Characterization of novel polyomaviruses from Bornean and Sumatran orangutans

Marlous J. Groenewoud¹,⁴, Zahra Fagrouch¹, Sabine van Gessel¹,⁵, Henk Niphuis¹, Kestutis Sasnauskas², Kristin S. Warren³, Jonathan L. Heeney¹,⁶, 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
⁴Present address: Department of Physiological Chemistry and Centre for Biomedical Genetics, University Medical Center, Utrecht, The Netherlands
⁵Present address: Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
⁶Present address: Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom

*Corresponding author.
The Netherlands

Tel: +31-15-2842661
Fax: +31-15-2842601
E-mail address: verschoor@bprc.nl

Running title: novel polyomaviruses from orangutans

Word count Summary: 124
Word count Text (including legends): 1799

Number of figures: 3
Number of supplementary figures: 3
ABSTRACT

Serological screening of sera from orangutans demonstrated a high percentage of sera that cross-reacted with SV40 polyomavirus antigens. Analysis of archival DNA samples from seventy-one Bornean and eight Sumatran orangutans with a broad-spectrum PCR assay resulted in the detection of polyomavirus infections in eleven animals from both species. Sequence analysis of the amplicons revealed considerable differences between the polyomaviruses from Bornean and Sumatran orangutans. The genome from two polyomaviruses, one from each species, was therefore amplified and sequenced. Both viral genomes revealed a characteristic polyomavirus architecture, but lacked an obvious agnogene. Neighbor-joining analysis positioned the viruses in a larger cluster together with viruses from bats, bovines, rodents, and several primate polyomaviruses from chimpanzees, African green monkeys, squirrel monkeys, and the human Merkel cell polyomavirus.
Polyomaviruses (PyV) are dsDNA viruses with a relatively small circular genome of approximately 5 kb. In humans, five different PyV have been described until now. The two best-studied human viruses, JCV and BKV, are associated with serious diseases in immunosuppressed persons (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), or in patients with Merkel cell carcinoma (MCC), a rare aggressive skin cancer, respectively (Allander et al., 2007; Feng et al., 2008; Gaynor et al., 2007).

Another polyomavirus, simian virus 40 (SV40), has also been detected in humans, but these data remain controversial (Martini et al., 2007). Its natural hosts are Asian macaques (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 tumors 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 (PML) in humans (Axthelm et al., 2004; Chretien et al., 2000; Horvath et al., 1992; Simon et al., 1992; Simon et al., 1999).

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

A panel of 95 sera from orangutans was screened in an ELISA using virus-like particles derived from SV40 VP1, which is the major capsid protein (Verschoor et al., 2008b). A substantial number of orangutan sera were positive for antibodies that reacted with this protein, indicating that orangutans can be infected with SV40, or a related polyomavirus. In total, 43% of the sera (41 out of 95) reacted positively in the test.

Next, archival DNA specimens which had been extracted from frozen blood were examined for evidence of polyomavirus infections. All blood samples from Bornean orangutans (Pongo pygmaeus) had previously been acquired from wild-caught animals, or animals housed at the Wanariset orangutan rehabilitation centre in Kalimantan, Indonesia, while those from the Sumatran species (Pongo abelii) had been collected from wild-caught and zoo animals. With a broad-spectrum PCR assay, targeted specifically to the VP1 gene (Johne et al., 2005), we analyzed 79 DNA samples obtained from 71 Bornean orangutans and eight Sumatran orangutans. PyV-like sequences of approximately 250 bp were detected in eight Bornean orangutans (11.4%), and in three Sumatran animals (37.5%). The PCR fragments were gel-purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corp, Orange, USA), and sequence analysis was performed directly on the purified amplicons (Baseclear BV, Leiden, The Netherlands). Alignment of the sequences revealed a clear distinction between viruses from both species. While the intraspecies nucleotide identity in the Sumatran and Bornean apes appeared limited, the sequences
derived from each species differed considerably (59-61% sequence identity) (Supplementary figure 1). This finding was confirmed by the amplification and sequence analysis of five full-length VP1 sequences derived from two Sumatran orangutans (SU77 and Pi), and three Bornean individuals (Cl, Ku, and Bo). Viral VP1 genes from the Sumatran animals differed 1 nucleotide on a total length of 1,143 nt (0.17%), while two VP1 variants were found in the Bornean orangutan, differing 10 nt from each other (0.9%). The deduced protein sequences were 100% similar within each orangutan species. The direct comparison of the Sumatran and Bornean VP1 genes again showed substantial differences. Alignment analysis revealed 63% nucleotide identity, while the inferred protein sequences were 27% different (Supplementary figures 2 and 3). Sequences have been submitted to EMBL Database under accession numbers FN356900 – FN356910.

