into several subtypes based on the amino acid sequence of the receptor binding domain of the envelope protein (25). Only subtype A is known to occur endogenously (7, 26) and is also found as an exogenous virus throughout Australia (11, 12). KoRV subtype B was first identified in a captive koala colony in Japan (referred to as KoRV-J) and was subsequently isolated from captive koalas in the San Diego Zoo (3, 27). It has since been detected in wild koalas throughout northern Australia (4, 5, 13, 28). KoRV subtypes C, D, E, and F were isolated from captive koalas internationally (27, 29), with KoRV-D subsequently found throughout wild northern koalas whereas KoRV-C has only been observed in northern Queensland and KoRV-F is more common in southern Queensland (13). Chappell et al. (25) identified a further three subtypes (G to I) using deep sequencing in wild southern Queensland koalas, and subtype K has recently been identified in captive Australian koalas (10). Several of these exogenous subtypes have also been reported in southern koalas (14, 28, 30). However, these results require confirmation as they were identified at around the detection limit for the methodology used (*Discussion*), and KoRV-B was not detected in southern Australia despite PCR screening of over 640 animals (6).

These recent studies suggest that KoRV subtype diversity is extensive, with some subtypes common throughout the koala's range while others are more restricted. The processes by which these subtypes arise are currently unknown. It is likely that the hypervariable region of the receptor-binding domain is subject to positive selection for mutations and recombination events with host loci including other endogenous retroviral elements that can alter receptor use and so overcome host cell resistance to superinfection, as shown to occur in endogenous feline leukemia virus (31, 32). In feline leukemia virus, subtype A is the only subtype to be transmitted and undergoes recombination within each host to produce subtype B (33, 34). However, KoRV subtypes appear to be exogenously transmitted predominately from mother to offspring (10, 13). This suggests that rather than resulting from repeated recombination, each subtype may have arisen from a rare event and then spread through koala populations via exogenous transmission. Investigating the patterns of genetic diversity and differentiation across the koala's geographic range should provide further insight into which of these processes is likely to predominate, with local differentiation among subtypes B to K in the absence of KoRV-A differentiation more consistent with exogenous transmission than repeated recombination.

In this study we analyzed isolated DNA from washed koala fecal samples to describe KoRV genetic diversity across the koala's natural range with the following specific aims: 1) confirm the geographic extent of KoRV endogenization and identify any populations where endogenization is not complete, 2) characterize the geographic patterns and distributions of KoRV subtypes to gain insights into subtype evolution and transmission dynamics, and 3) assess how KoRV genetic diversity both within and between populations varies over the koala's geographic range to inform our understanding of KoRV evolution. This study represents the most comprehensive investigation of KoRV biogeography to date, encompassing the koala's entire geographic range. By utilizing advanced population genetic analyses not commonly applied in virology, we uncover previously unidentified genetic patterns that challenge previous assumptions and highlight the complex host–virus interactions that occur as a retrovirus transitions from an exogenous, disease-causing virus to endogenous viral element.

Results

KoRV Incidence and Copy Number per Cell. To investigate KoRV's geographic patterns we successfully extracted DNA from 192 fecal wash samples collected from 20 populations throughout the entire koala geographic range of eastern Australia (*SI Appendix, Methods* and Fig. 1*A*). All samples collected from NSW and QLD were KoRV *pol*-positive as determined by qPCR (*SI Appendix, Methods* and Fig. 1*B*). By contrast, 25.8% (24 of 93) of samples collected from Victoria and SA were KoRV *pol*-positive with individual population incidence rates ranging from zero to 55.5% (Fig. 1*B*). Ulupna, situated on the Victorian side of the Murray River, which forms the NSW/Vic boarder, had the highest KoRV *pol* prevalence.

As identified previously (11), there was a distinct difference in KoRV *pol* proviral copy number per koala cell between the northern (QLD/NSW) and southern populations (Vic/SA). All northern koalas had estimated copy numbers greater than two, suggesting that KoRV was endogenous in those populations (35) (Fig. 1 *C* and *D*). Copy number estimates varied considerably among northern koalas, ranging from 5.7 to 97.3 (Fig. 1*D*). The lowest copy numbers were found in koalas from the Monaro region in southern NSW (mean = 8.1, $n = 8$) (Fig. 1*D*). Copy number estimates were considerably lower than the 133 previously estimated from examination of the genome of a northern koala (26) and the 139 to 199 previously estimated from blood/tissue samples (11). However, our estimated copy numbers for Port Macquarie koalas were consistent with those previously reported for fecal samples collected from the same region (35). Therefore, the variation in copy number between these studies may be due to differences in copy number between sample types due to reintegration of endogenous KoRV or exogenous infection in some tissues (36). However, it is conceivable that mismatches between the primers and some KoRV sequence variants may also contribute to the lower number of copies detected.

In the southern populations, estimated *pol* proviral copy number did not exceed 1 in any of the sampled koalas (Fig. 1*D*), which is inconsistent with the presence of proviral sequences in

