



# A recent gibbon ape leukemia virus germline integration in a rodent from New Guinea

Saba Mottaghinia<sup>a,b</sup> , Saskia Stenzel<sup>c,d</sup> , Kyriakos Tsangaras<sup>e</sup> , Nikolas Nikolaidis<sup>f</sup> , Michael Laue<sup>g</sup> , Karin Müller<sup>h</sup> , Henriette Hölscher<sup>a</sup> , Ulrike Löber<sup>i</sup> , Gayle K. McEwen<sup>a</sup>, Stephen C. Donnellan<sup>j</sup> , Kevin C. Rowe<sup>k</sup> , Ken P. Aplin<sup>l</sup>, Christine Goffinet<sup>c,d</sup> , and Alex D. Greenwood<sup>a,k,2</sup>

Edited by John Coffin, Tufts University, Boston, MA; received December 9, 2022; accepted November 30, 2023

Germline colonization by retroviruses results in the formation of endogenous retroviruses (ERVs). Most colonization's occurred millions of years ago. However, in the Australo-Papuan region (Australia and New Guinea), several recent germline colonization events have been discovered. The Wallace Line separates much of Southeast Asia from the Australo-Papuan region restricting faunal and pathogen dispersion. West of the Wallace Line, gibbon ape leukemia viruses (GALVs) have been isolated from captive gibbons. Two microbat species from China appear to have been infected naturally. East of Wallace's Line, the woolly monkey virus (a GALV) and the closely related koala retrovirus (KoRV) have been detected in eutherians and marsupials in the Australo-Papuan region, often vertically transmitted. The detected vertically transmitted GALV-like viruses in Australo-Papuan fauna compared to sporadic horizontal transmission in Southeast Asia and China suggest the GALV-KoRV clade originates in the former region and further models of early-stage genome colonization may be found. We screened 278 samples, seven bat and one rodent family endemic to the Australo-Papuan region and bat and rodent species found on both sides of the Wallace Line. We identified two rodents (*Melomys*) from Australia and Papua New Guinea and no bat species harboring GALV-like retroviruses. *Melomys leucogaster* from New Guinea harbored a genomically complete replication-competent retrovirus with a shared integration site among individuals. The integration was only present in some individuals of the species indicating this retrovirus is at the earliest stages of germline colonization of the *Melomys* genome, providing a new small wild mammal model of early-stage genome colonization.

endogenous retrovirus (ERV) | gibbon ape leukemia virus (GALV) | koala retrovirus (KoRV) | woolly monkey virus (WMV)

Retroviruses integrate into the host genome as part of their replication cycle. While the majority of integrations occur in somatic cells, several families of retroviruses have frequently integrated into the host germline, resulting in the development of vertically transmitted endogenous retroviruses (ERVs). Approximately 8 to 10% of vertebrate genomes are composed of ERVs (1–3). Not all retroviral families colonize vertebrate genomes with equal frequency. Gammaretroviruses, for example, the murine leukemia virus-related viruses, frequently colonize the germline in vertebrates when compared to other retroviral groups (4, 5). While most vertebrate retroviral colonization events were completed millions of years ago, Gibbon ape leukemia virus (GALV) and the closely related koala retrovirus (KoRV) represent gammaretroviruses that have colonized or have recently begun to colonize the genomes of a variety of mammals in Southeast Asia, the Australo-Papuan region (Australia, New Guinea), and Wallacea (including the Philippines apart from Palawan) (6). These regions represent historic biogeographical realms that have limited natural faunal dispersion among them as demarcated by the Wallace (1863), and Huxley (1868), and Lydekker (1896) Lines (Fig. 1). To date, according to the IUCN (The International Union for Conservation of Nature Red List of Threatened Species. v2020-3, <http://www.iucnredlist.org>), approximately eight bat families and three genera of murine rodents from the following two tribes have distributions that span the Wallace Line: i) Hydromyini: *Haeromys* (Borneo and Sulawesi), *Chiropodomys* (West of the Wallace Line in the Asian continental shelf), ii) Rattini: *Rattus* (from Asia to Philippines and Australia), *Maxomys* (from Asia to Sulawesi), and iii) Murini: *Mus musculus* (West of the Wallace line in Sunda shelf) (7).

GALVs are oncogenic viruses consisting of seven isolates (8–14), initially detected in captive white-handed gibbons (*Hylobates lar*) in Thailand. The basal GALV strain, the woolly monkey virus (WMV—previously known as simian sarcoma-associated virus) was isolated from a brown woolly monkey (*Legothrix lagothrica*) that was co-housed with infected gibbons and likely represents a gibbon to woolly monkey transmission (15, 16). GALVs are closely related to the recently identified ERVs in *Melomys* rodents, endemic to Australia (MbrV)

## Significance

Retroviruses colonize vertebrate genomes forming endogenous retroviruses. With very few exceptions, these colonization events are ancient. After screening 278 samples representing seven bat and one rodent family endemic to the Australo-Papuan region (Australia and New Guinea), we report the discovery of genomically intact and infectious retroviruses currently colonizing the genome of a *Melomys leucogaster* in New Guinea. This represents the second example, after the koala retrovirus (KoRV), of a retrovirus that has colonized the genome but retains a functional viral life cycle identified in the Australo-Papuan region.

Author contributions: S.M., C.G., and A.D.G. designed research; S.M., S.S., K.T., N.N., M.L., K.M., H.H., and C.G. performed research; G.K.M. and K.P.A. contributed new reagents/analytic tools; S.M., S.S., K.T., N.N., M.L., K.M., U.L., and C.G. analyzed data; S.D. sample collection and contribution; K.C.R. contributed by providing samples that are vital for this study; K.P.A. contributed by providing samples that are vital for this study; and S.M., S.S., N.N., M.L., K.M., U.L., S.C.D., K.C.R., C.G., and A.D.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Although PNAS asks authors to adhere to United Nations naming conventions for maps (<https://www.un.org/geospatial/mapsgeo>), our policy is to publish maps as provided by the authors.

<sup>1</sup>Deceased January 15, 2019.

<sup>2</sup>To whom correspondence may be addressed. Email: [greenwood@izw-berlin.de](mailto:greenwood@izw-berlin.de).

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2220392121/-/DCSupplemental>.

Published February 1, 2024.

(17) and the North Moluccas (MelWMV) (18, 19), forming a monophyletic clade with KoRV, which thus far is found exclusively in koalas (*Phascolarctos cinereus*). Extensive screening for GALV in wild gibbons failed to detect the virus (20), suggesting that all isolates in primates are derived from horizontal transmission during captivity. The recently characterized gammaretroviruses (FFRV1, HPG, MmGRV, and SaGRV) from wild-caught Australian bats (*Pteropus alecto*, *Macroglossus minimus*, and *Syconycteris australis*) form a clade basal to GALV-KoRV, while HIGRV and RhGRV, isolates from the Chinese bats (*Hipposideros larvatus* and *Rhinolophus hipposideros*) are a GALV sister clade (21, 22). The number of identified GALV-KoRV-related viruses in Australo-Papuan wildlife is greater than in Southeast Asia or China with only two isolates from bat species that have historically spanned the Wallace line (23, 24). Therefore, this complex distribution suggests the GALV-KoRV clade derives from the Australo-Papuan side of the Wallace Line and that the related viruses found in Southeast Asia or China represent either human-mediated spillover or infection of taxa in the Australo-Papuan region that can cross the Wallace Line.

