Endogenous Gammaretrovirus Acquisition in *Mus musculus* Subspecies Carrying Functional Variants of the XPR1 Virus Receptor

Devinka Bamunusinghe,a Qingping Liu,a Xiaoyu Lu,a Andrew Oler,b Christine A. Kozak

Laboratory of Molecular Microbiology,a Bioinformatics and Computational Biosciences Branch,b National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA

The xenotropic and polytropic mouse leukemia viruses (X-MLVs and P-MLVs, respectively) have different host ranges but use the same functionally polymorphic receptor, XPR1, for entry. Endogenous retroviruses (ERVs) of these 2 gammaretrovirus subtypes are largely segregated in different house mouse subspecies, but both MLV types are found in the classical strains of laboratory mice, which are genetic mosaics of 3 wild mouse subspecies. To describe the subspecies origins of laboratory mouse XP-MLV ERVs and their coevolutionary trajectory with their XPR1 receptor, we screened the house mouse subspecies for known and novel Xpr1 variants and for the individual full-length XP-MLV ERVs found in the sequenced C57BL mouse genome. The 12 X-MLV ERVs predate the origins of laboratory mice; they were all traced to Japanese wild mice and are embedded in the 5% of the laboratory mouse genome derived from the Asian *Mus musculus musculus* and, in one case, in the <1% derived from *M. m. castaneus*. While all 31 P-MLV ERVs map to the 95% of the laboratory mouse genome derived from P-MLV-infected *M. m. domesticus*, no C57BL P-MLV ERVs were found in wild *M. m. domesticus*, no C57BL P-MLV ERVs were found in wild *M. m. domesticus*. All *M. m. domesticus* mice carry the fully permissive XPR1 receptor allele, but all of the various restrictive XPR1 receptors, including the X-MLV-restricting laboratory mouse Xpr1<sup>n</sup> and a novel *M. m. castaneus* allele, originated in X-MLV-infected Asian mice. Thus, P-MLV ERVs show more insertion polymorphism than X-MLVs, and these differences in ERV acquisition and fixation are linked to subspecies-specific and functionally distinct XPR1 receptor variants.

Gammaretroviruses of two distinctive host range groups that use the same XPR1 receptor have been isolated from laboratory mice. The xenotropic and polytropic mouse leukemia viruses (X-MLVs and P-MLVs, respectively) both are able to infect cells of nonrodent species, although P-MLVs but not X-MLVs can infect cells of their laboratory mouse strain hosts (1–3). These host range differences are due to polymorphisms in the viral envelope and in the XPR1 receptor (4).

MLVs can be transmitted in mice as infectious virus (5, 6) or can be inherited as endogenous retroviruses (ERVs) inserted into the host germ line. X-MLV and P-MLV (collectively, XP-MLV) ERVs are present in the common strains of laboratory mouse; these strains carry 1 to 20 copies of X-MLVs and 10 to 30 copies of P-MLV ERVs (7, 8). The sequenced C57BL mouse genome has 3 distinct clades of X-MLV ERVs (Xmvs) (9), and P-MLV ERVs have been subgrouped as polytropic and modified polytropic murine viruses (Pmvs and Mpmvs, respectively) (10).

The XP-MLV ERVs are recent acquisitions in the *Mus* genome and are essentially restricted to the 4 house mouse lineages that diverged from the other *Mus* species about 0.5 MYA and that live in close association with humans (11). The ERV subtypes are largely segregated in different house mouse subspecies. The western *European Mus musculus domesticus* carries only P-MLVs, whereas *M. m. castaneus* of southeast Asia, *M. m. molossinus* of Japan, and *M. m. musculus* of eastern Europe and Asia all carry multiple copies of *Xmvs* and few copies of *Mpmvs* or *Pmvs*. The house mouse subspecies also carry XP-MLV ERV variants that appear to be recombinants (12).

XP-MLV infection is mediated by the XPR1 receptor, a transmembrane protein of unknown function that acts as a receptor for P-MLVs in mice and humans and for X-MLVs in humans but not most common strains of laboratory mice (13–15). The virus-infected *M. musculus* subspecies carry polymorphic variants of the XPR1 XP-MLV receptor with different sensitivities to XP-MLV subtypes (16). Four such variants of the *Xpr1* receptor have been found in species of *Mus*, 3 of which are restrictive and are found either in the mice that carry X-MLVs or in a species (*M. pahari*) that is sympatric with house mice that carry these viruses.

The classical inbred strains of laboratory mice have no wild mouse counterparts but were developed at the turn of the last century from fancy mice bred by hobbyists, and these strains have long been recognized as mosaics of several house mouse subspecies (17). The presence of multiple copies of both X- and P-MLV ERVs in the common inbred strains is consistent with this hybrid origin. Although a few of these common inbred strains carry the permissive XPR1 receptor of European *M. m. domesticus*, *Xpr1<sup>nm</sup>*'s, most of the laboratory strains are resistant to infection by X-MLVs, and this is due to a fifth XPR1 receptor variant, *Xpr1<sup>n</sup>*, that has not been found in any wild mouse (16).

