MURDOCH RESEARCH REPOSITORY

http://researchrepository.murdoch.edu.au

This is the author's final version of the work, as accepted for publication following peer review but without the publisher's layout or pagination.

Groenewoud, M.J., Fagrouch, Z., van Gessel, S., Niphuis, H., Bulavaite, A., Warren, K.S. , Heeney, J. L. and Verschoor, E. J. (2010) Characterization of novel polyomaviruses from Bornean and Sumatran orang-utans. Journal of General Virology, 91 (3). pp. 653-658.

http://researchrepository.murdoch.edu.au/3637

Copyright © 2010 SGM
It is posted here for your personal use. No further distribution is permitted.

JGV Papers in Press. Published November 18, 2009 as doi:10.1099/vir.0.017673-0

1 Characterization of novel polyomaviruses from Bornean and
2 Sumatran orangutans

3

4 Marlous J. Groenewoud[1,4], Zahra Fagrouch[1], Sabine van Gessel[1,5], Henk Niphuis[1],
5 Kestutis Sasnauskas[2], Kristin S. Warren[3], Jonathan L. Heeney[1,6], and Ernst J.
6 Verschoor[1,*]

7

8 [1] Department of Virology, Biomedical Primate Research Centre, Rijswijk,
9 The Netherlands
10 [2] Institute of Biotechnology, Vilnius, Lithuania
11 [3] School of Veterinary and Biomedical Sciences, Murdoch University, Perth,
12 Western Australia
13 [4] Present address: Department of Physiological Chemistry and Centre for Biomedical
14 Genetics, University Medical Center, Utrecht, The Netherlands
15 [5] Present address: Department of Human Genetics, Radboud University Nijmegen
16 Medical Centre, Nijmegen, The Netherlands
17 [6] Present address: Department of Veterinary Medicine, University of Cambridge,
18 Cambridge, United Kingdom

19

20 [*]Corresponding author.
21 E. J. Verschoor
22 Department of Virology
23 Biomedical Primate Research Centre
24 Lange Kleiweg 139
25 2288GJ Rijswijk

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.

1 Polyomaviruses (PyV) are dsDNA viruses with a relatively small circular
2 genome of approximately 5 kb. In humans, five different PyV have been described
3 until now. The two best-studied human viruses, JCV and BKV, are associated with
4 serious diseases in immunosuppressed persons (Jiang *et al.*, 2009). The others,
5 WUPyV, KIPyV , and Merkel cell PyV (McPyV), have been found in patients with
6 respiratory tract infections (WU and KI), or in patients with Merkel cell carcinoma
7 (MCC), a rare aggressive skin cancer, respectively(Allander *et al.*, 2007; Feng *et al.*,
8 2008; Gaynor *et al.*, 2007).

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

19 In addition to SV40, other polyomaviruses have been described in nonhuman
20 primates. African green monkeys (LPV or AgmPyV), chacma baboons (SA12),
21 squirrel monkeys (SquiPyV), and chimpanzees (ChPyV) are all naturally infected
22 with PyV, but no association with any disease in the healthy, immunocompetent host
23 has been reported until now (Johne *et al.*, 2005; Valis *et al.*, 1977; Verschoor *et al.*,
24 2008a; zur Hausen & Gissmann, 1979). The finding of polyomaviruses in apes
25 (chimpanzee), Old World monkeys (baboons, African green monkeys), and New

1 World monkeys (squirrel monkey) suggests an extensive spread of PyV in primates.
2 As part of an ongoing project with the aim to investigate polyomavirus infections in
3 nonhuman primates, we here report the detection and genetic characterization of
4 polyomaviruses from Bornean and Sumatran orangutans.

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

11 Next, archival DNA specimens which had been extracted from frozen blood
12 were examined for evidence of polyomavirus infections. All blood samples from
13 Bornean orangutans (*Pongo pygmaeus*) had previously been acquired from wild-
14 caught animals, or animals housed at the Wanariset orangutan rehabilitation centre in
15 Kalimantan, Indonesia, while those from the Sumatran species (*Pongo abelii*) had
16 been collected from wild-caught and zoo animals. With a broad-spectrum PCR assay,
17 targeted specifically to the VP1 gene(Johne *et al.*, 2005), we analyzed 79 DNA
18 samples obtained from 71 Bornean orangutans and eight Sumatran orangutans. PyV-
19 like sequences of approximately 250 bp were detected in eight Bornean orangutans
20 (11.4%), and in three Sumatran animals (37.5%). The PCR fragments were gel-
21 purified using the Zymoclean™ Gel DNA Recovery Kit (Zymo Research Corp,
22 Orange, USA), and sequence analysis was performed directly on the purified
23 amplicons (Baseclear BV, Leiden, The Netherlands). Alignment of the sequences
24 revealed a clear distinction between viruses from both species. While the intraspecies
25 nucleotide identity in the Sumatran and Bornean apes appeared limited, the sequences

1 derived from each species differed considerably (59-61% sequence identity)
2 (Supplementary figure 1). This finding was confirmed by the amplification and
3 sequence analysis of five full-length VP1 sequences derived from two Sumatran
4 orangutans (SU77 and Pi), and three Bornean individuals (Cl, Ku, and Bo). Viral VP1
5 genes from the Sumatran animals differed 1 nucleotide on a total length of 1,143 nt
6 (0.17%), while two VP1 variants were found in the Bornean orangutan, differing 10 nt
7 from each other (0.9%). The deduced protein sequences were 100 % similar within
8 each orangutan species. The direct comparison of the Sumatran and Bornean VP1
9 genes again showed substantial differences. Alignment analysis revealed 63%
10 nucleotide identity, while the inferred protein sequences were 27% different
11 (Supplementary figures 2 and 3). Sequences have been submitted to EMBL Database
12 under accession numbers FN356900 – FN356910.

