Foamy viruses (FVs) form a subfamily of retroviruses. Although FV cause a persistent infection, they are harmless, implying that the virus may be a valuable tool for viral vector development. FV have been isolated and cloned from a variety of primate (for references see below) and non-primate hosts (Holzscher et al., 1998; Tobaly-Tapiero et al., 2000; Winkler et al., 1997). Evolutionary studies would benefit from the characterization of the FV genomic diversity, but to date, only a few primate FVs from macaques (SFVmac) (Kupiec et al., 1991), African green monkey (SFVagm) (Renne et al., 1992), orang-utan (SFVora) (Verschoor et al., 2003), chimpanzees (SFVcpz) (Herchenroeder et al., 1994, 1995), spider monkey (SFVspm) (Thumer et al., 2007), marmoset and squirrel monkey (Pacheco et al., 2010) have been cloned and completely sequenced. Although the gorilla simian FV (SFVgor) has been previously isolated and partially sequenced (Bieniasz et al., 1995; Calattini et al., 2004), the complete sequence of the SFVgor represents the last missing piece of the FV genomic puzzle within the great apes. Since gorillas are the closest related primates to humans, next to chimpanzees, the complete sequence of SFVgor has the potential to provide additional insights into virus and primate evolution, which are known to be related by a history co-speciation (Switzer et al., 2004). Facilitated by only small differences between cellular restriction factors and between the immune system of gorillas and humans, gorilla-derived retroviruses can be transmitted to humans. This could result in a potential threat even with the virus being non-pathogenic in its natural host. These potential dangers are highlighted by the fact that, at least for SFVcpz and SFVora, such inter-species transmissions are not rare events and were even observed for SFVgor as well (Calattini et al., 2007; Heneine et al., 1998; Switzer et al., 2004). In this regard it should be noted that the prototype foamy virus (PFV), which was supposed to be a specific human isolate, is most probably a transmission from chimpanzees (Epstein, 2004).

To broaden our knowledge of FV evolution and genomic diversity, we have cloned and determined the sequence of the complete SFVgor. The SFVgor isolate, derived from a western lowland gorilla and described by Biensz et al. (1995), was propagated in baby hamster kidney cells (BHK-21) in a co-culture system, as previously described, in order to yield high infection rates (Biensz et al., 1995). Total DNA was isolated from the infected cells and served as template in the PCRs. The viral genome was amplified in two segments starting from the 5' LTR to the integrase (IN) region and from the IN to the 3' LTR. The PCR products were cloned into the PCR-XL-Topo (Invitrogen) vector in accordance with the manufacturer’s manual and three independent clones of each fragment were isolated and completely sequenced by primer walking. Sequence assembly was performed with DNASTAR Lasergene Software. Since some of the isolated clones contained frame-shift mutations in the pol region, we reamplified, cloned and resequenced this part of the genome. These sequences...
showed no frame-shifts or unexpected stop codons. To determine the complete sequences of both LTRs, and to verify the obtained sequence a primer located at the start of the 5′ LTR was used to amplify the genome up to the gag gene, whereas the 3′ end was amplified from the bel region until the 3′ end of the LTR. All sequencing primers used in this study and a detailed clone description is available upon request.

Sequence similarity analysis revealed the overall length of the LTR to be 1283 nt (Supplementary Table S1, available in JGV Online). Based on this analysis the U3, R, U5 regions were defined as being 944, 181 and 158 bp, respectively. A TATA box motif was identified from −26 to −21 in the LTR. Sequence comparison with the PFV Tas responsive element (TRE) (Rethwilm et al., 1987) revealed a highly conserved sub-sequence from −47 to −58. A splice-donor sequence was found in the R region, defining exon one as 51 nt. The primer-binding site is located, as expected, downstream of the LTR. A central polyurine-tract is located at 6490–6501 (Peters et al., 2008).

The position of the internal promoter (IP) was determined by similarity to PFV at nt position 9329 (Löchelt et al., 1993). A TATA box sequence at −20 to −15 was found upstream of the potential IP. In addition, due to the high similarity of the PFV and SFVgor sequences, we could determine a potential TRE at −151 to −133 of the IP (Kang & Cullen, 1998). This TRE shows a sequence identity of 18 of the 19 nt to its PFV counterpart. The 3′ polyurine-tract was determined directly before the 3′ LTR at 10949–10972. A poly(A) signal was identified from 12072 to 12077 in the 3′ LTR and it can be assumed that the poly(A) addition takes place at nt 12097 of the proviral DNA. This would define a genomic transcript with the poly(A) addition takes place at nt 12097 of the proviral DNA. This would define a genomic transcript with

This protein processing would give rise to an IN protein of 389 aa. The catalytic centre (YVDD) of the RT encompasses the aa residues 312–315. The zinc-finger motif of the IN was found at H\textsuperscript{H414}–H\textsuperscript{H418}–C\textsuperscript{C424}–C\textsuperscript{C428} and the (DD35E) catalytic centre motif at D\textsuperscript{D380}–D\textsuperscript{D397}–E\textsuperscript{E393}.

The env gene encodes a 987 aa protein (Supplementary Table S1). The FV Env precursor is cleaved in contrast to the ortho-retroviral Env by cellular furin-like proteases in a particle-associated leader peptide (LP), the surface (SU) and transmembrane (TM) proteins. The LP/SU cleavage is likely to occur at aa 126 (RIAR/ALRA) (Pfrepper et al., 1998), while the SU/TM cleavage site was determined by similarity at aa 568 (NQQR/KREVN) (Duda et al., 2004).