The infectious XP-MLVs represent a virus family that includes variants with atypical host range (18), and this diversity, together with corresponding polymorphisms in their XPR1 receptor, illustrates how genetic conflicts can produce diversifying selection in coevolving hosts and pathogens. To examine the coevolutionary pathways of these interacting entities, we examined the wild mouse origins and fixation of the individual XP-MLV ERVs found...
in the sequenced C57BL mouse genome. This was done to establish whether these ERV insertions originated through invasion by infection or by introgression. We screened these same mice for known and novel XPR1 receptor variants. We describe different subsets of 20 to 58 wild mouse DNA samples and the large number of samples from each subspecies, different nucleotide polymorphisms (SNPs) to define the local subspecific origin of North Carolina (http://msub.csbio.unc.edu) (19) to examine and XPR1 variants.

### MATERIALS AND METHODS

**Mice and DNAs.** Sources of mice and DNAs are listed in Table 1, as well as in Table S1 in the supplemental material. DNAs for some strains of laboratory mice were isolated from mice maintained in our laboratory. Mouse or DNA was also purchased from The Jackson Laboratory (Bar Harbor, ME). For *M. musculus* samples, DNAs were isolated from 33 wild-trapped or wild-derived mice in our laboratory or were obtained from other sources for 54 mice. DNAs were isolated from livers of wild-derived mice obtained in 1984 to 1986 from the randomly bred colonies of M. Potter (NCI, Bethesda, MD). Many of these colonies were subsequently transferred to The Jackson Laboratory, where they were inbred, and DNAs were also obtained from some of these strains. Wild-trapped mice from Lake Casitas, CA, were obtained from S. Rasheed (University of Southern California, Los Angeles) in 1986. Additional DNAs from wild-caught or wild-derived mice were obtained from M. Nachmann (University of Arizona, Phoenix), from R. Abe in 1991 (Naval Medical Research Institute, Bethesda, MD), from S. Chattopadhyay and H. Morse III (NIAID, Bethesda, MD), and from RIKEN BioResource Center (Ibaraki, Japan). This mouse DNA panel includes 36 classical inbred strains, 14 DNAs isolated directly from wild-trapped mice, and 73 DNAs from wild-derived laboratory colonies and wild-derived inbred strains of *M. musculus* or *M. spreatus*. **Priming and PCR.** Primers to identify 3' and 5' cell-virus junction fragments and the empty preintegration loci were designed for 19 Pmv, 12 Xmv, and 12 Mpmv (see Table S2 in the supplemental material). Junction fragments were generated using primers from flanking sequences and forward and reverse primers from the MLV long terminal repeat (LTR) sequence: 5'-CAGCTCGGTCTCGCTTCTG. The Pmv16 provirus was identified by an additional primer from the virus env, XT: 5'-TCAGGAGAAGGTGTTTGGAG. Because of limited DNA for many samples and the large number of samples from each subspecies, different subsets of 20 to 58 wild mouse DNAs were scored for each insertion. While most DNAs produced either the cell-virus junction fragment or the empty locus for each tested ERV insertion (Table 2), there were rare exceptions for which no primer set produced amplicons. These failures were generally subspecies specific, and analysis of the insertion site of one such example (*Mpmv9*) in the genome sequence of *M. m. castaneus* identified large deletions overlapping the original primer sequences (data not shown).

### TABLE 1 Inbred strains of laboratory mice grouped by related breeding history

<table>
<thead>
<tr>
<th>Strain family</th>
<th>Strains</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Zealand</td>
<td>NZB/B1NJ, NZL/Ltj, NZO/H1Ltj</td>
</tr>
<tr>
<td>Swiss</td>
<td>FVB/NJ, NFS/N, NOD/ShiLtj, NON/Ltj, SIM/R, SJL/J, SWR/J</td>
</tr>
<tr>
<td>Other</td>
<td>BUB/Bnu, CE/J, LP/J, PL/J, RIIs/J</td>
</tr>
</tbody>
</table>

* Collector Abbas Lathrop provided fancy mice to Castle that served as progenitors for the classical inbred strains. NZ mice were derived from outbred NIH Swiss colony or from other sources of Swiss mice.

### TABLE 2 Distribution of individual C57BL XP-MLV ERVs in laboratory strains and house mouse (*M. musculus*) subspecies