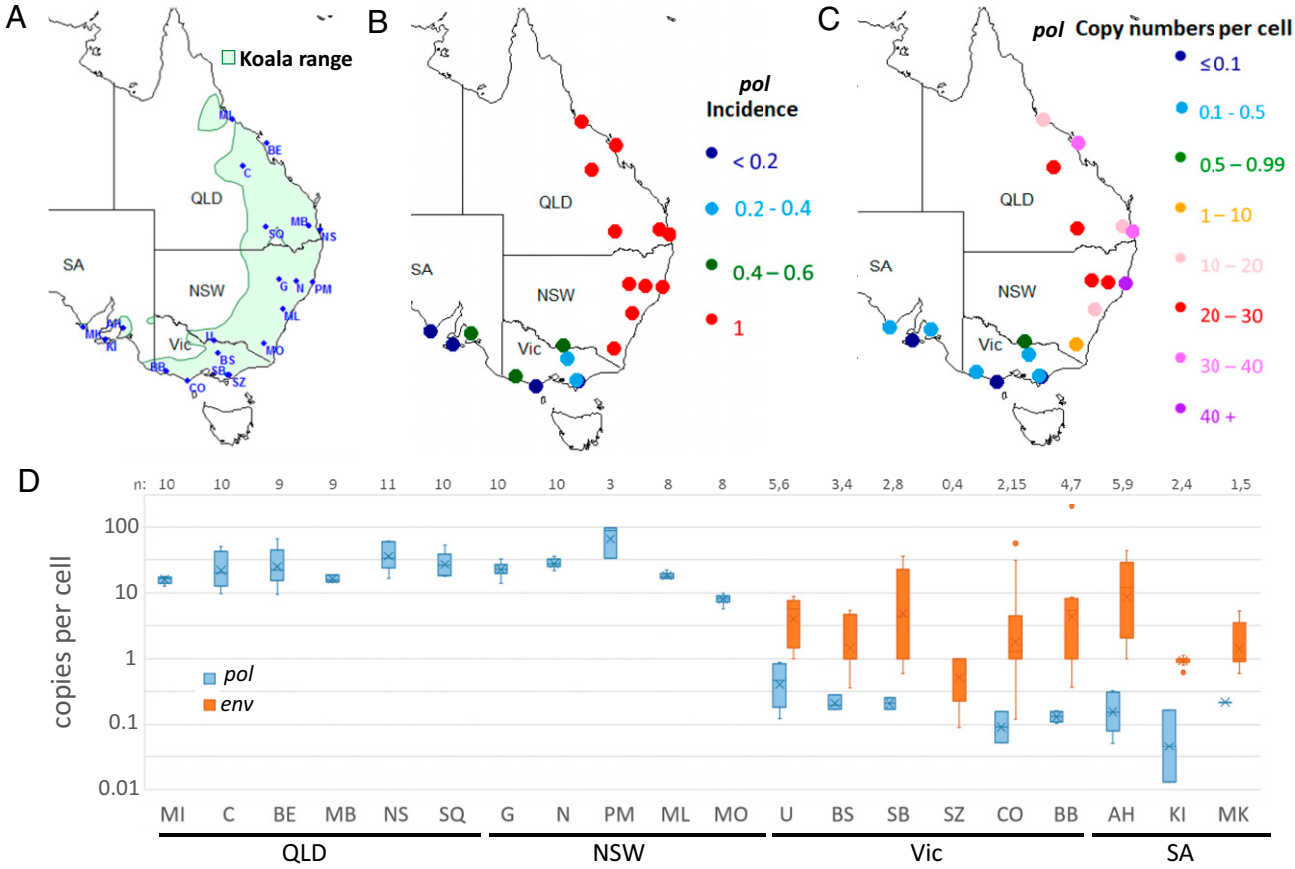

Fig. 1. (*A*) Geographic location, (*B*) KoRV *pol* incidence, and (*C*) average *pol* proviral copies per koala cell for each population. (*D*) Boxplots of the number of KoRV *pol* and *env* copies per koala cell, values shown for *pol*-positive koalas and southern koalas for which single *env* melt curve peaks were obtained, respectively. MI, Magnetic Island; C, Clermont; BE, St. Bee's Island; MB, Mt. Byron; NS, North Stradbroke Island; SQ, southwest Queensland; G, Gunnedah; N, Nowendoc; PM, Port Macquarie; ML, Mountain Lagoon; MO, Monaro; U, Ulupna; BS, Boho South; SZ, Strzelecki 1; SB, Strzelecki 2; CO, Cape Otway; BB, Bessiebelle; AH, Adelaide Hills; KI, Kangaroo Island; MK, Mikkira; QLD, Queensland; NSW, New South Wales; Vic, Victoria; SA, South Australia.

every cell as occurs when endogenous sequences are present. This strongly suggests that intact KoRV is not endogenized in those populations. However, it should be noted that these copy numbers are based on an estimate of 14 copies of β-actin in the koala diploid genome. If the actual number of β-actin copies was in fact 16, then the KoRV *pol* copy number in one of the Ulupna koalas would exceed 1. Therefore, the presence of intact endogenous KoRV at low incidence should not be discounted for the Ulupna population.

In stark contrast to the findings for KoRV *pol*, the incidence of KoRV-A *env* among the 97 southern koalas was 100% as determined by qPCR (*SI Appendix, Methods*). Of these koalas, 35 produced melt curve peaks at multiple temperatures, and 11 did not have a peak at the expected temperature for the original KoRV-A *env* fragment. This potentially indicates that these samples contained multiple sequence variants, preventing accurate copy number estimation. Of the remaining 62 samples, 46 (74.2%) had greater than 1 *env* copy per cell (mean = 11.2, range = 1.1 to 56.3, excluding an outlier from Bessiebelle that had 209 copies per cell) (Fig. 1*D*). This suggests that at least a fragment of the KoRV *env* gene is endogenous in the majority of southern koalas.

Geographic Distribution of KoRV Subtypes. To determine KoRV genetic diversity and subtype incidence across the koala's geographic range, Illumina deep sequencing was conducted on all KoRV *pol*-positive samples ($n = 124$) using an ~500-bp region of the KoRV *env* gene that includes the previously identified hypervariable region within the receptor-binding domain (*SI Appendix, Methods*) (25). To confirm that fecal wash samples could be used to identify subtype presence, we sequenced paired blood and fecal wash samples from eight captive koalas. Subtypes were unable to be detected in the feces when they occurred at less than 0.5% of *env* reads in plasma ($n = 6$). When subtypes were found above 0.5% abundance in the plasma, they were detected in the fecal washes on 65% (13/20) of occasions. Thus, in agreement with the findings of Quigley et al. (28), we determined that fecal wash samples provide a convenient and noninvasive tool for assessing KoRV subtype diversity, although not all subtypes present, particularly those at low abundance, are reliably detected.

The raw data from the *env* amplicon sequencing of the KoRV *pol*-positive samples are deposited in National Center for Biotechnology Information (NCBI) under Bioproject PRJNA813964. A ctotal of 486 unique *env* sequences were characterized after de novo clustering of the quality-controlled deep sequencing reads at 97% similarity and removal of singletons. Of the unique sequences, 173 (35.6%) were found to contain deletions, frameshifts, and/or nonsense mutations and were classified as nonfunctional. Among the remaining 313 intact *env* sequences, the previously identified subtypes A ($n = 87$), B ($n = 19$), C ($n = 15$), D ($n = 118$), F ($n = 1$), I ($n = 8$), and K ($n = 3$) were detected from a combination of Principle Coordinate Analysis (PCoA) (Tables 1 and 2 and *SI Appendix*, Fig. S1), nucleotide similarities (NTS) of the hypervariable region, and the in silico translated amino acid sequences of the hypervariable region. Subtypes E (12), G, and H (2) were not found. Among the subtype D sequences, the

Table 1. Incidence of KoRV subtypes among koala populations

Subtype	Total No. of koalas	No. of populations
A/B intermediate	28	7
B	12	3
C	9	3
D	60	10
Divergent D	19	7
A/D intermediate	10	4
F	2	1
D/F intermediate	6	3
I	4	1
K	1	1
L	6	1
M	4	1

PCoA analysis revealed 39 sequences that were not tightly clustered with the other D sequences, and many of those had low similarities to the D reference sequence (minimum 70.1% NTS). However, the mutations in these sequences were consistent with stepwise mutation from the remaining tightly clustered D sequences and showed amino acid similarity to the reference D sequence. Therefore, they were designated "divergent D" sequences. The translated amino acid sequences also revealed several intermediates between subtypes, with the first section of the hypervariable region showing similarity to one subtype and the remainder resembling a different subtype. These sequences included 17 A/B intermediates, 8 A/D intermediates, and 4 D/F intermediates (Table 2). Two distinct sequence clusters did not show similarity to any of the previously identified subtypes (*SI Appendix*, Figs. S2 and S3) and were designated subtypes L and M. The hypervariable region of subtype L showed some similarity to subtype F (71.6 to 82.0% NTS), while subtype M showed low similarity to all previously identified subtypes (maximum NTS 63.3 to 72.9%).

In northern koala populations, KoRV-A was the dominant subtype and found in all koalas, with abundance ranging between 83.5 and 100% of KoRV reads. In southern populations, KoRV-A was the only subtype detected ($n = 24$). The predominant KoRV-A sequence present in all animals clustered with 97% similarity to the original KoRV-A sequence isolated by Hanger et al. (37) and later confirmed to be endogenous in northern koalas (7). The majority (77.4%) of koalas also possessed additional sequences belonging to KoRV-A at low frequency (maximum 2.1%). Nonfunctional sequences were also detected in most (87.1%) koalas but at very low combined frequency (average = 0.4%, maximum = 5.2%).

To further investigate KoRV-A sequence similarity among koalas, the sequencing reads were reclustered at 100% similarity. This revealed four sequences that each accounted for the majority of reads in one or more koalas. One of these sequences was the original KoRV-A sequence and was found in all northern koalas and eight southern koalas. This sequence predominated in NSW and southern Queensland koalas but was generally at low abundance in north Queensland (Fig. 2). Another sequence that differed from the original sequence by one nucleotide was abundant among Queensland koalas but rare in NSW. A third sequence that also differed from the original KoRV-A by one nucleotide was most abundant in north NSW. Among southern *pol*-positive koalas, all except one animal from the Adelaide Hills had the same dominant sequence that differed by three nucleotides from the original KoRV-A (Fig. 2).