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’- AATCCTTACCCAGTATTCTTTGCTG-3’ and Bora-Rout 5’- TCTGTATTTTCATGCTTCCATTAG, and the inner set Bora-Fin 5’ATGATGCCACCTGTACAGGGACAAC-3’ and Bora-Rin 5’- CCACATATCGGTACATGACACATA-3’. To amplify OraPyV-Sum we used to outer primer set Pora-Fout 5’-AATGTCAGAAATCCATACCCTGTAACCTCC-3’ and Pora-Rout 5’-ATACCTTGGCAGACCTCTATATTGCTTAG-3’, and the inner primers Pora-Fin 5’-AGTCTTATGCTTAAATGCAAGGTCAGCC-3’ and Pora-Rin 5’-CTGCACAGCTTAGAAAAATGCCATCACC-3’. The 5 kb subgenome fragments were cloned in the pJET1.2 vector (CloneJET™ PCR Cloning Kit, Fermentas, Germany), and sequenced by using a primer-walking strategy. Finally, the
VP1 sequences and the 5 kb PCR fragments were combined to construct complete PyV genomes.

The viral genomes differed considerably in length. The virus from Ora-Bo (OraPyV-Bor) measured 5,168 bp, while the Sumatran isolate (OraPyV-Sum) was 5,358 bp in length, which is one of the longest primate polyomavirus genomes identified thus far. The genome architecture of the viruses is characteristic of that of the *Polyomaviridae* with an early region encoding the small T (stAg) and large T antigens (LTAg), and a late region with genes for the VP1, VP2, and VP3 structural proteins. Both viruses lack an agnogene open reading frame (orf) located 5' to the VP2 orf, which has been reported for several primate PyV (Khalili *et al.*, 2005; Verschoor *et al.*, 2008a).

The stAg and LTAg are transcribed from a single messenger RNA (mRNA). The stAg orf is relatively short and encodes proteins of 197 and 194 aa. in length for OraPyV-Bor and -Sum, respectively. The LTAgS are transcribed from a spliced mRNA. Putative splice-donor and -acceptor sites were calculated using SpliceView ([http://zeus2.itb.cnr.it/~webgene/wwwspliceview.html](http://zeus2.itb.cnr.it/~webgene/wwwspliceview.html)) (Rogozin & Milanesi, 1997) and GeneSplicer programs ([http://www.ccbcb.umd.edu/software/GeneSplicer/gene_spl.shtml](http://www.ccbcb.umd.edu/software/GeneSplicer/gene_spl.shtml)) (Pertea *et al.*, 2001), and the most likely sites were determined on the basis of comparison with known PyV LTAgS. The putative LTAgS vary significantly in size (693 and 735 aa. for Bor and Sum, respectively), and contain multiple domains common to other PyV LTAgS that are involved in viral DNA replication and host cell transformation (Figure 1.) (Sullivan *et al.*, 2000). Both LTAgS possess the conserved J domain (HPDKGG), which is important for efficient DNA replication and transformation. A pRB tumor suppressor-binding motif (LXCXE), which is critical for DNA replication is also
present, but in case of OraPyV-Sum the leucine (L) residue has been changed into an isoleucine (I). A zinc-binding motif (CX₂CX₇HX₈HX₉H) and an ATP-binding motif (GX₄GK) are also found in both LTags. Like in other polyomavirus LTags, a conserved binding motif similar to the conserved region 1 of the adenovirus E1A protein ((E/D)X₃LX(E/D)LX₂(L/I)) is found in the N-terminal part of the LTags (Pipas, 1992). As for the pRB motif, this domain is fully conserved in OraPyV-Bor, but in -Sum there are a glutamine (Q) and a valine (V) instead of a glutamic acid (E) or an aspartic acid (D). Finally, putative nuclear localization sequences (NLS) are also present in both proteins.