<table>
<thead>
<tr>
<th>Laboratory strains</th>
<th>No. positive for each ERV/no. tested*</th>
</tr>
</thead>
<tbody>
<tr>
<td>C57BL ERV</td>
<td>Laboratory strains</td>
</tr>
<tr>
<td>Xmv1</td>
<td>18/20</td>
</tr>
<tr>
<td>Xmv2</td>
<td>4/20</td>
</tr>
<tr>
<td>Xmv3</td>
<td>6/19</td>
</tr>
<tr>
<td>Xmv4</td>
<td>21/24</td>
</tr>
<tr>
<td>Xmv5</td>
<td>3/21</td>
</tr>
<tr>
<td>Xmv6</td>
<td>20/22</td>
</tr>
<tr>
<td>Xmv7</td>
<td>11/21</td>
</tr>
<tr>
<td>Xmv8</td>
<td>2/21</td>
</tr>
<tr>
<td>Xmv9</td>
<td>20/22</td>
</tr>
<tr>
<td>Xmv10</td>
<td>13/23</td>
</tr>
<tr>
<td>Xmv11</td>
<td>3/18</td>
</tr>
<tr>
<td>Xmv12</td>
<td>1/26</td>
</tr>
<tr>
<td>Xmv13</td>
<td>3/20</td>
</tr>
<tr>
<td>Xmv14</td>
<td>5/19</td>
</tr>
<tr>
<td>Xmv15</td>
<td>20/24</td>
</tr>
<tr>
<td>Xmv16</td>
<td>4/26</td>
</tr>
<tr>
<td>Xmv17</td>
<td>3/22</td>
</tr>
<tr>
<td>Xmv18</td>
<td>4/22</td>
</tr>
<tr>
<td>Xmv19</td>
<td>10/26</td>
</tr>
</tbody>
</table>

* Classical laboratory strains include strains derived from the Lathrop fancy mice, the NZ strains, and other strains not derived from Swiss mice. Blank entries represent DNA sets that failed to produce any PCR product (empty locus or junction fragments). Xmv, endogenous xenotropic MLV; Pnv, polytropic MLV; Mpmv, modified polytropic MLV.

**Identification of the subspecies origins of XP-MLV integration sites and XPR1 variants.** We used the Mouse Phylogeny Viewer at the University of North Carolina (http://msub.csbio.unc.edu) (19) to examine genomic segments containing each ERV integration site to provide insight into their wild mouse origins. This browser uses a set of diagnostic single-nucleotide polymorphisms (SNPs) to define the local subspecific origin along each autosome and the X chromosome for a set of 100 classical
laboratory strains and 98 wild-derived and wild-caught mice. This browser also provides information on regions of haplotype identity for the inbred strains and the SNP variants that define those regions. For this analysis, we used chromosome coordinates for each ERV and for Xpr1, determined for the NCBI37/mm9 reference assembly by BLAT searches (20) of flanking sequences using the UCSC genome browser (http://genome.ucsc.edu/).

Identification and characterization of a novel Xpr1 variant. Sequence reads (BAM format) were obtained from D. Halligan and P. Keightley (University of Edinburgh, United Kingdom) for 10 wild M. m. castaneus animals trapped in 10 different locations in northwest India (see Table S1 in the supplemental material). Sequences were aligned for the region of Xpr1 exon 13 using SAMtools (21), chromosome coordinates Chr1:15737324-15737461 according to the NCBI37/mm9 reference assembly. Files for each strain were converted to fastq format with bam2fastx (from the TopHat package (22)) and assembled with SOAPdenovo (http://soap.genomics.org.cn) (23) using the following parameters: SOAPdenovo63mer all -K 59. The resulting contigs and the corresponding sequence from GenBank accession number AF131097 were aligned with ClustalW using Jalview (24).

The QuickChange II XL site-directed mutagenesis kit (Stratagene, La Jolla, CA) was used to introduce the M. dunni Xpr1sxv variant into two previously described (20) of flanking sequences using the UCSC genome browser (http://genome.ucsc.edu/). The primer sets for all 43 ERVs produced the expected products or the empty locus. The empty preintegration locus was not identified for Xmv8 and Xmv9, two linked ERVs located on chromosome 4 at Chr4:145.68 and Chr4:146.6, respectively. These Xmv variants are inserted in a region of small local duplications; BLAT screens identified additional copies of the insertion sites for Xmv8 (for example, Chr4:145.26 and Chr4:146.60) and for Xmv9 (Chr4:145.45 and Chr4:145.7). Because of these additional copies, the sequences surrounding these Xmv could not be used to design primers to detect the empty locus.

All 12 Xmv were identified in multiple Lathrop-derived inbred strains (Table 2). The eight Swiss-derived strains generally lacked these 12 Xmv, the notable exceptions being NOR, SMLR, and FVB. These strains carry the linked Xmv8 and Xmv9 ERVs. These insertions map near the Fv1 retrovirus restriction gene (Chr4:147.2) (30), and all 3 strains carry the Fv1b allele at this locus, as does C57BL and other strains carrying Xmv8 and Xmv9 (31–33), suggesting a common origin for Fv1b and these 2 Xmv.

Not surprisingly, none of the C57BL Xmv were identified in wild-derived or wild-caught M. m. domesticus, mice that lack X-MLV env genes according to southern analysis (Table 2) (11). Similarly, none of these insertions were identified in European M. m. musculus, despite the presence of multiple X-MLV env genes in these mice (11). All 12 Xmv were, however, found in M. m. molossinus, and most were identified in multiple molossinus DNAs. M. m. molossinus is a natural hybrid of M. m. castaneus and M. m. musculus (34), and two Xmv, Xmv42 and Xmv8, were also found in M. m. castaneus, suggesting these ERVs predate the molossinus hybridization. The acquisition of these Xmv by laboratory strains likely results from the inclusion of Japanese mice, like the Japanese waltzer mouse, in the fancy mouse colonies used to develop the common inbred strains (29), and the failure to detect these ERVs in European M. m. musculus confirms the previously observed distinction between Asian and European M. m. musculus (17).