13 From two animals representing each species, *P. pygmaeus* Ora-Bo and *P.*
14 *abelii* Ora-Pi, the circular dsDNA genomes were amplified, essentially as described
15 previously (Verschoor *et al.*, 2008a). OraPyV-Bor was amplified with outer primer
16 pair Bora-Fout 5'- AATCCTTACCCAGTTACTTCTTTGCTG-3' and Bora-Rout 5'-
17 TCTGTATTTCATGCTTCCATCATTAG, and the inner set Bora-Fin
18 5'ATGATGCCACCTGTACAGGGACAAC-3' and Bora-Rin 5'-
19 CCACTATGTCACATGCTGACACATAC-3'. To amplify OraPyV-Sum we used to
20 outer primer set Pora-Fout 5'-AATGTCAGAAATCCATACCCTGTAACCTCC-3'
21 and Pora-Rout 5'-ATACCTTGGCAGACCTCTATATTGCTTAG-3', and the inner
22 primers Pora-Fin 5'-AGTCTTATGCCTAAAATGCAAGGTCAGCC-3' and Pora-
23 Rin 5'-CTGCACAGCTTAGAAAAAGTCCATCACC-3'. The 5 kb subgenome
24 fragments were cloned in the pJET1.2 vector (CloneJET™ PCR Cloning Kit,
25 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) (Rogozin & Milanesi, 1997) and GeneSplicer programs (http://www.cbcb.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

1 present, but in case of OraPyV-Sum the leucine (L) residue has been changed into an
2 isoleucine (I). A zinc- binding motif ($CX_2CX_7HX_3HX_2H$) and an ATP-binding motif
3 (GX_4GK) are also found in both LTags. Like in other polyomavirus LTags, a
4 conserved binding motif similar to the conserved region 1 of the adenovirus E1A
5 protein ($((E/D)X_3LX(E/D)LX_2(L/I))$) is found in the N-terminal part of the LTags
6 (Pipas, 1992). As for the pRB motif, this domain is fully conserved in OraPyV-Bor,
7 but in -Sum there are a glutamine (Q) and a valine (V) instead of a glutamic acid (E)
8 or an aspartic acid (D). Finally, putative nuclear localization sequences (NLS) are also
9 present in both proteins.

10 In contrast to most polyomaviruses, except for the mouse and squirrel monkey
11 PyV, the VP2/3 orfs terminate immediately after the nuclear transport signal (NLS),
12 resulting in relatively short proteins. The VP2 and VP3 antigens of OraPyV-Bor are
13 311 and 190 aa. in length, and those of -Sum 317 and 202 aa. Alignment of the VP1
14 proteins with other published polyomavirus VP1 proteins revealed various regions of
15 relative conservation, alternated by highly variable regions (Figure 2). Most variation
16 is located in the BC-, DE-, EF-, and HI-loops that protrude from the VP1 protein
17 structure and that mediate receptor-binding, and that contain principal antigenic
18 domains (Li *et al.*, 2000; Liddington *et al.*, 1991; Murata *et al.*, 2008; Neu *et al.*,
19 2008; Stehle *et al.*, 1996; Stehle & Harrison, 1997).

20 The excessive variation in the nucleotide sequences of VP1 precludes accurate
21 alignment and phyogenetic analysis. We therefore used the protein alignment to
22 evaluate the genetic relationships of the orangutan polyomaviruses with the other PyV
23 (Figure 3). The orangutan viruses cluster with a heterogeneous group of viruses that
24 have been isolated from bats, bovines, and rodents, and a variety of primate PyV from
25 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).

1 **References**

2 **Legends to Figures**

3

4 **Figure 1**

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

8

9 **Figure 2**

10 Comparison of VP1 proteins from orangutans with VP1 from other mammalian and
11 avian polyomaviruses. Grey-shaded boxes indicate the areas of more than 70% amino
12 acid identity. Underlined regions designate loops in the VP1 proteins. The GenBank
13 accession numbers of the viruses used are: NC_001515 (MuPyV), NC_001663
14 (HaPyV), NC_010277 (McPyV), M30540 (LPV), AY691168 (ChPyV), NC_001442
15 (BoPyV), NC_009951 (SquiPyV), NC_011310 (MyoPyV), NC_007922 (CrPyV),
16 NC_004800 (GoPyV), NC_007923 (FiPyV), AB453166 (BFDPyV), NC_001669
17 (SV40), NC_001699 (JCV), AY614708 (SA12), NC_001538 (BKV), EF127906
18 (KIPyV), and EF444549 (WUPyV). OraPyV-Bor and OraPyV-Sum genomes have
19 been deposited under accession numbers FN356900 and FN356901.

20

21 **Figure 3**

22 Phylogenetic tree constructed using the VP1 protein of avian and mammalian
23 polyomaviruses. Orangutan PyV are in bold. Mammalian viruses are indicated by
24 grey shading. Sequence alignments were made by using MacVector version 10.6. The
25 GapStreeze program (Los Alamos HIV Sequence Database; http://www.hiv.

1 lanl.gov/content/hiv-db/GAPSTREEZE/gap.html) was used to remove columns with a
2 gap tolerance of 20%. Phylogenetic analysis was performed by the neighbor-joining
3 method using the JTT matrix model as implemented in MEGA version 4 (Tamura *et*
4 *al.*, 2007). Bootstrap values (as % of 1000 re-samplings) are indicated. Bar, 0.2 amino
5 acid residue replacements per site.

6
7

8 **Allander, T., Andreasson, K., Gupta, S., Bjerkner, A., Bogdanovic, G., Persson,**
9 **M. A. A., Dalianis, T., Ramqvist, T. & Andersson, B. (2007).** Identification
10 of a third human polyomavirus. *Journal of virology* **81**, 4130-4136.