Among the northern koalas, KoRV-D was the most prevalent non–KoRV-A subtype and was detected in 60.6% (60/99) of koalas that were widely distributed across 10 of the 11 populations (Table 1 and Fig. 3*B*). Divergent D sequences and A/D intermediate sequences were also widely distributed across multiple populations, with divergent D sequences also found in a population where D was not detected (Fig. 3*B*). A/B intermediate sequences were more prevalent than KoRV-B sequences and were found in seven populations, including three in NSW where KoRV-B has not been detected (Fig. 3*A*). Different A/B intermediate sequences were detected in each population; however, the breakpoints between the A and B motifs were in similar positions among the majority of sequences (*SI Appendix*, Fig. S4). By contrast, KoRV-B was only recovered from three QLD populations. Subtype C was restricted to northern QLD, while subtypes I, L, and M were locally prevalent (40, 75, and 50%, respectively) but each restricted to a single sampled population (Fig. 3*C*). Subtype F itself was only found in two koalas from Mt. Byron (QLD) (Fig. 3*C*). However, D/F intermediates reached an incidence of 36.4% (4/11) on North Stradbroke Island and were found in a single koala in two other

Table 2. Geographic distribution of KoRV sequence clusters

	Northern populations		Southern populations		North and south	
	Unique to a single population	Shared across multiple populations	Unique to a single population	Shared across multiple populations	Shared between north and south populations	Total
A	30	23	5	36	7	87
A/B intermediate	16	1	0	0	0	17
B	18	1	0	0	0	19
C	9	6	0	0	0	15
D	111	7	0	0	0	118
A/D intermediate	6	2	0	0	0	8
F	1	0	0	0	0	1
D/F intermediate	4	0	0	0	0	4
I	8	0	0	0	0	8
K	3	0	0	0	0	3
L	25	0	0	0	0	25
M	8	0	0	0	0	8
Nonfunctional	71	34	17	58	7	173
Total	310	74	22	94	14	486

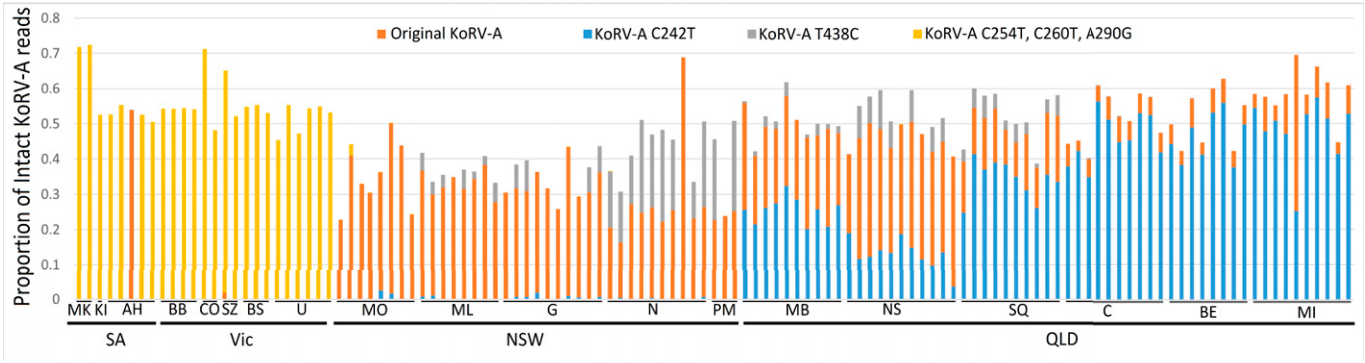

Fig. 2. Proportion of intact KoRV-A *env* reads from each KoRV *pol*-positive koala belonging to the four dominant sequences. Mutation codes are relative to the original KoRV-A sequence (37) with numbering relative to the beginning of the *env* gene. Population and state codes are as described in Fig. 1.

populations. Subtype K was only detected in a single koala from northern QLD (Fig. 3*C*).

env Sequence Sharing and Richness. Here we outline how the number of sequence clusters (sequence richness) detected differed between our study populations and between northern and southern koalas. We also investigate whether there were any geographic patterns to which populations shared sequences. Overall, we found that there was limited sharing of sequences between the north and the south, with higher sharing between populations within these regions. Northern populations had more unique sequences (not shared with other populations) than southern populations, leading to higher total sequence richness in the north than the south. Most of these unique sequences belong to the exogenous subtypes, KoRV-B–M, as KoRV-A sequence richness was actually higher in southern populations.

Of the 486 *env* sequence clusters identified, 312 (64.2%) were detected in multiple koalas, 162 (33.3%) were found in multiple populations, and only 14 were found in both northern and southern koalas (Table 2). Among these sequence clusters, 384 were found in northern koalas, and 116 were found in southern koalas. Similarly, among the 313 intact *env* sequence clusters, there were far more identified in northern koalas (279 sequence clusters) than in southern koalas (41 sequence clusters) with only a very small number (7 sequence clusters) found in both the north and the south (Table 2). This was not an effect of differing sample sizes between the north and the south, because if the same number of koalas were sequenced in the north as were sequenced in the south ($n = 24$), then it is expected that ~120 intact *env* sequence clusters would still have been detected in the north (as determined by rarefaction; Fig. 4*D*), which still greatly exceeds the number detected in the south.

When the study populations were compared pairwise, on average, considerably fewer sequence clusters were shared between a northern and a southern population than between two northern or two southern populations (average number of sequence clusters shared between two populations: north vs. south, 1.57; north vs. north, 6.78; south vs. south, 7.68; Fig. 4*A*). However, among northern populations or southern populations there was no evidence that populations that were geographically closer together shared more sequence clusters (Mantel test, correlation between the number of sequence clusters shared between two populations and the geographic distance between those populations, with significance determined by permutation; north, $P = 0.18$; south, $P = 0.13$).

The number of intact *env* sequence clusters found within a population was generally similar in the north and south, once differences in sample size were taken into account (Fig. 4*G*). This suggests that the higher total number of sequence clusters found in the north is due to different sequences being present in each population, rather than higher within-population richness (Table 2). An exception to this general finding is the Monaro population (NSW), where a higher number of sequence clusters were identified than for the other populations (Fig. 4*G*).

The patterns of KoRV-A sequence sharing among populations reflected those seen for all sequences combined, with fewer sequence clusters shared between the north and the south than between populations within those regions (Table 2 and Fig. 4*B*). However, in contrast to the patterns observed for all intact *env* sequences combined, a similar number of KoRV-A sequence clusters were identified among southern ($n = 38$) and northern ($n = 40$) koalas, despite fewer koalas being sampled in the south (Fig. 4*E*). This indicates that each southern koala carries more KoRV-A sequence clusters than do northern koalas, and this is reflected in a higher number of sequences being identified in each southern population than in northern populations, when the same number of koalas are sampled (Fig. 4*H*). Again, the Monaro population was found to be an

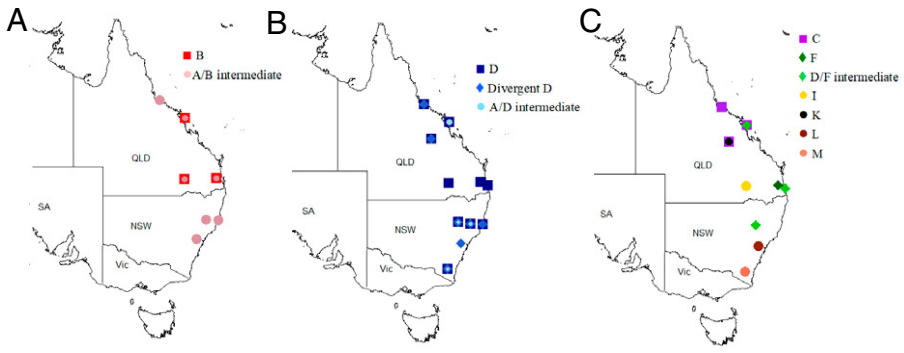

Fig. 3. Geographic distribution of (*A*) KoRV-B and A/B intermediates, (*B*) KoRV-D and A/D intermediates, and (*C*) the other detected subtypes.

