Characterization of novel polyomaviruses from Bornean and Sumatran orang-utans

Marlous J. Groenewoud,1† Zahra Fagrouch,1 Sabine van Gessel,1‡ Henk Niphuis,1 Aiste Bulavaite,2 Kristin S. Warren,3 Jonathan L. Heeney1§ and Ernst J. Verschoor1

1Department of Virology, Biomedical Primate Research Centre, Rijswijk, The Netherlands
2Institute of Biotechnology, Vilnius, Lithuania
3School of Veterinary and Biomedical Sciences, Murdoch University, Perth, Western Australia, Australia

Correspondence
Ernst J. Verschoor
verschoor@bprc.nl

Serological screening of sera from orang-utans demonstrated a high percentage of sera that cross-reacted with antigens of the polyomavirus (PyV) simian virus 40. Analysis of archival DNA samples from 71 Bornean and eight Sumatran orang-utans with a broad-spectrum PCR assay resulted in the detection of PyV infections in 11 animals from both species. Sequence analysis of the amplicons revealed considerable differences between the PyVs from Bornean and Sumatran orang-utans. The genome from two PyVs, one from each species, was therefore amplified and sequenced. Both viral genomes revealed a characteristic PyV architecture, but lacked an obvious agnogene. Neighbour-joining analysis positioned the viruses in a large cluster together with viruses from bats, bovines, rodents and several primate PyVs from chimpanzees, African green monkeys, squirrel monkeys and the human Merkel cell PyV.

Polyomaviruses (PyVs) are double-stranded DNA (dsDNA) viruses with a relatively small circular genome of approximately 5 kb. In humans, five different PyVs have been described to date. The two best-studied human viruses, JC virus (JCV) and BK virus (BKV), are associated with serious diseases in immunosuppressed people (Jiang et al., 2009). The others, WUPyV, KIPyV and Merkel cell PyV (McPyV), have been found in patients with respiratory tract infections (WU and KI viruses) and in patients with Merkel cell carcinoma, a rare aggressive skin cancer, respectively (Allander et al., 2007; Feng et al., 2008; Gaynor et al., 2007).

Another PyV, simian virus 40 (SV40), has also been detected in humans, but these data remain controversial (Martini et al., 2007). Its natural host is the Asian macaque (Koliaskina, 1963). SV40 may have entered the human population as a contaminant of polio vaccine batches that were prepared on primary macaque kidney cells (Sweet & Hilleman, 1960). It causes cell transformation in vitro, but also induces tumours in rodents. In its natural host, SV40 infection generally goes unnoticed, but when the host is immunosuppressed, SV40 can induce disease symptoms comparable to JCV-induced progressive multifocal leukoencephalopathy in humans (Axthelm et al., 2004; Chretien et al., 2000; Horvath et al., 1992; Simon et al., 1992, 1999).

In addition to SV40, other polyomaviruses have been described in non-human primates. African green monkeys [lymphotropic polyomavirus (LPV) or African green monkey PyV], chacma baboons (SA12 virus), squirrel monkeys [squirrel monkey PyV (SquiPyV)] and chimpanzees [chimpanzee PyV (ChPyV)] are all naturally infected with PyVs, but no association with any disease in the healthy, immunocompetent hosts has been reported until now (Johne et al., 2005; Valis et al., 1977; Verschoor et al., 2008a; zur Hausen & Gissmann, 1979). The finding of PyVs in apes (chimpanzees), Old World monkeys (baboons and African green monkeys) and New World monkeys (squirrel monkeys) suggests an extensive spread of PyVs in primates. As part of an ongoing project with the aim of investigating PyV infections in non-human primates, we report here the detection and genetic characterization of PyVs from Bornean and Sumatran orang-utans.

The GenBank/EMBL/DDBJ accession numbers for the VP1 sequences determined in this study are FN356900–FN356910.

Three supplementary figures showing additional sequence comparison data are available with the online version of this paper.

†Present address: Department of Physiological Chemistry and Centre for Biomedical Genetics, University Medical Center, Utrecht, The Netherlands.
‡Present address: Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.
§Present address: Department of Veterinary Medicine, University of Cambridge, Cambridge, UK.

Three supplementary figures showing additional sequence comparison data are available with the online version of this paper.

Three supplementary figures showing additional sequence comparison data are available with the online version of this paper.
A panel of 95 sera from orang-utans was screened in an ELISA using virus-like particles derived from SV40 VP1, the major capsid protein (Verschoor et al., 2008b). A substantial number of orang-utan sera were positive for antibodies that reacted with this protein, indicating that orang-utans can be infected with SV40 or a related polyomavirus. In total, 43% of the sera (41/95) reacted positively in the test.