11 **Axthelm, M. K., Koralnik, I. J., Dang, X., Wuthrich, C., Rohne, D., Stillman, I.**
12 **E. & Letvin, N. L. (2004).** Meningoencephalitis and demyelination are
13 pathologic manifestations of primary polyomavirus infection in
14 immunosuppressed rhesus monkeys. *Journal of neuropathology and*
15 *experimental neurology* **63**, 750-758.

16 **Chretien, F., Boche, D., Lorin de la Grandmaison, G., Ereau, T., Mikol, J.,**
17 **Hurtrel, M., Hurtrel, B. & Gray, F. (2000).** Progressive multifocal
18 leukoencephalopathy and oligodendroglioma in a monkey co-infected by
19 simian immunodeficiency virus and simian virus 40. *Acta neuropathologica*
20 **100**, 332-336.

21 **Feng, H., Shuda, M., Chang, Y. & Moore, P. S. (2008).** Clonal integration of a
22 polyomavirus in human Merkel cell carcinoma. *Science* **319**, 1096-1100.

23 **Gaynor, A. M., Nissen, M. D., Whiley, D. M., Mackay, I. M., Lambert, S. B., Wu,**
24 **G., Brennan, D. C., Storch, G. A., Sloots, T. P. & Wang, D. (2007).**
25 Identification of a novel polyomavirus from patients with acute respiratory
26 tract infections. *PLoS pathogens* **3**, e64.

Horvath, C. J., Simon, M. A., Bergsagel, D. J., Pauley, D. R., King, N. W., Garcea, R. L. & Ringler, D. J. (1992). Simian virus 40-induced disease in 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 polyomaviruses in human disease. *Virology* **384**, 266-273.

Johne, R., Enderlein, D., Nieper, H. & Muller, H. (2005). Novel polyomavirus detected in the feces of a chimpanzee by nested broad-spectrum PCR. *Journal of virology* **79**, 3883-3887.

Khalili, K., White, M. K., Sawa, H., Nagashima, K. & Safak, M. (2005). The agnoprotein of polyomaviruses: a multifunctional auxiliary protein. *J Cell Physiol* **204**, 1-7.

Koliaskina, G. I. (1963). [Virological and Serological Data on Latent Infection 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 infectivity. *Journal of virology* **74**, 11388-11393.

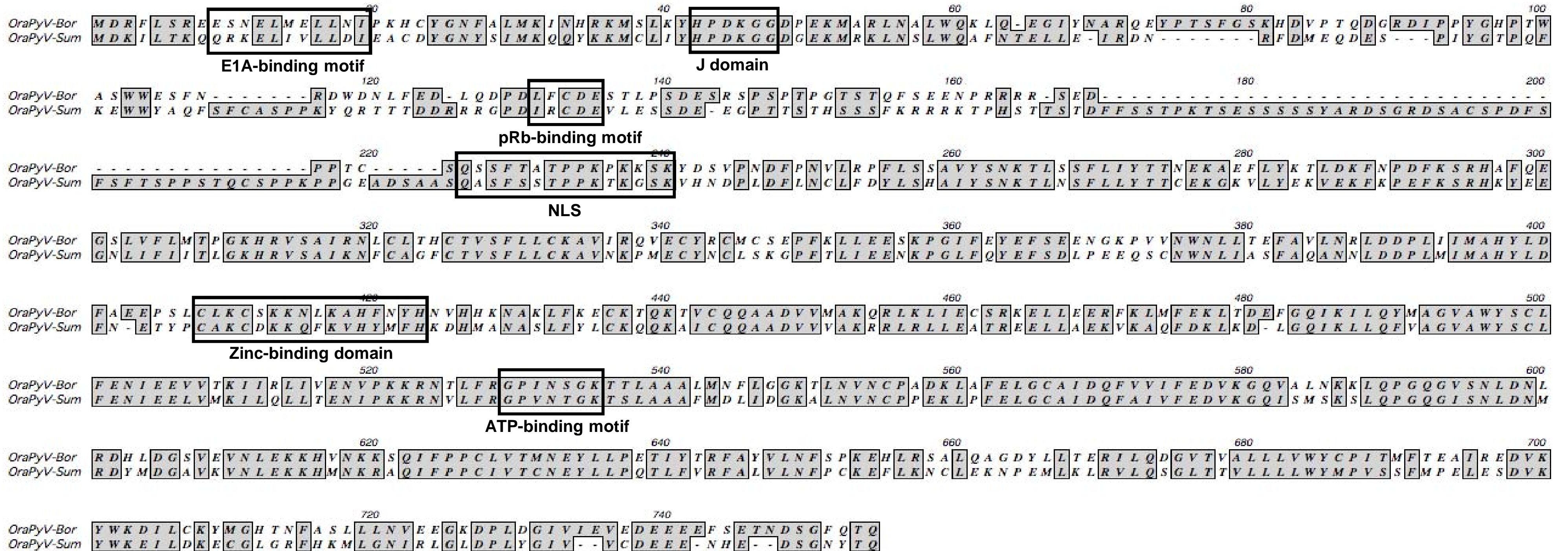
Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L. & Harrison, S. C. (1991). Structure of simian virus 40 at 3.8-A resolution. *Nature* **354**, 278-284.

Martini, F., Corallini, A., Balatti, V., Sabbioni, S., Pancaldi, C. & Tognon, M. (2007). Simian virus 40 in humans. *Infectious agents and cancer* **2**, 13.

