

## Phylogenetic Analysis of Polyomavirus BK Sequences

Preety M. Sharma,<sup>1</sup> Gaurav Gupta,<sup>2</sup> Abhay Vats,<sup>2</sup> Ron Shapiro,<sup>3</sup> and Parmjeet Randhawa<sup>1\*</sup>

*Division of Transplant Pathology, Department of Pathology,<sup>1</sup> and Division of Transplant Surgery, Department of Surgery,<sup>3</sup> University of Pittsburgh Medical Center, and Department of Pediatrics, Children's Hospital of Pittsburgh,<sup>2</sup> Pittsburgh, Pennsylvania*

Received 11 March 2006/Accepted 23 June 2006

**Polyomavirus BK (BKV) has emerged as an important pathogen in kidney transplant patients. Existing taxonomic classifications of BKV come from conventional DNA sequence alignments based on limited data derived from the VP1 gene. We have used a phylogenetic whole-genome approach to examine the pattern of diversity and evolutionary relationships between 45 BKV strains isolated from multiple clinical settings. This analysis supports the classification of BKV into six genotypes, of which types V and VI have not been previously recognized. BKV strains hitherto classified as type I are, in fact, quite heterogeneous, and several cluster with our newly defined genotypes V and VI. The sequence information needed for assigning genotypes can be captured by VP1, VP2, VP3, or large T-gene sequencing. The most polymorphic coding region in the viral genome is VP1, but significant variation is also present in the large T-antigen gene, wherein polymorphisms are found in 11.39% of all nucleotide sites, 46.22% of which are cluster specific. Type-specific amino acid changes within the VP1 region are predicted to map to the BC and DE loops. The number of taxonomically informative amino acid changes in the large T antigen exceeds even that of the VP1 region. Viral strains isolated from healthy subjects and from patients with human immunodeficiency virus infection, Wiskott-Aldrich syndrome, and vasculopathy with capillary leak syndrome formed distinct subclusters. However, within the kidney transplant population, BKV strains derived from patients with asymptomatic viruria did not show complete separation from strains associated with allograft nephropathy.**

Polyomavirus BK (BKV) has a worldwide seroprevalence of 60 to 80% in humans (10). Primary infection occurs in childhood (11) and leads to viral latency in the urogenital tract and mononuclear cells. Following reactivation, which mostly occurs in immunocompromised patients, the virus is excreted in the urine (5). BKV viruria in renal transplant recipients ranges between 10 to 60% and has been associated with ureteric stenosis (16) and BK viral nephropathy characterized by tubular necrosis and interstitial nephritis (32, 33). In bone marrow transplant recipients, BKV has been linked to hemorrhagic cystitis (13, 26, 34).

Characterization of genetic diversity of BK virus has biologic as well as clinical implications. This information is needed to seek potential relationships between viral genotype and clinical disease. Sequence data are required by diagnostic virology laboratories to ensure that primers and probes being used for amplification of viral DNA can successfully detect all naturally occurring viral strains. One of the most variable regions of the viral genome is the non-coding control region (NCCR), which shows insertions, deletions, duplications, and complex rearrangements involving enhancer and/or promoter sequences (28, 29). Significant variation is also recognized in the VP1 gene, which codes for VP1, a major structural protein that comprises approximately 80% of the total viral capsid protein. The VP1 protein bears important domains which interact with viral receptors on host cells. A single amino acid change in VP1 protein has been shown to result in increased pathogenicity

of mouse polyomavirus (15). Based on VP1 sequence data and restriction enzyme analysis, Jin et al. (20) designed a typing scheme whereby all the existing BKV isolates are classified into four genotypes, which have been designated types I, II, III, and IV. These genotypes correlated well with the serotypes defined by Knowles et al. (24). Knowledge of genetic variation in other regions of the viral genome is limited, even though a number of other functionally important proteins are coded by the virus. Thus, agnoprotein seems to be involved in multiple functions, including enhancing nuclear localization of VP1, viral capsid assembly, virion release, and DNA repair (22). VP2 and VP3 contribute to the scaffolding of the viral capsid architecture. The T antigen promotes viral replication, binds to tumor suppressor proteins Rb and p53, and stimulates host cell entry into the cell cycle (19).

There is only limited BKV whole-genome information published in the literature. Until recently, only three BKV whole genomic sequences were available, namely those from the MM, Dun, and AS strains. In 2004, Chen et al. published 15 whole-genome sequences derived by cloning from three individuals (8). We have now sequenced the complete genomes of BKV isolates from 20 renal transplant recipients. These whole genomic sequences and additional partial sequences published in the literature have been analyzed using phylogenetic methods. The principal aims of our study were to (i) determine the most informative viral genomic regions from a phylogenetic standpoint, (ii) define clades for epidemiological studies incorporating additional sequence data that became available after the publication of Jin et al.'s genotyping schema, and (iii) seek evolutionary relationships between BKV strains isolated from different geographical locations and clinical settings, including healthy individuals, pregnant women, renal transplant patients with asymptomatic viruria or BKV nephropathy, bone marrow

\* Corresponding author. Mailing address: Division of Transplant Pathology, E737 UPMC-Montefiore Hospital, 3459 Fifth Ave., Pittsburgh, PA 15213. Phone: (412) 647-7646. Fax: (412) 647-5237. E-mail: randhawapa@msx.upmc.edu.

TABLE 1. Sequencing primers for the BKV genome

Name	Position	Strand orientation	Primer sequence (5' to 3')
SeqB-1	511–487	Reverse	CTCTACAAAATTCCAGCAAAAAGCTC
SeqB-2	517–542	Forward	GACAGTGTAGACGGGAAAAACAAAAG
SeqB-3	810–831	Forward	CTAACTCCTCAAACATATGCTG
SeqB-4	871–850	Reverse	GCAGCAAACCCAGCAATAGCCC
SeqB-5	1154–1177	Forward	CTCACAGGAATTGCAGAGAAGAAC
SeqB-6	1327–1306	Reverse	TCAGCTACTTGTCTAACCATTG
SeqB-7	1484–1505	Forward	AAGAAGTCTCCTCAATGGATG
SeqB-8	1820–1796	Reverse	AGCATTTTTCTCTCTGGGCTATCAC
SeqB-9	1824–1847	Forward	CTGTTACAGCACAGCAAGAATTCC
SeqB-10	2102–2124	Forward	CTAAAAACCCAACAGCCCAGTCC
SeqB-11	2246–2222	Reverse	CCTCCTGTGAAAGTCCAAAATACC
SeqB-12	2330–2351	Forward	AAGCTGATAGCCTGTATGTTTC
SeqB-13	2575–2554	Reverse	TGCCATCAAACACCCTAACCTC
SeqB-14	2617–2639	Forward	GACAAACAGGGACAATTGCAAAC
SeqB-15	2861–2839	Reverse	AATCACAATGCTCTTCCCAAGTC
SeqB-16	2993–3015	Forward	TCCAGCCTTTCCTCCATTCAAC
SeqB-17	3224–3203	Reverse	CAATGAATGAGTATCCTGTCC
SeqB-18	3470–3491	Forward	CCACCACACAAATCTAATAACC
SeqB-19	3738–3717	Reverse	ACCAGGGAAGAAATGCTAACAG
SeqB-20	3960–3981	Forward	ACTTTGTCTCTACTGCATACTC
SeqB-21	4102–4081	Reverse	TGCCTTAACTAGATCCATAC
SeqB-22	4317–4338	Forward	TATACACAGCAAAGCAGGCAAG
SeqB-23	4448–4426	Reverse	ATTCTCAACACTCAACACCACCC
SeqB-24	4748–4726	Reverse	TGCTACTGCATTGACTGCTTCAC
SeqB-25	4733–4756	Forward	AGTCAATGCAGTAGCAATCTATCC
SeqB-26	4971–4948	Reverse	AATGGAGCAGGATGTAAAGGTAGC
SeqB-27	4996–5017	Forward	TTCATTTTATCCTCGTCGCCCC
SeqB-28	28–52	Forward	ATTTCCCAAAATAGTTTTGTAGGC
SeqB-29	128–106	Reverse	AATATATAAGAGGCCCTGAGCCGC
SeqB-30	332–308	Reverse	TGTCTGTCATGCACTTTCCTCCTG

transplant recipients with viruria or hemorrhagic cystitis, and systemic lupus erythematosus (SLE).

#### MATERIALS AND METHODS

**Clinical material.** Urine samples were collected with informed consent from 42 adult renal transplant (RT) recipients at the University of Pittsburgh Medical Center (UPMC), Pennsylvania, between 2001 and 2004. The sample collection protocol was approved by the University of Pittsburgh Institutional Review Board (protocol 000586). Patients typically received pretransplant induction with thymoglobulin or anti-CD52 antibody (Campath) followed by Tacrolimus monotherapy. All patients were regularly monitored for BKV replication in their urine and plasma using quantitative PCR as previously described (31). Based on the quantitative PCR results, these patients were classified as having BK viruria or BK viremia. All viremic patients were also viruric. In cases where it was clinically indicated, a rise in serum creatinine led to a renal allograft biopsy to rule out BKV nephropathy. The diagnosis of nephropathy was based on the presence of intranuclear viral inclusions in the tubular epithelium and confirmation of BKV DNA by *in situ* hybridization.

**Long-range PCR.** DNA was extracted from 5 ml of urine using the QIAGEN Maxi kit (Valencia, CA) and eluted in a final volume of 200  $\mu$ l of buffer AE, supplied by the manufacturer. The entire circular genome of BKV was then amplified in a single long-range PCR using a strategy that has been successful for JC virus (1). The protocol used was developed by the late Gerald Stoner at the National Institutes of Health. DNA was first digested by 1 U of BamHI (catalog no. R6021; Promega, Madison, WI) in a 10- $\mu$ l reaction containing 3  $\mu$ l of DNA and the supplied reaction buffer. The reaction mix was incubated at 37°C for 30 min, followed by 65°C for 20 min and then a 4°C soak. The BamHI digest was then used to amplify the full-length linearized BKV genome using 6 pmol each of primers which overlapped the BamHI site. The primer sequences were the following: BBam 1, GGG ATC CAG ATG AAA ACC TTA GGG G; BBam 2, TGG ATC CCC CAT TTC TGG GTT TAG G. Amplification of BamHI-digested DNA was performed in a 50- $\mu$ l reaction using *rTth* polymerase-based long-range PCR (catalog no. N808-0188; Applied Biosystems, Foster City, CA). PCR conditions consisted of a first denaturation step of 94°C for 4 min, 14 cycles of denaturation at 94°C for 40 s, annealing at 64°C for 6 min, 5 cycles of

denaturation at 94°C for 40 s, annealing at 64°C for 8 min, 5 cycles of denaturation at 94°C for 40 s, annealing at 64°C for 10 min, 5 cycles of denaturation at 94°C for 40 s, annealing at 64°C for 12 min, 5 cycles of denaturation at 94°C for 40 s, annealing at 64°C for 14 min, and a final elongation period at 72°C for 10 min, followed by a soak at 4°C. Cleanup of the long-range PCR product was performed using Millipore PCR Cleanup columns (catalog no. UFC7PCR50; Billerica, MA).

