



## Evolution of four BK virus subtypes

Andi Krumbholz<sup>a,1</sup>, Olaf R.P. Bininda-Emonds<sup>b,1,2</sup>, Peter Wutzler<sup>a</sup>, Roland Zell<sup>a,\*</sup>

<sup>a</sup> Institute of Virology and Antiviral Therapy, Friedrich Schiller University Jena, Hans-Knoell-Strasse 2, D-07745 Jena, Germany

<sup>b</sup> Institute for Systematic Zoology and Evolutionary Biology, Friedrich Schiller University Jena, Erbertstr. 1, D-07743 Jena, Germany

### ARTICLE INFO

#### Article history:

Received 12 March 2008

Received in revised form 27 May 2008

Accepted 30 May 2008

Available online 5 June 2008

#### Keywords:

Polyomavirus

BKV

Phylogenetic analysis

Relaxed molecular clock

### ABSTRACT

BK viruses (BKV) comprise four subtypes that are distinguishable by serological and molecular methods with the latter indicating up to four subgroups within subtype I. In this study, the phylogeny of all BKV subtypes was analyzed. Phylogenetic analyses of the viral structural protein VP1, concatenated sequences including the T-Ag, t-Ag, VP1 and VP2 genes and the entire coding region of BKV, each employing several tree inference methods, consistently revealed seven strongly supported clades that correlate with BKV subtypes and subgroups. In general, subtype II and III viruses were found to be sister groups, the genetic distances between which were significantly lower than those between either and the other BKV subtypes. Two benchmarks of human evolution (emergence of modern humans, 200,000 years ago; out-of-Africa migration, 100,000 years ago) were assumed for the internal calibration of BKV evolution. Utilization of either calibration point resulted in the diversification of most BKV subgroups coinciding with human radiation less than 50,000 years ago. Another approach, external calibration by linking BKV divergence to the evolution of mammals, was rejected as it advances BKV divergence prior to the emergence of modern humans.

© 2008 Published by Elsevier B.V.

## 1. Introduction

BK viruses (BKV) represent a discrete species within the genus *Polyomavirus* of the *Polyomaviridae*, a family of nonenveloped viruses with icosahedral capsids. The capsid encloses a circular double-stranded DNA genome of approximately 5100 nucleotides that is coated by host-cell histones. All polyomaviruses display a similar genome organization (Fig. 1). The genome can be functionally divided into three regions with those coding for the early and late proteins being separated by a non-coding region (NCR) that contains the origin of replication and promoters (Cole, 1996; Hirsch and Steiger, 2003; Hou et al., 2005; Jin and Gibson, 1996).

The early region consists of two open reading frames (ORFs), both of which encode the non-structural proteins designated T- and t-antigen (T-Ag, t-Ag). The two early transcripts are produced by alternative splicing from a common pre-mRNA prior to viral replication. Because both T-Ag and t-Ag use the same start codon, the N-terminal amino acids of these two proteins are identical (Cole, 1996; Jin and Gibson, 1996).

The late region of the BKV genome codes for the structural proteins VP1, VP2, and VP3 as well as for the agnoprotein (VPx). As with the early mRNAs, late transcripts are generated from a common pre-mRNA by alternative splicing (Cole, 1996). The major capsid protein VP1 is responsible for the antigenic variability among BKV isolates (Jin et al., 1993b). In contrast to the generally high degree of similarity of the entire coding region of VP1 (>95% similarity across all BKVs), similarity between amino acid residues 61 to 83 is only 61 to 70%. These latter amino acids probably constitute the epitope responsible for serological subtyping. Four BKV serogroups I–IV are distinguishable (Jin et al., 1993b; Knowles et al., 1989) that correlate with the division of BKVs into four major subtypes based on genotyping of this subgenomic region (Jin et al., 1993b).

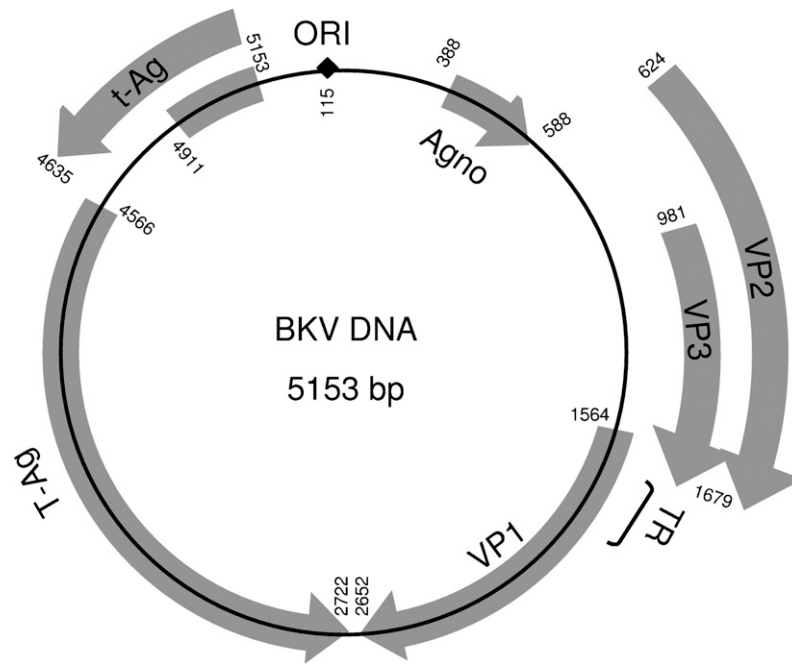
BKV was first isolated from the urine obtained from patient B.K. after renal transplantation (Gardner et al., 1971). Epidemiological studies revealed that up to 90% of the adult population worldwide is seropositive for BKV (Knowles, 2001). Primary BKV infection occurs early in childhood and appears to be asymptomatic. In rare cases, infection is associated with mild diseases of the upper respiratory tract, tonsillitis, slight pyrexia and transient cystitis (Hirsch, 2005). Acute infection is self-limited and proceeds to latent infection. It is generally believed that the urogenital tract is the primary site of latency (Hirsch, 2005). In addition, the viral genome was found in peripheral blood mononuclear cells (Mannon, 2004). In both a state

\* Corresponding author. Tel.: +49 3641 9395711; fax: +49 3641 9395702.

E-mail address: [roland.zell@med.uni-jena.de](mailto:roland.zell@med.uni-jena.de) (R. Zell).

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Current address: AG Systematics and Evolutionary Biology, IBU - Faculty V, Carl von Ossietzky University Oldenburg, 26111 Oldenburg, Germany.



**Fig. 1.** Genome organization of BKV. The genome of BKV is a circular double-stranded DNA of 5153 nucleotides (strain Dun). Six open reading frames (encoding the agnoprotein, the capsid proteins VP1, VP2 and VP3, early regulatory proteins large T-Ag and small t-Ag), the origin of replication (ORI) and the 287-nt typing region (TR) are indicated. Nucleotide numbers refer to Dun-numbering (Seif et al., 1979).

**Table 1**

BKV isolates compared in this study (based on and expanded from Krumbholz et al., 2006; Nishimoto et al., 2006)

Subtype	Isolate	Clinical state	Geogr. origin	Accession no.	Reference
Ia	DUN <sup>a</sup>	WAS	USA	NC_001538	Howley et al. (1975) and Seif et al. (1979)
Ia	UT <sup>a</sup>	Solid tumor	USA	DQ305492	Unpublished
Ia	PittNP4	RT	USA	DQ989802	Sharma et al. (2006)
Ia	PittNP5	RT	USA	DQ989804	Sharma et al. (2006)
Ia	PittVR4	RT	USA	DQ989812	Sharma et al. (2006)
Ia	PittVR9	RT	USA	DQ989807	Sharma et al. (2006)
Ia	KEN-1	No data available	Kenya	AB263926	Zheng et al. (2007)
Ia	KEN-4	No data available	Kenya	AB263928	Zheng et al. (2007)
Ia	CAF-5	No data available	Central Africa	AB263913	Zheng et al. (2007)
Ia	CAF-9	No data available	Central Africa	AB263914	Zheng et al. (2007)
Ia	CAF-15	No data available	Central Africa	AB263912	Zheng et al. (2007)
Ia	ZAF-1	No data available	South Africa	AB263938	Zheng et al. (2007)
Ib	DIK <sup>a</sup>	Acute tonsillitis	Netherlands	AB211369	Goudsmit et al. (1981)
Ib	WW <sup>a</sup>	RT	South Africa	AB211371	Chauhan et al. (1984)
Ib	HI-u5	HIV-2	West Africa	AY628225	Chen et al. (2004)
Ib	HI-u6	HIV-2	West Africa	AY628237	Chen et al. (2004)
Ib	HI-u8	HIV-2	West Africa	AY628238	Chen et al. (2004)
Ib	CAP-h2	CAP	USA	AY628226	Chen et al. (2004)
Ib	CAP-h5	CAP	USA	AY628228	Chen et al. (2004)
Ib	CAP-h8	CAP	USA	AY628229	Chen et al. (2004)
Ib	CAP-h22	CAP	USA	AY628227	Chen et al. (2004)
Ib	CAP-m2	CAP	USA	AY628231	Chen et al. (2004)
Ib	CAP-m5	CAP	USA	AY628232	Chen et al. (2004)
Ib	CAP-m9	CAP	USA	AY628233	Chen et al. (2004)
Ib	CAP-m13	CAP	USA	AY628230	Chen et al. (2004)
Ib	CAP-m18	CAP	USA	AY628224	Chen et al. (2004)
Ib	PittVM2	RT	USA	DQ989809	Sharma et al. (2006)
Ib	PittNP1	RT	USA	DQ989813	Sharma et al. (2006)
Ib	PittVR8	RT	USA	DQ989806	Sharma et al. (2006)
Ib	KOM-1	BMT	Japan	AB211373	Nishimoto et al. (2006)
Ib	KOM-5	BMT	Japan	AB211374	Nishimoto et al. (2006)
Ib	GBR-6	No data available	England	AB263922	Zheng et al. (2007)
Ib	KEN-3	No data available	Kenya	AB263927	Zheng et al. (2007)
Ib	NER-1	No data available	Niger	AB263932	Zheng et al. (2007)
Ib	PHL-6	No data available	Philippines	AB263933	Zheng et al. (2007)
Ib	PHL-7	No data available	Philippines	AB263934	Zheng et al. (2007)
Ib	MMR-6	No data available	Myanmar	AB263929	Zheng et al. (2007)
Ib	VNM-9	No data available	Vietnam	AB263937	Zheng et al. (2007)
Ib	JL <sup>a</sup>	BMT	Netherlands	AB211370	Pauw and Choufoer (1978)
Ib	HC-u2	Healthy	USA	AY628234	Chen et al. (2004)