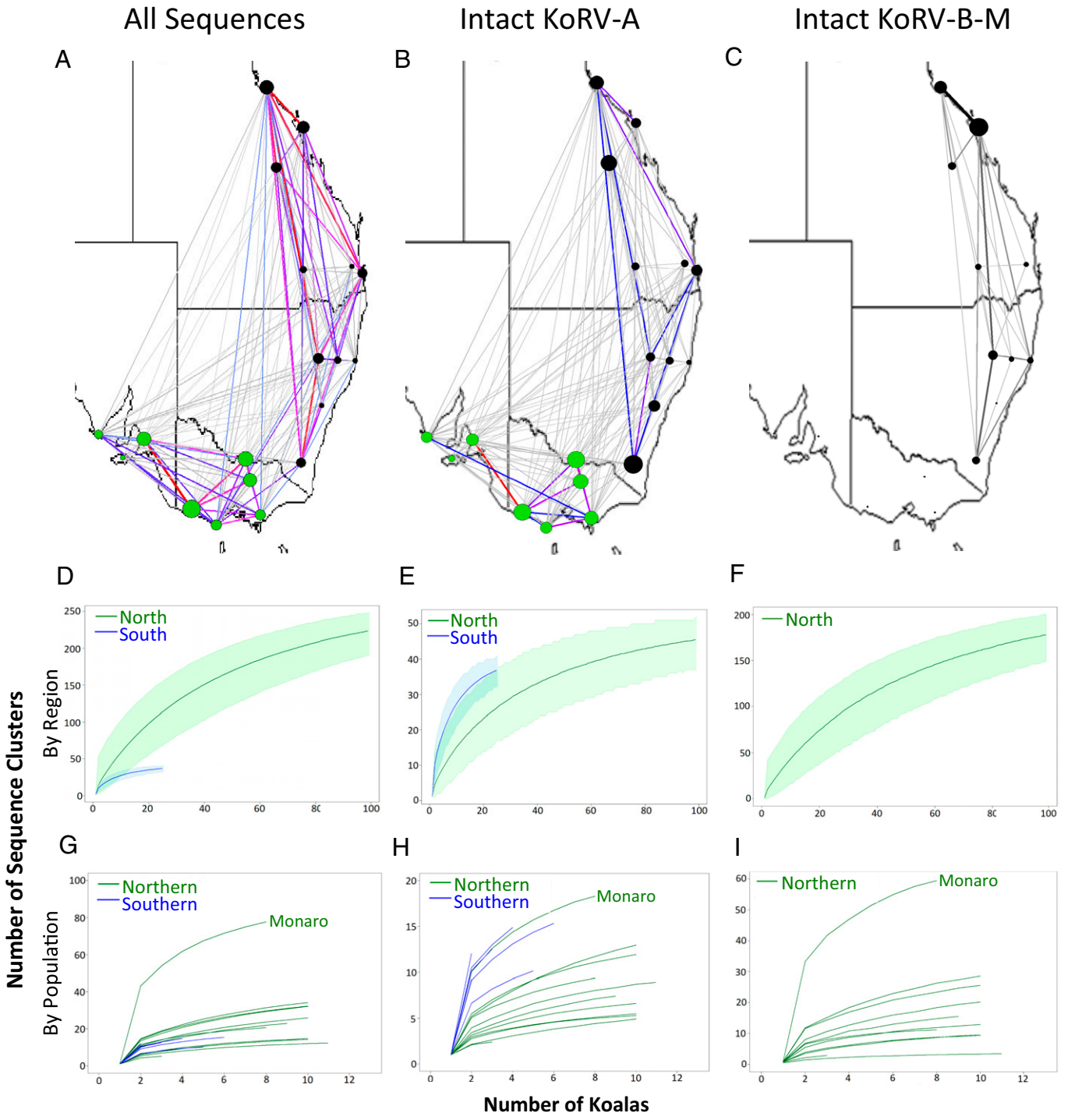

Fig. 4. *Env* sequence sharing between populations and sequence richness (number of unique sequence clusters) within populations and regions. Network diagrams of (*A*) all (intact and nonfunctional) KoRV *env*, (*B*) intact KoRV-A *env* (excluding the original KoRV-A cluster), and (*C*) intact KoRV-B-M *env* sequence clusters shared between populations. Nodes (circles) represent sites, with the size of the node indicating the number of sequence clusters shared with other populations. Black nodes indicate northern populations, and light green nodes indicate southern populations. A line joining two nodes indicates that those two populations shared sequence clusters, with the thickness of the line indicating the number of sequence clusters shared. Gray lines indicate 5 or fewer sequence clusters shared; gradient of blue to red lines indicates 6 to 20 clusters shared. (*D–I*) Rarefaction curves depicting KoRV sequence richness. The total number of intact KoRV *env* sequence clusters (*D* and *G*), intact KoRV-A *env* sequence clusters (*E* and *H*), and intact KoRV-B-M *env* sequence clusters (*F* and *I*) detected in a population (*G–I*) or region (north and south; *D–F*) as the number of koalas sampled is increased. The 95% confidence intervals are shown for *D–F*.

exception and had similar KoRV-A richness to the southern populations (Fig. 4*H*).

Compared to KoRV-A, there were far fewer KoRV-B–M sequences shared between populations (KoRV-A = 59.8%, KoRV-B–M = 7.5%; Fig. 4 *B* and *C* and Table 2). KoRV-B–M richness could not be compared between the north and the south due to the absence of the exogenous subtypes in the south. However, among northern populations, the Monaro population was again an outlier and had far more KoRV-B–M sequences than the other populations (Fig. 4*I*).

env Genetic Diversity throughout the Koala Range. Here we compare how sequence diversity differed between northern and southern koalas and determine if there were any geographic

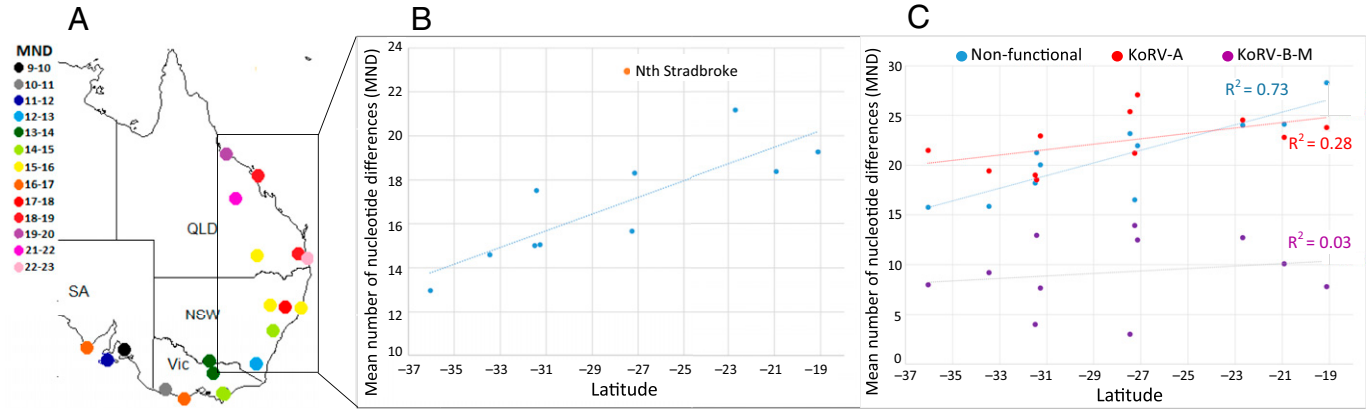

Fig. 5. (*A*) The MND among all sequence clusters detected within each population by geographic location and (*B*) for northern populations in relation to latitude. Dotted line indicates fitted regression, blue points indicate populations included in regression analysis, and the orange point indicates North (Nth) Stradbroke Island. (*C*) The mean number of pairwise nucleotide differences among KoRV-A, KoRV-B-M, and nonfunctional sequences are also shown in relation to latitude for the northern populations.

patterns to how diverse each population's sequences were. In order to assess genetic diversity in the *env* sequences, the number of nucleotides that differed between two sequences was calculated for all sequence pairs, excluding the hypervariable region (that is likely to be derived through recombination and thereby does not adhere to stepwise diversification; Fig. 5*A*). In general, the mean number of nucleotide differences (MND; Fig. 5*A*) for intact and nonfunctional *env* sequence pairs was significantly higher for sequence pairs found in the northern populations than for those found in the southern populations, although not by a large margin (mean ± SD: north, 17.35 ± 3.03; south, 13.13 ± 2.83; $P = 0.007$; Fig. 5*B*).

