Evolutionary Spread and Recombination of Porcine Endogenous Retroviruses in the Suiformes

Marcus Niebert† and Ralf R. Tönjes*
Paul-Ehrlich-Institut, Langen, Germany

Received 31 May 2004/Accepted 23 August 2004

Porcine endogenous retroviruses (PERV) pose a serious challenge in xenotransplantation, i.e., the transfer of cells, tissues, or organs from one species to another for therapeutic reasons, since they are present in all cells and in almost every individual pig (27) and are transmitted vertically.

Endogenous retroviruses have been detected in the genomes of all vertebrate species (4, 14), and their general organization corresponds to that of exogenous retroviruses (4, 40, 46); however, most are replication-incompetent, while only a minority are functional, as reported, e.g., for mice (5, 10, 41) and pigs (1, 2, 21, 26, 28, 29).

For gamma-type PERV, three different classes exist, designated PERV-A, PERV-B, and PERV-C (1, 23, 42). The first two classes (7, 42) productively infect human cells in vitro, thus posing a serious risk in xenotransplantation, while the latter (1) does not replicate on human cells. There are only minor genetic differences between the classes, being most prominent in the receptor-binding domain of the Env protein. In addition, there are two different types of long terminal repeats (LTRs) that significantly affect the replication properties of single viruses (37, 47) through binding of transcription factor NF-Y (36). PERV-A and PERV-B proviruses demonstrate LTRs that harbor distinct 39-bp repeats in U3 which enable high transcription levels and adaptation to new host cells by multimerization of these repeats, with the transcriptional activity being generally stronger if more repeats are present. On the other hand, PERV-A and PERV-C were found to display repeatless LTRs (37, 47). Although sequences homologous to the 39-bp repeat are present in these LTRs, they are not organized in a repetitive manner and do not show multimerization as a response to replication cycles in their hosts. This LTR type confers a very low transcriptional activity, and we propose that it originated as an adaptation to endogenous replication of PERV. This leaves PERV-A as the only PERV harboring both types of LTR.

We recently determined the age of PERV as having an upper limit of approximately 7.6 × 10^6 years, while the repeatless LTR type evolved approximately 3.4 × 10^6 years ago and is therefore the phylogenetically younger structure (45). The age determined for PERV correlates with the time of separation of pig species (Suidae, Sus scrofa) from their closest relatives, American-born peccaries (Tayassuidae, Pecari tajacu), 7.4 × 10^6 years ago. While the time of the phylogenetic split was calculated by using mitochondrial DNA sequences (19), which tend to underestimate the time, archaeological data suggest a split of Suidae and Tayassuidae in the Eocene epoch, about 15 MYA (43, 44).

To extend the above-mentioned study on PERV’s relative age, we analyzed a wide array of Suidae of Eurasian and African origin, as well as samples of the Tayassuidae as the closest evolutionary relatives of pigs, for the presence or absence of PERV. In anthropological studies, the presence or absence of human endogenous retroviruses in various hominoids has been used successfully to designate the age of the proviruses (17, 18, 39). Therefore, we aimed for a similar approach to define the age and spread of the different PERV genotypes more accurately.

**Sample acquisition.** Suiformes of various species were assayed. They are listed below in decreasing order of relationship to Sus scrofa. The epoch when the phylogenetic split occurred, as determined by archaeological evidence (43, 44) and genetic reconstruction (8, 11, 15, 24), is given in parentheses: Sus barbatus, Sus celebensis (Holocene), Potamochoerus porcus, Potamochoerus larvatus (Pleistocene), Phacochoerus africanus, Babyrussa babyrussa (Pliocene), and Tayassuidae pecari (Eocene). A detailed classification of the families of Suidae and Tayassuidae, including subfamilies and genera, is given in Fig. 1.