Table 1 (Continued)

Subtype	Isolate	Clinical state	Geogr. origin	Accession no.	Reference
Ib	HC-u5	Healthy	USA	AY628235	Chen et al. (2004)
Ib	HC-u9	Healthy	USA	AY628236	Chen et al. (2004)
Ib	PittVM1	RT	USA	DQ989794	Sharma et al. (2006)
Ib	PittVR1	RT	USA	DQ989795	Sharma et al. (2006)
Ib	PittVR2	RT	USA	DQ989796	Sharma et al. (2006)
Ib	PittVR3	RT	USA	DQ989797	Sharma et al. (2006)
Ib	PittVM3	RT	USA	DQ989798	Sharma et al. (2006)
Ib	PittNP2	RT	USA	DQ989799	Sharma et al. (2006)
Ib	PittNP3	RT	USA	DQ989800	Sharma et al. (2006)
Ib	PittVM4	RT	USA	DQ989801	Sharma et al. (2006)
Ib	PittVR5	RT	USA	DQ989803	Sharma et al. (2006)
Ib	PittVM5	RT	USA	DQ989805	Sharma et al. (2006)
Ib	PittVR10	RT	USA	DQ989808	Sharma et al. (2006)
Ib	PittVR6	RT	USA	DQ989810	Sharma et al. (2006)
Ib	PittVR7	RT	USA	DQ989811	Sharma et al. (2006)
Ib	FIN-10	RT	Finland	AB260032	Ikegaya et al. (2006)
Ib	FIN-11	RT	Finland	AB260029	Ikegaya et al. (2006)
Ib	FIN-13	RT	Finland	AB260028	Ikegaya et al. (2006)
Ib	FIN-14	RT	Finland	AB260031	Ikegaya et al. (2006)
Ib	FIN-23	RT	Finland	AB260030	Ikegaya et al. (2006)
Ib	ESP-2	No data available	Spain	AB263915	Zheng et al. (2007)
Ib	ETH-4	No data available	Ethiopia	AB263917	Zheng et al. (2007)
Ib	FNL-12	No data available	Finland	AB263918	Zheng et al. (2007)
Ib	FNL-22	No data available	Finland	AB263919	Zheng et al. (2007)
Ib	GBR-4	No data available	England	AB263921	Zheng et al. (2007)
Ib	GBR-8	No data available	England	AB263923	Zheng et al. (2007)
Ib	GBR-9	No data available	England	AB263924	Zheng et al. (2007)
Ib	ITA-5	No data available	Italy	AB263925	Zheng et al. (2007)
Ib	SWE-2	No data available	Sweden	AB263935	Zheng et al. (2007)
Ib	TUR-5	No data available	Turkey	AB263936	Zheng et al. (2007)
Ic	MT <sup>a</sup>	SLE	Japan	AB211372	Nishimoto et al. (2006) and Sugimoto et al. (1989)
Ic	KOM-6	BMT	Japan	AB211375	Nishimoto et al. (2006)
Ic	RYU-1	RT	Japan	AB211376	Nishimoto et al. (2006)
Ic	RYU-2	RT	Japan	AB211377	Nishimoto et al. (2006)
Ic	THK-6	RT	Japan	AB211378	Nishimoto et al. (2006)
Ic	THK-9	RT	Japan	AB211379	Nishimoto et al. (2006)
Ic	THK-9a	RT	Japan	AB217921	Takasaka et al. (2006a)
Ic	THK-11	RT	Japan	AB211380	Nishimoto et al. (2006)
Ic	TW-1	RT	Japan	AB211381	Nishimoto et al. (2006)
Ic	TW-1a	RT	Japan	AB217917	Takasaka et al. (2006a)
Ic	TW-1b	RT	Japan	AB217918	Takasaka et al. (2006a)
Ic	TW-2	RT	Japan	AB213487	Unpublished
Ic	TW-4	RT	Japan	AB211382	Nishimoto et al. (2006)
Ic	TW-5	RT	Japan	AB211383	Nishimoto et al. (2006)
Ic	TW-7	RT	Japan	AB211384	Nishimoto et al. (2006)
Ic	TW-8	RT	Japan	AB211385	Nishimoto et al. (2006)
Ic	TW-8a	RT	Japan	AB217920	Takasaka et al. (2006a)
Ic	NEC-8	No data available	China	AB263931	Zheng et al. (2007)
Ic	NEC-12	No data available	China	AB263930	Zheng et al. (2007)
II	SB <sup>b,a</sup>	Lymphoma	England	Z19536	Jin et al. (1993b)
II	J/1025/05	BMT	Germany	EF376992	This study
II	GBR-12	No data available	England	AB262920	Sugimoto et al. (1997) and Zheng et al. (2007)
II	ETH-3	No data available	Ethiopia	AB263916	Sugimoto et al. (1997) and Zheng et al. (2007)
III	AS <sup>a</sup>	Pregnant	England	M23122	Coleman et al. (1980) and Tavis et al. (1989)
III	KOM-3	BMT	Japan	AB211386	Nishimoto et al. (2006)
IV	IV <sup>b,a</sup>	RT	England	Z19535	Jin et al. (1993b)
IV	KOM-2	BMT	Japan	AB211387	Nishimoto et al. (2006)
IV	KOM-7	BMT	Japan	AB211388	Nishimoto et al. (2006)
IV	RYU-3	RT	Japan	AB211389	Nishimoto et al. (2006)
IV	THK-8	RT	Japan	AB211390	Nishimoto et al. (2006)
IV	TW-3	RT	Japan	AB211391	Nishimoto et al. (2006)
IV	TW-3a	RT	Japan	AB217919	Takasaka et al. (2006a)
IV	FIN-2	RT	Finland	AB260033	Ikegaya et al. (2006)
IV	FIN-4	RT	Finland	AB260034	Ikegaya et al. (2006)

WAS, Wiskott-Aldrich syndrome; BMT, bone marrow transplant; RT, renal transplant; SLE, systemic lupus erythematosus; CAP, capillary leak syndrome; HIV-2, HIV-2 infection.

<sup>a</sup> Reference isolate used for genotyping.

<sup>b</sup> Partial VP1 sequence.

of immunosuppression as well as in some healthy individuals virus may be reactivated leading to viral shedding in urine (Hirsch, 2005). In bone marrow recipients, BKV reactivation is linked to hemorrhagic cystitis, whereas severe tubulointerstitial nephritis and ureteric stenosis are causes of concern in renal transplant recipients

(Hirsch, 2005). Furthermore, there is evidence that BKV infection is a prominent risk factor for bladder carcinoma in immunocompetent individuals (Weinreb et al., 2006).

The regional distribution of the BKV subtypes has been established in several studies (Agostini et al., 1995; Baksh et al.,

2001; Chen et al., 2006; Di Taranto et al., 1997; Ikegaya et al., 2006; Jin et al., 1993a, 1995; Krumbholz et al., 2006; Takasaka et al., 2004). Subtype I is most prevalent, followed by subtype IV, with subtypes II and III occurring less frequently. BKV subtype IV was previously found to be prevalent in East Asia except for Japan (Chen et al., 2006), although there is evidence that the prevalence of this subtype was underestimated in other regions (Zheng et al., 2007). On the basis of the analysis of 121 sequences derived from urine specimens from Japanese renal and bone marrow transplant recipients, the predominant subtype I was described as consisting of three subgroups designated Ia, Ib and Ic (Takasaka et al., 2004). Whereas subgroup Ic is prevalent in Japan, Ib is widespread in Germany and in other European countries (Ikegaya et al., 2006; Krumbholz et al., 2006; Takasaka et al., 2004), and Ia appears to be prevalent in Africa (Zheng et al., 2007).