**DNA sequencing.** The long-range PCR product (30 to 90 ng of DNA) was then subjected to a cycle sequencing reaction using Big Dye chemistry (Big Dye Terminator Cycle Sequencing kit, v. 3.1; Applied Biosystems Inc., Foster City, CA) and 3.2 pmol of the appropriate forward or reverse sequencing primer. A set of 30 sequencing primers spanning the entire viral genome was designed for this project (Table 1). PCR conditions consisted of a first denaturation step of 96°C for 2 min, followed by 25 cycles of denaturation at 96°C for 30 s, annealing at 50°C for 15 s, and extension at 60°C for 4 min. Cleanup of the cycle sequencing PCR products was performed using CentriSep 96-well plates (catalog no. CS-961; Princeton Separations Inc., Adelphia, NJ) using the manufacturer's instructions. DNA sequencing was performed at the University of Pittsburgh Genomics and Proteomics Core Facility using an ABI 310 automated sequencer. Sequences were trimmed, analyzed, and assembled using Sequencher 4.2 (Gene Codes Corporation, Ann Arbor, MI). All base calls were verified manually and compared to the Dun reference sequence (GenBank accession no. V01108). The presence of nucleotide polymorphisms was accepted only if the chromatogram reading was unambiguous and the observed changes occurred in more than one overlapping sequence at that nucleotide position. All forward primers were run in duplicate to ensure multiple coverage of polymorphic sites. This paradigm corrects for the possibility of sequencing errors being introduced during PCR as a result of infidelity in DNA polymerase.

**Retrieval of published BKV sequences.** Twenty-five full-length BKV genome sequences were available from GenBank (Table 2), which included three reference strains, namely, Dunlop (Dun) (36), MM (46), and AS (44), nine cloned sequences from a patient with BKV vasculopathy (CAP), three clones from a human immunodeficiency virus type 2 (HIV-2)-positive patient (HI), three clones from a healthy control subject (HC) (8), six clones from renal transplant patients (42), and one clone from strain UT (GenBank accession no.

TABLE 2. BKV strains phylogenetically analyzed in this study

Sequence <sup>a</sup>	Location	Isolate name; no. of samples <sup>b</sup>	Illness; sample origin <sup>c</sup>	Cell culture status/clone status	Reference or source	Accession no.	
WG	United States	Dunlop	WA; urine	Yes/no	36	V01108	
	United States	MM	WA; urine and brain	Yes/unknown	46	V01109	
	United States	PittVR; 10	RT; urine	No/no	This study	This study	
	United States	PittNP; 5	RT; urine	No/no	This study	This study	
	United States	PittVM; 5	RT; urine	No/no	This study	This study	
	United States	CAP; 9	Capillary leak syndrome; muscle, heart	No/yes	8	AY628224, AY628226-AY628233	
	United States	HC; 3	Healthy; urine	No/yes	8	AY628234-AY628236	
	West Africa	HI; 3	HIV-2; urine	No/yes	8	AY628225, AY628237-AY628238	
	United Kingdom	AS	Pregnant; urine	Yes/yes	44	M23122	
	Japan	TW,THK; 6	RT; urine	No/yes	42	AB217917-AB217921, AB213487	
	United States	UT	Solid tumor; urine	Yes/yes	GenBank	DQ305492	
	VP1	England	IV	RT; urine	Yes/yes	20	Z19535
		England	GS	RT; urine	Yes/yes	20	Z19537
		Japan	MT	SLE; urine	No/yes	39	X56911
		Japan	KOM; 31	BMT; urine	No/yes	43	AB181539-AB181569
Japan		THK, TW, TU; 29	RT; urine	No/yes	43	AB181570-AB181598	
The Netherlands		Dik	Acute tonsillitis; urine	Yes/yes	40	X56912	
The Netherlands		JL	BMT; urine	Yes/yes	40	X56914	
South Africa		WW	RT; urine	No/yes	40	X56913	
Spain		BCN	Sewage	No/no	3	AF120244	
Spain		BCNU	Pregnant; urine	No/no	3	AF120245	
South Africa		PRETORIA1	Sewage	No/no	3	AF120246	
France		NANCY2	Sewage	No/no	3	AF120247	
VP1		Unspecified	160998	Sewage	No/no	4	AF356534
		South Africa	SA17	Sewage	No/no	4	AF356535
		United States	USA3	Sewage	No/no	4	AF356536
	Unspecified	RA	Sewage	No/no	4	AF356537	
	Spain	BCN10	Sewage	No/no	4	AF356543	
	England	SB	Lymphoma; urine	Yes/yes	20	Z19536	

<sup>a</sup> WG, whole genome of BK virus; VP1, major capsid protein of BK virus.

<sup>b</sup> Original isolate names are preserved wherever applicable.

<sup>c</sup> BMT, bone marrow transplant; CAP, capillary leak syndrome; RT, renal transplant; SLE, systemic lupus erythematosus; WA, Wiskott-Aldrich syndrome.

DQ305492). Partial sequences from the VP1 regions were obtained from a variety of sources, as listed in Table 2.

**Phylogenetic analysis.** Analysis was carried out for whole-genome sequences as well as sequences derived from agnogene, VP1, VP2, VP3, and the T-antigen region. Sequence alignment was performed with CLUSTAL W (45) at the EMBL-EBI website (<http://www.ebi.ac.uk/clustalw/>) (9) using default parameters, followed by manual adjustment using known landmarks in the viral genome. Sequences were numbered using the system of Seif et al., in which nucleotide position 1 is assigned to the nucleotide adjacent to the start codon of the T antigen (36). Following established conventions, the entire NCCR until the start codon of the agnogene was excluded from phylogenetic analysis for the following reasons: (i) the presence of insertions, deletions, and rearrangements make accurate alignment in this region difficult, (ii) it is recognized that rearrangement patterns in the NCCR do not correlate with conventional viral genotypes (21), and (iii) when large indels occur, one cannot decide if they represent single or multiple mutational events (14). Neighbor-joining (NJ) (35), maximum parsimony (MP), and unweighted pair-group method using arithmetic averages (UPGMA) (37) trees were constructed using MEGA version 3.1 (25). Complete or pairwise deletion of gaps obtained from whole-genome analysis did not alter the topology. Divergences were estimated with Kimura's two-parameter method. MP trees were based on the close-neighbor-interchange paradigm. The initial tree was constructed using random addition with 10 replicates. All phylogenetic trees were visualized using MEGA 3.1 Tree explorer (25). A bootstrap test with 1,000 replicates was used to estimate the confidence of the branching patterns of the trees by all three methods (12).

**RESULTS**

**Whole-genome sequences.** The whole-genome sequence analysis was based on 45 full-length BKV DNA sequences,

including 20 sequences from Pittsburgh and 25 from GenBank. After exclusion of the NCCR, the alignment consisted of 4,771 nucleotide sites, starting at the initiation codon of the agnogene protein gene. Ninety-one percent of the aligned positions were invariant. The initial Clustal alignment needed manual realignment for the MM sequence, since the published MM sequence begins at position 1842 in the VP1 coding region, unlike all other sequences that begin at the first base after the start codon of the T antigen. The final alignment required introduction of four gaps, which corresponded to the following extra nucleotides in the AS strain sequence: GC at positions 2680 and 2681 (VP1-T-antigen intergenic region) and AA at positions 4584 and 4585 (T-antigen intron).

Phylogenetic trees were made using MP, NJ, and UPGMA methods (Fig. 1 to 3). MP analysis was based on a total of 4,465 sites, of which 233 were parsimony informative. The MP original tree had a consistency index of 0.85, retention index of 0.93, and rescaled consistency index of 0.79. An MP tree constructed from the 45 aligned complete sequences is shown in Fig. 1. It is a bootstrap consensus tree derived from 31 equally parsimonious trees calculated by MEGA 3.1. There are six major clusters supported by bootstrap values of >50%. For simplicity, these six clusters are designated A, B, C, D, E, and F. To eliminate the possibility that the tree structure was skewed by inclusion of multiple isolates from the same patient,

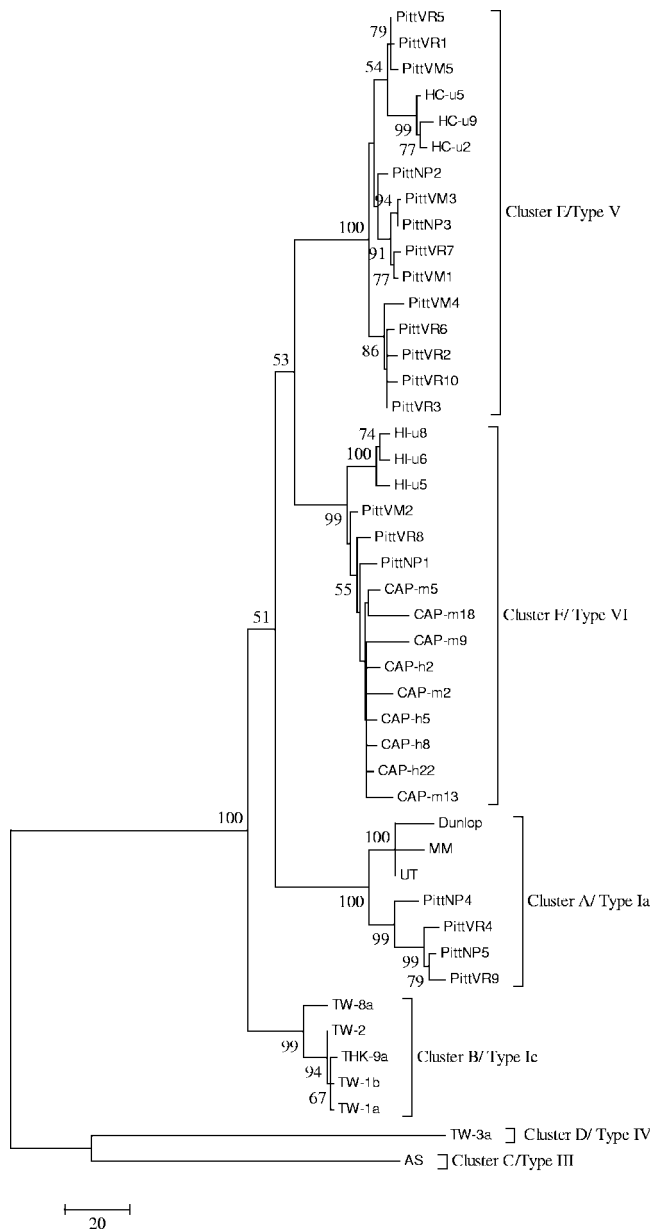


FIG. 1. An unrooted bootstrap consensus tree was constructed by the maximum parsimony method using 45 whole-genome sequences. The presence of six major phylogenetic clusters is noted. The numbers at the nodes are the bootstrap confidence levels (in percentages) obtained for 1,000 replicates (values  $\geq 50\%$  are indicated). Strain notations and origins are described in Table 1.

we also constructed trees using only a single consensus sequence from each patient for whom multiple sequences were available. The six major clusters were retained (data not shown).