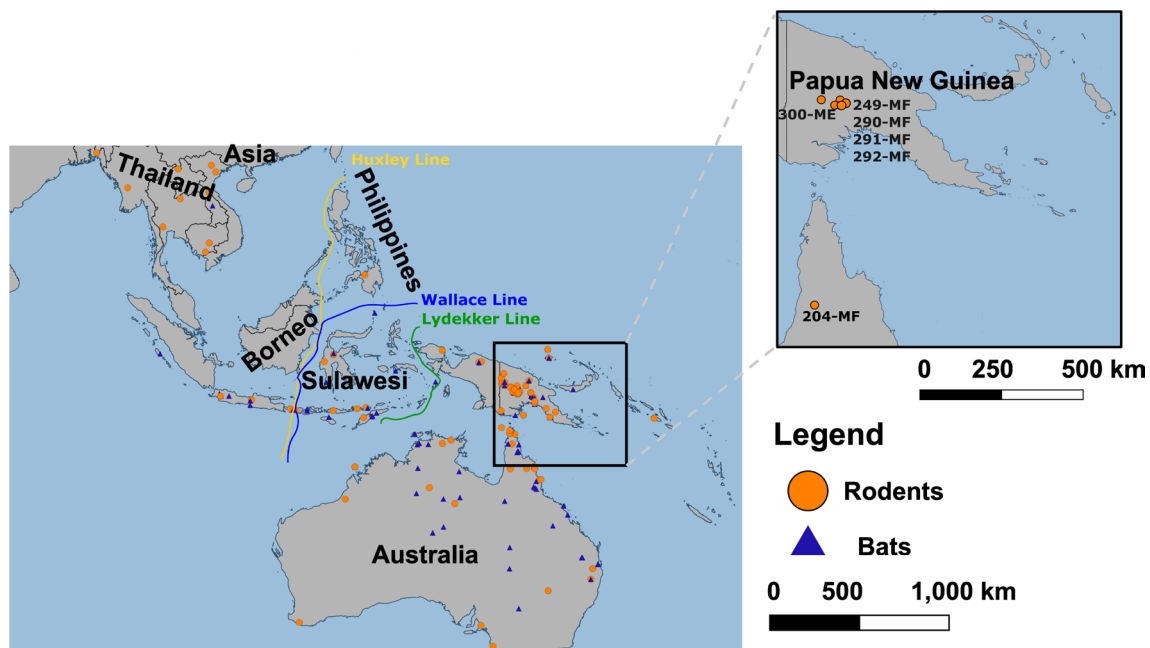
To identify potential reservoirs for GALV-KoRV viruses and evidence of germline colonization events in the Australo-Papuan and Wallacean regions, we used pan-GALV-KoRV PCR, hybridization capture viral enrichment and high throughput sequencing to screen bat and rodent species from both sides of the Wallace line (19). We identified genomically intact and infectious WMV, denoted complete melomys woolly monkey retrovirus (cMWMV), in some but not all populations of *Melomys leucogaster* endemic to New Guinea. Among individual *M. leucogaster* which were cMWMV positive, the retrovirus was found at a single shared integration site, indicating that it has been transmitted vertically and is a part of the *M. leucogaster* germline but has not become fixed in the species. No bats, including members of the same species found harboring GALV relatives in Australia and China, were positive. Structural modeling of the few variable amino acids among the retrieved sequences and WMV and in vitro infection models suggest cMWMVs identified are replication-competent. Our data suggest that cMWMV represents an additional model

to KoRV for exploring the earliest stages of retroviral germline colonization.

## Results

**WMVs in the Australo-Papuan Region.** The degenerate oligonucleotide primer set (KOGAWM-1) was designed to amplify the *gag* gene of any GALV, KoRV, and WMV. The DNA of *R. norvegicus* was used as a negative control as these viral clades are absent from *Rattus*. Various tissue samples ( $n = 278$ ) from seven bat families, and the murine rodents *Rattus* (spanning the Wallace Line), *Hydromys*, and *Melomys* (spanning the Lydekker Line) were PCR-screened for the presence of GALV and KoRV-like sequences (Fig. 1, Table 1, and Dataset S1). None of the 156 bat samples yielded an amplicon. However, liver and kidney tissue samples of five *M. leucogaster* ( $n = 10$ ), a *Rattus verecundus* ( $n = 5$ ), and a *Rattus niobe*\_sp.B ( $n = 5$ ) specimen collected between 1985 to 2014 in Western and Southern Highland Provinces and two *Melomys burtoni* samples ( $n = 7$ ) from Queensland of Australia each yielded an amplicon which had 89 to 100% sequence identity (Sanger sequencing) to WMV (Dataset S1 marked \*). *Melomys* is one of the most speciose genera (*Melomys* = 23 species, *Pseudomys* = 23 species, *Rattus* = 26 species) (25, 26) in the Australo-Papuan region with ongoing taxonomic revisions (27–30). Though currently confined to the east side of the Wallace line, the nearest relatives of the Australo-Papuan *Rattus* are found on Sulawesi Island, while the nearest relatives of *Melomys* (*Solomys*, *Protochromys*, *Paramelomys*, and *Uromys* in the *Uromys* division) are found among the Australo-Papuan “old endemic” rodents (27–30). The distribution and diet of the Papuan white-bellied melomys (*M. leucogaster*) overlap with those of the moss-forest rat (*R. niobe*) and slender rat (*R. verecundus*), especially along the New Guinean Central Cordillera (31).

**Hybridization Capture Viral Enrichment.** Samples (89, 201, 204, 246, 249, 290, 291, 292, and 300) that yielded an amplicon were used for Illumina library preparation and subsequent hybridization



**Fig. 1.** The approximate locality for rodent (orange circles) and bat (blue triangles) samples tested, corresponding to sample details in Dataset S1. The closeup map shows the approximate locality of the *M. leucogaster* (249, 290, 291, and 292) and *M. burtoni* (204) harboring cMWMV. Map created using the free and open source QGIS version 3.16.10-Hannover. The Lydekker (1896) (red dashes), Wallace (1863), and Huxley's extension to the Wallace Line (1868) (white line) were drawn manually.

**Table 1. Chiropteran and rodent genera tested for GALV-like viruses in the current study and number of each tested**

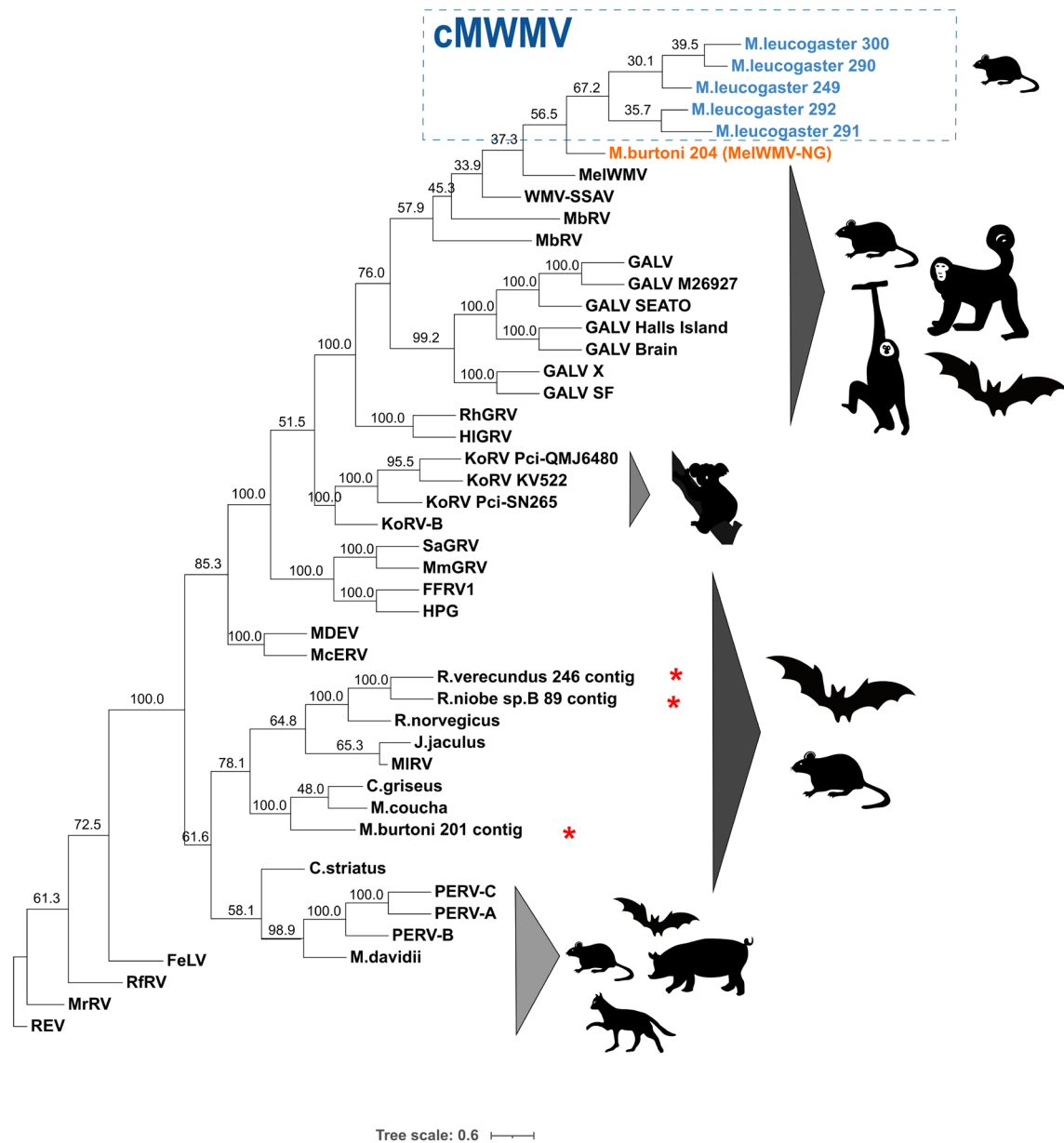
Order	Family	Genus	Number of samples	Yielded an amplicon
Chiroptera	Emballonuridae	<i>Emballonura</i>	5	no
Chiroptera	Emballonuridae	<i>Mosia</i>	1	no
Chiroptera	Emballonuridae	<i>Saccolaimus</i>	2	no
Chiroptera	Emballonuridae	<i>Taphozous</i>	3	no
Chiroptera	Hipposideridae	<i>Aselliscus</i>	1	no
Chiroptera	Hipposideridae	<i>Hipposideros</i>	14	no
Chiroptera	Miniopteridae	<i>Miniopterus</i>	12	no
Chiroptera	Molossidae	<i>Chaerephon</i>	2	no
Chiroptera	Molossidae	<i>Mormopterus</i>	6	no
Chiroptera	Pteropodidae	<i>Acerodon</i>	2	no
Chiroptera	Pteropodidae	<i>Aethalops</i>	1	no
Chiroptera	Pteropodidae	<i>Aproteles</i>	1	no
Chiroptera	Pteropodidae	<i>Chironax</i>	1	no
Chiroptera	Pteropodidae	<i>Cynopterus</i>	2	no
Chiroptera	Pteropodidae	<i>Dobsonia</i>	6	no
Chiroptera	Pteropodidae	<i>Eonycteris</i>	1	no
Chiroptera	Pteropodidae	<i>Macroglossus</i>	3	no
Chiroptera	Pteropodidae	<i>Nyctimene</i>	4	no
Chiroptera	Pteropodidae	<i>Paranyctimene</i>	2	no
Chiroptera	Pteropodidae	<i>Pteropus</i>	6	no
Chiroptera	Pteropodidae	<i>Rousettus</i>	5	no
Chiroptera	Pteropodidae	<i>Syconycteris</i>	4	no
Chiroptera	Rhinolophidae	<i>Rhinolophus</i>	8	no
Chiroptera	Vespertilionidae	<i>Arielulus</i>	1	no
Chiroptera	Vespertilionidae	<i>Chalinolobus</i>	6	no
Chiroptera	Vespertilionidae	<i>Hypsugo</i>	2	no
Chiroptera	Vespertilionidae	<i>Kerivoula</i>	5	no
Chiroptera	Vespertilionidae	<i>Murina</i>	3	no
Chiroptera	Vespertilionidae	<i>Myotis</i>	5	no
Chiroptera	Vespertilionidae	<i>Nyctophilus</i>	8	no
Chiroptera	Vespertilionidae	<i>Philetor</i>	1	no
Chiroptera	Vespertilionidae	<i>Phoniscus</i>	1	no
Chiroptera	Vespertilionidae	<i>Pipistrellus</i>	16	no
Chiroptera	Vespertilionidae	<i>Scoteanax</i>	1	no
Chiroptera	Vespertilionidae	<i>Scotophilus</i>	1	no
Chiroptera	Vespertilionidae	<i>Scotorepens</i>	6	no
Chiroptera	Vespertilionidae	<i>Vespadelus</i>	8	no
Rodentia	Muridae	<i>Hydromys</i>	9	no
Rodentia	Muridae	<i>Melomys</i>	38	7 positive
Rodentia	Muridae	<i>Rattus</i>	75	2 positive