A recent study which focused on the evolution of BKVs (Nishimoto et al., 2006) estimated that the clades corresponding to subtypes I, III and IV diverged approximately 4.4 million years ago. In addition, the BKV subgroup Ib was further subdivided into subgroups Ib-1 and Ib-2 (diverging from one another approximately 1.2 million years ago). The concept of four subtype I subgroups was first raised by an analysis of 52 full-length BKV genomes including two subtype II sequences (Zheng et al., 2007). Furthermore, host-linked evolution of polyomaviruses has also been investigated (Perez-Losada et al., 2006).

In the present study, the complete genome of the German subtype II virus J/1025/05 was compared to other available BKV sequences. A molecular genetic analysis as well as phylogenetic analyses of the capsid protein VP1 and concatenated sequences of T-Ag, t-Ag, VP1 and VP2 was conducted to study the relationship of J/1025/05 to the other BKV subtypes. Finally, we investigated the evolutionary dynamics of BKV assuming different calibration points and evolutionary rates.

## 2. Material and methods

### 2.1. Sample sources and sequencing

DNA of the complete genome of J/1025/05 was amplified from the urine sample of a 58-year-old German female bone marrow transplant recipient suffering from hemorrhagic cystitis. Subtyping of the virus was based on sequence analysis of the 287-nt VP1 typing region (i.e., nts 1650 to 1936, DUN numbering) as described previously (Krumbholz et al., 2006). For amplification, specific oligonucleotide primers were used in polymerase chain reaction (PCR) followed by sequencing of the purified amplicons with the respective oligonucleotides. PCR and sequencing methods followed those in Krumbholz et al. (2006). The DNA sequence of J/1025/05 was deposited in GenBank (acc. no. EF376992).

### 2.2. Phylogenetic analysis

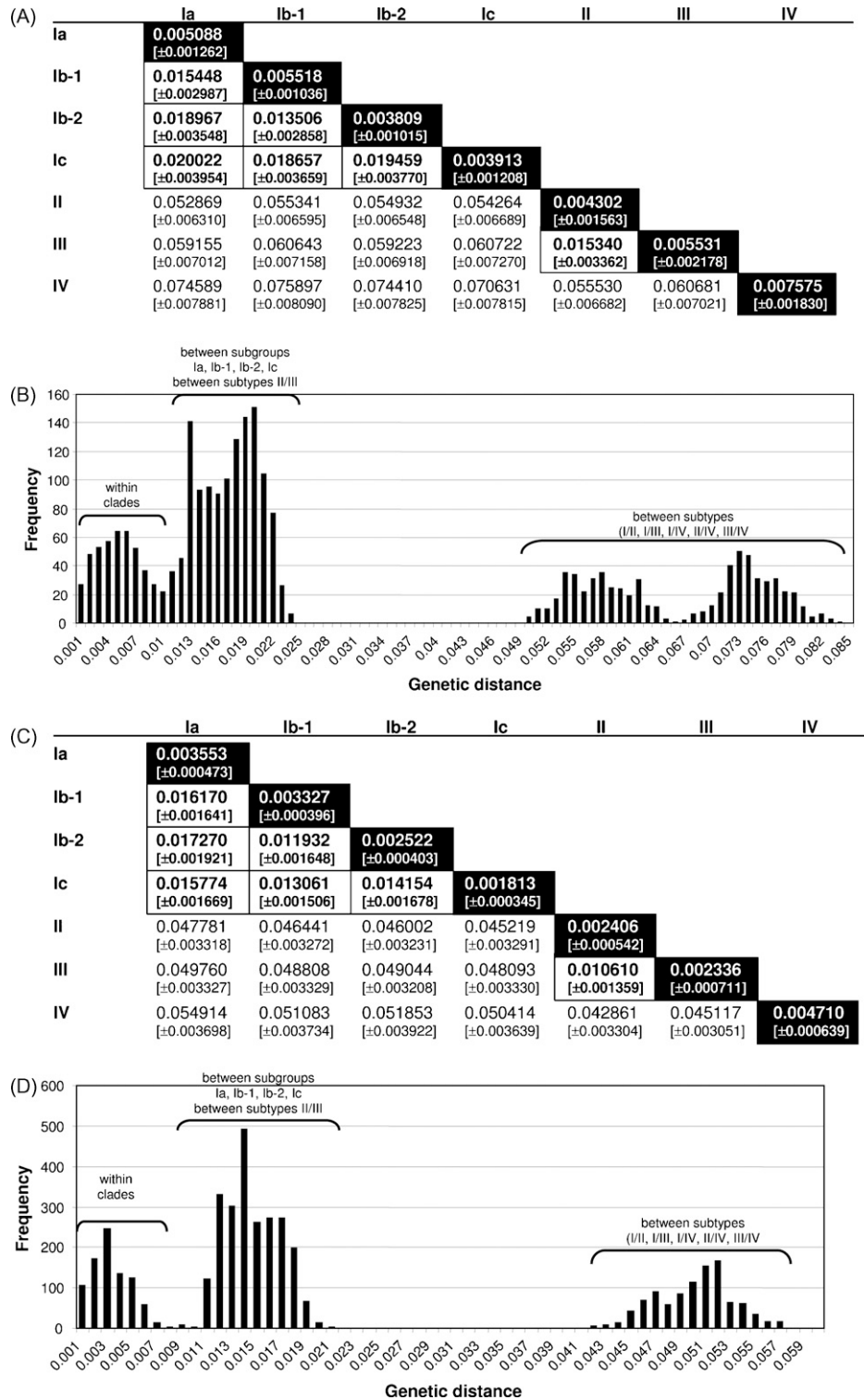
The complete genome sequence of J/1025/05, two partial VP1 sequences (reference strains IV and SB, GenBank acc. nos. Z19535 and Z19536) and 102 complete BKV genomes obtained from GenBank (valuation date 12 March 2007) were included in this study (listed in Table 1). For sequence comparisons, several data sets were analyzed. The entire coding region of the BKV genomes was aligned with ClustalW (Thompson et al., 1994) and improved by eye using MEGA 3.0 (Kumar et al., 2004). In addition, concatenated sequences (CS) comprising the major polyomavirus proteins T-Ag, t-Ag, VP1 and VP2 were also aligned. The VP3 gene, which has also often been examined in other studies, was not included here because it conforms to a truncated version of the VP2 gene that uses the same reading frame but another start codon;

therefore, it adds no non-redundant sequence information. Only unique BKV sequences ( $n=92$ ) were used for phylogenetic analyses. Sequence alignments of the VP1, VP2, T-Ag and t-Ag gene regions alone were also analyzed. In addition, sequences of a human JC virus isolate, GH-1 (AB038252), two simian polyomaviruses, SV40 (J02400) and SA-12 (DQ435829), as well as the murine pneumotropic virus (MPtV, NC\_001505) were included in the CS, VP1, VP2, T-Ag and t-Ag data sets for outgroup analysis, with the tree being rooted on MPtV. The entire coding region data set comprised only BKV sequences as some polyomaviruses apparently encode no agnoprotein and exhibit hypervariability in the intergenic region between the VP1 and the T-Ag gene.

The optimal nucleotide substitution model for the data sets was determined using MrAIC (Nylander, 2004) on the basis of the Akaike information criterion (AIC); for all data sets, the general time reversible model (Rodriguez et al., 1990) assuming a gamma distribution (GTR +  $\Gamma$ ) was indicated. In addition, the data sets for each gene were analyzed with ModelTEST/PAUP\* in order to test whether any genes evolved according to a strict molecular clock. ModelTEST selected GTR +  $\Gamma$  for VP1, TIM +  $\Gamma$ (clock) for VP2, GTR +  $\Gamma$  for T-Ag and TrN +  $\Gamma$ (clock) for t-Ag. Distance-based neighbor-joining (NJ) phylogenetic trees and unweighted maximum parsimony (MP) trees (summarized using fully resolved majority rule consensus) were calculated with PAUP\* v4.0b10 (Swofford, 2002). NJ analyses used GTR +  $\Gamma$  corrected distances. MP analyses employed a heuristic search strategy with 1000 random-addition sequences, steepest descent turned on, and a limit of 10,000 maxtrees. Maximum likelihood (ML) trees were inferred with RAxML-VI-HPC v2.2.3 (Stamatakis et al., 2005) under a GTR +  $\Gamma$  model using the default search parameters. Phylogenetic trees were also reconstructed employing Bayesian tree inference (BI) statistics as implemented in MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003) using a Metropolis-coupled Markov chain Monte Carlo (MC3) algorithm. Two runs each of four chains were run for 10,000,000 generations, and sampled every 5000 generations, discarding the first 50% of samples as burn-in. For the analyses of the CS data set, both the ML and BI computations used a partitioned model, whereby the model parameters were allowed to vary between the gene partitions. Nodal support within the trees was determined by bootstrap analyses (Felsenstein, 1985) based on 1000 replicates (NJ, MP, and ML) and identical search parameters to the main analyses (apart from only 100 random-addition sequences being used for the MP analyses), and by posterior probabilities derived from the post-burn-in generations using MrBayes. All trees were displayed with TreeView 1.6.6 (Page, 1996).

