Recombinant Origin of the Retrovirus XMRV:

Discovery, Analysis and Distribution of Two Ancestral Viruses

A dissertation submitted by

Oya Cingöz

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Genetics

Tufts University

Sackler School of Graduate Biomedical Sciences

February 2012

Run, rabbit run
Dig that hole, forget the sun
And when at last the work is done
Don't sit down, it's time to dig another one

Roger Waters
"Breathe"
1973

ABSTRACT

Xenotropic murine leukemia virus-related virus (XMRV) is a retrovirus, initially reported to be associated with human prostate cancer and chronic fatigue syndrome, although these findings have not been replicated. XMRV bears high sequence identity to a class of mouse retroviruses, called murine leukemia viruses (MLVs). We sought to determine the origin of XMRV and how it may have crossed species to infect the human population. Our work showed that XMRV was created through recombination between two endogenous MLVs, named PreXMRV-1 and PreXMRV-2, and that the recombination event that gave rise to XMRV occurred in the laboratory, during passaging of a prostate tumor xenograft in nude mice. The probability of generating the exact same recombinant more than once and independently is negligible, suggesting that all XMRV isolates described to date are derived from this unique recombination event. Furthermore, XMRV is not present as a single endogenous provirus in any of the wild or laboratory mouse strains tested. The strain distribution of PreXMRV-1 and PreXMRV-2 are quite different, making it unlikely that the two XMRV ancestors could have recombined independently in the wild to generate an infectious virus. Our results show that the association of XMRV with human disease is due to contamination of laboratory assays with virus originating from the initial recombination event.

ACKNOWLEDGEMENTS

The years leading up to the final work included in a dissertation constitutes the longest and the most difficult time of a Ph.D. From many months, sometimes years of failed experiments, trying to find a direction and interpreting data, one rises to such great heights when everything finally comes together. This transition would not have been possible without the help and support of those around me, both academically and personally.

First and foremost, I would like to thank my advisor, John Coffin, for all he has done. He took me in his laboratory as an inexperienced graduate student who had no idea about retroviruses. He gave me all the freedom I needed - sometimes more, allowed me to make my own mistakes and learn from them, taught me to ask for help, provided extremely valuable and to-the-point guidance when I needed it, and made me realize that, in the end, I had become a scientist. It was all I could have asked for.

Members of my thesis committee, Phil Hinds, Naomi Rosenberg and Gavin Schnitzler, have been very helpful over the years, regardless of whether my experiments worked or not, always coming up with ideas and suggestions. They have made my committee meetings fun, friendly, and something I looked forward to, by being absolutely involved in any project I took on and by having interesting discussion sessions. I have been very lucky to have such great researchers on my side.

I would like to thank Wayne Frankel for accepting to be my outside examiner, and for sticking with this decision even when all sensible ways of bringing him to Boston vanished. Being a former member of our laboratory, he has done so much influential work, I found myself referring to his papers over and over again. Thanks for being the

dream outside examiner. Thanks also for all the ideas, answering my endless questions, and sending me decade-old samples from the depths of your freezer.

Members of the Coffin Lab have created an amazing atmosphere to work in, for which I am grateful. I would like to thank my current labmates Ravi, Neeru and Zach for keeping me company, when so many of our members moved on to greener pastures. Please keep the lab going! I would also like to thank past members of the lab who were there during almost my entire time in the lab: Jamie, Julia, Irvin, Raul, Robbie and Andrew. You all have been wonderful to work with! Jamie, I especially miss our coffee breaks and spontaneous chats.

I have been a proud member of the Genetics Program, which would not have been possible if it was not for Phil Hinds and Dean Dawson, who both met with me on a Saturday six years ago, the only day I could be in the country. Thank you for bringing me here to Tufts. A special thanks goes to Evyan, who helped make the visit happen. Di and Libby, thank you so much for all your help at almost every step of my thesis defense. You are the ones that keep the program running. Cindy, thanks for your help with all things administrative. Although we did not have a department for the Genetics Program, the Micro Dept. has kindly accepted me as their own. Verna, it has been said so many times, for so many years, by so many people, but I will repeat it: You are great. You make things happen. The department owes you a lot. Teri, Jaime and Rachel, the department would not stand without your hard work and support, so thank you for making everything run smoothly.

The collaborations I have had with other researchers have been extremely valuable. Chris, even though we did not overlap in the lab, our old projects did. Thanks

for keeping such a great lab notebook and letting me into your project. We had many meetings and discussions with the Huber Lab here at Tufts, and our results have paved the way for establishing the origin of the viral sequences detected in human samples. Brigitte, Brendan, Albert and Madi it was a pleasure working with you. What was once my single, straightforward project evolved into a much bigger and much more interesting story, thanks to the contributions from Vinay Pathak's lab at the National Cancer Institute, Frederick. Tobi and Krista's findings explained one half of a mysterious virus, while our findings explained the other half. The collaboration between our labs shed light on the big picture, resulting in the publication of a great story, with great implications. It was "jigsaw falling into place." It was important. It was fun.

I have received help and supplies from many individuals, who I would like to acknowledge. We shared a floor with the Moore and Heldwein Labs, who have been very generous with the use of their equipment and supplies. Allen and Steve helped with flow cytometry experiments. Alfonse was always there to help with the many posters I have printed. Tufts University Sequencing Facility has received a large number of sequencing requests from me, and they have always been very accurate and timely.

I would like to thank all my friends in the Genetics Program and the Micro Department, although the names would be just too long to be listed here. Derick, I will miss our Settlers nights and the short breaks where we complain about many aspects of life as a graduate student. Ivana, you have been my partner in crime; we were the two international students who joined the program together, did our qualifying exam together, and now we are graduating together. Thanks for the company all through these years.

My friends outside of Tufts helped me have a "real life." Tuğba and Tamer, thank you for being our family in Boston to many of us lost souls away from the warmth of our real families. I thank you especially for opening your lovely home to me, so that I could have the much-needed peace and quiet to write part of this dissertation. Ebru and Duda, thank you for your kindness, generosity, moral support, and providing an escape whenever I needed one. I still remember the Vienna days. Beste, though you have left Boston the last few years, you are still the connection that ties us together. Lerzan, you were there with me through some of the hardest times, and I really appreciated it. I wonder if we will ever be able to travel together. Volkan, I have known you for over a decade, and I know I would not be here if it was not for you. Our name "Pıtır" will always stay with me. Gülfer, in a relatively short time, you became one of the closest. Thanks for understanding me in all aspects of life, keeping an eye on me, and for feeding me after many late nights spent in the lab.

I owe so much to my family.

My grandmother, Nesibe - the only grandparent I have. You are one brilliant, intelligent, funny and caring grandma. I miss our crazy laughs, your amazing meals, and the many summers we have spent together, when I was a little kid.

Banu, you made me realize how lucky we both are for having each other as a sister, to have spent our childhood together, to have someone to share those memories with, to have unconditional love and support whatever life brings us. It was great sharing all this with you. I miss you.

Sinan, thanks to you, I do not worry about Banu as I normally would. You have taught me a lot of things, though I know that was never your intention. You are one

fantastic father. My nephew, Oguz, though you are just too young to read it now, you have brought indescribable happiness to this family. Your smile brightens our days.

Mom and Dad, where I am today has only been possible through the endless sacrifices you have made. Thank you for believing in me, for all your support through the years, for letting me pursue whatever it was I was interested in - not just career-wise, for bringing me up as an independent person. And for letting me go.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION-I: RETROVIRUSES

Brief history of retroviruses

Retroviruses have been the subject of an extremely interesting field of research, with major contributions to our understanding of diverse biological processes such as cancer, infectious disease and evolution.

One of the major breakthroughs in molecular biology came from the study of retroviruses by Howard Temin and David Baltimore in 1970, when the retroviral reverse transcriptase enzyme was discovered. The central dogma of biology, which stated unidirectional flow of information from DNA to RNA to protein, was forever broken. This paradigm shift was awarded the Nobel Prize in Physiology or Medicine in 1975.

It was also through the study of retrovirus-induced tumors that oncogenes and their role in cancer were first discovered. Following the discovery of the first retrovirus-acquired oncogene in 1970, many other oncogenes and tumor suppressor genes were identified, changing the field of cancer research entirely.

Finally, with the discovery of HIV as the causative agent of AIDS, all scientific knowledge that accumulated over the years on retroviruses became extremely useful, for it was rapidly applied to the field of HIV research, allowing the development of potential cures and policies for prevention. Today, although it is still not possible to eliminate HIV from infected individuals, unprecedented treatment options are available.

Retroviral Classification and Nomenclature

Initial classification criteria for retroviruses included the shape of the particles produced as observed under the electron microscope, as well as their host range, a feature determined by their *env* genes. The International Committee on Taxonomy of Viruses (ICTV) criteria group retroviruses into seven genera: alpha-, beta-, gamma-, delta-, epsilonretroviruses, lenti- and spumaviruses, mainly based on the sequence of the *pol* gene, the best-conserved gene among retroviruses. Endogenous retroviruses (ERVs) are grouped into three more loosely defined classes, largely based on their relatedness to exogenous genera. ERVs that group with gammaretroviruses are called Class I, those that group with alpha-, beta-, delta- and lentiviruses are Class II, while those that cluster with spumaviruses are termed Class III. A phylogenetic tree constructed using the RT portion of the *pol* gene of retroviruses from the seven genera is shown in Fig. 1.

Typically, retroviral gene names are given three letter names, which are italicized in writing (e.g. *gag*). The cellular oncogenes occasionally acquired by retroviruses follow the same rule. Polyproteins derived from retroviral genes have the same names, but are capitalized and not italicized (e.g. Gag). Individual protein subunits created by processing of polyproteins are referred to with their first two letters, both capitalized (e.g. RT, for reverse transcriptase).

Genome Organization

One broad way of classifying retroviruses relies on the mode of expression of the retroviral genes. Retroviruses that carry the most basic set of genes, which are present in all non-defective members of the *Retroviridae* family, i.e. *gag, pro, pol* and *env* genes,

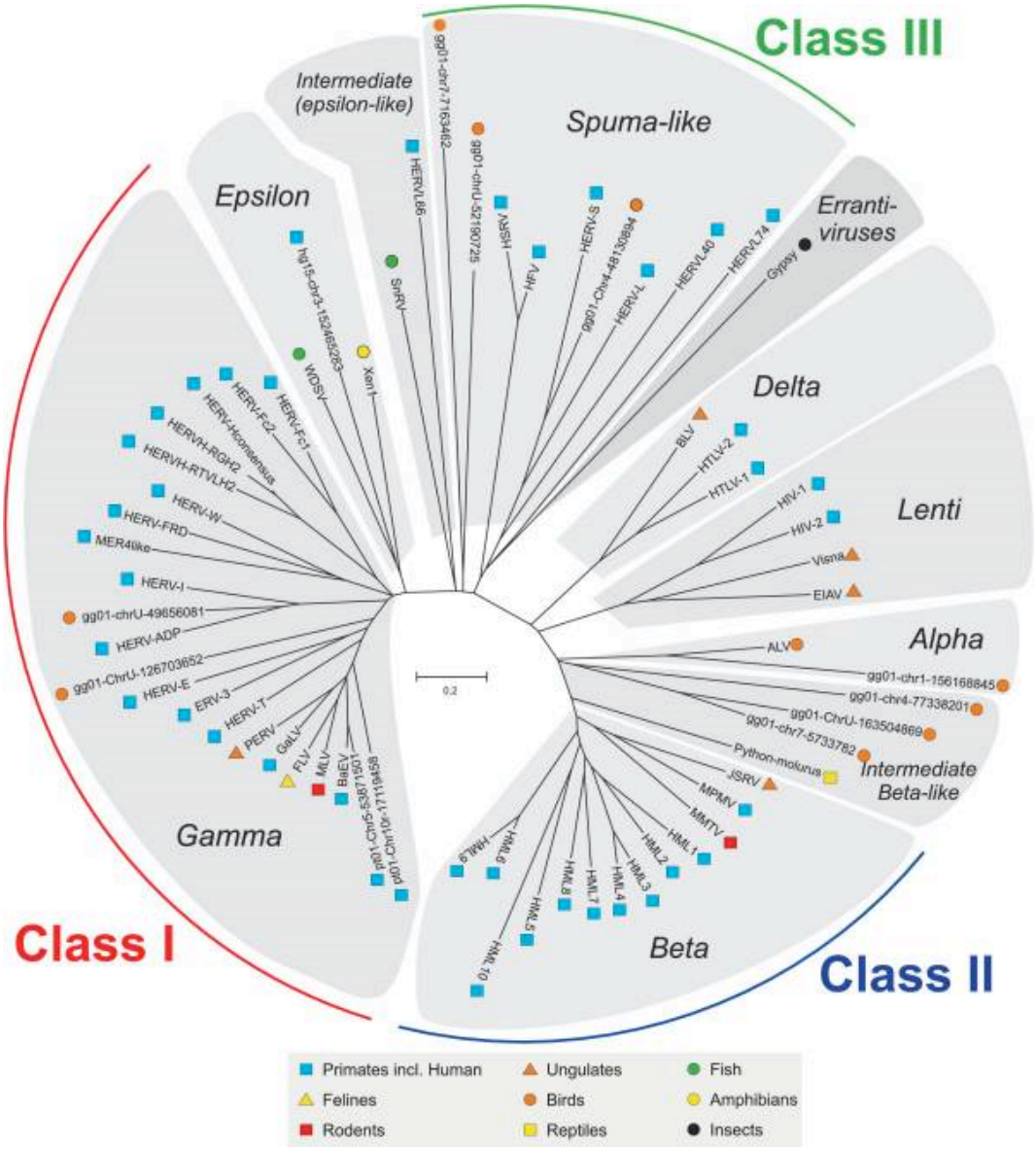


Figure 1. A neighbor-joining tree of the seven retroviral genera

An unrooted neighbor-joining tree based on the *pol* regions of the retroviral genera. Endogenous retroviruses are grouped into three more loosely defined classes, indicated on the periphery, based on their relatedness to exogenous viruses. Figure from (73), under the BioMed Central Open Access license agreement.

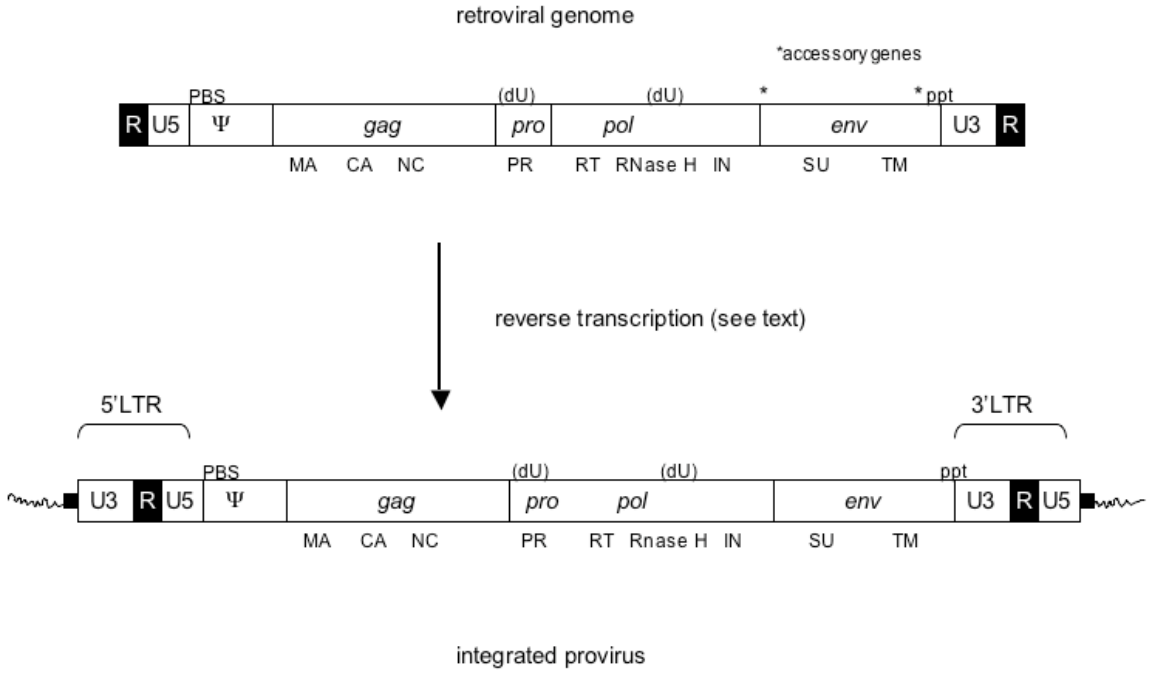


Figure 2. Retroviral genome structure

Retroviral genome structure before (upper panel) and after (lower panel) reverse transcription. R, repeat; U5 and U3, unique 5' and 3'; PBS, primer binding site; Ψ (psi), packaging signal; MA, matrix; CA, capsid, NC, nucleocapsid; PR, protease; RT, reverse transcriptase, IN, integrase; SU, surface; TM, transmembrane; dU, dUTPase domain, present in some retroviral genera; ppt, polypurine tract. Black boxes flanking the provirus represent target-site duplications (TSD). Figure from (64), with permission.

are termed simple retroviruses (Fig. 2). Retroviruses that carry additional "accessory" genes and modify the expression pattern to produce other protein products that help the virus at various steps during its replication cycle are named complex retroviruses.

The basic retroviral genes are always found in the same order: *gag, pro, pol* and *env*. However, the relative positions of the reading frames differ from one genus to the next. Accessory genes could be present between *pol* and *env*, downstream of *env*, or could divide their reading frames between different parts of the genome and be generated by splicing (28).

The retroviral genome is flanked by repeat sequences at each end, termed R. The 5' R sequence is followed by the U5 region, while the 3' R sequence is preceded by the U3 region (Fig. 2, upper panel). Due to strand transfer reactions during reverse transcription (explained in the section: Reverse Transcription), the U3 and U5 regions are duplicated, resulting in the formation of identical long terminal repeat (LTR) sequences, composed of U3-R-U5 that flank each end of the provirus (Fig. 2). The LTRs of a provirus are identical at the time of integration.

Retroviral Genes and Gene Products

A cartoon depicting the structure of a typical simple retrovirus particle is shown in Fig. 3. The *gag* gene encodes the structural proteins of the virion. The minimal set of proteins produced by the *gag* gene includes matrix (MA), capsid (CA) and nucleocapsid (NC). MA associates with lipid membranes, and targets the Gag polyprotein to the cellular membrane. CA monomers assemble to form the "capsid core" that "encapsidates" the retroviral genome. NC protein is found tightly associated to the genomic RNA inside the capsid core.

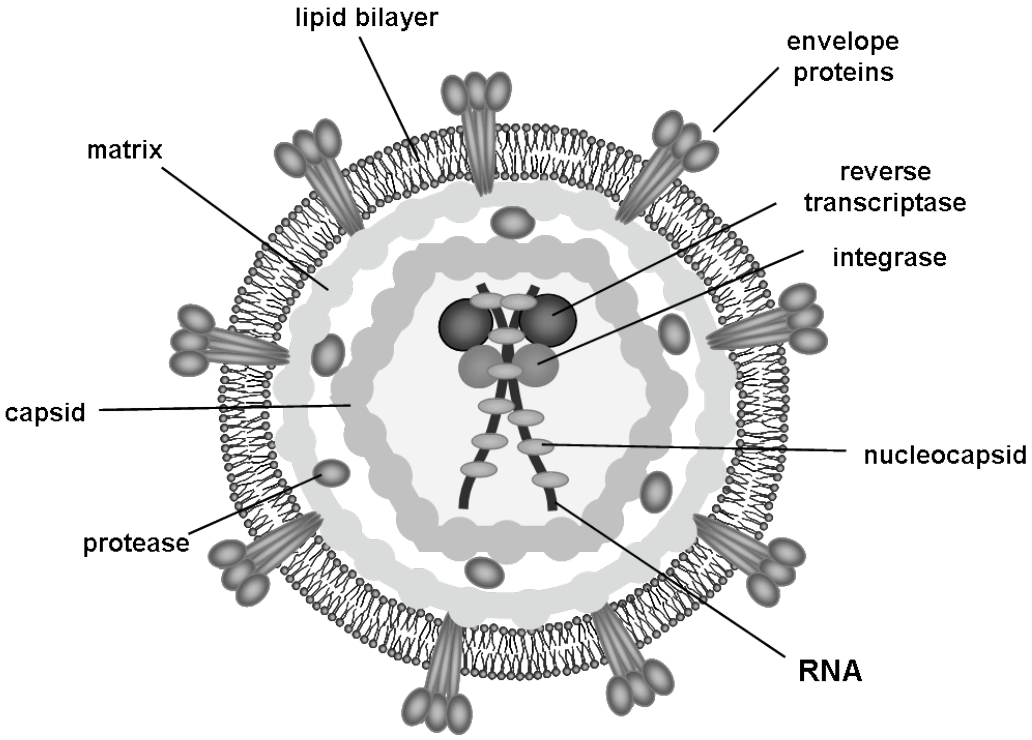


Figure 3. A retrovirus particle

The proteins of a simple retrovirus are shown. The lipid bilayer is derived from the plasma membrane of the host cell. Figure from (174), with permission.

The *pro* gene encodes the viral protease (PR), which is responsible for the proteolytic processing of viral polyproteins into individual subunits. The *pol* gene codes for the reverse transcriptase (RT) that also includes an RNaseH domain, and integrase (IN) proteins present in all infectious retroviruses. RT is an RNA- or DNA-dependent DNA polymerase, the hallmark of *Retroviridae* that earned the family its name. IN catalyzes the integration of the reverse transcribed double stranded viral DNA into the host genome.

The *env* gene codes for the Env glycoprotein made up of the surface (SU) and the transmembrane (TM) subunits. SU interacts specifically with the surface receptor expressed on the host cell, while TM ultimately leads to the fusion of the viral and cellular membranes, mediating entry of the virus particle into the cell.

Some of the accessory genes carried by various complex retroviruses include those that encode proteins that regulate the expression, splicing and transport of viral transcripts, modulate host gene expression and overcome host restriction factors, thereby enhancing viral infectivity (reviewed in (97)).

Retrovirus Replication Cycle

An overview of the retrovirus replication cycle is shown in Figure 4.

Entry and Uncoating

The retroviral replication cycle starts with the binding of the SU subunit of the Env protein on the retrovirus surface with its specific receptor on the target cell surface. This interaction is very specific and is required for virus entry into cells. It is therefore a major determinant of the host range of the virus. The binding of SU to the receptor

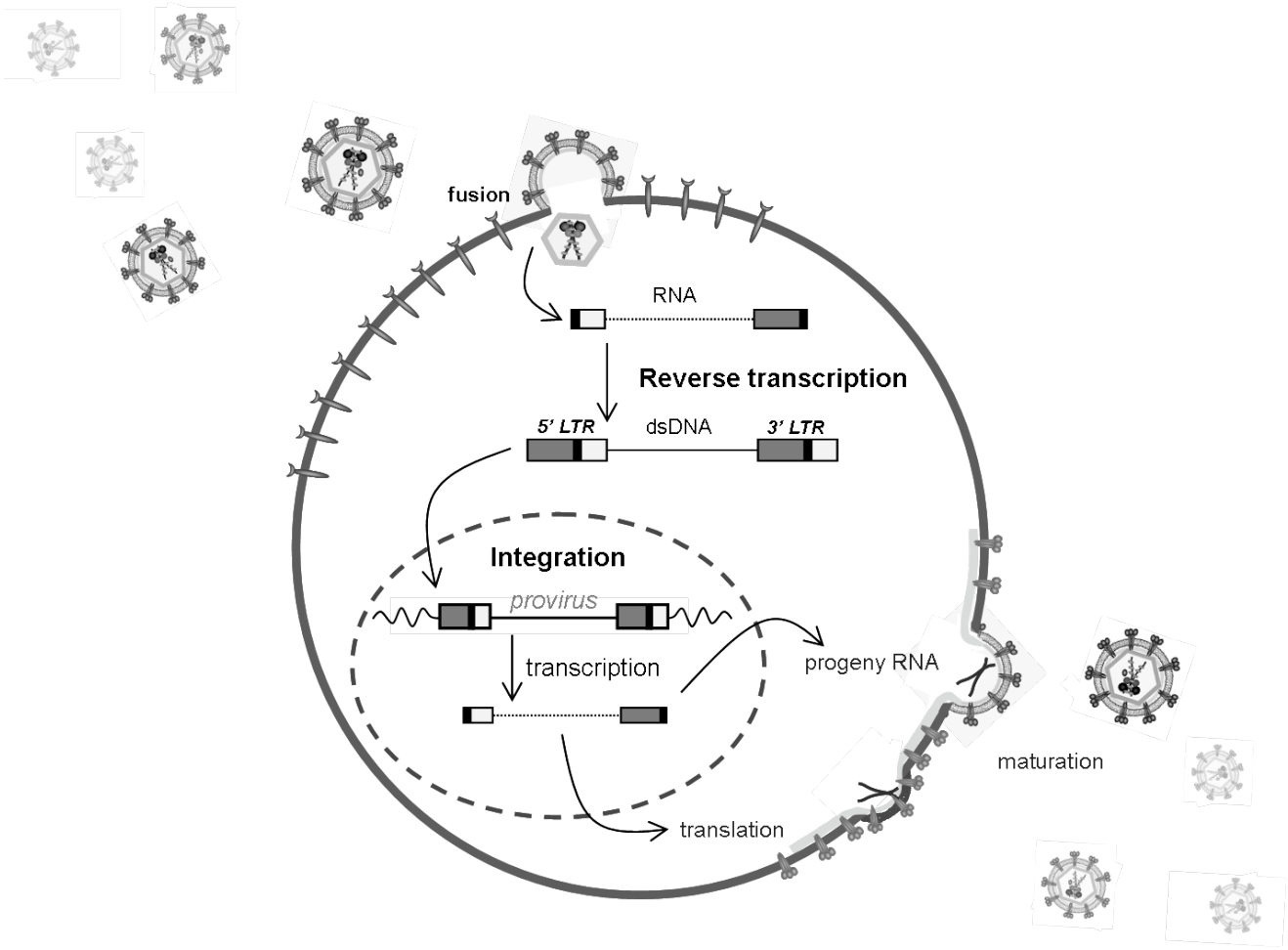


Figure 4. Overview of the retroviral replication cycle

Entry into the host cell is mediated by a specific interaction between the Env protein and a cellular receptor, releasing the viral core into the cytoplasm. Following uncoating and reverse transcription, the pre-integration complexes are transported to the nucleus, where integration of the viral cDNA into the host genome occurs. The integrated provirus is then transcribed and translated to give rise to retroviral proteins and genome, which are packaged into newly formed virions for another round of infection. Figure from (174), with permission.

induces conformational changes that ultimately lead to the fusion of the viral and cellular membranes either at the cell surface, or in an endocytic compartment following internalization, releasing the capsid core into the cytoplasm. The disassembly of the capsid core occurs after entry into the cell and before nuclear import, in a process referred to as uncoating. The precise timing and location of uncoating is unclear; however, there is evidence that the timely disassembly of the capsid is required for the completion of reverse transcription and successful nuclear entry (7).

Reverse Transcription

Reverse transcription is catalyzed by the reverse transcriptase (RT) enzyme packaged into the virions. An overview of the reverse transcription reaction is depicted in Fig. 5. The reaction starts when the minus strand synthesis is primed by a tRNA molecule, which binds to the primer binding site (PBS) sequence just downstream of the U5 region. The DNA is extended until the R region at the very 5' end of the viral genome where it stops, and is referred to as the minus strand strong stop DNA. Meanwhile, the RNaseH domain of RT degrades the genomic RNA strand of the newly formed RNA:DNA duplex.

The minus strand strong stop DNA then jumps to the identical R sequence found in the 3'end of the viral genome (termed the first strand transfer), where RT resumes to extend the minus-strand DNA, and RNaseH continues to degrade the RNA, except for the polypurine tract (PPT) which is somewhat resistant to digestion. The PPT then primes positive-strand synthesis by RT towards the 3' end of the genome, and RNaseH removes the tRNA primer at the 3' end, resulting in the formation of plus-strand strong stop DNA. The second strand transfer occurs when the two single stranded complementary PBS

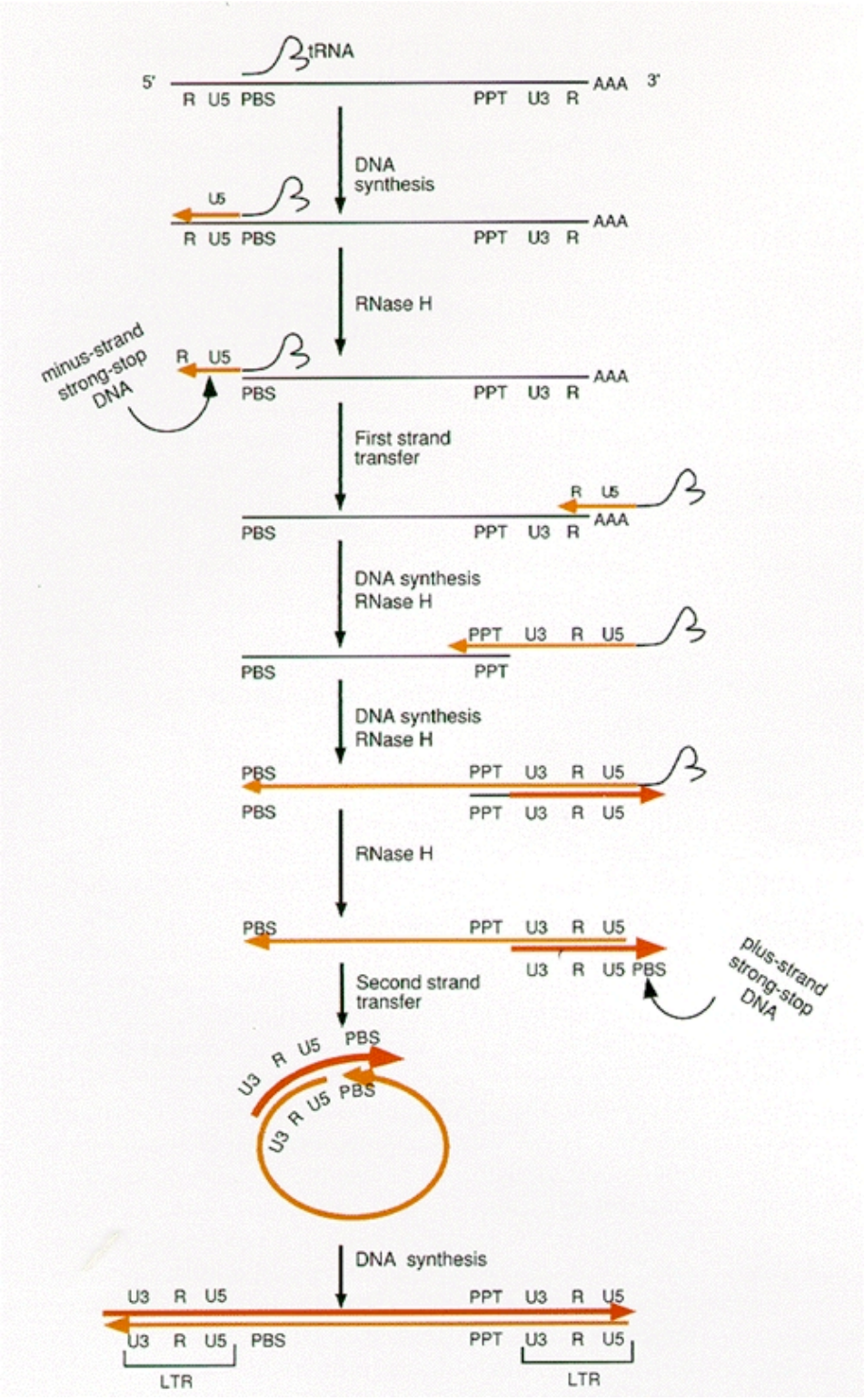


Figure 5. Reverse transcription of the retroviral genome

Black lines represent RNA genome, orange lines minus-strand DNA, and red lines plus-strand DNA. Figure from (28).

regions found at both ends of the newly synthesized DNA anneal to each other, allowing the synthesis of the rest of the positive and negative strands of DNA. The resulting double stranded DNA contains the entire retroviral genome, plus identical LTRs. The newly synthesized dsDNA associates with the retroviral integrase enzyme packaged in virions, as well as other cellular factors to form the integration-competent preintegration complex (PIC).

Because retroviruses typically carry two nearly identical RNA genomes, which copy is used for the synthesis of the proviral DNA is of little consequence. However, in case of virions produced by cells that were co-infected with two genetically distinct retroviruses, different RNA genomes could be packaged into the same virion (70). When such heterozygous virus particles infect a cell, the template switches that occur during reverse transcription between the different genomes may result in the production of chimeric genomes with stretches of identity to both parental genomes. Provided that there are stretches of sequence identity between the templates, homologous recombination can occur between two loci that are 1 kb apart in ~4% of templates, during a single cycle of reverse transcription, while non-homologous recombination happens less frequently (70).

Genetic recombination during reverse transcription has been proposed to occur in two ways: during minus-strand DNA synthesis (copy-choice model) or during plus-strand DNA synthesis (strand-displacement assimilation model), although there is evidence that the majority of recombination events are consistent with the first model (186). As a result of RT template switching, a variety of recombinant retroviral genomes could be created, each with specific crossover sites, hence with specific stretches of sequences from each parental virus. Such recombinant retroviruses have been observed for ALV, HIV and

MLV, among others (27, 169, 176). Recombination between different viral genomes, combined with the mutations introduced by the error-prone RT enzyme, are the major drivers of the genetic diversity seen in retroviral infections, creating a pool of distinct viral populations and allowing the selection of viruses that can replicate more efficiently, evade the immune system, and complement defective viruses.

Integration

PICs are translocated to the nucleus, where retroviral integration occurs, catalyzed by IN. Specifically, IN forms a tetramer that is associated with viral DNA ends, collectively known as the intasome. IN cleaves off two nucleotides from the 3' end of each viral DNA end, resulting in short 5' terminal overhangs. The target DNA is held in a severely bent conformation. IN then captures the target DNA, and catalyzes the strand transfer of 3' viral ends to opposite strands of the host target DNA. The specific number of bases between the attachment sites in the opposite strands varies from one retrovirus genus to the next. The crystal structure of the prototype foamyvirus IN has been largely informative in elucidating the mechanistic details of retroviral integration (58, 96). The integrated DNA copy of a retrovirus is referred to as a provirus.

Expression, Assembly and Budding

Once integrated, the provirus makes use of the host cell machinery to produce multiple copies of the viral genome and to express viral proteins, resulting in the production of viral particles. The LTRs of a provirus contain numerous transcription factor-binding sites, hormone response elements, promoters and enhancers, which allow efficient transcription of the provirus. Part of the nascent mRNA transcripts produced

from the provirus serve as the viral genome to be packaged, while other spliced and unspliced forms are translated by the cellular machinery to produce viral proteins. In case of MLVs, Gag and Gag-Pro-Pol polyproteins are produced from unspliced RNA, while Env protein is produced from singly spliced RNA. The RNA genome and proteins are packaged into virions, which are released from the cell by budding at the plasma membrane. After budding, the polyproteins are cleaved into individual subunits by the viral protease packaged into the virion, in a process called maturation cleavage. The mature virion is then ready for another round of infection.

Endogenous Retroviruses

As a natural step in their replication cycle, retroviruses must integrate into the genomes of the host cells they infect (28). Exogenous retroviruses can be transmitted between individuals in a population, i.e. horizontal transmission, or from one species to another, i.e. cross-species transmission (zoonosis). Through infection of the germ line, exogenous retroviruses can become a part of the host genome, leading to the generation of endogenous proviruses. Endogenous retroviruses (ERVs) are transmitted genetically from parent to offspring in a Mendelian fashion. All vertebrate species examined carry remnants of such prior retroviral infections in their genomes (28). Humans, for instance, carry some 80,000 sequences, or about 8% of our total genome, derived from retrovirus infections dating from some 40 million to a hundred thousand or so years ago (161).

Retroviruses can be deleterious for the organism they infect through various mechanisms, including insertional mutagenesis, deregulation of gene expression, and acquisition of oncogenes. The potentially detrimental effects of such infections are evidenced by the evolution of multiple lines of host defenses against invading

retroviruses. However, there exist numerous cases of millions of years of host-virus coevolution, where endogenous retroviruses have been "domesticated" to serve a useful function for the host, ranging from fending off infection by other retroviruses, to increasing the expression of a beneficial gene, thereby providing a novel function within the host through the acquisition of retroviral sequences. Host-virus coevolution is a most curious process, implicating endogenous retroviruses in an active role in the process of acquisition of new traits, rather than a passive bystander.

Endogenous Retroviruses of Mice

Like the human genome, nearly 40% of the mouse genome is composed of transposable elements, with ~10% of the genome derived from mouse endogenous retroviruses (mERVs). Mouse endogenous retroviruses have been historically grouped into three broad classes based on their sequence similarity to exogenous retroviruses. ERVs that group with the genera alpha-, beta-, delta-, and lentivirus in phylogenetic analyses belong to Class I; those that group with gamma- and epsilonretrovirus belong to Class II; and those that group with spumavirus belong to Class III (Fig. 6; for a review, see (144)).

Class I ERVs are the smallest class of mERVs, making up 0.7% of the mouse genome. Class I ERVs are closely related to gammaretroviruses, and include murine leukemia viruses (MLVs), a virus family that will be explored in detail in this report, and others including MuRRS, MmERV/VL30, McERV, MuRV-Y, MuERV-C elements, some of which contain extensive deletions in their genomes. Class II ERVs cover 3.14% of the mouse genome and include the mouse mammary tumor virus (MMTV), MusD and

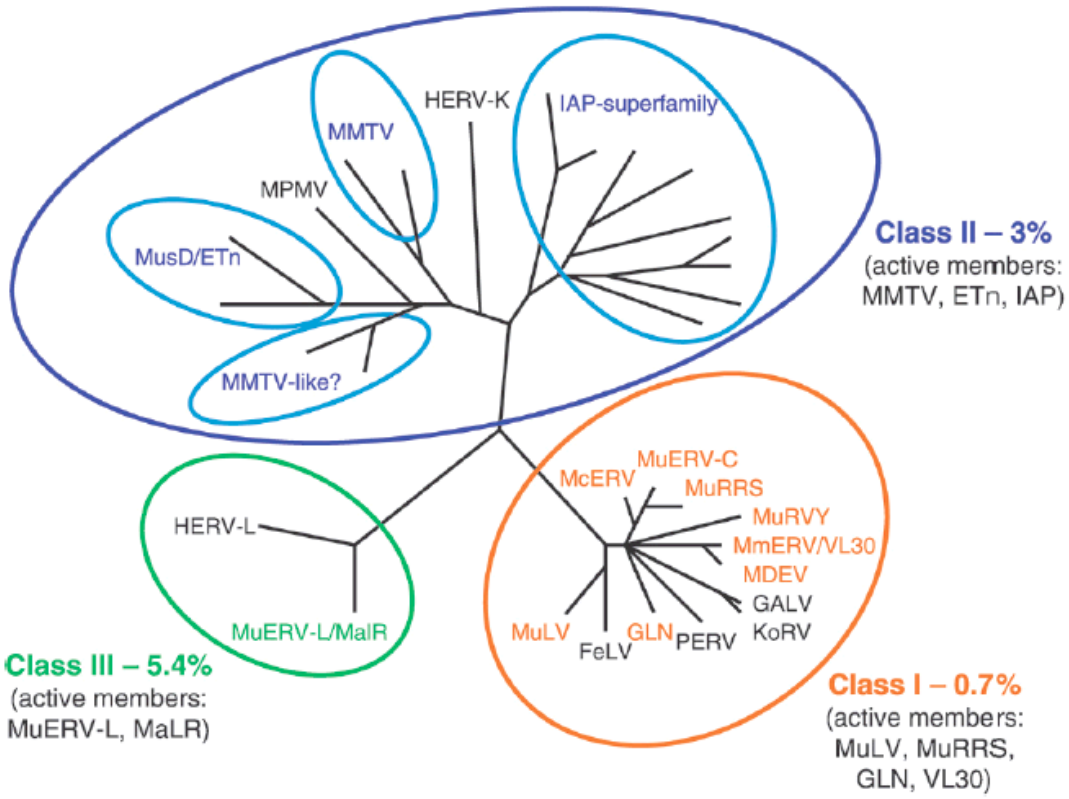


Figure 6. Phylogenetic analysis of RT domains of mouse endogenous retroviruses. Retroviruses from other organisms are included for comparison. Different retroviral classes are indicated with different colors. The values next to the class names indicate the percentage of the mouse genome that is occupied by each class. "Active members" refers to retroviral elements that are actively acquired, and not necessarily replication-competent. Figure from (144), with permission.

ETn (early transposon) elements, and intracisternal-A type particles (IAPs). Class III ERVs make up the largest class, which collectively make up about 5.4% of the mouse genome and include the mERV-L and MuERV-L/MalR transposable elements (Fig. 6). MalR elements are the most commonly found retroviruses in the mouse genome, and their LTRs share ~50% identity with that of mERV-L elements (100).

Mouse Phylogeny

The laboratory strains used today are largely derived from a mixture of three *Mus musculus* subspecies: *M. m. domesticus, M. m. musculus,* and *M. m. castaneus* (57). The latter two subspecies have interbred extensively in the wild, giving rise to the subspecies now referred to as *M. m. molossinus* (Fig. 7A). Besides these most recently derived subspecies, other, more distantly related, *Mus* subspecies exist, in different geographical locations (Fig. 7B). Colonization of the Americas and Australia by European mice (*M. m. domesticus*) has occurred after human travel between these continents became common.

Despite the extensive use of inbred laboratory mouse strains as model organisms for a plethora of biological systems, the genetic variation seen among inbred mice is very limited. Haplotype analyses show that inbred strains are largely derived from *M. m. domesticus*, with limited but variable contributions from *M. m. musculus,* and *M. m. castaneus* (183). Interestingly, the genomes of many wild-derived strains that were bred in the laboratory show haplotypes associated with inbred laboratory strains, suggesting unintentional crossbreeding (183). Such accidental contamination with inbred strains has been previously reported for several widely used strains (182), and can result in the appearance of new proviruses over time, due to genetic heterogeneity.