1 **Murata, H., Teferedegne, B., Sheng, L., Lewis, A. M., Jr. & Peden, K. (2008).**
2 Identification of a neutralization epitope in the VP1 capsid protein of SV40.
3 *Virology* **381**, 116-122.
4 **Neu, U., Woellner, K., Gauglitz, G. & Stehle, T. (2008).** Structural basis of GM1
5 ganglioside recognition by simian virus 40. *Proc Natl Acad Sci U S A* **105**,
6 5219-5224.
7 **Pertea, M., Lin, X. & Salzberg, S. L. (2001).** GeneSplicer: a new computational
8 method for splice site prediction. *Nucleic Acids Res* **29**, 1185-1190.
9 **Pipas, J. M. (1992).** Common and unique features of T antigens encoded by the
10 polyomavirus group. *Journal of virology* **66**, 3979-3985.
11 **Rogozin, I. B. & Milanesi, L. (1997).** Analysis of donor splice sites in different
12 eukaryotic organisms. *Journal of molecular evolution* **45**, 50-59.
13 **Simon, M. A., Chalifoux, L. V. & Ringler, D. J. (1992).** Pathologic features of SIV-
14 induced disease and the association of macrophage infection with disease
15 evolution. *AIDS Res Hum Retroviruses* **8**, 327-337.
16 **Simon, M. A., Ilyinskii, P. O., Baskin, G. B., Knight, H. Y., Pauley, D. R. &**
17 **Lackner, A. A. (1999).** Association of simian virus 40 with a central nervous
18 system lesion distinct from progressive multifocal leukoencephalopathy in
19 macaques with AIDS. *The American journal of pathology* **154**, 437-446.
20 **Stehle, T., Gamblin, S. J., Yan, Y. & Harrison, S. C. (1996).** The structure of
21 simian virus 40 refined at 3.1 A resolution. *Structure* **4**, 165-182.
22 **Stehle, T. & Harrison, S. C. (1997).** High-resolution structure of a polyomavirus
23 VP1-oligosaccharide complex: implications for assembly and receptor
24 binding. *Embo J* **16**, 5139-5148.

1 **Sullivan, C. S., Cantalupo, P. & Pipas, J. M. (2000).** The molecular chaperone
2 activity of simian virus 40 large T antigen is required to disrupt Rb-E2F
3 family complexes by an ATP-dependent mechanism. *Mol Cell Biol* **20**, 6233-
4 6243.
5 **Sweet, B. H. & Hilleman, M. R. (1960).** The vacuolating virus, S.V. 40. *Proc Soc*
6 *Exp Biol Med* **105**, 420-427.
7 **Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007).** MEGA4: Molecular
8 Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular*
9 *biology and evolution* **24**, 1596-1599.
10 **Valis, J. D., Newell, N., Reissig, M., Malherbe, H., Kaschula, V. R. & Shah, K. V.**
11 **(1977).** Characterization of SA12 as a simian virus 40-related papovavirus of
12 chacma baboons. *Infect Immun* **18**, 247-252.
13 **Verschoor, E. J., Groenewoud, M. J., Fagrouch, Z., Kewalapat, A., van Gessel,**
14 **S., Kik, M. J. & Heeney, J. L. (2008a).** Molecular characterization of the first
15 polyomavirus from a New World primate: squirrel monkey polyomavirus. *J*
16 *Gen Virol* **89**, 130-137.
17 **Verschoor, E. J., Niphuis, H., Fagrouch, Z., Christian, P., Sasnauskas, K.,**
18 **Pizarro, M. C. & Heeney, J. L. (2008b).** Seroprevalence of SV40-like
19 polyomavirus infections in captive and free-ranging macaque species. *J Med*
20 *Primatol* **37**, 196-201.
21 **Warren, K. S., Verschoor, E. J., Langenhuijzen, S., Heriyanto, Swan, R. A.,**
22 **Vigilant, L. & Heeney, J. L. (2001).** Speciation and intrasubspecific variation
23 of Bornean orangutans, Pongo pygmaeus pygmaeus. *Molecular biology and*
24 *evolution* **18**, 472-480.

1 **zur Hausen, H. & Gissmann, L. (1979).** Lymphotropic papovaviruses isolated from
2 African green monkey and human cells. *Med Microbiol Immunol* **167**, 137-
3 153.
4
5