Pairwise genetic distances (number of different nucleotides per site) were calculated for VP1 as well as for the BKV coding region with TreePuzzle (Schmidt et al., 2002) using the GTR +  $\Gamma$  model. Because MEGA 3.0 does not implement the GTR +  $\Gamma$  model, mean genetic distances  $\pm$  standard error (S.E.) were computed using the Tamura-Nei (TN) model, the next most complex model (Kumar et al., 2004).

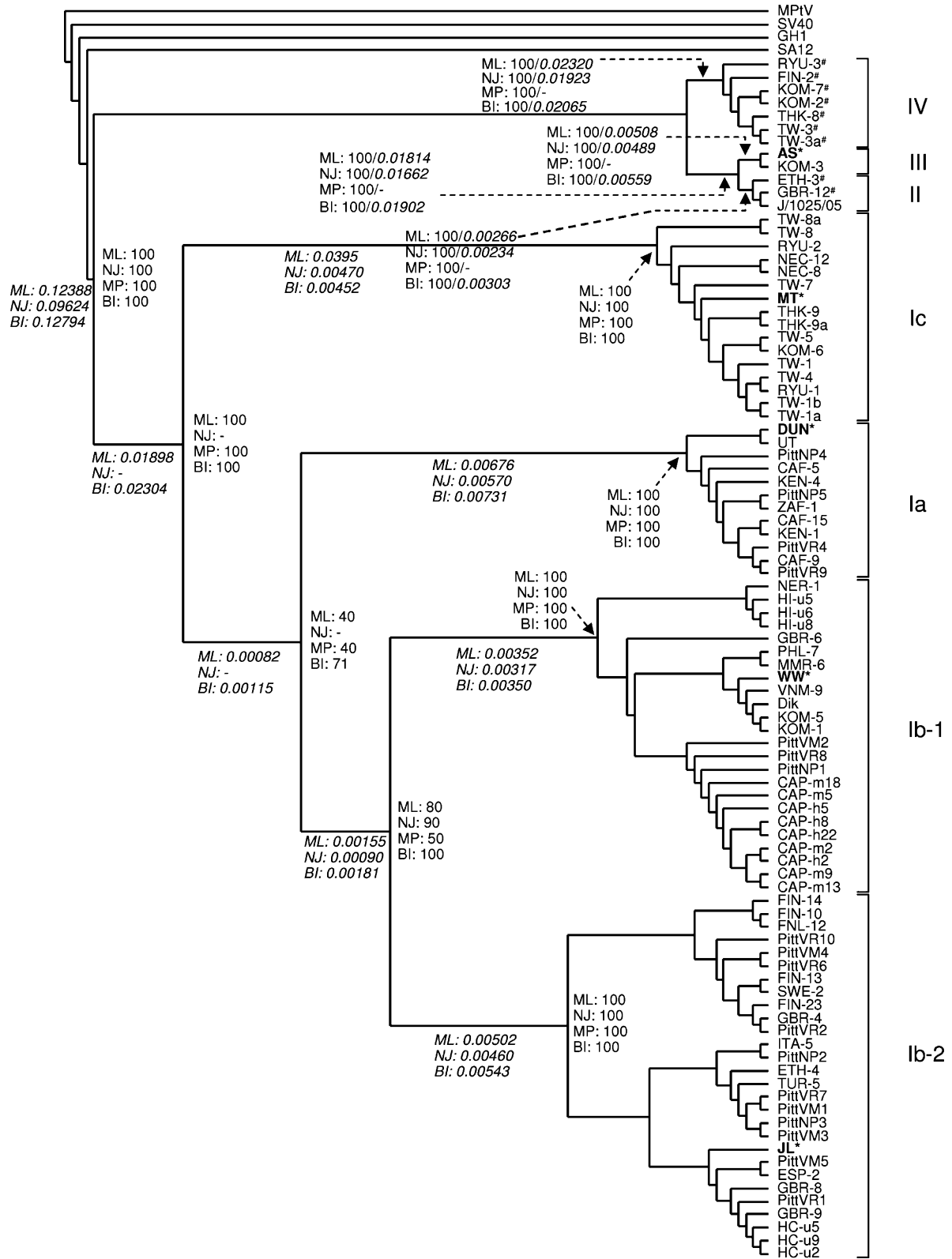
Divergence times within BKVs were inferred using Bayesian MCMC analysis and relaxed molecular clock models as implemented in BEAST 1.4.7 (Drummond and Rambaut, 2007). Data sets including BKV sequences (entire coding region, CS, VP1, VP2, T-Ag, t-Ag) and related polyomaviruses (JCV, SA-12, SV40, MPtV) were analyzed using the GTR +  $\Gamma$  substitution model, two partitions into codon positions (SRD06 model) for the CS, VP1, VP2, T-Ag and t-Ag data sets and three partitions for the entire coding region data set. A relaxed clock (uncorrelated lognormal) was assumed and the initial substitution rate was set to 1. Chain lengths varied from 10 to 100 million generations to yield significant results. Log parameters were sampled every 1000 generations.



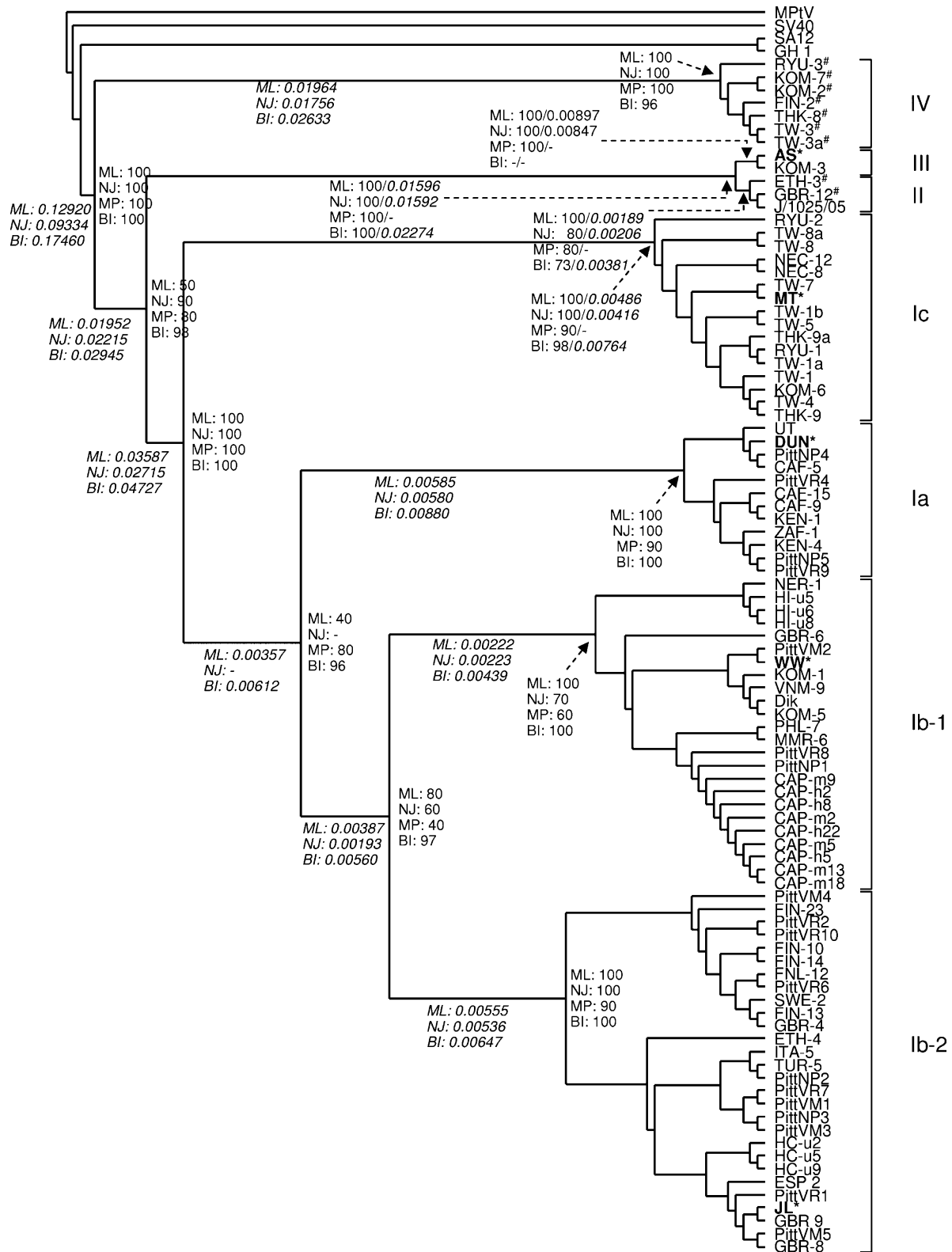
**Fig. 2.** Genetic relationships of BKV clades. Mean genetic distances  $\pm$  standard errors between and within clades (A, C) and frequency distribution of pairwise genetic distances (B, D) of 69 unique BKV VP1 sequences (A, B) and 92 unique BKV coding region sequences (C, D) computed using MEGA under the TN model. Mean values of intra-clade comparisons and comparisons between subtype I subgroups and between subtypes II and III are highlighted in black and white boxes, respectively. Pairwise genetic distances were computed with TreePuzzle using the GTR +  $\Gamma$  model.