C57BL Pmv and Mpmv are not found in house mouse subspecies. PCR analysis of 36 inbred mouse strains indicates that the majority of P-MLV ERVs (collectively termed M/pmvs) are carried by fewer than half of the inbred strains tested (Table 2), although several, like Pmv10, Pmv8, Pmv10, and Pmv13, are carried by nearly all strains, suggesting that their acquisition predates laboratory strain origins. Previous studies of Mus taxa have shown that P-MLV ERVs are largely restricted to the M. m. domesticus lineage of house mice (11), a subspecies which is globally distributed as a result of passive
transport by western Europeans. M/pmvs are also found in M. spretus, a species outside the house mouse lineage, but one that is sympatric with and interfertile with M. m. domesticus, suggesting its few M/pmvs have been acquired by interbreeding (35). Multiple P-MLV ERV env genes are also found in M. m. musculus from eastern Europe, likely reflecting the predominantly west-to-east introgression across the hybrid zone that separates the ranges of M. m. domesticus and M. m. musculus (36).

Not surprisingly, the Pmvs or Mpmvs were rarely identified by PCR in Asian subspecies which lack P-MLV env genes by Southern hybridization (Table 2) (11), but it was surprising that none of the 31 M/pmvs, including those present in nearly all laboratory strains, were reliably detected in the M. m. domesticus DNAs or in M. spretus, even though we sampled multiple geographically separated populations from 5 continents. This suggests a high level of insertional polymorphism and indicates that none of these ERVs were acquired at or soon after subspeciation.

**Solo LTRs: evidence of provirus loss by homologous recombination.** The mouse genome is replete with solo LTRs, resulting from the excision of the bulk of a retrovirus by homologous recombination, although for the recently acquired MLV ERV family, solo LTRs represent less than 20% of insertions (37). Solo LTRs at the 43 XP-MLV insertion sites should be identifiable by the presence of cell-LTR junction fragments together with an empty-locus PCR product of increased size (Fig. 1B). Within our set of 43 ERVs, we identified only one such deletion, of Pmv16. The larger-than-expected “empty” site for Pmv16 contained a solo LTR with the expected U3-R-U5 configuration and the 190-bp U3 insertion typical of P-MLV ERVs (data not shown). The full-length Pmv16 was found only in C57BL, and the solo LTR was found in half of the laboratory strains and none of the wild mice.

**Subspecific origins of genome segments carrying XP-MLV ERVs.** High-resolution genomic maps developed for 100 laboratory strains and 98 wild-derived and wild-trapped species (17)
have been used to infer the subspecific origins of DNA segments along each of the 19 mouse autosomes and the X Chr, and they showed that the genomes of the various inbred strains are 90 to 98% *M. m. domesticus*, 2 to 9% *M. m. musculus*, and 1% *M. m. castaneus*. This analysis also confirmed that Japanese *M. m. molossinus* is a natural hybrid of *M. m. musculus* and *M. m. castaneus* and determined that the *M. m. musculus* segments of the laboratory mouse genome derive from Japanese mouse (*M. m. molossinus*) progenitors (17) (Fig. 2). The smaller Asian mouse-derived segments are scattered over the laboratory mouse genome, and the pattern of this distribution differs from strain to strain. We used this resource to examine the insertion sites for each of the 43 C57BL ERVs.

**Complications from pervasive intersubspecific introgression in wild-derived strains.** Our PCR analysis identified several M/pmv, as well as Xmvxs, in rare *M. m. domesticus* samples. The exceptions involved wild-derived laboratory strains but no wild-trapped mice. This is an important distinction, because there is evidence of intersubspecific introgression in the wild-derived

**FIG 2** Distribution of XP-MLV ERVs and Xpr1 receptor alleles in Mus taxa and in *M. musculus*-derived laboratory strains. A phylogenetic tree shows the house mouse (*M. musculus*) radiation. The inbred strain genomes are largely *M. m. domesticus* derived, but some of these strains acquired the restrictive Xpr1 receptor and X-MLV ERVs from Japanese fancy mice (*M. m. molossinus*).
strains that is suggestive of undocumented cross-breeding with common laboratory strains (17). In particular, multiple (4–12) C57BL ERVs were found in the \textit{M. m. domesticus} wild-derived strains SOD1, RBB, RBA, and RBF, and all 4 strains contain substantial genomic regions that are nearly identical to classical inbred strains (17). The donor of these introgressed regions has been identified as C57BL for SOD1 and RBA, providing an explanation for the presence of several C57BL ERVs that were not found in any other \textit{M. m. domesticus} mice. That these ERVs resulted from contamination is supported by examination of genomic insertion sites for the \textit{Xmv} that were detected in these mice by PCR. In all cases (Fig. 4B shows representative data), the \textit{M. m. domesticus} wild-derived mice carrying \textit{Xmv}s are \textit{M. m. musculus} at these sites. Therefore, detection of these ERVs in clearly contaminated strains is not evidence of their wild mouse origins, so typing data from