PCR screening for GALV and KoRV-like viruses resulted in detection of amplicon for five *M. leucogaster* ( $n = 10$ ), two *M. burtoni* ( $n = 7$ ), one *R. verecundus* ( $n = 5$ ) and a *R. niobe\_sp.B* ( $n = 5$ ). Sample details are described in [Dataset S1](#).

capture enrichment. The target enrichment factor for each sample was calculated from the VIP (Virus Integrated Pipeline) coverage information of gammaretroviruses output plot ([SI Appendix, Table S1](#)).

The unique retroviral sequences identified here were assembled into contigs (refer to [Materials and Methods](#)) and aligned to all the KoRVs, GALVs (including MbERV and MelWMV), Australian and Asian GALV-like bat sequences, and the related gammaretroviruses ([SI Appendix, Table S2](#)). These alignments were used to perform

phylogenetic analysis to infer the evolutionary relationships among the viral sequences. Contigs from *R. niobe\_Sp.B* (89) and *R. verecundus* (246) formed a clade with *R. norvegicus* (LOC102557044), while partial sequence retrieved from one of the *M. burtoni* (201, contig-4 ~ 1,300 bp) grouped with viral outgroup sequences from the rodents *C. griseus* and *M. coucha*. The remaining *Melomys* consensus contigs formed a clade with WMV, while the Asian HIGRV and RhGRV formed a sister clade (Fig. 2). As described in Hayward et al. (22), we applied Gblocks to the full-genome alignments,



**Fig. 2.** The maximum likelihood phylogenetic relationship of cMWMV inferred from complete genomic nucleotide sequences of 46 gammaretroviruses. Node support was assessed by 1,000 rapid bootstrap pseudoreplicates and is indicated at each node. The newick file is visualized with the Interactive Tree Of Life (iTOL) v5 (32). Branch length is with an average of 0.6 nucleotide substitutions per site. The avian reticuloendotheliosis virus (REV) was used as an outgroup and the sequences used for alignments and phylogenetic analysis are listed in *SI Appendix, Table S2*. Silhouettes represent the host species. The viral contigs identified in this study are marked with a red asterisk. The cMWMV clade is marked with blue and *M. burtoni* 204 represents MelWMV-NG which is displayed in orange.

eliminating the divergent and poorly aligned regions and further compared it to our initial alignment. Tree topology was largely congruent for the full-genome alignment, individual genes, and the amino acid sequences (*SI Appendix, Fig. S2*). Thus, we conclude that the poor node support (ranging from 30 to 58) for the WMV-HIGRV-RhGRV clade is due to the low sequence diversity of these sequences rather than phylogenetic approaches employed.

**Characterization of Viral Integration Flanking Sites.** We extended the viral sequences into the host genome integration site for samples 249, 291, and 300. Identical host flanking sequences were found each with the same target duplication site for one integration site (*SI Appendix, Text file S1 and Fig. S3A*). The data suggest that there is only one cMWMV shared integration site in all of the samples for which integration site flanking sequence could be identified. The results were derived from different tissue samples of different *M.*

*leucogaster* that were collected in 1985, 1987, and 2014 from four different collection sites and two different provinces in Papua New Guinea (PNG). An identical integration site in multiple tissues in multiple individuals can most parsimoniously be explained by vertical transmission indicating the WMV-like sequence is an endogenous retrovirus (ERV). Flanking sequences could not be extended for one of the *M. leucogaster* (290) and *M. burtoni* (201 and 204) and therefore it could not be determined whether these viral sequences are ERVs or XRVs (exogenous retroviruses). However, the greater divergence of the *M. burtoni* sequence from *M. leucogaster* sequences, the closer relationship of the WMV-like sequence obtained from *M. burtoni* to the Indonesian MelWMV, and the observation that the MelWMV integration site is different from that of cMWMV in *M. leucogaster* (*SI Appendix, Fig. S3B and Text file S1*) suggests that the *M. burtoni* sequence obtained is not cMWMV but a related retrovirus like the defective *M. burtoni*



viruses MbRV and MelWMV. We termed this MelWMV variant MelWMV-NG (MelWMV New Guinea).

**Structural Characteristics of cMWMV.** The cMWMV has retained the typical gammaretroviral structure with a genome of ~ 8.5 kb and unlike MelWMV has a potentially functional *env* gene. The coding region is flanked by 5' and 3' untranslated LTRs (*SI Appendix, Fig. S4*). The conserved CETTG motif often found in highly infectious gammaretroviruses was identified in cMWMV (*SI Appendix, Fig. S5*). Further, cMWMV exhibited intact structural polyproteins (GAG, ENV) and functional POL with 97.3% pairwise identity to WMV and 57.6% to KoRV. The latter finding indicates a closer phylogenetic relationship of cMWMV to WMV. This outcome is consistent with the results from the BLAST search, multiple sequence alignments, and phylogenetic topology compared to KoRV-A.

Receptor binding mediated by the ENV is vital to the viral cellular entry process and the determinant factor in viral tropism. To predict the functional effects of the identified mutations, a computational strategy with five major criteria (*SI Appendix, Methods*) was used. Our prediction was that the probability of a mutation to have functional consequences increases with the number of criteria fulfilled. The results of these analyses, which are summarized in *SI Appendix, Table S3*, strongly suggest that the identified mutations most probably do not functionally alter the proteins.