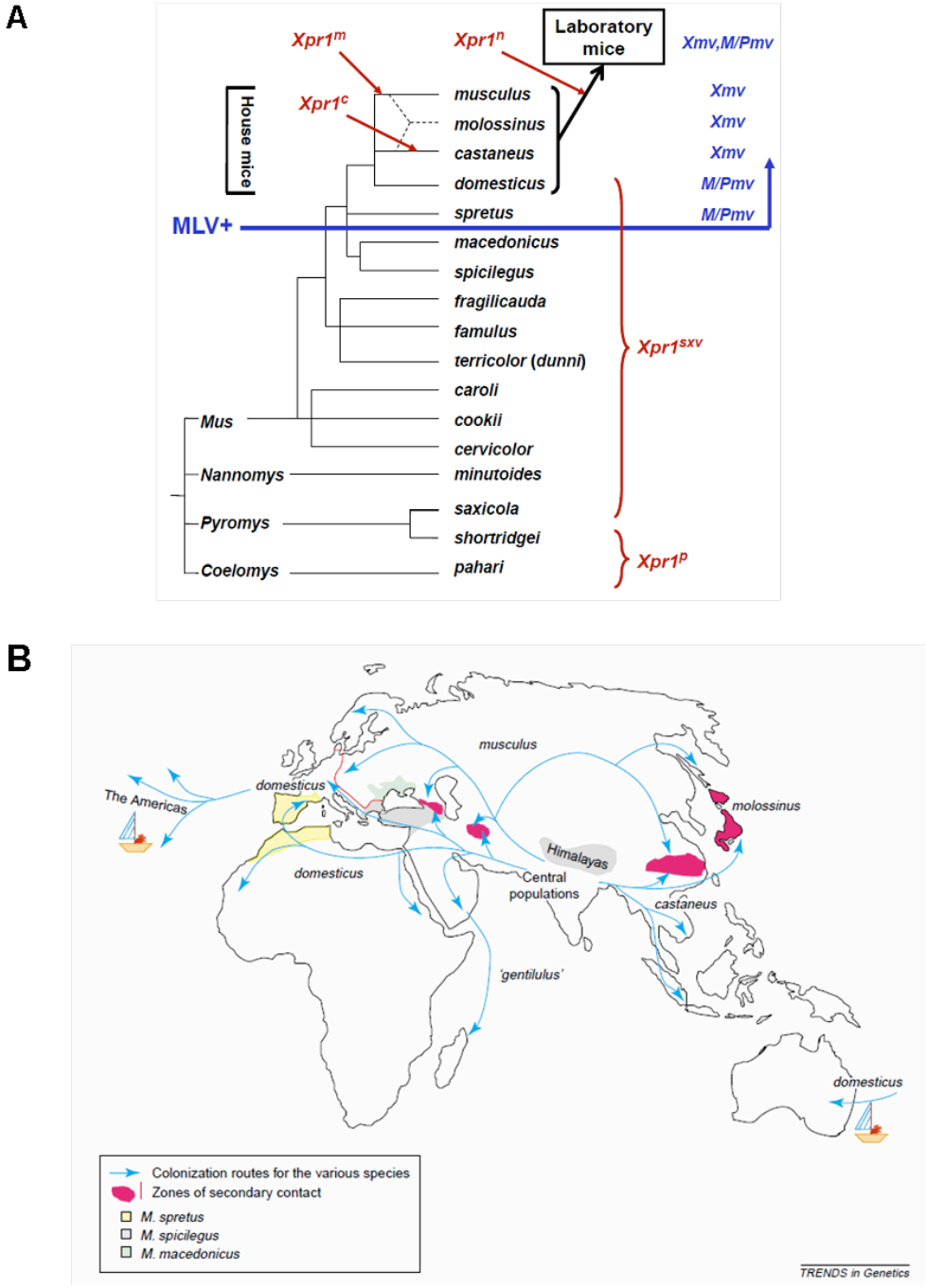


Figure 7. The distribution of Xpr1 alleles and MLVs in mice

(A) Evolutionary tree of the genus *Mus*. Modern laboratory strains are largely derived from *M. m. domesticus*, with variable contributions from *M. m. musculus* and *M. m. castaneus*. The distribution of nonecotropic MLV subgroups and known Xpr1 alleles are shown. Figure from (88), under the BioMed Central Open Access license agreement. (B) The geographical distribution of *Mus* subspecies. Figure from (57), with permission.

Gammaretroviruses

The gammaretrovirus genus includes a large number of retroviruses with the ability to infect diverse organisms from reptiles to primates. Gammaretroviruses are most closely related to the Class I ERVs (Fig. 6). All members of this genus identified so far use structurally similar but otherwise unrelated small molecule transporters with multiple transmembrane domains (for a review, see (150)). Mouse genomes contain a large number of more recently integrated endogenous retroviruses, dating from less than 1.5 million years ago, named murine leukemia viruses (MLVs), and the endogenization process is still continuing.

Many gammaretroviruses can and have jumped species, a few examples being gibbon ape leukemia virus (GaLV), simian sarcoma-associated virus (SSAV) and koala retrovirus (KoRV). GaLV and SSAV are likely to share origins with endogenous MLVs found in Southeast Asian mice (91, 144). Interestingly, KoRV is currently in the process of being endogenized in koalas living in Australia, brought to the continent very recently by human migration (153). Also thought to share origins with one or more MLVs in wild mice based on sequence, KoRV causes lymphoid malignancies in native koala populations (41).

Murine Leukemia Viruses

Host Range

Some of the best-studied viruses in the gammaretrovirus genus are murine leukemia viruses (MLVs). MLVs can be classified based on their host range and genomic sequence into ecotropic and nonecotropic classes. Ecotropic MLVs can only infect cells of murine origin. Non-ecotropic MLVs are further classified into xenotropic, polytropic

and modified polytropic subgroups; xenotropic MLVs are able to infect cells of non-murine origin only, while polytropic and modified polytropic MLVs can infect both murine and non-murine cells, and can be distinguished by a 27 bp insertion in the polytropic *env* gene. Amphotropic MLVs with a broad host range have been identified in some mice from a specific geographical location near California, although no known naturally occurring endogenous members of this subgroup exist (28).

Mapping and Distribution

Endogenous retroviruses provide a fossil record of prior infections, which are extremely useful as genetic markers, since they are passed on from parent to offspring and are easy to detect using standard laboratory techniques. In case of mice and MLVs, genomic DNA from many mouse strains can be analyzed for their MLV insertions by Southern blotting or the "unblot" technique previously developed in our laboratory (45, 46). An unblot is similar to a Southern blot, where genomic DNA is digested with a restriction enzyme, and run on an agarose gel. The gel is then dried down and hybridized with radiolabeled probes directed against parts of the retrovirus, allowing the detection of individual provirus fragments in the form of specific bands with characteristic sizes for each provirus. Alternatively, if the integration site is known, a simple PCR of host-virus junctions can determine whether an individual mouse has a specific provirus insertion. Previous work from our laboratory mapped a large number of MLV insertions with LTR- and *env*-specific probes in many laboratory mice, (43-46, 145, 146). In addition, using LTR-specific probes directed against the different U3 structures of MLV LTRs, the timing and location of insertions of various LTR subgroups were analyzed in wild mice,

with potential implications for the evolution and distribution of MLVs in wild mouse strains (Fig. 8) (43, 157, 158).

The availability of the mouse genome sequence made it possible to do database searches of published genomes for the presence of proviruses. One such study employed an *in silico* data mining approach, where the C57BL/6 mouse genome was searched using several probes specifically designed to detect the *env* genes of endogenous polytropic (Pmv), modified polytropic (Mpmv) and xenotropic (Xmv) proviruses (74). 49 such nonecotropic MLV proviruses were identified and mapped to the mouse genome, which allowed further analysis of these insertions for various properties such as complete provirus sequence, integration site, genome length, PBS type, target site duplication, and Apobec3 editing (74). In phylogenetic analyses, Pmv and Mpmv proviruses each formed a distinct monophyletic clade with a common ancestor, while Xmv proviruses failed to do so (Fig. 9), consistent with the notion that Xmv insertions may represent the oldest subgroup among nonecotropic MLVs, and that they have undergone more rounds of replication cycles, compared to Pmv and Mpmvs, prior to colonizing the mouse genome. Mutations consistent with Apobec3 editing were found for Pmv and Mpmv sequences, suggesting potential inactivation by this restriction factor, but this was not the case for Xmv proviruses. Therefore, the evolution of Xmvs seems to have followed a different path than those of Pmv and Mpmvs, perhaps by a mechanism of evading host restriction, or by colonizing species that failed to express considerable levels of such factors (74).

Receptor usage

MLVs, like other gammaretroviruses, use various types of cellular small molecule transporter molecules as receptors. Ecotropic MLVs use the mouse cationic amino acid

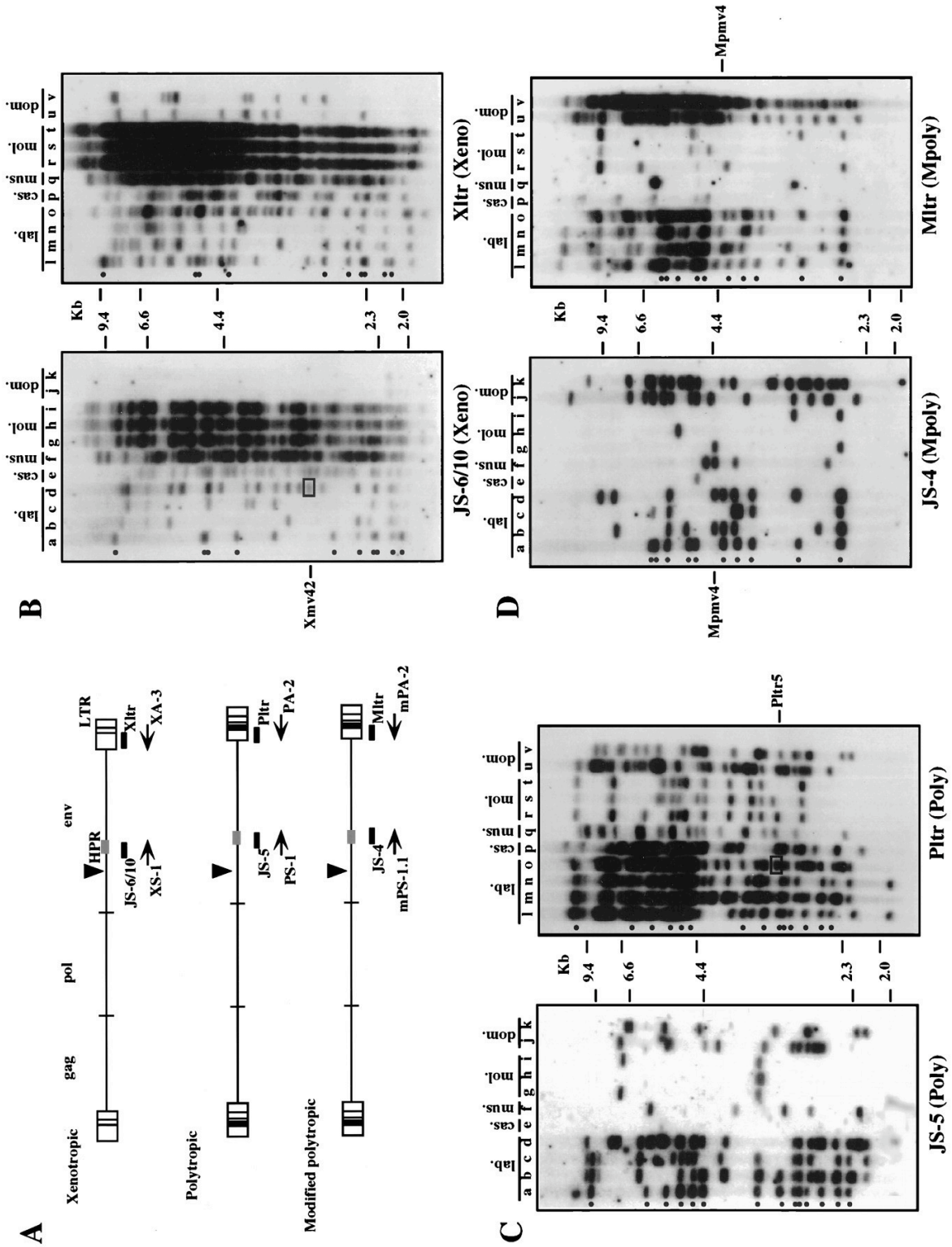


Figure 8. Distribution of nonecotropic MLV proviruses in wild mice

The unblot technique is described in Materials and Methods. (A) Location of *env* and LTR probes. Detection of (B) Xenotropic (Xmv), (C) Polytropic (Pmv) and (D) Modified polytropic (Mpmv) MLV fragments by unblots. Figure from (157), with permission.

transporter, (mCAT1) (4), while a few others with similar host range use the murine sodium-dependent myoinositol transporter (mSMIT1) (63, 156). Amphotropic MLVs use the sodium-dependent phosphate symporter, Pit2 (104). Polytropic and xenotropic MLVs use the molecule Xpr1, the mammalian homologue of the yeast *Syg1* gene, whose normal cellular function is as yet unknown (10, 151, 184).

As is the case with most host-pathogen arms races, evolution at the host-virus interface for MLVs and their mouse hosts is apparent on multiple levels. Xmv proviruses are considered to be the oldest MLV subgroup, a hypothesis supported by the genetic diversity among members, their failure to form a monophyletic clade and their ancestral location in phylogenetic analyses (Fig. 9) (74). During the long course of their coexistence with endogenous and exogenous MLVs, many *Mus* subspecies have evolved ways to cope with such assaults. One example of a resistance mechanism is the evolution of variants of the Xpr1 receptor so that the modified allele no longer supports virus entry, conferring a selective advantage on the host that carries the new variant. Endogenous X-MLVs derived from proviruses in those mice cannot infect the cells of their host organism, hence earning them the name xenotropic. Viruses have in turn responded by incorporating mutations in their *env* genes, allowing them to use the new version of the receptor and giving rise to the evolution of polytropic and modified polytropic MLVs, which are also carried in numerous copies by many mouse genomes. Polytropic Env proteins can infect the cells of many murine and non-murine species.

There are currently five known functionally distinct Xpr1 variants found in mice from diverse geographical locations, with differential susceptibilities to xenotropic and polytropic MLVs (88), suggesting that the overall evolutionary history is much more

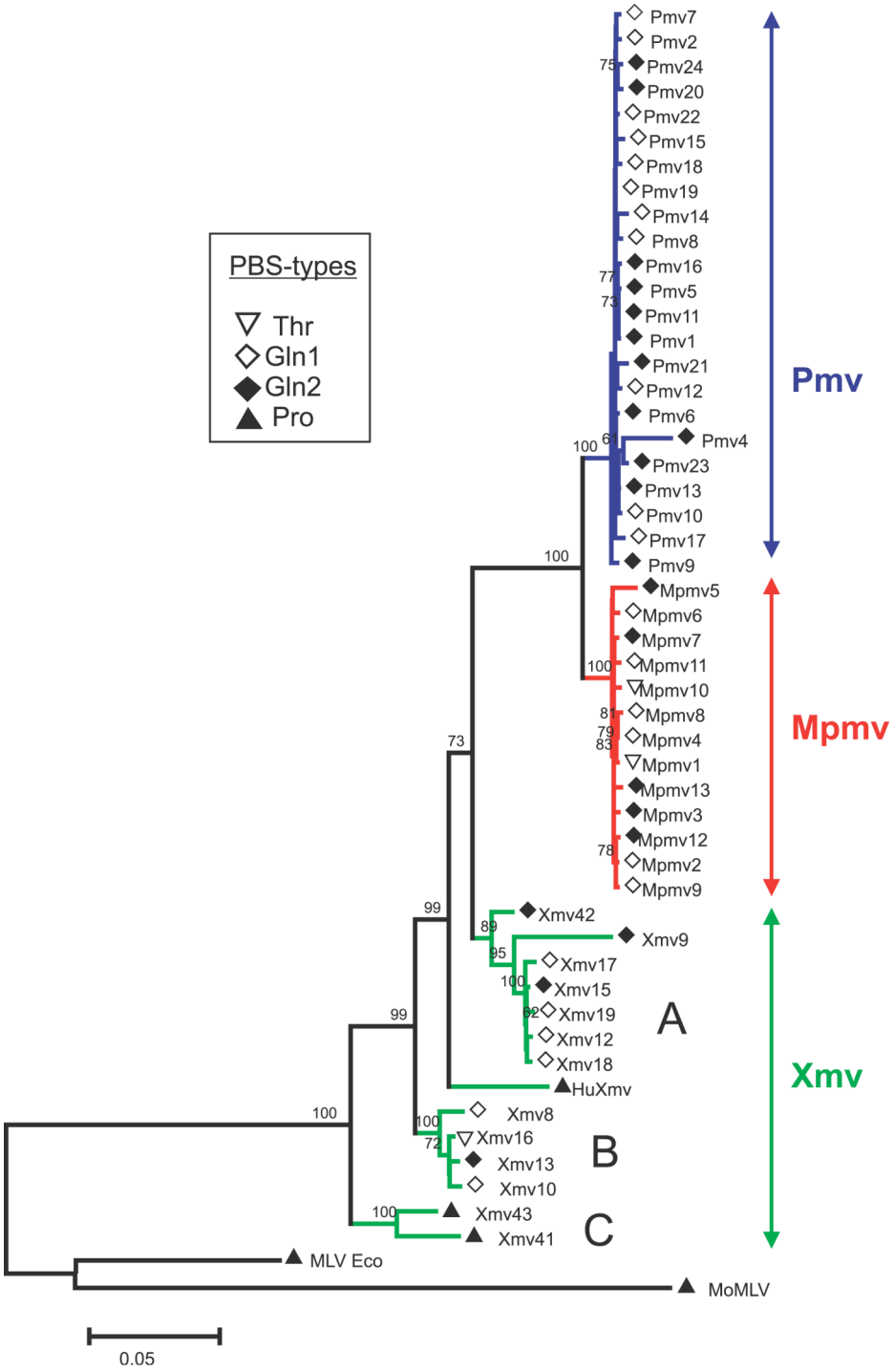


Figure 9. Phylogenetic analysis of nonecotropic MLVs in the C57BL/6 genome

Proviruses were detected using probe sequences specific for distinct subclasses of nonecotropic MLV *env* genes in BLAT searches, and complete proviral sequence information was used for phylogenetic reconstruction. Figure from (74), under the Creative Commons Attribution License.

complex. After the cloning of the mouse Xpr1 gene, several studies have mapped the domains and amino acid residues of the receptor critical for binding to the retroviral Env, using mutational analyses and chimeric Env proteins (99, 165, 180). Fig. 10 shows known Xpr1 variants present in geographically distinct populations of mice. The extracellular loops (ECL) 3 and 4 harbor the major determinants for viral entry; the important residues are highlighted and the differences between different alleles are shown. The most permissive allele is $Xpr1^{sxv}$, the most common variant among several *Mus* subspecies, while the other four restrictive alleles ($Xpr1^{n}$, $Xpr1^{p}$, $Xpr1^{c}$, $Xpr1^{m}$) have differential abilities to restrict or mediate entry of different MLV Env types (Fig. 10). Two of the restrictive alleles $Xpr1^{c}$ and $Xpr1^{m}$, parallel the species distribution of Xmv proviruses, suggesting they might have conferred a selective advantage to the mouse hosts (181). The only Xpr1 allele to fully restrict xenotropic MLV entry is $Xpr1^{n}$, which is only found in laboratory mice to date (181). The origin of this specific allele remains unknown, as it has not been detected in any wild mouse species yet.

Pathogenesis

Retroviruses have a long history of pathogenesis, having been associated with various types of malignancies, immunodeficiencies and neurological abnormalities. In fact, the first retrovirus was discovered in relationship to cancer, when an avian retrovirus was isolated from a tumor in chickens (127). There exist several mechanisms by which retroviruses can induce malignancies, including disruption of tumor suppressor genes, insertional activation of oncogenes and through acquisition of oncogenes from the host cell.

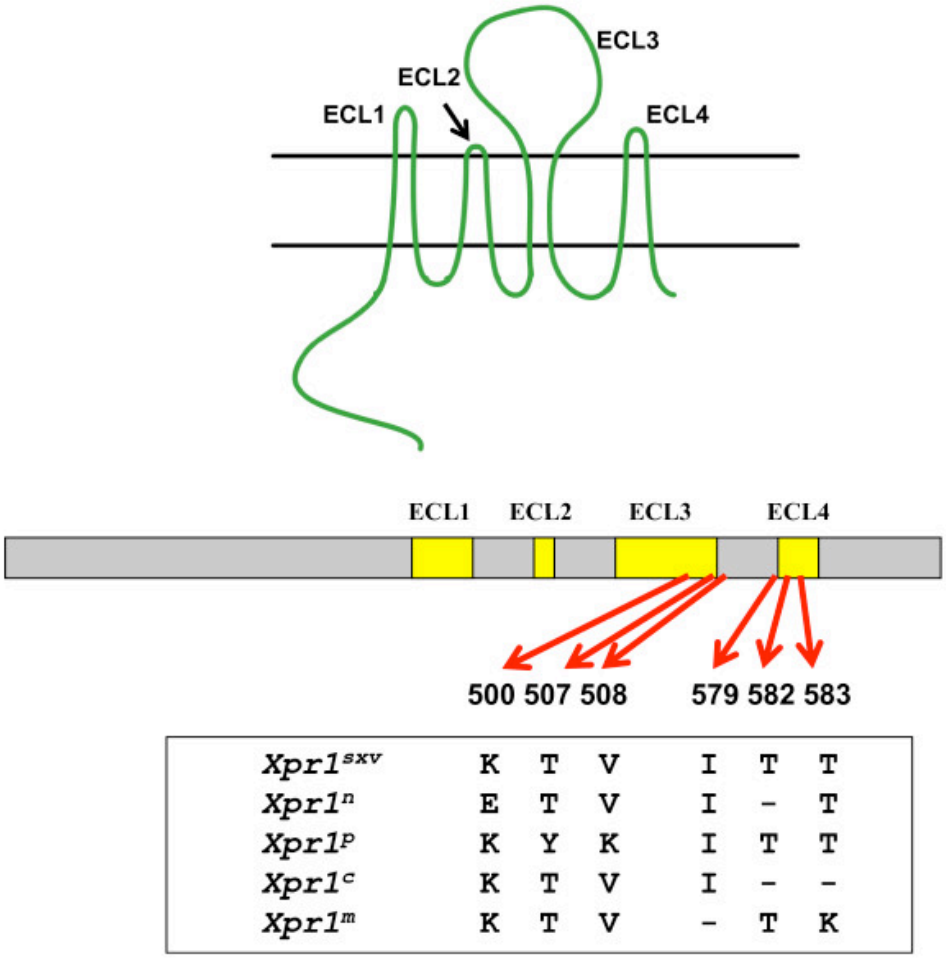


Figure 10. Mouse Xpr1 Variants

Putative structure of the Xpr1 cell surface receptor (upper panel) and the different mouse Xpr1 alleles found in distinct populations of mice. These alleles differ in their ability to support infection by xenotropic and polytropic MLVs, as well as two wild mouse isolates (CasE#1 and Cz524). $Xpr1^{sxv}$ stands for "susceptibility to xenotropic virus," fully permissive for X/P-MLVs, while other alleles are resistant to two or more of the mentioned viruses. $Xpr1^{n}$ is only permissive to P-MLVs. Critical amino acid residues important for Xpr1 function and host range are indicated, along with their positions. Figure from (88), under the BioMed Central Open Access license agreement.

Endogenous retroviruses are carried as part of the genome of the host species and are subject to the same forces of evolution as their host, resulting in the slow accumulation of mutations over time that can eventually render them inactive. However, there are intact endogenous proviruses that can be activated to produce infectious virus, through external stimuli, such as cellular stress and chemicals (1, 90), or through recombination with other endogenous or exogenous viruses (89). A well-characterized example of producing infectious retroviruses via recombination is the generation of highly pathogenic mink cell focus-forming (MCF) viruses that cause spontaneous thymic lymphomas in AKR mice (61, 147). A series of recombination events, whereby an endogenous Emv acquires the *env* gene from an endogenous Pmv, together with the U3 region of LTR from an Xmv, leads to the production of highly infectious recombinant viruses with a broad host range and increased oncogenic potential (61, 147), and demonstrates the potential of endogenous viruses to become mobilized.

Restriction Factors

Recent discoveries regarding the close interaction of retroviruses with their hosts have uncovered a complex evolutionary history. In large-scale screening studies, numerous host factors have been identified, which HIV-1, and possibly other retroviruses, depend on for successful infection of the host (17, 87, 188). We will not discuss the characterization and possible mechanism of action of these genes. On the other hand, several cellular genes that inhibit various steps in the retroviral replication cycle have been identified, in humans, other primates and mice. Such molecules are generally referred to as “restriction factors,” some of which are constitutively expressed in cells, allowing an intrinsic line of defense against retroviruses, while others are

induced by type I interferons (IFN), forming part of the immune system. The evolution of some restriction factors, especially in the coding regions critical for their function, have occurred extremely rapidly compared to what would be expected from the natural rate of evolution, suggesting that they have been under strong positive selection enforced by the continuous exposure of the host organism to infectious pathogens. We discuss four of these restriction factors below.

Fv1

One of the first clues regarding the presence of a host factor acting against retroviral replication came from studies of the susceptibility of inbred laboratory strains to leukemia caused by Friend virus (111). Among several such mouse genes identified that confer resistance to retroviruses, the Fv1 gene (for Friend virus susceptibility) was of particular interest, since its activity was dominant and seemed to act directly on the incoming virus, blocking a step in the replication cycle after reverse transcription and before integration (76).

There are two major alleles of Fv1: the $Fv1^{n}$ allele, carried by NIH Swiss mice, poses a block to infection by B-tropic MLVs, and $Fv1^{b}$, carried by BALB/c mice, impairs infection by N-tropic MLVs (60). This blockade is not absolute, as it can be saturated by high multiplicity of infection. Further studies described two more alleles of Fv1: $Fv1^{nr}$, which blocks B-tropic as well as some N-tropic MLVs, and the null allele, $Fv1^{0}$, with no restriction activity against either type of MLV. The Fv1 genetic locus was cloned in 1996 and was found to be derived from the *gag* gene of an endogenous retrovirus related to the MuERV-L family of endogenous retroviruses (13).

The target of Fv1 is the retroviral capsid, acting at a step after reverse transcription but prior to integration, although the exact mechanism of restriction has not been elucidated. The amino acid residues at positions 109 and 110 determine susceptibility or resistance of the retrovirus to the activity of Fv1, with position 110 being the critical determinant (33, 113). Through mutational analyses, Fv1 domains responsible for restriction activity have been mapped to the N- and C-terminal ends of the Fv1 gene product, as well as specific residues in the major homology region (15).

TRIM5α

HIV-1 is unable to infect many old world monkey cells due to a post-entry block (65). The gene responsible for this dominant restriction phenotype was identified as TRIM5α (148), which also poses an early post-entry block observed in many human and non-human primate cells (12, 32, 62, 159). TRIM5α directly recognizes the incoming retroviral CA, at the same residue (position 110) as Fv1, conferring specificity to the restriction phenotype (117). TRIM5α restriction of retroviruses occurs prior to completion of reverse transcription (177). Although the precise mechanistic details of restriction by TRIM5α have not been worked out, there is evidence that proteasome-mediated degradation plays an important role (19, 148).

APOBEC3G

Retroviral accessory genes can sometimes be dispensable for viral replication in certain tissue culture conditions; however they are strictly retained in the natural course of infection. For instance, the HIV-1 accessory protein "viral infectivity factor" (*vif*) is dispensable in some tissue culture systems, but is strictly maintained in natural infection.

The observation that *vif*-deficient strains of HIV-1 could replicate in certain cell lines (i.e. permissive cells), but not in other cell lines or primary cells (i.e. non-permissive cells) suggested the presence of a host factor which is differentially expressed between permissive vs. non-permissive cells (48, 139, 170). Heterokaryons produced between permissive and non-permissive cells were restrictive towards the replication of *vif*-deficient HIV-1, suggesting the presence of a dominant factor that inhibited replication in non-permissive cells, which is overcome by the Vif protein (95, 139). Using a cDNA subtraction strategy, which allowed the comparison of permissive and non-permissive cells, this host factor was identified as APOBEC3G (formerly CEM15) (134).

APOBEC3G is a cellular cytidine deaminase that acts on single stranded RNA. The name stands for "apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G." There are 11 APOBEC members in humans, with variable editing abilities (59). The ability of APOBEC3G to block *vif*-deficient HIV-1 occurs through two distinct mechanisms: 1) cytidine deaminase activity-dependent negative strand RNA editing (98) and 2) deaminase-independent mechanism (108). APOBEC3G does not specifically act on HIV-1, as it was shown to have broad activity against several retroviruses (98).

Tetherin

Studies with the HIV-1 accessory protein "Viral Protein U" (Vpu) allowed the identification of yet another antiviral restriction factor. Similar to the findings leading up to the discovery of APOBEC3G, the inability of HIV-1 defective for the *vpu* gene to replicate in certain human cells suggested the presence of an antiviral block (52). The restrictive phenotype was dominant, as shown by heterokaryons, pointing to the direction of an inhibitory factor (167). The phenotype seen in *vpu* deleted strains of HIV-1 was

seemingly mature virus particles, which bud from the plasma membrane but could not be released (106). The finding that protease treatment could free these virions implicated a cell surface protein that tethered formed virions to the plasma membrane. This factor was identified in 2008, given the name tetherin (aka BST2/CD317, with unknown function) (107), and showed broad activity against not just retroviruses, but other virus families as well (77), suggesting it evolved as a common antiviral defense mechanism.

INTRODUCTION-II: XMRV

Discovery of XMRV

Prostate cancer is one of the leading causes of illness and death among men. Family history of prostate cancer remains the greatest risk factor for this disease, highlighting the importance of genetic factors. In fact, numerous studies have established the association of a large number of genes, loci and polymorphisms with susceptibility to prostate cancer, although the significance or potential contribution of many of these candidates remain unresolved (109). One of the most well-established genes for susceptibility to hereditary prostate cancer (HPC) was mapped to the long arm of human chromosome 1 through genome-wide linkage analysis and given the name HPC1 (141). Cloning of the exact genomic locus followed, and the potential gene was identified as RNASEL (20).

RNASEL is an important interferon-stimulated antiviral factor in the innate immune response. Several mutations in the RNASEL gene have been associated with susceptibility to hereditary prostate cancer, including a low-activity variant caused by an Arg to Gln substitution at residue 462 (22). The fact that a low activity variant of a gene important for the antiviral pathway is linked to prostate cancer suggested the potential involvement of a viral agent. In an effort to establish a relationship between RNASEL activity and the development of prostate cancer, Urisman et al. analyzed primary tissues from prostate cancer patients with different RNASEL variants (163). Using a DNA-based microarray (the "Virochip") that contains well-conserved regions from ~950 sequenced virus genomes, and hybridizing PCR-amplified sequences from the tissues by random primers, the presence of a gammaretrovirus was indicated in a subset of prostate cancer

cases with the low-activity variant (R462Q) of RNASEL. Stretches of viral sequences were then amplified from prostate tissues and a full-length viral genome was obtained by assembling individual pieces of the genome. The virus was christened Xenotropic Murine Leukemia Virus-Related virus (or XMRV, for short) to connote its close relationship with xenotropic MLVs (163).

XMRV Genome Structure

The XMRV genome bears close resemblance to MLVs, with up to 95% identity along the entire length of its genome. The most closely related group is xenotropic MLVs denoted "X-MLV" for exogenous viruses, and "Xmv" for endogenous viruses (Fig. 9). However, the genome of XMRV is still clearly distinct from any MLV identified so far. It is found neither in the sequenced C57BL/6 mouse genome, nor among any of the endogenous or exogenous MLV sequences published in GenBank. The high percent sequence identity between XMRV and MLVs undoubtedly pointed to an origin in mice, and a potential cross-species transmission from mice to humans.

The full-length XMRV genome is 8185 nucleotides long, with intact open reading frames and a xenotropic *env* gene. As is the case with other MLVs, *gag-pro-pol* genes are expressed via translational read-through of the stop codon at the end of *gag*, encoding a polyprotein, which is later cleaved to give rise to individual viral proteins. The alternative "CUG" start codon in the *gag* leader region of XMRV is followed by a stop codon, which prevents the synthesis of the glycosylated Gag (glycoGag) protein (163).

Association of XMRV with Prostate Cancer

XMRV was first identified in human prostate cancer tissues. Parts of the viral genome were amplified by PCR, and later assembled into a full length clone (163). To provide further evidence of XMRV infection of primary human prostate tissues, XMRV DNA was detected by fluorescence in situ hybridization (FISH) and XMRV proteins were detected by immunohistochemistry (IHC) with a monoclonal antibody against SFFV Gag protein in the stromal cells surrounding primary prostate cancer tissues (163).

A second study demonstrating XMRV infection in human tissues and its association with prostate cancer came from histological analyses of prostate resection specimens (133). 334 prostate specimens were examined using quantitative PCR (Q-PCR) and immunohistochemistry (IHC) with anti-XMRV antiserum; a positive signal was detected in 6-23% of the samples, respectively (Table 1). Protein expression was primarily detected in malignant epithelial cells, suggesting potentially a direct link to tumorigenesis. Unlike the previous study, no association with the RNASEL genotype was found (133).

A third study used three different approaches to assess XMRV infection in patients with prostate cancer: PCR, FISH and serological detection of neutralizing antibodies (8). 11 out of 40 patients had neutralizing antibodies against XMRV. Seven samples had enough biological material to perform the three assays, which gave "unequivocal" results: five were positive, while two were negative, using all three assays (8). The use of multiple methods to detect XMRV infection in human samples was important, due to the lack of consensus in results from different labs after several reports were published (see the section: Controversies Regarding Disease Association).

Association of XMRV with Chronic Fatigue Syndrome

Chronic fatigue syndrome (CFS) is a debilitating disease of unknown etiology, characterized by extreme fatigue and a combination of several other symptoms, including pain, cognitive dysfunction and immune abnormalities. The diagnosis for CFS is often based on exclusion of other conditions, and there is no known biomarker for this disease, although several sets of international criteria have been established (21, 47). The absence of a biomarker, combined with the heterogeneity of the symptoms among different individuals, for many decades, has complicated the establishment of a unified definition for this disease, or a consensus regarding viable treatment options.

In 2009, three years after the identification of XMRV in human prostate cancer, the same virus was detected in blood cells of 67% of patients with CFS compared to only 3.7% of healthy controls (94). Several lines of evidence supported this finding. First, PCR on peripheral blood mononuclear cell (PBMC) DNA of CFS patients revealed that 68/101 (67%) contained the XMRV *gag* sequence. Second, viral proteins were detected using different antibodies or antisera, including: i) a rat monoclonal anti-SFFV Env antibody (175), ii) goat antisera against whole mouse NZB X-MLV, and iii) a rat mAb to MLV p30 (24). In an intracellular flow cytometry (IFC) assay, 19/30 CFS samples reacted with the anti-MLV p30, but none of the 16 healthy control samples did, and the results were confirmed by Western blots (94). Third, co-culturing activated lymphocytes taken from patients and LNCaP cells, a prostate cancer cell line highly susceptible to viral infections due to a defect in the IFN pathway (36), revealed the presence of infectious virus that could be transmitted via both cell-associated and cell-free routes. Fourth, 9/18 patient plasma samples contained antibodies that reacted with SFFV Env,

while none of the seven healthy donors did. Taken together, the findings pointed to a significant association of XMRV with CFS (94).

Although no causal relationship has been established between XMRV infection and the development of CFS symptoms, the detection of XMRV in such a high percentage of patients suggested that XMRV could at a minimum serve as a biomarker for CFS, given the variable definition and diagnostic criteria available for the disease. Furthermore, the presence of an infectious retrovirus in ~4% of the healthy population raised concerns about the safety of blood banks, leading to the organization of a multi-center study to address the issue (137).

XMRV Receptor Usage

Shortly after the discovery of XMRV, a full-length viral genome (clone VP62) was assembled from individual pieces amplified from prostate tissue RNA. In tissue culture experiments with LNCaP and DU145 prostate cancer cells, XMRV proved to be infectious and susceptible to the IFN-β antiviral pathway (34). The expression of the human xenotropic/polytropic MLV receptor (XPR1) was also required for infection, showing that XMRV is indeed a xenotropic MLV (34). Moreover, a rat fibroblast cell line expressing the rat Xpr1, which supports XMRV entry, was rendered more susceptible to XMRV infection when the human XPR1 was stably introduced (102).

Transforming Activity

The XMRV genome does not carry extra viral genes that have transforming activity; neither does it carry a host-derived oncogene. Investigation of transforming activity of XMRV revealed that expression of infectious virus in cultured fibroblasts and

epithelial cells only induced very rare transformed foci, consistent with occasional indirect transformation that sometimes occurs due to the activity of retroviral Env proteins (102). When XMRV was grown on HT-1080 fibrosarcoma cells, one of the few foci that formed was found to produce a transforming virus at high titers. Closer examination of the virus produced by this foci turned out to be XMRV that had acquired the Nras oncogene from the host cells (103), demonstrating once again the ability of retroviruses to acquire oncogenes from the cells they infect, resulting in high transformation potential.

Susceptibility of XMRV to Restriction Factors

Infectious XMRV was reported to be isolated from PBMCs of CFS patients (94), which are known to express restriction factors such as APOBEC3G (A3G), APOBEC3F (A3F) and tetherin, with known antiviral activity against a variety of viruses, including retroviruses (107, 134). The apparent replication of XMRV in PBMCs suggested that XMRV might have evolved a strategy to avoid host restriction, even though it is not known to encode any accessory proteins. Three studies investigating the susceptibility of XMRV to human and mouse restriction factors revealed that XMRV replication is, in fact, severely blocked by several host defenses (16, 56, 114).

Cell lines from different host species are highly variable in terms of supporting XMRV infection (56, 114). This variation does not depend on the variability in the level of XPR1 receptor expression in these cells, but rather the expression of various restriction factors. XMRV replication is restricted by human A3G, A3F and to some extent by A3B (56). It is also restricted by human tetherin, which has antiviral activity against a plethora of enveloped viruses (107). Moreover, XMRV does not seem to have evolved any means

of dealing with these host restriction factors, by acquiring accessory genes (like HIV-1 *vif*) or developing secondary functions for its essential genes (like HIV-2 *env*), making it highly unlikely that it would succeed in establishing a persistent infection in primary human tissues. Human TRIM5α does not have any effect on XMRV replication (56).

XMRV was also tested for its susceptibility to mouse restriction factors, even though it is not naturally found in any mouse species (56). Mouse Apobec3 and Tetherin can efficiently block XMRV. Interestingly, XMRV is inhibited by both $Fv1^N$ and $Fv1^B$ alleles, although its CA protein suggested that it might be B-tropic, based on the critical residue 110 (56). Restriction by both Fv1 alleles, to our knowledge, is unprecedented for an MLV.

In a second study by Paprotka et al., expression of human A3F and A3G, and mouse Apobec3 in virus producing cells severely inhibited XMRV infection by incorporating into virions and hypermutating the viral DNA (114). Moreover, analysis of various cell lines for the expression levels of A3F and A3G revealed that the prostate cancer cell lines 22R*v1*, DU145 and LNCaP, which support efficient XMRV replication, express much lower amounts of these restriction factors compared to the T-cell lines CEM and T9 (114).

Susceptibility of XMRV to Antiretroviral Drugs

There are currently close to 30 different drugs that are FDA-approved for antiretroviral therapy. With the potentially pathogenic status of XMRV for humans, drugs that were already available for other conditions were tested for their activity against XMRV in two studies. In a drug screening study, 45 compounds, 28 of which were previously approved for human use, were tested for their activity against XMRV

replication. XMRV is particularly susceptible to the integrase (IN) inhibitor raltegravir in a panel of human cell lines, including the MCF-7 (breast cancer) and LNCaP (prostate cancer) (140). A different integrase inhibitor (L-000870812) and two nucleoside reverse transcriptase (RT) inhibitors, zidovudine (AZT) and tenofovir (TDF), also blocked XMRV replication (140). Another study by Paprotka et al. analyzed a smaller set of 8 antiretrovirals, reaching the same conclusion that XMRV was most potently blocked by AZT, tenofovir and raltegravir (114). AZT was also found in a screen of 10 compounds by Sakuma et al. as a potential XMRV inhibitor (128).

Transcriptional Activity in Cell Lines

XMRV is an infectious virus; it can be grown in tissue culture following transfection with a full-length viral clone, or through incubating with supernatants from infected cells (23, 34, 124). The ability of XMRV to grow is highly dependent on the cell type used, however. While the levels of XMRV protein production and RT activity, indicating virus production, rise very rapidly in various prostate cancer cells, XMRV growth and spread in other cell types such as 293T are very poor (124). The difference in the amount of virus production could be explained by the differential XMRV LTR activity in prostate cancer cells compared to other cell types. Using luciferase reporter constructs, Rodriguez et al. showed that the XMRV LTR is extremely active in prostate cancer cells. The transcriptional activity of XMRV LTR in 293T cells could be increased to levels seen in prostate cancer cells by co-expressing androgen receptor (AR) and treating with an AR agonist such as dihydrotestosterone (DHT), suggesting that it is the transcriptional environment in these cells that made the XMRV LTR so active. The determinant for this transcriptional activation lies in the glucocorticoid response element

(GRE) in the U3 region of XMRV LTR: when a reporter construct with a mutated GRE sequence is used, the high transcriptional activity of the XMRV LTR in prostate cancer cells is lost (35, 124). These findings are consistent with the findings of Dong et al., where DHT treatment resulted in increased LTR activity in cells expressing AR, such as LNCaP cells, but not in DU145 cells, which do not express AR (35). In summary, the data suggest that the transcriptional environment in certain prostate cancer cells is very suitable for XMRV growth, and that XMRV LTR is very responsive to the effects of androgen in AR-expressing cells.

Integration Site Preference

Retroviral integration into the host genome exhibits only modest sequence specificity, but is not random. In an effort to investigate retroviral integration site targeting on a nucleotide scale, Holman et al. analyzed base preferences surrounding published integration sites of HIV-1, ASLV (avian sarcoma/leukosis virus) and MLV from experimental infections (69). For the three viruses examined, highly significant symmetrical base preferences were found at positions immediately surrounding the integration site, with the significance decreasing as the distance from the integration site increased (69).