OraPyV-Bor
OraPyV-Sum
E1A-binding motif
J domain
pRb-binding motif
NLS
Zinc-binding domain
ATP-binding motif

```
                   20                  40                  60                  80                  100                 120                 140
OraPyV-Bor   MAPKRKGEGCR---------KT--CP--------QPKPVPKLIVKGNVQVLGLK-TGPDSVTQIECFLNPRMGEND------EHKGWDMDTVKKLLLLTA-WQ----MTTYRKEIPTYSTARIPLPMLNEDLTCNTLMMWEAVSVKTEVVGVSTL
OraPyV-Sum   MSCKRKRGGDSKTKCTPCPTRKK-CP--------TAPVPVKLIVKGGVEVLDIR-TGPDSITTIEAFLNPRMGFN-------QSSEQNYGFSDKITTATSRSE----DKPKQNTLPCYSCARIALPLLNEDLTCASLQMWEAVSVKTEVVGISSF
ChPyV        MAPPRKRARCVSTPTKVK-CVPKKCP--------VPTPVPKLLVKGGVEVLNII-TGPDSTTEIELYLEPRMGINSP-TG--DKKEW-YGYSEVIHHADG-YD----NNLLNIQMPQYSCARVQLPMLNTDMTCDTLMMWEAVSCKTEIVGIGSL
LPV          MAPQRKRQDG--------A-CKKTCP--------IPAPVPRLLVKGGVEVLEVR-TGPDAITQIEAYLNPRMGNN-------IPSEDLYGYSNSINTAFSKAS----DTPNKGTLPCYSVAVIKLPLLNEDMTCDTILMWEAVSVKTEVVGISSL
SquiPyV      MPLKRKSASA-------------------PRK--TPQEVPRLLISGGVEVLGLK-TGPDSITEVEAFLNPRMGLE--------NTDNHYGYSENVTVAKKKDD----DKPLKNQLPCYSVAKIELPMLNEDLTNDTILMWECISVKTEVVGINTL
MePyV        MAPKRKASSTCKTPKRQCIPKPGCCP--------NVASVPKLLVKGGVEVLSVV-TGEDSITQIELYLNPRMGVN---SPDLPTTSNWYTYTYDLQPKGS-SP----DQPIKENLPAYSVARVSLPMLNEDITCDTLQMWEAISVKTEVVGISSL
SV40         MAPTKRKGS---------------CPGAAPKKPKEPVQVPKLVIKGGIEVLGVK-TGVDSFTEVECFLNPQMGNP---------DEHQKGLSKSLAAEKQFTD----DSPDKEQLPCYSVARIPLPNLNEDLTCGNILMWEAVTVKTEVIGVTAM
SA12         MPPAKRKGE---------------CPGAAPKKPKDPVRVPKLLIRGGVEVLEVK-TGVDSITEVECFLTPEMGDA---------NEHLRGYSQRVTCDVTFEN----DAPQKKTLPCYSTARIPLPNLNEDLTCGNILMWEAVTVKTEVLGVTSM
BKV          MAPTKRKGE---------------CPGAAPKKPKEPVQVPKLLIKGGVEVLEVK-TGVDAITEVECFLNPEMGDP---------DENLRGFSLKLSAENDFSS----DSPERKMLPCYSTARIPLPNLNEDLTCGNLLMWEAVTVQTEVIGITSM
JCV          MAPTKRKGER-----------------------KDPVQVPKLLIRGGVEVLEVK-TGVDSITEVECFLTPEMGDP---------DEHLRGFSKSISISDTFES----DSPNRDMLPCYSVARIPLPNLNEDLTCGNILMWEAVTLKTEVIGVTSL
KI PyV       MSCTPCRPQKRLT----RPRSQ----------VPRVQTLATEVKKGGVEVLAAVPLSEETEFKVELFVKPVIGNTTAAQDGREPTPHYWSISSAIHDKESGSSIKVEETPDADTTVCYSLAEIAPPDIPNQVSECDMKVWELYRMETELLVV-PL
WU PyV       MACTAKPACTAKPGR--SPRSQ----------PTRVQSLPKQVRKGGVDVLAAVPLSEETEFKVELFVKPVIGNA----EG--TTPHYWSISSPLKTAEAANV----TPDADTTVCYSLSQVAPPDIPNQVSECDMLIWELYRMETEVLVL-PV
MuPyV        MAPKRKSGVS---KCET-KCTKA-CP--------RPAPVPKLLIKGGMEVLDLV-TGPDSVTEIEAFLNPRMGQPPT-PESLTEGGQYYGWSRGINLATSDTE----DSPGNNTLPTWSMAKLQLPMLNEDLTCDTLQMWEAVSVKTEVVGSGSL
HaPyV        MAPKRKSGASS--RCAN-PCGKP-CP--------KPANVPKLIMRGGVGVLDLV-TGEDSITQIEAYLNPRMGQN-K-PGTGTDG-QYYGFSQSIKVNSSLTA----DEVKANQLPYYSMAKIQLPTLNEDLTCDTLQMWEAVSVKTEVVGVGSL
MyoPyV       MAPKKRSS--------------------------APSEVPKLIISGGIEVLSVR-TGPDSITTIEAYLNPRMGQP--------PESDFYGFSDNITVSQSFEA----DQPQIKEIPCYSMAKINLPLLNEDITCDTILMWEAISVKTEVVGISSL
BoPyV        MSRMRKNMN----------------P-PKKGLKGQPSPVPKLIIKGGIEVLGLR-TGPDSTTTIELFLNPRMGQS--------TESEYYGFSDNQRGSTSRTD----EDLISAELPRYSLGVVQLPLLNEKLTDDVLLMWEAVSCKTEVVGVNTL
FiPyV        MAPKKGNGS---------------CP--------RPQQVPKLIVKGGIEVLDVK-TGPDSTTTIEAYLNPRVGQN-------------WGFSTEITVASNGYN----DAPHLTEIPCYSCARISLPLLNEDITCPTLLMWEAVSVKTEVVGISSM
BFDPyV       MS-QKGKGS---------------CP--------RPQQVPRLLVKGGIEVLDVK-SGPDSITTIEAYLQPRPGQK-------------NGYSTVITVQAEGYQ----DAPHSTEVPCYSCARIPLPTINDDITCPTLLMWEAVSVKTEVVGVSSI
CrPyV        MGPKRPRAA-------------------------GPSPVPRLLIKGGIEVLEVK-TGPDSFTTIEAYLNPRMGLD-------------YGFSEAITVAHTLNP----DVPPKKQLPTYSCARIGLPELNDDMTKLQILMWEAVSCKTEVVGITSL
GoPyV        MAPKVKRPRN------------------------GPVPVPRLLVKGGIEVLGVR-TGPDSTTTIEAYLNPRMGTD-------------NGFSQAVTVATSLNP----DVPPKAELPCYSCARIGLPMLNEDMTTPEILMWEAVSVKTEVVGVTTM
```