### TABLE 3 Region of shared ancestry defined by SNPs around Xmv15

<table>
<thead>
<tr>
<th>M. musculus subspecies</th>
<th>Mouse</th>
<th>Mismatched SNPs</th>
<th>Shared SNPs</th>
<th>Uniquely shared SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>molossinus</td>
<td>JF1</td>
<td>2</td>
<td>38</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>90.5</td>
</tr>
<tr>
<td>musculus</td>
<td>BAG94</td>
<td>5</td>
<td>35</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>76.2</td>
</tr>
<tr>
<td>castaneus</td>
<td>IN13</td>
<td>12</td>
<td>24</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>47.4</td>
</tr>
<tr>
<td>domesticus</td>
<td>MWN1026</td>
<td>21</td>
<td>52.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
</tbody>
</table>

\(a\) Comparisons for the 40 SNPs in a region including Xmv15 (Chr9:62112348-62297739); SNPs were from http://msub.csbio.unc.edu (19). The 9 Xmv15\(^{+}\) inbred strains (C57BL, NZB, NZL, RF, LT, C57BR, C58, and AKR) share all 40 SNPs, and these were compared to representative DNAs of the 4 subspecies. The unique shared SNP subset excludes the 19 SNPs for which the 9 inbreds and 4 wild mice are identical.

FIG 4 Subspecific origins of \textit{Xpr1} alleles and XP-MLV ERVs and haplotype diversity at insertion sites. (A) Regions of haplotype similarity at a representative P-MLV ERV, \textit{Mpmv6}. The black line marks the Chr 1 insertion site, and the different track colors represent different shared haplotypes. The \textit{Mpmv6} insertion lies in a region of haplotype similarity (tan color) shared by the 12 positive strains as determined by Mouse Phylogeny Viewer (19). (B and C) The unexpected detection of \textit{Xmv12} in \textit{M. m. domesticus} strains RBB and SOD1 and \textit{Mpmv6} in \textit{M. m. molossinus} strains MOLD and MOLG is because their insertion sites in these strains are \textit{M. m. musculus} and \textit{M. m. domesticus} derived, respectively, due to genomic contamination. (D) Subspecific origins of the two laboratory mouse \textit{Xpr1} alleles and two linked X-MLVs. In panels B to D, ERV sites are indicated by yellow lines, and track colors represent 3 \textit{Mus musculus} subspecies. To the right are PCR typing data for each marker.
these 4 strains was deleted from Table 2. Similarly, insertion sites for several M/mpv ERVs unexpectedly found in Asian wild-derived strains were found within contaminating M. m. domesticus segments (Fig. 4C). This also raises concerns about the few P-MLV ERVs detected in rare wild mouse DNAs (Table 2); all of these cases were wild-derived strains either known to be contaminated (PERA/C, Pmv10 and Mpmv10; SF, Mpmv6; CALB, Mpmv6 and Pnv12), known to be derived from contaminated lineages (MOLC, Mpmv11 and Mpmv4), or not examined for subspecific origins and for possible contamination (BQC, Pmv12; VEJ and Viborg, Mpmv10; JJD, Pmv1; HAF, Pmv11). Thus, the rare positives for these ERVs are not convincing evidence of their wild mouse origins.

**Wild mouse origin of Xpr1**. Two of the 5 known alleles of the Mus XP-MLV Xpr1 receptor are found in the classical and Swiss strains of laboratory mice (Fig. 2). The more common allele, Xpr1n, supports entry of P-MLVs but not X-MLVs. The fully permissive wild mouse allele, Xpr1v, is found in a few of the classical laboratory strains (33). In previous screenings of wild house mice, we identified Xpr1ui in all tested M. m. domesticus animals from Eurasia and the Americas as well as in aboriginal Mus species, but we did not find a single wild mouse carrying Xpr1n (16).

To further investigate the wild mouse origins of the two laboratory mouse Xpr1 alleles, we examined the subspecies origins of the chromosome segment carrying the Xpr1 site on Chr 1 in the 29 strains typed for Xpr1 (Fig. 4D). The 4 strains with the permissive Xpr1ui allele are M. m. domesticus at the Chr 1 Xpr1 locus, consistent with the observation that M. m. domesticus carries this receptor variant. The site of the Xpr1 locus, however, is M. m. musculus derived in the mice carrying Xpr1n. This indicates that the restrictive Xpr1n allele is derived from X-MLV-infected M. m. musculus. This same region of Chr 1 carries 2 Xmvs (Bsvl1/Xmv43 and Xmv41), and analysis of this larger strain set further illustrates the earlier point that the Xmvs map to M. m. musculus-derived genomic regions (Fig. 4D).