Cluster A contains reference strains MM and Dun, which have been conventionally been classified as subtype Ia. It also contains two Pitt isolates from patients with BKV nephropathy (PittNP4 and PittNP5) and two Pitt isolates from patients with BK viruria (PittVR4 and PittVR9). These Pitt isolates branched off from the Dun, MM, and UT strains, supported

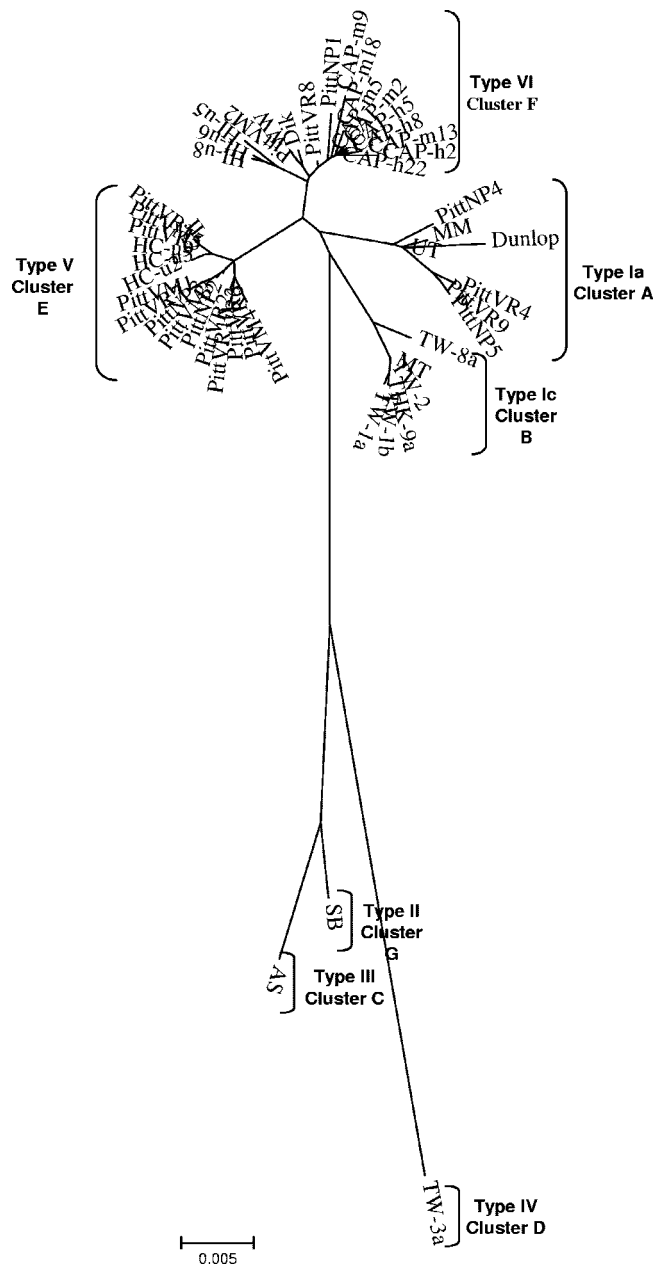


FIG. 2. Unrooted consensus NJ tree using 1,000 bootstrap replicates constructed from the VP1 gene region from 45 whole-genome sequences and the reference strains. The six phylogenetic clustering patterns were reproduced by the VP1 gene. No distinct subgroup Ib cluster was identified. JL clustered with subtype V strains, and WW and Dik clustered with subtype VI. Strain notations and origins are described in Table 1.

with a bootstrap value of 100%. Cluster B was represented by five viral isolates (TW-1a, TW-1b, TW-2, TW-8b, and THK-9a) classified as subtype Ic by Takasaka et al. (43). TW-8a branched out from the other isolates in the cluster with a bootstrap support of 99%. Within the subcluster formed by the remaining isolates in cluster B, TW-2 separated from TW-1a, TW-1b, and THK-9a with a bootstrap value of 94%. Cluster C was represented by strain AS, a prototype III strain according



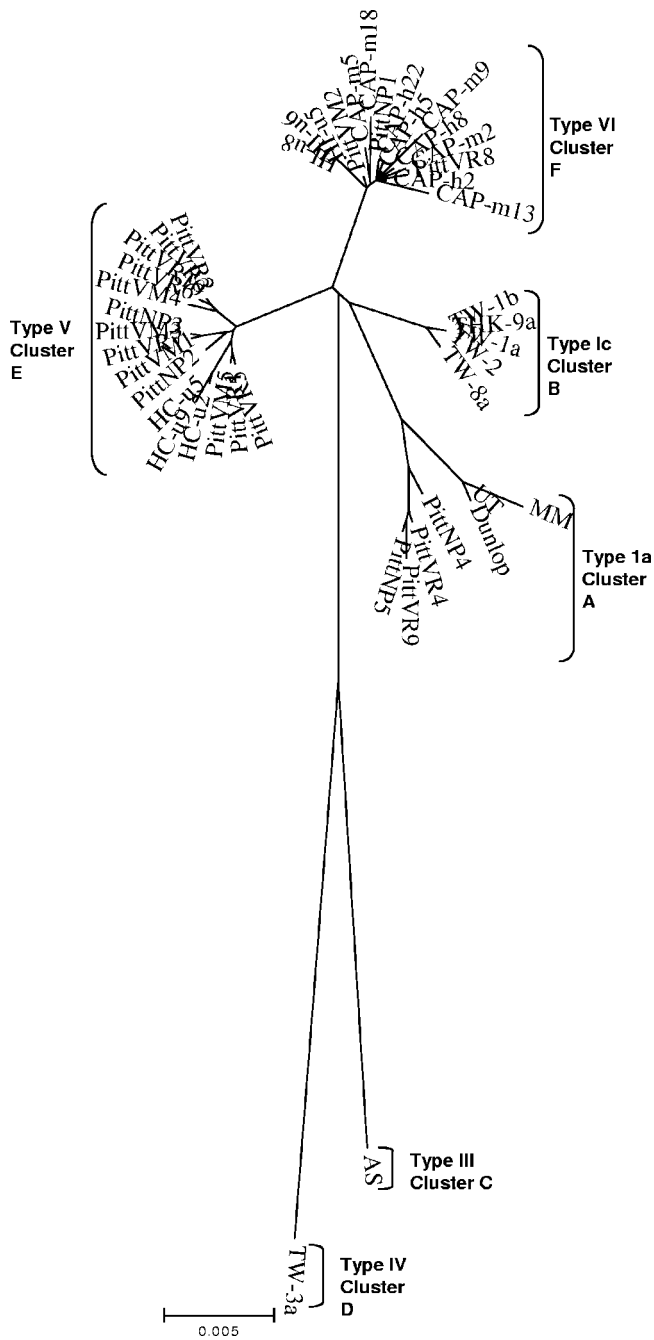


FIG. 3. Unrooted consensus NJ tree constructed from the large T-gene region derived from 45 whole-genome sequences. A phylogenetic clustering similar to that obtained with the whole-genome data was observed from the large T data analysis. Strain notations and origins are described in Table 1.

to the Jin et al. schema (20). Cluster D was represented by a TW-3a isolate classified as subtype IV by Takasaka et al. (42). No other type III or type IV whole-genome sequences are currently available in the literature. A whole-genome sequence corresponding to Jin type II genotype has not been published to date.

All the remaining isolates were found to form two clusters, E and F. These two clusters, supported by high bootstrap

values, separated out from the major subtypes Ia, Ic, III, and IV and are, therefore, proposed to represent newly recognized genotypes V and VI, respectively. Cluster E, representing the proposed genotype V, consists of several Pitt strains associated with viremia (PittVR1, PittVR2, PittVR3, PittVR5, PittVR6, PittVR7, and PittVR10), two associated with nephropathy (PittNP2 and PittNP3), four associated with viremia (PittVM1, PittVM3, PittVM4, and PittVM5), and three clones derived from a healthy subject (HC-u2, HC-u5, and HC-u9). Interestingly, healthy control sequences (HC-u5, HC-u2, and HC-u9) from a subject in the Boston area formed a subcluster separate from the Pitt isolates. Cluster F, representing proposed genotype VI, consists primarily of cloned sequences derived from a single patient with BKV vasculopathy and capillary leak syndrome (CAP-m2, CAP-m5, CAP-m9, CAP-m13, CAP-m18, CAP-mh2, CAP-m5, CAP-m8, and CAP-m22), a single subject with HIV infection (HI-u5, HI-u6, and HI-u8), and Pitt isolates with viremia (PittVR8), viremia (PittVM2), and nephropathy (PittNP1). The HIV sequences formed a subcluster separate from the remaining sequences in this cluster, supported with a bootstrap value of 99%. Additional data are needed to determine if this separation reflects a specific disease association or the common origin of all HIV clones from the same patient. All the clones from the CAP patient formed a cluster separate from the renal transplant sequences from Pittsburgh.

The primary branching of BKV strains into six major clusters seen in the MP tree was quite well reproduced by the NJ tree, with bootstrap values of >80%. Indeed, even second-order subclusters were identical in both trees, with some bootstrap values as high as 100%. The existence of six major clusters was further confirmed by phylogenetic analysis using the UPGMA method (data not shown).

**VP1 sequences.** In order to determine if the major phylogenetic clusters A, B, C, D, E, and F observed with whole-genome sequences could be reproduced by using gene-specific data, we constructed trees using subsets of the whole-sequence information. Initially, complete 1,089-bp alignments were constructed from the 45 whole-genome sequences. This alignment showed that 86.13% of the positions were invariant. The bootstrap consensus tree constructed by the NJ method was in agreement with the whole-genome tree. The six major clusters A, B, C, D, E, and F contained the same viral strains shown in Fig. 1 and 2. MP analysis produced 164 equally parsimonious trees, differing from each other and from the NJ tree only in second-order branching patterns within the clusters. The aforementioned analysis did not allow us to ascertain the relationship of our clusters V and VI to several previously published and only partially sequenced strains which had been classified by the Jin schema. Hence, we constructed another tree (Fig. 2) of 50 complete VP1 gene sequences, including the 45 strains shown in Fig. 1; type II strain SB (17); type Ib strains WW (7, 18, 40), DIK (18, 40), and JL (30, 40); and type Ic strain MT (39, 40). In this VP1 region tree, Jin genotype II (SB) and genotype III (AS) appeared to branch off from a common ancestor, supporting the separate recognition of these two genotypes. However, the type Ib JL strain grouped with genotype V strains in cluster E, whereas WW and DIK type Ib strains clustered with genotype VI strains in cluster F. The MT strain grouped with subtype Ic strains in cluster B.