Specifically, out of the 46 differences identified, only 10 are predicted to be radical substitutions according to both BLOSUM 65 and 80 scores. This finding suggests that only a few mutations are physico-chemically different enough to suggest functional changes. Furthermore, only four mutations were found in a highly conserved amino acid position based on conservation level analysis of each position. This finding also suggests that the positions of these mutations are highly variable, thus, they might have a lower probability of being functionally important. Additionally, the two different prediction algorithms (SIFT and PROVEAN) that were used found that only eight substitutions are suggested to alter protein function with only two of them predicted with high confidence by both algorithms. Analysis to determine whether any of the mutations change an amino acid of known function, suggested that only 12 mutations were found in known functional motifs, but none was predicted to alter the amino acid properties of the motif. Lastly, homology modeling of parts of the proteins was to test whether a mutation alters the protein local conformation. The latter analysis revealed that only four of the identified mutations might affect local topology (*SI Appendix, Fig. S6 and Table S3*). Although a few of the identified mutations were predicted to alter some characteristics of the protein, none of them fulfilled more than three of the above-mentioned criteria, supporting the notion that these mutations most probably do not drastically alter protein function. Only two changes each were observed in the VRA and VRB region of the ENV protein relative to WMV (*SI Appendix, Fig. S5*). Our ENV modeling suggests that like WMV (33), cMWMV may employ PiT-1 cellular receptors. However, we cannot exclude that these changes or changes in other viral domains affect receptor usage by cMWMV.

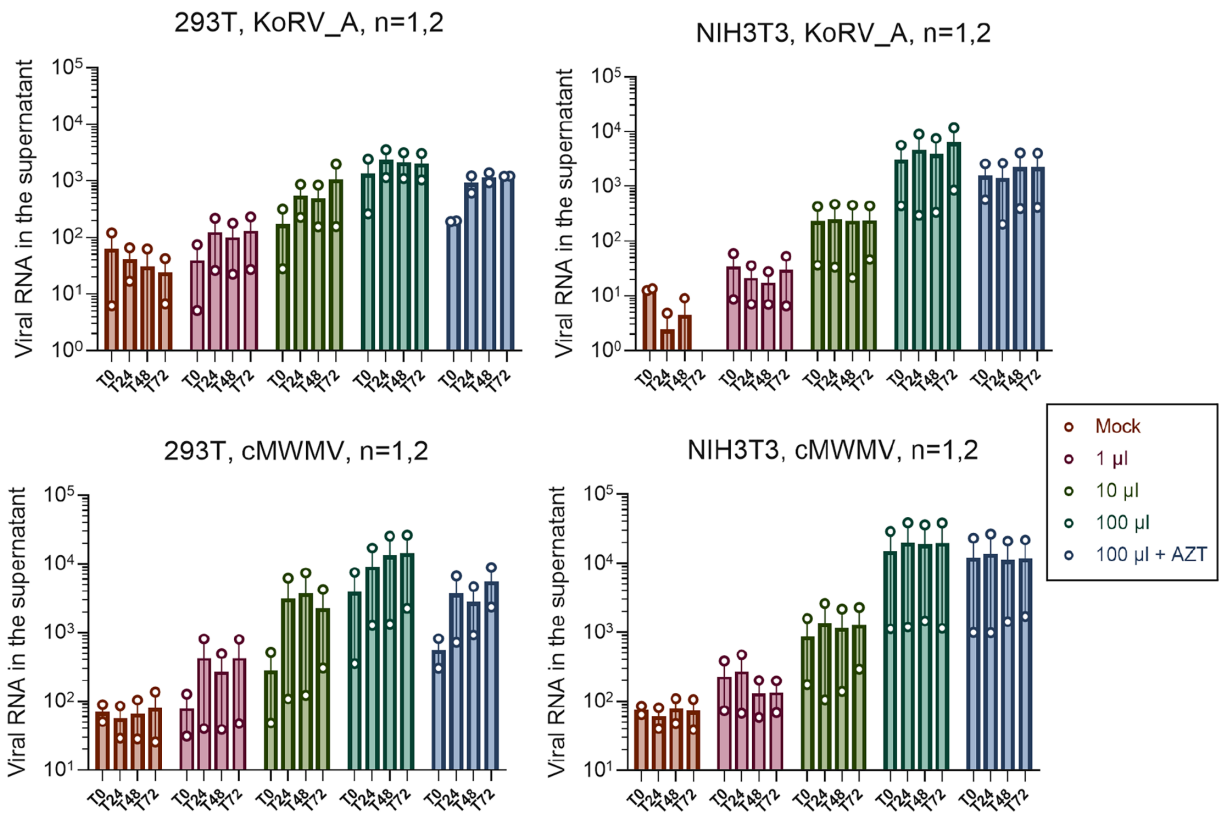
To evaluate the computational predictions, we performed cell culture experiments employing viral vectors for cMWMV and KoRV-A as a control by infecting NIH3T3 and HEK293T cells in triplicates and culturing over 5 d. cMWMV and KoRV-A (Fig. 3) viral kinetics as measured by qPCR of viral RNA in the supernatant demonstrated weak increase in viral titer. However, upon 24 hours post infection (hpi), representing the post-wash time point, production of viral copies increased by twofold to threefold at 48 hpi and peaked at 72 hpi resulting in a threefold to fivefold increase for all virus inocula. These data suggest that cMWMV and KoRV-A are

replication-competent in HEK293T cells (Fig. 3). In NIH3T3 cells, increase of viral RNA was documented 48 hpi (Fig. 3). Furthermore, infection in presence of the reverse-transcriptase inhibitor azidothymidine (AZT) effectively diminished viral RNA production (280-fold for cMWMV and twofold for KoRV-A) in HEK293T cells and NIH3T3 cells for both viruses, confirming de novo synthesis of viral DNA at indicated time points in the absence of AZT (Fig. 3). Recognizing that other studies have observed NIH3T3 cell resistance to KoRV-A infection (22), we confirmed replication through a variety of additional experiments (*SI Appendix, Methods*). First, the experiments were performed independently with two NIH3T3 cell lines, one cultured in the lab of the authors (Charité) for years and one a newly obtained culture from the American Tissue Type Collection (ATCC). Increased viral expression was detected in both cell culture batches (ATCC results for NIH3T3 shown in Fig. 3). A and B domains of PiT-1, which are essential for virus entry (34, 35), were sequenced to determine whether available NIH3T3 cultures have diverged in sequence. However, the sanger sequences of our cultures were identical to the reported PiT-1 *Mus musculus* mRNA sequence (M73696.1) (*SI Appendix, Fig. S8*). Finally, we used a long fragment inverse PCR fragment approach coupled with PacBio sequencing to determine whether and where integration sites occurred for cMWMV and KoRV infected HEK293T and NIH3T3 cell lines. Evidence of retroviral integration would support that viral particles had infected the cells, reverse transcribed, and integrated over the course of the cell culture infection experiments. Multiple integration sites were identified for both viruses for all cell lines demonstrating that reverse transcription and viral integration had occurred demonstrating retroviral replication had taken place (*Dataset S2*). The infection observed for NIH3T3 cells in contrast to similar studies using viral expression constructs may be attributable to the sensitivity of the TaqMan-based RT-PCR approach employed in the current study which can detect even minute amounts of expression.

Immunostaining of NIH3T3 cells with antibodies against PiT-1 and PiT-2 cellular receptors suggests that both receptors are present (*SI Appendix, Fig. S7*). We cannot therefore determine whether cMWMV exclusively binds to PiT-1 or is able to use PiT-2 cellular receptors as well to infect mouse cells such as NIH3T3 cells. Given that single amino acid changes can have major functional consequences for virus-receptor interaction (36), we cannot exclude that cMWMV can additionally or exclusively use another receptor entirely. Electron microscopy (EM) of cMWMV in both human HEK293T and mouse NIH3T3 cell lines revealed an electron-dense polygonal core which is enclosed by a spherical envelope (Fig. 4A and *SI Appendix, Fig. S9*). The particle morphology is similar to the morphology of KoRV-A particles propagated in the same cell lines (*SI Appendix, Fig. S9 A and D*) and corresponds to the C-type morphology of retroviruses (37, 38). We rarely found budding structures at the surface of the cells (Fig. 4B) which indicate the typical virus particle assembly observed with this morpho-type (37, 38). However, the density below the membrane of the budding site is not separated by a clearly visible gap from the membrane like in typical C-type viruses and appears more similar to the budding in Human T-lymphotropic virus 1 or Bovine leukemia virus, which share a similar particle morphology than C-type viruses (37, 38). Overall, our results suggest that like exogenous GALVs, cMWMV is capable of completing a retrovirus life cycle and forming infectious virions (Figs. 3 and 4).