There was no distinct geographic pattern in sequence diversity for southern populations (Fig. 5*A*), whereas sequence diversity (MND) within the northern populations significantly decreased with distance from the equator ($R^2 = 0.48$, $P = 0.018$). The naturally occurring koala population (not introduced) on North Stradbroke Island had higher sequence diversity than expected from its latitude, and the strength of the correlation between MND and latitude was considerably improved by its exclusion ($R^2 = 0.72$, $P = 0.001$; Fig. 5*B*). When considering the different types of sequence clusters separately, nonfunctional sequence diversity was strongly correlated with latitude ($R^2 = 0.73$, $P < 0.001$; Fig. 5*C*). There was a trend for intact KoRV-A *env* diversity to increase with proximity to the equator ($R^2 = 0.28$, $P = 0.09$), but there was no evidence of a correlation between latitude and sequence diversity among subtypes B to M ($R^2 = 0.03$, $P = 0.58$; Fig. 5*C*).

The genetic center of all populations appears to be the original KoRV-A sequence as it had the lowest MND to other sequences in each population. The Monaro was the only exception, with a divergent D sequence and the KoRV-A original sequence having similar MND (8.72 and 8.74, respectively). Once the hypervariable region was excluded, 18 sequences belonging to subtypes B ($n = 1$), C ($n = 2$), D ($n = 8$), divergent D ($n = 3$), D/F intermediate ($n = 1$), I ($n = 1$), and L ($n = 1$) were identical to the original KoRV-A sequence. This suggests that the exogenous subtypes have arisen from the endogenous KoRV-A sequence, likely through recombination with other exogenous retroviruses and/or the host's own genes. Additionally, the MND between KoRV-B to M sequences was far lower than among KoRV-A sequences across all populations, suggesting they have more recently diversified (Fig. 5*C*).

env Sequence Genetic Differentiation. Analysis of molecular variance (AMOVA) provides a statistical framework to compare the genetic similarity of populations that takes into account both sequence sharing and sequence similarity (the pairwise nucleotide differences among the sequence clusters). This analysis revealed small but significant KoRV *env* genetic differentiation between northern and southern koalas (variance attributed to differences between regions = 2%; $P = 0.004$; Table 3). There was no differentiation between southern populations (variance attributed to differences between southern populations = 0%; $P = 0.755$), while there was significant differentiation accounting for 8% of the total genetic variance across northern populations ($P = 0.001$; Table 3). How different the northern populations were from one another varied between pairs of northern populations (*SI Appendix*, Table S1); however, populations that were geographically more distant were not more genetically different (Mantel test between geographic distance and PhiPT; a measure intraindividual variation: $P = 0.208$; *SI Appendix*, Fig. S5). Instead, the diversity of KoRV sequences within each northern population was distinct (*SI Appendix*, Table S1), with the Monaro considerably more distinct than the other populations (average pairwise PhiPT: Monaro = 0.14; other populations = 0.03 to 0.08).

When considering the different types of sequence clusters separately, KoRV-A differentiation was nonsignificant among northern and southern populations, while there was a low but significant level of differentiation among regions (Table 3). There was also no differentiation among southern populations for the nonfunctional sequence clusters, while there was low but significant differentiation among northern populations (Table 3). By contrast, there was considerable differentiation in KoRV-B to M subtype sequence clusters between northern populations (despite the exclusion of the hypervariable region from the analysis; variance attributed = 19%; $P < 0.001$; Table 3). Thus, the total *env* genetic differentiation observed between northern populations can primarily be attributed to KoRV-B to M.

The Effect of Population Incidence and Within-Koala Abundance on Genetic Differentiation. AMOVA can also account for whether shared or similar sequences are abundant or rare within the populations being compared. When the incidence of (number of koalas carrying) the sequence clusters within each population was included in the AMOVA analysis (instead of assessment based on sequence cluster presence only), overall differentiation among northern populations increased slightly from 8 to 10%, while very low but significant differentiation was found among southern populations (Table 3). Further, when the

Table 3. Summary of molecular variance among populations within regions and between regions

	Between northern populations	Between southern populations	Between north and south
Variation between populations by sequence presence only			
All sequences	8 (0.001)	0 (0.755)	2 (0.004)
KoRV-A	1 (0.18)	0 (0.993)	4 (< 0.001)
KoRV-B-M	19 (< 0.001)	NA	NA
Nonfunctional	3 (0.001)	0 (0.859)	3 (< 0.001)
Variation between populations by sequence and incidence			
All sequences	10 (0.001)	1 (0.012)	2 (0.001)
Variation between populations by sequence, incidence, and abundance			
All sequences	26 (0.001)	0 (0.995)	26 (0.001)

Data are presented as % (*P* value). NA = not applicable.

sequences were also weighted by their average within koala abundance in the AMOVA, differentiation among northern populations increased markedly to 26%, while there remained no differentiation among southern populations (Table 3). This marked increase in differentiation among northern populations reflects the trend for the more abundant sequence clusters within koalas to be found in fewer populations (regression z value = −1.753, $P = 0.080$; *SI Appendix*, Fig. S6*B*).

Among sequence clusters detected in northern koalas, those that had a high incidence within a population were also found at higher abundance within koalas from that population (regression z value = 3.277, $P = 0.001$; *SI Appendix*, Fig. S6*A*). Therefore, between-koala differences in KoRV *env* sequence composition that accounted for 7% of genetic variation based on incidence ($P = 0.001$) became nonsignificant when within koala abundance was taken into account (genetic variance = 1%, $P = 0.554$). These locally abundant sequence clusters predominantly belonged to KoRV-B to M, with only 2 of the 13 most prevalent sequence clusters within northern populations found to belong to KoRV-A (after exclusion of the ubiquitous original KoRV-A cluster).

Discussion

Our assessment of the biogeographic distribution of KoRV sequence diversity provides insights into KoRV evolution and the processes underlying the emergence of novel subtypes and has implications for koala conservation. Here we discuss each of these topics in turn and outline how our findings challenge several previous assumptions to contribute to an improved understanding of KoRV.

KoRV Evolution. In agreement with previous studies (6, 11, 12, 28), our *pol* gene incidence and copy number findings show that all koalas in the northern Australian states of QLD and NSW appear to carry intact, endogenous KoRV. However, in contrast to assumptions made in previous studies based solely on *pol* proviral copy numbers (11, 35) or incorrectly calculated KoRV-A *env* proviral copy numbers (28), our results suggest that the majority of, and potentially all, southern koalas carry partial KoRV-A sequences within their genomes. While partial sequences consisting of the terminal ends of the KoRV genome, including partial *env*, have previously been identified (14, 15), our work shows that such sequences are widespread and additionally that all southern koalas carry an identical *env* sequence which is distinct from the originally described KoRV-A sequence (37) at three nucleotides.

The presence of partial endogenous KoRV sequences across southern koalas but not endogenous full-length KoRV-A has implications for the resounding question of how and where KoRV evolved. Given the relatively young age of KoRV-A [less than 50,000 y (2)], the successful elimination of the large number of full-length copies of KoRV seen in northern koalas from the genomes of southern koalas is unlikely, even when considering the severe population bottleneck that has occurred in southern populations (13, 23). Alternatively, higher KoRV sequence diversity in northern Queensland is consistent with a longer history of KoRV infection (Fig. 6) and suggests that intact KoRV-A first entered the koala genome in northern Australia. It is therefore possible that full-length KoRV-A may not have yet spread to southern koalas, or it may be present in only a very low number of animals. However, if this is the case, then it raises the question of how partial replication incompetent endogenous KoRV sequences arose in southern koalas if they did not result from the degradation of intact KoRV sequences. One explanation could be that historically, intact KoRV did spread in low copy numbers into Victoria and was subsequently lost during the population bottleneck while a replication incompetent version of the virus became fixed or near fixed.