BC-loop

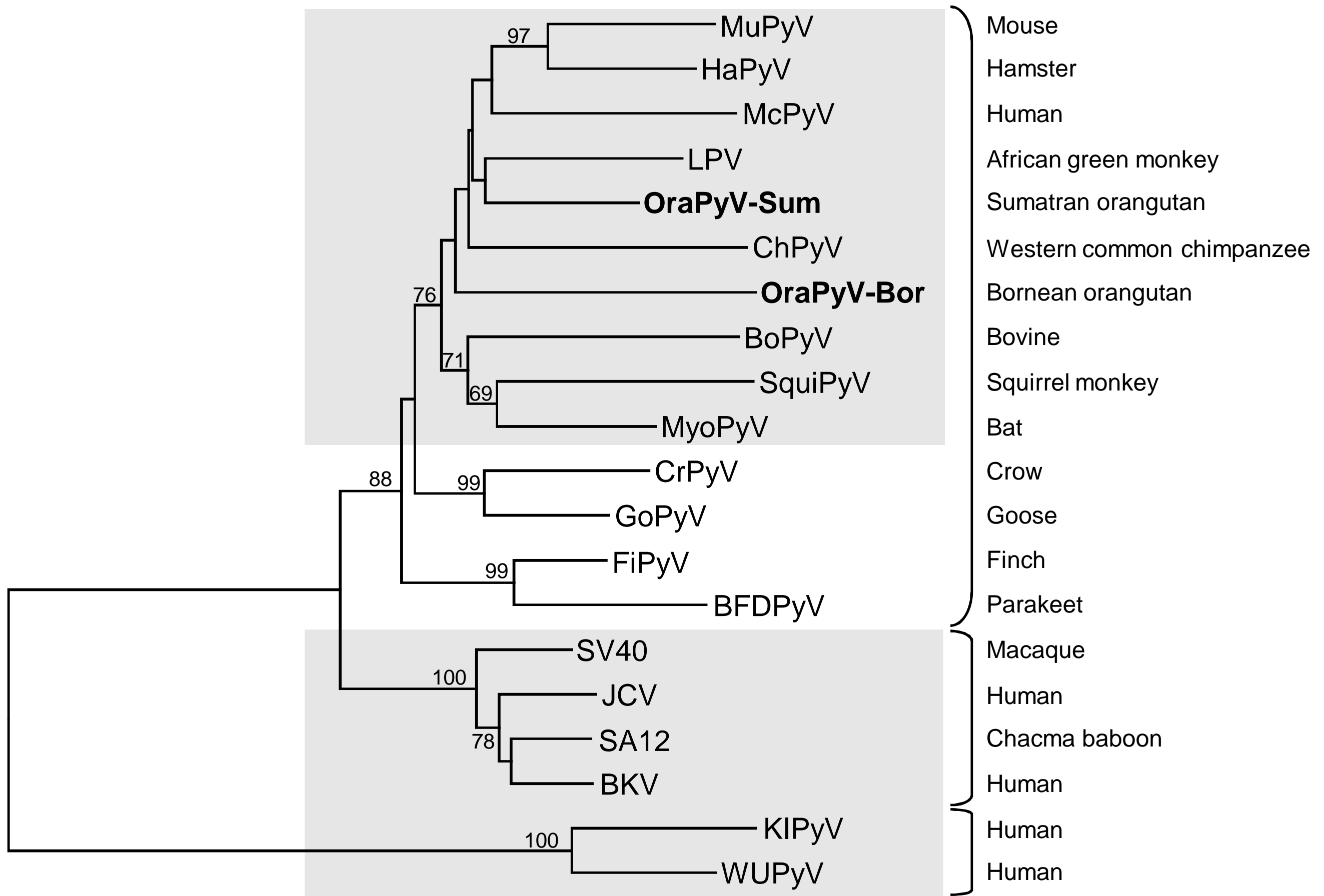
```
                   160                 180                 200                 220                 240                 260                 280                 300
OraPyV-Bor   VNCHMPSKRMYDDQ---GIGFPIEGMNFHMFAVGGEPLDLQFITSNYKTEY-------KDKYVGPDGKSSTQ----QALNTAYKSKLLKDGAFPVECWCPDPFKNENSRYYGSYTGGQSTPPVMQFTNVTTVLLDENWVVGPLCKGDSLYVSAC
OraPyV-Sum   VNVHSAAKKETDSQ---GPALPIEGLNYHMFAVGGEPLELQGLVMNYEAKYNENS---STVISIKKVTNADMTAKNQVLDMAAKATLDSDALYPVEMWCPDPSKNENTRYFGSYTGGLTTPPVLQFTNVTTVLLDENG-VGPLCKGDGLFLSCA
ChPyV        ISVHLLEAKMAAKEGGDGPSQPIEGMNYHMFAVGGEPLDLQGIESNALTKYASAIPPKTIHPNDIAKLAEEEKPQLQGLVPKAKARLDKDGFYPIEEWSPGPSRNENSRYFGSFVGGLNTPPNLQFTNAVTTVLLDENG-VGPLCKGDGLFVSAA
LPV          VNLHQGGKYIYGSS---SGCVPVQGTTYHMFAVGGEPLELQGLVASSTATY-------PDDVVAIKNMKPGN----QGLDPKAKALLDKDGKYPVEVWCPDPSKNENTRYYGSFTGGATTPPVMQFTNSVTTVLLDENG-VGPLCKGDKLFLSCA
SquiPyV      VNTHSYGKRDGE-----SPSMPIVGLNYHFFAVGGEPIEMQYIVQNFQCDY-------PVGVVAMKPSPLST----QVLDPKQKSKLTGDGIFPIECWAPDPSKNENARYFATYTGGLNTPPVLNITNVTTILLNENG-VGPLCKGDQLHIAAD
MePyV        INVHYWDMKRVHDY---GAGIPVSGVNYHMFAIGGEPLDLQGLVLDYQTEYPKTTNGGPITIETVLGRKMTP--KNQGLDPQAKAKLDKDGNYPIEVWCPDPSKNENSRYYGSIQTGSQTPTVLQFSNTLTTVLLDENG-VGPLCKGDGLFISCA
SV40         LNLHSGTQKTHEN----GAGKPIQGSNFHFFAVGGEPLELQGVLANYRTKY-------PAQTVTPKNATVDS----QQMNTDHKAVLDKDNAYPVECWVPDPSKNENTRYFGTYTGGENVPPVLHITNTATTVLLDEQG-VGPLCKADSLYVSAV
SA12         LNLHAEAQKVHDN----GAGRPVQGSNFHFFSVGGEPLELQGVLHNYRTTY-------PEGTIAPKNPTAES----QVMNTEHKAYLDKTGAYPVECWVPDPSRNENTRYFGTYTGGENVPPVLHVTNTATTVLLDEQG-VGPLCKADSLYVSAV
BKV          LNLHAGSQKVHEH----GGGKPIQGSNFHFFAVGGEPLEMQGVLMNYRSKY-------PDGTITPKNPTAQS----QVMNTDHKAYLDKNNAYPVECWVPDPSRNENARYFGTFTGGENVPPVLHVTNTATTVLLDEQG-VGPLCKADSLYVSAA