**A novel XPR1 variant in M. m. castaneus.** The house mouse lineages originated and diverged from an ancestral population on the Indian subcontinent, which carries many of the alleles found in peripheral Eurasian populations, but it is more genetically variable (38, 39). It is likely that the various XP-MLVs were acquired by this ancestral population along with the various Mus Xpr1 receptors; founders of the M. musculus subspecies with virus and receptor variants then moved to their current ranges. To determine if this ancestral population carries additional Xpr1 variants not found in the peripheral populations, we screened genomic sequences from wild-caught M. m. castaneus mice from northern India for Xpr1 sequence variation.

The four M. musculus Xpr1 alleles have receptor-critical replacement or deletion mutations in the putative fourth extracellular loop (ECL4) encoded by exon 13 and a substitution in ECL3 (exon 11) at position 500 (16). Alignment of the contigs from 10 sequenced M. m. castaneus mice to the ancestral Xpr1 exon 13 sequence (AF131097) demonstrates that 7 mice are homozygous for the deletion of ECL4 amino acid residues TTFKP, a deletion that is diagnostic of the Xpr1n receptor identified in M. m. castaneus animals trapped in Thailand (Fig. 5A). Only the deletion allele was detected in strain H24, so it is also likely to be homozygous for the deletion; however, as there is low sequence coverage for this strain in this region, it is difficult to rule out the possibility of heterozygosity. In all cases, this deletion is linked to a nearby C→T polymorphism resulting in a Thr→Ile replacement compared to the reference genome; this change lies in the 7th putative transmembrane domain. Mice H12 and H14 lack the Xpr1n TTFKP deletion and are similar to the ancestral type, although they also have a T→A substitution within this segment, resulting in the altered protein sequence TTKP. Strain H15 is heterozygous for these 2 alleles. In exon 11 (ECL3), all 10 mice carry the permissive residue K500 at this receptor-critical position.

In order to assess the effect of this novel F584I substitution on receptor function, we introduced it into two expression clones of the XPR1 receptor. One clone contained the fully permissive Xpr1ui, and the second carried the replacement mutation, K500E. Mus X-MLV receptor function requires either K500 or T582; both sites have to be mutated to restrict X-MLV entry, and both sites are mutated in Xpr1n mice, which are X-MLV resistant (40). Thus, we introduced K500E into a second construct together with F584I to determine if the F584I mutation, like T582Δ of Xpr1n, is unable to support infection in the presence of K500E.

The Xpr1 constructs were expressed in XP-MLV-resistant E36 hamster cells and infected with 4 pseudoviruses with Env glycoproteins having different XPR1 receptor sensitivities. Xpr1ui, F584I showed the permissive parental phenotype, but in the presence of K500E, pseudovirus infection was significantly reduced for 2 of the 4 viruses, AKR6 X-MLV and Case#1 (P = 0.0011 and P < 0.0001, respectively) (Fig. 5B). This indicates that residues at 584 contribute to the virus binding site, and that F584I can reduce receptor function in some receptor sequence contexts. We term this second M. m. castaneus receptor variant Xpr1u2.

**Xpr1 variation in Asian M. musculus.** Three of the restrictive XPR1 genes identified in Mus have different deletions in the receptor-determining ECL4 of this membrane protein. In our previous screening of wild mouse DNAs (16), we typed most mice using a PCR-based assay designed to identify these 3 diagnostic deletions. Because our PCR screening may have missed replacement mutations like Xpr1u2, we sequenced this receptor-determining segment of 31 DNAs from the 3 X-MLV-positive subspecies. No additional novel variants were identified, and virtually all sequences were true to subspecies type and/or their geographical distribution (Table 4). These sequenced DNAs included 12 Japanese mice classified as M. m. molossinus, 5 of which were determined to carry Xpr1n; this confirms our earlier conclusion that Xpr1n arose in Japanese mice.

**DISCUSSION.** The mammalian genome is populated with various families of ERVs that can alter gene function and genome structure and that are capable of forming stable associations with their hosts over evolutionary time. Understanding how these elements originate, spread, and diversify is fundamental to our understanding of evolution and the virus-host relationship. The present study was done to provide insights into the origins of the XP-MLVs that have become fixed in laboratory mice and to describe their subspecies and strain distribution in conjunction with the evolution of functional variants of their receptor.

Novel ERV families originate as the result of exogenous integration events. Active ERVs can produce additional ERVs, and the expansion of most ERVs is thought to occur soon after their introduction into the germ line as the result of recurrent reinsertion and reinsertion of active elements in hosts that encode functional receptors. Host populations with active ERVs are characterized by...
some degree of insertional polymorphism. With time, ERVs accumulate mutations, and they become less active and less capable of generating new insertions.