Further research is required to determine how and when these partial KoRV sequences invaded southern koala populations and also to ascertain the role other endogenous and exogenous viral elements played in the origin of KoRV. Löber et al. (15) identified partial KoRV sequences in the koala genome flanking a partial sequence of an older degraded retroelement (PhERV) and named these sequences recKoRV based on their hypothesis that such sequences were the result of recombination with intact KoRV-A. However, analysis of KoRV-A sequence insertions within the reference koala genome reveals a clear difference between those insertions comprising truncated KoRV LTRs (12 to 35% divergence) and those of full-length KoRV-A (<2% divergence) (*SI Appendix*, Fig. S7), suggesting that the truncated sequences may be considerably older and may predate the emergence of full-length KoRV-A. Therefore, the genealogical relationship among the recKoRV sequences, partial KoRV sequences, and full-length KoRV-A is not clear. Simmons et al. (11) identified a similar virus to KoRV in an Australian native rodent (*Melomys burtoni*), and other similar viruses have recently been detected in flying foxes (38, 39). While none of these viruses are the proximate source of KoRV, they demonstrate that multiple closely related retroviruses exist and that there is a wealth of viral biodiversity that may have contributed to contemporary patterns of both full-length and partial KoRV sequences.

It appears that koala population history and demography have impacted the geographic distribution of KoRV. Our assessment of proviral copy numbers, subtype profiles, and KoRV genetic

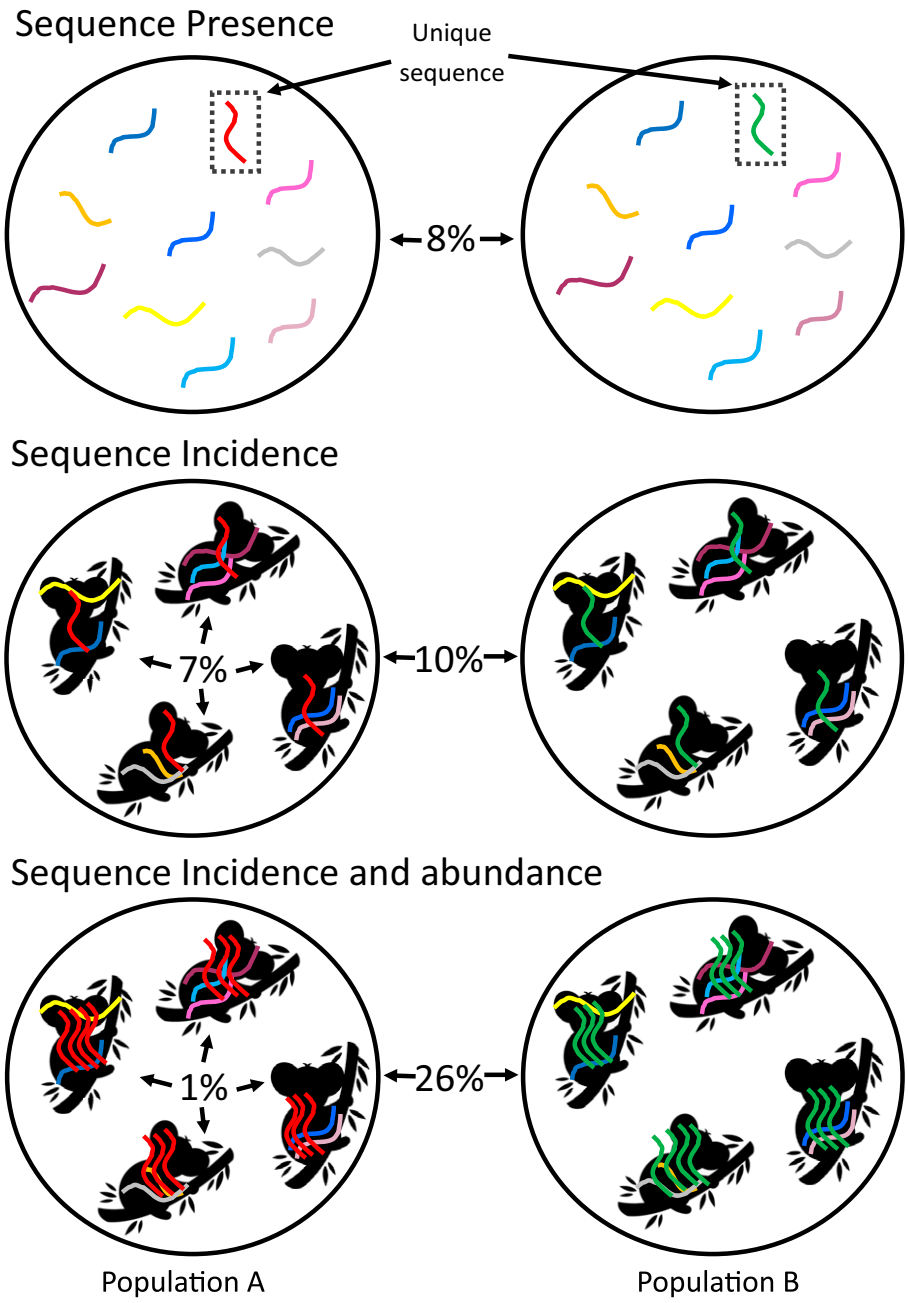

Fig. 6. Schematic of proposed KoRV evolution and spread.

diversity all showed abrupt changes at the Victoria/NSW border. Koalas in the Ulupna population on the Victorian boarder had low genetic diversity and were genetically similar to other koalas in Victoria, while the genetic diversity of the Monaro population in southern NSW was consistent with that of other NSW koalas (*SI Appendix*, Fig. S8 and Table S2). This is consistent with the reintroduction of koalas throughout Victoria and SA from southern offshore islands (11, 21). This historic population bottleneck can explain the absence of KoRV genetic differentiation and the high rate of sequence cluster sharing (despite small sample sizes) among southern populations. Among northern koalas, KoRV *pol* copy number, genetic diversity, and subtype profiles in the Monaro population were found to be divergent from the other populations. Recent phylogenetic analysis of koala populations throughout Australia has revealed a shallow lineage division around the Sydney harbor basin in central NSW, which may indicate a past biogeographic barrier in that region (20). Such a barrier would have reduced migration between the Monaro koala population and those populations farther north. In turn, this would lead to reduced introgression of endogenous KoRV sequences into the Monaro as well as fewer introductions of novel exogenous variants, producing the observed level of KoRV differentiation.

The Emergence of Novel Subtypes. We did not detect any non–KoRV-A subtypes in any of the koalas we sampled from southern Australia (Victoria and SA). This is consistent with previous PCR-based analyses that have not detected KoRV-B in Victoria (6) but is at odds with the results of two deep sequencing studies of koalas in the Mt. Lofty Ranges, South Australia (14, 30), and one of koalas in Victoria (28). The first of the Mt. Lofty Ranges studies (14) inferred the presence of the non–KoRV-A subtypes from the pseudoalignment of a very small number of reads to the hypervariable regions of the different subtypes. Given the small read counts, these assignments are likely to be erroneous, and our reanalysis of these

data (available at the European Nucleotide Archive with the accession number PRJEB21505) using competitive read mapping revealed no compelling evidence for the presence of non–KoRV-A subtypes in their southern koalas. In the second study focusing on koalas within the Mt. Lofty Ranges study (30), there was remarkable similarity between the subtype sequences detected in their Queensland and South Australian study sites, with 93% of sequences shared and an 80% correlation between the number of koalas carrying each sequence in the two populations (calculated from *SI Appendix*). Further, all non–KoRV-A sequences identified in the Victorian koalas were also found in northern koalas from that study (*SI Appendix*) (28). Given that the sequence sharing between two adjacent Queensland captive populations that exchange koalas was only 46% (10), high sequence sharing between northern and southern populations would be unexpected, and it is more likely that the low abundance detection of the non–KoRV-A subtypes in the Mt. Lofty Ranges and Victorian studies was due to cross-contamination during sample preparation or sequencing.