JCV          MNVHSNGQATHDN----GAGKPVQGTSFHFFSVGGEALELQGVLFNYRTKY-------PDGTIFPKNATVQS----QVMNTEHKAYLDKNKAYPVECWVPDPTRNENTRYFGTLTGGENVPPVLHITNTATTVLLDEFG-VGPLCKGDNLYLSAV
KI PyV       VNALGNTN---------GVVHGLAGTQLYFWAVGGQPLDVVGVTPTDKYKG-------PTTYTINPPGDPRTLHV-YNSN-TPKAKVTSE-RYSVESWAPDPSRNDNCRYFGRVVGGAATPPVVSYGNNSTIPLLDENG-IGILCLQGRLYITCA
WU PyV       LNAGILTT---------GGVGGIAGPQLYFWAVGGQPLDVLGLAPTEKYKG-------PAQYTVNPKTNGTVPHV-YSSSETPRARVTNE-KYSIESWVADPSRNDNCRYFGRMVGGAATPPVVSFSNNSTIPLLDENG-IGILCLQGRLYITCA
MuPyV        LDVHGFNKPTDTVNT-KGISTPVEGSQYHVFAVGGEPLDLQGLVTDARTKYKEEG---VVTIKTITKKDMVN--KDQVLNPISKAKLDKDGMYPVEIWHPDPAKNENTRYFGNYTGGTTTPPVLQFTNTLTTVLLDENG-VGPLCKGEGLYLSCV
HaPyV        LNVHGYGSRSETKD--IGISKPVEGTTYHMFAVGGEPLDLQGLVQNYNANY-EAA---IVSIKTVTGKAMTS--TNQVLDPTAKAKLDKDGRYPIEIWGPDPSKNENSRYYGNFTGGTGTPPVMQFTNTLTTVLLDENG-VGPLCKGDGLYLSAA
MyoPyV       CNVHSAVHREFENE---GAGFPVQGLNFHFFAVGGEPLELQMVVDNHKCRY-------PAGLAALQQAPKTA----QVLDPQLKANLIKDGTFPVEVWCPDPAKNENSRYYGSYTGGLTTPPVLQFTNVTTILLNENG-VGPLCKGDGLFLSCA
BoPyV        TTCHGYKKRYSPSAG-QGSAMPIEGINYHFFAVGGEPLEIQFICEDFKAPYH------PTETIVPPKDKLSN--KSQVLDPTLKGILDKDGVYPVECWCPDPSKNENTRYFGTYTGGVSTPPVLQFTNVTTILLDENG-VGPLCKADKLYITAA
FiPyV        LNMHSYGLRAFGGY---GGGYTIEGSHIHFFSVGGEPLDLQGLMQNHSTQY-------PSPLVGPKKPDGTTDDSAQVLNPIYKAKLDKDATYPIECWCPDPSRNENSRYFGSYTGGVETPPVLSFTNTSTTILLDENG-VGPLCKGDGLYLSSA
BFDPyV       LNMHSGAFRAFNGY---GGGFTICGPRIHFFSVGGEPLDLQACMQNSKTVY-------PAPLIGPGEGEPRETA--QVLDTGYKARLDKDGLYPIECWCPDPAKNENTRYYGNLTGGPETPPVLAFTNTTTILLDENG-VGPLCKGDGLFLSAA
CrPyV        TNLHSGMKRTVSGY---GAGIPIEGLNLHFFAIGGEPLDLQGLVQNYTATY-------PEGVIAPKLADAKA----QILDPTLKAKLDKDGTYPVECWSPDPSRNENTRYFGSYTGGTQTPPVLPIGNSVTTVLLDENG-IGPLCKGDGLYISAV
GoPyV        CNVHSASIRMNGGY---GVGRPIEGLNCHMFAVGGEPLELQGCVQNWSTTY-------PSGVVAPPLKDAKA----QVLDPGLKARLDKDGAYPVECWCPDPSRNENTRYFGTYTGGQQTPPVLPFTNTVTTVLLDENG-VGPLCKGDGLYLSCV
```