Two distinct patterns of ERV acquisition and fixation were observed for X-MLVs and P-MLVs, and these patterns are generally consistent with their different tropisms and with the XPR1 receptors carried by their hosts (Fig. 2). As shown here, all of the C57BL X-MLV ERVs are found in Japanese wild mice, as previously demonstrated for one of them, Bxv1/Xmv43 (33). This suggests limited ongoing retrotransposition, which is not surprising, because although many X-MLVs are active (4, 41), there is an entry block imposed by the laboratory mouse Xpr1n receptor, a receptor that we show here to have originated in Japanese wild mice, just as the M. m. musculus-derived genomic segments in which these ERVs are embedded have been determined to be Japanese (17).

In contrast, we failed to convincingly trace any of the P-MLV ERVs to wild mouse progenitors. This is surprising for 2 reasons. First, these ERVs are present in high copy numbers in all M. m.
domesticus mice (11), and we sampled M. m. domesticus mice trapped at multiple sites on 5 continents. Second, all 31 insertion sites were in regions of shared haplotypes, indicating that insertion predates the origins of laboratory mice. The unexpected failure to identify any of the 31 C57BL M/pmvs in M. m. domesticus suggests that these P-MLV ERVs either were fixed in M. m. domesticus subpopulations not sampled for this analysis or were acquired during the centuries-long development of the fancy mouse colonies that served as progenitors of the laboratory strains.

Insertional polymorphism is characteristic of recently acquired and active families of ERVs carried by hosts with permissive receptors, and P-MLV-positive wild and inbred mice carry the P-MLV permissive receptors, Xpr1\textsuperscript{sv} and Xpr1\textsuperscript{v}. However, this correlation is complicated by the fact that no infectious viruses corresponding to \textit{Pmv} or \textit{M/pmvs} have been isolated; all infectious MLVs with P-MLV host range are recombinants between these P-MLV ERVs and infectious viruses of other host range subgroups (42, 43). The reason for the failure of P-MLV ERVs to produce infectious virus is not known. Although many of these ERVs have open reading frames for all coding regions, their LTRs carry a negative regulatory element and a 190-bp insertion of unknown functional significance that has not been found in any infectious MLV (44, 45). Thus, although functional defects in the \textit{M/pmvs} LTRs may be responsible for the absence of infectious \textit{Pmv}- and \textit{M/pmvs}-like viruses, at least some of these LTRs have promoter activity (46). The failure to produce infectious virus suggests that the mechanism for P-MLV ERV retrotransposition is independent of recombination, and in fact, P-MLV ERVs can amplify in the presence of ecotropic MLV (E-MLV) infection, because P-MLV genomes are preferentially packaged in E-MLV particles that then use the mCAT-1 receptor (47, 48). Such retrotransposition in the absence of infection would be expected to amplify ERVs with fatal defects, but our panel of 31 \textit{M/pmvs} contains no obvious examples of this, and the E-MLVs needed to facilitate such retrotransposition have not been found in M. m. domesticus except in California, where M. m. domesticus has recently interbred with E-MLV-positive M. m. castaneus introduced from Asia (11). Further work on P-MLV ERV activity in inbred and wild mice is needed to resolve this issue.

MLVs have not been detected in species other than mice. This is surprising given the global distribution of MLV-infected mice, the ready transmissibility of MLVs in mice housed together (6), and the fact that the ubiquitously expressed XPR1 acts as an X-MLV receptor in most mammals (16). The failure to find MLV ERVs in other mammals could reflect the fact that retrotransposition in most sequenced genomes is incompletely annotated and the fact that few species in direct contact with Asian virus-producing mice have been sequenced. Also, virus transmission and the likelihood of endogenization can be influenced by multiple host factors. The factors that govern trans-species transmission are poorly understood, although some host restriction factors, like APOBEC3, are effective against XP-MLVs (49). Interestingly, while APOBEC3 is blocked by the MLV glycosylated Gag (50), this antagonist is not encoded by P-MLV ERVs, which, unlike \textit{Xmvs}, show evidence of APOBEC3 editing (9, 51). Thus, the species specificity of XP-MLV ERVs may be influenced by factors in addition to receptor polymorphism.

Ever since the discovery of ERVs and the more recent recognition that ERVs and other transposable elements (TEs) make up nearly half of the mammalian genome, there has been speculation that they play critical but largely unrecognized roles in their hosts. While some specific ERVs have been recruited to serve specific host functions (e.g., Fv1 and syncytins), the vast majority of individual ERVs and TEs have no such defined roles, although these elements clearly can participate in genomic remodeling and are important players in host gene regulatory networks, especially in early development and during stress responses (52, 53). The XP-MLV ERVs described here have already survived purging by natural selection but are not fixed. The insertional polymorphism that produces variation in the ERV content of mouse strains and species could influence host processes governed by individual ERVs or by coordinately regulated ERV subtypes. ERV insertions are estimated to cause 10% of spontaneous mutations in mice (54), and ERV polymorphisms have been associated with changes in gene expression, especially in genes with differential expression across strains (37, 55), suggesting that insertional polymorphic ERVs like the XP-MLVs contribute to these strain-specific phenotypic differences.