Three different calibration points were examined independently. The first “external” calibration used a date of 23 million years to date the split between SV40 and the human polyomaviruses in line with the estimated time of divergence between their respective hosts, the Old World monkeys (excluding humans,

Cercopithecoidea) and humans (Hominoidea) (Glazko and Nei, 2003; Kumar and Hedges, 1998). This date is supported by several multigene studies, with an emphasis on nuclear markers: Kumar and Hedges (1998):  $23.3 \pm 1.2$  Myr, 56 nuclear genes; Glazko and Nei (2003): 23 Myr, range 21–25, determined with several different



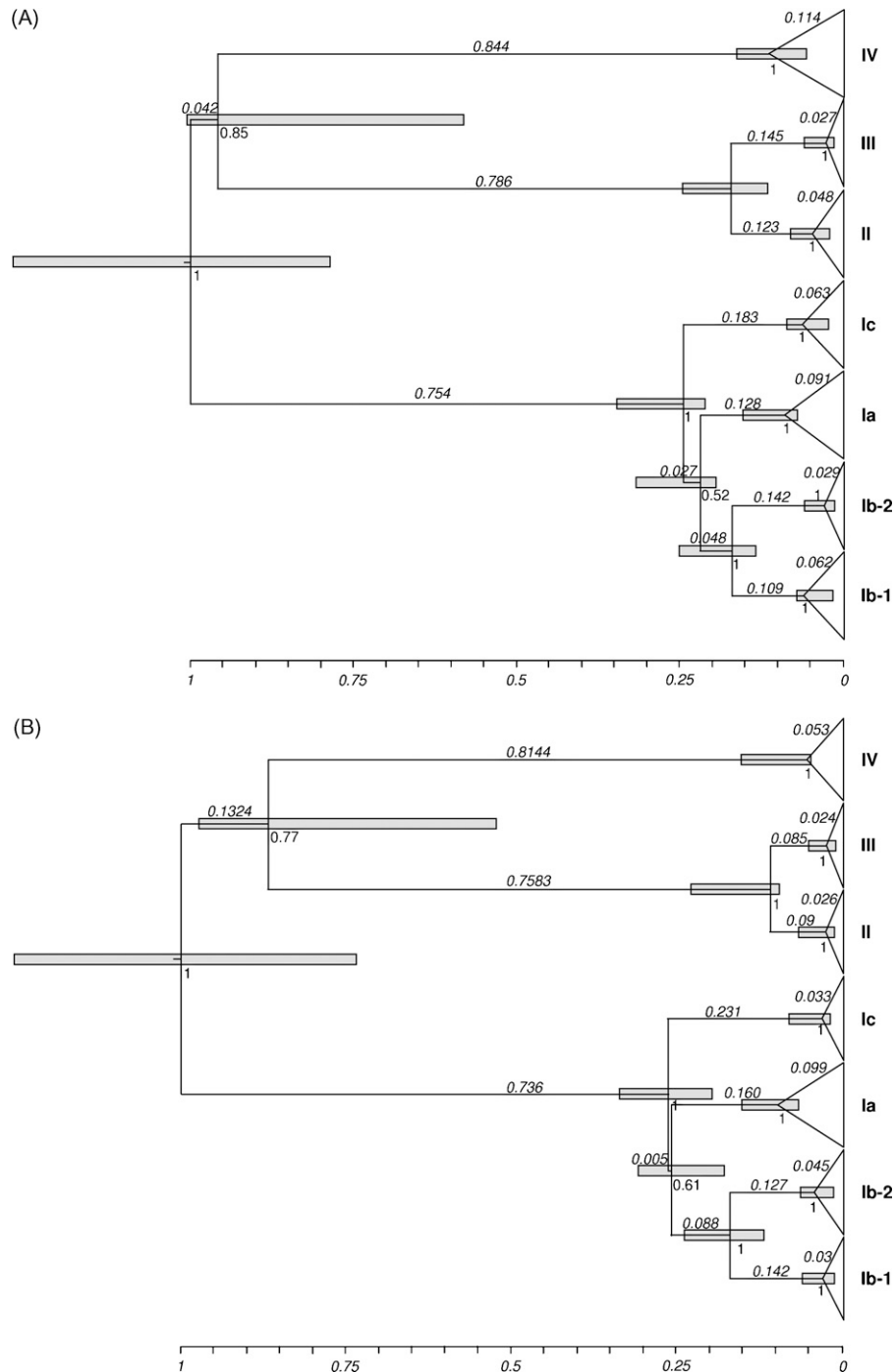
**Fig. 3.** Phylogenetic relationships of polyomavirus concatenated sequences (CS) as exemplified by the results of a ML analysis (sources in Table 1). Phylogenetic analysis reveals seven BKV clades. Branch lengths are indicated (in italics). The statistical support of node resolution based on bootstrap frequencies or posterior probabilities is given. \* indicates reference isolates used for genotyping. # indicates isolates that were genotyped using the 287-nt VP1 typing region of the respective reference strains (i.e., strain SB for genotype II and strain IV for genotype IV).



**Fig. 4.** Phylogenetic relationships of polyomavirus VP1 sequences as exemplified by the results of a ML analysis (sources in Table 1). Phylogenetic analysis reveals seven BKV clades. Branch lengths are indicated (in *italics*). The statistical support of node resolution based on bootstrap frequencies or posterior probabilities is given. \* indicates reference isolates used for genotyping. # indicates isolates that were genotyped using the 287-nt VP1 typing region of the respective reference strains (i.e., strain SB for genotype II and strain IV for genotype IV).

data sets; Hasegawa et al. (2003):  $25.5 \pm 2.7$  Myr, using the data set of Murphy et al. (2001) which consisted of 12 nuclear encoded protein genes and two mitochondrial ribosomal RNA gene segments from 64 species.

The two remaining calibrations were internal to BKV, dating initial BKV diversification (i.e., the separation of BKV subtypes IV and I, II, III) according to important events in human history, either the emergence of modern humans (Cann et al., 1987; McDougall



**Fig. 5.** Consensus tree of BKV sequences. The consensus tree of the entire coding region (A) and concatenated sequence data sets (B) under the relaxed clock analysis with a GTR +  $\Gamma$  substitution model. Node bars illustrate the 95% highest posterior density intervals for each divergence time. Relative branch times are presented (in *italics*). Numbers below nodes indicate posterior probabilities. The branches of subtypes and subgroups were collapsed.

et al., 2005) 200,000 years ago or the timing of their dispersal out of Africa 100,000 years ago as it was done for JCV (Hatwell and Sharp, 2000; Sugimoto et al., 2002; Takasaka et al., 2006b).

The evolution rates (substitutions/site/year) within BKV were calculated from the nucleotide diversity  $\Pi$  (estimated with DnaSP4, Rozas et al., 2003) and the mean rates which were determined with BEAST for each data set. Date estimates were visualized with FigTree v1.1.2 (Drummond and Rambaut, 2007).

### 3. Results

#### 3.1. Genetic relationship of J/1025/05 to the BKV subtypes

In a first set of phylogenetic analyses, unrooted trees were reconstructed using only BKV sequences to analyze the relationships of subtypes and subgroups. For these analyses, the genome of J/1025/05 was compared to 102 complete BKV sequences, including the recently described subtype II viruses GBR-12 and



ETH-3 (Zheng et al., 2007). The trees reconstructed using MP, ML and BI analyses of each of the VP1-encoding and entire coding regions (data not shown) were all consistent with the existence of seven BKV clades (i.e. “clans” sensu Wilkinson et al., 2007).

The mean distances  $\pm$  S.E. within and between these seven clades are presented in Fig. 2. For the VP1 encoding region, the intra-clade genetic distances ranged from 0.0038 to 0.0076. Values spanning from 0.052 to 0.076 were computed for the inter-clade genetic distances between BKV subtypes I/II, I/III, I/IV, II/IV, and III/IV, but not subtypes II/III. The genetic distances between the latter subtypes (0.013 to 0.020) were similar to those of the subtype I subgroup comparisons. For the entire coding region, the intra-clade genetic distances were lower, ranging from 0.0018 to 0.0047. Between subtype I subgroups and between subtypes II and III, the genetic distances ranged from 0.01061 to 0.01727. The distances between BKV subtypes I/II, I/III, I/IV, II/IV, and III/IV ranged from 0.042 to 0.055. Frequency plots of pairwise comparisons of 69 unique VP1 sequences and 92 unique coding region sequences, respectively, revealed a discontinuous frequency distribution of the genetic distances within and between the BKV subtypes and subgroups (Fig. 2).

### 3.2. Phylogenetic relationships and evolution of the BKV subtypes

Analysis of the VP1 and CS data sets revealed a consistent pattern regardless of the method of analysis (NJ, MP, ML or BI) with

13 major nodes being recovered that enable the distinction of seven BKV clades (see Figs. 3 and 4 which show the ML trees as an example). The only significant deviations from the general pattern were (i) alternative resolutions of the subtypes II/III/IV for the MP and ML analyses of the CS data set (Fig. 3), and (ii) for the subgroup Ia and Ic branches for the NJ analyses (data not shown).