Retroviral integration site selection is also affected by the many features of the host genome, including the presence of nucleosomes, genes, chromatin structure, DNase hypersensitive regions, bent DNA, GC content, and primary sequence preferences (reviewed in (18)). Studies analyzing nearly 900 integration sites of murine leukemia viruses in HeLa cells revealed a strong preference for integration near the transcriptional start sites of genes (178). XMRV being most closely related to MLVs, Kim et al.

investigated whether XMRV had the same integration site preference as other MLVs (82). DU145 prostate cancer cell line was infected with XMRV, and 472 unique integration sites were mapped to the human genome. The genomic locations of XMRV integration was consistent with a strong preference for regions with features associated with open chromatin, such as transcription start sites, CpG islands and DNase hypersensitive sites. A later study by the same group also showed 13 out of 15 clonal cell lines expanded from infected DU145 cells had the correct 4 bp target site duplication (TSD), typically observed in MLV integration (83). A weak, palindromic consensus sequence was also detected in the mapped XMRV integration sites (83).

A large-scale integration mapping study was conducted in primary CD4+ T-lymphocytes infected with XMRV or an MLV-based vector. Analysis of over 32,000 unique integration sites by 454 pyrosequencing revealed that integration was favored at the predicted nucleosome locations, near the region facing the major groove of DNA (126), suggesting that a preference for DNase I hypersensitive regions did not necessarily imply integration into nucleosome-free regions. Importantly, the integration site preference of XMRV does not show differences compared to other MLVs, essentially stating, once again, the fact that XMRV indeed behaves like an MLV.

Animal Models for XMRV

The pathogenic potential of XMRV for humans prompted research into creating animal models to study the properties of XMRV infection *in vivo*. An inbred laboratory mouse model for XMRV infection did not seem plausible for two reasons. First, most laboratory strains of mice carry the $Xpr1^{n}$ receptor variant that is non-permissive to XMRV infection (88). However, a large-scale analysis of wild-derived strains revealed

that many wild mouse strains carry different Xpr1 alleles, some of which support XMRV entry. Second, MLV proviruses found in the genomes of inbred strains would most likely complicate the analysis during the course of infection, as endogenous proviruses could be expressed and/or mobilized by XMRV. Some wild mice strains distantly related to inbred laboratory strains carry either very few or no endogenous MLVs. For these reasons, the usage of wild mouse strains as an animal model for XMRV became an attractive possibility.

Sakuma et al. examined the potential of the wild-derived strain *Mus pahari* to serve as a small animal model (130). After showing that the *M. pahari* cells were susceptible to infection by XMRV, adult and neonatal mice were injected with XMRV. Proviral DNA was detected in the blood cells and in the spleen and brain. Moreover, XMRV-specific neutralizing antibodies could be detected in the plasma from infected mice. Viral genomes isolated from infected mice showed extensive G to A hypermutation, consistent with previous data regarding the mouse Apobec3-based restriction of XMRV (16, 56, 114).

Another study to create an animal model for XMRV infection employed Indian rhesus macaques (112). XMRV was infectious in the macaques tested, establishing a persistent infection in specific organs and tissues after intravenous injection, even though the level of cell free virus was mostly below the limit of detection after the initial phase of viremia. Strikingly, XMRV was targeted preferentially to prostatic epithelium and the reproductive tract, consistent with its suspected role in the development of prostate cancer (112).

Detection of Integration Sites in Human Samples

One of the most compelling pieces of evidence to demonstrate infection of human cells by a retrovirus is to identify integration sites in the genomic DNA from the infected tissue. To do this, Kim et al. examined XMRV-positive prostate tumor tissues from biopsies, and cloned the junction between the human genome and the provirus, providing further support to the previous data on infection of human tissues (83). 14 XMRV integration sites in the human genome were cloned.

XMRV integration sites from human tissues were cloned in the same study that determined hundreds of integration sites from experimentally infected DU145 cells (82). Surprisingly, a study by Garson et al. later showed that at least 2 out of these 14 integration sites were in fact derived from XMRV infected DU145 cells (49). The two integration loci were identical base-to-base to the loci mapped in tissue culture experiments, and since it was statistically close to impossible that integration would occur at exactly the same spot in the human genome, the flanking host sequences were most likely derived from DU145 cells (49). While this finding did not necessarily rule out the authenticity of the remaining 12 clones, it pointed to the possibility of PCR contamination for those sites as well.

A Human Prostate Cancer Cell Line Produces XMRV

Association of XMRV with prostate cancer suggested a role for XMRV in the development of this disease, although XMRV had not yet been found in prostate carcinoma cells. Analysis of a human prostate cancer cell line, 22R*v1*, revealed the production of a gammaretrovirus particle by electron microscopy (84). Moreover, the

supernatant from 22R*v1* cells produced infectious virus that could rescue and transfer a defective retroviral vector, suggesting the presence of infectious virus. Further analysis of the viral genome revealed that it was identical to the XMRV sequences previously detected in patient samples (84, 114).

The 22R*v1* cell line was established by passaging a prostate tumor as xenografts in nude mice (105, 119, 142, 172). The production of XMRV by 22R*v1* cells was assumed to indicate that the original prostate cancer patient, whose tumor eventually gave rise to the cell line, was infected with this virus (84). Further studies from our laboratory proved this assumption to be incorrect (explained in the Results section).

Lack of Sequence Variation between Isolates

Retroviruses are rapidly evolving entities, due to the highly error-prone nature of reverse transcriptase enzyme and the selection forces from the host species that result in the amplification of mutants that can escape host restriction factors and immune responses, which the virus must overcome in order to successfully replicate. Such a selection process is particularly evident in the case of HIV infection, where treatment with a single antiretroviral drug results in the rapid selection of drug-resistant mutants, which were presumably already present among the virus pool in the individual prior to therapy.

Full-length XMRV sequences detected from prostate cancer patients (VP62, VP42 and VP35) (163) and CFS patients (WPI-1106, WPI-1178) (94) showed ~ 99% identity to each other. This was quite an unexpected result, considering the many rounds of reverse transcription the virus must have gone through in different individuals. Even though the viruses showed single nucleotide differences at various positions along their

genomes, the degree of variation among isolates was not entirely consistent with XMRV being a replicating and transmissible human retrovirus. In fact, analysis of XMRV sequences produced by 22R*v1* cells showed broader diversity compared to the sequences reportedly isolated from patients, suggesting that the 22R*v1* isolates must be ancestral to all clinical isolates (71). Although other studies that amplified XMRV fragments from 22R*v1* genomic DNA did not find the same degree of variation (Mary Kearney and Tobias Paprotka, unpublished results), the lack of variation among the clinical isolates still remained a significant concern. To put things in perspective, the diversity of viral sequences present in a single HIV infected individual is far broader than the diversity of XMRV sequences detected in different individuals, at different times, from different geographical locations (71).

Detection of MLV-Related Sequences in CFS Patients

The identification of XMRV in the cells and serum from a large fraction of CFS patients prompted numerous studies on samples previously collected from patients that were kept in tissue repositories. While almost all of these studies failed to find XMRV, one study by Lo et al. reported the detection of not XMRV, but other MLV-related sequences from the PBMCs of 32/37 CFS patients and only in 3/44 healthy controls (92). The results were assumed to imply that the findings of Lombardi et al. had been replicated, and that there were a number of variable MLV sequences potentially circulating in the human population. Unlike the Lombardi et al. study, however, the only method used to analyze samples was by PCR, and no full-length virus genome was amplified. No proteins, antibodies or infectious virus was recovered (92).

The sequences detected were quite variable compared to each other and to XMRV; they were either identical or had very high sequence identity to known endogenous mouse proviruses found in over a hundred copies per mouse cell, which immediately raised concerns about potential mouse DNA contamination. Although the authors used an assay to amplify mouse mitochondrial DNA to control for mouse DNA contamination and found none, the sensitivity of the assay and the discrepancy of mitochondrial copy number calculations caused the possibility of mouse DNA contamination to remain a valid concern. Numerous studies conducted following the Lo et al. study highlighted the rightful prudence of such concerns.

Mouse DNA Contamination of Laboratory Reagents

Despite the inclusion of apparently adequate controls in studies like that of Lo et al., mouse DNA contamination remains a significant concern (92). Extremely small amounts of DNA, from as little as one one-hundredth of a cell, contain enough provirus sequences to yield false positives with internal provirus specific primers. In fact, two recent studies have described the detection of MLV-like sequences in samples from patients and/or healthy controls, where every positive sample turned out to be positive for mouse genomic DNA (110, 123) (see Results). Several other studies further supported the idea of mouse DNA contamination, where potential sources of MLV genome contaminants, most likely from mouse genomic DNA, were discovered in commercially available laboratory reagents and kits, particularly *Taq* DNA polymerase containing a mouse monoclonal antibody (38, 71, 85, 131, 135, 162, 187). It has also been suggested that microtomes used for both laboratory and clinical samples could carry traces of mouse DNA over to human pathology samples, possibly including the fixed and archived

prostate cancer specimens analyzed by Robinson et al (123). Cross-contamination from a laboratory robot used for XMRV more than a year previously has also been reported (135). Taken together, these findings call for caution before interpreting the data and the need for very sensitive assays to detect mouse DNA contamination, when endogenous MLV sequences are detected by PCR from human samples.

Controversies Regarding Disease Association

The first reports describing the association of XMRV with prostate cancer and CFS seemed to rely on compelling evidence. The percentage of positives in disease cases were much higher than that of controls, full length viral sequences were PCR amplified (94, 163), XMRV DNA and protein were detected in primary human tissues (163), and infectious virus was recovered from the blood cells of CFS patients (94). However, subsequent studies failed to reproduce not only the association of XMRV infection with prostate cancer and CFS, but also its presence within the human population entirely (reviewed in (30)). Initial studies that failed to find the virus were approached with caution; rather than doubting the presence of XMRV in the human population, many potential explanations were provided, including geographical differences, description of the patient cohort and variation in the detection techniques used. When more and more studies reporting the absence of XMRV in large patient cohorts from around the world started to accumulate, however, the status of XMRV as a human pathogen was seriously challenged.

Table 1 lists all XMRV disease association studies published as of September 2011. The studies listed include those that analyzed the association of XMRV with prostate cancer, CFS, other diseases of poorly described etiology, and those that assess

the prevalence of XMRV in high-risk individuals, such as patients with compromised immune systems. The majority of such studies failed to replicate the findings of the initial reports, despite the use of numerous highly sensitive methods in large patient populations. Unfortunately, studies that yield data supporting the lack of association or of a virus in a given population are almost always less interesting to the public than positive studies. However, two very well designed replication studies by Shin et al. and Knox et al. merit further discussion, since they undoubtedly laid to rest a long-standing argument, which stated that the lack of detection of XMRV resulted from the variation in the detection techniques used or the selection criteria for the patient population (85, 135).

The only report that has claimed to find XMRV-positive CFS patients came from a research institute (Whittemore-Peterson Institute (WPI), Reno, Nevada) and the commercial branch of the same institute (VIPdx, Reno, Nevada) (94). Knox et al. used two patient populations (85): The first population (P1) consisted of the WPI cohort used in the original CFS association study, where banked samples were available for analysis and 26/41 samples had previously tested positive by WPI/VIPdx (94). The second patient population (P2) was selected by the history of reported XMRV positive status (26/29), again by WPI and VIPdx, but this time blood was freshly collected from the patients and analyzed. The analyses included PCR and RT-PCR, assays for infectious virus isolation and detection of virus-specific antibodies. No evidence for XMRV was found in any of the patient samples tested using all of the mentioned methods, which would have detected the virus if it were present. Interestingly, 19/41 patients in the first cohort had their blood drawn on the same day by the same phlebotomist; one tube of blood being sent to VIPdx, and the other tube to the Wisconsin Viral Research Group medical practice.

		Positive signal[1]			
Reference	**Disease**	**% Cases (n)**	**% Controls (n)**	**Tissue Type**	**Location[2]**
Urisman et al. PLoS Path. 2006	PC	10.5% (86)	–	prostate	USA
Fischer et al. J.Clin.Virol 2008	PC	0.95% (105)	1.4% (70)	prostate	Northern Europe
Schlaberg et al. PNAS 2009	PC	6% (233)	4% (101)	prostate	USA
Hohn et al. Retrovirology 2009	PC	<0.15% (589)	–	prostate, blood	Germany
Arnold et al. Urology 2010	PC	27.5% (40)	–	prostate, blood	USA
Aloia et al. Cancer Res. 2010	PC	0% (596)	0% (452)	prostate	USA
Martinez-Fierro et al. BMC Cancer 2010	PC	0% (55)	1.3% (75)	prostate	Mexico
Sakuma et al. Retrovirology 2011	PC	4.6% (110)[3]	2.5% (40)[3]	prostate	USA
Furuta et al. Retrovirology 2011	PC, CFS	2.4% (67)	1.6% (500)	blood	Japan
Hohn et al. PLoS One 2010	PC, CFS, MS	0% (151)	0% (40)	blood	Germany
Kearney et al. Adv Virol 2011	PC	0% (134)	–	prostate, blood	USA
Verhaegh et al. Prostate 2011	PC	4% (74)	–	prostate	The Netherlands
Switzer et al. PLoS One 2011	PC	1.9% (162)	–	prostate	USA
Akgul et al. Med Microbiol Immunol. 2011	PC	0% (85)	–	prostate	Germany
Lombardi et al. Science 2009	CFS	67% (101)	3.7% (218)	blood	USA
Erlwein et al. PLoS One 2010	CFS	0% (186)	–	blood	UK
Lo et al. PNAS 2010	CFS	86.5% (37)	6.8% (44)	blood	USA
Groom et al. Retrovirology 2010	CFS	0% (170)	0% (395)	blood	UK
van Kuppeveld et al. BMJ 2010	CFS	0% (32)	0% (43)	blood	The Netherlands
Oakes et al. Retrovirology 2010	CFS	1.8% (112)[3]	52.8% (36)[3]	blood	USA
Robinson et al. Retrovirology 2010	CFS	4.9% (427)[3]	–	blood	UK, Thailand, Korea
Henrich et al. J Infect Dis 2010	CFS, HIV, other[4]	0% (293)	–	blood	USA
Switzer et al. Retrovirology 2010	CFS	0% (51)	0% (97)	blood	USA
Satterfield et al. Retrovirology 2011	CFS	0% (45)	0% (42)	blood	USA
Shin et al. J Virol 2011	CFS	0% (100)	0% (200)	blood	USA
Knox et al. Science 2011	CFS	0% (70)	–	blood	USA
Simmons et al. Science 2011	CFS	0% (15)	0% (15)	blood	USA
Cool et al. Virology 2011	CFS	0% (113)	–	LT, BM, blood	Canada
McCormick et al. Neurology 2008	ALS	0% (25)	0% (14)	blood, CSF	USA
Cornelissen et al. PLoS One 2010	HIV	0% (54)	–	blood	The Netherlands
Barnes et al. J Infect Dis 2010	HIV, HepC	0% (230)	–	blood	UK, Europe
Satterfield et al. Mol Autism 2010	Autism	0% (205)	0% (184)	blood	USA, Italy
Jeziorski et al. Retrovirology 2010	PID, SpA	0% (161)	–	blood	France
Fischer et al. Emerg Infect Dis 2010	RTI	8.8% (192)	3.7% (137)	resp. secretions	Germany, Asia
Delviks-Frankenberry et al. Adv Virol 2011	HIV lymphoma	0% (26)	–	blood	USA
Spindler et al. Adv Virol 2011	HIV	0% (166)	0% (166)	blood	USA
Waugh et al. Cancer Epid Biomarkers 2011	Lymph. malignancy	0% (507)	–	blood	UK

Table 1. Association of XMRV with human disease

A list of published studies investigating the association of XMRV with human disease (as of September 2011). PC, prostate cancer; CFS, chronic fatigue syndrome; CSF, cerebrospinal fluid; ALS, amyotrophic lateral sclerosis; HepC, Hepatitis C; PID, pediatric idiopathic diseases; SpA, spondyloarthritis; RTI, respiratory tract infection; LT, lymphoid tissue; BM, bone marrow. [1]Positive signal is based on the methods used in the study, and does not necessarily mean the detection of XMRV specifically, esp. when no sequences were reported. [2]Location the samples were collected and/or kept, only loosely indicating where the patients were from, but not necessarily the nationality of the population tested. [3]In these cases, mouse DNA contamination of assays was reported. [4]Other conditions include patients with rheumatoid arthritis, organ transplants, or admitted for general medical care.

In contrast to the findings of Knox et al., 53% of the patient samples were reported as being XMRV-positive by VIPdx, seriously calling into question the validity of this non-FDA approved commercial test.

Shin et al. analyzed blood samples from 100 CFS patients and 200 healthy controls for the presence of XMRV using multiple methods, including four different TaqMan qPCR assays, ELISA assay to detect XMRV proteins and growing the virus in tissue culture (135). In addition, samples from 14 patients, previously reported by WPI to be XMRV-positive, were freshly collected by a third party phlebotomist and tested using the same assays. Every single sample tested was negative, using molecular, serological and culturing techniques, including those previously reported as positive by WPI/VIPdx.

These two studies used the same patient population and the same assays as the original study linking XMRV to CFS, ruling out the argument that lack of detection was due to the use of different methods or different patient populations. Considering the large number of negative studies published until then, where no XMRV was found in large-scale human screening studies (Table 1), a more credible hypothesis was supported; that the detection of XMRV may have been, in fact, due to assay contamination.

In the present study, we investigated the origin of XMRV. We identified two previously unknown ancestral MLV proviruses, which generated XMRV by recombination in the laboratory. We analyzed the distribution of these proviruses in wild and inbred mice, and found the species distribution to be rather different, only being found in the same mouse in a few laboratory strains. Our findings indicate that XMRV was created in the laboratory, and its detection in human samples represents an artifact, rather than an infectious virus circulating in human population.

MATERIALS & METHODS

Mouse DNA Samples

Some of the mouse DNA samples were kindly provided by the following individuals: *M. cervicolor popaeus* (J53), *M. caroli* (J136), *M. cookii* (J135), *M. spicilegus* (Halbturn) (J131) from Christine Kozak (NIAID, Bethesda, MD); *M. macedonicus* (XBS), *M. cervicolor cervicolor* (CRV), *M. musculus bactrianus* (BIR), *M. famulus* (FAM), *M. platythrix* (PTX), *M. spicilegus* (ZRU), *M. musculus musculus* (MPB) from François Bonhomme (ISE, Montpellier, France); C57BR/cd, LPT/Le NZW/Lac, PWK/Ph, WMP/Pas and BALB/c (Sigma) from Greg J. Towers (UCL, London, UK); Harlan Sprague Dawley (Hsd) and NIH-Swiss nude mouse DNAs were obtained from Harlan Laboratories, Indianapolis, IN. *M. dunni* DNA was prepared from tail fibroblast (MDTF) cells. NCRNU-M DNA was purchased from Taconic Farms. NU/NU DNAs were prepared from spleens purchased from Charles River Laboratories (Wilmington, MA). 14 AKXL recombinant inbred mouse DNAs were kindly given by Wayne Frankel (The Jackson Laboratory, Bar Harbor, ME). All other mouse genomic DNA samples were purchased from The Jackson Laboratory (Bar Harbor, ME). A list of all mouse strains used in this study is given as part of Table 3.

Polymerase Chain Reaction

All PCRs were performed using either a BioRad C1000 Thermal cycler or an MJ Research Peltier Thermal Cycler PTC200. For screening experiments and host-virus junction DNA amplification, 50 ng of mouse genomic DNA was used per 25 µl reaction. GAGr is specific for the XMRV *gag* leader deletion, XmU3f is specific for the 2-nt

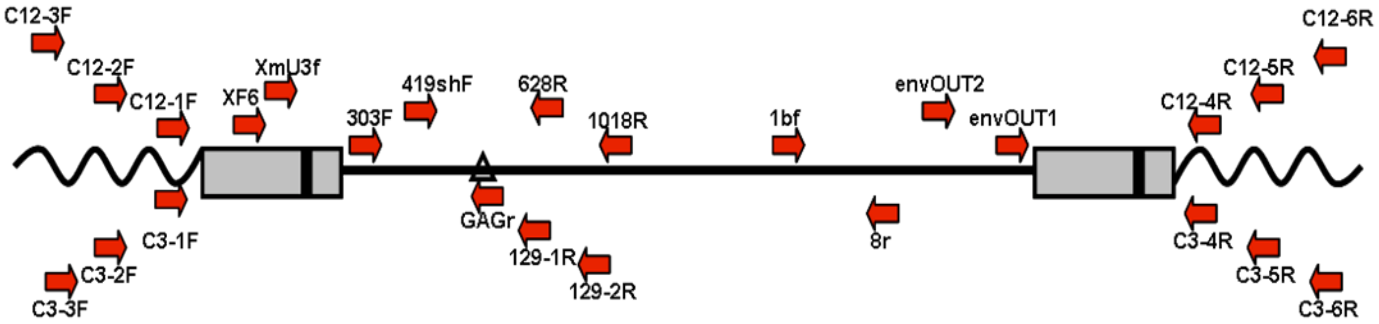
insertion in the U3 region of the XMRV LTR, while XF6 just upstream of XmU3f in the U3 region would amplify many endogenous MLVs. To amplify PreXMRV-2 host-virus junctions, flanking primers for chromosome 12 were used in combination with internal provirus primers 129-1R and envOUT1 respectively. To amplify PreXMRV-1 host-virus junctions, flanking primers for chromosome were used with XgR2 and XF6, respectively. β-actin was amplified using BA-F and BA-R primer set, which amplify both mouse and human β-actin. Mouse IAP elements were amplified using IAP-F and IAP-R. Site directed mutagenesis was performed using the Quikchange kit (Stratagene), according to manufacturer's recommendations.

Primers and Probes

A list of the primers and probes used in this study are given in Fig. 11, together with a cartoon showing the positions of primers used for provirus and flanking DNA amplification.

Single Genome Sequencing

Three fold serial dilutions of mouse genomic DNA were amplified with primer set 303F and 1018R in the first round, then with 419shF and 628R in the second round, 10 reactions per dilution. The dilution at which 3/10 reactions are positive was selected, nested PCR was performed in 96-well format, and reactions were analyzed by electrophoresis on pre-cast 2% E-gels (Invitrogen). Positive reaction products were digested with BsaAI (NEB); fragments that remained undigested were isolated using gel purification kit (Qiagen) and sequenced to confirm the *gag* leader deletion.



Name	Bases	Sequence (5' - 3')
Provirus primers (XMRV, PreXMRV-1, PreXMRV-2)		
GAGr	21	TCCCCCAACAAAGCCACTCCA
XmU3f	22	GTCCTAGCCCTATAAAAAAGGGG
XF6	20	GTACCCGCGCTTTTTGCTCC
envOUT-1	24	CTGACCCAACAGTATCACCAACTC
envOUT-2	23	CAGGCTGCTAAACCTGGTAGATG
XgR1	25	GCGACTCAGTCTATCGGATGACTGG
XgR2	28	AGAGACAAAGACAAAACGATCGCCGGCC
129-1R	26	GCGGTTTCGGCGTAAAACCGAAAGCA
129-2R	26	GGTCTGTCCCATGATCTCGAGAACAC
1bf	19	AGGCATTCCCGACCAAGCG
8r	19	CTGGATGCTACCGGAGCCC
Unblot Probe		
gL1	30	TCGGACTTTTTGGAGTGGCTTTGTTGGGGG
PreXMRV-1 flanking primers (Chr. 3)		
C3-1F	23	CCACCACATATACGTACACCTTC
C3-2F	22	GTAGCCATCAATGAGTTGTGAC
C3-3F	23	TGTCAAGCTGACTCGACCAGAGT
C3-4R	25	GGATCTTCCAGTAGAACTATGTCC
C3-5R	24	GTGATTGGGTTACCTCACTTAGGC
C3-6R	22	GCAGTTTCTGGATGGTCATTCC

Name	Bases	Sequence (5' - 3')
PreXMRV-2 flanking primers (Chr. 12)		
C12-1F	24	TGCTGGACAGAATCTCTGGTCTCT
C12-2F	22	ATGGGCTGTGATCTGACAGAAC
C12-3R	23	TGAGTTCGAGGACAGCCTGATCT
C12-4R	22	GATACTCAAGTGGTTCCCACCC
C12-5R	28	CAGGCACAAAGTTGAAACAGAGACTCAG
C12-6R	25	ATGTGCCTGAGACTATGAATGGGAG
IAP primers		
IAP-F	22	ATAATCTGCGCATGAGCCAAGG
IAP-R	22	TCTGGTCTGTGGTGTTCTTCCT
Beta-actin primers		
BA-F	20	AGCCATGTACGTAGCCATCC
BA-R	20	CTCTCAGCTGTGGTGGTGAA
Single genome sequencing		
303F	24	GCTAACTAGATCTGTATCTGGCGG
419shF	22	ATCAGTTAACCTACCCGAGTCG
628R	21	GGTAGTTACGGTCTGTCCCAT
1018R	23	CTTTATAGAGGGGGTAAGGGCAG

Figure 11. Oligos used in the study

Primers for detecting XMRV, PreXMRV-1 and PreXMRV-2 are depicted in the cartoon (upper panel, not drawn to scale). Names, lengths and sequences of the primers and the one probe used in this study are listed (lower panel).

Integration Site Mapping

The integration sites for PreXMRV-1 and PreXMRV-2 in the mouse genome were identified using the GenomeWalker™ Kit (Clontech) according to the manufacturer's recommendations, with some modifications. To clone the PreXMRV-2 integration site, DBA/2J genomic DNA was digested with six restriction enzymes separately (NEB), and fragments with overhangs were blunted with the Quick Blunting Kit (NEB). Adaptors supplied with the kit were ligated to fragment ends to construct genomic DNA libraries, which were serially diluted to avoid generation of recombinants during PCR. Nested PCR was performed with two sets of adaptor-specific primers (AP1 and AP2) provided with the Genome Walker™ kit and provirus-specific primers, 129-1R and 129-2R, on diluted libraries. Fragments from the first round of PCR were also subjected to BsaAI digest (NEB) to prevent most proviruses without the *gag* leader deletion from being amplified in the second round. A single band was observed after the second PCR, which was cloned into pCR4-TOPO (Invitrogen). Clones were screened for the presence of >1.2-kb insert as well as the absence of the restriction site by BsaAI digest by PCR; those that passed both tests were sequenced. Five such clones revealed identical fragments containing the *gag* leader deletion, a complete 5' LTR and flanking cellular DNA, which mapped to mouse chromosome 12.

To clone the integration site of PreXMRV-1, C57L/J DNA was subjected to the same process as described for PreXMRV-2. Two of the restriction enzymes chosen were ApaLI and Hpy166II, both of which would digest most C57BL/6 MLVs in the LTR upstream of the provirus primer binding site and prevent their amplification, but would leave PreXMRV-1 intact to be amplified. Nested PCR was performed with AP1 and AP2

supplied with the kit, and provirus-specific primers XgR1 and XgR2. Fragments were cloned, sequenced and screened as mentioned for PreXMRV-2. The correct clones with PreXMRV-1 LTR and flanking sequence mapped the provirus to mouse chromosome 3.

Cell Lines

In general, cells were maintained in DMEM F-12 medium (Gibco) supplemented with 10% FBS and 1X Pen/Strep. LNCaP cells (CRL-1740) were grown in RPMI-1640. 293-iGFP-puro (DERSE-293) and LNCaP-iGFP-puro (DERSE-LNCaP) cells were kindly given by KyeongEun Lee and Vineet KewalRamani (NCI, Frederick), and kept in DMEM + 10% FBS + 1 μg/ml puromycin and RPMI + 10% FBS + 1 μg/ml puromycin, respectively.

Transfections

Cells were plated the day before transfection to reach ~50% confluency at the day of transfection. The reaction described is for a 35 mm cell culture plates, for larger volumes, the reaction volume has been scaled proportionately. The DNA to be transfected was mixed with 6 μl CaCl2 (2.5M) and filled with water to a total volume of 60 μl in a microcentrifuge tube. In a separate tube, 60 μl of 2XHBS was placed. The DNA+$CaCl_2$+water mixture was added dropwise to 2XHBS while bubbling the HBS with a pipette. The combined mixture was vortexed, incubated at room temperature for 20-30 minutes to allow complexes to form, and added dropwise to the media present on the cells. Transfection medium was replaced with fresh medium after 12-16 hours of incubation. Assays were performed after an additional 24 hours.

Virus Production and Infection

Viruses were produced either by transient transfection or from cell lines chronically infected with the viruses. Transfected cells were allowed 48 hours for expression and release of virus particles from cells. Supernatant was collected, filtered through 0.45μ filters, and frozen at -80ºC or used directly on cells. To concentrate the virus, supernatants were placed on a sucrose cushion (20% sucrose solution + 1mM EDTA) and ultracentrifuged at 30,000 rpm for 90 minutes to pellet the virus. The pellet was resuspended in 100 μl PBS and used for infections or further assays.

Unblots

Unblots were performed as described previously, with slight modifications (145). The gL1 probe for detecting the *gag* leader deletion was end labeled with ^{32}P γ-ATP using polynucleotide kinase (NEB) at 37^{o}C for 1 hr and the labeled probe was cleaned using Micro Bio-Spin chromatography columns (BioRad). Mouse genomic DNA (15 μg) was either directly digested with BsmI or Eco53KI (NEB) or was first obtained by whole genome amplification from much smaller quantities using the REPLI-g Midi Kit (Qiagen) and phenol extraction. Digested genomic DNA samples were electrophoresed on a 0.9% agarose gel for 20-30 hours, the gel was dried under vacuum, and DNA was visualized by EtBr staining. The dried gel was denatured for 15 min in 0.5 M NaOH, 1.5 M NaCl, neutralized for 15 min in 1 M Tris-HCl, 1.5 M NaCl, pH=8.0, and hybridized overnight with the radiolabeled probe (7.5 x 10^6 cpm) in hybridization buffer (5xSSPE, 0.1% SDS, pH=7.4). The hybridized gel was washed (2xSSC, 0.1% SDS), and subjected to autoradiography.

RESULTS

Disclaimer: The following work has been or is in the process of being published in various scientific journals. I have taken the liberty to copy paragraphs verbatim from such material, with permission. Some of the experiments described in this section were performed in the laboratories of our collaborators: Dr. Vinay Pathak's group at the National Cancer Institute, Frederick, MD and Dr. Brigitte Huber's group at Tufts University, Boston, MA. These data have been included for the sake of completeness, and they formed parts of the following publications:

1. Paprotka T*, Frankenberry KAD*, **Cingöz O***, Martinez A, Kung HS, Tepper CG, Hu WS, Fivash M, Coffin JM, and Pathak VK. "Recombinant Origin of the Retrovirus XMRV" Science, 2011, Jul 1;333(6038):97-101. (*Equal contribution)
2. **Cingöz O,** Paprotka T, Frankenberry KAD, Pathak VK and Coffin JM. "Discovery, analysis and distribution of the two XMRV ancestors." Journal of Virology, 2011 Virol. 2012 Jan;86(1):328-38.
3. **Cingöz O** and Coffin JM. "Endogenous Murine Leukemia Viruses: Relationship to XMRV and MLVs Detected in Human DNA Samples" Advances in Virology, 2011. Article ID 940210. *Review*.
4. Tipper CH*, **Cingöz O***, Coffin JM. "*Mus spicilegus* endogenous retrovirus HEMV uses murine sodium-dependent myo-inositol transporter 1 as a receptor" [Under Review], 2012. (*Equal contribution)
5. Oakes B, Tai AK, **Cingöz O**, Henefield MH, Levine S, Coffin JM, Huber BT. "Contamination of human DNA samples with mouse DNA can lead to false detection of XMRV-like sequences." Retrovirology. 2010;7(1):109.
6. Robinson MJ, Erlwein OW, Kaye S, Weber J, **Cingöz O**, Patel A, Walker MM, Kim WJ, Uiprasertkul M, Coffin JM, McClure MO. "Mouse DNA contamination in human tissue tested for XMRV." Retrovirology. 2010;7(1):108.

A Sensitive Assay for Detection of Mouse DNA

As noted in the introduction section, Lo et al. reported the detection of not XMRV, but MLV-like sequences in blood samples from patients with CFS (92). Similarly, two studies by our collaborators analyzed DNA from the blood of CFS patients and frozen prostate tissues, for the presence of XMRV and MLV-like sequences. In the first study, Robinson et al. found that 14/282 of UK prostate cancer cases, 5/139 of Korean and 2/6 of Thai cases tested positive for amplification with XMRV primers (123). In the second study by Oakes et al., positive amplification results were obtained in very few samples from CFS patients, but in many of the samples from healthy controls, which had been processed at a different time with a slightly different protocol (110). We will explain the second study in more detail.

Oakes et al. analyzed frozen blood samples from CFS patients for the presence of XMRV and/or MLV sequences. In a qPCR assay, the cases or the poorly matched control samples were all negative (110). Interestingly, a nested PCR assay using XMRV *gag* primers revealed 19/36 positives in control samples, but only 2/112 of CFS samples. Sequencing of short virus-derived products revealed extensive variation in sequence, although all were closely related and some were identical to known endogenous mouse proviruses. Fig. 12 shows a neighbor-joining tree derived from the detected patient sequences (names starting with TH; each sequence given a unique number), as well as the corresponding *gag* region from known endogenous and exogenous MLV sequences.

The patient sequences detected in this study clustered with various MLV subgroups (Fig. 12), similar to what would be expected if a DNA template with a mixed

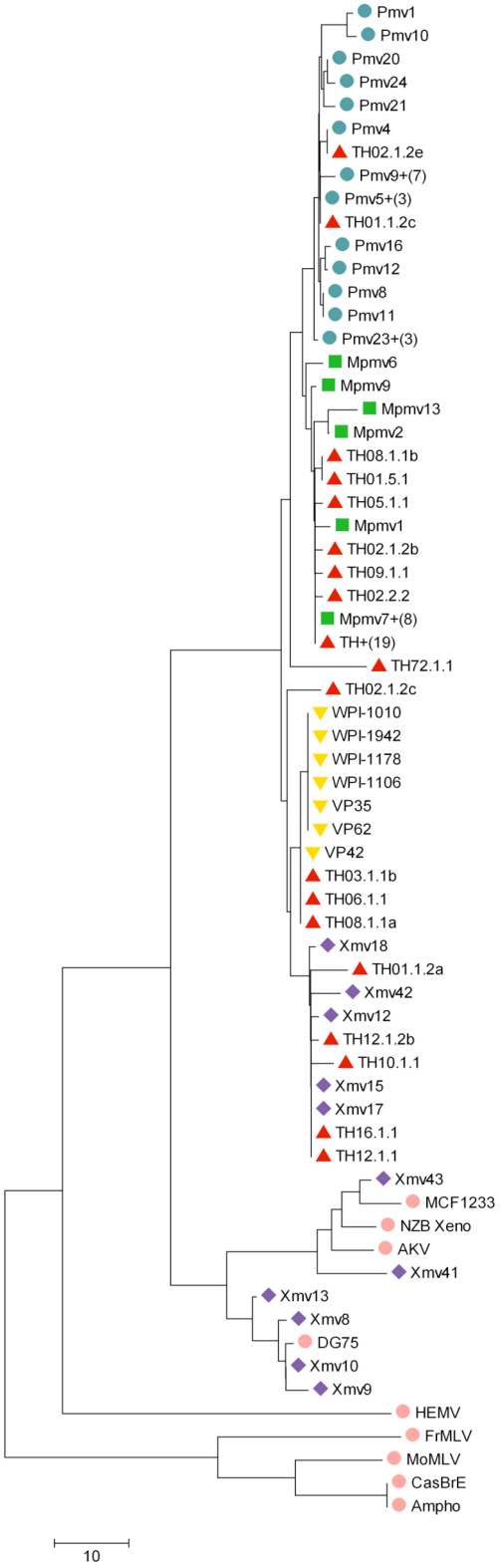


Figure 12. A neighbor-joining tree of gag fragments detected from CFS patients

382 bp *gag* fragments from patient samples (TH#, red) are shown with the corresponding region from XMRV isolates (yellow), endogenous MLVs of C57BL/6 (Xmv, purple; Pmv, blue; Mpmv, green) and some exogenous MLVs (pink). Figure from (110), under the BioMed Central Open Access license agreement.

population of MLV sequences (such as mouse DNA) were to be included in the reaction. In fact, this experiment was done, where trace amounts of mouse DNA were included in a PCR and amplified using MLV primers (Mary Kearney, unpublished results, and (25)). A comparison of neighbor-joining trees constructed from the sequences detected by 1) Lo et al. (CFS patients) 2) Oakes et al. (CFS patients) and 3) Kearney et al. (mouse DNA) is shown in Fig. 13. As with the Lo et al. sequences, those of Oakes et al. fit perfectly within the endogenous MLV phylogeny.

To test whether the presence of these MLV sequences could be explained by contamination of the assay with mouse DNA, we developed a very sensitive assay to detect mouse DNA sequences that relied on the presence of intracisternal A-type particles (IAPs), LTR retroelements found over a thousand copies per mouse genome (120), and not cross-reactive with any sequence in the human genome despite the presence of distantly related IAP elements. The assay relied on simple PCR detection with primers directed to a conserved region in the LTR of mouse IAP elements. The samples were also tested for the presence of mouse mitochondrial DNA.

Amplification results indicated that there was a perfect correlation between the presence of a positive PCR *gag* signal and of mouse DNA (Table 2), suggesting that the detected sequences represent MLV proviruses amplified from the contaminating mouse DNA, rather than infection of humans with a gammaretrovirus (110). Although the exact source of mouse DNA contamination could not be fully determined, there were several recent examples of laboratory reagents found to contain trace amounts of mouse DNA, resulting in the generation of false positive results in PCR amplification assays with MLV primers (38, 85, 131, 135, 162, 187). Taken together, our results demonstrated the

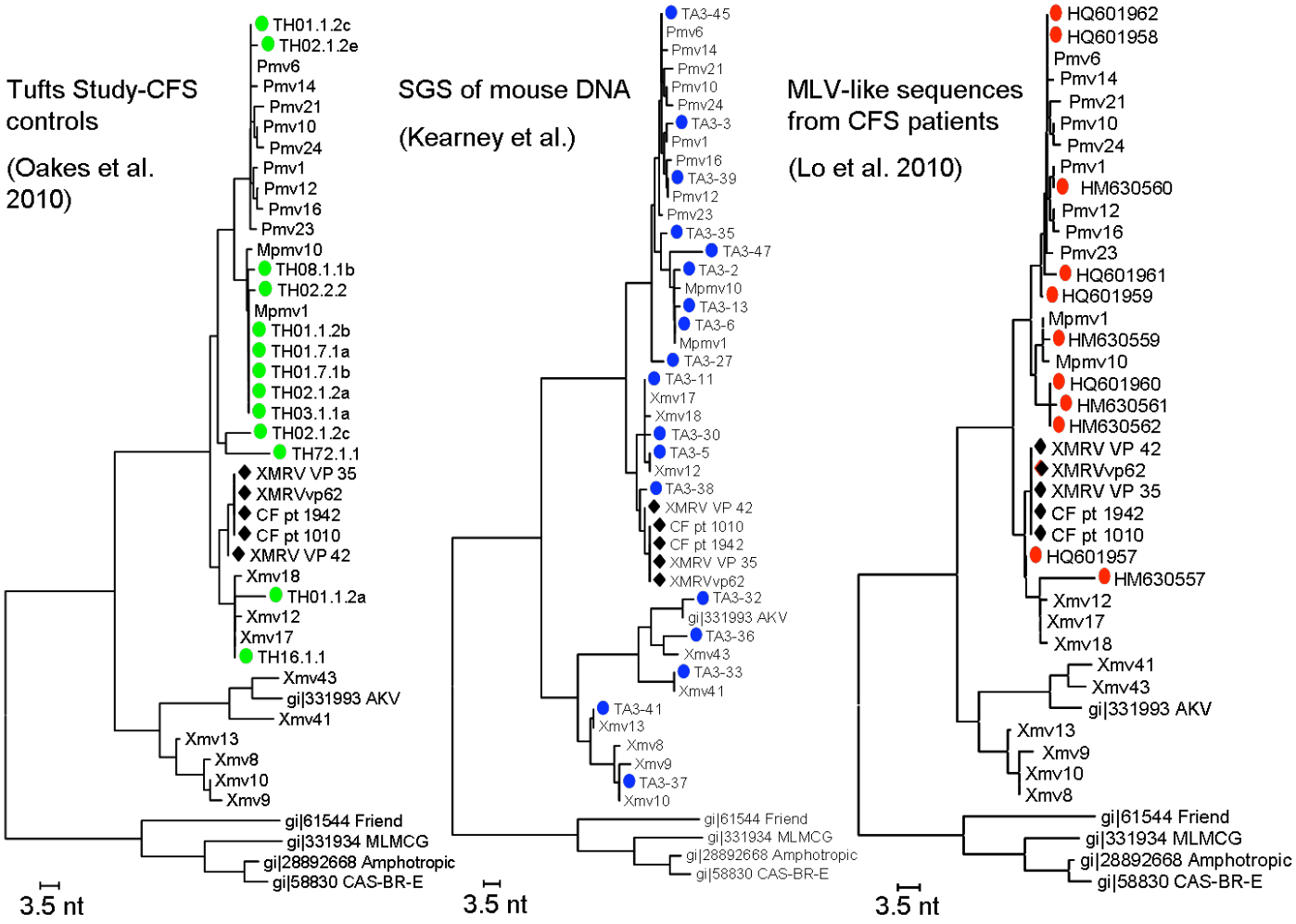


Figure 13. Phylogenetic analysis of XMRV/MLV sequences

Left panel; Sequences detected in samples from banked CFS patients by Oakes et al. (110). Middle panel; Sequences detected by single genome sequencing on serial dilutions of mouse DNA (Mary Kearney, unpublished results). Right panel; MLV-like sequences detected in the study by Lo et al. (92). Note the similar positions occupied by MLV sequences in the three different studies, as would be expected if MLV sequences were to be amplified from small amounts of mouse genomic DNA. Figure from (25). (The authors retain the copyright.)

XMRV GAG	*cox*	IAP	CFS Patients # of Samples (n = 112)	Percent	Healthy Controls** # of Samples (n = 36)	Percent
+	+	+	2*	1.8	17	47.2
-	-	-	53	47.3	12	33.3
+	-	-	0	0	0	0
+	-	+	0	0	2	5.6
+	+	-	0	0	0	0
-	+	+	10	9.0	1	2.8
-	-	+	47	42.0	4	11.1
-	+	-	0	0	0	0

*One CFS sample from 2005 collection, and one CFS sample from 2010 collection. All the other CFS samples were collected in 2005.
**All collected in 2009-2010.