The published literature contains a number of geographi-

TABLE 3. Number of nucleotide differences in each gene

Gene	Sequence length, in bp ( <i>n</i> )	Total no. of nucleotide differences <sup>a</sup> ( <i>d</i> )	<i>d/n</i> ( $\times 100$ )	No. of nucleotide differences defining clusters <sup>b</sup> ( <i>c</i> )	<i>c/d</i> <sup>c</sup> ( $\times 100$ )
Agno	201	9	4.48	3	33.33
VP2	1,056	90	8.52	30	33.33
VP3	699	72	10.30	26	36.11
VP1	1,089	151	13.86	52	34.44
Large T	2,089	238	11.39	110	46.22
Small t	518	40	7.72	20	50
Total		600		241	

<sup>a</sup> Total number of nucleotide differences in respective genes.

<sup>b</sup> Number of nucleotide differences that distinguished the six clusters in the phylogenetic analysis of whole genomes.

<sup>c</sup> Percentage of number of variable sites contributing to the six clusters in the respective genes.

cally diverse sequences of unique origin, such as pregnant subjects (6), patients with systemic lupus erythematosus (SLE) (2), Japanese renal (RT) or bone marrow transplant (BMT) recipients (43), and metropolitan sewage (PRETORIA1, NANCY2, 160998, SA17, USA3, RA, BCN, and BCN10) (3, 4). These are all partial sequences in different segments of VP1, which could not be included in the tree depicted in Fig. 2 or aligned together to construct a single VP1 tree. Hence, additional phylogenetic trees were constructed (data not shown) using partial VP1 sequences from all viral strains shown in Fig. 1 and additional sequence data from one of the following four groups: (i) nucleotides 1650 to 1936 (287 bp) for Japanese renal and bone marrow transplant isolates; (ii) nucleotides 1714 to 1873 (160 bp) for sewage isolates 160996, SA17, RA, and USA3; (iii) nucleotides 1943 to 2190 (248 bp) for the pregnant subject sequence BCNU and sewage isolates PRETORIA1, NANCY2, BCN, and BCN10; and (iv) nucleotides 1880 to 2093 (214 bp) for two SLE isolates.

The six phylogenetic clusters were resolved by the analysis of the Japanese isolates and sewage isolates (248 bp long). Only partial separation of the clusters was obtained by the analysis of the SLE (214 bp long) and the sewage isolates (160 bp long), indicating insufficient sequence data information. No separate clustering of the sewage or SLE isolates was observed. In the analysis of the Japanese isolates, BMT isolates KOM-1, KOM-5, KOM-9, and KOM-19 and RT isolate THK-2, classified earlier as subtype Ib by Takasaka et al. (43), were observed to cluster with type VI isolates (42). As previously reported by Takasaka et al. (43), KOM-3 and RYU-7 grouped as type III, and THK-3, THK-8, TW-3, TW-3a, TW-9, KOM-2, KOM-12, KOM-22, RYU-3, and RYU-5 grouped as type IV. The remaining Japanese BKV isolates clustered as type Ic. Japanese BKV strains and strains originating in Pittsburgh clustered separately by NJ, UPGMA, and MP methods.

**Large T, small t, agnogene, VP2, and VP3 sequences.** Analysis restricted to large T antigen successfully resolved clusters A through F with high bootstrap values (85% by NJ method) (Fig. 3). These major clusters could also be recognized in trees made from VP2 and VP3 sequences (data not shown). However, if agnogene or small t sequences alone were examined, the complete resolution of the six major clusters was not observed. This is likely due to the smaller size of these sequence

data sets as well as greater evolutionary conservatism of the corresponding genes.

**Gene polymorphisms in phylogenetic clusters.** An attempt was made to define the polymorphic sites that help distinguish the major clusters in our phylogenetic trees. Table 3 lists the gene-specific distribution of these sites. Table 4 provides a detailed listing of all polymorphisms which can assist viral genotyping. The only polymorphic sites listed are those shared by all of the isolates in a particular cluster. It is apparent that the variations are not randomly distributed but seem to be arranged in "hot spots." As expected, the most polymorphic region is VP1, where 13.86% of all sites showed variation, although only 34.44% of these helped distinguish between clusters (Table 3). Significant variation was also found in the large T-antigen gene, wherein polymorphisms were found in 11.39% of all nucleotide sites, 46.2% of which were cluster specific. This degree of genetic variability in the BKV T-antigen gene has not been previously documented. Agnogene was the most conserved coding area in the viral genome.

**Amino acid variations in phylogenetic clusters.** Amino acid replacements corresponding to different phylogenetic clusters and viral strains are listed in Tables 5 and 6, respectively. A total of 44 amino acid differences were identified; 13 were located in large T, 10 in VP2, 9 in VP1, 8 in VP3, and 2 each in small t and agnoprotein, respectively. Although the crystal structure of the BKV VP1 has not been elucidated, amino acid similarity with simian virus 40 VP1 protein (19) suggests a crystal structure similar to crystallized simian virus 40 VP1 protein (27). Using this assumption, 7/9 cluster-specific amino acid changes observed in BKV strains map to the exposed surface of the VP1 molecule. Six of these amino acid variations (positions 61, 62, 68, 69, 74, and 77) localize to the BC loop, while the seventh amino acid change (position 138) maps to the DE loop (Table 5). The newly proposed genotypes V and VI were distinguished at four amino acid positions. These can be represented as follows (in the form of amino acid in genotype V, amino acid position in gene, and amino acid in genotype VI): agnoprotein, L14V; VP2, K318Q; VP3, K199Q; and large T, S354T.

## DISCUSSION

This study extensively examines phylogenetic relationships between polyomavirus BKV strains obtained from different clinical sources and geographical locations. Full-length BKV sequences from only 13 subjects have been published to date. These include 3 strains, Dun, AS, and MM, that were sequenced after in vitro cell cultivation, a UT strain (GenBank accession no. DQ305492), 6 sequences obtained from renal transplant patients in Japan (43), and 15 sequences representing multiple clones of viral DNA amplified directly from the urine of three subjects (8). The latter 15 sequences were derived from a patient with BKV vasculopathy, an HIV-infected patient, and an HIV-negative healthy subject (8). Our study substantially increases knowledge of BKV genomic diversity by providing full-length BKV sequences derived from 20 different renal transplant recipients with BK viruria, viremia, or nephropathy.

Phylogenetic analysis of all 45 full-length sequences established six clusters, designated A, B, C, D, E, and F. Cluster A