Next, archival DNA specimens that had been extracted from frozen blood were examined for evidence of PyV infection. All blood samples from Bornean orang-utans (*Pongo pygmaeus*) had been acquired previously from wild-caught animals or animals housed at the Wanariset Rehabilitation Centre in Kalimantan, Indonesia, whilst those from the Sumatran species (*Pongo abelii*) had been collected from wild-caught and zoo animals. Using a broad-spectrum PCR assay targeted specifically to the VP1 gene (John et al., 2005), we analysed 79 DNA samples obtained from 71 Bornean orang-utans and eight Sumatran orang-utans. PyV-like sequences of approximately 250 bp were detected in eight Bornean orang-utans (11.3%) and in three Sumatran animals (37.5%). The PCR fragments were gel purified using a Zymoclean Gel DNA Recovery kit (Zymo Research), and sequence analysis was performed on the purified amplicons (Baseclear BV). Alignment of the sequences revealed a clear distinction between the viruses from the two species. Whilst the intraspecies nucleotide identity in the Sumatran and Bornean apes was high, it differed considerably between the two species (59–61% sequence identity) (see Supplementary Figure S1, available in JGV Online). This finding was confirmed by amplification and sequence analysis of five full-length VP1 sequences derived from two of the Sumatran orang-utans (SU77 and Pi) and three Bornean individuals (Cl, Ku and Bo). Viral VP1 genes from the Sumatran animals differed by 1 nt over a total length of 1143 nt (0.1%), whilst two VP1 variants were found in the Bornean orang-utans, differing by 10 nt from each other (0.9%). The deduced protein sequences were 100% similar within each orangutan species. A direct comparison of the Sumatran and Bornean VP1 genes again showed substantial differences. Alignment analysis revealed 63% nucleotide identity, whilst the inferred protein sequences differed by 27% (see Supplementary Figures S2 and S3, available in JGV Online).

From two animals representing each species, *P. pygmaeus* Ora-Bo and *P. abelii* Ora-Pi, the circular dsDNA genomes were amplified, essentially as described previously (Verschoor et al., 2008a). OraPyV-Bor was amplified with outer primer pair Bora-Fout (5'-ATACCTTACCCAGTTACTGCTTTGCTG-3') and Bora-Rout (5'-TCTGTATTTCATGCTTCCATCATTAG-3'), and the inner set Bora-Fin (5'-ATAGTGCCACCTGTACGGGACAC-3') and Bora-Rin (5'-CACCTATGTCACTGCTGACACATAC-3'). To amplify OraPyV-Sum, we used the outer primer set Ora-Fout (5'-AATTCGAAAAATCATCCGGTCACCGTCC-3') and Ora-Rout (5'-ATACCTGGCAAGCCCTATCATTGCTTTAG-3'), and the inner primers Ora-Fin (5'-AGTCTTATGCCTAAATGCAAGGTCAGCC-3') and Pora-Rin (5'-CTGCACAGCCTGTAGAAAAAGTCCCATCAC-3'). The 5 kb subgenomic fragments were cloned in the pJET1.2 vector (CloneJET PCR Cloning kit; Fermentas) and sequenced using a primer-walking strategy. Finally, the VP1 sequences and the 5 kb PCR fragments were combined to construct complete PyV genomes.

The viral differences considered in length. The virus from Ora-Bo (OraPyV-Bor) was 5168 bp, whilst the Sumatran isolate (OraPyV-Sum) was 5358 bp, which is one of the longest primate PyV genomes identified to date. The genome architecture of the viruses was characteristic of members of the family *Polyomaviridae*, with an early region encoding the small T (STAg) and large T (LTAg) antigens, and a late region with genes for the VP1, VP2 and VP3 structural proteins. Both viruses lacked an agnogene open reading frame (ORF) located 5' to the VP2 ORF, which has been reported for several primate PyVs (Khalili et al., 2005; Verschoor et al., 2008a).

The STAg and LTAg were transcribed from a single mRNA. The STAg ORF was relatively short and encoded proteins of 197 and 194 aa for OraPyV-Bor and OraPyV-Sum, respectively. The LTAg was transcribed from a spliced mRNA. Putative splice-donor and splice-acceptor sites were calculated using SpliceView (http://zeus2.itb.cnr.it/~webgene/wwwspliceview.html) (Rogozin & Milanesi, 1997) and GeneSplicer (http://www.ccb.cbq.edu/software/GeneSplicer/gene_spl.shtml) (Pertea et al., 2001) programs, and the most likely sites were determined on the basis of comparison with known PyV LTAg. The putative LTAg varied significantly in size (693 and 735 aa for Bor and Sum, respectively) and contained multiple domains common to other PyV LTAg that are involved in viral DNA replication and host cell transformation (Fig. 1) (Sullivan et al., 2000). Both LTAg possessed the conserved J domain (HPDKGG), which is important for efficient DNA replication and transformation. A pRB tumour suppressor-binding motif (LXCXE), which is critical for DNA replication, was also present, but in the case of OraPyV-Sum, the leucine (L) residue was changed to an isoleucine (I). A zinc-binding motif (CXXCXXHXXHXHXH) and an ATP-binding motif (GXGK) were also found in both LTAg. As in other PyV LTAg, a conserved binding motif similar to conserved region 1 of the adenovirus E1A protein [(E/D)X3LX(E/D)LX2(L/I)] was found in the N-terminal region of the LTAg (Pipas, 1992). As with the pRB motif, this domain was fully conserved in OraPyV-Bor, but in OraPyV-Sum there was a glutamine (Q) and a valine (V) instead of a glutamic acid (E) or an aspartic acid (D). Finally, a putative nuclear localization sequence (NLS) was also present in both proteins.