The Tas ORF was defined from 9577 to 10476 giving rise to a protein of 299 aa (Supplementary Table S1). This protein shows a strikingly high similarity in the DNA-binding domain to the Tas sequence of PFV (Supplementary Tables S2 and S3, available in JGV Online), which is reflected by a high conservation of the TREs in both promoters.

The Bet ORF encompasses the 5′ Tas reading frame from nt 9577 to 9838 (Supplementary Table S1). The latter is joined by splicing with nt 10140, resulting in a Bet ORF of 1443 nt and 480 aa, respectively. The Bet amino acid sequence, with 80% similarity, is highly homologous to the PFV counterpart.

To determine the evolutionary relationship between SFVgor and other simian or non-primate FV, we reconstructed the evolutionary history of the concatenated gag, pol and env amino acid sequences for all cloned FV and compared this to the host evolutionary relationships inferred from their mitochondrial DNA sequences. We employed Bayesian Markov chain Monte Carlo (MCMC) analysis implemented in BEAST (Wang & Mulligan, 1999) to estimate posterior distributions of time-measured phylogenetic trees (Drummond & Rambaut, 2007). To calibrate both the host and viral phylogenies under the assumption of ancient co-speciation, we identified two host calibration dates available from the fossil record that could be applied to nodes with consistent relative divergence times (Supplementary Material, available in JGV Online). The host and viral evolutionary histories are summarized on the same timescale in Fig. 1. All primate FV sequences clustered according to their host species, which would argue for a long-lasting co-evolutionary process. However, the relationships among the feline (F), bovine (B) and equine (E) FV, with FFV basal to BFV and EFV, did not mirror the host phylogeny, which places the horse lineage basal to the bovine and feline cluster. It has, however, been shown that relationships among these hosts are not consistent among different genomic segments (Drummond & Rambaut, 2007).

The matching virus and primate host trees are a prerequisite to the assumption of co-speciation, but also the ages of the corresponding common ancestors are expected to be concordant. If this was not the case then the
matching phylogenies could also be explained by preferential host switching (Prasad et al., 2008), which postulates that viruses are more likely to jump between related host species. With a few exceptions, there is generally a wide overlap among the credible intervals for the virus and host divergence times (Fig. 1). The PFV-SFVcpz divergence is for example younger than the respective host divergence times, but this can be expected as the divergence is the result of viral evolution within chimpanzee species that forms geographical isolated clusters. PFV is in fact the result of a cross-species transmission from chimpanzees to humans rather than co-speciation (Charleston & Robertson, 2002). Retroviruses are actively crossing into humans exposed to non-human primates, from chimpanzees as well as other non-human primates (Switzer et al., 2004; Wolfe et al., 2004). Furthermore, there are also many examples of FV cross-species transmission among non-human primates, e.g. from cercopithecus monkeys to chimpanzees (Epstein, 2004) and from colobus monkeys to chimpanzees (Liu et al., 2008). Although this complicates the history of FV, not all cross-species transmission will lead to successful onwards transmission in the new host and co-speciation may still hold as a long-term evolutionary process.

Assuming ancient co-speciation in primate hosts, the viral rate of amino acid substitution was estimated to be $1.31 \times 10^{-8}$ (1.14 x 10^{-8} – 1.48 x 10^{-8}) substitutions per site per year. This is a remarkable low rate of evolution for a virus with RT that is as error-prone as other retroviruses. It has been suggested that FV may replicate through clonal expansion, analogous to primate T-lymphotropic viruses (Leendertz et al., 2008; Verschoor et al., 2003), thereby avoiding the introduction of many RT-induced mutations over short timescales. However, although most tissues harbour latent provirus in naturally infected rhesus macaques, viral replication does take place in oral mucosa (Murray et al., 2008), which complicates an explanation of low evolutionary rates by clonal expansion.

We decided to analyse further the evolution of SFVgor by using sequences derived from primary isolates, since differences from the complete SFVgor sequence to the isolates would argue for an adaptation to the cell culture conditions. We determined sequence variations up to 5% in the 425 bp IN fragment. The complete SFVgor sequence showed a high homology to previously published wild-type isolates within a range of 95.5% nt identity. The published sequences of the 425 bp IN fragment of SFVcpz were aligned as control and showed divergence in the same range. In addition, differences in the same order were found previously for SFVcpz (Wattel et al., 1995).

Several studies have suggested that FV sequence diversity can be used as a marker for host evolution. The sequence data presented here appear to confirm this assumption. However, comparing our trees to the trees previously published for the 425 bp fragment in the IN coding region,
an obvious difference in the grouping of the SFVgor can be observed. Bieniasz et al. (1995) grouped this fragment in the branch near to SFVmac and SFVagm, in contrast to whole Pol sequence, which is grouped in the SFVcpz branch. This indicates that small sequence fragments may lead to different phylogenetic trees, which would argue for the use of complete sequences in order to get reliable results.

In summary, this report underlines that FVs are, indeed, old viruses, which display a long co-evolution with their respective hosts. Differences between infected animals account for less than 5% sequence deviation. The reason why FV evolution is such a slow process remains to be discovered.

Acknowledgements

We would like to thank Tanja Schied for technical support and Sylvia Geubig for help with the preparation of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (B0006/1-2 and SFB 479). P.L. was supported by a postdoctoral fellowship from the Fund for Scientific Research (FWO) Flanders.

References