TABLE 4. Genotyping sites for the 45 whole-genome sequences

Gene (Dun numbering) <sup>a</sup>	Nucleotide position <sup>a</sup> (bp)	BKV genotype <sup>b,c</sup>						Gene (Dun numbering) <sup>a</sup>	Nucleotide position <sup>a</sup> (bp)	BKV genotype <sup>b,c</sup>																														
		Ia	Ic	II	III	IV	V			VI	Ia	Ic	II	III	IV	V	VI																							
Agno (388-588)	396	G	.	.	.	<u>A</u>	.	.	2199	T	.	.	.	<u>C</u>	.	.	2223	G	.	.	.	<u>A</u>	.	.																
	427	G	.	.	.	.	<u>C*</u>	.	2235	T	.	.	.	<u>A</u>	.	.	2238	C	<b>G</b>	.	.	.	.	.																
	455	A	.	.	.	.	<u>G</u>	.	2259	C	.	.	<u>T</u>	.	.	2274	G	.	<u>T</u>	<u>T</u>	<u>A</u>	.	<u>/C</u>	.																
	598	C	.	.	.	<u>G</u>	.	.	2277	C	.	.	.	<u>T</u>	.	.	2325	T	.	<u>A</u>	<u>G</u>	<u>G</u>	.	.																
	616	C	.	.	.	<u>T</u>	.	.	2337	T	.	.	<u>C</u>	.	.	2370	C	.	<u>G</u>	<u>G</u>	<u>A</u>	.	.	.																
VP2 (624-1679)	734	T	.	.	<u>C</u>	.	.	.	2406	A	.	.	.	<u>G</u>	.	.	2413	G	.	.	.	<u>C</u>	.	.																
	781	G	.	.	<u>C</u>	.	.	.	2457	T	.	.	.	<u>C</u>	.	.	2541	A	.	.	.	<u>G</u>	.	.																
	815	T	.	.	<u>A</u>	.	.	.	Intergenic region (VP1 T antigen)	2550	A	.	<u>T</u>	<u>T</u>	<u>G</u>	.	.	2559	T	<u>C</u>	<u>G</u>	<u>A</u>	<u>C</u>	.	<u>*</u>	<u>C*</u>														
	894	A	<u>G</u>	.	.	.	.	2619		C	.	.	.	.	.	.	<u>*</u>	<u>T*</u>	2678	C	.	.	<u>T</u>	.	.															
	VP3 (981-1679)	986	T	.	.	.	<u>A</u>	.		.	2698	G	.	.	.	<u>A</u>	.	.	2703	A	.	.	.	<u>T</u>	.	.														
1022		T	.	.	.	.	<u>A*</u>	.		2704	G	.	.	.	<u>C</u>	<u>A</u>	.	2705	C	.	.	.	<u>A</u>	.	.															
1023		<u>C</u>	<u>T</u>	.	<u>T</u>	<u>T</u>	<u>T</u>	<u>T</u>		2706	C	.	.	.	<u>A</u>	.	.	2707	A	.	.	.	<u>C</u>	.	.															
1064		<u>C</u>	.	.	<u>T</u>	.	.	.	2708	C	.	.	<u>G</u>	<u>A</u>	<u>G*</u>	<u>*</u>	2711	T	.	.	.	<u>G</u>	.	<u>/C</u>																
1091		T	.	.	.	.	.	<u>C*</u>	2715	G/ <u>C</u>	<u>G</u>	.	<u>G</u>	<u>T</u>	<u>G</u>	<u>G</u>	2719	G	.	.	.	<u>C</u>	.	.																
1109	C	.	.	.	.	<u>T*</u>	.	2761	A	.	<u>G</u>	.	.	.	.	Large T (2722-4566)	2788	A	.	.	.	<u>G</u>	.	.	2791	A	.	.	.	<u>G</u>	.	.	2799	C	.	.	.	<u>T</u>	.	.
1133	G	.	.	<u>A</u>	.	.	.	2801	G	.	.	.	<u>A</u>	.	.		2802	A	.	.	.	<u>G</u>	.	.	2816	T	.	.	.	<u>G</u>	.	.								
1146	<u>T</u>	<u>G</u>	.	<u>G</u>	<u>G</u>	<u>G</u>	<u>G</u>	2817	C	.	.	.	<u>A</u>	.	.		2819	G	.	.	.	<u>T</u>	.	.	2863	G	.	.	.	<u>T</u>	.	.								
1166	<u>G</u>	<u>A</u>	.	.	<u>G</u>	<u>G</u>	<u>G</u>	2819	G	.	.	.	<u>T</u>	.	.		2908	<u>C</u>	<u>T</u>	.	<u>T</u>	<u>T</u>	<u>T</u>	<u>T</u>	2920	A/ <u>G</u>	<u>G</u>	.	<u>A</u>	<u>T</u>	<u>/G</u>	<u>G</u>								
1169	<u>G</u>	<u>A</u>	.	<u>A</u>	<u>A</u>	<u>A</u>	<u>A</u>	2802	A	.	.	.	<u>G</u>	.	.		3036	T	.	.	.	<u>G</u>	.	.	3039	C	.	.	.	<u>A</u>	.	.								
1172	A	.	.	<u>G</u>	.	.	.	2802	A	.	.	.	<u>G</u>	.	.	3058	A	.	.	.	<u>G</u>	.	.	3070	C	.	.	.	<u>T</u>	.	<u>/A</u>									
1187	T	.	.	.	<u>C</u>	.	.	2791	A	.	.	.	<u>G</u>	.	.	3079	C	<u>T</u>	.	<u>A</u>	.	.	3081	A	.	.	.	<u>G</u>	.	.										
1193	A	.	.	<u>G</u>	.	.	.	2799	C	.	.	.	<u>T</u>	.	.	3100	A	<u>G</u>	.	<u>T</u>	<u>T</u>	.	.	3100	A	<u>G</u>	.	<u>T</u>	<u>T</u>	.	.									
1199	C	.	.	<u>T</u>	.	.	.	2801	G	.	.	.	<u>A</u>	.	.	3121	C	.	.	.	<u>T</u>	.	.	3124	T	.	.	.	<u>A</u>	.	.									
1223	T	.	.	<u>C</u>	.	.	.	2802	A	.	.	.	<u>G</u>	.	.	3124	T	.	.	.	<u>A</u>	.	.	3139	T	.	.	.	<u>C</u>	.	.									
1247	T	.	.	<u>C</u>	.	.	.	2816	T	.	.	.	<u>G</u>	.	.	3151	T	.	.	.	<u>A</u>	.	.	3157	G	.	.	.	<u>A</u>	.	.									
1272	<u>C</u>	<u>G</u>	.	<u>G</u>	<u>G</u>	<u>G</u>	<u>G*</u>	2817	C	.	.	.	<u>A</u>	.	.	3172	<u>T</u>	<u>C</u>	.	<u>C</u>	<u>C</u>	<u>C</u>	<u>C</u>	3178	<u>T</u>	.	.	.	<u>C</u>	<u>C</u>	<u>C</u>									
1322	A	.	.	.	<u>T</u>	<u>*</u>	<u>G*</u>	2819	G	.	.	.	<u>T</u>	.	.	3190	<u>T</u>	<u>C</u>	.	<u>C</u>	<u>C</u>	<u>C</u>	<u>C</u>	3193	<u>T</u>	<u>C</u>	.	<u>C</u>	<u>C</u>	<u>C</u>										
1337	T	.	.	.	<u>A</u>	.	.	2820	A	.	.	.	<u>G</u>	.	.	3193	C	.	.	.	<u>A</u>	.	.	Large T (2722-4566)	3195	G	.	.	.	<u>A</u>	.	.								
1343	T	.	.	.	<u>A</u>	.	.	2821	T	.	.	.	<u>G</u>	.	.	3196	G	.	.	.	<u>A</u>	.	.		3202	A	.	.	.	<u>G</u>	.	.								
1367	T	.	.	<u>C</u>	<u>A</u>	.	.	2821	C	.	.	.	<u>T</u>	.	.	3205	G	.	.	.	<u>C</u>	.	<u>*</u>		<u>A*</u>	3232	G	.	.	<u>C</u>	<u>A</u>	.	.							
1389	G	.	.	.	<u>C</u>	.	.	2822	A	.	.	.	<u>G</u>	.	.	3238	T	<u>A</u>	.	.	.	.	.		3250	G	.	.	.	<u>T</u>	.	.								
1412	C/T	<u>C</u>	.	<u>T</u>	<u>G</u>	<u>C</u>	<u>C</u>	2822	A	.	.	.	<u>G</u>	.	.	3265	A	.	.	.	<u>G</u>	.	.		3283	<u>T</u>	<u>A</u>	.	<u>A</u>	<u>A</u>	<u>A</u>	<u>A</u>								
1427	A	.	.	.	<u>G</u>	.	.	2822	A	.	.	.	<u>G</u>	.	.	3289	A	.	.	.	<u>G</u>	.	.	3303	A	.	.	.	<u>G</u>	.	.									
1429	G	.	.	.	<u>A</u>	<u>C</u>	<u>C</u>	2822	A	.	.	.	<u>G</u>	.	.	3315	A	<u>G</u>	.	.	.	.	.	3376	T	.	.	.	<u>C</u>	.	.									
VP1 (1564-2652)	1575	C	.	.	.	.	<u>A*</u>	<u>*</u>	2761	A	.	.	.	.	.	Large T (2722-4566)	3195	G	.	.	.	<u>A</u>	.	.	3379	T	.	.	.	<u>C</u>	.	.								
	1704	G	.	.	.	<u>A</u>	.	.	2788	A	.	.	.	<u>G</u>	.		3196	G	.	.	.	<u>A</u>	.	.	3202	A	.	.	.	<u>G</u>	.	.								
	1722	C	.	.	.	<u>T</u>	.	.	2791	A	.	.	.	<u>G</u>	.		3205	G	.	.	.	<u>C</u>	.	<u>*</u>	<u>A*</u>	3232	G	.	.	<u>C</u>	<u>A</u>	.	.							
	1747	A	.	.	.	<u>G</u>	.	.	2799	C	.	.	.	<u>T</u>	.		3238	T	<u>A</u>	.	.	.	.	.	3250	G	.	.	.	<u>T</u>	.	.								
	1766	T	.	.	<u>A</u>	.	.	.	2801	G	.	.	.	<u>A</u>	.		3265	A	.	.	.	<u>G</u>	.	.	3283	<u>T</u>	<u>A</u>	.	<u>A</u>	<u>A</u>	<u>A</u>	<u>A</u>								
1767	A	.	.	<u>G</u>	.	.	.	2802	A	.	.	.	<u>G</u>	.	3289	A	.	.	.	<u>G</u>	.	.	3303	A	.	.	.	<u>G</u>	.	.										
VP1 (1564-2652)	1768	A	.	.	<u>C</u>	.	.	.	2802	A	.	.	.	<u>G</u>	.	3315	A	<u>G</u>	.	.	.	.	3376	T	.	.	.	<u>C</u>	.	.										
	1769	A	.	.	.	<u>G</u>	.	.	2816	T	.	.	.	<u>G</u>	.	3379	T	.	.	.	<u>C</u>	.	.	Large T (2722-4566)	3195	G	.	.	.	<u>A</u>	.	.								
	1770	G	.	.	<u>C</u>	<u>A</u>	.	.	2817	C	.	.	.	<u>T</u>	.	3196	G	.	.	.	<u>A</u>	.	.		3202	A	.	.	.	<u>G</u>	.	.								
	1784	A	.	.	.	<u>C</u>	.	.	2819	G	.	.	.	<u>T</u>	.	3205	G	.	.	.	<u>C</u>	.	<u>*</u>		<u>A*</u>	3232	G	.	.	<u>C</u>	<u>A</u>	.	.							
	1794	C	.	.	<u>G</u>	.	.	.	2820	A	.	.	.	<u>G</u>	.	3238	T	<u>A</u>	.	.	.	.	.		3250	G	.	.	.	<u>T</u>	.	.								
	1833	C	.	.	<u>T</u>	.	.	.	2822	A	.	.	.	<u>G</u>	.	3265	A	.	.	.	<u>G</u>	.	.		3283	<u>T</u>	<u>A</u>	.	<u>A</u>	<u>A</u>	<u>A</u>	<u>A</u>								
	1851	C/G	<u>C</u>	<u>G</u>	<u>G</u>	<u>A</u>	<u>C</u>	<u>C</u>	2822	A	.	.	.	<u>G</u>	.	3289	A	.	.	.	<u>G</u>	.	.	3303	A	.	.	.	<u>G</u>	.	.									
	1854	C	.	.	.	<u>T</u>	.	.	2822	A	.	.	.	<u>G</u>	.	3315	A	<u>G</u>	.	.	.	.	.	3376	T	.	.	.	<u>C</u>	.	.									
	1857	T	.	.	<u>C</u>	.	.	.	2822	A	.	.	.	<u>G</u>	.	3379	T	.	.	.	<u>C</u>	.	.	Large T (2722-4566)	3195	G	.	.	.	<u>A</u>	.	.								
	1869	C	.	.	.	<u>T</u>	.	.	2822	A	.	.	.	<u>G</u>	.	3196	G	.	.	.	<u>A</u>	.	.		3202	A	.	.	.	<u>G</u>	.	.								
	1905	A	.	.	.	<u>G</u>	.	.	2822	A	.	.	.	<u>G</u>	.	3205	G	.	.	.	<u>C</u>	.	<u>*</u>		<u>A*</u>	3232	G	.	.	<u>C</u>	<u>A</u>	.	.							
	1908	T	.	.	.	.	<u>A*</u>	<u>*</u>	2822	A	.	.	.	<u>G</u>	.	3238	T	<u>A</u>	.	.	.	.	.		3250	G	.	.	.	<u>T</u>	.	.								
	1965	A	.	.	.	<u>G</u>	.	.	2822	A	.	.	.	<u>G</u>	.	3265	A	.	.	.	<u>G</u>	.	.		3283	<u>T</u>	<u>A</u>	.	<u>A</u>	<u>A</u>	<u>A</u>	<u>A</u>								
	1977	G	.	.	.	<u>C</u>	.	.	2822	A	.	.	.	<u>G</u>	.	3289	A	.	.	.	<u>G</u>	.	.	3303	A	.	.	.	<u>G</u>	.	.									
	1989	A	<u>G</u>	<u>C</u>	<u>T</u>	<u>C</u>	<u>T</u>	<u>T</u>	2822	A	.	.	.	<u>G</u>	.	3315	A	<u>G</u>	.	.	.	.	.	3376	T	.	.	.	<u>C</u>	.	.									
	1992	A	<u>G</u>	.	.	.	.	.	2822	A	.	.	.	<u>G</u>	.	3379	T	.	.	.	<u>C</u>	.	.	Large T (2722-4566)	3195	G	.	.	.	<u>A</u>	.	.								
	1998	T</																																						