In total, seven of eight computations (CS: MP, NJ, ML, BI; VP1: MP, NJ, ML) resolved the seven BKV clades with the eighth analysis (VP1: BI) resolving all but two (subtypes II and III being part of a polytomy). Most nodes were supported by the four methods with high bootstrap values or posterior probabilities (Figs. 3 and 4). Phylogenetic analyses of the T-Ag, t-Ag and VP2 ORFs yielded deviant tree topologies when compared to the CS and VP1 data sets (data not shown). The seven BKV clades were generally obtained in MP and ML trees from each data set, whereas only the NJ tree of the T-Ag data set yielded them. In all computations, SA-12 was always more closely related to BKV than to SV40 among the outgroup sequences (Figs. 3 and 4).

Divergence time estimates derived from the relaxed molecular clock analyses are presented in Fig. 5 and Table 2. Assuming the separation of SV40 and human polyomaviruses 23 Myr ago, JC virus (strain GH-1) and the ancestors of BKV were inferred to have separated some 13 Myr ago (CS data set, data not shown). The BKV radiation began 2.2 Myr ago, when subtype I and subtypes II/III/IV diverged and, at which time, the archetypal non-coding control

**Table 2**  
Evolution of BKV subtypes/subgroups and evolution rates within BKV

	Entire coding region	CS	VP1	VP2 <sup>a</sup>	T-Ag <sup>a</sup>
Mean rate	0.962	0.954	0.929	0.941	0.951
St. dev. of mean	0.001443	0.001598	0.000914	0.001507	0.001321
95% highest posterior density interval	0.817–1.068	0.796–1.08	0.709–1.078	0.734–1.076	0.767–1.098
Nucl. diversity II (substitutions/site)	0.02013	0.01914	0.02541	0.01523	0.02164
Evolution rate (substitutions/site/year)					
Hypothesis I	n.d.	8.1991E–9	2.0090E–8	3.7844E–9	7.8940E–9
Hypothesis II	9.6825E–8	9.1298E–8	1.1803E–7	7.1657E–8	1.0290E–7
Hypothesis III	1.9365E–7	1.8260E–7	2.3606E–7	1.4133E–7	2.0580E–7
Divergence events					
Relative node ages					
Separation of I and II/III/IV	1.0	1.0	1.0	1.0	1.0
Separation of II/III and IV	0.958	0.868	0.755	0.498	0.923
Separation of II and III	0.114	0.109	0.173	0.179	0.241
Separation of Ic and Ia/Ib	0.246	0.264	0.245	–	–
Separation of Ia and Ib	0.219	0.26	0.218	–	–
Separation of Ib-1 and Ib-2	0.171	0.172	0.15	0.146	0.303
Hypothesis I: coevolution of BKV with hominids (node ages in Myr)					
Separation of I and II/III/IV	n.d.	2.227	1.175	3.787	2.607
Separation of II/III and IV	n.d.	1.932	0.887	1.885	2.407
Separation of II and III	n.d.	0.244	0.203	0.677	0.629
Separation of Ic and Ia/Ib	n.d.	0.588	0.288	–	–
Separation of Ia and Ib	n.d.	0.578	0.256	–	–
Separation of Ib-1 and Ib-2	n.d.	0.383	0.176	0.551	0.791
Hypothesis II: evolution of modern human (node ages in Myr)					
Separation of I and II/III/IV	0.2	0.2	0.2	0.2	0.2
Separation of II/III and IV	0.192	0.174	0.151	0.1	0.185
Separation of II and III	0.034	0.022	0.035	0.036	0.048
Separation of Ic and Ia/Ib	0.049	0.053	0.049	–	–
Separation of Ia and Ib	0.044	0.052	0.044	–	–
Separation of Ib-1 and Ib-2	0.034	0.034	0.03	0.029	0.061
Hypothesis III: out-of-Africa model (node ages in Myr)					
Separation of I and II/III/IV	0.1	0.1	0.1	0.1	0.1
Separation of II/III and IV	0.096	0.087	0.075	0.05	0.092
Separation of II and III	0.017	0.011	0.017	0.018	0.024
Separation of Ic and Ia/Ib	0.025	0.026	0.025	–	–
Separation of Ia and Ib	0.022	0.026	0.022	–	–
Separation of Ib-1 and Ib-2	0.017	0.017	0.015	0.015	0.03

<sup>a</sup>n.d., not done. VP2 and T-Ag yielded different tree topologies for the subtype I evolution

region already had evolved. Diversification of BKV subtypes was completed with the emergence of subtypes II and III diverging 244,000 years ago at which point all seven BKV clades had evolved.

Much more recent divergence times were obtained by calibrating initial BKV diversification in line with either of the two human evolutionary events, each of which is an order of magnitude younger than the date inferred for this node using the external calibration point (2.2 Myr ago). When the separation of subtype I and subtypes II, III, IV was assumed to occur 200,000 years ago, the segregation of subtype IV and subtypes II, III is inferred to have occurred 174,000 years ago, with diversification of all BKV subtypes and subgroups having been completed 22,000 years ago. Assuming instead the out-of-Africa hypothesis (Cavalli-Sforza, 1998) for the calibration of BKV diversification, the evolution of the seven BKV clades started some 100,000 years ago and ended 11,000 years ago (Table 2, Fig. 5B). A similar divergence time estimate was obtained with the entire coding region data set (Table 2, Fig. 5A).

Across genes, inferred rates of evolution within BKV (obtained using BEAST) differed only by about a factor of 1–3× for any given calibration point, whereas differences were greater (6–22×) across calibration methods for a given gene (Table 2). Rate variation between genes can be explained by different selective forces acting on the capsid proteins (VP1, VP2, VP3), the DNA-binding protein T-Ag and other viral proteins (agno, t-Ag) (see below as well). All rates ranged between  $3.7844 \times 10^{-9}$  and  $2.3606 \times 10^{-7}$  substitutions/site/year. These values are considerably slower than the rate of 2 to  $5 \times 10^{-5}$  obtained by Chen et al. (2004) based on serial sampling from a single patient approximately 50 years after his primary infection, lending further support to suggestions that the latter rate is overinflated (Takasaka et al., 2006a).

## 4. Discussion

### 4.1. Phylogenetic relationship of J/1025/05 to the BKV subtypes

A previous serological study revealed four BKV serotypes (Knowles et al., 1989) and evidence exists that a characteristic amino acid stretch ranging from amino acids 61 to 83 of the VP1 gene is the molecular basis of the antigenic diversity of BKV (Jin et al., 1993b). Phylogenetic studies using distance-based NJ algorithms of a 287-nt subgenomic region of the VP1 gene revealed these same four genetic clusters (Jin, 1993; Jin et al., 1993a,b; Takasaka et al., 2004). Thus, a genetic subtyping scheme was proposed that correlates with the BKV serotypes. In addition, subgroups within subtype I (designated Ia, Ib, Ic) were distinguished on the basis of the 287-nt typing region. However, unambiguous assignment of several viral sequences to a subgroup failed using this short sequence (Krumbholz et al., 2006). ML analyses employing concatenated BKV sequences and single ORFs of additional clinical samples revealed two subgroups within group Ib, designated Ib-1 and Ib-2 (Nishimoto et al., 2006). Another study investigating the regional distribution of BKV lineages considered the coding region of up to 52 complete genomes and proposed seven major BKV clusters at the genome level using NJ tree inference (Nishimoto et al., 2006; Zheng et al., 2007).

In the present study, the data set comprised the complete genomes of 102 BKV strains retrieved from GenBank, containing 92 unique sequences representing all four recognized BKV subtypes. A variety of algorithms based on distances (NJ) or discrete criteria (MP, ML, BI) were applied. For VP1 as well as for concatenated sequences and the entire coding region, all models support seven BKV clusters (Figs. 3–5), with all calculations indicating BKV subtypes II and III to be sister groups. This close relationship is also supported by the mean genetic distance between the subtypes

being similar to those between the subgroups within subtype I (Fig. 2). Seven BKV clusters were also observed with VP2, T-Ag and t-Ag data sets although tree topologies varied significantly. The data demonstrate a reliable BKV typing and subgrouping based on the fast NJ algorithm if the complete VP1 ORF rather than the 287-nt typing region is investigated. This may improve diagnostic characterization of BKV.

Despite their genetic similarity, subtypes II and III are held to be serologically distinct as defined in a previous study using the prototype strains SB and AS (Knowles et al., 1989). Whereas a discontinuous frequency distribution of pairwise genetic distances proved helpful for species distinction in previous studies (Van Regenmortel et al., 1997), it fails to discriminate subtypes II and III of BKV, with our frequency plots demonstrating a clear clustering of the genetic distances between subtype I subgroups and also between subtypes II and III. Therefore, our data indicate that successful BKV subclassification requires the consideration of two criteria: (i) serological typing to distinguish among the four subtypes; alternatively, the 287-nt VP1 gene fragment can be sequenced as a proxy and (ii) further analyses that include at least sequencing of the complete VP1 gene. However, for precise and unambiguous subtyping and subgrouping, it appears to be necessary to study the entire BKV coding region or, minimally, the concatenated sequences of the major genes.