**Phylogenetic analysis.** Recombination is an important process that has impact on biological evolution at different levels.
Recombination reshuffles existing variation and even creates new variants at the amino acid level. It breaks down linkage disequilibrium and, especially, has a significant impact on the evolution of viral pathogens (12, 13). Common phylogenetic analysis depicts the history of a number of sequences as a tree where two sequences share a common ancestor. This model fails when recombination is present, because a sequence now is no longer limited to a single ancestor. But this poses a problem only if one tries to reconstruct a phylogenetic tree from a sample for which it is unknown if recombination has occurred. Here, we try to determine whether recombination played a role in PERV evolution. A number of methods have been developed to detect recombination, and their relative performance has been analyzed (9, 30, 48). In accordance with these comparisons, we have chosen SimPlot because it performed well in PERV analysis depicting the history of a number of sequences as a tree through time, the determined sequences represent independent snapshots of different specimens of different species. Analyses of modern pigs suggest a number of (6 to 10) replication-competent proviruses (26), 30 to 50 full-length PERV (1, 2, 28, 29), and 100 to 200 loci encompassing partial proviruses (35). Even under the assumption of less frequent PERV integration in older Suiformes, there should be enough sequences for detection by PCR. Thus, we believe that the limited number of animals should not pose a problem when interpreting samples testing negative. A cytochrome b phylogenetic tree was generated to calibrate the relative age of PERV in different Suiformes samples (Fig. 2A), including the archaeological data on Suiformes fossils. This tree is in line with previous studies using cytochrome b sequences (31) or archaeological data (43, 44), but includes three additional families not analyzed hitherto (S. celebensis, P. porcus, and P. larvatus). PERV sequences are completely undetectable in T. pecari and in B. babyrussa (Eocene and Miocene epochs, respectively), while the earliest presence of PERV happens in P. africanus associated with the late Miocene or early Pliocene epoch (3.5 to 7.5 MYA) (Fig. 2A and B), which confirms our recent study of the age of PERV (45). Only PERV-A can be detected in samples of P. africanus, while the earliest appearance of PERV-B is in the slightly younger P. porcus associated with the early Pleistocene epoch (3.5 to 7.5 MYA) (Fig. 2A and B). In contrast, PERV-C is detected for the first time in the much younger S. barbatus of the early Holocene epoch (0.1 to 1.5 MYA) (Fig. 2A and B). The repeatless LTR, on the other hand, has been detectable since appearing in P. larvatus and P. porcus of the late Pleistocene epoch (1.5 to 3.5 MYA) (Fig. 2B) (45). While we therefore assume a separate origin for PERV-C and the repeatless LTR, both could have emerged together as discussed below.

Patience et al. (28) also analyzed a series of Suiformes for the presence of beta- and gamma-retroviruses, and while our results match at large, we disagree on some minor points. As it is difficult to acquire exotic DNA samples in the first place, we were not able to repeat the study with the exact set of Suiformes used by Patience et al.; some species may differ in subspecies, and only a limited number of specimens were available for analysis. In addition, the primer sets used by Patience et al. and by us differ in sequence and therefore may lead to different results, particularly when only minor sequence variations have an impact on the detection limit. Hence, these differences may account for all minor disagreements. In contrast to our own findings, Patience et al. detected PERV-C in the sample of P. larvatus of the Pleistocene epoch, therefore shifting the emergence of PERV-C by one epoch to 1.5 to 3.5 MYA, though coinciding with the appearance of the repeatless LTR.

**Analysis of interclass mosaicism.** SimPlot, an interactive 32-bit program for Windows, was created to plot sequence similarity versus position (32). Briefly, SimPlot calculates and plots the percent identity of the query sequence to a panel of reference sequences in a sliding window, which is moved stepwise across the alignment. The window and step sizes are adjustable. Alignments were analyzed for recombination breakpoints by maximization of $\chi^2$ as previously described (34, 38).

In general, the homologies between PERV-A and PERV-C are approximately 85%, while the similarities between PERV-B and either PERV-A or PERV-C barely exceed 70%. In general, gamma-type retroviruses, including PERV, share a common homology of approximately 60%. This fact and the occurrence of repeatless LTRs in both PERV-A and PERV-C but not in PERV-B (see above), leads to the assumption of a common evolutionary origin for these two classes. A possible recombination of PERV classes is most likely in the LTR or env sequences, and analysis revealed negligible sequence variations in gag and pol (data not shown).

Detection of recombination was carried out by comparing PERV env sequences obtained from various Suiformes to reference sequences for PERV-A (AJ133817) (7), PERV-B (AJ133818) (7), and PERV-C (AF038600) (1). Some PERV env sequences obtained from “old” Suiformes showed the PERV-A or PERV-C sequence and an analysis was therefore
not necessary, while A/C recombinant sequences obtained from these samples showed a high degree of homology. Because of this finding and to keep the representation in Fig. 3 as simple as possible, the analyses shown in Fig. 3 were displayed with only one representative sequence per species.

The first recombination event between PERV envelope genes is detectable in *S. celebensis*, affecting the 3’ end of a PERV-A env, which acquires some sequence homology of PERV-C (Fig. 3 B). This change affects mostly the C-terminal region of the env gene, but does not change the class of the env gene (Fig. 3 A). Therefore, *S. celebensis* is still considered negative for PERV-C as revealed for the other evolutionary older Suiformes (Fig. 2 B). This observation has two important implications. The change leads to a slightly truncated R peptide, with this R peptide structure being associated with repeatless LTRs in PERV-A proviruses. In addition, a truncated cytoplasmic tail has been shown to increase the fusogenic potential of several gamma-type retroviruses (6, 16, 33) and of PERV-A and PERV-B (3). As PERV-C seems to harbor only repeatless LTRs, it is conceivable that recombination happened between only one repeatless PERV-A provirus and the unknown PERV-C progenitor. The changed R peptide may have played an essential role in the origin of PERV-C by enabling the incorporation of the modified envelope into the capsid in the first place or by offsetting some of the limitations posed by the less active LTR through larger fusogenicity. However, the latter would also apply to repeatless PERV-A proviruses (3). A second recombination event is detectable with *S. barbatus*, affecting the receptor binding domain between nucleotide positions 300 and 900 of the env gene (Fig. 3 C), thus creating a new class of PERV (Fig. 3 A) and enabling the detection of PERV-C for the first time (Fig. 2 B). In samples from *S. scrofa*, PERV-C was detectable with reasonable abundance, as the homology with evolutionary older forms is already as high as 98%. Therefore, mostly point mutations are sufficient to change the query sequences to match the env reference sequence almost perfectly (Fig. 3 D).