Table 2. Correlation of MLV sequence detection with mouse DNA contamination

DNA samples from CFS patients (n=112) and healthy controls (n=36) were tested by PCR for the presence of MLV *gag* (column 1), mouse mitochondrial DNA (column 2) and mouse IAP elements (column 3). In all cases, samples that were positive for MLV *gag* sesquences by PCR were also positive for mouse DNA. Table from (110), under the BioMed Central Open Access license agreement.

ease with which assays could be contaminated with mouse DNA, complicating the distinction of false positive results from true human infections.

Searching Mouse Genomes for XMRV-like Elements

The sequence similarity between XMRV and murine leukemia viruses (MLVs) strongly suggests that XMRV originated from one or more MLVs. Despite the degree of sequence identity, the XMRV genome is clearly distinct from all known Xmv proviruses, and it is not found in the sequenced C57BL/6 mouse genome. We hypothesized that XMRV, or an ancestral provirus, must be present in the genomes of other inbred or wild mouse strains. We developed several assays to screen various mouse genomes in an effort to identify the ancestral provirus(es) that gave rise to XMRV. Four strategies were employed to search for XMRV-like sequences in mouse genomes.

1. XMRV-Specific PCR Assay

The XMRV genome shows up to ~95% identity with some MLVs, found as both endogenous and exogenous viruses (163). To detect XMRV-like sequences in mouse DNA, first a highly specific PCR assay was designed that selectively amplified the 5' region of the XMRV genome, but not any of the endogenous MLVs in C57BL/6 (74) under the conditions used (115). Due to the sequence similarity between XMRV and endogenous MLVs, primer selection for PCR was restricted to regions of the provirus that were sufficiently divergent. One of the features that distinguish the XMRV genome from most other MLVs is a 24-bp deletion in its *gag* leader region, relative to exogenous and some endogenous MLVs (or a 15-bp deletion relative to most endogenous MLVs) (Fig. 14A). This deletion was initially thought to be specific for XMRV, but has recently

been detected in a provirus in the whole genome shotgun (WGS) sequence of an inbred strain, 129X1/SvJ (GenBank ID: AAHY01591888.1) (31), and later in a provirus in a few other strains using deep sequencing (71). To specifically detect XMRV, a reverse primer (GAGr) spanning the *gag* leader deletion was used in combination with a forward primer (XmU3f), which includes a 2-bp insertion unique to XMRV U3 region (Fig. 14A). The sensitivity and specificity of the assay was assessed by amplification of XMRV sequences from as little as 30-100 pg of genomic DNA (~5-16 cell equivalents; ca 100-300 proviruses) from the XMRV-infected 22R*v1* cell line in the absence or presence of 50 ng of C57BL/6 DNA, respectively (Fig. 14).

In the initial screen with the XmU3f-GAGr primer set, XMRV-specific sequences were not found in any of the 46 wild-derived or 48 laboratory mouse strains tested (115) (for a complete list, see Table 3), while the 22Rv1 control cell line was always positive, implying that no endogenous provirus 100% identical to XMRV was present in these strains. For each strain tested, β-actin was successfully amplified from the genomic DNAs, ruling out the possibility that the lack of detection resulted from low DNA quality or amplification associated problems (Fig. 14C).

The partial U3 region included in the 129X1/SvJ WGS sequence revealed that the XMRV-specific 2-bp insertion was not present in the 129X1/SvJ provirus (Fig. 14A). Moreover, there was a 6-bp deletion in the XmU3f primer-binding site, rendering this primer ineffective for amplification of that provirus under such stringent conditions. Therefore, a different forward primer was designed (XF6) with a perfect match to both XMRV and the 129X1/SvJ WGS sequence, which yielded a clear band in both wild-derived and laboratory strains when used with the GAGr primer (Fig. 14C).

Laboratory mice			Wild-derived mice			
Strain (n=48)	**PreXMRV-1**	**PreXMRV-2**	**Subspecies**	**Strain (n=46)**	**PreXMRV-1**	**PreXMRV-2**
129P1/ReJ	-	+	*M. m. domesticus*	SF/CamEi	-	+
129P3/J	-	+		SK/CamEi	-	+
129S1/SvImJ	-	+		SK/CamRk	-	-
129X1/SvJ	-	+		PERA/Ei	-	-
A/J	-	-		PERC/Ei	-	-
AKR/J	-	-		CALB/Rk[a]	+	-
AKR/J nude	-	-		WSB/Ei	-	-
B6.129/J	-	-		BIK	-	-
B6CByF1/J nude	-	-		ZALENDE/Ei	-	-
BALB/c nude	-	-		TIRANO/Ei	-	-
BALB/cByJ	-	-		Poschiavinus	-	-
BALB/cJ	-	-		BFM	-	-
BTBR/J	-	+		WMP/Pas	-	-
C3H/HeJ	-	+	*M. m. castaneus*	CTA	-	-
C57BL/6J	-	-		CASA/Rk	-	-
C57BR/cdJ	+	+		CAST/Ei	+	-
C57L/J	+	-	*M. m. molossinus*	MOLC	+	-
C58/J	+	-		MOLD/Rk	+	-
CBA/J	-	+		MOLE/Rk	+	-
CByB6F1/J nude	-	-		MOLF/Ei	+	-
CByJ.Cg/J nude	-	-		MOLG/DN	+	-
CE/J	-	-		MSM/Ms	+	-
CWD/LeJ	-	-		JF1/Ms	-	-
DBA/1J	-	-	*M. m. musculus*	CZECH/I	-	-
DBA/2J	-	+		CZECH/II	-	-
HRS/J hr/+	-	+		SKIVE/Ei	-	-
Hsd nude*	+/-	+/-		MPB	-	-
I/LnJ	-	-		PWK/Ph	-	-
LP/J	-	+	*M. spretus*	SFM	-	-
LPT/LeJ	-	+		SPRET/Ei	-	-
MA/MyJ	-	+	*M. spicilegus*	J131	-	-
NCRNU*	-	+/-		PANCEVO/Ei	-	-
NFS/N	-	+		ZRU	-	-
NIH Swiss*	-	+	*M. caroli*	KAR	-	-
NIH-III nude	-	+		CAROLI/Ei	-	-
NU/J	-	+		J135	-	-
NU/NU*	+/-	+/-	*M. cookii*	COK	-	-
NUJM nude	-	+		J136	-	-
NZB/B1NJ	-	-	*M. cervicolor*	CRV	-	-
NZW/LacJ	+	-		J53	-	-
P/J	-	-	*M. platythrix*	PTX	-	-
RIIIS/J	-	-	*M. bactrianus*	BIR	-	-
SJL/J	-	+	*M. famulus*	FAM	-	-
SJLSmn.AK nude	-	+	*M. macedonicus*	XBS	-	-
SM/J	-	-	*M. dunni*	MDTF	-	-
ST/bJ	-	+	*M. pahari*	*Mus pahari*/Ei	-	-
STOCK Ces1c nude	-	-				
SWR/J	-	+		Positives (%)	8 (17.4%)	2 (4.3%)
Positives (%)	6 (12.5%)	25 (52.1%)				

Table 3. Mouse strains tested for the presence of XMRV, PreXMRV-1 and -2

XMRV is not found as a single provirus in any mouse strain tested. [a]CALB/Rk carries haplotypes from *M. m. castaneus* (72). *Outbred strains; individual mice may differ in the number and distribution of proviruses they harbor (Fig. 26). The +/- sign means some members are positive, while some are negative (outbred strains). Table from (26), with permission.

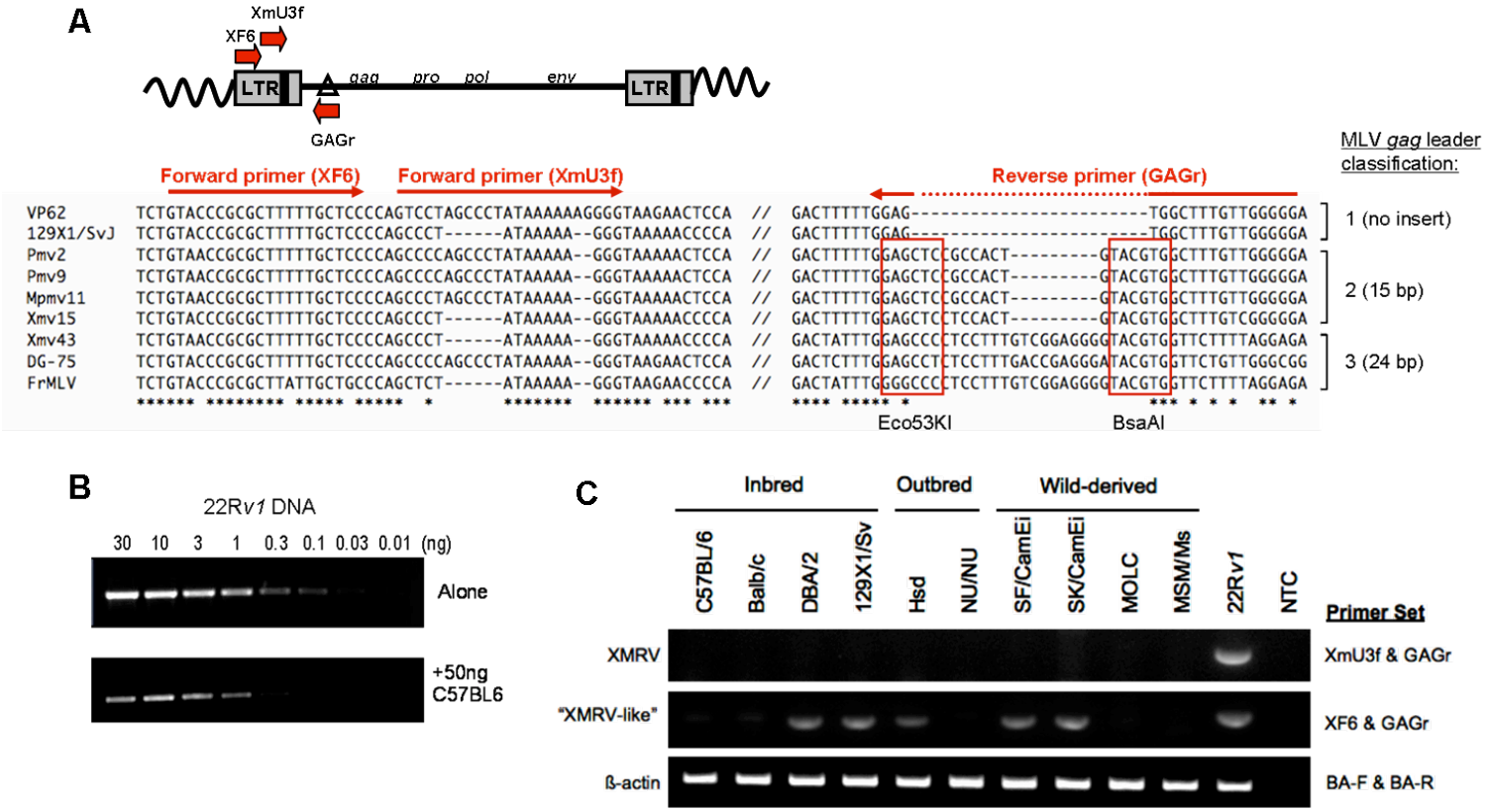


Figure 14. Screening of mouse DNA for XMRV and XMRV-like elements by PCR

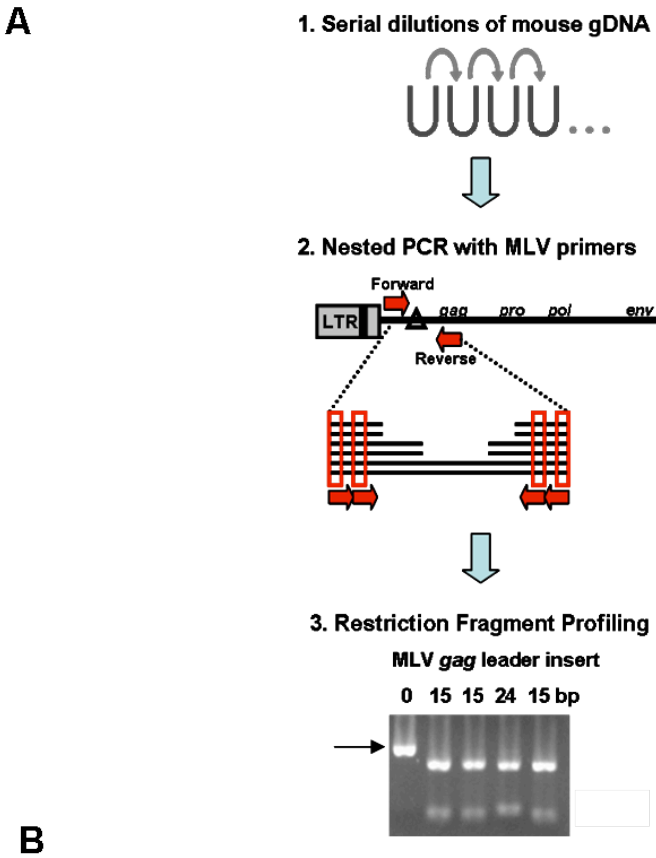
(A) Schematic representation of primers used (upper panel), with the exact locations indicated by arrows (lower panel). VP62 is the reference XMRV isolate (163). MLV gag leader regions can be divided into three subgroups based on the length of the insert in the gag leader. Unique restriction sites BsaAI and Eco53KI, which distinguish proviruses in the first group (no insert) from the rest (15 or 24 bp insert) are shown in boxes. (B) Sensitivity and specificity of the PCR assay. Serial dilutions of genomic DNA from 22Rv1 cells were amplified with the XMRV-specific primer set (GAGr and XmU3f) in the presence or absence of 50 ng of C57BL/6 genomic DNA. (C) The absence of XMRV (top) and presence of an XMRV-like provirus in a panel of representative mouse strains. The primer sets used for each assay is indicated on the right. Note that the NU/NU and Hsd nude mice are outbred strains; only one representative of each is shown. Larger numbers of individual mice from these outbred strains are shown in Fig. 26. Figure from (26), with permission.

The PCR products from all strains were sequenced and confirmed to be identical to each other, differing from the 129X1/SvJ sequence at two positions (nt 157 and 263 of the WGS file), where the deposited 129X1/SvJ sequence has an extra C and T nucleotide, respectively. The presence of the 24-bp deletion in an endogenous provirus of this mouse strain (129X1/SvJ) suggested the presence of XMRV-like sequences in mouse genomes, even though the rest of the sequence differed from XMRV significantly. XMRV itself was not present in any mouse genome as a full-length endogenous provirus.

2. Single Genome Sequencing

To obtain more accurate sequence information and to rule out potential PCR recombination between different proviruses, a single genome sequencing (SGS) approach was developed. MLV genomes were classified into three groups based on their *gag* leader profiles: those that contain 1) no insert, 2) a 15-bp, or 3) a 24 -bp insert (Fig. 14A). The first group includes all XMRV isolates, as well as the uncharacterized provirus fragments identified by deep sequencing (71), while the second and third groups include all C57BL/6 proviruses and some exogenous MLVs (74). A diagnostic BsaAI restriction site was identified at the junction of the deletion, which would selectively digest a large fraction of the sequences in the last two groups (Fig. 14A). This selection method eliminated most "unwanted" proviruses and focused on those with the 24-bp deletion, plus a few with a polymorphism at the restriction site.

An overview of the single genome sequencing approach is shown in Fig. 15A. Briefly, serial dilutions of genomic DNA were subjected to nested PCR with well-conserved MLV primers upstream and downstream of the gag leader deletion, to the point where one in three wells was positive (see Materials and Methods). Positive



B

```
NCRNU fragment      ACCCGAGTCGGACTTTTTGGAG-----------------------TGGCTTTGTTGGGGGACG
NU/J fragment       ACCCGAGTCGGACTTTTTGGAG-----------------------TGGCTTTGTTGGGGGACG
DBA/2J fragment     ACCCGAGTCGGACTTTTTGGAG-----------------------TGGCTTTGTTGGGGGACG
129x1/SvJ WGS       ACCCGAGTCGGACTTTTTGGAG-----------------------TGGCTTTGTTGGGGGACG
DBA/2J WGS-1        ACCCGAGTCGGACTTTTTGGAG-----------------------TGGCTTTGTTGGGGGACG
DBA/2J WGS-2        ACCCGAGTCGGACTTTTTGGAG-----------------------TGGCTTTGTTGGGGGACG
DBA/2J WGS-3        ACCCGAGTCGGACTTTTTGGAG-----------------------TGGCTTTGTTGGGGGACG
Pmv9                ACCCGAGTCGGACTTTTTGGAGCTCCGCCACTGT---------ACGTGGCTTTGTTGGGGGACG
Xmv15               ACCCGAGTCGGACTTTTTGGAGCTCCTCCACTGT---------ACGTGGCTTTGTCGGGGGACG
Xmv43               GACCGATTCGGACTATTTGGAGCCCCTCCTTTGTCGGAGGGGTACGTGGTTCTTTTAGGAGACG
DG-75               ACCCGTCTCGGACTCTTTGGAGCCTCTCCTTTGACCGAGGGATACGTGGTTCTGTTGGGCGGCG
                      ***  ******* *******                      *** * * *  ** * **
```

Figure 15. Detection of XMRV-like elements by single genome sequencing

(A) Outline of the SGS approach. BsaAI digestion profiling (Fig. 14A) distinguishes fragments with the deletion from those without. (B) Alignment of sequences from three strains positive for the *gag* leader deletion (NU/J, NCRNU and DBA/2J), the whole genome shotgun (WGS) sequences from the 129X1/SvJ and DBA/2J genomes in GenBank and some other MLV sequences as a reference. Asterisks indicate bases conserved at a given position for all MLV fragments analyzed. Figure from (26), with permission.

reactions were analyzed by BsaAI digest. Fragments that remained undigested (i.e. those with the *gag* leader deletion) were analyzed further by sequencing, and the presence of the deletion (or a polymorphism in the restriction site) was confirmed. Using this approach, a total of seven mouse strains were analyzed, three of which contained XMRV-like *gag* leader regions (Fig. 15B), providing further evidence that such elements are present as endogenous viruses in the genomes of several inbred mouse strains.

3. Detection of Provirus Fragments in Unblots

To investigate the distribution of XMRV-like sequences in mouse genomes and to provide further support to the PCR screening data, genomic hybridization blots were performed using the unblotting technique (158). Briefly, genomic DNAs from various mouse strains were digested with a restriction enzyme, run on an agarose gel that was then dried and hybridized with a *gag* leader deletion-specific radiolabeled probe (see Materials & Methods). The enzyme Eco53KI was chosen because it would selectively digest proviruses without the *gag* leader deletion at the probe binding site, leaving proviruses with the deletion intact until the next downstream restriction site in the *pro* gene (Fig. 14A and 16A). A single high-intensity band was detected in six out of ten strains analyzed, in agreement with the genome screening results by PCR (Fig. 16B).

4. GenBank Database Searches

To investigate whether other mouse strains besides 129X1/SvJ also carried the *gag* leader deletion (31, 71), an extensive search of the NCBI Trace Archives was performed by querying the *gag* leader region of XMRV. This search returned an additional three hits in the DBA/2J genome and one unassigned sequence with exactly

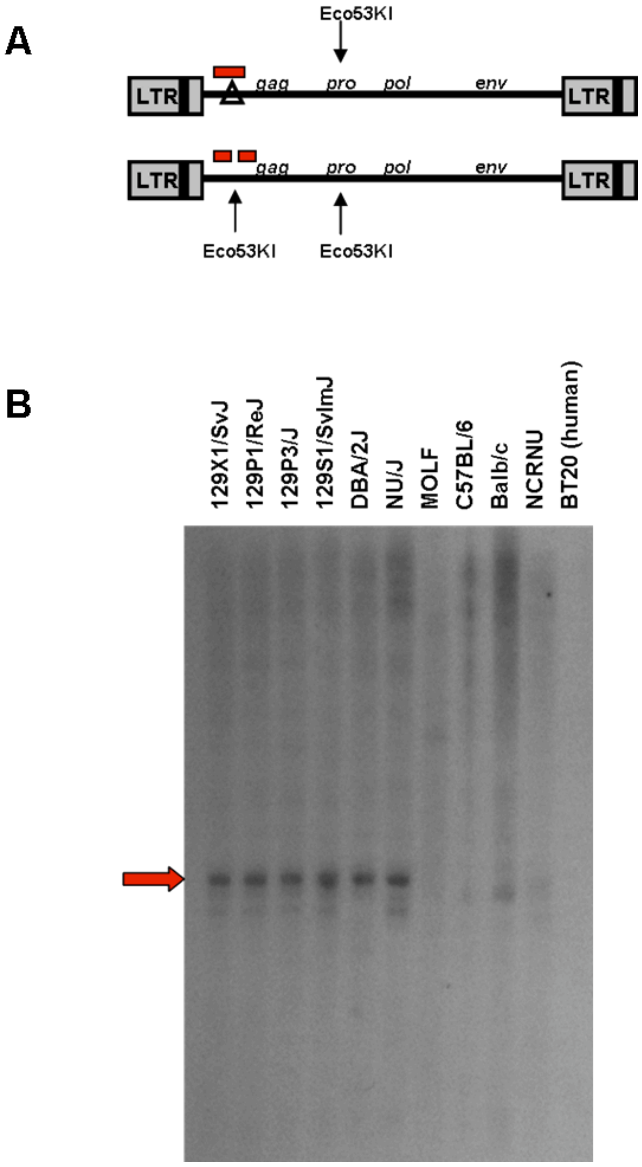


Figure 16. Detection of XMRV-like elements by genomic blot hybridization

(A) Schematic representation of the Eco53K1 sites and the gL1 probe binding site in known MLVs. Upper panel: XMRV-like, lower panel: most MLVs (B) Unblots on various mouse genomic DNA samples digested with Eco53KI and hybridized with a radiolabeled *gag* leader deletion specific probe. A distinct provirus fragment corresponding to PreXMRV-2 was detected in six strains (arrow). Figure from (26), with permission.

the same deletion (GenBank IDs: ti:1098845661, 1096851092, 1096808503, 1039132836). None of these sequence files extended beyond the LTR, providing no information about the integration site. Nevertheless, the presence of this deletion in multiple sequencing reads further confirmed our results and those of others (71), that XMRV-like endogenous sequences are indeed present in the genomes of multiple inbred strains of mice.

In addition to the trace archive database, a BLAST search of the genomes of additional mouse strains sequenced by the Sanger Institute was performed (for a list of strains, see Fig. 17A) (79). Unfortunately, it seems likely that the genome construction of these newly sequenced strains was based on the reference C57BL/6 genome, causing proviruses absent from the C57BL/6 genome to be left out from the final assembly, even though they are found in the sequenced strains. For example, in a ligation-mediated PCR approach aimed at mapping a specific MLV insertion (discussed further in the Results section), we identified the integration site of a different MLV provirus in the 129X1/SvJ strain. Amplification of host-virus junctions from 129X1/SvJ genomic DNA confirmed the presence of this provirus in this strain, as well as in three closely related strains: 129P1/ReJ, 129P3/J and 129S1/SvImJ. The junction fragments were sequenced to determine the exact integration site. BLAST searches of several strains sequenced by the Sanger institute, which included three 129-derived strains (Fig. 17A), failed to detect the junction fragment in any of the assembled genome sequences. Searching instead for the flanking sequence returned the expected genomic locus with no provirus (Fig. 17B). The sequence at the exact integration site in the whole genome assembly was complicated by a series of mixed bases, indicating incomplete base reads. This discrepancy is likely

A

1. 129P2
2. 129S1/SvImJ
3. 129S5
4. A/J
5. AKR/J
6. BALB/cJ
7. C3H/HeJ
8. C57BL/6NJ
9. CAST/EiJ
10. CBA/J
11. DBA/2J
12. LP/J
13. NOD/ShiLtJ
14. NZO/HiLtJ
15. PWK/PhJ
16. Spretus/EiJ
17. WSB/EiJ

B

Figure 17. List of mouse strains sequenced by the Sanger Institute

(A) List of mouse strains sequenced by the Sanger Institute, UK (79). (B) One example of an experimentally detected MLV provirus not present in the C57BL/6 genome (Query), but likely found in the 129P2 strain (Subject), which is not included in the assembly. Arrow represents the exact integration site sequenced from four related inbred strains (129X1/SvJ, 129P1/ReJ, 129P3/J and 129S1/SvImJ), and is complicated by a series of mixed base codes (other than A T G C, based on IUPAC codes). Searches in the 129S1/SvImJ genome also revealed the absence of this provirus in the final genome sequence, even though it is in the genome (data not shown).

caused by viral sequences read at the start and end of the provirus, but cannot be assigned to the reference C57BL/6 genome because it does not contain that insertion at that specific site, resulting in multiple base reads at the integration site, which manifests itself as unspecified base codes in the final assembly. Although the sequence alignment shown in Fig. 17B is 129P2, a strain we did not test, the presence of this insertion was confirmed in 129S1/SvImJ, whose genome sequence also lacked it. This approach was therefore not pursued further.

Mapping the Integration Site

Multiple independent lines of evidence implied the presence of an XMRV-like element in various mouse genomes. Further analysis required determination of the integration site. A ligation-mediated PCR approach was used to amplify and clone fragments of DBA/2J genomic DNA that contained the specific *gag* leader deletion, including the complete 5' LTR and flanking sequence, which mapped to chromosome 12 on the C57BL/6J genome sequence. The experimental details regarding the mapping of this provirus is described in Materials and Methods. Knowledge of the flanking sequence allowed the design of primer pairs to amplify host-virus junctions from mouse genomic DNA (Fig. 18A).

The integration site of the provirus was confirmed by amplifying both 5' and 3' junctions from the DBA/2J genomic DNA, using flanking chromosomal primers and internal provirus primers (Fig. 18B and 14A). Both junction fragments contained a full-length LTR of 539 bp and the expected chromosome 12 flanking region as confirmed by sequencing, with a 4-bp target site duplication (TSD) typically found in MLV insertions. The *gag* leader region included in the 5' junction fragment was also sequenced again,

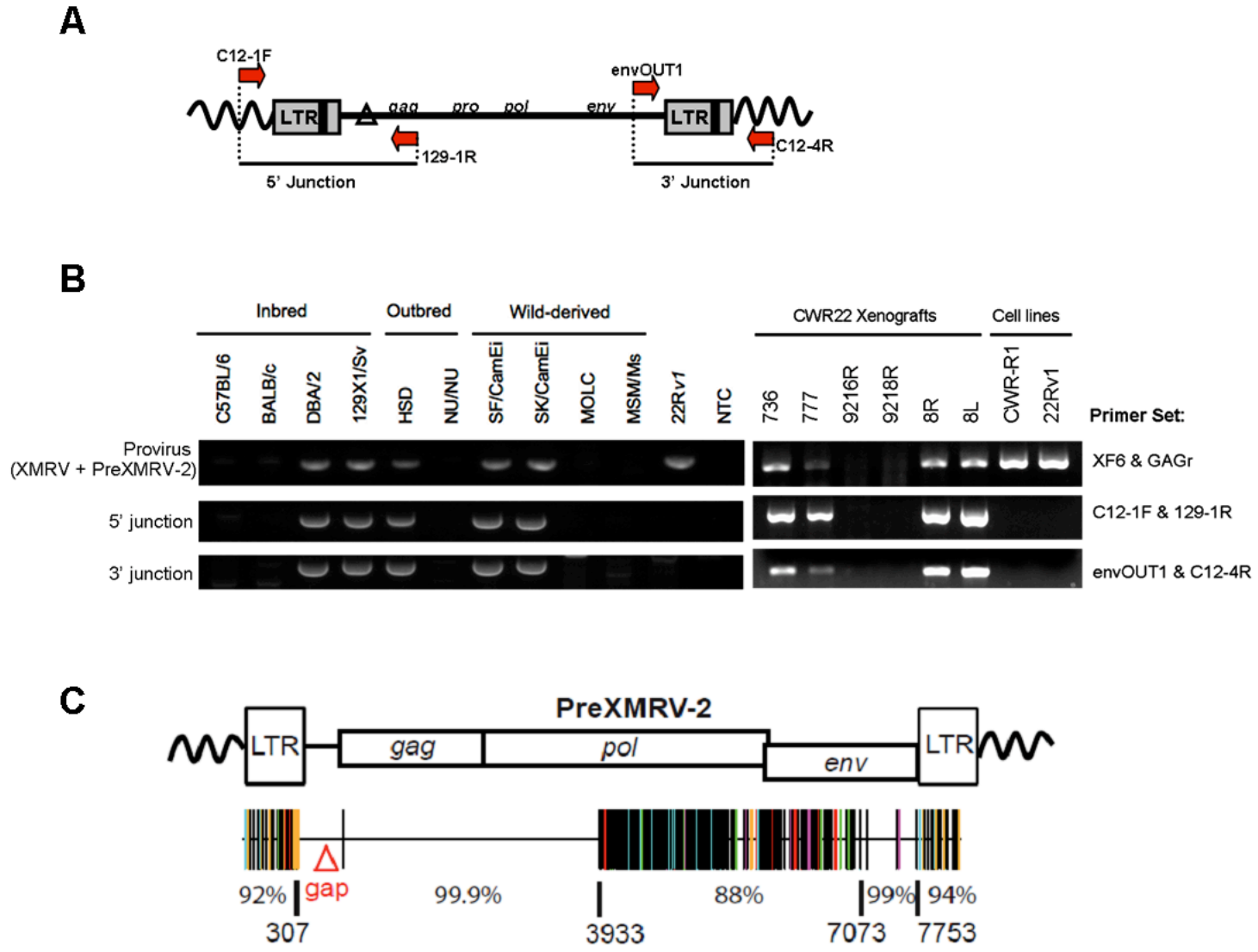


Figure 18. Detection of XMRV-related sequences (PreXMRV-2) in mouse strains.

(A) PCR approach to amplify host-virus junctions from genomic DNA. (B) Representative mouse strains and early xenograft samples were analyzed for the presence of PreXMRV-2 (for a complete list, see Table 3). In 22R*v1* and CWR-R1 cells, the fragment detected by internal provirus primers is XMRV. (C) Hypermut plots indicate nucleotide mismatches relative to XMRV as color-coded vertical lines. Percent identity to consensus XMRV for different regions of each provirus is indicated below. Nucleotide numbers refer to the 22Rv1-XMRV sequence (FN692043) (114). Figures modified from (26) and (115), with permission.

confirming the presence of the 24-bp deletion. It is important to note that the provirus integration site is found within a rather complex, repetitive and poorly sequenced region, making it impossible to distinguish the empty pre-integration site from the provirus-occupied integration site.

The identified *gag*-leader deletion-containing provirus is characterized in detail in the following sections, however, the most striking feature is a ~3.6 kb stretch encompassing the *gag* leader region, *gag* and part of *pol* genes, which is 100% identical to XMRV released from the 22R*v1* cells (114), and only 1 bp different from several XMRV isolates (Fig. 18C). In addition, a ~700 bp region of env is 99% identical to XMRV, however, the LTRs and the remaining viral genome differ by 6 to 12% from consensus XMRV (115). This provirus was named PreXMRV-2 (GenBank ID: FR871850.1), due to its role in the generation of XMRV (115), explained in more detail in the following sections.

Generation of the 22R*v1* Cell Line

The discovery that a human prostate cancer cell line produces infectious XMRV and contains ≥10 proviral copies per cell was assumed to suggest that the patient, whose tumor gave rise to the cell line, was infected with XMRV at the onset (84). To determine the origin of XMRV, we investigated the genesis of the 22R*v1* cell line (142). This cell line was derived in 1999, from a xenograft (CWR22) that was established from a primary prostate tumor in 1992 at Case Western Reserve University (CWRU) and serially passaged in nude mice (Fig. 19) (105, 119). To elucidate the origin of the virus in 22R*v1* cells, we analyzed various passages of the CWR22 xenograft, as well as a subline of the CWR22 xenograft (2152) from which the 22R*v1* cell line was established (142), and

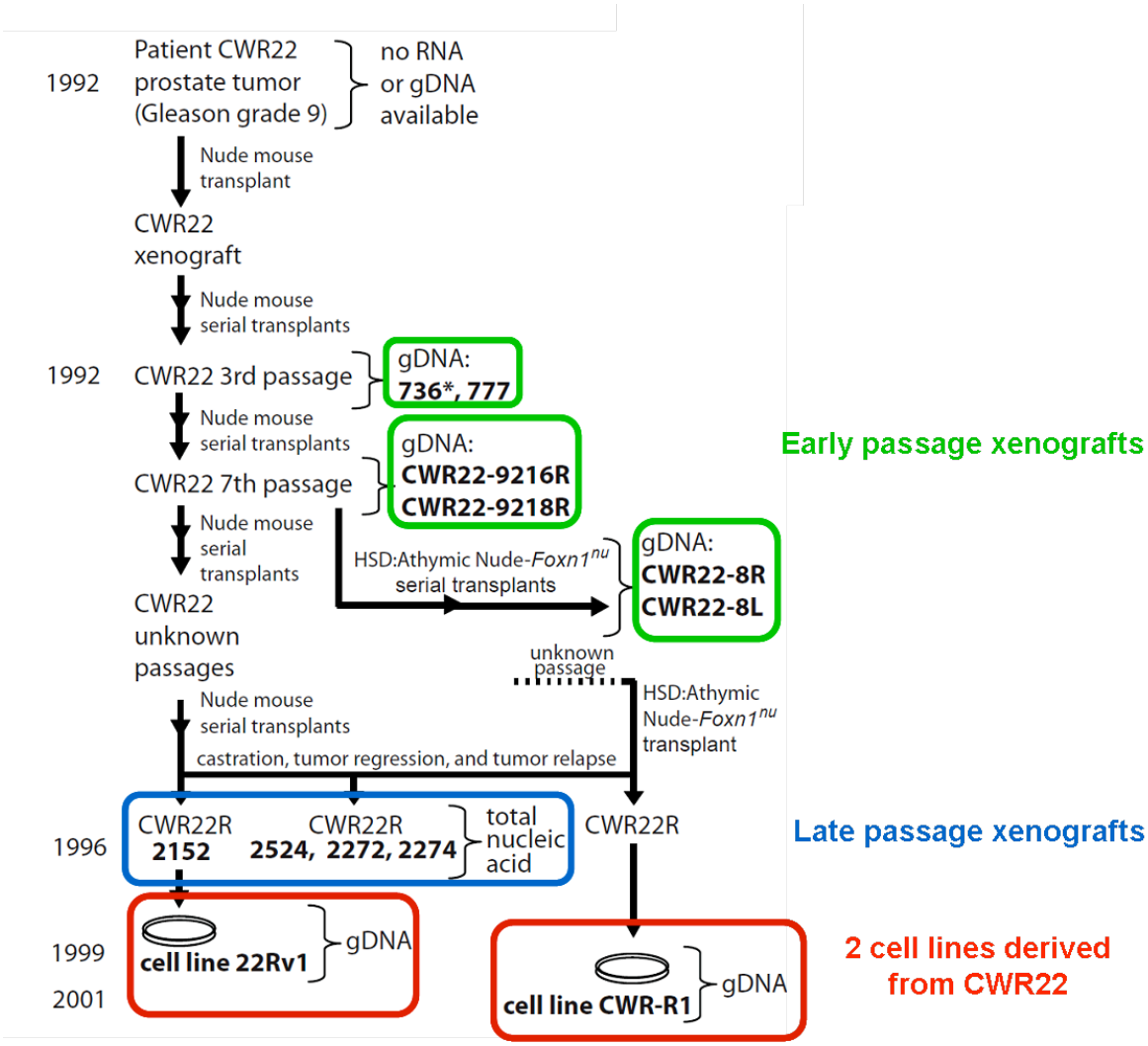


Figure 19. Characterization of CWR22 xenografts and XMRV-related sequences

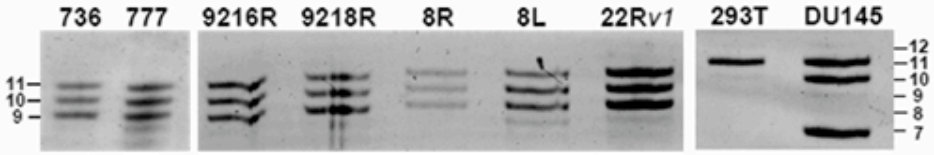
(A) Genesis of 22Rv1 and CWR-R1 cell lines. Bold letters indicate samples from which genomic DNA (gDNA) or total nucleic acid was available for analysis. XMRV-positive samples are boxed. Asterisk (*) indicates unknown early passage. Figure modified from (115), with permission.

another prostate cancer cell line, CWR-R1, which was also derived from CWR22 (54). Fig. 19 traces the timeline of the serial xenograft transplants of CWR22 up to the derivation of the cell lines 22R*v1* and CWR-R1 and indicates (bold letters) the samples that were available for analysis. Outbred nude mouse strain(s) maintained by Charles River (NU/NU) and Harlan Laboratories (Harlan Sprague Dawley [Hsd]) are likely to have been used for *in vivo* passages of the xenograft. DNA samples from passage 3 (777) and an unknown early passage (736) were obtained along with samples from the 7th passage, CWR22-9216R and CWR22-9218R. A xenograft tumor from the early 7th passage was independently propagated at the University of California, Davis (UCD) using Hsd nude mice (CWR22-8R and 8L). Total nucleic acid from relapsed androgen-independent tumors (CWR22R) 2152, 2524, 2272, and 2274 and the 22R*v1* and CWR-R1 cell lines was also available for analysis (Fig. 19) (105). All xenograft samples were traced, found, requested and obtained by the Pathak Lab (115).

CWR22 Xenografts are Derived from the Same Person

To ensure the identity of the tumor DNA samples analyzed, we verified that the xenograft samples (736, 777, 9216R, 9218R, 8R and 8L) and the 22R*v1* or CWR-R1 cell lines were all derived from the same person by performing short tandem repeat (STR) analysis (Fig. 20A). We included human cell lines 293T and DU145 as controls. Allele patterns for the control cell lines were as expected for all 7 loci (data not shown). Allele patterns for 22R*v1* and CWR-R1 were consistent with previous reports and American Type Culture Collection (ATCC) (86, 164). For 22R*v1* and CWR-R1, 5/7 and 6/7 allele patterns matched the xenograft alleles, respectively. The probabilities that the xenografts and the two cell lines have the same allele patterns for these loci by chance are

A



B

	Xenografts: 736, 777, 9216R, 9218R, 8R, and 8L		22R*v1*		CWR-R1	
locus	**alleles**	**frequency, 1 in ___**	**alleles**	**frequency, 1 in ___**	**alleles**	**frequency, 1 in ___**
D5S818*	**11, 12**	3.91	**11,** 13	9.50	**11, 12**	3.91
D7S820†	**9,10,11**	11.60	**9,10,11**	11.60	**9,10,**12	11.60
D13S317	**8, 12**	15.94	**9, 12**	26.39	**8, 12**	15.94
D16S539	**12, 12**	9.77	**12, 12**	9.77	**12, 12**	9.77
TH01	**6, 9.3**	7.39	**6, 9.3**	7.39	**6, 9.3**	7.39
TPOX	**8, 8**	4.11	**8, 8**	4.11	**8, 8**	4.11
CSF1PO	**10, 11**	5.86	**10, 11**	5.86	**10, 11**	5.86
Random Match Probability, 1 in ___ :		1,259,877		5,062,112		1,259,877
Random chance that the xenografts and 22R*v1* or R1 would have the same allele pattern:				1.6×10^{-13}		6.3×10^{-13}

*STR analysis at locus D5S818 was not determined for sample 736.
†Frequency for D7S820 was determined using only the 9 and 10 allele frequencies.

Figure 20. Short tandem repeat (STR) analysis of CWR22 xenografts

(A) Representative D7S280 allele pattern of xenografts and 22Rv1 and CWR-R1 cell lines. An allelic ladder is shown on the sides of the gel. (B) STR analyses of the six xenografts (736, 777, 9216R, 9218R, 8R and 8L) and the 22R*v1* and CWR-R1 cell lines were compared at 7 different loci for lineage determination. Figure from (115), with permission.

1.6×10^{-13} and 6.3×10^{-13}, respectively. The analysis for locus D7S820 involves chromosome 7, for which the xenografts CWR22 and CWR22R are trisomic (164). Since full trisomy 7 is lethal in early gestation, only alleles 9 and 10, which are common among the xenografts and CWR22-derived cell lines, were used in the frequency calculation. The frequency for full trisomy 7 would be negligible in the human population. Furthermore, it is unknown whether the trisomy was originally from the CWR22 prostate tumor or was amplified during serial xenograft transplantations. All allele frequencies were determined using the published frequencies for Caucasian-Americans.

Detection of XMRV-like Sequences in Xenografts

To quantify the amount of XMRV DNA in the CWR22 xenografts, a real-time PCR primer-probe set was developed that specifically detected XMRV *env* and excluded murine endogenous proviruses present in BALB/c and NIH3T3 genomic DNA (Fig. 21A). Quantitative PCR of 22R*v1* DNA was used to estimate 20 proviruses per cell (data not shown); and to generate a standard curve. The CWR22 xenografts had significantly fewer copies of XMRV *env* (<1-3 copies/100 cells) compared to the 22R*v1* cells (2000 copies/100 cells). The CWR-R1 cell line had ~3000 XMRV-like sequence copies per 100 cells, and the NU/NU and Hsd nude mice, thought to have been used to passage the CWR22 xenograft, had 58 and 68 copies per 100 cells, respectively (Fig. 21A). Since xenograft tumors are expected to contain a mixture of human and mouse cells, we quantified the amount of mouse DNA by analyzing mouse intracisternal A-type particle (IAP) DNA as previously described (110, 123). All 6 xenografts were positive for mouse DNA with an estimated level around 0.3-1% (Fig. 21B), which was comparable to the <1-3 XMRV *env* sequences per 100 cells detected in the same samples (Fig. 21A).