TABLE 4—Continued

Gene (Dun numbering) <sup>a</sup>	Nucleotide position <sup>a</sup> (bp)	BKV genotype <sup>b,c</sup>						Gene (Dun numbering) <sup>a</sup>	Nucleotide position <sup>a</sup> (bp)	BKV genotype <sup>b,c</sup>							
		Ia	Ic	II	III	IV	V			VI	Ia	Ic	II	III	IV	V	VI
	3400	G	.	.	.	<u>A</u>	.	.		4299	G	.	.	.	<u>A</u>	.	.
	3409	T	.	.	<u>A</u>	.	.	.		4315	A	.	.	<u>G</u>	.	.	.
	3424	<u>C</u>	T	.	T	T	T	T		4330	G	.	.	<u>A</u>	.	.	.
	3433	T	.	.	.	<u>A</u>	/C	.		4339	G	.	.	<u>T</u>	<u>A</u>	.	.
	3469	A	.	.	.	<u>G</u>	.	.		4417	T	.	.	.	.	*	<u>C</u> *
	3481	A	.	.	<u>G</u>	.	.	.		4435	T	.	.	.	.	*	<u>G</u> *
	3501	G	.	.	<u>A</u>	.	.	.		4459	T	.	.	.	<u>C</u>	.	.
	3502	T	.	.	<u>G</u>	.	.	.		4462	T	.	.	.	<u>C</u>	.	.
	3511	T	.	.	<u>G</u>	.	.	.		4483	T	.	.	<u>C</u>	.	.	.
	3523	G	<u>A</u>	.	T	T	.	.									
	3532	A	.	.	.	<u>G</u>	.	.		Large T intron (4567-4910)	4570	T	<u>G/C</u>	.	.	.	.
	3535	T	.	.	.	<u>C</u>	.	.		4575	A	.	.	.	.	<u>C</u> *	*
	3562	<u>G</u>	A	.	A	A	A	A		4592	C	.	.	.	.	.	.
	3570	T	.	.	<u>C</u>	.	.	.		4596	T	.	.	<u>C</u>	.	.	.
	3577	C	.	.	<u>A</u>	.	.	.		4597	T	A	.	.	<u>C</u>	A	A
	3580	A	.	.	.	<u>G</u>	.	.		4598	A	.	.	T	<u>C</u>	.	.
	3589	T	.	.	<u>A</u>	.	.	.		4599	A	<u>C</u>	.	T	T	.	.
	3619	G	.	.	<u>A</u>	.	.	.		4600	T	.	.	.	<u>A</u>	.	.
	3622	C	.	.	<u>T</u>	.	.	.		4601	A	.	.	.	<u>T</u>	.	.
	3634	A	.	.	<u>C</u>	T	.	.		4603	T	.	.	.	<u>A</u>	.	.
	3652	T	.	.	.	.	*	<u>C</u> *		4606	A	.	.	.	<u>T</u>	*	<u>C</u> *
	3654	G	.	.	.	.	<u>A</u> *	*									
	3673	T	.	.	.	.	<u>C</u> *	*		Small t (4635-5153)	4607	T	.	.	<u>C</u>	<u>A</u>	.
	3709	<u>G</u>	A	.	A	A	A	A		4608	T	.	.	.	<u>C</u>	.	.
	3715	T	.	.	.	<u>G</u>	.	.									
	3736	G	.	.	.	.	<u>A</u> *	*		Large T (4911-5153)	4609	A	.	.	.	<u>T</u>	.
	3749	G	.	.	.	.	<u>C</u> *	*		4610	T	.	.	<u>C</u>	<u>A</u>	.	.
	3757	T	.	.	.	<u>C</u>	.	.		4613	A	<u>C</u>	.	<u>T</u>	.	.	.
										4620	T	.	.	.	<u>G</u>	.	.
Large T (2722-4566)	3772	A	.	.	<u>T</u>	<u>C</u>	.	.		4626	T	.	.	.	<u>G</u> *	*	.
	3781	T	.	.	.	<u>C</u>	.	.		4662	T	.	.	.	<u>C</u>	.	.
	3829	A	.	.	.	<u>G</u>	.	.		4692	A	.	.	<u>G</u>	.	.	.
	3844	<u>G</u>	A	.	A	A	A	A		4704	T	.	.	<u>G</u>	.	.	.
	3859	C	.	.	.	T	.	.		4734	G	.	.	<u>A</u>	.	.	.
	3871	A	.	.	.	<u>G</u>	.	.		4764	G	.	.	<u>T</u>	.	.	.
	3877	G	.	.	<u>A</u>	.	.	.		4797	C	.	.	.	<u>T</u>	.	.
	3904	C	.	.	.	.	*	<u>T</u> *		4815	T	.	.	.	<u>C</u>	.	.
	3916	T	.	.	.	<u>C</u>	.	.		4830	G	.	.	<u>A</u>	.	.	.
	3942	A	.	.	.	<u>G</u>	.	.		4877	G	.	.	<u>A</u>	.	.	.
	3955	C	.	.	.	<u>T</u>	.	.		4878	G	.	.	<u>A</u>	.	.	.
	3985	A	.	.	.	<u>T</u>	.	.		4921	C	.	.	<u>T</u>	.	.	.
	3997	A	.	.	.	<u>G</u>	.	.		4944	A	.	.	.	<u>G</u>	.	.
	4021	C	.	.	.	<u>T</u>	.	.		4947	A	<u>G</u>	.	.	.	.	.
	4036	A	.	.	<u>G</u>	.	.	.		5047	C	.	.	<u>T</u>	.	.	.
	4050	A	.	.	.	<u>G</u>	.	.		5055	<u>A</u>	G	.	G	G	G	G
	4063	G	.	.	.	<u>A</u>	.	.		5061	T	.	.	.	<u>C</u>	.	.
	4075	<u>A</u>	T	.	T	T	T	T		5076	<u>A</u>	G	.	G	G	G	G
	4080	G	.	.	.	<u>A</u>	.	.		5103	T	.	.	.	.	<u>C</u> *	*
	4081	G	.	.	.	<u>A</u>	.	.		5141	G	.	.	<u>A</u>	.	.	.
	4090	T	.	.	<u>C</u>	.	.	.		5142	A/C	C	.	<u>G</u>	C	C	C
	4298	T	.	.	.	<u>A</u>	.	.									

<sup>a</sup> Nucleotide base numbering is the same as that for the BKV Dunlop GenBank sequence (accession no. V01108).

<sup>b</sup> Genotyping sites specific to a BKV subtype are indicated in boldface type. Wherever sequence data for reference type II are unavailable, positions distinguishing available subtypes are underlined.

<sup>c</sup> The sites differentiating the newly recognized genotypes V and VI are marked with an asterisk. A period indicates identity with the Dunlop nucleotide sequences at that position. A dash indicates no sequence data were available. A slash indicates that different strains within a genotype had more than one nucleotide present at that position (for example, G/C).

contained isolates exclusively from genotype Ia. Cluster B (subtype Ic) was represented by isolates from renal transplant patients from Japan. Cluster C (subtype III) included strain AS, which is the only genotype III whole-genome sequence published to date. Cluster D (subtype IV) was represented by TW-3a, a Japanese renal transplant isolate. All the remaining isolates were divided among the clusters E and F. Phylogenetic analysis of VP1, VP2, VP3, and large-T-derived gene-specific sequences by three different methods corroborated our whole-genome analysis, demonstrating the existence of two previously unrecognized clusters, E and F, which we propose to call ge-

notypes V and VI, respectively. These new genotypes seem to have independently branched off from the four other clusters, A, B, C, and D. Due to the lack of an appropriate outgroup for BKV phylogenetic analysis, it is not possible for us to determine the order in which different BKV genotypes have evolved.

Based on conventional alignments of VP1 sequences, existing BKV strains have been classified into genotypes I, II, III, and IV (7, 18, 37). Genotype I has been subdivided into subgroups Ia (Dun and MM strains), Ib (Dik, JL, and WW strains), and Ic (MT) (43). Phylogenetic analysis of the VP1



TABLE 5. Amino acid changes differentiating clusters in 45 whole-genome tree

Protein	Amino acid position <sup>a</sup>	Amino acid change by cluster and genotype <sup>b</sup>							
		A, Ia	B, Ic	G, II	C, III	D, IV	E, V	F, VI	
Agno	14	V	.	-	.	.	<b>L</b>	.	
	23	K	.	-	.	<b>R</b>	.	.	
VP2	53	S	.	-	<b>T</b>	.	.	.	
	91	I	<b>V</b>	-	.	.	.	.	
	<u>175</u>	<b>S</b>	<b>A</b>	-	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>	
	<u>217</u>	<b>Q</b>	<b>E</b>	-	<b>D</b>	<b>D</b>	<b>E</b>	<b>E</b>	
	<u>240</u>	<b>R</b>	.	-	<b>H</b>	<b>Q</b>	.	.	
	<u>248</u>	<b>S</b>	.	-	.	<b>R</b>	.	.	
	<u>256</u>	<b>E</b>	.	-	.	<b>Q</b>	.	.	
	<u>263</u>	<b>D</b>	.	-	.	<b>E</b>	.	.	
	<u>269</u>	<b>S</b>	.	-	.	<b>N</b>	<b>T</b>	<b>T</b>	
	<u>318</u>	<b>Q</b>	.	-	.	.	<b>K</b>	.	
	VP3	<u>56</u>	<b>S</b>	<b>A</b>	-	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>
<u>98</u>		<b>Q</b>	<b>E</b>	-	<b>D</b>	<b>D</b>	<b>E</b>	<b>E</b>	
<u>121</u>		<b>R</b>	.	-	<b>H</b>	<b>Q</b>	.	.	
<u>129</u>		<b>S</b>	.	-	.	<b>R</b>	.	.	
<u>137</u>		<b>E</b>	.	-	.	<b>Q</b>	.	.	
<u>144</u>		<b>D</b>	.	-	.	<b>E</b>	.	.	
<u>150</u>		<b>S</b>	.	-	.	<b>N</b>	<b>T</b>	<b>T</b>	
<u>199</u>		<b>Q</b>	.	-	.	.	<b>K</b>	.	
VP1		61	E	<b>/K<sup>c</sup></b>	D	D	<b>N</b>	.	.
	62	N	.	.	.	<b>D</b>	.	.	
	68	L	.	.	<b>Q</b>	.	.	.	
	69	K	.	.	<b>H</b>	<b>R</b>	.	.	
	74	N	.	.	.	<b>T</b>	.	<b>/Y</b>	
	77	S	.	D	E	D	.	.	
	138	E	.	.	.	<b>D</b>	.	.	
	225	F	<b>L</b>	Y	Y	Y	.	.	
	284	A	.	.	.	<b>P</b>	.	.	
Large T	<u>36</u>	<b>R</b>	.	-	<b>K</b>	.	.	.	
	<u>78</u>	<b>S</b>	.	-	<b>N</b>	.	.	.	
	171	<b>Q</b>	.	-	.	<b>L</b>	.	.	
	244	<b>H</b>	.	-	.	<b>Y</b>	.	.	
	354	<b>T</b>	.	-	.	.	<b>S</b>	.	
Large T	365	E	.	-	.	<b>D</b>	.	.	
	414	I	.	-	<b>V</b>	.	.	.	
	591	A	.	-	<b>S</b>	.	.	.	
	592	T	.	-	<b>K</b>	<b>Q</b>	.	<b>/A</b>	
	664	A	.	-	<b>D</b>	.	.	.	
	665	E	.	-	<b>S</b>	.	.	.	
	670	S	.	-	.	<b>L</b>	.	.	
	671	D	.	-	<b>N</b>	.	.	.	
Small t	<u>36</u>	<b>R</b>	.	-	<b>K</b>	.	.	.	
	<u>78</u>	<b>S</b>	.	-	<b>N</b>	.	.	.	