## Discussion

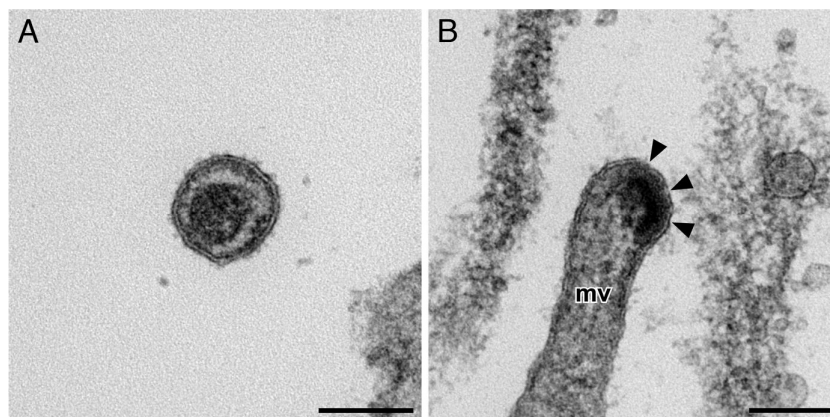
The cMWMV was detected in *M. leucogaster* in a subset of provinces (Southern Highland and Western Provinces) within its distribution in PNG but was not detected in Chimbu, Gulf, and Sandaun



**Fig. 3.** Replication kinetics of KoRV-A (*Top*) and cMWMV (*Bottom*) in the mouse NIH3T3 (obtained from ATCC) and human HEK293T cells, based on two biological replicates. Both cell lines were either treated with 10  $\mu$ M of the reverse-transcriptase inhibitor AZT prior to infection with 100  $\mu$ L of virus-suspension or infected with increasing volume of virus-suspension without AZT pre-treatment. At indicated time points, virus-containing supernatant was harvested for cDNA synthesis, following qPCR. De novo synthesis of viral DNA was measured and concentrations of the samples were calculated using standards of known DNA concentrations.

Provinces. Identical integration sites were detected among multiple tissues and multiple individuals from the same region which strongly indicates the virus is endogenous and that these individuals inherited the virus as a genomic locus. The absence of cMWMV from other populations indicates that it is not yet fixed in *M. leucogaster* and unlike MelWMV, has managed to retain intact open reading frames (ORFs) for all protein coding and non-coding viral sequences. In addition to retaining intact ORFs, cMWMV-encoded polypeptides that likely maintained their original functions. This is supported by the minimal alterations in the predicted function and structure of the GAG, POL, and ENV proteins because none of the identified mutations are predicted to result in functional

changes between the cMWMV and WMV. This high level of structural and functional conservation could suggest that cMWMV, similar to WMV, might employ the ubiquitous sodium-dependent phosphate transporter (PiT-1, also known as SLC20A1) protein as cellular receptor. While we were unable to determine whether cMWMV uses PiT-1 exclusively or can also use PiT-2 as a receptor, its ability to infect both NIH3T3 and HEK293T cells productively, and to form viral particles, and bud from the cell membrane suggest that this ERV is potentially infectious in vivo. The conservation of the PiT-1 receptor across mammals (39) could also explain why such diverse species including, rodents, bats, primates, and marsupials have been infected by relatives of this viral clade.



**Fig. 4.** EM of thin sections through cMWMV-infected NIH3T3 cells. (A) A single cMWMV particle in the extracellular space close to a cell surface. The particle is limited by a distinct bi-layered bio-membrane, which is studded by dot-like outer surface structures and underlaid by a thin dense matrix layer. The core (=capsid) is dense and polygonal, like in other gammaretrovirus-like particles. (B) Budding of a virus particle at the tip of a microvillus (*mv*) with the characteristic semi-annular density (*arrowheads*) below the plasma membrane. (Scale bars, 100 nm.)

MelWMV is an endogenized WMV in an isolated *M. burtoni* subspecies (Halmahera) in Indonesia. It has large deletions in the *env* and *pol* genes suggesting that it is no longer capable of producing viral particles nor re-integrating in the host genome without a helper virus. MbrRV is a WMV-like sequence identified in *M. burtoni* in Australia. However, neither the full genome nor its status as an ERV has been determined. We detected a MelWMV-like virus in *M. burtoni* samples from Australia. However, it lacked the large-scale deletions of MelWMV suggesting that both it and MbrRV represent at least three distinct integrations into the *M. burtoni* genome. The closer relationship of MbrRV, MelWMV, and the *M. burtoni* sequence identified in this study and the observation that the MelWMV integration site is different from that of cMWMV in *M. leucogaster* suggest that the three *M. burtoni* viruses represent distinct colonization events from that of cMWMV in *M. leucogaster*.

Broad-scale phylogenomics indicates that 53% of all gammaretroviral-derived ERVs come from rodents (4), suggesting that rodents have transmitted XRVs and their integrated counterparts among mammals for millions of years. In laboratory mice, ongoing colonization continues to contribute to endogenization, whereas in most other species, this process was completed millions of years ago (5). The frequency of cross-species transmission and that rodent-derived ERVs are not monophyletic, suggests that rodents may still be a source of novel endogenizing retroviruses both among rodents and non-rodent mammals. GALV is thought to have been iatrogenically transmitted to captive gibbons as a result of experimental contamination with human material from New Guinea (11, 40). Most viruses ancestral to GALV-KoRV derive from murine rodents such as Asian *Mus caroli* ERV (41), and the frequency of GALV-KoRV ERVs and XRVs in Australo-Papuan rodents suggests an overall rodent origin for this viral group.

In contrast, several bat families such as Hipposideridae (23), Pteropodidae, and Rhinolophidae (24) are documented to have crossed the Wallace Line but only few GALV-like retroviral sequences have been reported in these species. In Queensland, Australia, two *P. alecto* (21, 22) were found to harbor a GALV-like retrovirus though the same species was negative for GALV and KoRV relatives in PNG samples tested in our study. The seven Australian bat species (*M. minimus*, *P. alecto*, *P. conspicillatus*, *P. macrotis*, *P. poliocephalus*, *P. scapulatus*, and *P. vampyrus*) tested by Simmons et al. (2014) did not yield any GALV or KoRV-like gammaretrovirus. We found no evidence for any GALV-like sequences in bats from Australia, Indonesia, Laos, PNG, or Timor-Leste ( $n = 156$ ). While rodents in the Australo-Papuan region regularly have detectable GALV relatives, bats rarely carry such sequences and likely as a result of independent cross-species transmission, such as *R. ferrumequinum* retrovirus which may have a tree-shrew origin (42). HIGRV and RhGRV were isolated from the pooled fecal and pharyngeal samples from Chinese *H. larvatus* and *R. hipposideros* bats such that the prevalence of the viruses is unclear (22). All 156 bat samples (ca. 120 species) in the current study, including 14 Australo-Papuan *Hipposideros* and eight Australo-Papuan *Rhinolophus*, tested negative for GALV or KoRV-related viruses. Thus, WMVs in *Melomys* rodents are common but in bats exclusively exogenous and are sporadically detected, suggesting recent transmission.

Based on the phylogenetic analysis, bats and rodents are host to viruses in both basal and crown positions within the GALV-KoRV clade. Several bat sequences were successive sister lineages to the GALV-KoRV clade while the ancestral sequences of the GALV-KoRV clade are associated with rodent hosts. This interpretation is consistent with the evolutionary history of the entire gammaretroviral group, which shows a transition from rodent to non-rodent lineages (4). While viruses identified in bats are GALV related and those previously identified in *M. burtoni* are degraded ERVs, cMWMV

is a derived, completely intact WMV with 98.9% nucleotide identity to WMV. The phylogeny indicates that all the GALVs either represent derived WMVs or a clade that recently split from WMV. KoRV represents an older lineage and its exogenous counterpart may no longer exist as KoRV began colonizing the koala genome at least 50,000 y ago (43). Why no other taxa within the region carry sequences with higher KoRV identity is unclear, a question that could be answered by further taxa screening. This outcome may have to do either with chance that led to lack of germline invasion in the original host, or some unknown aspect of WMV biology that allows for a more opportunistic endogenization in rodents than other members of the GALV or KoRV clades. It should be noted that applying various alignment and phylogenetic approaches failed to adequately resolve the relationships between RhGRV-HIGRV bat-derived clade and WMV-GALV clade. The high sequence identity among viruses across the genome makes phylogenetic resolution difficult, emphasizing the very close relationship and minimal divergence among these viral sequences (SI Appendix, Fig. S4).