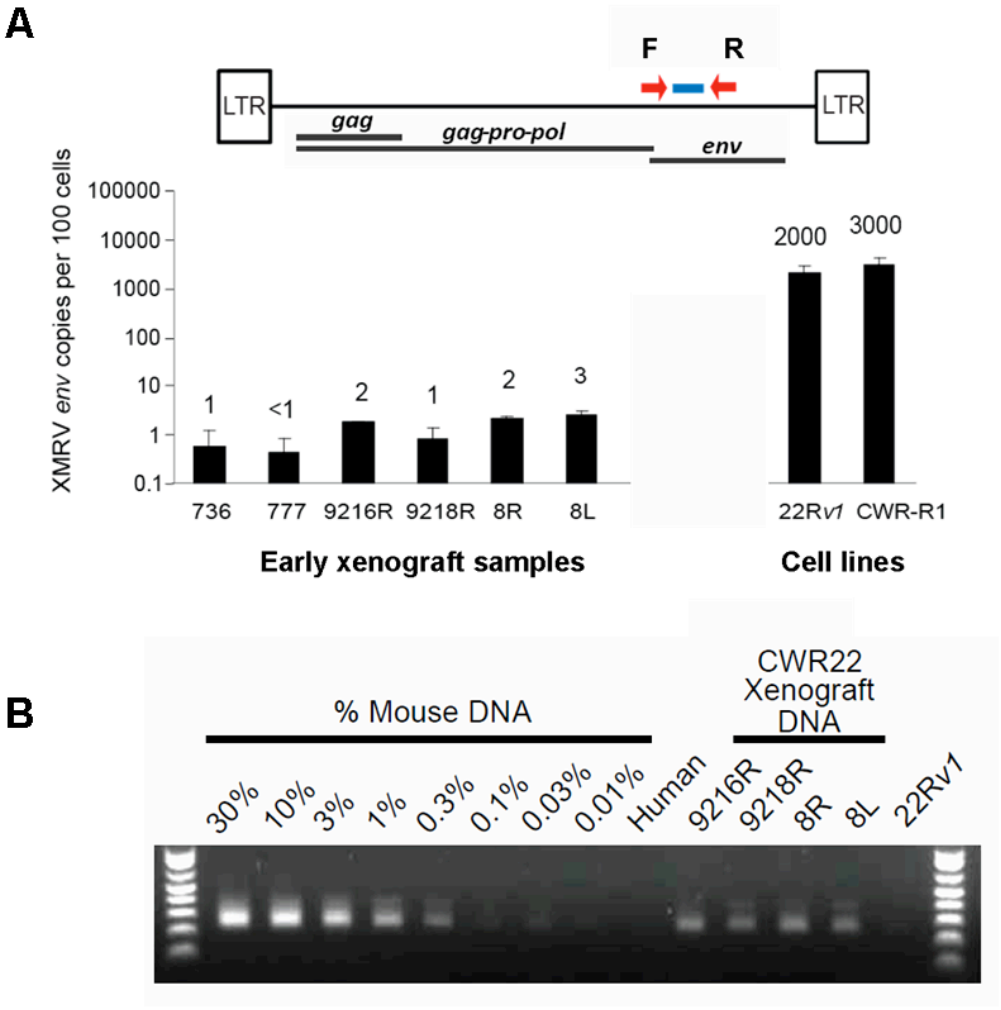


Figure 21. PCR assay to detect XMRV and IAP sequences

(A) Quantitative real-time PCR to detect XMRV env sequences. Probe is shown in blue, primers in red arrows. Calculated copies per 100 cells are indicated above each bar. (B) IAP assay to detect and quantify the amount of mouse DNA present in the xenograft genomic DNAs. Figure from (115), with permission.

We characterized XMRV and related sequences in the xenografts, cell lines, and nude mouse strains by PCR and DNA sequencing (Fig. 22A). Using primers previously used to clone and sequence XMRV from 22R*v1* cells (114), we determined that the XMRV genome in the CWR-R1 and 22R*v1* cell lines were identical (Fig. 22A and 23). Next, we developed several primer sets to specifically amplify XMRV sequences and exclude known endogenous murine retroviruses, some of which was shown in Fig. 14. Primers that specifically amplified XMRV were used to perform PCR on DNA from the late-passage xenografts 2152, 2524, 2272 and 2274; sequencing confirmed the presence of these XMRV sequences in these tumors (Fig. 22A and 19). We also used the XMRV-specific primer sets to amplify and sequence DNA from early-passage xenografts (736, 777, 8L, 8R, 16R, and 18R; Fig. 19B). XMRV *env*, but not *gag* sequences were present, indicating that the early xenografts did not contain XMRV (Fig. 24). The differential results with *gag* versus *env* primers on early and late xenografts were intriguing; we sought to find an explanation by examining the source of the positive XMRV signals.

Identification of PreXMRV-1

The presence of a positive XMRV *env* signal detected by Q-PCR in the early xenografts (Fig. 21) but the absence of any signal for XMRV *gag* sequences by PCR (Fig. 24) seemed contradictory. To address this discrepancy, we amplified and sequenced overlapping provirus fragments by primer walking (Fig. 22B). We found that early xenografts contained a previously undescribed XMRV-related provirus that we have named PreXMRV-1 (Fig. 25). The complete sequence of PreXMRV-1 was determined from all xenografts, the NU/NU and Hsd strains, and the CWR-R1 cell line, all of which were identical.

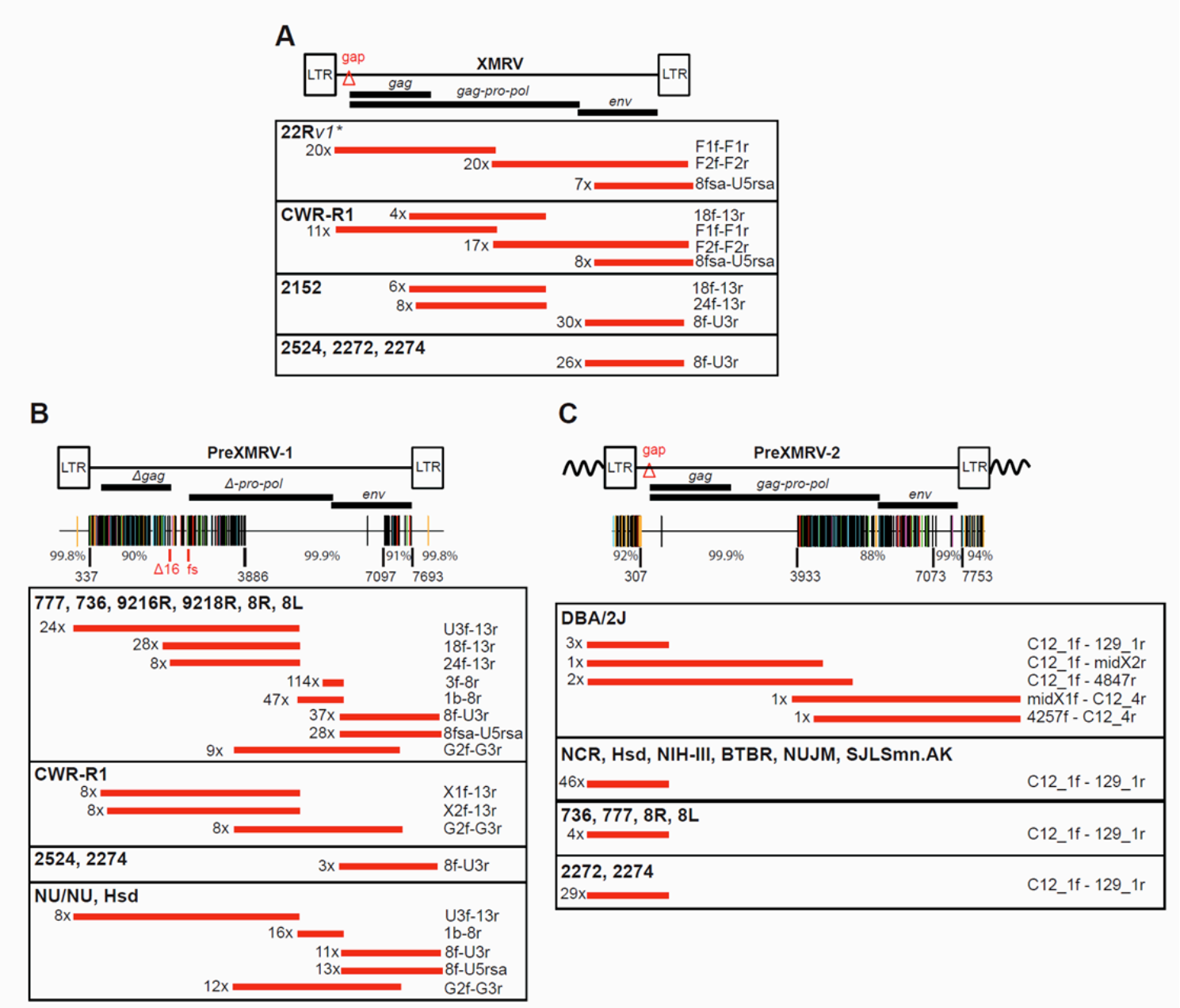


Figure 22. PCR and sequencing analysis of XMRV, PreXMRV-1 and PreXMRV-2

PCR and sequencing analysis of XMRV (A), PreXMRV-1 (B), and PreXMRV-2 (C) from xenografts and cell lines. Single nucleotide differences of PreXMRV-1 and PreXMRV-2 compared to consensus XMRV (1 nt difference compared to XMRV-22R*v1*; FN692043 (114)) are indicated as vertical bars using a modified Hypermut plot (125). Cloned and sequenced PCR products from each indicated source are shown as red bars along with the primer sets used for amplification. Figure from (115), with permission.

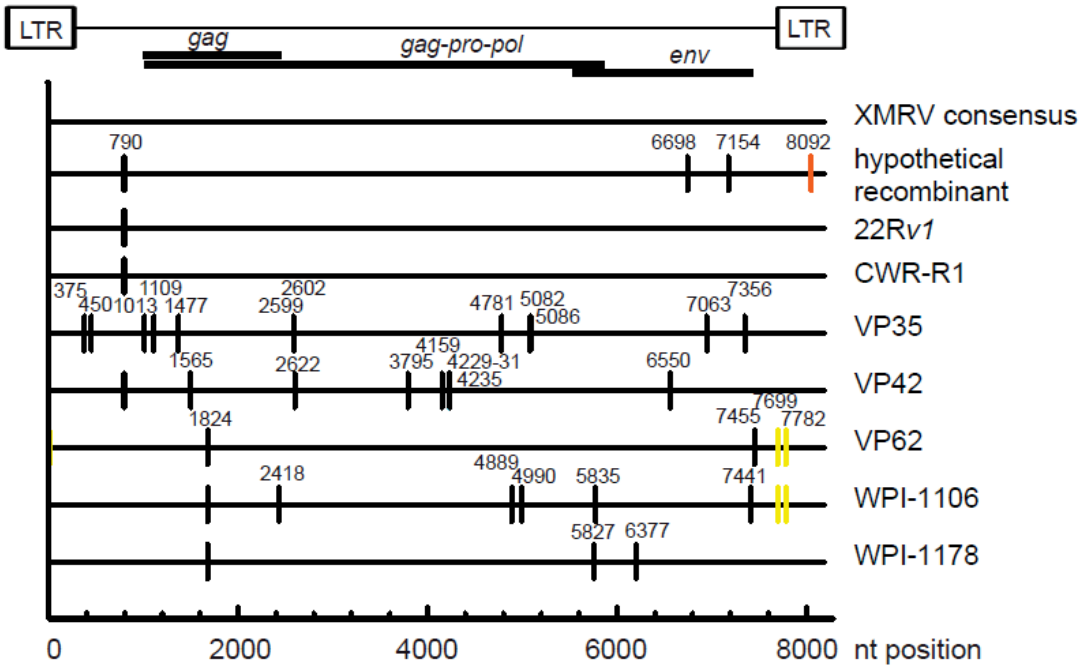


Figure 23. Diversity of XMRV sequences

Multiple alignments comparing the predicted recombinant with the XMRV consensus sequence and all available XMRV full-length genome sequences from cell lines (22Rv1, CWR-R1), PC patients (VP35, VP42, VP62), and CFS patients (WPI-1106, WPI-1178) are shown (94, 114, 163). Differences relative to the consensus sequence are indicated with black vertical bars and deletions with orange vertical bars. The 22Rv1 XMRV sequence differs from consensus XMRV by only 1 nt (at position 790), which is polymorphic among the multiple proviruses in this cell line. Note that differential carryover of A or G into the various isolates implies at least two original contamination events. XMRV sequences from cell lines 22Rv1 and CWR-R1 are identical. Also note that the predicted recombinant would differ from all XMRV sequences only at positions 6698, 7154, and 8902. Figure from (115), with permission.

PreXMRV-1 and consensus XMRV differed by only one nucleotide in a 3211-nt stretch of the genome encoding the 3' half of *pol* and 5' 2/3 of *env* (Fig. 23 and 25C). In addition, the 532-bp PreXMRV-1 LTR was nearly identical to XMRV; PreXMRV-1 had a single adenine deletion relative to XMRV in a run of 6 adenines. The two genomes differed by 10% over the remaining 3.5-kb stretch of *gag-pro-pol* and by 9% in a 600-nt stretch at the 3' end of *env* (Fig. 25C). Late-passage xenografts 2524 and 2274, but not 2152 and 2272, contained PreXMRV-1 (Fig. 24). The detection of low levels of XMRV *env* sequence in early xenografts (Fig. 21A) can therefore be attributed to the PreXMRV-1 proviruses present in the contaminating mouse DNA (see IAP band in Fig. 21B). These results indicate that PreXMRV-1 is an endogenous murine provirus that is present in the NU/NU and Hsd nude strains, but neither of these strains contains XMRV.

The integration site of PreXMRV-1 was cloned from C57L/J genomic DNA, using the same approach for cloning the integration site of PreXMRV-2, with slight modifications (see Materials and Methods). Sequencing the upstream flanking sequence, which was cloned together with the PreXMRV-1 LTR, revealed that PreXMRV-1 insertion is found on mouse chromosome 3. Amplification of host-virus junctions from both ends followed by sequencing confirmed that the location indeed corresponded to PreXMRV-1, with a TSD of "GCAG" (Fig. 25A and B). However, prior to the integration site cloning, all analysis for the presence of PreXMRV-1 was performed by a specific PCR primer set that amplified both PreXMRV-1 and XMRV. Comparing these results with XMRV-specific primers alone confirmed the identity of the band obtained in all cases.

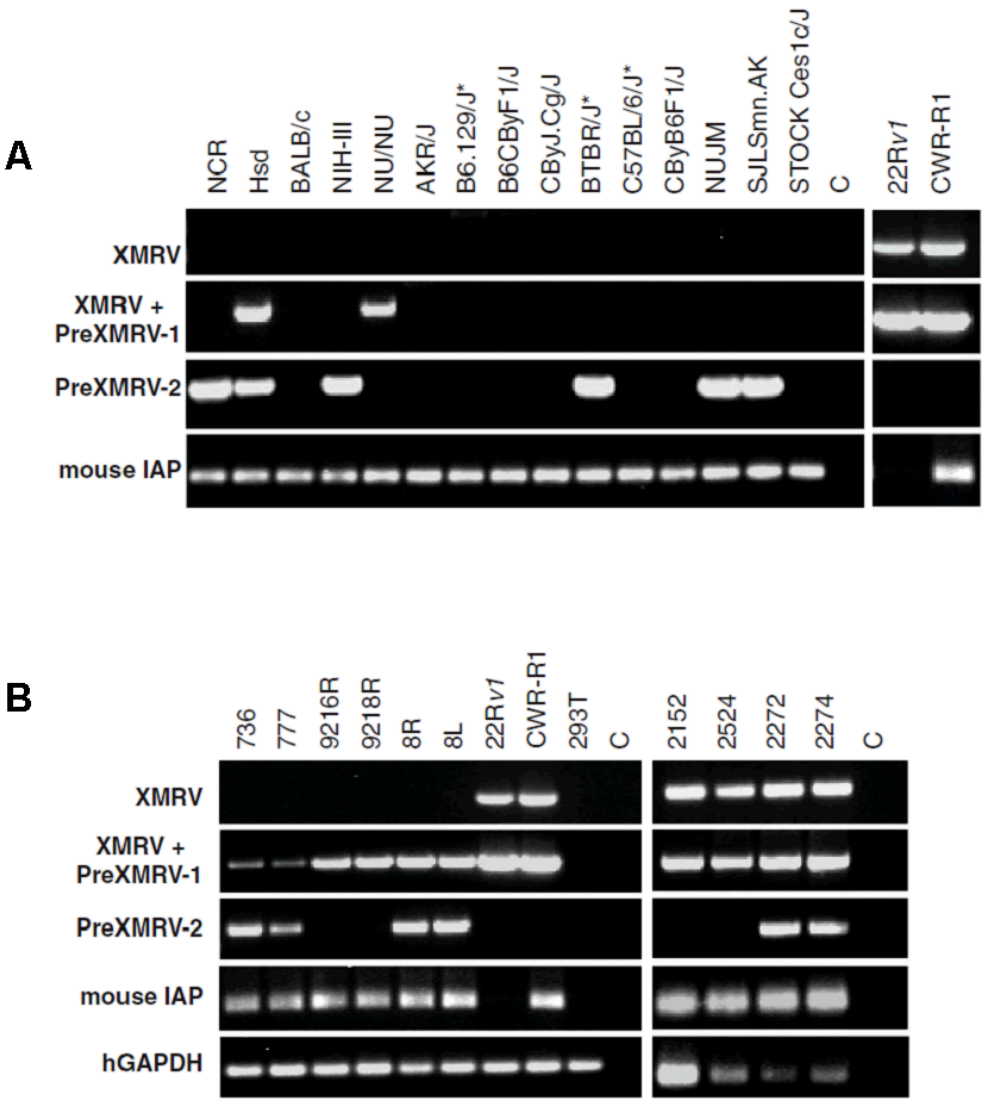


Figure 24. PCR analysis of xenograft samples and prostate cancer cell lines

Xenograft samples and prostate cancer cell lines were analyzed by PCR for the presence of XMRV, PreXMRV-1, and PreXMRV-2. Mouse IAP and human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) served as positive controls for the presence of mouse and human DNA, respectively. The primer set used to detect PreXMRV-1 can also detect XMRV (row 2); however XMRV primers (row 1) would not detect PreXMRV-1. For ease of comparison, the 22Rv1 and CWR-R1 gel lanes from (B), which were run in parallel, are duplicated in (A). *These mouse strains are not nude. Figure from (115), with permission.

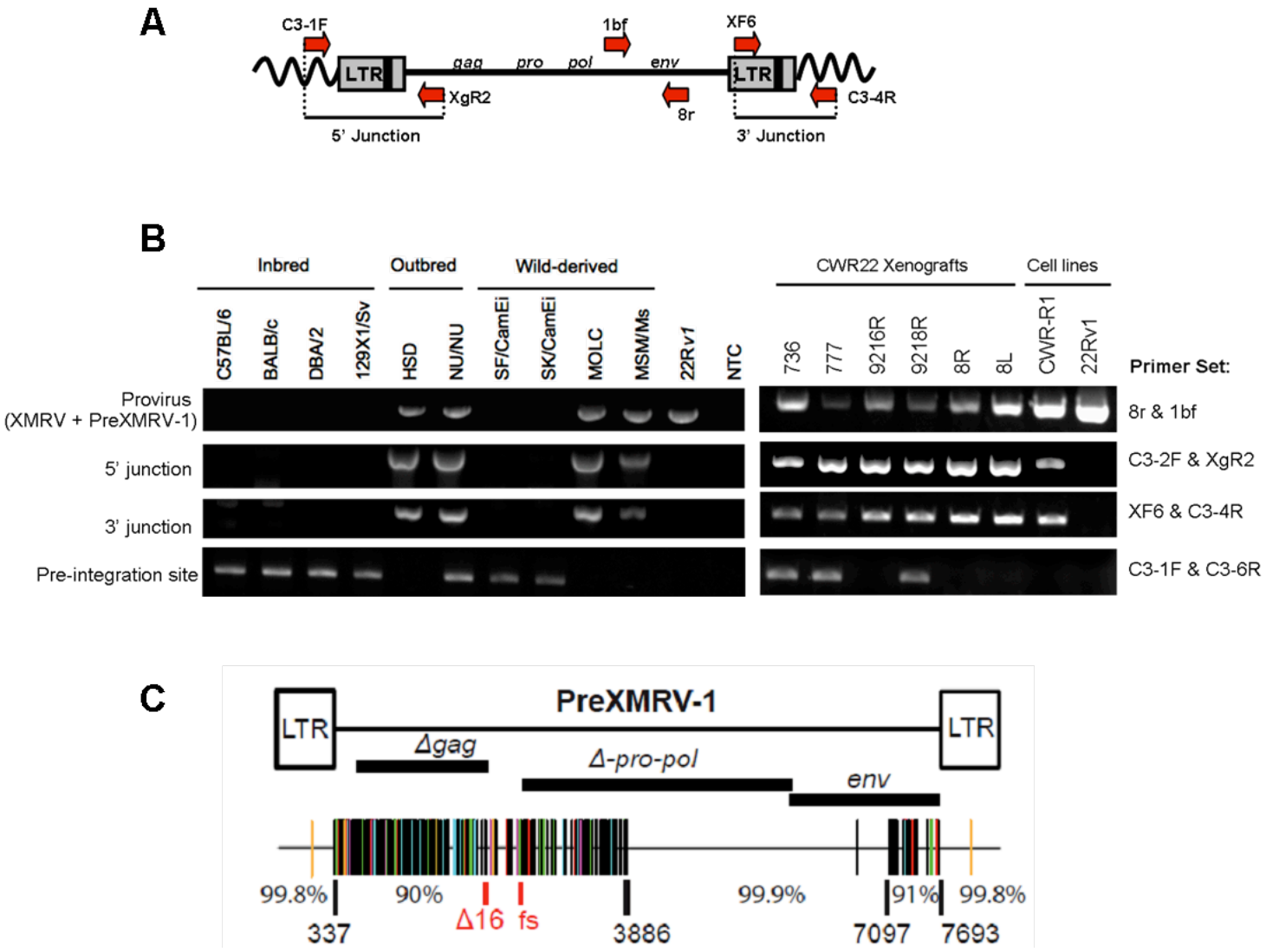


Figure 25. Detection of PreXMRV-1 in mouse strains

(A) PCR approach to amplify host-virus junctions from genomic DNA. (B) Representative mouse strains and xenograft samples were analyzed for the presence of PreXMRV-1 (for a complete list, see Table 3). In 22R*v1* cells, the fragment detected by internal provirus primers is solely XMRV. In CWR-R1 cells, the internal provirus signal is from both XMRV and PreXMRV-1 (C) Hypermut plots indicate nucleotide mismatches relative to XMRV as color-coded vertical lines. Percent identity to consensus XMRV for different regions of each provirus is indicated. below. Nucleotide numbers refer to the 22Rv1-XMRV sequence (FN692043) (114). Figures modified from (26) and (115), with permission.

Both Hsd nude and NU/NU outbred strains are likely to have been used for passaging the prostate tumor xenografts that later gave rise to the 22R*v1* cell line (115). Because both are outbred strains, the number and distribution of specific proviruses may differ among individual mice. To clarify this point, the distribution of XMRV, PreXMRV-1 and PreXMRV-2 was analyzed in a large panel of Hsd nude (n=49), NU/NU (n=5), as well as NIH Swiss mice (n=10), another outbred strain. The NIH Swiss strain was included because it might have contributed to the outbred nude mouse strains (51). Among the Hsd nude mice analyzed, 27% were positive for PreXMRV-1 only, 10% were positive for PreXMRV-2 only, and 53% contained both proviruses, consistent with the null allele frequencies of 0.45 and 0.61, respectively (Fig. 26A and Table 4). Thus, there were enough individual mice within this outbred colony that carried both XMRV parental proviruses with potential for recombination between the two during xenograft experiments (Fig. 26A). This result is consistent with the observation that some (but not all) of the mouse tissues associated with the tumor xenografts contained both PreXMRV-1 and PreXMRV-2 (Fig. 24B). NU/NU outbred mice (n=5) were also variable for the presence of PreXMRV-1 and PreXMRV-2; 2/5 mice contained both PreXMRV-1 and PreXMRV-2 (Fig. 26B). NIH-Swiss mice were more homogenous; all tested samples contained PreXMRV-2, while none contained PreXMRV-1 (Fig. 26C). Most importantly, XMRV was not present in any of the 64 outbred samples tested, confirming previous data on the absence of XMRV in various inbred and wild-derived strains (Table 3, and (115)). Once again, although the detection of PreXMRV-1 in the genomes of the outbred mice relied on internal provirus primers that would have detected XMRV as well, the absence of signal with XMRV-specific primers (while the XMRV-infected 22Rv1 control cell

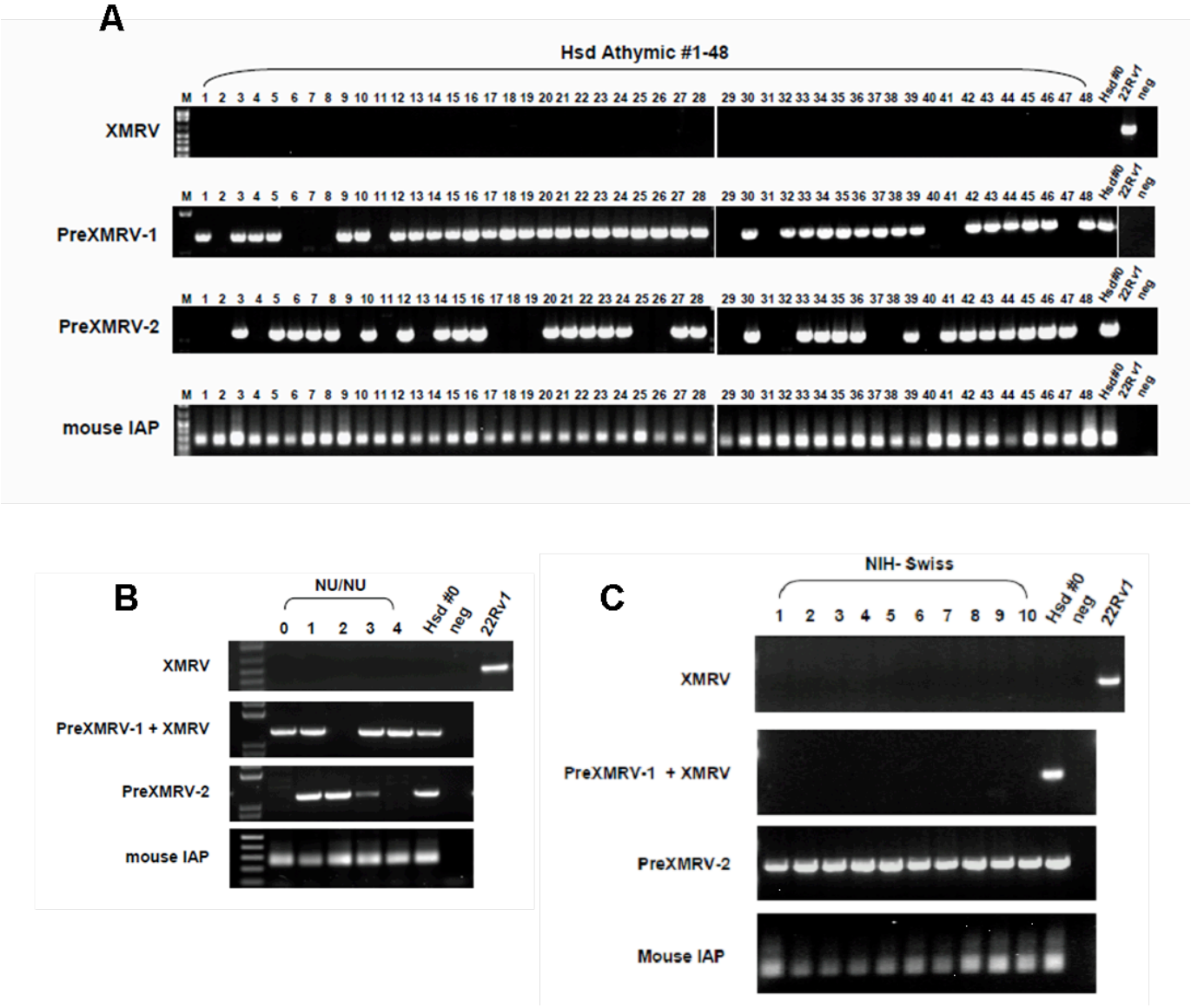


Figure 26. Variability of PreXMRV-1 and PreXMRV-2 in outbred strains

(A) Hsd nude mice (n=49) and (B) NU/NU mice, and (C) NIH Swiss mice (n=10) were analyzed for the presence of XMRV, PreXMRV-1 and PreXMRV-2 by PCR. Mouse IAP elements served as amplification controls. The Hsd and NU/NU samples used in Figs. 11, 15 and 21 are referred to as #0 in each case. Allele frequencies calculated from these results are shown in Table 4. The presence of PreXMRV-2, but not PreXMRV-1 in NIH Swiss mice is consistent with the detection of the former, but not the latter, in NIH3T3 cells (101). Figures modified from (26) and (115), with permission.

line was positive) showed that the detected sequence belongs to PreXMRV-1. Moreover, after cloning of the PreXMRV-1 integration site, our data were confirmed by amplification of the host-virus junctions, with identical results (data not shown). Although not shown here, the strains shown in Fig. 26 were also analyzed for the PreXMRV-1 preintegration site, allowing us to differentiate homozygotes from heterozygotes (Table 4).

Analysis of Xenografts and Mouse Hosts for XMRV, PreXMRV-1 and PreXMRV-2

The xenografts derived from CWR22 were passaged in nude mice, although precise records of the mouse strains used were not available (105, 119). To infer which mouse strains may have been used for the xenograft experiments, and to examine their potential contribution to the resulting XMRV infection, we analyzed various mouse strains for the presence XMRV, PreXMRV-1, and PreXMRV-2. We screened 15 mouse strains, including 12 nude mice, using i) XMRV-specific primers (Fig. 14), ii) primers that amplified XMRV or PreXMRV-1, iii) PreXMRV-2-specific primers, iv) PreXMRV-1-specific primers, after its integration site was cloned, and v) in some cases, PreXMRV-1 empty integration site primers (Fig. 17, 24 and 25). None of the mouse strains contained XMRV and only the Hsd and the NU/NU outbred nude strains contained PreXMRV-1. Six of the 15 mouse strains contained PreXMRV-2, but only the Hsd nude mice contained both PreXMRV-1 and PreXMRV-2. It should be noted that, since Hsd nude and NU/NU are outbred strains, individual mice might differ in their endogenous proviruses. The 22R*v1* cell line contained only XMRV as confirmed by sequence analysis; however the CWR-R1 cell line contained both XMRV and PreXMRV-1. The

	Frequency in Outbred Mouse Strains											
	Hsd (n=49)				NU/NU (n=5)				NIH Swiss (n=10)			
Provirus	+/+	+/-	-/-	AF[1]	+/+	+/-	-/-	AF[1]	+/+	+/-	-/-	AF[1]
PreXMRV-1	0.31	0.49	0.20	0.45	0.40	0.40	0.20	0.40	0	0	1	1
PreXMRV-2	0.63[2]		0.37	0.61	0.60[2]		0.40	0.63	1[2]		0	<0.05

Table 4. Frequency of PreXMRV-1 and PreXMRV-2 in outbred mice

[1]Frequency of the null (-/-) allele.

[2]Heterozygotes could not be distinguished from homozygotes since repeat regions made it impossible to specifically detect the unoccupied integration site.

Table from (26), with permission.

CWR-R1 cell line has been reported to contain contaminating mouse stromal cells (164) (see IAP signal, Fig 24A), which are probably the source of the PreXMRV-1 sequences.

We used the same primer sets to determine the distribution of XMRV, PreXMRV-1 and PreXMRV-2 in early and late xenografts. None of the early xenografts (736, 777, 9216R, 9218R, 8R and 8L), but all of the late xenografts (2152, 2524, 2272, and 2274) and both cell lines were positive for XMRV (Fig. 24B, 25and 18). In our initial experiments, the primers used to detect PreXMRV-1 could also detect XMRV; however, sequencing analysis of the PCR products from all of the early xenografts detected only PreXMRV-1, whereas both XMRV and PreXMRV-1 were detected from the late xenografts 2524 and 2274 (Fig. 24). These results were later confirmed in the early xenograft samples by amplification of the PreXMRV-1 integration site from both 5' and 3' junctions, and the heterozygosity of some of the mouse hosts for PreXMRV-1 was confirmed by amplifying the preintegration site (Fig. 25). Amplification with PreXMRV-2-specific primers revealed the presence of this provirus in early xenografts 736, 777, 8R and 8L, and late xenografts 2272 and 2274 (Fig. 18, 24B and 22C). The variable detection level of PreXMRV-2 in the late xenografts could be due to individual differences in the outbred mice, and by extension, in the mouse DNA in these samples.

Recombination between PreXMRV-1 and PreXMRV-2

Comparison of the PreXMRV-1 and PreXMRV-2 sequences revealed that the regions of near identity to XMRV are reciprocal and largely non-overlapping. We therefore hypothesized that recombination between these two retroviruses resulted in the formation of XMRV. As shown in Fig. 27, reverse transcriptase template switching events during minus-strand DNA synthesis can form a recombinant that is essentially

identical to the sequences of all of the XMRVs reported to date, and differing from the consensus XMRV by only 4 nucleotides. The six switching events occurred in 20-73 nt stretches that are identical between PreXMRV-1 and PreXMRV-2 (Fig. 27; red numbers, and Fig. 28). These regions of identity allow template switching by reverse transcriptase during strand synthesis. Switching events 2 and 3 do not affect the amino acid sequence of the predicted recombinant and are unlikely to contribute to its fitness. Of the four nucleotide differences between the predicted recombinant and consensus XMRV, only the A>G change at position 790, which is absent in XMRV from 22R*v1* and CWR-R1 cell lines, results in a conservative valine to isoleucine amino acid substitution and is found in some XMRV isolates; the other 3 substitutions are silent. The insertion of an A at position 8092 occurred within a run of 6 adenines; such frameshift mutations commonly occur in such homopolymers during retroviral replication (116).

Variation between the Predicted Recombinant and Patient-Derived XMRV Sequences

To date, there are five full-length XMRV sequences available that were reportedly isolated from patients (Fig. 23) (94, 163). We considered the possibility that these patient-derived XMRV sequences represent independent recombination events. All of these viruses have the same six crossover sites that are predicted to occur between PreXMRV-1 and PreXMRV-2; therefore, the same six crossovers plus additional crossovers between PreXMRV-1 and PreXMRV-2 would have to have taken place to generate the XMRV sequences with some of these nucleotide differences. Of the 41 total nucleotide differences present in all five patient-derived XMRVs, 35 nucleotide

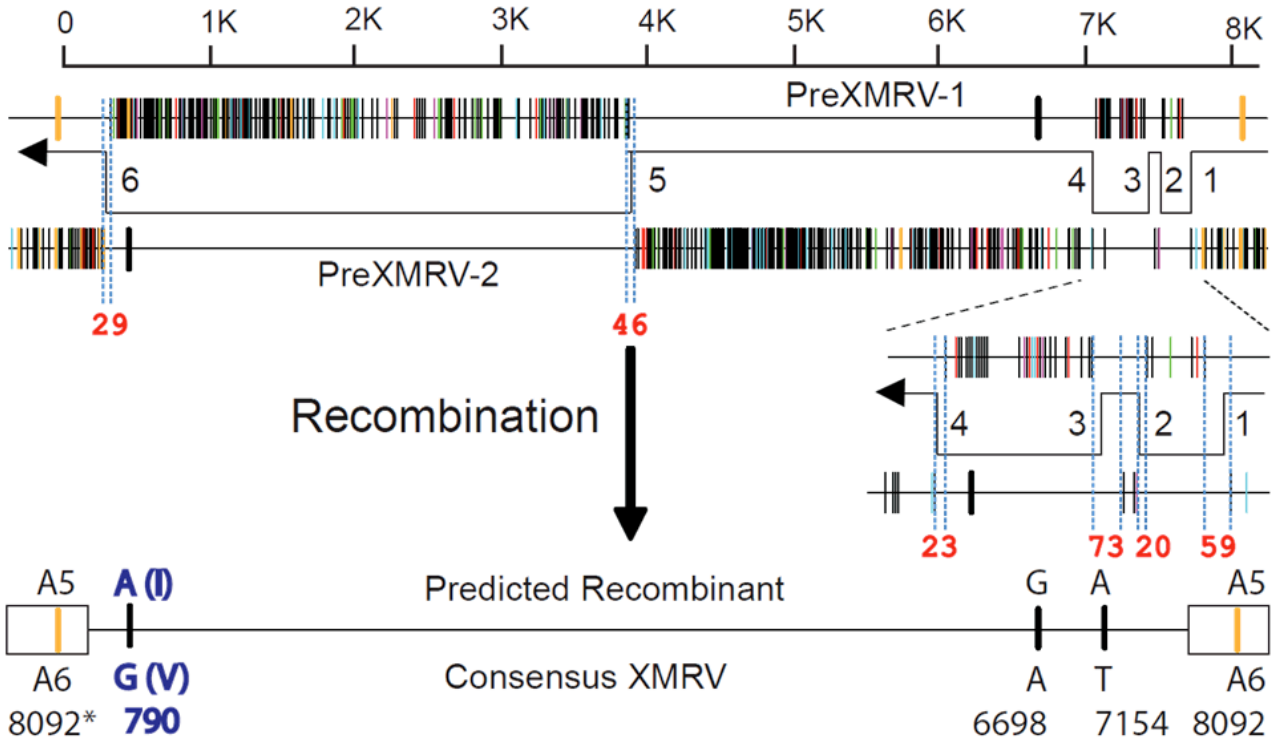


Figure 27. Predicted recombinant between PreXMRV-1 and -2 is identical to XMRV

Alignment of Hypermut plots of PreXMRV-1 and PreXMRV-2 reveals the reciprocal and largely nonoverlapping regions of near identity to XMRV. The direction of minus strand DNA synthesis catalyzed by reverse transcriptase and the predicted template-switching events (numbered 1 to 6) are shown. The lengths of nucleotide identity within the presumed template-switching regions are indicated in red numbers. The predicted recombinant and four nucleotide differences with consensus XMRV are shown. The nucleotide numbers refer to numbers of the 22R*v1* XMRV (GenBank ID: FN692043). Note that nucleotide 8092 is within the U3 region and is present in both LTRs (boxes). A5 and A6 refer to homopolymeric runs of five and six adenines, respectively. The A>G change at 790 results in an isoleucine (I) to valine (V) substitution. Figure from (115), with permission.

Figure 28. Regions of reverse transcriptase (RT) template switching events

The RT template switching events are inferred to be involved in the generation of the predicted XMRV progenitor. Identical sequences between PreXMRV-1 and PreXMRV-2 within the crossover sites. All recombination sites between PreXMRV-1 (top) and PreXMRV-2 (bottom) are shown. Nucleotide mismatches relative to XMRV are indicated in red and the position of the first (or last) mismatched nucleotide is indicated relative to the XMRV-22R*v1* genome. Figure from (115), with permission.

substitutions are not present in either PreXMRV-1 or PreXMRV-2; thus, these nucleotide differences cannot be explained by additional crossovers between these two parental viruses. The remaining 6 nucleotide differences are present in one of the parents; however, none of these substituted sites is flanked by blocks of 20-nt identity. Therefore, it is unlikely that these nucleotide differences would arise through additional recombination events. It is much more likely that these differences represent errors during PCR, errors during sequencing, or variation that arose as a result of passage of the virus in another cell line. The possibility that most of these nucleotide differences represent PCR and/or sequencing errors is strengthened by a comparison of different VP62 sequences available in GenBank. The first deposited VP62 sequence (DQ399707.1) differs from the 22Rv1 sequence by 16 nucleotides; however, a later deposited sequence (EF185282.1) differs from the 22Rv1 sequence by only 4 nucleotides.

Phylogenetic Analyses of PreXMRV-1 and PreXMRV-2

Phlyogenetic analyses of various MLV provirus genomes are shown in Figures 29-32. In the complete provirus genome PreXMRV-1 groups with xenotropic MLVs and PreXMRV-2 groups on a separate branch closest to both polytropic and modified polytropic MLVs. As has been previously noted (163), XMRV sequences group in a distinct clade within the Xmv subgroup (Fig. 29). When only the region of identity between PreXMRV-1 and XMRV is considered (encompassing nt 3933-7073), the XMRV branch now includes PreXMRV-1, while PreXMRV-2 does not change its position compared to full-length genome tree (Fig. 30). Likewise, when only the region of identity between PreXMRV-2 and XMRV is considered (encompassing nt 337-3886),

PreXMRV-2 groups with the other XMRV sequences and PreXMRV-1 does not change its position in the tree (Fig. 31). Analysis of the LTR sequences alone (between nt 7726-8185), where XMRV has a 1-nt insertion compared to PreXMRV-1 in a run of six adenines, shows that once again PreXMRV-1 groups tightly with the XMRV sequences (Fig. 32). The PreXMRV-2 LTR clusters somewhat differently within the Xmv group, suggestive of a recombinant origin. In fact, the genome features of PreXMRV-2 support the hypothesis that this provirus may be a recombinant between a xenotropic and a polytropic MLV. Furthermore, phylogenetic analysis supports the predicted recombinant virus as the precursor of the virus in the CWR22 xenografts, the 22Rv1 and CWR-R1 cell lines, and all XMRVs isolated and sequenced from patients (see inset in Fig. 29). In fact, the estimated sequence of events that led to the creation of XMRV correlates perfectly with timeline of the detection of XMRV sequences supposedly from patients.

Strain Distribution PreXMRV-1 and PreXMRV-2

Despite the identification of two MLV proviruses with large stretches of identity to XMRV, and a mechanism by which they would give rise to XMRV, it remained a possibility that the recombination event might have occurred much earlier during mouse evolution, such that a provirus identical to XMRV could be present in the genome of a mouse strain that was later somehow transferred to humans. While the absence of a PCR signal with XMRV-specific primers in the mouse strains analyzed argued against this possibility, we sought to determine the strain distribution of the two XMRV ancestors, PreXMRV-1 and PreXMRV-2, in wild and inbred mice.