DE-loop

EF-loop

```
                   320                 340                 360                 380                 400                 420                 440                 460
OraPyV-Bor   DIVGFYV-ANDGSMKYRGLPRYFDITLRKRIVKNPYPVTSLLNNLFTGMMPPVQGQPMDSRLFTQVEEVRVYDGNRK---DFQENPDMVRYIDQF-GQDQT------RPPIRMH-----------------------------
OraPyV-Sum   DICGFFTNAD-GSKQYRGLPRYFNVTLRKRNVRNPYPVTSLLSSLFSSLMPKMQGQPMEG-RDAQVEEVRVYEG-TE---SVTADPDMRRFIDKF-GQKQT------DVPQ--------------------------------
ChPyV        DICGVMVKADNEAIRYRGLSRYFKVTLRKRAVKNPYPITSLLGSLFTGLMPKMDGQPMTG-PDAQVEEVRIYQG-KE---GLPADPDMKRYIDQF-GQEQT------PTPTPAAPAA--------------------------
LPV          DIAGVHTNYS-ETQSWRGLPRYFNVTLRKRIVKNPYPVSSLLNSFFSGLMPQIQGQPMEG-VSGQVEEVRIFEG-TE---GLPGDPDLNRYVDKF-CQNQT------VLPVSNDM-----------------------------
SquiPyV      DICGFHIEED-NKYKYRGLPRYFKLVLRKRVVKNPYPVSTLLNTLFTQMSPNVQGQDMT----KQVEEVKIYEG-TE---KLPGDPDMIRFRNQF-GQEET------VIPVITN------------------------------
MePyV        HIVGFLFKTS-GKMALHGLPRYFNVTLRKIWVKNPYPVVNLINSLFSNLMPKVSGQPMEG-KDNQVEEVRIYEG-SE---QLPGNPDIVRFLDKF-GQEKTVYPKPSVAPAAVTFQSNQQDKGKAPLKGPQKASQKESQTQQLL
SV40         DICGLFTNTS-GTQQWKGLPRYFKITLRKRSVKNPYPISFLLSDLINRRTQRVDGQPMIG-MSSQVEEVRVYED-TE---ELPGDPDMIRYIDEF-GQTTT------RMQ---------------------------------
SA12         DICGLFTNSS-GTQQWRGLSRYFKISLRKRSVKNPYPISFLLSDLINRRTTRVQGQPMYG-MNAQVEEVRVYDG-TE---ELPGDPDMMRYIDKY-GQHQT------KML---------------------------------
BKV          DICGLFTNSS-GTQQWRGLARYFKIRLRKRSVKNPYPISFLLSDLINRRTQRVDGQPMYG-MESQVEEVRVFDG-TE---RLPGDPDMIRYIDKQ-GQLQT------KML---------------------------------
JCV          DVCGMFTNRS-GSQQWRGLSRYFKVQLRKRRVKNPYPISFLLTDLINRRTPRVDGQPMYG-MDAQVEEVRVFEG-TE---ELPGDPDMMRYVDKY-GQLQT------KML---------------------------------
KI PyV       DMLGTAN-----SRIHTPMARFFRLHFRQRRVKNPFTMNVLYKQVFNRPTETVDAQVGVT-EVTMVEEIGPLPPSIQ--TTLPTSVNLTQLPRTVTLQSQA------PLLNTQQNSK--------------------------
WU PyV       DLLGVNK-----NRVHTGLSRFFRLHFRQRRVRNPYTINLLYKQVFNKPADDISGQLQVT-EVTMTEETGPLPPTVEGNVGVPTTSNLSHLPATVTLQATG------PILNTQG------------------------------
MuPyV        DIMGWRVTRNYDVHHWRGLPRYFKITLRKRWVKNPYPMASLISSLFNNMLPQVQGQPMEG-ENTQVEEVRVYDG-TE---PVPGDPDMTRYVDRF-GKTKT------VFPGN--------------------------------
HaPyV        DVMGWYIEYNSAGWHWRGLPRYFNVTLRKRWVKNPYPVTSLLASLYNNMLPTIEGQPMEG-EAAQVEEVRIYEG-TE---AVPGDPDVNRFIDKY-GQQHTKPP---AKPAN------------------------------
MyoPyV       DICGFRTQDN-KKMQFRGLPRYFNVTLRKRLVKNPYPVSSLLNTLFSDMMPKVQGQVMEG-DKAQVKEVRVYEG-VE---RLPGDPDMIRYRDKY-GQEKT------KIPAVAN------------------------------
BoPyV        DICGFLTQPN-DQQQFRGLPRYMSVTLRKRLVKNPYPIASILTSLFTNSLPPVTSQEMD----KQVEEVRIYQG-VE---GLPGDPDMVRYINKF-GQEET------CIPK--------------------------------
FiPyV        DVAGTFVQQTSQKQYWRGLPRYFNITLRKRAVKNPYPVTGLLTSLFNGLMPRMQGQSMSG-PTAQVEEVRVYEG-ME---GLPGDPTMERHLDQQ-GQSHT-------NPAQD-------------------------------
BFDPyV       DVAGTYVDQRG-RQYWRGLPRYFSIQLRKRNVRNPYPVSGLLNSLFNDLMPRMTGQSMQG-SDAQVEEVRVYEG-ME---GLA--PEID-------MPPK-------APR--------------------------------
CrPyV        DIVGLYTAQYDETQYFRGLPRYFSVTLRKRVVKNPYPISSLLTSLFGNILPRMDGQPMQG-KDGQVEEVRIYQG-VE---PLPGDPTLTRTVDRY-GQNEP------VLPHRE-------------------------------
GoPyV        DICGFYSEQYSQKQHFRGLPRYFSVSLRKRLVRNPYPVSSLLTSLFTSIIPKMQGQPMVG-DNSQVEEVRVYEG-VE---PLPGDPDISRFVDRF-GEQHT------HMPSN--------------------------------
```

HI-loop


97
MuPyV
HaPyV
McPyV
LPV
OraPyV-Sum
ChPyV
76
OraPyV-Bor
BoPyV
71
SquiPyV
69
MyoPyV
88
99
CrPyV
GoPyV
99
FiPyV
BFDPyV
SV40
100
JCV
78
SA12
BKV
100
KIPyV
WUPyV
0.2
Mouse
Hamster
Human
African green monkey
Sumatran orangutan
Western common chimpanzee
Bornean orangutan
Bovine
Squirrel monkey
Bat
Crow
Goose
Finch
Parakeet
Macaque
Human
Chacma baboon
Human
Human
Human