In contrast to most polyomaviruses, except for the mouse and squirrel monkey PyV, the VP2/3 orfs terminate immediately after the nuclear transport signal (NLS), resulting in relatively short proteins. The VP2 and VP3 antigens of OraPyV-Bor are 311 and 190 aa. in length, and those of -Sum 317 and 202 aa. Alignment of the VP1 proteins with other published polyomavirus VP1 proteins revealed various regions of relative conservation, alternated by highly variable regions (Figure 2). Most variation is located in the BC-, DE-, EF-, and HI-loops that protrude from the VP1 protein structure and that mediate receptor-binding, and that contain principal antigenic domains (Li et al., 2000; Liddington et al., 1991; Murata et al., 2008; Neu et al., 2008; Stehle et al., 1996; Stehle & Harrison, 1997).

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

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

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

Legends to Figures

Figure 1
Alignment of large T antigens of OraPyV-Bor and OraPyV-Sum. Grey shaded boxes indicate sequence similarities and identities between the proteins. The open boxes designate conserved domains.

Figure 2
Comparison of VP1 proteins from orangutans with VP1 from other mammalian and avian polyomaviruses. Grey-shaded boxes indicate the areas of more than 70% amino acid identity. Underlined regions designate 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_001699 (JCV), AY614708 (SA12), NC_001538 (BKV), EF127906 (KIPyV), and EF444549 (WUPyV). OraPyV-Bor and OraPyV-Sum genomes have been deposited under accession numbers FN356900 and FN356901.

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

Allander, T., Andreasson, K., Gupta, S., Bjerkner, A., Bogdanovic, G., Persson,

pathologic manifestations of primary polyomavirus infection in
immunosuppressed rhesus monkeys. *Journal of neuropathology and
experimental neurology* **63**, 750-758.

Chretien, F., Boche, D., Lorin de la Grandmaison, G., Ereau, T., Mikol, J.,
leukoencephalopathy and oligodendroglioma in a monkey co-infected by
simian immunodeficiency virus and simian virus 40. *Acta neuropathologica*
**100**, 332-336.


Gaynor, A. M., Nissen, M. D., Whiley, D. M., Mackay, I. M., Lambert, S. B., Wu,
Identification of a novel polyomavirus from patients with acute respiratory
Horvath, C. J., Simon, M. A., Bergsagel, D. J., Pauley, D. R., King, N. W.,
rhesus monkeys with simian acquired immunodeficiency syndrome. The
American journal of pathology 140, 1431-1440.

Jiang, M., Abend, J. R., Johnson, S. F. & Imperiale, M. J. (2009). The role of

detected in the feces of a chimpanzee by nested broad-spectrum PCR. Journal
of virology 79, 3883-3887.

agnoprotein of polyomaviruses: a multifunctional auxiliary protein. J Cell
Physiol 204, 1-7.

Caused by Ob40 (Sv40) Virus in Macaca Rhesus Monkeys.]. Voprosy
virusologii 29, 450-452.

Li, P. P., Nakanishi, A., Tran, M. A., Salazar, A. M., Liddington, R. C. &
Kasamatsu, H. (2000). Role of simian virus 40 Vp1 cysteines in virion

Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L. & Harrison, S.
284.

Martini, F., Corallini, A., Balatti, V., Sabbioni, S., Pancaldi, C. & Tognon, M.


J domain

pRb-binding motif

E1A-binding motif

Zinc-binding domain

ATP-binding motif

NLS