### 4.2. Evolution of BKV subtypes

The lack of a fossil record for viruses means that any divergence time estimates must instead be calibrated by other means. In principal, an internal calibration should yield optimal results. However, previous attempts to determine the rate of nucleotide substitution in BKV (e.g., Chen et al., 2004) using serial sampling have been controversial (Chen et al., 2004; Takasaka et al., 2006a). We would add that the value derived from Chen et al. (2004) is based on only four nucleotide substitutions over a period of 50 years. Both the total number of substitutions observed and the time scale are far too small, meaning that the inferred rate is liable to be affected severely by stochastic variation. In addition, accumulation of substitutions may not be the only explanation of the sequence heterogeneities upon serial sampling: dual or multiple BKV infections of the respective individuals cannot be excluded. Assuming an evolution rate of 2 to  $5 \times 10^{-5}$  as estimated by Chen et al. (2004), less than 1000 years are sufficient to yield nucleotide diversities ranging from 0.015 to 0.025 as observed for the seven BKV subtypes/subgroups (depending on the data set which included 69 to 92 unique sequences of the respective gene region; compare Table 2). Therefore, this approach appears to yield unreliable divergence time estimates.

One alternative is to assume a strict molecular clock, where the average rate of substitution is derived from other, closely related viruses. The obvious rate heterogeneity between viral groups, if not within a given viral genome itself, renders this procedure highly suspect in most cases, however. Another option presents itself when there is apparent phylogenetic codivergence between the viruses and their host species, namely to use the divergence times of the host species if they are known. This procedure is also not without its problems, with host switching and lineage sorting potentially disrupting strict phylogenetic codivergence. In fact, Perez-Losada et al. (2006) uncovered inconsistencies in their study of polyomaviruses that suggested that noncodivergence events might have occurred in the evolution of the murine pneumotropic virus MPtV, the baboon polyomavirus SA-12, the African green monkey polyomavirus and probably of BKV as well. Therefore, the SA-12 sequence appears to be unsuited for the calibration of the local molecular clock as was done previously (Nishimoto et al.,

2006). Indeed, evidence for noncodivergence throughout polyomaviruses is mounting: (i) the newly described human polyomaviruses WU and KI clearly cluster apart from the BK and JC polyomaviruses (Allander et al., 2007; Gaynor et al., 2007); (ii) likewise, phylogenetic relationships of polyomaviruses isolated from Old World monkeys, chimpanzee, a New World monkey and rodents reveal inconsistencies (Allander et al., 2007; Verschoor et al., 2008); and (iii) in contrast to previous reports, a recent study on the closely related JC virus provided no evidence for its codivergence with humans (Shackelton et al., 2006).

In light of these reports, the codivergence of SV40 (isolated from Rhesus monkey cells) and the human BKV with that of Old World monkeys (Cercopithecoidea) and hominids (Hominoidea) should be viewed cautiously. Even taking the 95% highest posterior density intervals into account, the time scale provided using this calibration point clearly dates the evolution and divergence of BKVs (2.2 Myr ago) far prior to the appearance and radiation of modern humans (0.195 Myr ago; McDougall et al., 2005), and does not offer an explanation for the observed relationships between BKV lineages and human populations (Zheng et al., 2007). There are several reports on the prevalence of BKV subtypes and subgroups (Agostini et al., 1995; Baksh et al., 2001; Chen et al., 2006; Di Taranto et al., 1997; Ikegaya et al., 2006; Jin et al., 1993a, 1995; Krumbholz et al., 2006; Takasaka et al., 2004), and the data indicate a complex regional distribution pattern. An even earlier timescale for BKV evolution was obtained by Nishimoto et al. (2006). On the basis of concatenated sequences of 28 BKV isolates representing subtypes I, III and IV and instead assuming BKVs and SA-12 to have separated 30 Myr ago, these authors found the separation of the subtypes (albeit as an unresolved polytomy) to have occurred 4.4 Myr ago, with the splitting of subtype I subgroups estimated at approximately 1.2 Myr ago. Both sets of analyses show the potential limitations in having to use very distant calibration points (23 or 30 Myr ago), and ones that are based on the divergence of a different viral species (either SV40 or SA-12). Thus, the clock-like analyses are unable to account for any extreme lineage-specific rate heterogeneity that might have occurred since this time and would confound the divergence-time inferences.

A final option is to use a calibration point internal to BKV. One such calibration point ties BKV diversification to the out-of-Africa migration of modern humans approximately 100,000 years ago in an analogous fashion to previous works on JCV (Hatwell and Sharp, 2000; Sugimoto et al., 2002; Takasaka et al., 2006b). This approach would yield evolution rates ranging from  $1.41 \times 10^{-7}$  to  $2.36 \times 10^{-7}$  depending on the data set. Such values are in the same order of magnitude as the evolution rates estimated for the human JC polyomaviruses ( $4 \times 10^{-7}$ , see Hatwell and Sharp, 2000; Sugimoto et al., 2002). A second possibility involves linking BKV diversification to the emergence of modern humans 200,000 years ago. Utilization of either calibration point resulted in the diversification of BKV subtypes II and III and subgroups of subtype I occurring clearly less than 50,000 years ago. This model appears to better explain the observed geographical prevalences of BKV subtypes and groups. However, the model again makes a strong, if more taxonomically restricted, assumption of codivergence and tends to be limited to the diversification of BKV subtypes and subgroups. In other words, the model probably cannot be applied reliably to earlier steps in human polyomavirus evolution for the same reasons as why an external calibration point tends to fail within BKV. Nevertheless, in the absence of a suitable and more recent external calibration point, the timescales from these calibration points remain our best current estimates for the evolution of BKV.

## Acknowledgements

The authors are grateful to Martina Müller for excellent technical assistance. OBE was supported through a Heisenberg Fellowship from the Deutsche Forschungsgemeinschaft (BI 825/2-1; BI 825/3-2).

## References

- Agostini, H.T., Brubaker, G.R., Shao, J., Levin, A., Ryschkewitsch, C.F., Blattner, W.A., Stoner, G.L., 1995. BK virus and a new type of JC virus excreted by HIV-1 positive patients in rural Tanzania. *Arch. Virol.* 140, 1919–1934.
- Allander, T., Andreasson, K., Gupta, S., Bjerkner, A., Bogdanovic, G., Persson, M.A.A., Dalianis, T., Ramqvist, T., Andersson, B., 2007. Identification of a third human polyomavirus. *J. Virol.* 81, 4130–4136.
- Baksh, F.K., Finkelstein, S.D., Swalsky, P.A., Stoner, G.L., Ryschkewitsch, C.F., Randhawa, P., 2001. Molecular genotyping of BK and JC viruses in human polyomavirus-associated interstitial nephritis after renal transplantation. *Am. J. Kid. Dis.* 38, 354–365.
- Cann, R.L., Stoneking, M., Wilson, A.C., 1987. Mitochondrial DNA and human evolution. *Nature* 325, 31–36.
- Cavalli-Sforza, L.L., 1998. The DNA revolution in population genetics. *Trends Genet.* 14, 60–65.
- Chauhan, S., Lecatsas, G., Harley, E.H., 1984. Genome analysis of BK (WW) viral DNA cloned directly from human urine. *Intervirology* 22, 170–176.
- Chen, Q., Zheng, H.Y., Zhong, S., Ikegaya, H., He, H.X., Wei, W., He, Y.Y., Kobayashi, N., Honjo, T., Takasaka, T., Takahashi, S., Kitamura, T., Yogo, Y., 2006. Subtype IV of the BK polyomavirus is prevalent in East Asia. *Arch. Virol.* 151, 2419–2429.
- Chen, Y., Sharp, P.M., Fowkes, M., Kocher, O., Joseph, J.T., Korolnik, I.J., 2004. Analysis of 15 novel full-length BK virus sequences from three individuals: evidence of a high intra-strain genetic diversity. *J. Gen. Virol.* 85, 2651–2663.
- Cole, C.N., 1996. Polyomavirinae: the viruses and their replication. In: Fields, B.N., Knipe, D.M., Howley, P.M. (Eds.), *Fundamental Virology*, vol. 3. Lippincott-Raven Publishers, Philadelphia.
- Coleman, D.V., Wolfendale, M.R., Daniel, R.A., Dhanjal, N.K., Gardner, S.D., Gibson, P.E., Field, A.M., 1980. A prospective study of human polyomavirus infection in pregnancy. *J. Inf. Dis.* 142, 1–8.
- Di Taranto, C., Pietropaolo, V., Orsi, G.B., Jin, L., Sinibaldi, L., Degener, A.M., 1997. Detection of BK polyomavirus genotypes in healthy and HIV-positive children. *Eur. J. Epidemiol.* 13, 653–657.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791.
- Gardner, S.D., Field, A.M., Coleman, D.V., Hulme, B., 1971. New human papovavirus (B.K.) isolated from urine after renal transplantation. *Lancet* 1, 1253–1257.
- Gaynor, A.M., Nissen, M.D., Whitley, D.M., Mackay, I.M., Lambert, S.B., Wu, G., Brennan, D.C., Storch, G.A., Sloots, T.P., Wang, D., 2007. Identification of a novel polyomavirus from patients with acute respiratory tract infections. *PLoS Pathog.* 3, e64.
- Glazko, G.V., Nei, M., 2003. Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* 20, 424–434.
- Goudsmit, J., Baak, M.L., Sletterus, K.W., Van der Noordaa, J., 1981. Human papovavirus isolated from urine of a child with acute tonsillitis. *Br. Med. J. (Clin. Res. Ed.)* 283, 1363–1364.
- Hasegawa, M., Thorne, J.L., Kishino, H., 2003. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet. Syst.* 78, 267–283.
- Hatwell, J.N., Sharp, P.M., 2000. Evolution of human polyomavirus JC. *J. Gen. Virol.* 81, 1191–1200.
- Hirsch, H.H., 2005. BK virus: opportunity makes a pathogen. *Clin. Infect. Dis.* 41, 354–360.
- Hirsch, H.H., Steiger, J., 2003. Polyomavirus BK. *Lancet Infect. Dis.* 3, 611–623.
- Hou, J., Jensen, P.J., Major, E.O., Zur Hausen, H.J., Almeida, J., Van der Noordaa, D., Walker, D., Lowy, D., Bernard, U., Butel, J.S., Cheng, D., Frisque, R.J., Nagashima, K., 2005. Polyomaviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desseisenberger, U., Ball, L.A. (Eds.), *Virus Taxonomy*. VIIIth Report of The International Committee on Taxonomy of Viruses. Academic Press, Amsterdam, pp. 231–238.
- Howley, P.M., Khoury, G., Byrne, J.C., Takemoto, K.K., Martin, M.A., 1975. Physical map of the BK virus genome. *J. Virol.* 16, 959–973.
- Ikegaya, H., Saukko, P.J., Terti, R., Metsarinne, K.P., Carr, M.J., Crowley, B., Sakurada, K., Zheng, H.Y., Kitamura, T., Yogo, Y., 2006. Identification of a genomic subgroup of BK polyomavirus spread in European populations. *J. Gen. Virol.* 87, 3201–3208.
- Jin, L., 1993. Rapid genomic typing of BK virus directly from clinical specimens. *Mol. Cell. Probe* 7, 331–334.
- Jin, L., Gibson, P.E., 1996. Genomic function and variation of human polyomavirus BK (BKV). *Rev. Med. Virol.* 6, 201–214.
- Jin, L., Gibson, P.E., Booth, J.C., Clewley, J.P., 1993a. Genomic typing of BK virus in clinical specimens by direct sequencing of polymerase chain reaction products. *J. Med. Virol.* 41, 11–17.