Our findings suggest that new exogenous KoRV subtypes arise from endogenous KoRV-A under selection to escape host suppression and allow superinfection where functional KoRV-A is ubiquitous. In all populations (with the exception of the Monaro) the original endogenous KoRV-A sequence cluster was the genetic center of sequence diversity, with the representative sequences from clusters of six other subtypes found to be identical to that sequence after the hypervariable region was excluded. Additionally, the other subtypes were only detected in populations where KoRV-A was present and had a copy number above 1. These findings on their own could lend support to the hypothesis that endogenous KoRV-A repeatedly gives rise to the other subtypes within hosts, as occurs in feline leukemia virus (33, 34). However, we found that several of the subtypes had localized geographic distributions, with considerable KoRV genetic differentiation between northern populations due to the non–KoRV-A subtypes. This finding suggests that different subtypes arise sporadically across the geographic range and become locally prevalent through selection and exogenous transmission (Fig. 7), which is known to occur at least between mothers and their offspring (10, 13).

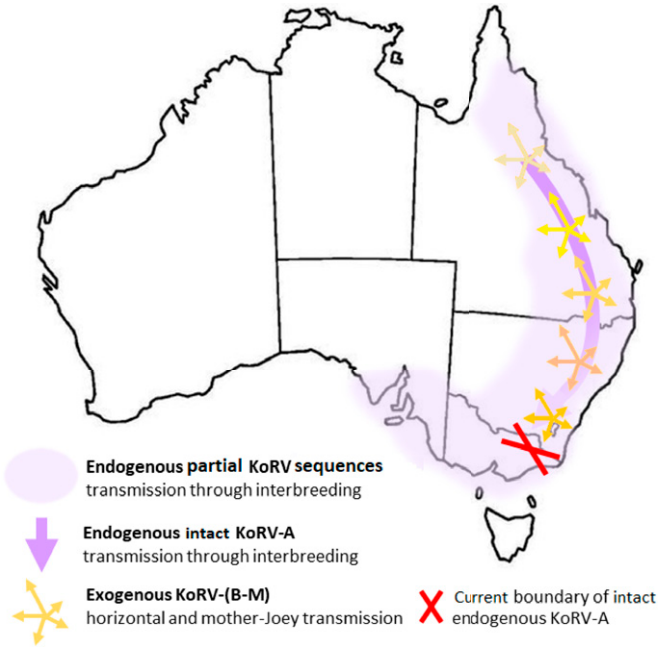

Fig. 7. Schematic of proposed KoRV evolution hypothesis.

Implications for Koala Conservation. The koala is now listed as endangered by the Australian Federal Government and faces severe decline in northeastern Australia due to fire, habitat loss, and disease (40, 41). The characterization of KoRV subtype diversity across the koala's range will likely contribute to the management of disease in the future, particularly for northern koala populations. Little is currently known about the disease association of the majority of the KoRV subtypes. However, there is some evidence that KoRV-B may be more pathogenic than KoRV-A as it has been associated with Chlamydiosis and neoplasia in southern Queensland koalas (3–5). Therefore, it is of concern that KoRV-B has recently been detected in NSW and Queensland by deep sequencing (13, 28) and that we have also identified KoRV-B in northern Queensland and A/B intermediates in NSW. As unique A/B intermediates were detected in each population, they may have been produced by local hybridization instead of widespread transmission of intermediate sequences. Alternatively, since directionality of recombination cannot be assumed and the A/B intermediates were more widely detected than KoRV-B, these intermediate sequences may in fact be ancestral, and KoRV-B is the product of two recombination events. Further investigation of the clinical significance of all the subtypes is needed to more robustly ascertain how their distributions are likely to impact koala health in different areas of Australia. However, it is clear that as the exogenous subtypes are currently geographically restricted, the translocation of koalas among northern populations would lead to the introduction of novel subtypes to a region, with potentially negative impacts on disease.

Whether replication-competent, non–germline-integrated KoRV is present in southern koalas is an important question with implications for how southern populations are managed. In the south, our estimated number of *pol* copies per cell was markedly lower than 1 and was detected in around 26% of animals. Such findings have previously been interpreted as the presence of functional exogenous KoRV in some southern koalas (11, 35); however, it should be noted that the full-length KoRV genome has never been identified in southern koala populations. Nevertheless, a previous study in over 640 koalas did find an association between KoRV-A *pol* positivity and poor body condition in Victorian koalas (6), with a significance level of $P = 0.008$, suggestive of the presence of replication competent KoRV-A. Studies to determine the presence of the entire functional KoRV genome in the south should be high priority to determine whether functional exogenous KoRV is a threat to koalas in the south.

The stark difference in KoRV incidence and diversity between northern and southern koala populations highlights priorities for management and potential containment of KoRV. In a recent study investigating the transmission of exogenous KoRV sequences among captive northern koala populations (10), we found low sequence sharing between nonrelated, cohoused animals and no evidence of sexual transmission. Transmission of exogenous KoRV sequences was overwhelmingly from mother to offspring. Unless an alternate mode of transmission is at play within southern koalas, neither the exogenous subtypes nor endogenous KoRV-A are anticipated to spread into the south quickly, and translocations between northern and southern populations are not advised. Instead, koala populations at the boundary of the northern and southern koala populations (around the NSW/Vic border) are of high interest. The Monaro population in southern NSW displays low proviral copy numbers (28; this study) and yet high sequence richness. This could be attributed to more recent endogenization and a greater expression of KoRV due to an

absence of strong host suppression. Additionally, the presence of koalas with *pol* proviral copy numbers approaching 1 in the Ulupna population could be the result of endogenous functional KoRV-A recently reaching the border through migrant koalas from southern NSW.

Methods

Koala fecal samples were collected between April and September 2016, from 20 locations spanning the koala's current wild geographic range. Total DNA was extracted from surface washes of the fecal samples following the general approach of Wedrowicz et al. (38). Proviral copy numbers per cell were estimated by qPCR amplification of either the KoRV *pol* gene 110-bp fragment or *env* gene 97-bp fragment and normalized against a 123-bp fragment of the koala β-actin gene, using previously published primers (25, 39). KoRV subtype analysis was conducted by Illumina sequencing of the envelope hypervariable region as per published methods (25). The geographic distribution of KoRV subtypes was determined from the *env* Illumina deep sequencing data. The complete methodology used is included in *SI Appendix*.

The fieldwork was carried out under Western Sydney University ethics approval (A11253) and with appropriate permits from the New South Wales (SL101722), Queensland (WITK17277716), Victorian (10007714), and South Australian State (U26533-1) governments. Collection of samples from captive koalas was carried out under University of Queensland ethics approval (AE36153).

Data, Materials, and Software Availability. Sequence data is available; Michaela D.J. Blyton, Geographic patterns of koala retrovirus genetic diversity, endogenization and subtype distributions, NCBI Sequence Read Archive(accession no. PRJNA813964), depositied March 9th 2022, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA813964 (42).

ACKNOWLEDGMENTS. The following people assisted with study site selection and fieldwork: Mark Boorman, Bill Ellis, Sean FitzGibbon, Ben Barth, Dalene Adam, Huiying Wu, Cheyne Flanagan, Mathew Crowther, Caroline Marschner, Laura Schmertmann, Daniel Lunney, Mark Krockenberger, Valentina Mella, Kellie Leigh, Faye Wedrowicz, James Fitzgerald, Kath Handasyde, Jack Pascoe, Emily Hynes, Robyn Molsher, and Jason VanWeenen. We thank Peter Smouse for his advice on the statistical analysis of KoRV genetic diversity and differentiation. The fieldwork was funded through a grant from the Australian Research Council (LP140100751) in partnership with Evolva Biotech A/S, the NSW Office of Environment and Heritage, and The Conservation Ecology Trust. The molecular analysis was funded by the Australian Research Council grant (DP180103362).