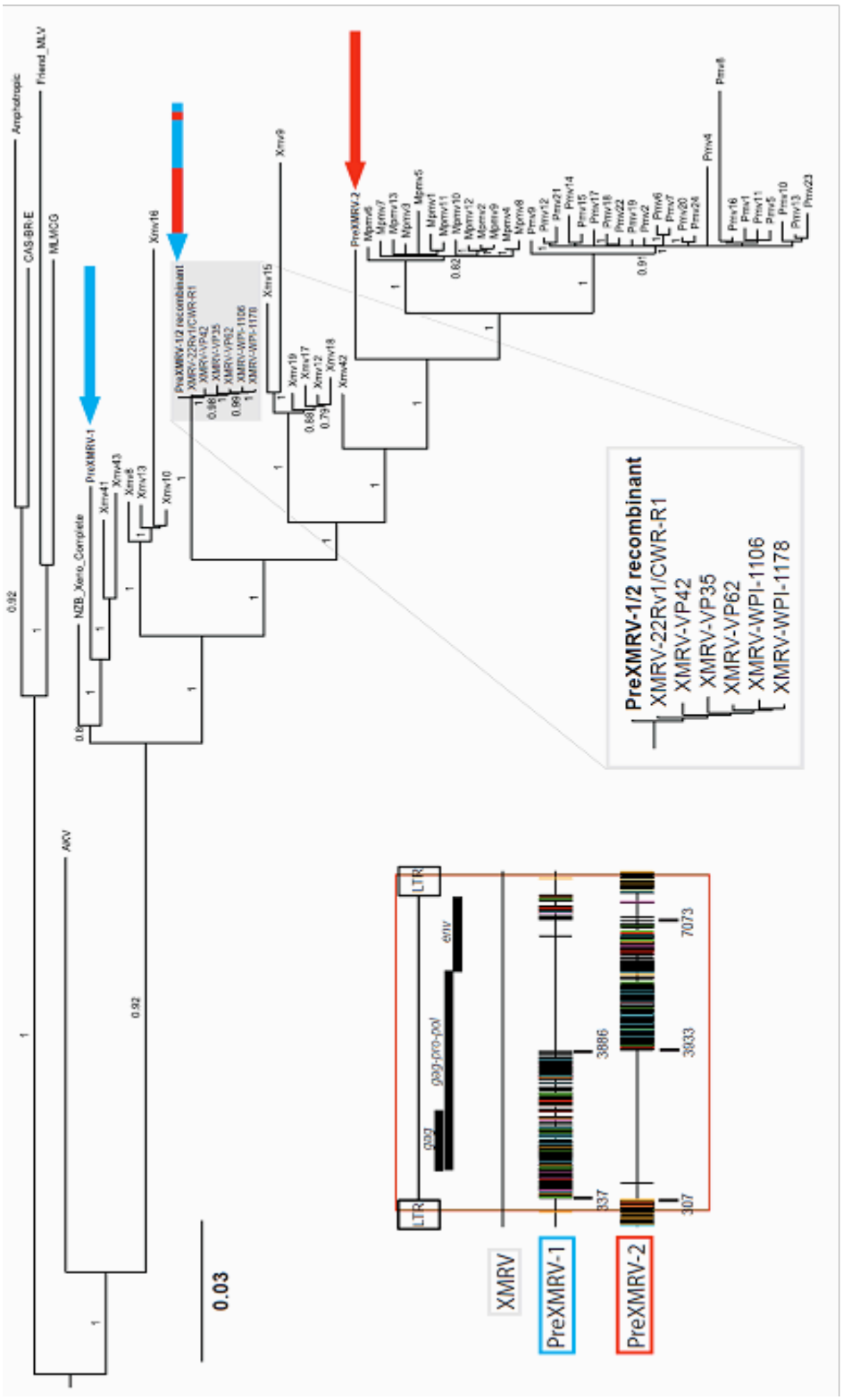


Figure 29. Analysis of the complete genome sequences

Inset shows the enlarged XMRV clade. Blue arrow, PreXMRV-1; red arrow, PreXMRV-2; blue/red arrow, inferred recombinant. Inset shows the XMRV clade, with the inferred recombinant occupying an ancestral position relative to all known XMRV isolates. Figure from (115), with permission.

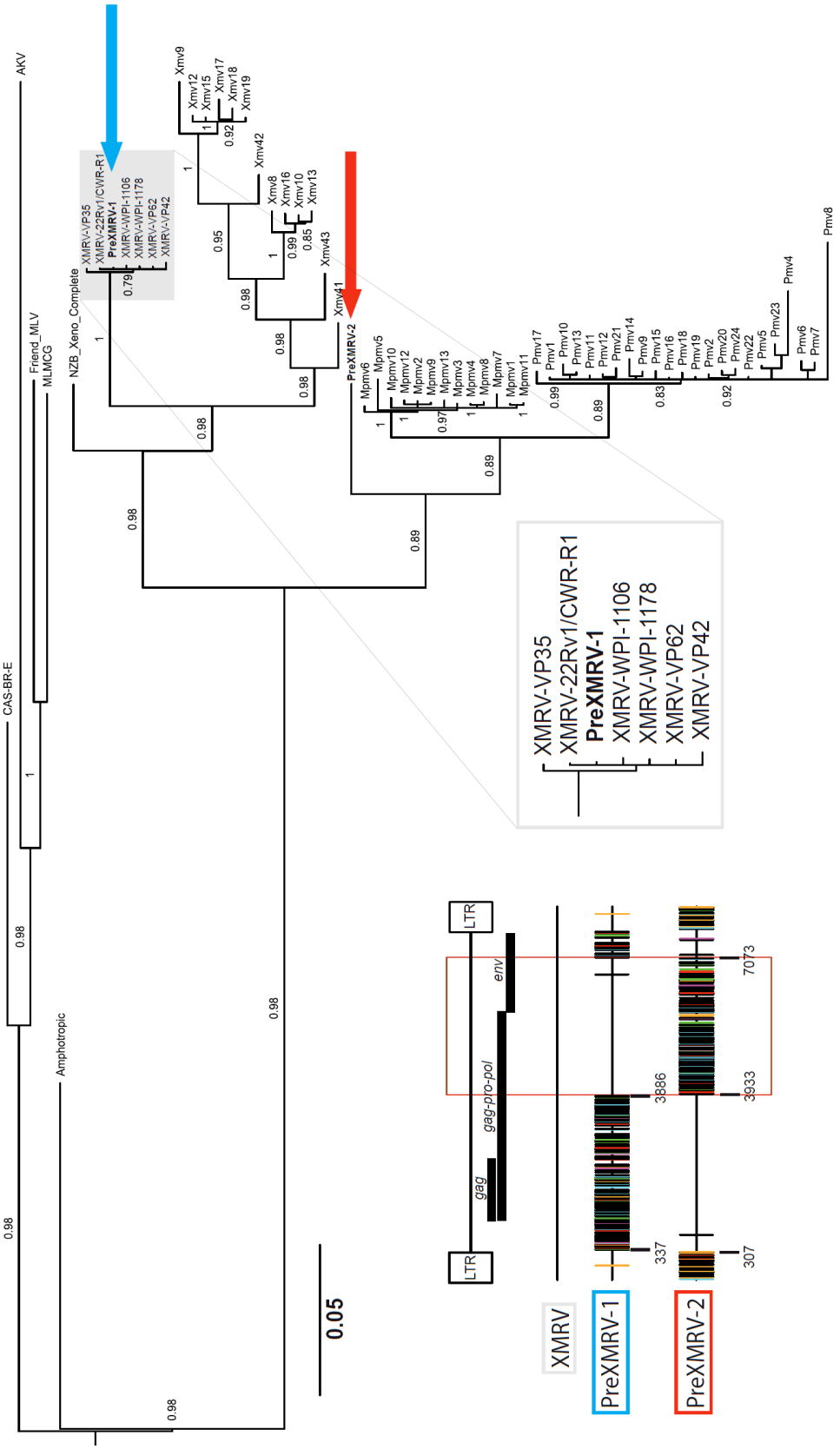


Figure 30. Analysis of the region encompassing nt 3933-7073

Figure from (115), with permission.

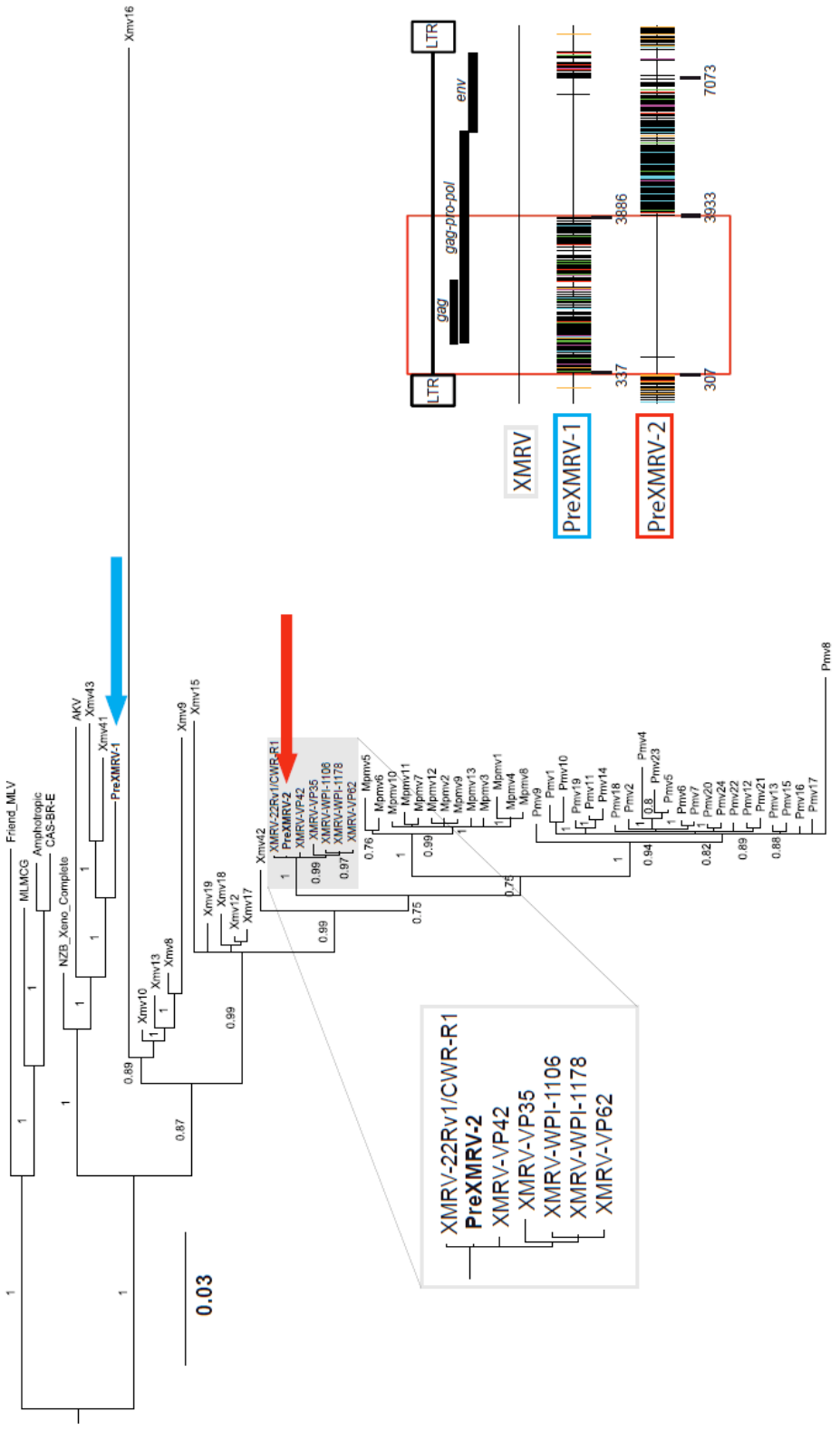


Figure 31. Analysis of the region encompassing nt 337-3886

Figure from (115), with permission.

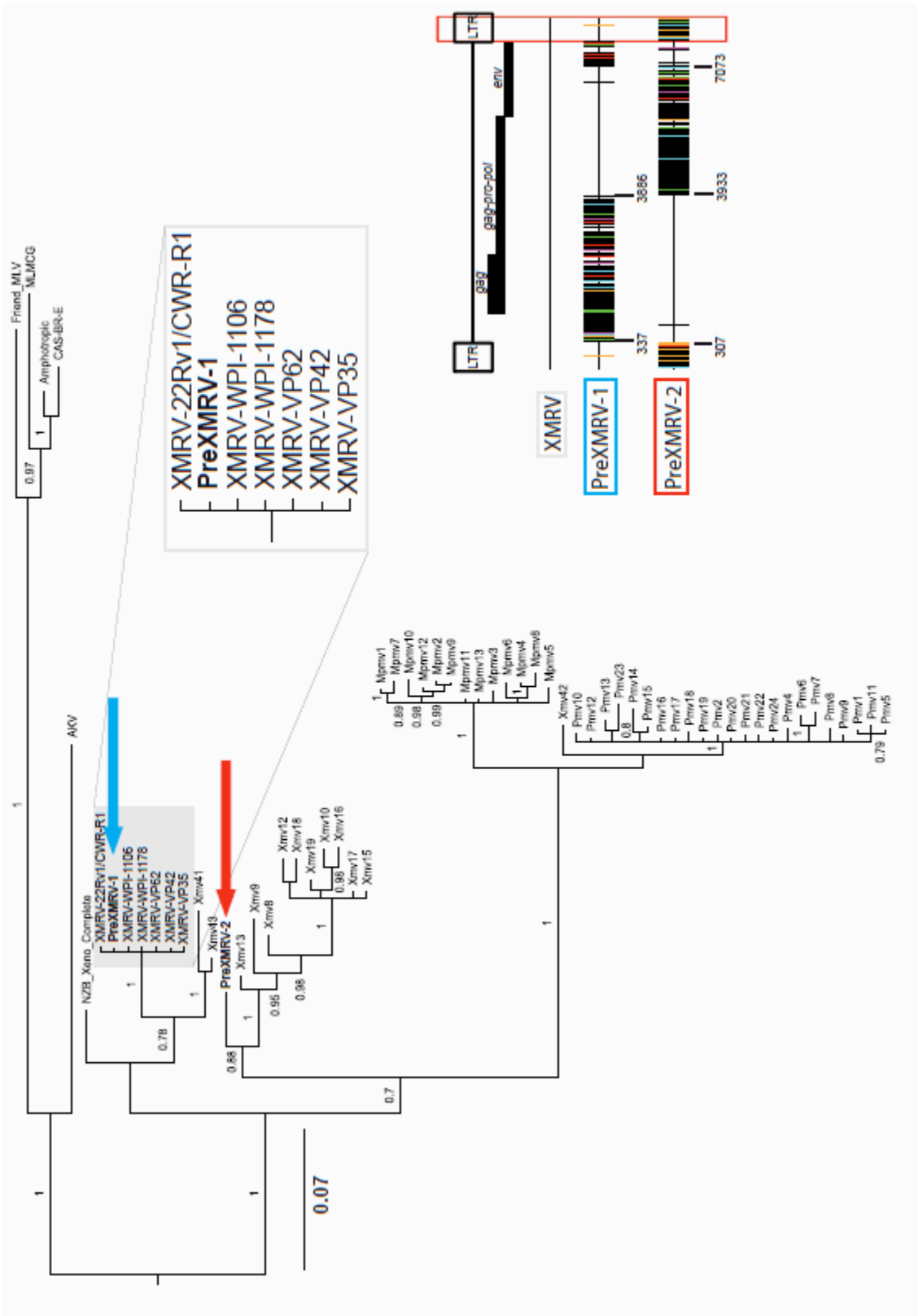


Figure 32. Analysis of the LTR sequences, encompassing nt 7726-8185

Figure from (115), with permission.

Genomic DNA samples from 48 laboratory and 46 wild-derived strains were screened for the presence of PreXMRV-1 and PreXMRV-2. As shown earlier in Fig. 24A and Table 3, no provirus identical to XMRV was present in any of the mouse strains tested (115). These strains were also screened for the presence of PreXMRV-1 and PreXMRV-2 using junction amplification. In the present analysis, 23 inbred, 2 outbred and 2 wild-derived strains were positive for PreXMRV-2 (Table 3). Its strain distribution agreed perfectly with the earlier mouse genome screening results with *gag* leader deletion specific primers (Fig. 18B), consistent with PreXMRV-2 being the only provirus in these strains with the characteristic deletion. PreXMRV-1 was present in 4 inbred, 2 outbred and 8 wild-derived strains (Table 3). Only one inbred strain (C57BR/cd) and two outbred strains (Hsd nude and NU/NU) contained both proviruses.

PreXMRV-1 seems to be quite rare among inbred strains; to date it has been detected only in four inbred strains, C57L/J, C58/J, C57BR/cd and NZW/Lac. Although PreXMRV-2 is absent from the sequenced C57BL/6 mouse genome, it was present in nearly half of the inbred strains tested, suggesting that it is a common insertion. The two wild-derived strains that harbor PreXMRV-2 belong to *M. m. domesticus* subspecies, SF/CamEi and SK/CamEi (Table 3). *M. m. domesticus* is native to the near-East, Europe and Africa, and later colonized the Americas and Australia as human travel between these continents became common (57). PreXMRV-1 was detected in several strains of Asian mice, *M. m. castaneus* and *M. m. molossinus*, the former being an ancestral species that gave rise to the latter by hybridization with *Mus musculus*, and also in the CALB/Rk strain, which carries haplotypes from both *M. m. domesticus* and *M. m. castaneus* (72, 183).

The Xpr1 Receptors in Hsd nude, NU/NU, C57BR/cd and CWR22 Xenograft Mouse DNA are Non-Permissive to XMRV Entry

X-MLV and P-MLVs use different variants of the Xpr1 receptor (Fig. 33) to gain entry into cells (180, 181). Many laboratory strains contain the $Xpr1^n$ variant, which does not support X-MLV entry, while the P-MLV can use this variant to infect cells. We determined whether XMRV, which has a xenotropic *env*, would have been able to infect the cells of the outbred mice used for passaging the tumor. Exon 13 of the Xpr1 gene, corresponding to the ECL4 region, was amplified and sequenced from Hsd nude (n=21) and NU/NU (n=5) mice (Fig. 33). Every individual mouse tested between these two strains had the $Xpr1^n$ allele, which is non-permissive to Xmv entry, suggesting that the XMRV recombinant would not have been able to infect these mice. Exons 10, 11 and 12 of the Xpr1 gene, corresponding to the ECL3 region, were also sequenced from Hsd nude (n=4) and NU/NU (n=3) mice, and carried the $Xpr1^n$ allele. All early xenografts, which contained some host mouse DNA from the original transplant, carried the non-permissive $Xpr1^n$ allele. The absence of a functional receptor to support Xmv entry shows that XMRV could not have infected the cells of the host mouse it was created in and provides further support for the requirement of the transplanted human cells for infection. Lastly, the C57BR/cd mouse, which is positive for both PreXMRV-1 and PreXMRV-2, contained the $Xpr1^n$ allele as well (Fig. 33). This is consistent with a previous report by Baliji et al, which showed that C57/BR, a strain derived from C57BR/cd, also contains the $Xpr1^n$ allele (9).

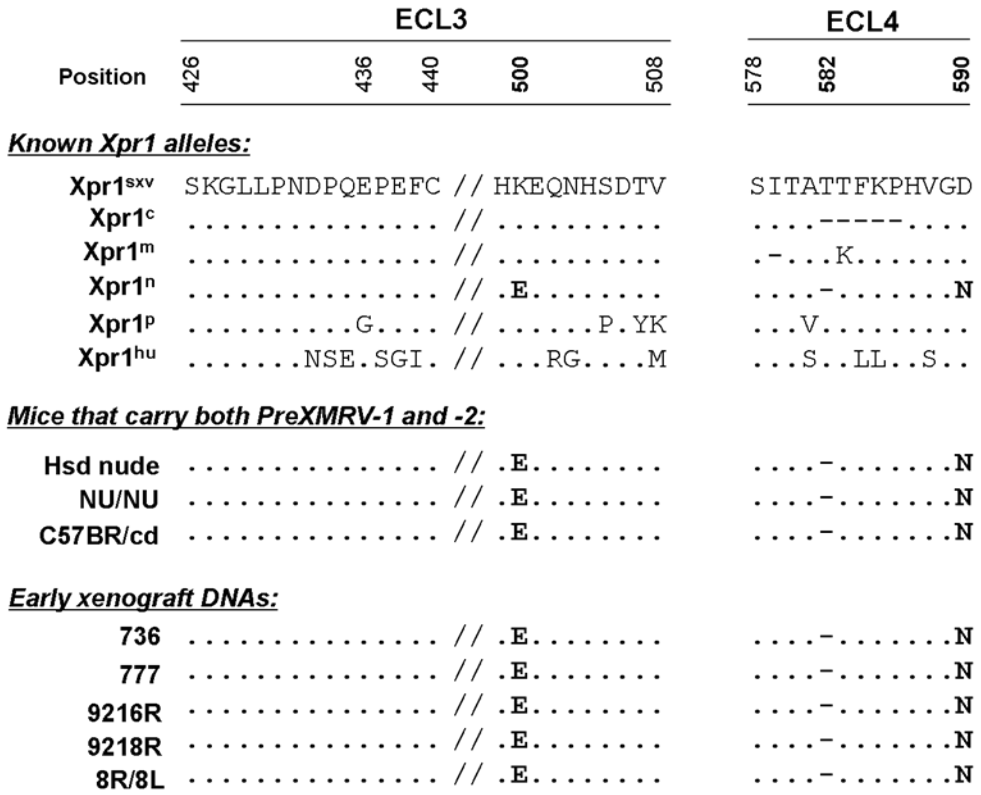


Figure 33. Xenograft DNAs carry the Xpr1^{n} allele

Xenografts, outbred Hsd nude and NU/NU,and inbred C57BR/cd strains carry the Xpr1^{n} receptor allele. Known variants of the XPR1 receptor are shown (upper panel), with the amino acid residues in the ECL3 and ECL4 loops critical for xenotropic MLV Env binding highlighted in bold. ECL3 and ECL4 loops of the XPR1 receptor for Hsd nude, NU/NU and C57BR/cd carry the Xpr1^{n} allele (middle panel), which does not support entry by a xenotropic Env protein (88). Early xenografts containing either PreXMRV-1 alone, or both proviruses together (see Fig. 24) carry the Xpr1^{n} allele in the mouse tissues found associated with the xenografts (lower panel). Figure from (26) with permission.

Characterization of the PreXMRV-2 Genome

The RNA genome of PreXMRV-2 is 8193-bp long, while the provirus insertion in the mouse genome is 8662-bp long with identical LTRs (539 bp) (GenBank ID: FR871850). The genome has intact *gag-pro-pol* and *env* open reading frames with no obvious deleterious mutations, suggesting that it could potentially encode infectious virus. The most striking feature is a ~3.5-kb stretch encompassing the *gag* leader region, *gag* and part of *pol* genes, 100% identical to XMRV released from the 22R*v1* cells (114), while the rest of the provirus differs by ~11% (115).

PreXMRV-2 carries xenotropic type LTRs of 539-bp in length, which lack the 150- or 190-bp insertion found in Pmv and Mpmv LTRs, respectively (81). Amplification and sequencing of complete 5' LTR regions from 20 positive strains revealed that the LTRs in different mouse strains are also identical, flanked by the same "GGAA" TSD. 3' LTRs and flanking chromosomal region from three strains were sequenced and the two LTRs were identical in all cases.

Analysis of the PreXMRV-2 genome revealed two intact open reading frames encoding *gag-pro-pol* and *env*, as expected. The Gag-Pro-Pol polyprotein would be generated via translational read-through of the stop codon (UAG) at the end of *gag*, as is the case with XMRV and other MLVs. As with XMRV, a mutation in the gag leader would prevent synthesis of the glycosylated Gag protein expressed by many MLVs (163). The tRNA primer binding site (PBS) is complementary to Gln2, commonly used by many MLV subgroups (29, 74). The genome of PreXMRV-2 from positions 297 to 3921 is identical to XMRV released from the 22R*v1* cell line, which includes most of the *gag* leader region, as well as the entire *gag* and part of the *pol* genes (115).

The stretch of PreXMRV-2 genome that did not contribute to the final XMRV recombinant starts within the 5' region of *pol* gene that encodes RT, around amino acid 450 of RT, before the RNase H domain. Thus, the RT of XMRV is actually a recombinant between those of PreXMRV-1 and PreXMRV-2. BLAST searches using the 3' half of PreXMRV-2 genome revealed that part of the provirus sequence had been previously identified in NFS/N mice (40). The 2587 bp NO1 clone described in that study (GenBank ID: AY219565.2) corresponds to the 3' end of *pol*, *env* and U3 regions of PreXMRV-2, and differs from PreXMRV-2 by only a single nucleotide, resulting in an amino acid change from Trp to Arg at residue 46 of Env.

The PreXMRV-2 *env* gene encodes a protein of 645 amino acids, grouping more closely in phylogenetic analyses with Pmv and Mpmv *env* sequences than with Xmv *env*, although not closely with either group (40, 115). It contains a unique 12-bp insert in the SU region described previously by Evans et al., as well as the 27-bp insertion characteristic of Pmv *env* genes relative to those of Mpmvs (40). Despite the presence of this insert, this provirus would not have reacted with probes previously designed to detect polytropic *env* genes (JS5) due to several mismatches in the probe binding site (145). This difference most likely prevented the detection of PreXMRV-2 in previous studies with polytropic *env* probes to analyze the chromosomal location and distribution of nonecotropic MLVs in mice (44, 145, 157).

Characterization of the PreXMRV-1 Genome

The RNA genome of PreXMRV-1 is 8197 bp long (GenBank ID: FR871849) with typical xenotropic MLV features. PreXMRV-1 donated its LTRs and the 3' part of its *pol* and *env* genes to the resulting XMRV recombinant; therefore we will not go into

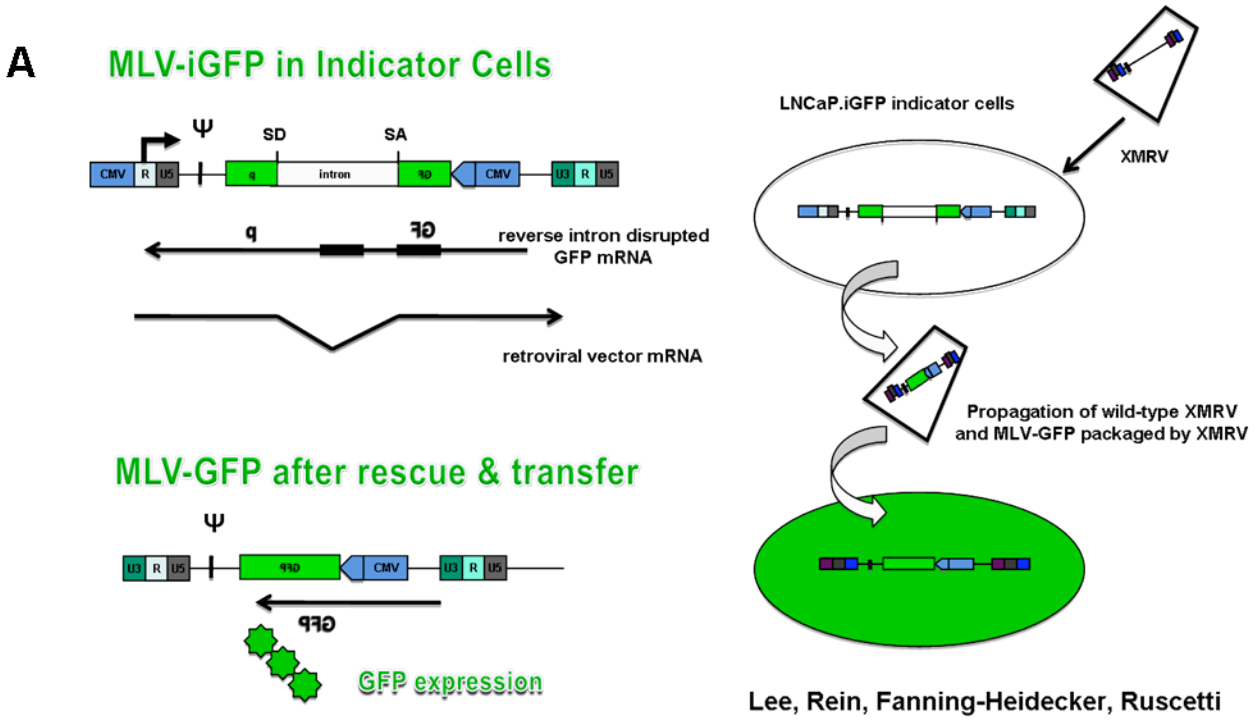
detail regarding these regions of the PreXMRV-1 genome. PreXMRV-1 is replication defective because of a 16-nt deletion in *gag* and a +1 frameshift mutation in *pol* (Fig. 25). The defective regions of PreXMRV-1 were apparently rescued by the corresponding region of PreXMRV-2 to generate XMRV (115).

PreXMRV-1, like XMRV, has a xenotropic *env* gene, which would have been detected in previous studies from our lab analyzing nonecotropic MLV distribution in wild and inbred mice using *env* probes (45, 146). Therefore, we compared the strain distribution of PreXMRV-1 and the expected sizes of EcoRI and PvuII restriction fragments in unblots of mouse genomic DNA to a large number of endogenous xenotropic MLVs previously identified in our laboratory using recombinant inbred (RI) strains (45). The closest provirus candidate based on these parameters was Xmv39, previously identified in C57L/J strain using the AKXL RI mice (a cross between AKR/J and C57L/J strains) (45). We tested the 14 recombinant inbred AKXL mice used in the original study for the presence of PreXMRV-1, to compare the distribution of Xmv39 with that of PreXMRV-1, which would allow us to determine if they represent the same provirus. Our data revealed that these two proviruses did not represent the same insertion (data not shown). The chromosomal location of Xmv39 was not known at the time of its publication; however, one of the markers that segregated with Xmv39 was later mapped to mouse Chr 10, whereas we have mapped PreXMRV-1 to mouse Chr 3, suggesting that the two Xmv proviruses are different insertions in the mouse genome.

Infectious Properties of PreXMRV-2

Although PreXMRV-1 is replication defective, PreXMRV-2 has an intact genome devoid of deleterious mutations that could potentially encode an infectious virus. To test

whether PreXMRV-2 was indeed infectious, the two halves of the provirus were amplified using flanking chromosomal primers and internal provirus primers. The full-length provirus was cloned into a TOPO vector and transfected into indicator cells to monitor for virus production. We used 293-iGFP-puro (DERSE-293) and LNCaP-iGFP-puro (DERSE-LNCaP) indicator cells, which contain a defective GFP cassette that can be mobilized by an infectious retrovirus, allowing its expression in the second round of infection (Fig. 34A). Briefly, DERSE cells express a defective retroviral vector with a GFP cassette inserted in the reverse orientation, which also contains an intron inserted in the forward orientation relative to the provirus. The GFP gene can be mobilized by an infectious virus entering the cell, allowing visualization of the infected cells under a microscope (Fig. 34A). Transfection of DERSE-293 cells with full-length clones of XMRV, Bxv-1 and PreXMRV-2 showed that, while XMRV and Bxv-1 are infectious, PreXMRV-2 is not (Fig. 34B). Even though the time point shown in Fig. 34B is day 4, the results were still negative at day 8. This finding was further confirmed by a different, canine-based, indicator cell line (DIG cells) (23), where the results were still negative at day 16 (Tobias Paprotka, unpublished observation). Taken together, PreXMRV-2 does not seem to be infectious, even though it contains intact open reading frames.



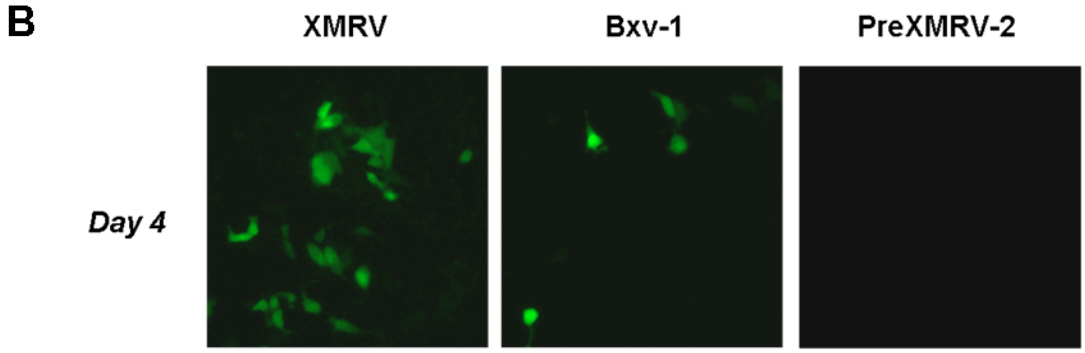


Figure 34. Overview of the DERSE indicator cells

(A) Cells expressing a defective retroviral vector with a GFP cassette inserted in the reverse orientation that also contains an intron inserted in the forward orientation relative to the provirus. The GFP gene can then be mobilized by an infectious virus that enters the cell, allowing visualization of infected cells under a microscope (Figure from Vineet KewalRamani). (B) Complete proviral genomes of XMRV, Bxv-1 and PreXMRV-2 were transfected into 293-based DERSE cells, GFP expression was scored at day 4.

DISCUSSION & FUTURE DIRECTIONS

The identification of XMRV, which was initially thought to be a novel human gammaretrovirus, closely related to endogenous retroviruses of mice, pointed to the possibility of a cross-species transmission, a very interesting phenomenon from an evolutionary perspective. The finding that the presence of this novel retrovirus was associated with two human diseases further increased the interest of the scientific community in the context of public health and infectious disease.

Initial studies that identified XMRV in certain patient cohorts and investigated the association of XMRV with human disease found compelling experimental evidence (94, 163). However, numerous subsequent studies failed to reproduce the original findings, despite the use of adequate assays and controls (5, 30, 37, 42, 53, 55, 67, 68, 75, 129, 132, 149, 166, 168). Science progresses as a continuum, albeit discontinuously; novel findings are always based on previous research, and previous results have to be reproducible. Science is objective; correlation does not mean causation, no matter how desperate for an answer a patient cohort might be.

The doubts about the connection of XMRV with human disease were strengthened by other findings besides negative results in various patient cohorts. The degree of sequence diversity in reported patient samples was not consistent with the presence of an infectious retrovirus, which would be expected to have a high error rate. Moreover, several reports described contamination of laboratory reagents with mouse DNA, raising concerns about the true source of MLV-like sequences amplified from banked patient samples (92). In addition, a fraction of the previously detected integration

sites in the human genome turned out to be from experimentally infected cells, calling into question the authenticity of the rest of the reported integration sites (49, 82, 83).

In the present study, we examined the origin of XMRV and what seemed to be the source of the zoonotic transmission to humans. Solely based on the sequence of the retroviral genome, XMRV was widely accepted to have derived from one or more MLVs, commonly found as endogenous viruses in mouse genomes, even though no known MLV provirus identical to XMRV was known.

One of the first clues regarding the origin of XMRV came in 2009, with the discovery that the human prostate cancer cell line 22R*v1* produces a virus that is identical in sequence to XMRV detected in patient samples. This finding was taken to mean that the original tumor that gave rise to the cell line was infected with XMRV. We examined the history and generation of the 22R*v1* cell line to figure out if this supposition was really the case. Careful analysis of samples obtained during the generation of this cell line revealed that early xenograft passages of the tumor did not contain XMRV, whereas the later passages did, suggesting that XMRV infected the tumor after initial passaging and prior to the generation of this cell line (115). Moreover, we identified two endogenous MLV proviruses from the xenografts, PreXMRV-1 and PreXMRV-2, which contained long, non-overlapping regions of identity to XMRV, such that recombination between the two MLVs would create a virus identical to XMRV, providing the most likely explanation of how and when XMRV arose. The host mouse DNA found associated with the tumor contained both XMRV parental proviruses, but not XMRV, strengthening the argument that recombination between these two endogenous MLVs led to the generation of XMRV.

Here, we discuss the implications of our findings and future research directions.

The Distribution of PreXMRV-1 and PreXMRV-2

Identification of the two XMRV parental proviruses, PreXMRV-1 and PreXMRV-2, in the xenograft samples that gave rise to the 22Rv1 cell line presented the most likely scenario of how and when XMRV was formed; i.e. through recombination between the two proviruses during passaging of the xenografts in nude mice. However, there remained the alternative possibility that the recombination event between the two ancestors occurred during mouse evolution, the resulting XMRV recombinant became endogenized, and then found its way to the patient whose tumor cells later became the 22R*v1* cell line. We explored this alternative scenario, and found it to be very unlikely, for reasons explained below.

Our screen of ~170 individual mice from 48 laboratory strains and 46 wild-derived strains from all over the world failed to detect XMRV as an endogenous provirus. Thus it is very unlikely that the original CWR22 tumor was infected with XMRV derived from an endogenous XMRV provirus. In addition, the NU/NU and Hsd mice are the only mice among the 12 nude mouse strains that we analyzed that contain both PreXMRV-1 and PreXMRV-2 (Fig. 26). Furthermore, no wild mouse analyzed carried both proviruses. Taken together, the possibility is remote that the CWR22 tumor was originally infected with XMRV, which would have to come from an extremely rare putative endogenous XMRV provirus, and by coincidence, was transplanted into rare nude mice that harbor both parental proviruses. Although the possibility that the CWR22 tumor was initially infected with XMRV can never be completely excluded, we conclude

that it is far more likely that XMRV arose through recombination between PreXMRV-1 and PreXMRV-2 during passaging of the CWR22 xenograft in nude mice.

The distribution of PreXMRV-1 and PreXMRV-2 in mice is rather different. Besides the two outbred strains mentioned above, the only other mouse strain that harbors both proviruses is an inbred strain, C57BR/cd. Interestingly, C57BR is among the inbred strains with the greatest *M. m. musculus* content on a largely *M. m. domesticus* background (183). It is particularly noteworthy that the two proviruses found their way into laboratory mice from wild subspecies found on two different continents, Asia (*M. m. molossinus and M. m. castaneus*) and Europe (*M. m. domesticus*), and are, therefore, very unlikely to have been present in the same mouse until their recent crossbreeding to create the fancy and laboratory mouse strains (171). The fact that neither provirus is fixed in the subspecies where it occurs suggests that both are of fairly recent origin, having integrated after the origin of these subspecies less than one million years ago, and prior to the radiation of *M. m. domesticus* in the Americas or the hybridization that gave rise to *M. m. molossinus* (57, 185). As mentioned above, not a single wild mouse strain was found to carry both PreXMRV-1 and PreXMRV-2, making it extremely unlikely that the two parents would be present in the same mouse in the wild, recombine, and become established as an endogenous MLV, let alone cross species to infect humans.

Particular Issues Regarding the Detection of PreXMRV-2

The presence, in a number of mouse strains, of a provirus carrying the characteristic 24-bp *gag* leader deletion has been previously noted (31). Since the strain distribution of that provirus by PCR using *gag* leader deletion-specific primers is in perfect agreement with that of PreXMRV-2, we think that this is the only provirus in

these strains with the characteristic deletion. Proviruses with the *gag* leader deletion were also detected in four inbred strains by deep sequencing (71). Although those fragments most likely correspond to PreXMRV-2, the short length of the deep sequence reads (167 bp) precludes a firm conclusion.

We cannot predict whether the size of the band observed in our unblots corresponds to the expected band size for PreXMRV-2 based on the genomic locus from the C57BL/6 database, because even the current mouse genome build (version 37) contains large stretches of N's, making it impossible to infer the presence of a restriction site beyond the available flanking sequence information (data not shown). It is also important to note that PreXMRV-2 would have been detected with xenotropic LTR specific probes (Xltr) in previous studies from our laboratory analyzing the distribution of nonecotropic MLVs in mice (43, 157), but not with the Xmv subclass-specific probes (KT-51 and KT-55) due to a 3-bp mismatch with both probes. PreXMRV-2, despite the presence of a polytropic *env* gene, also would not have been detected with the polytropic *env*-specific probes (JS5), also due to a three base mismatch at the probe-binding site.

Recombination between PreXMRV-1 and PreXMRV-2

The generation of recombinant viruses with similar crossover patterns can occur multiple times, reflecting selection for advantageous properties such as receptor usage, expression in target cells, evasion of host restriction and repair of defects in one or the other parental virus, resulting in a selective replicative advantage gained by the virus through bringing together of different genomic stretches. In the present case, PreXMRV-1 has obvious defects in *gag* and *pol* that were repaired by the corresponding regions in PreXMRV-2. Even in such cases, however, the exact crossover sites vary from one

recombinant to the next, such as the case with MCF viruses in AKR mice (147). It is worth emphasizing that the probability of generating the same XMRV recombinant with the same crossovers independently more than once is extremely small (115). In fact, among the many hypothetical recombinants that could form between PreXMRV-1 and PreXMRV-2, there are ~10^{34} possible recombinants (without regard to the number of crossover events per round of replication), of which about 10^{17} would generate a virus with identical amino acids to XMRV in all proteins, but with a different genome (data not shown). In other words, assuming that there is a selection pressure from the human host acting on every single amino acid residue of the virus, thus imposing a requirement for recombination to occur at certain sites, it would still be possible to generate a virus with identical proteins by 10^{17} different ways. Such a flexible window of redundancy for generating a virus with identical proteins further highlights the biological improbability of creating the exact same recombinant twice. In a hypothetical situation where, by an extremely remote chance, such an event did occur, the recombinant (XMRV) would be unable to reinfect the mouse target cells and spread, since the host mice used for xenografting carry the $Xpr1^{n}$ receptor variant non-permissive for XMRV infection (Fig. 33) (180, 181). Once again, if we explore the remote possibility that the specific recombination event occurred in a mouse strain that carried both proviruses, the recombinant would still be unable to infect those mouse cells, since all three strains that carry both proviruses, Hsd nude, NU/NU and C57BR/cd mice, carry the non-permissive allele. The presence of the grafted human tumor tissue therefore gave the recombinant the opportunity to infect and propagate.

It will be very interesting to explore the properties of various recombinants that form when the two XMRV ancestors, PreXMRV-1 and PreXMRV-2, are co-expressed in tissue culture. There are multiple stretches of identity between the two proviruses with many potential regions for recombination. Any recombinant that can avoid the inclusion of the 16-bp deletion and the frameshift mutation in PreXMRV-1 could potentially be infectious, as all remaining regions of both PreXMRV genomes are devoid of obvious defects. However, even though PreXMRV-2 has intact open reading frames, preliminary results indicate it is not replication-competent (Fig. 34). From the fact that PreXMRV-2 has identical sequences in *gag* and the 5' half of *pol* to the replication-competent XMRV, and that recombination between two defective viruses requires that they be able to complement one another when co-expressed to make infectious virions, we can infer that the PreXMRV-2 *gag* and *pol* genes should encode functional proteins. Analyzing the sequences of various recombinants formed by coexpression of the two *in vitro* will show which parts of retroviral proteins are conserved in the generated viruses, giving us clues as to which regions are selected for, in order to overcome potential blocks against their replication.