<sup>a</sup> Amino acid numbering is the same as that for the GenBank BKV Dunlop sequence (accession no. V01108).

<sup>b</sup> Amino acid changes specific to a BKV genotype are indicated in boldface type. A period indicates identity with the BKV Dunlop amino acid sequence at that position. The 3' region of genes encoding VP2 and VP3 and the 5' coding region of genes encoding large T and small t overlap. VP3 protein has its C terminus in common with VP2, and small t protein has its N-terminal region in common with large T. The respective regions are transcribed in the same reading frame, encoding the same amino acids. Amino acid substitutions in the above VP2/VP3 common region and large T/small t common region are underlined.

<sup>c</sup> A slash indicates that different strains within a genotype have more than one amino acid present at that position (for example, **/K**).

TABLE 6. Amino acid pattern observed in VP1 protein

Subtype <sup>a</sup>	Strain	Amino acid at the indicated position in VP1 protein <sup>b</sup>		
		82	340	362
Ia	Dunlop, MM, UT, PittVR4, PittVR9, PittNP4, PittNP5	E	R	L
Ic	TW-1b, TW-2, strain MT, THK-9a	.	K	.
	TW-1a	Q	K	.
	TW-8a	D	K	.
II	SB	D	Q	V
III	AS	D	Q	V
IV	TW-3a	D	K	V
V	PittVR2, PittVR3, PittVR6, PittVR10, PittVM2	D	.	.
	PittVR7, PittNP2, PittNP3, PittVM1, PittVM3	.	K	.
	PittVR1, PittVR5, PittVM5, JL strain, HC-u2, HC-u5, HC-u9	.	.	V
	CAP-m2, CAP-m5, CAP-m9, CAP-m13, CAP-m18, CAP-h2, CAP-h5, CAP-h8, CAP-h22	.	.	.
VI	HI-u5, HI-u6, HI-u8, PittVM2, PittVR8, PittNP1	.	.	.

<sup>a</sup> BKV subtypes are represented by strains used in this study.

<sup>b</sup> Amino acid numbering is the same as that for the VP1 protein in the Dunlop strain (GenBank accession no. CAA24299). Amino acids similar to those of Dunlop are represented by periods at the respective positions.

gene sequences in our study supports the existence of genotypes Ia and Ic. However, no separate cluster corresponding to genotype Ib was identified. Of BKV strains designated Ib previously, Dik and WW clustered with type VI strains, while JL clustered with type V strains. As noted by Chen et al., the recognition of genotype Ib is based essentially on differences at three nucleotide positions, 1698, 1809, and 1923 (8). Genotype Ib-defining substitutions at these three sites do not result in amino acid changes and are not expected to result in distinct viral serotypes. Hence, genotype Ib may not be a biologically relevant taxonomic subgroup. In contrast, Chen et al. (8) did observe genotype Ic to represent a distinct phylogenetic cluster, similar to our analysis in which the MT strain clusters with other Ic strains published by Takasaka et al. (42). The amino acid sequence of subtype Ic was unique at position 225 in VP1 protein and position 91 in VP2 protein. Genotypes V and VI, as defined by us, have distinctive nucleotide substitutions at positions 427, 1022, 1091, 1109, 1322, 1575, 1908, 2076, 2127, 2559, 2619, 2708, 3205, 3652, 3654, 3673, 3736, 3749, 3904, 4417, 4435, 4575, 4606, 4626, and 5103 (Table 4). These nucleotide substitutions result in four nonsynonymous amino acid changes: one each in agnoprotein, VP2, VP3, and large T protein.

Although not informative for genotyping, amino acids at positions 82, 340, and 362 in the VP1 protein showed an interesting pattern of changes, which suggested that amino acid substitutions at these three locations might result in type-defining changes in three-dimensional protein configurations. In genotype II, III, and IV isolates, amino acids at all three positions were substituted compared to the Dunlop reference Ia sequence (Table 6), whereas type VI strains showed no substitution at any of these three positions. In genotype Ic, amino acid substitutions occurred only at positions 82 and 340,

whereas in type V any one of the three sites could be mutated. Amino acid position 82 can be predicted to map to the BC loop, whereas positions 340 and 362 were mapped to predicted "C-insert" and "C loop" of the C-terminal region, respectively (27). Since the BC loop is believed to interact with the cellular receptor for BKV, it can be speculated that the genotype-specific amino acid changes might alter BKV tissue tropism. By the same token, changes in the C-terminal region may have implications with regard to the efficiency of viral capsid assembly. The biologic basis for the amino acid substitution constraints observed at positions 82, 340, and 362 is not clear, since the BC loop and C terminus are not known to interact with each other.

BKV is primarily a kidney pathogen. However, a BKV (Yale) strain has been amplified from a leukemia patient who had lytic infection in both the kidney and brain (38). A partial VP1 sequence of this unique strain reportedly showed three mutations within the VP1 gene (positions 1687, 1702, and 1908), which distinguished this strain from other type I sequences. Many of the genotype V BKV strains from Pittsburgh were identical to the Yale strain at positions 1687 and 1908, but none showed the G1702C mutation. The latter mutation results in a Glu-to-Gln mutation, which leads to predicted changes in  $\beta$  structure of the coded protein and a postulated increase in the tropism of BKV for brain tissue. The G1702 mutation, however, has also been reported by Chen et al. (8) in one of three clones derived from an HIV-infected patient who is not known to have viral encephalitis.

In our follow-up renal transplant patients in the clinic, we have observed that while 30% of patients develop asymptomatic BK viruria, only 1 to 2% develop viral nephropathy. Whole-genome phylogenetic analysis does not suggest any major evolutionary distinction between viral strains obtained from patients with or without nephropathy, since sequences derived from both clinical categories did not form distinct clusters. This suggests that the intensity of immunosuppression and genetic susceptibility of the host immune system, possibly regulated by host gene polymorphisms, are the principal determinants of whether or not a particular patient will develop BKV-mediated tissue injury in the kidney.

Small data sets of viral sequences derived from healthy patients, HIV-infected subjects, Wiskott-Aldrich syndrome patients, and patients with BKV vasculopathy with capillary leak syndrome did form discrete subclusters. More sequences from such patients are needed to determine whether this finding reflects specific disease association or a distinct epidemiologic origin of the BKV strains selected for study. It is pertinent to note that sequences derived from the patient with capillary leak syndrome did not show any functionally significant mutations (8). A few amino acid changes were found in the VP1 region, but they did not result in any predicted change in VP1 serotype or cellular receptor interaction. Likewise, amino acid changes in the T antigen did not affect the DNA binding domain, host range domain, phosphorylation sites, or any other critical part of this multifunctional molecule.

A phylogenetic separation of Japanese and Pittsburgh BKV isolates from renal transplant patients was seen on analysis of partial VP1 sequences. However, this observation may reflect overrepresentation of subtypes Ic and IV in the Japanese isolates and genotype V in the Pittsburgh BKV strains. It is, thus,

unclear whether this separation is due to geographic or genotypic distinction. However, the related polyomavirus JC is known to have specific viral genotypes associated with particular geographical regions (23, 43).

The NCCR includes the regulatory region enhancer and promoter sequences as well as origin of replication (ori). Genetic changes in the regulatory region are known to occur during viral growth in vitro, and it is believed that these changes can promote the selection of tissue culture-adapted strains (7, 8, 41). Whether similar changes occur in the coding region of the viral genome is not known. To address this question, we compared the DNA sequences of BKV Dunlop (36), AS (44, 46), UT (41), and MM (46) strains with 42 other strains that were sequenced directly after DNA amplification from clinical specimens. The cell-cultured strains MM, Dun, UT, and AS differed from the other strains at position 420 in the agnogene (A→T) and at position 5142 in the T-antigen/small t common gene region (A/G→C). The large T-antigen region also showed several differential substitutions at positions 3034, 3592, 3874, and 4980, with A, T, C, and C in the cell-cultured strains and T, A, T, and T in remaining strains. This observation suggests that in vitro culture can result in nucleotide substitutions in the coding regions of the viral genome. Since all nucleotide alterations were synonymous, the functional effect of these substitutions is doubtful. Nonetheless, these observations demonstrate that BKV is a rapidly evolving virus in vitro and in vivo. No nucleotide differences specific for the cell-cultured strains were observed in VP1, VP2, and VP3 genes.

In summary, our data document the phylogenetic diversity of BKV and have established the existence of clades not previously recognized in the literature. Currently, there is only limited whole-genome sequence data available for many important subject categories, including healthy individuals, pregnant women, patients with systemic lupus erythematosus or HIV infection, and bone marrow transplant recipients. Additional whole-genome sequence information is also needed for BKV genotypes II, III, and IV and for viral strains associated with metropolitan sewage. Expanded phylogenetic analyses with these additional BKV sequences will provide a global and statistically more robust classification schema for classifying BKV into types and subtypes, as has been accomplished for polyomavirus JC (17). Such an effort would allow better definition of the pathogenicity and tissue tropism of specific BKV viral strains encountered in nature. The phylogenetic diversity of BKV also has important implications for clinical diagnostics. Microbiology laboratories should remain alert to the possibility of continuing evolution of BKV, and the potential need for future modifications in currently utilized PCR assays, to ensure that all clinically relevant occurring viral strains can be successfully detected in clinical samples.

#### ACKNOWLEDGMENTS

This work was supported by NIH grants RO-1 AI-51227-01, AI-063360, and AI-060602.