In conclusion, multiple germline colonizing retroviruses have been detected in mammals in the Australo-Papuan region, a very rarely observed process outside this region. In particular, endogenous WMVs have been found frequently among *Melomys* species across their biogeographical distribution. In all cases, these viruses have endogenized in their rodent hosts but only regionally. Our finding suggests that like KoRV in its koala host, the endogenization process in *M. leucogaster* is at the earliest stages, providing an additional wildlife model of the complex process of germline colonization by exogenous retroviruses. Bats show a more regionally distinct and discontinuous prevalence even within the same species and may only be sporadically infected by contact with rodent reservoirs. There is no evidence of endogenization of GALVs in bats. Nonetheless, GALV-like viruses have been transmitted as far from the Wallace Line as Southeastern China and New Guinea. GALV-like viruses appear to be circulating, evolving, and endogenizing in the endemic New Guinean rodent population and occasionally transmitting to other vertebrates resulting in viruses that endogenize in non-rodent mammals such as KoRV in koalas. The biodiversity within New Guinea is immense including within *Melomys*, which contains many defined species ( $n = 23$ ), taxonomically uncharacterized populations, and close relatives in the *Uromys* division (7) most of which have yet to be screened for viruses. Our results suggest that the region will be of particular interest for further identifying germline integration events and assessing the limits to which the Wallace Line prevents viral spill-over into Southeast Asia and beyond.

## Materials and Methods

**Samples and DNA Extraction.** A total of 278 rodent ( $n = 122$ ) and bat ( $n = 156$ ) samples from the South Australian Museum (SAM) were analyzed. These samples were collected between 1981 and 2017 and represent seven bat families from 37 genera and ca. 120 species, three rodent genera from the family Muridae, representing ca. 38 species with six of them found on both sides of the Wallace Line (details in the SI Appendix). DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Germany) according to the manufacturer's protocol for frozen or ethanol-preserved blood, hair, and tissue samples.

**GALV-KoRV PCR Screening.** Degenerate primer set KOGAWM-1F 5'-CCCCTAATCGACCTCASTGG-3' and KOGAWM-1R 5'-RTATCTCTATARGCCTCCAT-3' (product size ~200 bp) were used to screen samples for GALV and KoRV-related retroviruses by PCR and visualized using electrophoresis. Details are in SI Appendix.

**Illumina Library Construction and Target Enrichment Hybridization Capture and Sequencing.** Illumina libraries were generated using standard methods and were hybridized with customized 70-mer biotinylated



oligonucleotide meta-viral-baits (probes). Sequencing of enriched libraries was performed with the Illumina MiSeq platform with v2 reagent kit. Details are described in *SI Appendix*.

**Bioinformatics Analysis and Virus Classifications.** The raw sequencing reads were demultiplexed, adaptor sequences, low-quality reads (quality cutoff 20 and minimum read length of 30 nt), and duplicates were removed and merged using Cutadapt v1.15 (44), Trimmomatic v0.27 (45), Picard v1.4 (<http://broadinstitute.github.io/picard>), and BBMerge (46), respectively. Two pipelines were applied for the identification and assembly of viral reads (*SI Appendix, Fig. S1*): VIP (47) in sense mode and Genome Detective (48), a web-based bioinformatics pipeline (further details are found in the *SI Appendix*).

**Phylogenetic Analysis and Integration Site Mapping.** Multiple nucleotide alignments of the consensus sequences with thirty-seven genome sequences of gammaretroviruses (*SI Appendix, Table S2*) were performed using default settings in MUSCLE (49). Statistical selection of the best-fit model for the phylogenetic analysis performed using jModelTest (50). Bayesian phylogenetic inference was produced using Markov Chain Monte Carlo for 1,000,000 iterations in MrBayes v3.2.7 (51). A Maximum likelihood (ML) tree was constructed with rapid bootstrapping (1,000 replicates) and GTRGAMMA substitution rate in Randomized Accelerated ML (RAxML v8.2.11) (52). Retroviral flanking read sequences were mapped using Geneious (further details are described in the *SI Appendix*).

**Retroviral Protein Structure Modeling.** The structure characteristics of the cMWMV viral genome were examined in comparison to the WMV genome. SWISS-MODEL server (53) was used for the prediction of the three-dimensional structures of WMV and cMWMV (249, a representative sequence with high sequence coverage). From the output structures predicted, only high-quality protein models as defined by QMEAN4 (54) values were considered for further analysis. Both WMV and cMWMV genomes produced high-quality structures in various domains for all three viral polypeptides (GAG, POL, and ENV). Pairwise structural alignment, superimposition, and figure design were performed using PyMol v2.4 (55, 56). Further details are described in *SI Appendix*.

**Virus Production and Immunofluorescence Microscopy of Cells.** The 8,459 bp construct for cMWMV (*SI Appendix, Text file S2*) was based on a majority consensus sequence of all individual identified consensus sequences (shown in Fig. 1) from *M. leucogaster* and the sequence identified from *M. burtoni*. While minor sequence variation was observed among the consensus sequences used to generate the overall consensus sequence, none altered any amino acids and therefore, the functional results resulting from the use of the overall consensus sequence should be representative for the individual isolates. The cMWMV and KoRV-A (AB721500) genomes were chemically synthesized and sub-cloned in pUC57 vector (GenScript, China). These constructs were used to transfect NIH Swiss mouse embryonic fibroblasts (NIH3T3) and Human embryonic kidney (HEK293T) cells (*SI Appendix*). To determine whether tropism of cMWMV is comparable to GALVs or, similar to WMV, is restricted to PiT-1, HEK293T and NIH3T3 cells were immunostained for PiT-1 and PiT-2 proteins as described in *SI Appendix*.

**Taqman RT-qPCR.** We designed primers and fluorescent probes on *pol* gene of cMWMV and *env* gene of KoRV-A. TaqMan primers and probes with 5'-6-FAM and 3'-BBO650 modifications were synthesized (Biomers, Germany). Viral RNA extraction and complementary DNA constructions were performed as described in *SI Appendix*. Quantification of absolute viral copies was performed with the LightCycler 480 II system (Roche, Germany). Previous cell culture infection studies of KoRV-A have suggested that NIH3T3 cells are resistant to infection (57, 58). However, we hypothesize that the Taqman RT-qPCR assay is more sensitive than the previously employed reporter systems used, which may explain why we detected weak replication of KoRV-A in NIH3T3 cells.

1. L. Bromham, The human zoo: Endogenous retroviruses in the human genome. *Trends Ecol. Evol.* **17**, 91–97 (2002).
2. J. P. Stoye, Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat. Rev. Microbiol.* **10**, 395–406 (2012).
3. F. K. Geis, S. P. Goff, Silencing and transcriptional regulation of endogenous retroviruses: An overview. *Viruses* **12**, 884 (2020).

**Thin-Section EM.** HEK293T and NIH3T3 cells were infected with indicated volumes of cMWMV- and KoRV-A-containing supernatants (*SI Appendix*). At 48 hpi, cells were fixed as described in *SI Appendix*. Thin-section microscopy and image processing was performed at the Laboratory for Diagnostic EM of Infectious Pathogens Robert Koch-Institut.

**PiT-1 Sequencing from NIH3T3 Cells.** We designed primers based on *Mus musculus* mRNA of PiT-1 (GenBank accession M73696.1) and amplified motifs A and B of the receptor as indicated in *SI Appendix*.

**DNA Sonication Inverse PCR to Identify cMWMV and KoRV-A Integration Sites in HEK293T and NIH3T3 Cell Line Infection Experiments.** Sonication-based inverse PCR, followed by PacBio sequencing was conducted as described in *SI Appendix* to identify virus integration sites.