In contrast to most PyVs, except for mouse PyV and SquiPyV, the VP2/3 ORFs terminated immediately after the NLS, resulting in relatively short proteins. The VP2 and VP3 antigens of OraPyV-Bor were 311 and 190 aa in length, and those of OraPyV-Sum were 317 and 202 aa, respectively. Alignment of the VP1 proteins with other published PyV
**Fig. 1.** Alignment of LTAg of OraPyV-Bor and OraPyV-Sum. Grey-shaded boxes indicate sequence similarities and identities between the proteins. Open boxes indicate conserved domains. Gaps are indicated by dashes.
**Fig. 2.** Comparison of VP1 proteins from orang-utans with VP1 proteins from other mammalian and avian polyomaviruses. Grey-shaded boxes indicate the areas of more than 70% amino acid identity. Underlined regions indicate loops in the VP1 proteins. The GenBank accession numbers of the viruses used are NC_001515 (MuPyV), NC_001663 (HaPyV), NC_010277 (McPyV), M30540 (LPV), AY691168 (ChPyV), NC_001442 (BoPyV), NC_009951 (SquiPyV), NC_011310 (MyoPyV), NC_007922 (CrPyV), NC_004800 (GoPyV), NC_007923 (FiPyV), AB453166 (BFDPyV), NC_001669 (SV40), NC_001538 (BKV), EF127906 (KIPyV) and EF444549 (WUPyV) (see Fig. 3 for virus hosts). Gaps are indicated by dashes.
VP1 proteins revealed various regions of relative conservation, alternating with highly variable regions (Fig. 2). Most variation was located in the BC, DE, EF and HI loops that protrude from the VP1 protein structure, mediate receptor-binding and contain principal antigenic domains (Li et al., 2000; Liddington et al., 1991; Murata et al., 2008; Neu et al., 2008; Stehle & Harrison, 1997; Stehle et al., 1996).

The excessive variation in the nucleotide sequences of VP1 precluded accurate alignment and phylogenetic analysis. We therefore used the protein alignment to evaluate the genetic relationships of the orang-utan PyVs with the other PyVs (Fig. 3). The orang-utan viruses clustered with a heterogeneous group of viruses that have been isolated from bats, cows and rodents and a variety of primate PyVs from chimpanzees, African green monkeys, New World squirrel monkeys and humans (bootstrap value of 76 %). The viruses in this group are well-separated from the avian PyVs, WUPyV and KIPyV, and from a tight cluster formed by JCV, BKV, SA12 and SV40.

Orang-utans from the islands of Borneo and Sumatra diverged approximately 1.1 million years ago, and they have been physically separated for 10 000–15 000 years (Warren et al., 2001). From their position in the tree, it becomes less evident that the Bornean and Sumatran PyVs have evolved from the same ancestral virus. Instead, a scenario describing two independent transmissions from as yet unknown hosts appears more likely. In view of this, it is interesting to note that PyVs have been characterized from rodents and bats. Both groups of animals are found worldwide, together representing over 60 % of mammal species, and are notorious for their role as vectors for zoonotic transmission of viruses. However, as our current knowledge about PyVs from primates, and also other mammals, is far from complete, more data concerning the prevalence of PyVs in other species will be needed to address this issue properly.

**Fig. 3.** Phylogenetic tree constructed using the VP1 protein of avian and mammalian PyVs. The orang-utan PyVs are in bold. Mammalian viruses are indicated by grey shading. Sequence alignments were made using MacVector version 10.6. The GapStreeze program (Los Alamos HIV Sequence Database; http://www.hiv.lanl.gov/content/sequence/GAPSTREEZE/gap.html) was used to remove columns with a gap tolerance of 20 %. Phylogenetic analysis was performed by the neighbour-joining method using the JTT matrix model as implemented in MEGA version 4 (Tamura et al., 2007). Bootstrap values (as percentages of 1000 resamplings) are indicated. Bar, 0.2 amino acid residue replacements per site.

**Acknowledgements**

This study was supported by the European Community Research Infrastructures Program grant RII3-CT-2006-026155 ‘European Primate Network: Specialized Infrastructures and Procedures for Biological and Biomedical Research’ (EUPRIM-NET).

**References**