- Jin, L., Gibson, P.E., Knowles, W.A., Clewley, J.P., 1993b. BK virus antigenic variants: sequence analysis within the capsid VP1 epitope. *J. Med. Virol.* 39, 50–56.
- Jin, L., Pietropaolo, V., Booth, J.C., Ward, K.H., Brown, D.W., 1995. Prevalence and distribution of BK virus subtypes in healthy people and immunocompromised patients detected by PCR-restriction enzyme analysis. *Clin. Diagn. Virol.* 3, 285–295.
- Knowles, W., Gibson, P.E., Gardner, S.D., 1989. Serological typing scheme for BK-like isolates of human polyomavirus. *J. Med. Virol.* 28, 118–123.
- Knowles, W.A., 2001. The epidemiology of BK virus and the occurrence of antigenic and genomic subtypes. In: Khalili, K., Stoner, G.L. (Eds.), *Human Polyomavirus: Molecular and Clinical Perspectives*. Wiley, New York, pp. 359–408.
- Krumbholz, A., Zell, R., Egerer, R., Sauerbrei, A., Helming, A., Gruhn, B., Wutzler, P., 2006. Prevalence of BK virus subtype I in Germany. *J. Med. Virol.* 78, 1588–1598.
- Kumar, S., Hedges, S.B., 1998. A molecular timescale for vertebrate evolution. *Nature* 392, 917–920.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinf.* 5, 150–163.
- Mannon, R.B., 2004. Polyomavirus nephropathy: what have we learned? *Transplantation* 77, 1313–1318.
- McDougall, I., Brown, F.H., Fleagle, J.G., 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* 433, 733–736.
- Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294, 2348–2351.
- Nishimoto, Y., Takasaka, T., Hasegawa, M., Zheng, H.Y., Chen, Q., Sugimoto, C., Kitamura, T., Yogo, Y., 2006. Evolution of BK virus based on complete genome data. *J. Mol. Evol.* 63, 341–352.
- Nylander, J.A.A., 2004. MrAIC.pl. In: *Evolutionary Biology Centre, Uppsala University*, City, p. Program distributed by the author.
- Page, R.D.M., 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12, 357–358.
- Pauw, W., Choufoer, J., 1978. Isolation of a variant of BK virus with altered restriction endonuclease pattern. *Arch. Virol.* 57, 35–42.
- Perez-Losada, M., Christensen, R.G., McClellan, D.A., Adams, B.J., Viscidi, R.P., Demma, J.C., Crandall, K.A., 2006. Comparing phylogenetic codivergence between polyomaviruses and their hosts. *J. Virol.* 80, 5663–5669.
- Rodriguez, F., Oliver, J.L., Marin, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142, 485–501.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.
- Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497.
- Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502–504.
- Seif, I., Khoury, G., Dhar, R., 1979. The genome of human papovavirus BKV. *Cell* 18, 963–977.
- Shackelton, L.A., Rambaut, A., Pybus, O.G., Holmes, E.C., 2006. JC virus evolution and its association with human populations. *J. Virol.* 80, 9928–9933.
- Sharma, P.M., Gupta, G., Vats, A., Shapiro, R., Randhawa, P., 2006. Phylogenetic analysis of polyomavirus BK sequences. *J. Virol.* 80, 8869–8879.
- Stamatakis, A., Ludwig, T., Meier, H., 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21, 456–463.
- Sugimoto, C., Hara, K., Taguchi, F., Yogo, Y., 1989. Growth efficiency of naturally occurring BK virus variants in vivo and in vitro. *J. Virol.* 63, 3195–3199.
- Sugimoto, C., Hasegawa, M., Kato, A., Zheng, H.Y., Ebihara, H., Taguchi, F., Kitamura, T., Yogo, Y., 2002. Evolution of human polyomavirus JC: implications for the population history of humans. *J. Mol. Evol.* 54, 285–297.
- Sugimoto, C., Kitamura, T., Guo, J., Al-Ahdal, M.N., Shchelkunov, S.N., Otova, B., Ondrejka, P., Chollet, J.Y., El-Safi, S., Ettayebi, M., Gresenguet, G., Kocagoz, T., Chaiyarasamee, S., Thant, K.Z., Thein, S., Moe, K., Kobayashi, N., Taguchi, F., Yogo, Y., 1997. Typing of urinary JC virus DNA offers a novel means of tracing human migrations. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9191–9196.
- Swofford, D.L., 2002. PAUP\*. *Phylogenetic Analysis Using Parsimony (\* and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts, City.
- Takasaka, T., Goya, N., Ishida, H., Tanabe, K., Toma, H., Fujioka, T., Omori, S., Zheng, H.Y., Chen, Q., Nukuzuma, S., Kitamura, T., Yogo, Y., 2006a. Stability of the BK polyomavirus genome in renal-transplant patients without nephropathy. *J. Gen. Virol.* 87, 303–306.
- Takasaka, T., Goya, N., Tokumoto, T., Tanabe, K., Toma, H., Ogawa, Y., Hokama, S., Momose, A., Funyu, T., Fujioka, T., Omori, S., Akiyama, H., Chen, Q., Zheng, H.Y., Ohta, N., Kitamura, T., Yogo, Y., 2004. Subtypes of BK virus prevalent in Japan and variation in their transcriptional control region. *J. Gen. Virol.* 85, 2821–2827.
- Takasaka, T., Kitamura, T., Sugimoto, C., Guo, J., Zheng, H.Y., Yogo, Y., 2006b. Phylogenetic analysis of major African genotype (Af2) of JC virus: implications for origin and dispersals of modern Africans. *Am. J. Phys. Anthropol.* 129, 465–472.
- Tavis, J.E., Walker, D.L., Gardner, S.D., Frisque, R.J., 1989. Nucleotide sequence of the human polyomavirus AS virus, an antigenic variant of BK virus. *J. Virol.* 63, 901–911.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- Van Regenmortel, M.H., Bishop, D.H., Fauquet, C.M., Mayo, M.A., Maniloff, J., Calisher, C.H., 1997. Guidelines to the demarcation of virus species. *Arch. Virol.* 142, 1505–1518.
- Verschoor, E.J., Groenewoud, M.J., Fagrouch, Z., Kewalapat, A., van Gessel, S., Kik, M.J., Heeney, J.L., 2008. Molecular characterization of the first polyomavirus from a New World primate: squirrel monkey polyomavirus. *J. Gen. Virol.* 89, 130–137.
- Weinreb, D.B., Desman, G.T., Amolat-Apiado, J.M., Burstein, D.E., Godbold, J.H., Johnson, E.M., 2006. Polyoma virus infection is a prominent risk factor for bladder carcinoma in immunocompetent individuals. *Diagn. Cytopathol.* 34, 201–203.
- Wilkinson, M., McLnerney, J.O., Hirt, R.P., Foster, P.G., Embley, T.M., 2007. Of clades and clans: terms for phylogenetic relationships in unrooted trees. *Trends Ecol. Evol.* 22, 114–115.
- Zheng, H.Y., Nishimoto, Y., Chen, Q., Hasegawa, M., Zhong, S., Ikegaya, H., Ohno, N., Sugimoto, C., Takasaka, T., Kitamura, T., Yogo, Y., 2007. Relationships between BK virus lineages and human populations. *Microbes Infect.* 9, 204–213.