**Data, Materials, and Software Availability.** All data are included in the main text and the *SI Appendix*. Raw sequence data are deposited to SRA and are accessible with BioProject [PRJNA870749](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA870749) (59). Sequences of cMWMV have been deposited in the GenBank database with accession [ON903268](https://www.ncbi.nlm.nih.gov/nuccore/ON903268) (isolate 249) (60), [OP921767](https://www.ncbi.nlm.nih.gov/nuccore/OP921767) (isolate 290) (<https://www.ncbi.nlm.nih.gov/nuccore/OP921767>) (61), [OP921768](https://www.ncbi.nlm.nih.gov/nuccore/OP921768) (isolate 291) (<https://www.ncbi.nlm.nih.gov/nuccore/OP921768>) (62), [OP921769](https://www.ncbi.nlm.nih.gov/nuccore/OP921769) (isolate 292) (<https://www.ncbi.nlm.nih.gov/nuccore/OP921769>) (63), [OP921770](https://www.ncbi.nlm.nih.gov/nuccore/OP921770) (isolate 300) (<https://www.ncbi.nlm.nih.gov/nuccore/OP921770>) (64), and [OP921771](https://www.ncbi.nlm.nih.gov/nuccore/OP921771) which represents MelWMV-NG (isolate 204) (<https://www.ncbi.nlm.nih.gov/nuccore/OP921771>) (65). BioSample accessions [SAMN37868416](https://www.ncbi.nlm.nih.gov/biosample/SAMN37868416) (<https://www.ncbi.nlm.nih.gov/biosample/SAMN37868416>) (66), [SAMN37868417](https://www.ncbi.nlm.nih.gov/biosample/SAMN37868417) (<https://www.ncbi.nlm.nih.gov/biosample/SAMN37868417>) (67), [SAMN37868418](https://www.ncbi.nlm.nih.gov/biosample/SAMN37868418) (<https://www.ncbi.nlm.nih.gov/biosample/SAMN37868418>) (68), [SAMN37868419](https://www.ncbi.nlm.nih.gov/biosample/SAMN37868419) (<https://www.ncbi.nlm.nih.gov/biosample/SAMN37868419>) (69), linked to the same BioProject shows PacBio sequences of HEK293T cells infected with cMWMV, NIH3T3 cells infected with cMWMV, HEK293T cells infected with KoRV-A and NIH3T3 cells infected with KoRV-A, respectively, which were used to identify the integration sites of the viruses. The alignment for Fig. 2 is accessible at [10.6084/m9.figshare.21704216](https://www.figshare.com/figure/21704216) (70). The consensus sequence used for in vitro experiments is provided in *SI Appendix, Text file S2* along with the retrieved flanking sequences (*SI Appendix, Text file S1*).

**ACKNOWLEDGMENTS.** This project was supported by Grant 3924-12-1 from Deutsche Forschungsgemeinschaft. Much of the extensive sample set that was used in this project was collected over the years by the late Dr. K.P.A. Lars Möller at Robert Koch Institute performed most of the EM of thin sections. The essential KoRV positive controls were provided by Dr. Farhid Hemmatzadeh and Dr. Tamsyn Stephenson from School of Animal and Veterinary Science at University of Adelaide. We thank Jenny Jansen at Charité-Universitätsmedizin Berlin for excellent technical assistance, Badru Mugerwa at Leibniz Institute for Zoo and Wildlife Research and Tony Daubiné for their help in illustrating the map and vector base figures.

Author affiliations: <sup>a</sup>Department of Wildlife Diseases, Leibniz Institute for Zoo and Wildlife Research, Berlin 10315, Germany; <sup>b</sup>Centre International de Recherche en Infectiologie, Université de Lyon, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, École Nationale Supérieure de Lyon, Lyon F-69007, France; <sup>c</sup>Institute of Virology Charité-Universitätsmedizin Berlin, Berlin D-10117, Germany; <sup>d</sup>Department of Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool L3 5QA, United Kingdom; <sup>e</sup>Department of Life and Health Sciences, University of Nicosia, Nicosia CY-2417, Cyprus; <sup>f</sup>Department of Biological Science, Center for Applied Biotechnology Studies, and Center for Computational and Applied Mathematics, College of Natural Sciences and Mathematics, California State University Fullerton, Fullerton, CA 92834-6850; <sup>g</sup>Advanced Light and Electron Microscopy (ZBS 4), Centre for Biological Threats and Special Pathogens, Robert Koch Institute, Berlin D-13353, Germany; <sup>h</sup>Max-Delbrück Center for Molecular Medicine Helmholtz Association, Berlin 13125, Germany; <sup>i</sup>South Australian Museum, North Terrace, Adelaide SA 5000, Australia; <sup>j</sup>Sciences Department, Museums Victoria, Melbourne, VIC 3001, Australia; and <sup>k</sup>School of Veterinary Medicine, Freie Universität Berlin, Berlin 14163, Germany

4. A. Hayward, M. Grabherr, P. Jern, Broad-scale phylogenomics provides insights into retrovirus-host evolution. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20146–20151 (2013).
5. C. Stocking, C. A. Kozak, Murine endogenous retroviruses. *Cell. Mol. Life Sci.* **65**, 3383–3398 (2008).
6. J. R. Ali, L. R. Heaney, Wallace's line, Wallacea, and associated divides and areas: History of a tortuous tangle of ideas and labels. *Biol. Rev. Camb. Philos. Soc.* **96**, 922–942 (2021), 10.1111/brv.12683.