No full-length, infectious, endogenous polytropic MLV has been described to date, although there are numerous examples of endogenous and exogenous viruses that acquire *env* genes from endogenous polytropic MLVs to generate an infectious recombinant virus (39, 61, 78). It has been generally thought that polytropic LTRs may have an intrinsic defect, perhaps relating to the characteristic insert sequences. To our knowledge, no studies to date have explicitly addressed this issue. Xenotropic LTRs are strong drivers of transcription, contain multiple promoter and enhancer elements (154)

and Xmvs can donate parts of their LTRs during generation of recombinant viruses in mice (66, 147). A well-known example is the case of Bxv1 (Xmv43), parts of whose U3 region contribute to recombinant MCF viruses that induce lymphoma in AKR and HRS mice (80, 147, 155). Interestingly, both PreXMRV-1 and PreXMRV-2 have xenotropic LTRs, but it was PreXMRV-1 that contributed its LTR to the final XMRV recombinant, perhaps providing a selective advantage for replication in human cells. The XMRV LTR, which is identical to the PreXMRV-1 LTR except for a single nucleotide, behaves quite differently from that of the exogenous Moloney MLV in terms of transcriptional activity and hormone-responsiveness (35, 124), although it is not known how it compares to LTRs from endogenous MLVs. Reporter assays comparing the different LTR types and an in-depth analysis of hormone response elements within the LTRs will give us a better understanding of the transcriptional properties of the different endogenous MLVs.

Susceptibility of PreXMRV-1 and PreXMRV-2 to Restriction Factors

An interesting property of XMRV relates to its restriction by cellular proteins. XMRV CA resembles B-tropic MLV based on the critical CA residue at position110, however its replication is restricted by both $Fv1^N$ and $Fv1^B$ alleles, which is quite unusual for an MLV and, to our knowledge, has not been observed for any other retrovirus (56). This region of XMRV was donated by PreXMRV-2. Therefore, we would expect PreXMRV-2 to have the same Fv1 susceptibility as XMRV. PreXMRV-1 *gag* is defective due to a 16-bp deletion. The reconstructed PreXMRV-1 CA protein is also reminiscent of B-MLV, with a glutamic acid residue at position 110.

XMRV was not restricted by human TRIM5α (56, 102), so we would not expect PreXMRV-2 to be restricted by this host factor either. However, establishing the actual

restriction properties of a reconstructed PreXMRV-1 CA will require further experiments.

XMRV replication is also blocked by many human restriction factors, lending further support to the notion that XMRV is unlikely to overcome the many host blocks present *in vivo*. Both human APOBEC3G (A3G) and APOBEC3F (A3F) proteins have been shown to impair XMRV replication, with evidence for virion incorporation and G to A hypermutation in vitro (16, 56, 114) and in vivo (Kearney et al. In prep). XMRV is also blocked by human tetherin (56), known to be expressed in human PBMCs. It will be very interesting to explore the restriction properties of PreXMRV-1 and PreXMRV-2, and the different experimental recombinants that may arise *in vitro* between the two.

Detection of MLVs as possible human pathogens

The initial reports of association between endogenous MLV-related viruses and human disease were attractive because of their biological plausibility: related viruses cause a similar variety of diseases in mouse models (28), close contact between mice and humans can result in zoonotic infection, and the virus isolated is highly infectious for at least some human tumor cell lines (14, 124). As the studies presented here unfolded, however, a number of experimental issues regarding the possibility of detection of endogenous proviral sequences and their confusion with replicating viruses infecting human patients came to light. There are a number of lessons that should be heeded in the development of future studies.

First, low levels of mouse DNA contamination are very commonly found in laboratory reagents. In some cases, this DNA can be attributed to the inclusion of mouse-derived products, such as mouse monoclonal antibodies in PCR reagents or used for

isolation of cells (85, 110, 123, 135, 162); in others, the source of mouse DNA is less than clear, but given the close relationship of human and wild mouse activity, it is not hard to imagine that mice can leave traces in many places that can find their way into almost any laboratory reagent or supply.

A second, related, issue regards the provision of appropriate controls. Given the apparent sporadic nature of this sort of contamination, it is absolutely essential that controls and patient samples be exactly matched, not only for personal characteristics, such as age, gender, geography, etc., but also in the reagents and materials (tubes, needles, etc.) used to obtain the assay samples. Unfortunately, such caution is often not the case, particularly in retrospective studies like Lo et al. (92). Clinical samples are often collected as at least two separate groups, the simplest example being patient vs. control samples. An interesting phenomenon is the detection of a contaminant in one set of samples and not the other, resulting in false association of virus with human disease. The discovery of a laboratory contaminant in some assays would still not provide a full explanation for the detection of samples in one group but not the other. Differential association between two different groups of clinical specimens can occur, even if the detected entity is an artifact. In fact, in the study by Oakes et al. (110), MLV sequences were detected preferentially in healthy controls drawn at a later time and processed by a slightly different method than those from CFS cases. One possible explanation for such erroneous associations is that the two groups might have been handled differently, collected at different times by different people at different locations using different reagents, supplies, or methods. Furthermore, blinding of investigators to which samples are cases and which are controls for such studies is crucial. Examples of such false

associations have persisted, even when independent laboratories had confirmed findings after the initial report (173).

Third, even in the absence of contaminating mouse DNA, a different complication arises from retrovirus contaminated cell lines used in the laboratory. There are multiple documented cases of such contamination events, which appear to be quite common among laboratories working with retroviruses (71, 84, 118, 122, 152). Even in laboratories that do not work with retroviruses, there are examples of cell lines producing replication competent retroviruses. Contamination of cell lines with retroviruses could occur through cross contamination of previously uninfected cells with a virus grown or handled on other cells nearby. As long as the virus is replication competent and can establish an efficient infection, it could eventually take over the entire culture even with trace amounts of starting virus. Other potential sources of retrovirus infection of cell lines can be mislabeling of frozen infected cells that are later thawed by someone else, or through liquid nitrogen freezers if they created a pool of viruses over the years that could somehow be transferred from one tube to the next.

Finally, it is also important to point out that most published "XMRV-specific" PCR assays, in fact detect sequences that are identical between either PreXMRV-1 or 2 and XMRV, and would yield a false positive result if confronted with mouse DNA containing one or the other of these proviruses. Only PCR assays that rely on primers flanking one of the crossover points can be considered to be truly XMRV-specific (Fig. 27) (115). In fact, of all the PCR assays published to date, almost none of them are truly specific, since the primer sets would detect either PreXMRV-1 or PreXMRV-2 (8, 37, 53, 92, 94, 133, 149, 163). To our knowledge, the only exception is the primer set used by

Hohn et al., where the first round of nested PCR spans a crossover site between PreXMRV-1 and PreXMRV-2 in the gag leader region (67). In the mentioned study, XMRV was not detected in a large number of prostate cancer patients from Germany (67).

The Current Status of XMRV Research

Through the work of many laboratories around the world, the connection of XMRV with prostate cancer, chronic fatigue syndrome, or any other human disease was eventually ruled out. Our work has shed light on the origin of XMRV; we showed that it was inadvertently created in the laboratory. These findings prompted many researchers to go back to their initial studies and check the constructs, samples, and assays, to find an explanation as to why the experiments may have yielded a positive signal.

In one such study, Dr. Robert Silverman's group, who were coauthors on the Lombardi et al. paper - the initial publication linking XMRV to CFS (94), retested the original samples they had received from the WPI for XMRV DNA sequences by PCR. In agreement with their initial findings, they detected XMRV-specific sequences in some of the patient samples. At that stage, it had become clear that XMRV was not a human retrovirus, so they explored the source of these sequences further, in order to understand where the false positive signal was coming from. PCR amplification with primers for the neomycin gene revealed sequences identical to the XMRV plasmid (VP62 clone) commonly used in experiments as a positive control, in the exact same samples that tested positive for XMRV-specific sequences (136). Following these findings, the data contributed by Dr. Silverman's group, as well as their names, were retracted from the original Lombardi et al. paper (94, 136).

The partial retraction was followed by a full editorial retraction of the entire paper by Science editor-in-chief Bruce Alberts (3). The full retraction, in turn, was followed by a retraction of the Lo et al. paper, the study linking the presence of MLV-like sequences in the blood of CFS patients (93). Looking back on the short history of XMRV research: First, a large number of studies failing to find a connection of XMRV with human disease were published (Table 1). Then came the inability to reliably detect the virus in blinded studies, even by laboratories that originally described the connection (138). Finally, the two main papers that had made the connection of XMRV/MLV with human disease were retracted (3, 93). These advances inevitably lead to the conclusion that XMRV is not a human pathogen.

Going back to the previous studies, one can try to predict - and in some cases prove - what might have caused the detection of false positive signals. Potential explanations for the detection of false positives in disease association studies are shown in Table 5.

Reference	Possible Reason for Positive Signal
Urisman et al. PLoS Path. 2006	PCR contamination
Fischer et al. J.Clin.Virol 2008	N/A
Schlaberg et al. PNAS 2009	Unspecific antibody assay
Hohn et al. Retrovirology 2009	N/A
Arnold et al. Urology 2010	Unspecific antibody assay
Aloia et al. Cancer Res. 2010	N/A
Martinez-Fierro et al. BMC Cancer 2010	N/A
Sakuma et al. Retrovirology 2011	PCR contamination
Furuta et al. Retrovirology 2011	Unspecific antibody assay
Hohn et al. PLoS One 2010	N/A
Kearney et al. Adv Virol 2011	N/A
Verhaegh et al. Prostate 2011	PCR contamination
Switzer et al. PLoS One 2011	Unspecific antibody assay
Akgul et al. Med Microbiol Immunol. 2011	N/A
Lombardi et al. Science 2009	RETRACTED (3)
Erlwein et al. PLoS One 2010	N/A
Lo et al. PNAS 2010	RETRACTED (93)
Groom et al. Retrovirology 2010	N/A
van Kuppeveld et al. BMJ 2010	N/A
Oakes et al. Retrovirology 2010	PCR contamination
Robinson et al. Retrovirology 2010	PCR contamination
Henrich et al. J Infect Dis 2010	N/A
Switzer et al. Retrovirology 2010	N/A
Satterfield et al. Retrovirology 2011	N/A
Shin et al. J Virol 2011	N/A
Knox et al. Science 2011	N/A
Simmons et al. Science 2011	N/A
Cool et al. Virology 2011	N/A
McCormick et al. Neurology 2008	N/A
Cornelissen et al. PLoS One 2010	N/A
Barnes et al. J Infect Dis 2010	N/A
Satterfield et al. Mol Autism 2010	N/A
Jeziorski et al. Retrovirology 2010	N/A
Fischer et al. Emerg Infect Dis 2010	PCR contamination
Delviks-Frankenberry et al. Adv Virol 2011	N/A
Spindler et al. Adv Virol 2011	N/A
Waugh et al. Cancer Epid Biomarkers 201	N/A

Table 5. Potential explanation for the detection of false positives

The details of previous studies investigating the association of XMRV with human disease were given in Table 1. Predictions as to why a positive signal may have been detected are shown here. Numbers in parentheses represent the reference number. N/A: not applicable - where no significant positive signal was detected.

PreXMRV-1 and PreXMRV-2 represent yet another addition to the growing list of endogenous MLV proviruses of mice. Their significance stems from the recombination between the two that ultimately led to the generation of XMRV during the passaging of a human prostate tumor as xenografts in mice. To our knowledge, no other MLV, recombinant or otherwise, acquired by human tissues upon transplantation into mice has ever received this much attention, even though such retroviral acquisitions by heterologous cells occurs frequently during passage of human cells in mice, and has been known to occur for more than 30 years (2, 11, 50, 160, 179). Even more frequent is when cell lines become unintentionally infected with retroviruses, which often go unnoticed for many years (6, 71, 84, 118, 121, 143, 152).

The overall lesson to be learned here is that extreme measures are required to avoid false associations of MLVs with disease, including: 1) Rigorous use of highly sensitive assays for detecting mouse DNA contamination. 2) Frequent testing of all cell cultures used in the laboratory for undetected infection with an MLV or another virus; 3) Use of controls that are exactly contemporaneous to the cases, and obtained by precisely the same methods using the same materials and reagents. As a few recent papers indicate (85, 135), these conditions are not easy to achieve, but only laboratories that do so can make credible claims to the discovery of new human infections.

REFERENCES

1. **Aaronson, S. A., J. W. Hartley, and G. J. Todaro.** 1969. Mouse Leukemia Virus: Spontaneous Release by Mouse Embryo Cells after Long-term in vitro Cultivation. Proc Natl Acad Sci USA **64:**87-94.
2. **Achong, B. G., P. A. Trumper, and B. C. Giovanella.** 1976. C-type virus particles in human tumours transplanted into nude mice. Br J Cancer **34:**203-206.
3. **Alberts, B.** 2011. Retraction. Science **334:**1636.
4. **Albritton, L. M., L. Tseng, D. Scadden, and J. M. Cunningham.** 1989. A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. Cell **57:**659-66.
5. **Aloia, A. L., K. S. Sfanos, W. B. Isaacs, Q. Zheng, F. Maldarelli, A. M. De Marzo, and A. Rein.** 2010. XMRV: A New Virus in Prostate Cancer? Cancer Research **70:**10028-10033.
6. **Antoine, M., B. Wegmann, and P. Kiefer.** 1998. Envelope and Long Terminal Repeat Sequences of an Infectious Murine Leukemia Virus from a Human SCLC Cell Line: Implications for Gene Transfer. Virus Genes **17:**157-168.
7. **Arhel, N.** 2010. Revisiting HIV-1 uncoating. Retrovirology **7:**96.
8. **Arnold, R. S., N. V. Makarova, A. O. Osunkoya, S. Suppiah, T. A. Scott, N. A. Johnson, S. M. Bhosle, D. Liotta, E. Hunter, F. F. Marshall, H. Ly, R. J. Molinaro, J. L. Blackwell, and J. A. Petros.** 2010. XMRV Infection in Patients With Prostate Cancer: Novel Serologic Assay and Correlation With PCR and FISH. Urology **75:**755-761.
9. **Baliji, S., Q. Liu, and C. A. Kozak.** 2011. Common Inbred Strains of the Laboratory Mouse That Are Susceptible to Infection by Mouse Xenotropic Gammaretroviruses and the Human-Derived Retrovirus XMRV. J. Virol. **84:**12841-12849.
10. **Battini, J.-L., J. E. J. Rasko, and A. D. Miller.** 1999. A human cell-surface receptor for xenotropic and polytropic murine leukemia viruses: Possible role in G protein-coupled signal transduction. Proc Natl Acad Sci USA **96:**1385-1390.
11. **Beattie, G. M., A. F. Knowles, F. C. Jensen, S. M. Baird, and N. O. Kaplan.** 1982. Induction of sarcomas in athymic mice. Proc Natl Acad Sci USA **79:**3033-3036.
12. **Besnier, C., Y. Takeuchi, and G. Towers.** 2002. Restriction of lentivirus in monkeys. Proceedings of the National Academy of Sciences **99:**11920-11925.
13. **Best, S., P. L. Tissier, G. Towers, and J. P. Stoye.** 1996. Positional cloning of the mouse retrovirus restriction gene Fvl. Nature **382:**826-829.
14. **Bhosle, S., S. Suppiah, R. Molinaro, Y. Liang, R. Arnold, W. Diehl, N. Makarova, J. Blackwell, J. Petros, D. Liotta, E. Hunter, and H. Ly.** 2010. Evaluation of Cellular Determinants Required for In Vitro Xenotropic Murine Leukemia Virus-Related Virus Entry into Human Prostate Cancer and Noncancerous Cells. J. Virol. **84:**6288-6296.
15. **Bishop, K. N., M. Bock, G. Towers, and J. P. Stoye.** 2001. Identification of the Regions of Fv1 Necessary for Murine Leukemia Virus Restriction. J. Virol. **75:**5182-5188.

16. **Bogerd, H. P., F. Zhang, P. D. Bieniasz, and B. R. Cullen.** 2011. Human APOBEC3 proteins can inhibit xenotropic murine leukemia virus-related virus infectivity. Virology **410:**234-239.
17. **Brass, A. L., D. M. Dykxhoorn, Y. Benita, N. Yan, A. Engelman, R. J. Xavier, J. Lieberman, and S. J. Elledge.** 2008. Identification of Host Proteins Required for HIV Infection Through a Functional Genomic Screen. Science **319:**921-926.
18. **Bushman, F. D.** 2003. Targeting Survival: Integration Site Selection by Retroviruses and LTR-Retrotransposons. Cell **115:**135-138.
19. **Campbell, E. M., O. Perez, J. L. Anderson, and T. J. Hope.** 2008. Visualization of a proteasome-independent intermediate during restriction of HIV-1 by rhesus TRIM5Œ±. The Journal of Cell Biology **180:**549-561.
20. **Carpten, J., N. Nupponen, S. Isaacs, R. Sood, C. Robbins, J. Xu, M. Faruque, T. Moses, C. Ewing, E. Gillanders, P. Hu, P. Bujnovszky, I. Makalowska, A. Baffoe-Bonnie, D. Faith, J. Smith, D. Stephan, K. Wiley, M. Brownstein, D. Gildea, B. Kelly, R. Jenkins, G. Hostetter, M. Matikainen, J. Schleutker, K. Klinger, T. Connors, Y. Xiang, Z. Wang, A. De Marzo, N. Papadopoulos, O. P. Kallioniemi, R. Burk, D. Meyers, H. Gronberg, P. Meltzer, R. Silverman, J. Bailey-Wilson, P. Walsh, W. Isaacs, and J. Trent.** 2002. Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. Nat Genet **30:**181-184.
21. **Carruthers, B. M., M. I. van de Sande, K. L. De Meirleir, N. G. Klimas, G. Broderick, T. Mitchell, D. Staines, A. C. P. Powles, N. Speight, R. Vallings, L. Bateman, B. Baumgarten-Austrheim, D. S. Bell, N. Carlo-Stella, J. Chia, A. Darragh, D. Jo, D. Lewis, A. R. Light, S. Marshall-Gradisbik, I. Mena, J. A. Mikovits, K. Miwa, M. Murovska, M. L. Pall, and S. Stevens.** 2011. Myalgic encephalomyelitis: International Consensus Criteria. Journal of Internal Medicine **270:**327-338.
22. **Casey, G., P. J. Neville, S. J. Plummer, Y. Xiang, L. M. Krumroy, E. A. Klein, W. J. Catalona, N. Nupponen, J. D. Carpten, J. M. Trent, R. H. Silverman, and J. S. Witte.** 2002. RNASEL Arg462Gln variant is implicated in up to 13% of prostate cancer cases. Nat Genet **32:**581-583.
23. **Chaipan, C., K. A. Dilley, T. Paprotka, K. A. Delviks-Frankenberry, N. J. Venkatachari, W.-S. Hu, and V. K. Pathak.** 2011. Severe Restriction of Xenotropic Murine Leukemia Virus-Related Virus Replication and Spread in Cultured Human Peripheral Blood Mononuclear Cells. J. Virol. **85:**4888-4897.
24. **Chesebro, B., W. Britt, L. Evans, K. Wehrly, J. Nishio, and M. Cloyd.** 1983. Characterization of monoclonal antibodies reactive with murine leukemia viruses: use in analysis of strains of friend MCF and Friend ecotropic murine leukemia virus. Virology **127:**134-148.
25. **Cingöz, O., and J. Coffin.** 2011. Endogenous Murine Leukemia Viruses: Relationship to XMRV and Related Sequences Detected in Human DNA Samples Advances in Virology **2011:**940210.
26. **Cingöz, O., T. Paprotka, K. A. Delviks-Frankenberry, S. Wildt, W.-S. Hu, V. K. Pathak, and J. M. Coffin.** 2011. Characterization, mapping and distribution of the two XMRV parental proviruses. J. Virol. **In press.**

27. **Clavel, F., M. D. Hoggan, R. L. Willey, K. Strebel, M. A. Martin, and R. Repaske.** 1989. Genetic recombination of human immunodeficiency virus. J. Virol. **63:**1455-1459.
28. **Coffin, J. M., S. H. Hughes, and H. Varmus.** 1997. Retroviruses. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
29. **Colicelli, J., and S. P. Goff.** 1986. Isolation of a recombinant murine leukemia virus utilizing a new primer tRNA. J. Virol. **57:**37-45.
30. **Cornelissen, M., F. Zorgdrager, P. Blom, S. Jurriaans, S. Repping, E. van Leeuwen, M. Bakker, B. Berkhout, and A. C. van der Kuyl.** 2010. Lack of Detection of XMRV in Seminal Plasma from HIV-1 Infected Men in The Netherlands. PLoS ONE **5:**e12040.
31. **Courgnaud, V., J.-L. Battini, M. Sitbon, and A. L. Mason.** 2010. Mouse retroviruses and chronic fatigue syndrome: Does X (or P) mark the spot? Proc Natl Acad Sci USA **107:**15666-15667.
32. **Cowan, S., T. Hatziioannou, T. Cunningham, M. A. Muesing, H. G. Gottlinger, and P. D. Bieniasz.** 2002. Cellular inhibitors with Fv1-like activity restrict human and simian immunodeficiency virus tropism. Proceedings of the National Academy of Sciences **99:**11914-11919.
33. **DesGroseillers, L., and P. Jolicoeur.** 1983. Physical mapping of the Fv-1 tropism host range determinant of BALB/c murine leukemia viruses. J. Virol. **48:**685-696.
34. **Dong, B., S. Kim, S. Hong, J. Das Gupta, K. Malathi, E. A. Klein, D. Ganem, J. L. DeRisi, S. A. Chow, and R. H. Silverman.** 2007. An infectious retrovirus susceptible to an IFN antiviral pathway from human prostate tumors. Proc Natl Acad Sci USA **104:**1655-1660.
35. **Dong, B., and R. H. Silverman.** 2010. Androgen Stimulates Transcription and Replication of Xenotropic Murine Leukemia Virus-Related Virus. J. Virol. **84:**1648-1651.
36. **Dunn, G. P., K. C. F. Sheehan, L. J. Old, and R. D. Schreiber.** 2005. IFN Unresponsiveness in LNCaP Cells Due to the Lack of JAK1 Gene Expression. Cancer Research **65:**3447-3453.
37. **Erlwein, O., S. Kaye, M. O. McClure, J. Weber, G. Wills, D. Collier, S. Wessely, and A. Cleare.** 2010. Failure to Detect the Novel Retrovirus XMRV in Chronic Fatigue Syndrome. PLoS ONE **5:**e8519.
38. **Erlwein, O., M. J. Robinson, S. Dustan, J. Weber, S. Kaye, and M. O. McClure.** 2011. DNA Extraction Columns Contaminated with Murine Sequences. PLoS ONE **6:**e23484.
39. **Evans, L. H., A. S. M. Alamgir, N. Owens, N. Weber, K. Virtaneva, K. Barbian, A. Babar, F. Malik, and K. Rosenke.** 2009. Mobilization of Endogenous Retroviruses in Mice after Infection with an Exogenous Retrovirus. J. Virol. **83:**2429-2435.
40. **Evans, L. H., M. Lavignon, M. Taylor, and A. S. M. Alamgir.** 2003. Antigenic Subclasses of Polytropic Murine Leukemia Virus (MLV) Isolates Reflect Three Distinct Groups of Endogenous Polytropic MLV-Related Sequences in NFS/N Mice. J. Virol. **77:**10327-10338.

41. **Fiebig, U., M. G. Hartmann, N. Bannert, R. Kurth, and J. Denner.** 2006. Transspecies Transmission of the Endogenous Koala Retrovirus. J. Virol. **80:**5651-5654.
42. **Fischer, N., H. Olaf, S. Claudia, K. H. C. Felix, H. Hartwig, A. Martin, and S. Thorsten.** 2008. Prevalence of human gammaretrovirus XMRV in sporadic prostate cancer. Journal of Clinical Virology **43:**277-283.
43. **Frankel, W. N., and J. M. Coffin.** 1994. Endogenous nonecotropic proviruses mapped with oligonucleotide probes from the long terminal repeat region. Mamm Genome **5:**275-81.
44. **Frankel, W. N., J. P. Stoye, B. A. Taylor, and J. M. Coffin.** 1990. A linkage map of endogenous murine leukemia proviruses. Genetics **124:**221-36.
45. **Frankel, W. N., J. P. Stoye, B. A. Taylor, and J. M. Coffin.** 1989. Genetic analysis of endogenous xenotropic murine leukemia viruses: association with two common mouse mutations and the viral restriction locus Fv-1. J. Virol. **63:**1763-74.
46. **Frankel, W. N., J. P. Stoye, B. A. Taylor, and J. M. Coffin.** 1989. Genetic identification of endogenous polytropic proviruses by using recombinant inbred mice. J. Virol. **63:**3810-21.
47. **Fukuda, K., S. E. Straus, I. Hickie, M. C. Sharpe, J. G. Dobbins, A. Komaroff, and G. International Chronic Fatigue Syndrome Study.** 1994. The Chronic Fatigue Syndrome: A Comprehensive Approach to Its Definition and Study. Annals of Internal Medicine **121:**953-959.
48. **Gabuzda, D. H., K. Lawrence, E. Langhoff, E. Terwilliger, T. Dorfman, W. A. Haseltine, and J. Sodroski.** 1992. Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. J. Virol. **66:**6489-6495.
49. **Garson, J., P. Kellam, and G. Towers.** 2011. Analysis of XMRV integration sites from human prostate cancer tissues suggests PCR contamination rather than genuine human infection. Retrovirology **8:**13.
50. **Gautsch, J. W., A. F. Knowles, F. C. Jensen, and N. O. Kaplan.** 1980. Highly efficient induction of type C retroviruses by a human tumor in athymic mice. Proc Natl Acad Sci USA **77:**2247-2250.
51. **Gershwin, M. E., B. Merchant, M. C. Gelfand, J. Vickers, A. D. Steinberg, and C. T. Hansen.** 1975. The natural history and immunopathology of outbred athymic (nude) mice. Clinical Immunology and Immunopathology **4:**324-340.
52. **Göttlinger, H. G., T. Dorfman, E. A. Cohen, and W. A. Haseltine.** 1993. Vpu protein of human immunodeficiency virus type 1 enhances the release of capsids produced by gag gene constructs of widely divergent retroviruses. Proceedings of the National Academy of Sciences **90:**7381-7385.
53. **Gray, E. R., J. A. Garson, J. Breuer, S. Edwards, P. Kellam, D. Pillay, and G. J. Towers.** 2011. No Evidence of XMRV or Related Retroviruses in a London HIV-1-Positive Patient Cohort. PLoS ONE **6:**e18096.
54. **Gregory, C. W., R. T. Johnson, J. L. Mohler, F. S. French, and E. M. Wilson.** 2001. Androgen Receptor Stabilization in Recurrent Prostate Cancer Is Associated with Hypersensitivity to Low Androgen. Cancer Research **61:**2892-2898.

55. **Groom, H., V. Boucherit, K. Makinson, E. Randal, S. Baptista, S. Hagan, J. Gow, F. Mattes, J. Breuer, J. Kerr, J. Stoye, and K. Bishop.** 2010. Absence of xenotropic murine leukaemia virus-related virus in UK patients with chronic fatigue syndrome. Retrovirology **7:**10.
56. **Groom, H. C. T., M. W. Yap, R. P. Galão, S. J. D. Neil, and K. N. Bishop.** 2010. Susceptibility of xenotropic murine leukemia virus-related virus (XMRV) to retroviral restriction factors. Proc Natl Acad Sci USA **107:**5166-5171.
57. **Guénet, J.-L., and F. Bonhomme.** 2003. Wild mice: an ever-increasing contribution to a popular mammalian model. Trends in Genetics **19:**24-31.
58. **Hare, S., S. S. Gupta, E. Valkov, A. Engelman, and P. Cherepanov.** 2010. Retroviral intasome assembly and inhibition of DNA strand transfer. Nature **464:**232-236.
59. **Harris, R. S., and M. T. Liddament.** 2004. Retroviral restriction by APOBEC proteins. Nat Rev Immunol **4:**868-877.
60. **Hartley, J. W., W. P. Rowe, and R. J. Huebner.** 1970. Host-Range Restrictions of Murine Leukemia Viruses in Mouse Embryo Cell Cultures. J. Virol. **5:**221-225.
61. **Hartley, J. W., N. K. Wolford, L. J. Old, and W. P. Rowe.** 1977. A new class of murine leukemia virus associated with development of spontaneous lymphomas. Proc Natl Acad Sci USA **74:**789-92.
62. **Hatziioannou, T., D. Perez-Caballero, A. Yang, S. Cowan, and P. D. Bieniasz.** 2004. Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5Œ±. Proceedings of the National Academy of Sciences of the United States of America **101:**10774-10779.
63. **Hein, S., V. Prassolov, Y. Zhang, D. Ivanov, J. Lohler, S. R. Ross, and C. Stocking.** 2003. Sodium-Dependent myo-Inositol Transporter 1 Is a Cellular Receptor for Mus cervicolor M813 Murine Leukemia Virus. J. Virol. **77:**5926-5932.
64. **Henzy, J.** 2011. Retroviral Envelope Glycoproteins: Structure and Evolution. Ph.D. Dissertation.
65. **Hofmann, W., D. Schubert, J. LaBonte, L. Munson, S. Gibson, J. Scammell, P. Ferrigno, and J. Sodroski.** 1999. Species-Specific, Postentry Barriers to Primate Immunodeficiency Virus Infection. J. Virol. **73:**10020-10028.
66. **Hoggan, M. D., C. E. Buckler, J. F. Sears, W. P. Rowe, and M. A. Martin.** 1983. Organization and stability of endogenous xenotropic murine leukemia virus proviral DNA in mouse genomes. J. Virol. **45:**473-7.
67. **Hohn, O., H. Krause, P. Barbarotto, L. Niederstadt, N. Beimforde, J. Denner, K. Miller, R. Kurth, and N. Bannert.** 2009. Lack of evidence for xenotropic murine leukemia virus-related virus (XMRV) in German prostate cancer patients. Retrovirology **6:**92.
68. **Hohn, O., K. Strohschein, A. U. Brandt, S. Seeher, S. Klein, R. Kurth, F. Paul, C. Meisel, C. Scheibenbogen, and N. Bannert.** 2010. No Evidence for XMRV in German CFS and MS Patients with Fatigue Despite the Ability of the Virus to Infect Human Blood Cells *In Vitro*. PLoS ONE **5:**e15632.
69. **Holman, A. G., and J. M. Coffin.** 2005. Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus

integration sites. Proceedings of the National Academy of Sciences **102:**6103-6107.

70. **Hu, W. S., and H. M. Temin.** 1990. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proceedings of the National Academy of Sciences **87:**1556-1560.
71. **Hue, S., E. Gray, A. Gall, A. Katzourakis, C. Tan, C. Houldcroft, S. McLaren, D. Pillay, A. Futreal, J. Garson, O. Pybus, P. Kellam, and G. Towers.** 2010. Disease-associated XMRV sequences are consistent with laboratory contamination. Retrovirology **7:**111.
72. **Ideraabdullah, F. Y., E. de la Casa-Esperón, T. A. Bell, D. A. Detwiler, T. Magnuson, C. Sapienza, and F. P.-M. de Villena.** 2004. Genetic and Haplotype Diversity Among Wild-Derived Mouse Inbred Strains. Genome Research **14:**1880-1887.
73. **Jern, P., G. Sperber, and J. Blomberg.** 2005. Use of Endogenous Retroviral Sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. Retrovirology **2:**50.
74. **Jern, P., J. P. Stoye, and J. M. Coffin.** 2007. Role of APOBEC3 in Genetic Diversity among Endogenous Murine Leukemia Viruses. PLoS Genet **3:**e183.
75. **Jeziorski, E., V. Foulongne, C. Ludwig, D. Louhaem, G. Chiocchia, M. Segondy, M. Rodiere, M. Sitbon, and V. Courgnaud.** 2010. No evidence for XMRV association in pediatric idiopathic diseases in France. Retrovirology **7:**63.
76. **Jolicoeur, P., and D. Baltimore.** 1976. Effect of Fv-1 gene product on proviral DNA formation and integration in cells infected with murine leukemia viruses. Proceedings of the National Academy of Sciences **73:**2236-2240.
77. **Jouvenet, N., S. J. D. Neil, M. Zhadina, T. Zang, Z. Kratovac, Y. Lee, M. McNatt, T. Hatziioannou, and P. D. Bieniasz.** 2009. Broad-Spectrum Inhibition of Retroviral and Filoviral Particle Release by Tetherin. J. Virol. **83:**1837-1844.
78. **Jung, Y. T., T. Wu, and C. A. Kozak.** 2003. Characterization of Recombinant Nonecotropic Murine Leukemia Viruses from the Wild Mouse Species Mus spretus. J. Virol. **77:**12773-12781.
79. **Keane, T. M., L. Goodstadt, P. Danecek, M. A. White, K. Wong, B. Yalcin, A. Heger, A. Agam, G. Slater, M. Goodson, N. A. Furlotte, E. Eskin, C. Nellaker, H. Whitley, J. Cleak, D. Janowitz, P. Hernandez-Pliego, A. Edwards, T. G. Belgard, P. L. Oliver, R. E. McIntyre, A. Bhomra, J. Nicod, X. Gan, W. Yuan, L. van der Weyden, C. A. Steward, S. Bala, J. Stalker, R. Mott, R. Durbin, I. J. Jackson, A. Czechanski, J. A. Guerra-Assuncao, L. R. Donahue, L. G. Reinholdt, B. A. Payseur, C. P. Ponting, E. Birney, J. Flint, and D. J. Adams.** 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature **477:**289-294.
80. **Khan, A. S.** 1984. Nucleotide sequence analysis establishes the role of endogenous murine leukemia virus DNA segments in formation of recombinant mink cell focus-forming murine leukemia viruses. J. Virol. **50:**864-71.
81. **Khan, A. S., and M. A. Martin.** 1983. Endogenous murine leukemia proviral long terminal repeats contain a unique 190-base-pair insert. Proc Natl Acad Sci USA **80:**2699-703.

82. **Kim, S., N. Kim, B. Dong, D. Boren, S. A. Lee, J. Das Gupta, C. Gaughan, E. A. Klein, C. Lee, R. H. Silverman, and S. A. Chow.** 2008. Integration Site Preference of Xenotropic Murine Leukemia Virus-Related Virus, a New Human Retrovirus Associated with Prostate Cancer. J. Virol. **82:**9964-9977.
83. **Kim, S., A. Rusmevichientong, B. Dong, R. Remenyi, R. H. Silverman, and S. A. Chow.** 2010. Fidelity of Target Site Duplication and Sequence Preference during Integration of Xenotropic Murine Leukemia Virus-Related Virus. PLoS ONE **5:**e10255.
84. **Knouf, E. C., M. J. Metzger, P. S. Mitchell, J. D. Arroyo, J. R. Chevillet, M. Tewari, and A. D. Miller.** 2009. Multiple Integrated Copies and High-Level Production of the Human Retrovirus XMRV (Xenotropic Murine Leukemia Virus-Related Virus) from 22Rv1 Prostate Carcinoma Cells. J. Virol. **83:**7353-7356.
85. **Knox, K., D. Carrigan, G. Simmons, F. Teque, Y. Zhou, J. Hackett, X. Qiu, K.-C. Luk, G. Schochetman, A. Knox, A. M. Kogelnik, and J. A. Levy.** 2011. No Evidence of Murine-Like Gammaretroviruses in CFS Patients Previously Identified as XMRV-Infected. Science **333:**94-97.
86. **Kochera, M., T. W. Depinet, T. P. Pretlow, J. M. Giaconia, N. L. Edgehouse, T. G. Pretlow, and S. Schwartz.** 1999. Molecular cytogenetic studies of a serially transplanted primary prostatic carcinoma xenograft (CWR22) and four relapsed tumors. The Prostate **41:**7-11.
87. **König, R., Y. Zhou, D. Elleder, T. L. Diamond, G. M. C. Bonamy, J. T. Irelan, C.-y. Chiang, B. P. Tu, P. D. De Jesus, C. E. Lilley, S. Seidel, A. M. Opaluch, J. S. Caldwell, M. D. Weitzman, K. L. Kuhen, S. Bandyopadhyay, T. Ideker, A. P. Orth, L. J. Miraglia, F. D. Bushman, J. A. Young, and S. K. Chanda.** 2008. Global Analysis of Host-Pathogen Interactions that Regulate Early-Stage HIV-1 Replication. Cell **135:**49-60.
88. **Kozak, C.** 2010. The mouse "xenotropic" gammaretroviruses and their XPR1 receptor. Retrovirology **7:**101.
89. **Kozak, C. A., J. W. Hartley, and H. C. Morse, 3rd.** 1984. Laboratory and wild-derived mice with multiple loci for production of xenotropic murine leukemia virus. J. Virol. **51:**77-80.
90. **Levy, J. A., and T. Pincus.** 1970. Demonstration of biological activity of a murine leukemia virus of New Zealand black mice. Science **170:**326-7.
91. **Lieber, M. M., C. J. Sherr, G. J. Todaro, R. E. Benveniste, R. Callahan, and H. G. Coon.** 1975. Isolation from the asian mouse Mus caroli of an endogenous type C virus related to infectious primate type C viruses. Proc Natl Acad Sci USA **72:**2315-2319.
92. **Lo, S.-C., N. Pripuzova, B. Li, A. L. Komaroff, G.-C. Hung, R. Wang, and H. J. Alter.** 2010. Detection of MLV-related virus gene sequences in blood of patients with chronic fatigue syndrome and healthy blood donors. Proc Natl Acad Sci USA **107:**15874-15879.
93. **Lo, S.-C., N. Pripuzova, B. Li, A. L. Komaroff, G.-C. Hung, R. Wang, and H. J. Alter.** 2012. Retraction for Lo et al., Detection of MLV-related virus gene sequences in blood of patients with chronic fatigue syndrome and healthy blood donors. Proceedings of the National Academy of Sciences **109:**346.

94. **Lombardi, V. C., F. W. Ruscetti, J. Das Gupta, M. A. Pfost, K. S. Hagen, D. L. Peterson, S. K. Ruscetti, R. K. Bagni, C. Petrow-Sadowski, B. Gold, M. Dean, R. H. Silverman, and J. A. Mikovits.** 2009. Detection of an Infectious Retrovirus, XMRV, in Blood Cells of Patients with Chronic Fatigue Syndrome. Science **326:**585-589.
95. **Madani, N., and D. Kabat.** 1998. An Endogenous Inhibitor of Human Immunodeficiency Virus in Human Lymphocytes Is Overcome by the Viral Vif Protein. J. Virol. **72:**10251-10255.
96. **Maertens, G. N., S. Hare, and P. Cherepanov.** 2010. The mechanism of retroviral integration from X-ray structures of its key intermediates. Nature **468:**326-329.
97. **Malim, M. H., and M. Emerman.** 2008. HIV-1 Accessory Proteins--Ensuring Viral Survival in a Hostile Environment. Cell Host & Microbe **3:**388-398.
98. **Mangeat, B., P. Turelli, G. Caron, M. Friedli, L. Perrin, and D. Trono.** 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature **424:**99-103.
99. **Marin, M., C. S. Tailor, A. Nouri, S. L. Kozak, and D. Kabat.** 1999. Polymorphisms of the Cell Surface Receptor Control Mouse Susceptibilities to Xenotropic and Polytropic Leukemia Viruses. J. Virol. **73:**9362-9368.
100. **McCarthy, E., and J. McDonald.** 2004. Long terminal repeat retrotransposons of Mus musculus. Genome Biology **5:**R14.
101. **Mendoza, R., A. E. Vaughan, and A. D. Miller.** 2011. The left half of XMRV is present in an endogenous retrovirus of NIH/3T3 Swiss mouse cells. J. Virol.
102. **Metzger, M. J., C. J. Holguin, R. Mendoza, and A. D. Miller.** 2010. The Prostate Cancer-Associated Human Retrovirus XMRV Lacks Direct Transforming Activity but Can Induce Low Rates of Transformation in Cultured Cells. J. Virol. **84:**1874-1880.
103. **Metzger, M. J., and A. D. Miller.** 2010. Acutely Transforming Retrovirus Expressing Nras Generated from HT-1080 Fibrosarcoma Cells Infected with the Human Retrovirus XMRV. J. Virol. **84:**7908-7910.
104. **Miller, D. G., and A. D. Miller.** 1994. A family of retroviruses that utilize related phosphate transporters for cell entry. J. Virol. **68:**8270-8276.
105. **Nagabhushan, M., C. M. Miller, T. P. Pretlow, J. M. Giaconia, N. L. Edgehouse, S. Schwartz, H.-J. Kung, R. W. de Vere White, P. H. Gumerlock, M. I. Resnick, S. B. Amini, and T. G. Pretlow.** 1996. CWR22: The First Human Prostate Cancer Xenograft with Strongly Androgen-dependent and Relapsed Strains Both in Vivo and in Soft Agar. Cancer Research **56:**3042-3046.
106. **Neil, S. J. D., S. W. Eastman, N. Jouvenet, and P. D. Bieniasz.** 2006. HIV-1 Vpu Promotes Release and Prevents Endocytosis of Nascent Retrovirus Particles from the Plasma Membrane. PLoS Pathog **2:**e39.
107. **Neil, S. J. D., T. Zang, and P. D. Bieniasz.** 2008. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature **451:**425-430.
108. **Newman, E. N. C., R. K. Holmes, H. M. Craig, K. C. Klein, J. R. Lingappa, M. H. Malim, and A. M. Sheehy.** 2005. Antiviral Function of APOBEC3G Can Be Dissociated from Cytidine Deaminase Activity. Current Biology **15:**166-170.