Like the P-MLV and X-MLV ERVs, functional variants of the XPR1 receptor show a subspecies-specific distribution. The 3 major house mouse lineages generally carry distinctive and subspecies-specific XPR1 variants (Xpr1\textsuperscript{sv} in M. m. domesticus, Xpr1\textsuperscript{v} in M. m. molossinus and M. m. musculus, and Xpr1\textsuperscript{v} and Xpr1\textsuperscript{v} in M. m. castaneus), indicating a distribution that roughly coincides with the house mouse radiation and with the acquisition of the different MLV ERVs (Fig. 2). Here, we established that the restrictive allele of laboratory mice, Xpr1\textsuperscript{v}, originated in the Asian mice that also supplied the laboratory mouse \textit{Xmvs}, and we identified a novel receptor variant in M. m. castaneus from northern India, bringing to five the number of restrictive alleles found in Asian mice. Because the house mouse subspecies are derived from an ancestral population in India which is more genetically variable than the peripheral populations, further characterization of these mice is likely to identify novel virus subtypes as well as receptor variants.

When retroviruses infect new hosts, coevolutionary pressures produce an "arms race" that results in adaptive mutations in pathogen and host, eventually allowing for some degree of peaceful coexistence. The restrictive alleles of Xpr1 would be expected to interfere with XP-MLV spread and any deleterious consequences of infection, like disease induction. The presence of all 5 restrictive alleles in mice carrying X-MLVs or in mice sympatric with X-MLV-infected house mice (\textit{M. pahari}) suggests that X-MLV infection has driven fixation of these restrictive receptors. On the other hand, for \textit{M. m. domesticus}, the persistence of the permissive XPR1 receptor in these P-MLV-positive mice could have contributed to the observed insertional polymorphism but may also reflect the fact that horizontal spread of P-MLVs is not necessarily XPR1 receptor dependent, rendering XPR1 variation irrelevant for defense in this subspecies.

ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, Bethesda, MD.

We thank Alicia Buckler-White and Ronald Pliska for sequencing.

REFERENCES

leukemia-virus associated with development of spontaneous lymphomas. 


XPR1 receptor. Retrovirology 7:101.


Congenital transmission of murine leukemia virus from wild mice prone 

to the development of lymphoma and paralysis. J. Natl. Cancer Inst. 62: 

63–70.


retroviruses. J. Virol. 61:1037–1044.

7. O’Neill RR, Khan AS, Hoggan MD, Hartley JW, Martin MA, Repaske 

R. 1986. Specific hybridization probes demonstrate fewer xenotropic than 
mink cell-focus-forming murine leukemia virus env-related sequences in 

8. Frankel WN, Stoye JP, Taylor BA, Coffin JM. 1990. A linkage map of 


1371/journal.pgen.0030183.

10. Stoye JP, Coffin JM. 1987. The four classes of endogenous murine leu-

cemia virus: structural relationships and potential for recombination. J. 
Virol. 61:2659–2669.

mink cell-focus-forming, and two types of ecotropic provisional genes. J. 
Virol. 61:3082–3088.

nonecotropic murine leukemia viruses in wild mice. J. Virol. 72:8289– 
8300.

characterization of a cell surface receptor for xenotropic and polytropic 

for xenotropic and polytropic murine leukemia viruses: possible role in 
96:1385–1390.

15. Yang Y-L, Guo L, Xu S, Holland CA, Kitamura T, Hunter K, Cunning-
ham JM. 1999. Receptors for polytropic and xenotropic mouse leukemia 

2010. Evolution of functional and sequence variants of the mammalian 
XPR1 receptor for mouse xenotropic gammaretroviruses and the human-

derived XMV. J. Virol. 84:11970–11980.

F, Nellaker C, Keane TM, Yalcin B, Wong K, Agam A, Belgard TG, Flint 
J, Adams DJ, Frankel WN, Ponting CP. 2012. The genomic landscape 
shaped by selection on transposable elements across 18 mouse strains. 

Talwar GP, Bonhomme F. 1996. Origin and radiation of the house 


of the cell surface receptor control mouse susceptibilities to xenotropic 

21. Kozak CA, Hartley JW, Morse HC, III. 1984. Laboratory and wild-
derived mice with multiple loci for production of xenotropic murine leu-


Biochemical evidence that MCF murine leukemia viruses are envelope (env) 

23. Chattopadhyay SK, Cloyd WM, Linemeyer DL, Lander MR, Rands E, 
Lowry DR. 1982. Cellular origin and role of mink cell focus-forming 

24. Khan AS, Martin MA. 1983. Endogenous murine leukemia proviral long 

25. Flanagan JR, Krieg AM, Max EE, Khan AS. 1989. Negative control 
region at the 5' end of murine leukemia virus long terminal repeats. Mol. 

The 3' terminal repeat of a transcribed yet defective endogenous 
retroviral sequence is a competent promoter of transcription. J. Virol. 
61:625–632.

27. Evans LH, Alamgir AS, Owens N, Weber N, Vintanova K, Barbian B, 
Babar A, Malik F, Rosenke K. 2009. Mobilization of endogenous retro-

viruses in mice after infection with an exogenous retrovirus. J. Virol. 83: 
2429–2435.