7. K. C. Rowe *et al.*, Oceanic islands of Wallacea as a source for dispersal and diversification of murine rodents. *J. Biogeography* **46**, 2752–2768 (2019).
8. T. G. Kawakami, P. M. Buckley, Antigenic studies on gibbon type-C viruses. *Transplant. Proc.* **6**, 193–196 (1974).
9. T. G. Kawakami, G. V. Kollias Jr., C. Holmberg, Oncogenicity of gibbon type-C myelogenous leukemia virus. *Int. J. Cancer* **25**, 641–646 (1980).
10. R. C. Gallo *et al.*, Isolation and tissue distribution of type-C virus and viral components from a gibbon ape (Hylobates lar) with lymphocytic leukemia. *Virology* **84**, 359–373 (1978).
11. G. J. Todaro *et al.*, Infectious primate type C viruses: Three isolates belonging to a new subgroup from the brains of normal gibbons. *Virology* **67**, 335–343 (1975).
12. S. P. Snyder, D. L. Dungworth, T. G. Kawakami, E. Callaway, D. T. Lau, Lymphosarcomas in two gibbons (Hylobates lar) with associated C-type virus. *J. Natl. Cancer Inst.* **51**, 89–94 (1973).
13. I. Parent *et al.*, Characterization of a C-type retrovirus isolated from an HIV infected cell line: Complete nucleotide sequence. *Arch. Virol.* **143**, 1077–1092 (1998).
14. G. Burtonboy, N. Delferriere, B. Moussef, M. Heusterspreute, Isolation of a C-type retrovirus from an HIV infected cell line. *Arch. Virol.* **130**, 289–300 (1993).
15. G. H. Theilen, D. Gould, M. Fowler, D. L. Dungworth, C-type virus in tumor tissue of a woolly monkey (Lagothrix spp.) with fibrosarcoma. *J. Natl. Cancer Inst.* **47**, 881–889 (1971).
16. L. G. Wolfe, R. K. Smith, F. Deinhardt, Simian sarcoma virus, type 1 (Lagothrix): Focus assay and demonstration of nontransforming associated virus. *J. Natl. Cancer Inst.* **48**, 1905–1908 (1972).
17. G. Simmons, D. Clarke, J. McKee, P. Young, J. Meers, Discovery of a novel retrovirus sequence in an Australian native rodent (Melomys burtoni): A Putative Link between gibbon ape leukemia virus and Koala Retrovirus. *PLoS One* **9**, e106954 (2014).
18. B. Michaux, Biogeology of Wallacea: Geotectonic models, areas of endemism, and natural biogeographical units. *Biol. J. Linn. Soc.* **101**, 193–212 (2010).
19. N. Alfano *et al.*, Endogenous Gibbon Ape Leukemia Virus identified in a rodent (Melomys burtoni subsp.) from Wallacea (Indonesia). *J. Virol.* **90**, 8169–8180 (2016).
20. J. L. Siegal-Willott *et al.*, Evaluation of captive gibbons (Hylobates spp., Nomascus spp., Symphalangus spp.) in North American Zoological Institutions for Gibbon Ape Leukemia Virus (GALV). *J. Zoo Wildl. Med.* **46**, 27–33 (2015).
21. L. McMichael *et al.*, A novel Australian flying-fox retrovirus shares an evolutionary ancestor with Koala, Gibbon and Melomys gamma-retroviruses. *Virus Genes* **55**, 421–424 (2019).
22. J. A. Hayward *et al.*, Infectious KoRV-related retroviruses circulating in Australian bats. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9529–9536 (2020).
23. S. W. Murray *et al.*, Molecular phylogeny of hipposiderid bats from Southeast Asia and evidence of cryptic diversity. *Mol. Phylogenet. Evol.* **62**, 597–611 (2012).
24. T. Kingston, S. J. Rossiter, Harmonic-hopping in Wallacea's bats. *Nature* **429**, 654–657 (2004).
25. R. M. Timm *et al.*, A new species of Rattus (Rodentia: Muridae) from Manus Island, Papua New Guinea. *J. Mammal.* **97**, 861–878 (2016).
26. M. D. Database, Mammal Diversity Database. (2020). <https://doi.org/10.5281/zenodo.4139818> (9 March 2021).
27. P.-H. Fabre *et al.*, New record of Melomys burtoni (Mammalia, Rodentia, Murinae) from Halmahera (North Moluccas, Indonesia): A review of Moluccan Melomys. *Mammalia* **82**, 218–247 (2018).
28. K. C. Rowe, M. L. Reno, D. M. Richmond, R. M. Adkins, S. J. Steppan, Pliocene colonization and adaptive radiations in Australia and New Guinea (Sahul): Multilocus systematics of the old endemic rodents (Muroidea: Murinae). *Mol. Phylogenet. Evol.* **47**, 84–101 (2008).
29. L. M. Bryant, S. C. Donnellan, D. A. Hurwood, S. J. Fuller, Phylogenetic relationships and divergence date estimates among Australo-Papuan mosaic-tailed rats from the Uromys division (Rodentia: Muridae). *Zoologica Scripta* **40**, 433–447 (2011).
30. E. Geffen, K. C. Rowe, Y. Yom-Tov, Reproductive rates in Australian rodents are related to phylogeny. *PLoS One* **6**, e19199 (2011).
31. T. F. Flannery, *Mammals of New Guinea* (Cornell University Press, 1995).
32. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
33. Y.-T. Ting, C. A. Wilson, K. B. Farrell, G. Jilani Chaudry, M. V. Eiden, Simian sarcoma-associated virus fails to infect Chinese hamster cells despite the presence of functional Gibbon Ape Leukemia Virus receptors. *J. Virol.* **72**, 9453–9458 (1998).
34. S. V. Johann, M. van Zeijl, J. Cekleniak, B. O'Hara, Definition of a domain of GLVR1 which is necessary for infection by gibbon ape leukemia virus and which is highly polymorphic between species. *J. Virol.* **67**, 6733–6736 (1993).
35. K. B. Farrell, J. L. Russ, R. K. Murthy, M. V. Eiden, Reassessing the role of region A in Pit1-mediated viral entry. *J. Virol.* **76**, 7683–7693 (2002).
36. M. Reinisová *et al.*, A single-amino-acid substitution in the Tvbs1 receptor results in decreased susceptibility to infection by avian sarcoma and leukosis virus subgroups B and D and resistance to infection by subgroup E in vitro and in vivo. *J. Virol.* **82**, 2097–2105 (2008).
37. M. V. Nermut, D. J. Hockley, Comparative morphology and structural classification of retroviruses. *Curr. Top. Microbiol. Immunol.* **214**, 1–24 (1996).
38. C. S. Goldsmith, Morphologic differentiation of viruses beyond the family level. *Viruses* **6**, 4902–4913 (2014).
39. L. Beck *et al.*, The phosphate transporter Pit1 (Slc20a1) revealed as a new essential gene for mouse liver development. *PLoS One* **5**, e9148 (2010).
40. K. Brown, R. E. Tarlinton, Is gibbon ape leukaemia virus still a threat? *Mammal Rev.* **47**, 53–61 (2017).
41. M. M. Lieber *et al.*, Isolation from the asian mouse Mus caroli of an endogenous type C virus related to infectious primate type C viruses. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2315–2319 (1975).
42. J. Cui, G. Tachedjian, L.-F. Wang, Bats and rodents shape mammalian retroviral phylogeny. *Sci. Rep.* **5**, 16561 (2015).
43. 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).
44. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* **17**, 10 (2011).
45. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
46. B. Bushnell, J. Rood, E. Singer, BBMerge—Accurate paired shotgun read merging via overlap. *PLoS One* **12**, e0185056 (2017).
47. Y. Li *et al.*, VIP: An integrated pipeline for metagenomics of virus identification and discovery. *Sci. Rep.* **6**, 23774 (2016).
48. M. Vilsker *et al.*, Genome detective: An automated system for virus identification from high-throughput sequencing data. *Bioinformatics* **35**, 871–873 (2018).
49. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
50. D. Posada, jModelTest: Phylogenetic model averaging. *Mol. Biol. Evol.* **25**, 1253–1256 (2008).
51. J. P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
52. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
53. A. Waterhouse *et al.*, SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
54. P. Benkert, M. Biasini, T. Schwede, Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343–350 (2011).
55. K. Tsangaras *et al.*, Hybridization capture reveals evolution and conservation across the entire Koala retrovirus genome. *PLoS One* **9**, e95633 (2014).
56. W. L. Delano, The PyMOL molecular graphics system. Schrödinger LLC (Version 2, 2002), [www.pymol.org](http://www.pymol.org).
57. T. Shojima *et al.*, Construction and characterization of an infectious molecular clone of Koala retrovirus. *J. Virol.* **87**, 5081–5088 (2013).
58. 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).
59. Leibniz Institute for Zoo and Wildlife Research, A recent gibbon ape leukemia virus germline integrations in a rodent from New Guinea. NCBI BioProject. <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA870749>. Deposited 18 August 2022.
60. S. Mottaghinia *et al.*, Melomys woolly monkey virus isolate 249-MF, partial genome. NCBI Genbank. <https://www.ncbi.nlm.nih.gov/nuccore/ON903268.1/>. Deposited 7 November 2023.
61. S. Mottaghinia *et al.*, Gammaretrovirus sp. isolate ABTC 45799 gag protein (gag) gene, complete cds; nonfunctional pol protein gene, complete sequence; and envelope protein (env) gene, complete cds. NCBI Genbank. <https://www.ncbi.nlm.nih.gov/nuccore/OP921767>. Deposited 9 January 2024.
62. S. Mottaghinia *et al.*, Gammaretrovirus sp. isolate ABTC 46056 gag protein (gag), pol protein (pol), and envelope protein (env) genes, complete cds. NCBI Genbank. <https://www.ncbi.nlm.nih.gov/nuccore/OP921768>. Deposited 9 January 2024.
63. S. Mottaghinia *et al.*, Gammaretrovirus sp. isolate ABTC 44487 gag protein (gag) gene, complete cds; nonfunctional pol protein gene, complete sequence; and envelope protein (env) gene, complete cds. NCBI Genbank. <https://www.ncbi.nlm.nih.gov/nuccore/OP921769>. Deposited 9 January 2024.
64. S. Mottaghinia *et al.*, Gammaretrovirus sp. isolate ABTC 137299 gag protein (gag) gene, complete cds; nonfunctional pol protein gene, complete sequence; and envelope protein (env) gene, complete cds. NCBI Genbank. <https://www.ncbi.nlm.nih.gov/nuccore/OP921770>. Deposited 9 January 2024.
65. S. Mottaghinia *et al.*, Gammaretrovirus sp. isolate ABTC 24240 nonfunctional gag-pol protein gene, partial sequence; and envelope protein (env) gene, partial cds. NCBI Genbank. <https://www.ncbi.nlm.nih.gov/nuccore/OP921771>. Deposited 9 January 2024.
66. Leibniz institute for zoo and wildlife diseases, SRX22123073: WGS of inverse PCR cMWMV infected HEK293 cells. NCBI SRA. <https://www.ncbi.nlm.nih.gov/sra/?term=SAMN37868416>. Deposited 9 January 2024.
67. Leibniz institute for zoo and wildlife diseases, SRX22123074: WGS of inverse PCR cMWMV infected NIH3T3 cells. NCBI SRA. <https://www.ncbi.nlm.nih.gov/sra/?term=SAMN37868417>. Deposited 9 January 2024.
68. Leibniz institute for zoo and wildlife diseases, SRX22123075: WGS of inverse PCR KoRV-A infected HEK293 cells3 NCBI SRA. NCBI SRA. <https://www.ncbi.nlm.nih.gov/sra/?term=SAMN37868418>. Deposited 9 January 2024.
69. Leibniz institute for zoo and wildlife diseases, SRX22123076: WGS of inverse PCR KoRV-A infected NIH3T3 cells. NCBI SRA. <https://www.ncbi.nlm.nih.gov/sra/?term=SAMN37868419>. Deposited 9 January 2024.
70. S. Mottaghinia, Fig\_2\_nt.seq.alignment.fasta related to manuscript "A Recent Gibbon Ape Leukemia Virus Germline Integration in a Rodent from New Guinea". Figshare. [https://figshare.com/articles/dataset/Fig\\_2\\_nt\\_seq\\_alignment\\_fasta/21704216/2](https://figshare.com/articles/dataset/Fig_2_nt_seq_alignment_fasta/21704216/2). Deposited 3 February 2022.