109. **Nupponen, N. N., and J. D. Carpten.** 2001. Prostate Cancer Susceptibility Genes: Many Studies, Many Results, No Answers. Cancer and Metastasis Reviews **20:**155-164.
110. **Oakes, B., A. Tai, O. Cingöz, M. Henefield, S. Levine, J. Coffin, and B. Huber.** 2010. Contamination of human DNA samples with mouse DNA can lead to false detection of XMRV-like sequences. Retrovirology **7:**109.
111. **Odaka, T., and T. Yamamoto.** 1965. Inheritance of susceptibility to Friend mouse leukemia virus. 11. Spleen foci method applied to test the susceptibility of crossbred progeny between a sensitive and a resistant strain. Jpn. J. Exp. Med. **35:**311-314.
112. **Onlamoon, N., J. Das Gupta, P. Sharma, K. Rogers, S. Suppiah, J. Rhea, R. J. Molinaro, C. Gaughan, B. Dong, E. A. Klein, X. Qiu, S. Devare, G. Schochetman, J. Hackett, Jr., R. H. Silverman, and F. Villinger.** 2011. Infection, Viral Dissemination, and Antibody Responses of Rhesus Macaques Exposed to the Human Gammaretrovirus XMRV. J. Virol. **85:**4547-4557.
113. **Ou, C. Y., L. R. Boone, C. K. Koh, R. W. Tennant, and W. K. Yang.** 1983. Nucleotide sequences of gag-pol regions that determine the Fv-1 host range property of BALB/c N-tropic and B-tropic murine leukemia viruses. J. Virol. **48:**779-784.
114. **Paprotka, T., N. J. Venkatachari, C. Chaipan, R. Burdick, K. A. Delviks-Frankenberry, W.-S. Hu, and V. K. Pathak.** 2010. Inhibition of Xenotropic Murine Leukemia Virus-Related Virus by APOBEC3 Proteins and Antiviral Drugs. J. Virol. **84:**5719-5729.
115. **Paprotka*, T., K. A. Delviks-Frankenberry*, O. Cingöz*, A. Martinez, H.-J. Kung, C. G. Tepper, W.-S. Hu, M. J. Fivash, J. M. Coffin, and V. K. Pathak.** 2011. Recombinant Origin of the Retrovirus XMRV. Science **333:**97-101.
116. **Pathak, V. K., and H. M. Temin.** 1990. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. Proc Natl Acad Sci USA **87:**6019-6023.
117. **Perez-Caballero, D., T. Hatziioannou, A. Yang, S. Cowan, and P. D. Bieniasz.** 2005. Human Tripartite Motif 5{alpha} Domains Responsible for Retrovirus Restriction Activity and Specificity. J. Virol. **79:**8969-8978.
118. **Pizzato, M.** 2010. MLV glycosylated-Gag is an infectivity factor that rescues Nef-deficient HIV-1. Proc Natl Acad Sci USA **107:**9364-9369.
119. **Pretlow, T. G., S. R. Wolman, M. A. Micale, R. J. Pelley, E. D. Kursh, M. I. Resnick, D. R. Bodner, J. W. Jacobberger, C. M. Delmoro, J. M. Giaconia, and T. P. Pretlow.** 1993. Xenografts of Primary Human Prostatic Carcinoma. Journal of the National Cancer Institute **85:**394-398.
120. **Qin, C., Z. Wang, J. Shang, K. Bekkari, R. Liu, S. Pacchione, K. A. McNulty, A. Ng, J. E. Barnum, and R. D. Storer.** 2010. Intracisternal A particle genes: Distribution in the mouse genome, active subtypes, and potential roles as species-specific mediators of susceptibility to cancer. Molecular Carcinogenesis **49:**54-67.
121. **Raisch, K. P., V. M. Kushnaryov, S. E. Grossberg, and L. W. Cashdollar.** 1998. Constitutive Production of a Murine Retrovirus in the Human B-Lymphoblastoid Cell Line, DG-75. Virology **250:**135-139.

122. **Raisch, K. P., M. Pizzato, H.-Y. Sun, Y. Takeuchi, L. W. Cashdollar, and S. E. Grossberg.** 2003. Molecular cloning, complete sequence, and biological characterization of a xenotropic murine leukemia virus constitutively released from the human B-lymphoblastoid cell line DG-75. Virology **308:**83-91.
123. **Robinson, M., O. Erlwein, S. Kaye, J. Weber, O. Cingoz, A. Patel, M. Walker, W.-J. Kim, M. Uiprasertkul, J. Coffin, and M. McClure.** 2010. Mouse DNA contamination in human tissue tested for XMRV. Retrovirology **7:**108.
124. **Rodriguez, J. J., and S. P. Goff.** 2010. Xenotropic Murine Leukemia Virus-Related Virus Establishes an Efficient Spreading Infection and Exhibits Enhanced Transcriptional Activity in Prostate Carcinoma Cells. J. Virol. **84:**2556-2562.
125. **Rose, P. P., and B. T. Korber.** 2000. Detecting hypermutations in viral sequences with an emphasis on G ,Üí A hypermutation. Bioinformatics **16:**400-401.
126. **Roth, S. L., N. Malani, and F. D. Bushman.** 2011. Gammaretroviral Integration into Nucleosomal Target DNA In Vivo. J. Virol. **85:**7393-7401.
127. **Rous, P.** 1911. A sarcoma of the fowl transmissible by an agent separable from the tumor cells. The Journal of Experimental Medicine **13:**397-411.
128. **Sakuma, R., T. Sakuma, S. Ohmine, R. H. Silverman, and Y. Ikeda.** 2009. Xenotropic murine leukemia virus-related virus is susceptible to AZT. Virology **In Press, Corrected Proof**.
129. **Sakuma, T., S. Hue, K. Squillace, J. Tonne, P. Blackburn, S. Ohmine, T. Thatava, G. Towers, and Y. Ikeda.** 2011. No Evidence of XMRV in Prostate Cancer Cohorts in the Midwestern United States. Retrovirology **8:**23.
130. **Sakuma, T., J. M. Tonne, K. A. Squillace, S. Ohmine, T. Thatava, K.-W. Peng, M. A. Barry, and Y. Ikeda.** 2011. Early Events in Retrovirus XMRV Infection of the Wild-Derived Mouse Mus pahari. J. Virol. **85:**1205-1213.
131. **Sato, E., R. Furuta, and T. Miyazawa.** 2010. An endogenous murine leukemia viral genome contaminant in a commercial RT-PCR Kit is amplified using standard primers for XMRV. Retrovirology **7:**110.
132. **Satterfield, B., R. Garcia, H. Jia, S. Tang, H. Zheng, and W. Switzer.** 2011. Serologic and PCR testing of persons with chronic fatigue syndrome in the United States shows no association with xenotropic or polytropic murine leukemia virus-related viruses. Retrovirology **8:**12.
133. **Schlaberg, R., D. J. Choe, K. R. Brown, H. M. Thaker, and I. R. Singh.** 2009. XMRV is present in malignant prostatic epithelium and is associated with prostate cancer, especially high-grade tumors. Proc Natl Acad Sci USA **106:**16351-16356.
134. **Sheehy, A. M., N. C. Gaddis, J. D. Choi, and M. H. Malim.** 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature **418:**646-650.
135. **Shin, C. H., L. Bateman, R. Schlaberg, A. M. Bunker, C. J. Leonard, R. W. Hughen, A. R. Light, K. C. Light, and I. R. Singh.** 2011. Absence of XMRV Retrovirus and Other Murine Leukemia Virus-Related Viruses in Patients with Chronic Fatigue Syndrome. J. Virol. **85:**7195-7202.
136. **Silverman, R. H., J. Das Gupta, V. C. Lombardi, F. W. Ruscetti, M. A. Pfost, K. S. Hagen, D. L. Peterson, S. K. Ruscetti, R. K. Bagni, C. Petrow-Sadowski,**

B. Gold, M. Dean, and J. A. Mikovits. 2011. Partial Retraction. Science **334:**176.

137. **Simmons, G., S. A. Glynn, J. A. Holmberg, J. M. Coffin, I. K. Hewlett, S.-C. Lo, J. A. Mikovits, W. M. Switzer, J. M. Linnen, M. P. Busch, and X. S. R. W. G. for the Blood.** 2011. The Blood Xenotropic Murine Leukemia Virus–Related Virus Scientific Research Working Group: mission, progress, and plans. Transfusion **51:**643-653.
138. **Simmons, G., S. A. Glynn, A. L. Komaroff, J. A. Mikovits, L. H. Tobler, J. Hackett, N. Tang, W. M. Switzer, W. Heneine, I. K. Hewlett, J. Zhao, S.-C. Lo, H. J. Alter, J. M. Linnen, K. Gao, J. M. Coffin, M. F. Kearney, F. W. Ruscetti, M. A. Pfost, J. Bethel, S. Kleinman, J. A. Holmberg, M. P. Busch, and X. S. R. W. G. for the Blood.** 2011. Failure to Confirm XMRV/MLVs in the Blood of Patients with Chronic Fatigue Syndrome: A Multi-Laboratory Study. Science **334:**814-817.
139. **Simon, J. H. M., N. C. Gaddis, R. A. M. Fouchier, and M. H. Malim.** 1998. Evidence for a newly discovered cellular anti-HIV-1 phenotype. Nat Med **4:**1397-1400.
140. **Singh, I. R., J. E. Gorzynski, D. Drobysheva, L. Bassit, and R. F. Schinazi.** 2010. Raltegravir Is a Potent Inhibitor of XMRV, a Virus Implicated in Prostate Cancer and Chronic Fatigue Syndrome. PLoS ONE **5:**e9948.
141. **Smith, J. R., D. Freije, J. D. Carpten, H. Gr√∂nberg, J. Xu, S. D. Isaacs, M. J. Brownstein, G. S. Bova, H. Guo, P. Bujnovszky, D. R. Nusskern, J.-E. Damber, A. Bergh, M. Emanuelsson, O. P. Kallioniemi, J. Walker-Daniels, J. E. Bailey-Wilson, T. H. Beaty, D. A. Meyers, P. C. Walsh, F. S. Collins, J. M. Trent, and W. B. Isaacs.** 1996. Major Susceptibility Locus for Prostate Cancer on Chromosome 1 Suggested by a Genome-Wide Search. Science **274:**1371-1374.
142. **Sramkoski, R., T. Pretlow, J. Giaconia, T. Pretlow, S. Schwartz, M.-S. Sy, S. Marengo, J. Rhim, D. Zhang, and J. Jacobberger.** 1999. A new human prostate carcinoma cell line, 22R*v1*. In Vitro Cellular and Developmental Biology - Animal **35:**403-409.
143. **Stang, A., E. Petrasch-Parwez, S. Brandt, R. Dermietzel, H. Meyer, K. Stuhler, S.-T. Liffers, K. Uberla, and T. Grunwald.** 2009. Unintended spread of a biosafety level 2 recombinant retrovirus. Retrovirology **6:**86.
144. **Stocking, C., and C. Kozak.** 2008. Endogenous retroviruses. Cellular and Molecular Life Sciences **65:**3383-3398.
145. **Stoye, J. P., and J. M. Coffin.** 1988. Polymorphism of murine endogenous proviruses revealed by using virus class-specific oligonucleotide probes. J. Virol. **62:**168-75.
146. **Stoye, J. P., and J. M. Coffin.** 1987. The four classes of endogenous murine leukemia virus: structural relationships and potential for recombination. J. Virol. **61:**2659-69.
147. **Stoye, J. P., C. Moroni, and J. M. Coffin.** 1991. Virological events leading to spontaneous AKR thymomas. J. Virol. **65:**1273-85.

148. **Stremlau, M., C. M. Owens, M. J. Perron, M. Kiessling, P. Autissier, and J. Sodroski.** 2004. The cytoplasmic body component TRIM5[alpha] restricts HIV-1 infection in Old World monkeys. Nature **427:**848-853.
149. **Switzer, W., H. Jia, O. Hohn, H. Zheng, S. Tang, A. Shankar, N. Bannert, G. Simmons, R. M. Hendry, V. Falkenberg, W. Reeves, and W. Heneine.** 2010. Absence of evidence of Xenotropic Murine Leukemia Virus-related virus infection in persons with Chronic Fatigue Syndrome and healthy controls in the United States. Retrovirology **7:**57.
150. **Tailor, C. S., D. Lavillette, M. Marin, and D. Kabat.** 2003. Cell surface receptors for gammaretroviruses. Curr Top Microbiol Immunol. **281:**29-106.
151. **Tailor, C. S., A. Nouri, C. G. Lee, C. Kozak, and D. Kabat.** 1999. Cloning and characterization of a cell surface receptor for xenotropic and polytropic murine leukemia viruses. Proc Natl Acad Sci USA **96:**927-932.
152. **Takeuchi, Y., M. O. McClure, and M. Pizzato.** 2008. Identification of Gammaretroviruses Constitutively Released from Cell Lines Used for Human Immunodeficiency Virus Research. J. Virol. **82:**12585-12588.
153. **Tarlinton, R. E., J. Meers, and P. R. Young.** 2006. Retroviral invasion of the koala genome. Nature **442:**79-81.
154. **Temin, H. M.** 1981. Structure, variation and synthesis of retrovirus long terminal repeat. Cell **27:**1-3.
155. **Thomas, C. Y., R. Khiroya, R. S. Schwartz, and J. M. Coffin.** 1984. Role of recombinant ecotropic and polytropic viruses in the development of spontaneous thymic lymphomas in HRS/J mice. J. Virol. **50:**397-407.
156. **Tipper, C. H., C. E. Bencsics, and J. M. Coffin.** 2005. Characterization of Hortulanus Endogenous Murine Leukemia Virus, an Endogenous Provirus That Encodes an Infectious Murine Leukemia Virus of a Novel Subgroup. J. Virol. **79:**8316-8329.
157. **Tomonaga, K., and J. M. Coffin.** 1998. Structure and distribution of endogenous nonecotropic murine leukemia viruses in wild mice. J. Virol. **72:**8289-300.
158. **Tomonaga, K., and J. M. Coffin.** 1999. Structures of endogenous nonecotropic murine leukemia virus (MLV) long terminal repeats in wild mice: implication for evolution of MLVs. J. Virol. **73:**4327-40.
159. **Towers, G., M. Bock, S. Martin, Y. Takeuchi, J. P. Stoye, and O. Danos.** 2000. A conserved mechanism of retrovirus restriction in mammals. Proceedings of the National Academy of Sciences **97:**12295-12299.
160. **Tralka, T. S., C. L. Yee, A. B. Rabson, N. A. Wivel, K. J. Stromberg, A. S. Rabson, and J. C. Costa.** 1983. Murine type C retroviruses and intracisternal A-particles in human tumors serially passaged in nude mice. J Natl Cancer Inst. **71:**591-9.
161. **Tristem, M.** 2000. Identification and Characterization of Novel Human Endogenous Retrovirus Families by Phylogenetic Screening of the Human Genome Mapping Project Database. J. Virol. **74:**3715-3730.
162. **Tuke, P. W., K. I. Tettmar, A. Tamuri, J. P. Stoye, and R. S. Tedder.** 2011. PCR Master Mixes Harbour Murine DNA Sequences. Caveat Emptor! PLoS ONE **6:**e19953.

163. **Urisman, A., R. J. Molinaro, N. Fischer, S. J. Plummer, G. Casey, E. A. Klein, K. Malathi, C. Magi-Galluzzi, R. R. Tubbs, D. Ganem, R. H. Silverman, and J. L. DeRisi.** 2006. Identification of a Novel Gammaretrovirus in Prostate Tumors of Patients Homozygous for R462Q *RNASEL* Variant. PLoS Pathog **2:**e25.
164. **van Bokhoven, A., A. Caires, M. D. Maria, A. P. Schulte, M. S. Lucia, S. K. Nordeen, G. J. Miller, and M. Varella-Garcia.** 2003. Spectral karyotype (SKY) analysis of human prostate carcinoma cell lines. The Prostate **57:**226-244.
165. **Van Hoeven, N., and A. Miller.** 2005. Use of different but overlapping determinants in a retrovirus receptor accounts for non-reciprocal interference between xenotropic and polytropic murine leukemia viruses. Retrovirology **2:**76.
166. **van Kuppeveld, F. J. M., A. S. d. Jong, K. H. Lanke, G. W. Verhaegh, W. J. G. Melchers, C. M. A. Swanink, G. Bleijenberg, M. G. Netea, J. M. D. Galama, and J. W. M. van der Meer.** 2010. Prevalence of xenotropic murine leukaemia virus-related virus in patients with chronic fatigue syndrome in the Netherlands: retrospective analysis of samples from an established cohort. BMJ **340:**c1018-.
167. **Varthakavi, V., R. M. Smith, S. P. Bour, K. Strebel, and P. Spearman.** 2003. Viral protein U counteracts a human host cell restriction that inhibits HIV-1 particle production. Proceedings of the National Academy of Sciences **100:**15154-15159.
168. **Verhaegh, G. W., A. S. de Jong, F. P. Smit, S. A. Jannink, W. J. G. Melchers, and J. A. Schalken.** Prevalence of human xenotropic murine leukemia virus-related gammaretrovirus (XMRV) in dutch prostate cancer patients. The Prostate **71:**415-420.
169. **Vogt, P. K.** 1971. Genetically stable reassortment of markers during mixed infection with avian tumor viruses. Virology **46:**947-52.
170. **von Schwedler, U., J. Song, C. Aiken, and D. Trono.** 1993. Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. J. Virol. **67:**4945-4955.
171. **Wade, C. M., and M. J. Daly.** 2005. Genetic variation in laboratory mice. Nat Genet **37:**1175-1180.
172. **Wainstein, M. A., F. He, D. Robinson, H.-J. Kung, S. Schwartz, J. M. Giaconia, N. L. Edgehouse, T. P. Pretlow, D. R. Bodner, E. D. Kursh, M. I. Resnick, A. Seftel, and T. G. Pretlow.** 1994. CWR22: Androgen-dependent Xenograft Model Derived from a Primary Human Prostatic Carcinoma. Cancer Research **54:**6049-6052.
173. **Weiss, R.** A cautionary tale of virus and disease. BMC Biology **8:**124.
174. **Wildschutte, J. H.** 2011. A study of polymorphic endogenous retroviruses in humans: Implications in host health and genome evolution. Ph.D. Dissertation.
175. **Wolff, L., R. Koller, and S. Ruscetti.** 1982. Monoclonal antibody to spleen focus-forming virus-encoded gp52 provides a probe for the amino-terminal region of retroviral envelope proteins that confers dual tropism and xenotropism. J. Virol. **43:**472-481.
176. **Wong, P. K., and J. A. McCarter.** 1973. Genetic studies of temperature-sensitive mutants of Moloney-murine leukemia virus. Virology **53:**319-326.

177. **Wu, X., J. L. Anderson, E. M. Campbell, A. M. Joseph, and T. J. Hope.** 2006. Proteasome inhibitors uncouple rhesus TRIM5Œ± restriction of HIV-1 reverse transcription and infection. Proceedings of the National Academy of Sciences **103:**7465-7470.
178. **Wu, X., Y. Li, B. Crise, and S. M. Burgess.** 2003. Transcription Start Regions in the Human Genome Are Favored Targets for MLV Integration. Science **300:**1749-1751.
179. **Wunderli, H., D. D. Mickey, and D. F. Paulson.** 1979. C-type virus particles in human urogenital tumours after heterotransplantation into nude mice. Br J Cancer **39:**35-42.
180. **Yan, Y., R. C. Knoper, and C. A. Kozak.** 2007. Wild Mouse Variants of Envelope Genes of Xenotropic/Polytropic Mouse Gammaretroviruses and Their XPR1 Receptors Elucidate Receptor Determinants of Virus Entry. J. Virol. **81:**10550-10557.
181. **Yan, Y., Q. Liu, K. Wollenberg, C. Martin, A. Buckler-White, and C. A. Kozak.** 2010. Evolution of Functional and Sequence Variants of the Mammalian XPR1 Receptor for Mouse Xenotropic Gammaretroviruses and the Human-Derived Retrovirus XMRV. J. Virol. **84:**11970-11980.
182. **Yang, H., T. A. Bell, G. A. Churchill, and F. Pardo-Manuel de Villena.** 2007. On the subspecific origin of the laboratory mouse. Nat Genet **39:**1100-1107.
183. **Yang, H., J. R. Wang, J. P. Didion, R. J. Buus, T. A. Bell, C. E. Welsh, F. Bonhomme, A. H.-T. Yu, M. W. Nachman, J. Pialek, P. Tucker, P. Boursot, L. McMillan, G. A. Churchill, and F. P.-M. de Villena.** 2011. Subspecific origin and haplotype diversity in the laboratory mouse. Nat Genet **43:**648-655.
184. **Yang, Y.-L., L. Guo, S. Xu, C. A. Holland, T. Kitamura, K. Hunter, and J. M. Cunningham.** 1999. Receptors for polytropic and xenotropic mouse leukaemia viruses encoded by a single gene at Rmc1. Nat Genet **21:**216-219.
185. **Yonekawa, H., K. Moriwaki, O. Gotoh, N. Miyashita, Y. Matsushima, L. M. Shi, W. S. Cho, X. L. Zhen, and Y. Tagashira.** 1988. Hybrid origin of Japanese mice "Mus musculus molossinus": evidence from restriction analysis of mitochondrial DNA. Molecular Biology and Evolution **5:**63-78.
186. **Zhang, J., L.-Y. Tang, T. Li, Y. Ma, and C. M. Sapp.** 2000. Most Retroviral Recombinations Occur during Minus-Strand DNA Synthesis. J. Virol. **74:**2313-2322.
187. **Zhang, Y.-A., A. Maitra, J.-T. Hsieh, C. M. Rudin, C. Peacock, C. Karikari, R. A. Brekken, V. Stastny, B. Gao, L. Girard, I. Wistuba, E. Frenkel, J. D. Minna, and A. F. Gazdar.** 2011. Frequent detection of infectious xenotropic murine leukemia virus (XMLV) in human cultures established from mouse xenografts. Cancer Biology & Therapy **12**.
188. **Zhou, H., M. Xu, Q. Huang, A. T. Gates, X. D. Zhang, J. C. Castle, E. Stec, M. Ferrer, B. Strulovici, D. J. Hazuda, and A. S. Espeseth.** 2008. Genome-Scale RNAi Screen for Host Factors Required for HIV Replication. Cell host & microbe **4:**495-504.

APPENDIX

Mus spicilegus endogenous retrovirus HEMV uses murine sodium-dependent myo-inositol transporter 1 as a receptor

Christopher H. Tipper[1,†,*], Oya Cingöz[1,2,*], John M. Coffin[1,2,§]

[1]Department of Molecular Biology and Microbiology and [2]Genetics Program, Tufts University, Boston, Massachusetts 02111

* The authors contributed equally to this work.

† *Present address: Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Boston, Massachusetts 02115*

§ Corresponding author.

Mailing address: 136 Harrison Ave, Boston, Massachusetts, 02111.

Phone: (617) 636-6528

Fax: (617) 636-4086

Email: john.coffin@tufts.edu

ABSTRACT

We sought to determine the relationship between two recent additions to the MLV subgroups: the *Mus cervicolor* isolate M813 and the *Mus spicilegus* endogenous retrovirus HEMV. Though divergent in sequence, both viruses share an Env protein with similarly curtailed VRA and VRB regions, and both are restricted to mouse cells. HEMV and M813 displayed reciprocal receptor interference, suggesting that they share a common receptor. Expression of the M813 receptor mSMIT1 allowed previously non-permissive cells to be infected by HEMV, indicating that mSMIT1 also serves as a receptor for HEMV. Our findings add HEMV as a second member to the MLV subgroup that uses mSMIT1 to gain entry into cells.

The diversity of receptor usage by murine leukemia viruses (MLVs) and their close relatives reflects an active recent evolutionary history. All MLV subgroups identified to date have related *env* genes and use structurally similar but unrelated small molecule transporters as receptors (18). M813, an exogenous MLV isolated from *Mus cervicolor*, represents a novel MLV subgroup that uses the murine sodium-dependent myo-inositol transporter 1 (mSMIT1) for entry (10). The *Mus spicilegus* endogenous retrovirus HEMV (20) is one of the few gammaretroviruses that has yet to have its receptor identified (for a review see (18)).

HEMV was initially identified in our lab by a comprehensive analysis of endogenous MLV sequences from various *Mus* subspecies (21). Analysis of the LTRs of multiple endogenous MLV-related proviruses from wild-derived mice revealed differences exceeding the scope of the previous standard classification of proviruses found in inbred strains (4-7, 17, 21). The provirus containing an unusual LTR found only in the genome of *M. spicilegus* was named HEMV, and further analyses showed that the locus represented an intact provirus (21). HEMV bears hallmarks of an ancient endogenous retrovirus, in that it is fixed in its host species, the sequences of *gag* and *pol* occupy positions near the root of the MLV phylogenetic tree, and the LTR has the simplest structure of all known MLVs (21). However, further study revealed that it is a recent insertion into the *M. spicilegus* genome, and that the cloned HEMV provirus is capable of producing infectious virus (20).

The coding sequence of the HEMV *env* gene is substantially shorter than that of most other gammaretroviruses, due to differences in the VRA and VRB regions of SU that encode the receptor binding domain. Among other known gammaretroviruses, this

property was seen only in M813. Though highly divergent in other regions of the genome, including TM, HEMV and M813 share VRA and VRB regions of similar length (Fig. 1). Overall, HEMV and M813 *env* genes show 69% identity at the nucleotide level, corresponding to 72% at the protein level, comparable to the degree of amino acid identity in VRA (76%), and VRB (70%) regions. The most prominent difference between the two Env proteins is a three-residue gap in HEMV relative to M813 in the proline rich region (Fig. 1). Closer examination of the SU phylogenetic tree revealed that, although the two viral SU regions share a common branch in the tree, the branches leading to the viruses from their last common node are quite long, suggesting a distant relationship (20).

The species tropism of HEMV is best described as strictly ecotropic, reflecting its ability to infect only cells of mouse origin, including *M. spicilegus* (Table 1) (20). M813 has been reported to infect mouse and rat cells, although infection of the latter is three orders of magnitude less efficient (16). To better understand the receptor usage of HEMV and M813, we examined the species tropism of both viruses.

The M813 *env* gene was constructed from the published sequence (16) by linking regularly staggered 120 bp oligonucleotides in sequential PCR reactions, and cloning into pCR2.1-TOPO (Invitrogen) to create pM813-TOPO. Mismatches from the published sequence (GenBank ID: AF327437) were repaired by site-directed mutagenesis (Stratagene). The HEMV *env* gene from pHEMV18 (21), the M813 *env* gene from pM813-TOPO and the 10A1 *env gene* from pB6 (13) were cloned into pSV-psi-minus-E-MLV (12) to create psi-HEMV, psi-M813 and psi-10A1 constructs, respectively. Co-transfection of each construct with pLacPuro (15) or MLV-GFP (14) into 293T cells resulted in the production of single-round infectious $LacZ^{+}Puro^{R}$ or GFP^{+} viruses bearing

the specified *env* genes. Following infection of target cells, titers of the virus innocula were determined either by counting LacZ+ (staining blue with X-gal) cells under a light microscope, or by GFP expression via flow cytometry.

As previously reported, both HEMV and M813 were able to infect all mouse cell lines tested, derived from several different *Mus* species, whereas non-murine cells, including human, simian, feline, canine, avian and hamster cells were not infected by either virus (Table 1) (16, 20). In contrast to published data, we found M813 to be incapable of infecting rat cells, despite using at least one identical cell line (Rat1 cells; Table 1). Even though the published infection efficiency was very low in this species, it was within the sensitivity of the described assay as set by infection levels of truly resistant target cells. The reason for this discrepancy is unknown, but we believe it unlikely to be due to low M813 titers since an infection with several orders of magnitude less efficiency would still yield detectable titers. 10A1 pseudotypes were able to infect all cell lines tested with the exception of CHO-K1, while those with the ecotropic (Moloney) MLV Env were restricted to mouse and some rat cell lines, as expected (Table 1).

The similarities in the SU regions of the *env* proteins of HEMV and M813, as well as the identical host range seen in our experiments prompted us to use interference assays to test whether the two viruses were members of the same subgroup; i.e., used the same receptor for infection. HEMV and M813 *env* genes were cloned into pNCS (3), to generate replication-competent recombinant viruses with the specified Env proteins. NIH3T3 cells chronically infected with these viruses, as well as viruses with a variety of other receptor specificities, were then challenged with the $LacZ^{+}Puro^{R}$ HEMV, M813, 10A1 or ecotropic pseudotyped virions. M813 exhibited the same interference pattern as

reported in the literature (16), completely interfering with itself. No significant drop in titers of either M813 or HEMV was observed when the cells were pre-infected with other MLV subgroups including ecotropic, polytropic, amphotropic and 10A1 MLVs (Table 2), implying that both M813 and HEMV use a receptor different from that used by any of the tested subgroups. When NIH3T3 cells were infected with replication-competent M813, HEMV was completely blocked from infection. Likewise, chronic infection of NIH3T3 cells with HEMV abolished M813 infection (Table 2). This type of reciprocal interference is often seen with viruses that use a common entry receptor. Prior infection with HEMV or M813 did not have any effect on infection by 10A1 or ecotropic virus, suggesting they use a separate receptor for entry.

The reciprocal interference between HEMV and M813 implied usage of a common entry receptor. To test this possibility directly, cDNA of the mSMIT1 gene, the entry receptor for M813, was amplified from an NIH 3T3 cDNA library (11) and cloned into the retroviral expression vector pLPCX (Clontech) to create pLPCX-mSMIT. mSMIT transducing virus was made by co-transfecting 293T cells with pLPCX-mSMIT, pMLVgagpol and a VSV-G expression vector. Canine Cf2 Th cells were then infected either by mSMIT transducing virus, or by virus made with "empty" pLPCX, and selected with 10 μg/ml puromycin. Puromycin-resistant colonies were expanded and challenged by GFP-expressing MLV pseudotypes, and infection was scored by flow cytometry. The expression of mSMIT1 rendered non-permissive Cf2 Th cells susceptible to infection by both HEMV and M813 (Table 3), confirming mSMIT1 as a receptor for HEMV. MoMLV, capable of infecting only murine cells, remained unable to infect Cf2 Th cells,

whereas 10A1-MLV, which has a broad host range, was able to infect canine cells regardless of mSMIT1 expression.

The possibility that HEMV and M813 might use the same receptor was suggested by the similarity and phylogenetic relationship of their SU proteins, relative to other gammaretroviruses, and by the unusually short receptor-binding regions (VRA and VRB) shared by these viruses. They also have a common host range: unlike other MLVs they infect only cells from *Mus* species. Although only partial sequence is available for the M813 genome (limited to *env* and portions of *pol* and the LTR), available evidence suggests that the two viruses are not closely related outside of SU. Unlike SU, the TM region is relatively well conserved among gammaretroviruses, and has been previously used to establish viral relationships, most often congruent with RT, but not as precisely (1). Phylogenetic analysis places the M813 TM well away from HEMV, post branching of FeLV (20). The LTRs of the two viruses share about 77% sequence identity.

Three major hypotheses as to how HEMV and M813 came to belong to the same subgroup despite the obvious divergence of most of their genomes present themselves. The first is that HEMV and M813 share a common, although distant, ancestor that used an ancestral form of mSMIT1, and their divergence reflects evolution as an exogenous virus infecting different host species in different parts of the world while preserving the use of the same receptor. The second hypothesis is that both viruses have independently arrived in this subgroup through convergent evolution, both SU proteins being the result of jumping a similar evolutionary hurdle. The third possibility is that the two viruses evolved separately from a distant gammaretrovirus ancestor, and acquired the same receptor-binding domain by a relatively recent recombination event. It is worth noting

that the discordance between the SU and TM regions suggests the possibility of at least one recombination event in the evolutionary history of these viruses. Such events are well documented to alter the host range of MLVs, such as the 10A1 virus (8), as well as during replication of endogenous ecotropic viruses in some strains of mice (2, 9). In this case, recombination would require that all elements of the final virus be present either as exogenous or endogenous viruses in a single host. Given the rather different geographic ranges of *M. spicilegus* and *M. cervicolor* (Eastern Europe and Southern Asia), from which these two viruses were isolated, it is likely that other, intermediate viruses may exist. Identifying *pol* and TM regions intermediate to these positions might prove fruitful to understanding this relationship.

Although it is common for distantly related viruses to retain the use of cognate receptors, for instance in the case of GaLV and FeLV-B (19), and incremental changes in receptor usage are achievable both *in vivo* and *in vitro*, as seen in co-receptor usage by HIV-1 and HIV-2, understanding how a true subgroup switch occurs remains elusive. The establishment of HEMV and M813 as members of the same subgroup may have important implications for the evolution and spread of gammaretroviruses. Their relationship represents an experimentally tractable investigation of a subgroup's origin and evolutionary history. Investigating how the two viruses came to be members of the same subgroup will depend upon further exploration of wild-mouse populations, the examination of carrier species, and finally, the possible identification of more members of this subgroup.

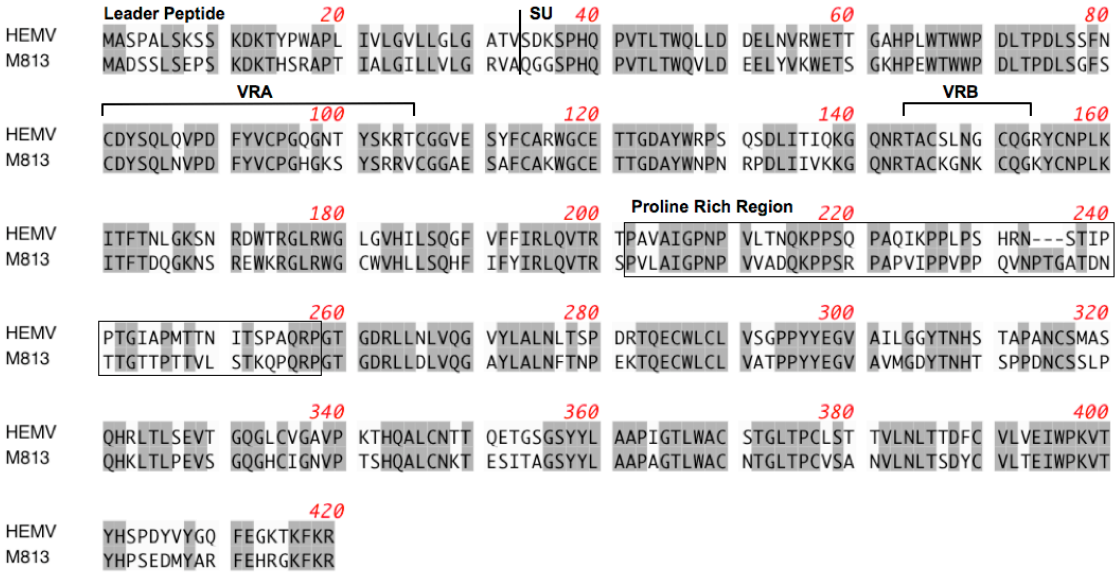


Figure 1. Relationship between HEMV and M813 Env proteins

Comparison of the SU regions of HEMV and M813 Env proteins. The leader peptide, VRA and VRB domains and the proline-rich region are indicated; the sequence represents the region upstream of the furin cleavage site.

TABLE 1. HEMV and M813 host ranges are restricted to mouse cells[a].

	Titer (IU/ml) of viral pseudotypes			
Cell Line	**HEMV**	**M813**	**MoMLV**	**10A1**
Mouse				
NIH3T3	$3.2x10^5$	$1.5x10^5$	$3.6x10^5$	$5.1x10^5$
MMK	$7.0x10^5$	$1.2x10^5$	$4.8x10^5$	$3.2x10^5$
SC1	$1.7x10^5$	$1.9x10^5$	$3.0x10^5$	$2.3x10^5$
M. dunni	$1.4x10^6$	$3.4x10^5$	$2.0x10^5$	$4.8x10^5$
M. spicilegus	$1.7x10^5$	$4.1x10^4$	$1.7x10^5$	$1.3x10^5$
MC3T3	$2.1x10^5$	$7.3x10^4$	$2.9x10^5$	$2.7x10^5$
Rat				
Rat1	<1	<1	$8.5x10^4$	$2.8x10^5$
REF	<1	<1	$1.9x10^2$	$7.2x10^5$
Hamster				
E36	<1	<1	<1	$2.8x10^5$
CHO-K1	<1	<1	<1	<1
Primate				
COS1	<1	<1	<1	$8.7x10^4$
293T	<1	<1	<1	$4.4x10^5$
293TmCAT1	<1	<1	$4.0x10^5$	$2.5x10^5$
Avian				
DF1	<1	<1	<1	$2.4x10^5$
QT6	<1	<1	<1	$1.7x10^5$
Cat				
FEF	<1	<1	<1	$3.4x10^5$
Dog				
D17	<1	<1	<1	$2.3x10^5$

[a] The MLV-based pLacPuro vector construct was cotransfected with clones expressing the indicated *env* genes. Titers were determined by X-gal staining. 10A1 and MoMLV pseudotypes served as controls. The data represent at least two independent experiments.

TABLE 2. Receptor interference between MLV Env proteins.

Pre-infecting Virus	Relative titer of Challenge Virus[a]			
	HEMV	**M813**	**Ecotropic**	**10A1**
None	1	1	1	1
HEMV	$<3.1\times10^{-6}$	$<6.5\times10^{-6}$	0.42	1.51
M813	$<3.1\times10^{-6}$	$<6.5\times10^{-6}$	2.03	0.13
Ecotropic	0.54	0.17	$<2.6\times10^{-6}$	0.73
Amphotropic	0.50	0.98	0.64	0.84
Polytropic	0.15	0.26	0.20	0.13
10A1	1.08	0.24	0.32	$<1.96\times10^{-4}$

[a] Calculated as the ratio of titers on pre-infected cells and on uninfected cells.

TABLE 3. HEMV and M813 use mSMIT1 as a receptor.

	Percent GFP-positive cells[a]				
Target Cell	**HEMV**	**M813**	**Ecotropic**	**10A1**	**No Envelope**
NIH 3T3	22.9 +/- 0.7	8.2 +/- 0.2	35.5 +/- 1.4	17.3+/- 1.3	3.3 +/- 0.5
Cf2 Th LPCX "empty"	2.0 +/- 0.4	1.6 +/- 0.4	2.0 +/- 0.1	15.9 +/- 0.4	1.8 +/- 0.4
Cf2 Th LPCX-mSMIT	11.0 +/- 0.2	5.2 +/- 0.2	1.9 +/- 0.2	14.6 +/- 0.6	1.9 +/- 0.3

[a]Percent infection of 10,000 gated viable cells, threshold set at 1.0% FITC in uninfected cells.

APPENDIX REFERENCES

1. **Benit, L., P. Dessen, and T. Heidmann.** 2001. Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. J Virol **75:**11709-19.
2. **Bosselman, R. A., F. van Straaten, C. Van Beveren, I. M. Verma, and M. Vogt.** 1982. Analysis of the env gene of a molecularly cloned and biologically active Moloney mink cell focus-forming proviral DNA. Journal of Virology **44:**19-31.
3. **Colicelli, J., and S. P. Goff.** 1988. Isolation of an integrated provirus of Moloney murine leukemia virus with long terminal repeats in inverted orientation: integration utilizing two U3 sequences. J Virol **62:**633-6.
4. **Frankel, W. N., and J. M. Coffin.** 1994. Endogenous nonecotropic proviruses mapped with oligonucleotide probes from the long terminal repeat region. Mamm Genome **5:**275-81.
5. **Frankel, W. N., J. P. Stoye, B. A. Taylor, and J. M. Coffin.** 1990. A linkage map of endogenous murine leukemia proviruses. Genetics **124:**221-36.
6. **Frankel, W. N., J. P. Stoye, B. A. Taylor, and J. M. Coffin.** 1989. Genetic analysis of endogenous xenotropic murine leukemia viruses: association with two common mouse mutations and the viral restriction locus Fv-1. J Virol **63:**1763-74.
7. **Frankel, W. N., J. P. Stoye, B. A. Taylor, and J. M. Coffin.** 1989. Genetic identification of endogenous polytropic proviruses by using recombinant inbred mice. J Virol **63:**3810-21.
8. **Han, J. Y., P. M. Cannon, K. M. Lai, Y. Zhao, M. V. Eiden, and W. F. Anderson.** 1997. Identification of envelope protein residues required for the expanded host range of 10A1 murine leukemia virus. J Virol **71:**8103-8.
9. **Hartley, J. W., N. K. Wolford, L. J. Old, and W. P. Rowe.** 1977. A new class of murine leukemia virus associated with development of spontaneous lymphomas. Proc Natl Acad Sci U S A **74:**789-92.
10. **Hein, S., V. Prassolov, Y. Zhang, D. Ivanov, J. Lohler, S. R. Ross, and C. Stocking.** 2003. Sodium-dependent myo-inositol transporter 1 is a cellular receptor for Mus cervicolor M813 murine leukemia virus. J Virol **77:**5926-32.
11. **Kitamura, T., M. Onishi, S. Kinoshita, A. Shibuya, A. Miyajima, and G. P. Nolan.** 1995. Efficient screening of retroviral cDNA expression libraries. Proc Natl Acad Sci U S A **92:**9146-50.
12. **Mann, R., R. C. Mulligan, and D. Baltimore.** 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. Cell **33:**153-9.
13. **Ott, D., R. Friedrich, and A. Rein.** 1990. Sequence analysis of amphotropic and 10A1 murine leukemia viruses: close relationship to mink cell focus-inducing viruses. J Virol **64:**757-66.
14. **Perron, M. J., M. Stremlau, B. Song, W. Ulm, R. C. Mulligan, and J. Sodroski.** 2004. TRIM5alpha mediates the postentry block to N-tropic murine leukemia viruses in human cells. Proceedings of the National Academy of Sciences of the United States of America **101:**11827-11832.

15. **Pfeiffer, J. K., R. S. Topping, N. H. Shin, and A. Telesnitsky.** 1999. Altering the intracellular environment increases the frequency of tandem repeat deletion during Moloney murine leukemia virus reverse transcription. J Virol **73:**8441-7.
16. **Prassolov, V., S. Hein, M. Ziegler, D. Ivanov, C. Munk, J. Lohler, and C. Stocking.** 2001. Mus cervicolor murine leukemia virus isolate M813 belongs to a unique receptor interference group. J Virol **75:**4490-8.
17. **Stoye, J. P., and J. M. Coffin.** 1988. Polymorphism of murine endogenous proviruses revealed by using virus class-specific oligonucleotide probes. J Virol **62:**168-75.
18. **Tailor, C. S., D. Lavillette, M. Marin, and D. Kabat.** 2003. Cell surface receptors for gammaretroviruses. Curr Top Microbiol Immunol **281:**29-106.
19. **Takeuchi, Y., R. G. Vile, G. Simpson, B. O'Hara, M. K. Collins, and R. A. Weiss.** 1992. Feline leukemia virus subgroup B uses the same cell surface receptor as gibbon ape leukemia virus. Journal of Virology **66:**1219-1222.
20. **Tipper, C. H., C. E. Bencsics, and J. M. Coffin.** 2005. Characterization of Hortulanus Endogenous Murine Leukemia Virus, an Endogenous Provirus That Encodes an Infectious Murine Leukemia Virus of a Novel Subgroup. J. Virol. **79:**8316-8329.
21. **Tomonaga, K., and J. M. Coffin.** 1999. Structures of endogenous nonecotropic murine leukemia virus (MLV) long terminal repeats in wild mice: implication for evolution of MLVs. J Virol **73:**4327-40.

- FIN -