We thank Sujata Patel for technical assistance and Basavaraju Sankarappa for help with Clustal W software. Christopher Cubitt and Caroline Ryschkeiwisch at the NIH generously provided an established experimental protocol for BKV whole-genome sequencing.

## REFERENCES

1. Agostini, H. T., and G. L. Stoner. 1995. Amplification of the complete polyomavirus JC genome from brain, cerebrospinal fluid and urine using pre-PCR restriction enzyme digestion. *J. Neurovirol.* **1**:316–320.
2. Bendiksen, S., O. P. Rekvig, M. Van Ghelue, and U. Moens. 2000. VP1 DNA sequences of JC and BK viruses detected in urine of systemic lupus erythematosus patients reveal no differences from strains expressed in normal individuals. *J. Gen. Virol.* **81**:2625–2633.
3. Bofill-Mas, S., M. Formiga-Cruz, P. Clemente-Casares, F. Calafell, and R. Girones. 2001. Potential transmission of human polyomaviruses through the gastrointestinal tract after exposure to virions or viral DNA. *J. Virol.* **75**:10290–10299.
4. Bofill-Mas, S., S. Pina, and R. Girones. 2000. Documenting the epidemiologic patterns of polyomaviruses in human populations by studying their presence in urban sewage. *Appl. Environ. Microbiol.* **66**:238–245.
5. Bressollette-Bodin, C., M. Coste-Burel, M. Hourmant, V. Sebillé, E. Andre-Garnier, and B. M. Imbert-Marcille. 2005. A prospective longitudinal study of BK virus infection in 104 renal transplant recipients. *Am. J. Transplant.* **5**:1926–1933.
6. Chang, D., M. Wang, W. C. Ou, M. S. Lee, H. N. Ho, and R. T. Tsai. 1996. Genotypes of human polyomaviruses in urine samples of pregnant women in Taiwan. *J. Med. Virol.* **48**:95–101.
7. Chauhan, S., G. Lecatsas, and E. H. Harley. 1984. Genomic analysis of (WW) viral DNA cloned directly from human urine. *Intervirology* **22**:170–176.
8. Chen, Y. P., P. M. Sharp, M. Fowkes, O. Kocher, J. T. Joseph, and I. J. Koralnik. 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.
9. Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. Gibson, D. Higgins, and J. D. Thompson. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**:3497–3500.
10. Demeter, L. M. 1995. JC, BK, and other polyomaviruses; progressive multifocal leukoencephalopathy, p. 1400–1406. *In* G. L. Mandel, J. E. Bennett, and R. Dolin (ed.), *Principles and practice of infectious diseases*. Churchill Livingstone, New York, N.Y.
11. DiTaranto, C., V. Pietropaolo, G. B. Orsi, L. Jin, L. Sinibaldi, and A. M. Degener. 1997. Detection of BK polyomavirus genotypes in healthy and HIV-positive children. *Eur. J. Epidemiol.* **13**:653–657.
12. Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
13. Fioriti, D., A. M. Degener, M. Mischitelli, M. Videtta, A. Arancio, S. Sica, F. Sora, and V. Pietropaolo. 2005. BKV infection and hemorrhagic cystitis after allogeneic bone marrow transplant. *Int. J. Immunopharmacol.* **18**:309–316.
14. Forsman, Z. H., J. A. Lednický, G. E. Fox, R. C. Willson, Z. S. White, S. J. Halvorson, C. Wong, A. M. Lewis, and J. S. Butel. 2004. Phylogenetic analysis of polyomavirus simian virus 40 from monkeys and humans reveals genetic variation. *J. Virol.* **78**:9306–9316.
15. Freund, R., R. L. Garcea, R. Sahli, and T. L. Benjamin. 1991. A single-amino-acid substitution in polyomavirus VP1 correlates with plaque size and hemagglutination behavior. *J. Virol.* **65**:350–355.
16. Gardner, S. D., E. F. MacKenzie, C. Smith, and A. A. Porter. 1984. Prospective study of the human polyomaviruses BK and JC and cytomegalovirus in renal transplant recipients. *J. Clin. Pathol.* **37**:578–586.
17. Gibson, P. E., and S. D. Gardner. 1983. Strain differences and some serological observations on several isolates of human polyomaviruses. *Prog. Clin. Biol. Res.* **105**:119–132.
18. Goudsmit, J., M. L. Baak, K. W. Sletterus, and J. Van der Noorda. 1981. Human papovavirus isolated from urine of a child with acute tonsillitis. *Br. Med. J.* **283**:1363–1364.
19. Imperiale, M. J. 2001. The human polyomavirus: an overview, p. 53–71. *In* K. Khalili and G. L. Stoner (ed.), *Human polyomaviruses*. Wiley-Liss Inc., New York, N.Y.
20. Jin, L., P. E. Gibson, J. C. Booth, and J. P. Clewley. 1993. Genomic typing of BK virus in clinical specimens by direct sequencing of polymerase chain reaction products. *J. Med. Virol.* **41**:11–17.
21. Jobs, D. V., S. C. Chima, C. F. Ryschkewitsch, and G. L. Stoner. 1998. Phylogenetic analysis of 22 complete genomes of the human polyomavirus JC virus. *J. Gen. Virol.* **79**:2491–2498.
22. Khalili, K., W. Martyn, S. Hirofumi, K. Nagashima, and M. Safak. 2005. The agnoprotein of polyomaviruses: a multifunctional auxiliary protein. *J. Cell Phys.* **204**:1–7.
23. Knowles, W. A. 2001. The epidemiology of BK virus and the occurrence of antigenic and genomic subtypes, p. 527–560. *In* K. Khalili and G. L. Stoner (ed.), *Human polyomaviruses: molecular and clinical perspectives*. Wiley-Liss, New York, N.Y.
24. Knowles, W. A., P. E. Gibson, and S. D. Gardner. 1989. Serological typing scheme for BK-like isolates of human polyomavirus. *J. Med. Virol.* **28**:118–123.
25. Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**:150–163.
26. Leung, A. Y., C. K. Suen, A. K. Lie, R. H. Liang, K. Y. Yuen, and Y. L. Kwong. 2001. Quantification of polyoma BK viruria in hemorrhagic cystitis complicating bone marrow transplantation. *Blood* **98**:1971–1978.
27. Liddington, R. C., Y. Yan, J. Moulai, R. Sahli, T. L. Benjamin, and S. C. Harrison. 1991. Structure of simian virus 40 at 3.8-Å resolution. *Nature* **354**:278–284.
28. Moens, U., T. Johansen, J. I. Johnsen, O. M. Seternes, and T. Traavik. 1995. Noncoding control region of naturally occurring BK virus variants: sequence comparison and functional analysis. *Virus Genes* **10**:261–275.
29. Moens, U., and M. Van Ghelue. 2005. Polymorphism in the genome of non-passaged human polyomavirus BK: implications for cell tropism and the pathological role of the virus. *Virology* **331**:209–231.
30. Pauw, W., and J. Choufoer. 1978. Isolation of a variant of BK virus with altered restriction endonuclease pattern. *Arch. Virol.* **57**:35–42.
31. Randhawa, P., A. Ho, R. Shapiro, A. Vats, P. Swalsky, S. Finkelstein, J. Uhrmacher, and K. Weck. 2004. Correlates of quantitative measurement of BK polyomavirus (BKV) DNA with clinical course of BKV infection in renal transplant patients. *J. Clin. Microbiol.* **42**:1176–1180.
32. Randhawa, P. S., S. Finkelstein, V. Scantlebury, R. Shapiro, C. Vivas, M. Jordan, M. M. Pickin, and A. J. Demetris. 1999. Human polyoma virus-associated interstitial nephritis in the allograft kidney. *Transplantation* **67**:103–109.
33. Randhawa, P. S., K. Khaleel-Ur-Rehman, P. A. Swalsky, A. Vats, V. Scantlebury, R. Shapiro, and S. Finkelstein. 2002. DNA sequencing of viral capsid protein VP-1 region in patients with BK virus interstitial nephritis. *Transplantation* **73**:1090–1094.
34. Rice, S. J., J. A. Bishop, J. Apperley, and S. D. Gardner. 1985. BK virus as cause of hemorrhagic cystitis after bone marrow transplantation. *Lancet* **ii**:844–845.
35. Saitou, N., and M. Nei. 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
36. Seif, I., G. Khoury, and R. Dhar. 1979. The genome of human papovavirus BKV. *Cell* **18**:963–977.
37. Sneath, P. H. A., and R. R. Sokal. 1973. *Numerical taxonomy*. Freeman, San Francisco, Calif.
38. Stoner, G. L., R. Alappan, D. V. Jobs, C. F. Ryschkewitsch, and M. L. Landry. 2002. BK virus regulatory region rearrangements in brain and cerebrospinal fluid from a leukemia patient with tubulointerstitial nephritis and meningoencephalitis. *Am. J. Kidney Dis.* **39**:1102–1112.
39. Sugimoto, C., K. Hara, F. Taguchi, and Y. Yogo. 1989. Growth efficiency of naturally occurring BK virus variants in vivo and in vitro. *J. Virol.* **63**:3195–3199.
40. Sugimoto, C., K. Hara, F. Taguchi, and Y. Yogo. 1990. Regulatory DNA sequence conserved in the course of BK virus evolution. *J. Mol. Evol.* **31**:485–492.
41. Sundsfjord, A., T. Johansen, T. Flaegstad, U. Moens, P. Villand, S. Subramani, and T. Traavik. 1990. At least two types of control regions can be found among naturally occurring BK virus strains. *J. Virol.* **64**:3864–3871.
42. Takasaka, T., N. Goya, H. Ishida, K. Tanabe, H. Toma, T. Fujioka, S. Omori, H. Y. Zheng, Q. Chen, S. Nukuzuma, T. Kitamura, and Y. Yogo. 2006. Stability of the BK polyomavirus genome in renal-transplant patients without nephropathy. *J. Gen. Virol.* **87**:303–306.
43. Takasaka, T., N. Goya, T. Tokumoto, K. Tanabe, H. Toma, Y. Ogawa, S. Hokama, A. Momose, T. Funyu, T. Fujioka, S. Omori, H. Akiyama, Q. Chen, H. Y. Zheng, N. Ohta, T. Kitamura, and Y. Yogo. 2004. Subtypes of BK virus prevalent in Japan and variation in their transcriptional control region. *J. Gen. Virol.* **85**:2821–2827.
44. Tavis, J. E., D. L. Walker, S. N. Gardner, and R. J. Frisque. 1989. Nucleotide sequence of the human polyomavirus AS virus, and antigenic variant of BK virus. *J. Virol.* **63**:901–911.
45. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 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.
46. Yang, R. C., and R. Wu. 1979. BK virus DNA: complete nucleotide sequence of a human tumor virus. *Science* **206**:456–462.