There was no detectable recombination event between PERV-B and either PERV-A or PERV-C in the samples (data not shown), but recent publications of PERV sequences suggest that the different PERV classes still recombine (20, 49). Recombination has been well documented for RNA viruses (50) and most likely involves similarity-assisted template switching (22, 25). The frequency of successful intertypic genetic exchanges is determined by, among other factors, (i) properties inherent to the process of viral replication, that is, the error susceptibility of viral reverse transcriptase and cellular RNA polymerase II; (ii) the frequency of cross-species
FIG. 3. Analysis of PERV envelope mosaic isolates showing similarity plots generated by SimPlot (32). PERV-A–PERV-C recombinant sequences obtained from old Suiformes showed a high degree of homology; therefore, the analysis was displayed with a single representative
transmission; (iii) the viability of the recombinant progeny; and (iv) the evolutionary gain.

Conclusion. PERV originated in African members of the Suidae family about 7.5 MYA (Fig. 2A and B), with repeat-harboring U3 sequences representing the exclusive type of LTR. The repeatless LTR developed during the early Pliocene epoch (3.5 MYA) in African Suidae, with its weak transcriptional activity most likely being an adaptation to an endogenous replication cycle. It is difficult to conclude whether PERV-A and PERV-B developed independently or whether both virus classes originated from a similar event like the A and C recombination described here. If so, the event took place too early to be clarified in this study. PERV-C, being much more closely related to PERV-A than to PERV-B, did not arise in the same epoch as PERV-A and PERV-B, but originated nearly 3.5 million years later due to a recombination event between PERV-A and an unknown ancestor. While the A and C recombination coincides with the appearance of the repeatless LTR, we assume that these events are independent from each other. Furthermore, we suggest that the recombination process leading to PERV-C involved an unknown ancestor and a PERV-A variant with a repeatless LTR.

Oligonucleotide sequences. Porcine cytochrome b sequences were amplified by nested PCR with outer forward primer 5′-GCT TAC CCT TTC CAA CTA GGC TTC TCC-3′ and outer reverse primer 5′-TTG CAA GAT CTT TAA TGG GAC GAG-3′ and inner forward primer 5′-CAC ACA CTA GCA CAA TGG ATG CC-3′ and inner reverse primer 5′-GAG GAT ACT GAT ATT CGG ATG ATT AT-3′ by using a regimen of 35 cycles of denaturation (94°C for 30 s), annealing (58°C for 30 s), and elongation (72°C for 90 s). PERV polymerase sequences were detected by PCR with the specific primer pair PERV-pol-forward (5′-TTG ACT TGG GAG TGT AGG GCT TCG TCA AAG ATG-3′) and PERV-pol-reverse (5′-GAT ACT AAT ATT CGG ATT GTT AT-3′) by using a regimen of 35 cycles of denaturation (94°C for 1 min), annealing (58°C for 1 min), and elongation (72°C for 3 min) as described in reference 7. For enhanced sensitivity, a nested PCR was performed with the same regimen with the following primer pair: PERV-pol-forward-nested (5′-GGT AAC CCA CTC GTT TTC TGG TCA-3′) and PERV-pol-reverse-nested (5′-GGT CCG TGT AGG GCT TCG TCA AAG ATG-3′). The chances of detecting polymorphic genes of non-PERV gamma-retroviruses or even from those of other related virus classes are negligible. PERV envelope gene class-specific detection was done with the same PCR regimen used for cytochrome b detection with the following primer pairs: PERV-A forward (5′-ATC CTA CCA GGT TTA ATA ATC AAT-3′) and PERV-A reverse (5′-GAT TAA AGG CTT CAG TGT GG-3′), PERV-B forward (5′-GGA TAA ATG GTA TGA GCT GG-3′) and PERV-B reverse (5′-GCT CAT AAA CCA CAG CAT TAC TAT-3′), and PERV-C forward (5′-CAC CTA TAC CAG CTC TGG ACA ATT-3′) and PERV-C reverse (5′-TAA ACA ACC AGG CTC CAT TCT AAA-3′). Envelope gene reference sequences used in SimPlot analysis were taken from previously published, full-length viruses PERV-A (AJ133817) (7), PERV-B (AJ133818) (7), and PERV-C (AF038600) (1).