Author affiliations: [a]School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, QLD 4072, Australia; [b]Hawkesbury Institute for the Environment, Western Sydney University, Richmond, NSW 2753, Australia; and [c]Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, St. Lucia, QLD 4072, Australia

1. M. Bock, J. P. Stoye, Endogenous retroviruses and the human germline. *Curr. Opin. Genet. Dev.* **10**, 651–655 (2000).
2. 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).
3. 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).
4. B. L. Quigley, V. A. Ong, J. Hanger, P. Timms, Molecular dynamics and mode of transmission of Koala Retrovirus (KoRV) as it invades and spreads through a wild Queensland koala population. *J. Virol.* **92**, e01871–e01817 (2018).
5. 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).
6. 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).
7. R. E. Tarlinton, J. Meers, P. R. J. N. Young, Retroviral invasion of the koala genome. *Nature* **442**, 79–81 (2006).
8. M. C. Ávila-Arcos *et al.*, One hundred twenty years of koala retrovirus evolution determined from museum skins. *Mol. Biol. Evol.* **30**, 299–304 (2012).
9. N. M. Oliveira, H. Satija, I. A. Kouwenhoven, M. V. Eiden, Changes in viral protein function that accompany retroviral endogenization. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17506–17511 (2007).
10. B. A. Joyce, M. D. J. Blyton, S. D. Johnston, P. R. Young, K. J. Chappell, Koala retrovirus genetic diversity and transmission dynamics within captive koala populations. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2024021118 (2021).
11. G. S. Simmons *et al.*, Prevalence of koala retrovirus in geographically diverse populations in Australia. *Aust. Vet. J.* **90**, 404–409 (2012).
12. F. Wedrowicz, J. Mosse, W. Wright, F. E. Hogan, Using non-invasive sampling methods to determine the prevalence and distribution of *Chlamydia pecorum* and koala retrovirus in a remnant koala population with conservation importance. *Wildl. Res.* **45**, 366–380 (2018).
13. B. L. Quigley *et al.*, Koala retrovirus in Northern Australia shows a mixture of stable endogenization and exogenous lineage diversification within fragmented koala populations. *J. Virol.* **95**, e02084–e02020 (2021).
14. R. E. Tarlinton *et al.*, Differential and defective transcription of koala retrovirus indicates the complexity of host and virus evolution. *J. Gen. Virol.* **103**, 001749 (2022).
15. U. Löber *et al.*, Degradation and remobilization of endogenous retroviruses by recombination during the earliest stages of a germ-line invasion. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8609–8614 (2018).
16. B. A. Houlden, P. R. England, A. C. Taylor, W. D. Greville, W. B. Sherwin, Low genetic variability of the koala Phascolarctos cinereus in south-eastern Australia following a severe population bottleneck. *Mol. Ecol.* **5**, 269–281 (1996).
17. A. M. Seymour *et al.*, "High effective inbreeding coefficients correlate with morphological abnormalities in populations of South Australian koalas (*Phascolarctos cinereus*)" in *Animal Conservation Forum* (Cambridge University Press, 2001), pp. 211–219.
18. Q. Lau, W. Jaratlerdsiri, J. E. Griffith, J. Gongora, D. P. Higgins, MHC class II diversity of koala (*Phascolarctos cinereus*) populations across their range. *Heredity* **113**, 287–296 (2014).
19. W. Ellis, P. Hale, F. Carrick, Breeding dynamics of koalas in open woodlands. *Wildl. Res.* **29**, 19–25 (2002).
20. S. R. Kjeldsen *et al.*, Genomic comparisons reveal biogeographic and anthropogenic impacts in the koala (*Phascolarctos cinereus*): A dietary-specialist species distributed across heterogeneous environments. *Heredity* **122**, 525–544 (2018).
21. P. Menkhorst, "Hunted, marooned, re-introduced, and contracepted: A history of Koala management in Victoria" in *Too Close for Comfort: Contentious Issues in Human-wildlife Encounters*, D. Lunney, L. Munn, W. Meikle, Eds. (Royal Zoological Society of NSW, Mosman, Australia, 2008), pp. 73–92.
22. L. E. Neaves *et al.*, Phylogeography of the koala,(Phascolarctos cinereus), and harmonising data to inform conservation. *PLoS One* **11**, e0162207 (2016).
23. K. T. Arrildt, S. B. Joseph, R. Swanstrom, The HIV-1 env protein: A coat of many colors. *Curr. HIV/AIDS Rep.* **9**, 52–63 (2012).
24. R. A. Taplitz, J. M. Coffin, Selection of an avian retrovirus mutant with extended receptor usage. *J. Virol.* **71**, 7814–7819 (1997).
25. K. J. Chappell *et al.*, Phylogenetic diversity of koala retrovirus within a wild koala population. *J. Virol.* **91**, e01820–e01816 (2017).
26. M. Hobbs *et al.*, Long-read genome sequence assembly provides insight into ongoing retroviral invasion of the koala germline. *Sci. Rep.* **7**, 15838 (2017).
27. T. Shojima *et al.*, Identification of a novel subgroup of koala retrovirus from koalas in Japanese zoos. *J. Virol.* **87**, 9943–9948 (2013).
28. B. L. Quigley, F. Wedrowicz, F. Hogan, P. Timms, Phylogenetic and geographical analysis of a retrovirus during the early stages of endogenous adaptation and exogenous spread in a new host. *Mol. Ecol.* **30**, 2626–2640 (2021).
29. W. Xu, K. Gorman, J. C. Santiago, K. Kluska, M. V. Eiden, Genetic diversity of koala retroviral envelopes. *Viruses* **7**, 1258–1270 (2015).
30. 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).
31. Y. Anai *et al.*, Infectious endogenous retroviruses in cats and emergence of recombinant viruses. *J. Virol.* **86**, 8634–8644 (2012).
32. M. A. Stewart *et al.*, Nucleotide sequences of a feline leukemia virus subgroup A envelope gene and long terminal repeat and evidence for the recombinational origin of subgroup B viruses. *J. Virol.* **58**, 825–834 (1986).
33. J. Overbaugh, N. Riedel, E. A. Hoover, J. I. Mullins, Transduction of endogenous envelope genes by feline leukaemia virus in vitro. *Nature* **332**, 731–734 (1988).
34. J. L. Rohn, M. L. Linenberger, E. A. Hoover, J. Overbaugh, Evolution of feline leukemia virus variant genomes with insertions, deletions, and defective envelope genes in infected cats with tumors. *J. Virol.* **68**, 2458–2467 (1994).
35. F. Wedrowicz, T. Saxton, J. Mosse, W. Wright, F. E. J. C. G. R. Hogan, A non-invasive tool for assessing pathogen prevalence in koala (Phascolarctos cinereus) populations: Detection of Chlamydia pecorum and koala retrovirus (KoRV) DNA in genetic material sourced from scats. *Conserv. Genet. Resour.* **8**, 511–521 (2016).
36. M. A. Hashem *et al.*, Transmission of koala retrovirus from parent koalas to a joey in a Japanese zoo. *J. Virol.* **94**, e00019–e00020 (2020).
37. 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).
38. F. Wedrowicz, M. Karsa, J. Mosse, F. E. Hogan, Reliable genotyping of the koala (Phascolarctos cinereus) using DNA isolated from a single faecal pellet. *Mol. Ecol. Resour.* **13**, 634–641 (2013).
39. M. E. H. Kayesh *et al.*, Molecular dynamics of koala retrovirus infection in captive koalas in Japan. *Arch. Virol.* **164**, 757–765 (2019).
40. G. Gordon, F. Hrdina, R. Patterson, Decline in the distribution of the koala *Phascolarctos cinereus* in Queensland. *Aust. Zool.* **33**, 345–358 (2006).
41. V. Gonzalez-Astudillo, R. Allavena, A. McKinnon, R. Larkin, J. Henning, Decline causes of Koalas in South East Queensland, Australia: A 17-year retrospective study of mortality and morbidity. *Sci. Rep.* **7**, 42587 (2017).
42. M. D. J. Blyton, Geographic patterns of koala retrovirus genetic diversity, endogenization and subtype distributions. NCBI BioProject. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA813964. Deposited 9 March 2022.