

# Genotyping Schemes for Polyomavirus BK, Using Gene-Specific Phylogenetic Trees and Single Nucleotide Polymorphism Analysis<sup>∇</sup>

Chunqing Luo,<sup>1</sup> Marta Bueno,<sup>2</sup> Jeffrey Kant,<sup>1,3</sup> Jeremy Martinson,<sup>4</sup> and Parmjeet Randhawa<sup>1\*</sup>

Departments of Pathology,<sup>1</sup> Computational Biology,<sup>2</sup> Human Genetics,<sup>3</sup> University of Pittsburgh, Pittsburgh, Pennsylvania, and Infectious Diseases and Microbiology,<sup>4</sup> Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania

Received 15 October 2008/Accepted 15 December 2008

**BK virus (BKV) genotyping has been historically based on nucleotides 1744 to 1812 in the VP1 gene. We reevaluated this practice by making BKV whole-genome and gene-specific phylogenetic trees as well as performing single nucleotide polymorphism (SNP) analysis of 162 sequences available in the public domain. It was found that currently known BKV subtypes and subgroups can no longer be reliably determined by sequencing certain partial gene sequences. Phylogenetic trees based on large T-antigen (LTA) allow separation of subtype I into subgroups Ia, Ib1, Ib2, and Ic, with bootstrap values of 100%, which are better than bootstraps obtained using VP1 sequences (bootstrap values of 71 to 97%). Subtype IV can be subdivided into subgroups, but LTA bootstrap values (33 to 80%) are lower than those obtained by whole-genome analysis (68 to 87%). Subtypes V and VI provisionally identified earlier on the basis of more limited sequence data are better classified as subgroups Ib2 and Ib1, respectively. LTA positions 3634, 3772, 3934, and 4339 can serve as a minimal SNP set to distinguish between the four major BKV subtypes. No subtype II-, IVa-, or IVb-defining SNPs are available in the VP1 gene. However, the overall congruence of viral strain classification based on either VP1 or LTA phylogenetic analysis indicates that these two areas of the viral genome are genetically linked. Interstrain genetic recombination between distant loci in the VP1 and LTA areas is not a common event.**

Polyomavirus BK (BKV) belongs to the family *Polyomaviridae*. Virions are 45 nm in diameter, with a 5-kb circular double-stranded genome. The viral genome is arranged in three general regions: the noncoding control region (NCCR), the early coding region (coding for the small t antigen and large T antigen [LTA]), and the late coding region coding for the viral capsid proteins (VP1, VP2, and VP3) and agnoprotein (7, 32). The NCCR contains the replication origin and regulatory elements that are important activators of viral transcription. The LTA promotes viral replication by binding to tumor suppressor proteins Rb (retinoblastoma) and p53 and stimulates host cell entry into the cell cycle (8, 12, 28, 38). VP1, VP2, and VP3 are structural proteins required for the assembly of complete virions.

Diseases associated with BKV infection include allograft nephropathy, ureteric stricture, hemorrhagic cystitis, and rare cases of disseminated infection. The role played by viral genomic variation in the pathogenesis of these clinical syndromes is not clear. Nonetheless, it has been shown that subtype I BKV strains grow more efficiently in human renal epithelial cells than subtype IV strains (26). In mice, specific mutations in the VP1 region have been associated with increased pathogenicity (2). Hence, there is a need for additional studies correlating viral subtype with clinical parameters. In addition, knowledge of viral subtypes is needed for quality assurance of diagnostic assays performed in molecular biology laboratories (13). There is evidence that subtypes III and IV do

not amplify well with some PCR primer sets currently in use. Genotyping is also important for tracing infection trails in epidemiologic investigations, documenting infections with multiple viral strains, defining the immune response to viral infection, and designing vaccines with the broadest possible protection. Finally, biologists use subtypes to study how viruses originate, spread globally from the point of initial geographical localization, and evolve into different strains under the forces of natural selection.

The first genotyping schema for BKV described by Jin et al. in 1993 was based on a very short segment of the VP1 gene (nucleotides 1744 to 1812) (15, 16). Four major subtypes (I through IV) were recognized. As our knowledge of BKV genomic diversity expanded, difficulties were encountered in assigning viral strains to existing subtypes (1, 27). Investigators proposed additional subgroups within the four major subtypes, such as subgroups Ia and Ib (35); Ic (36); and IVa, IVb, and IVc (14, 25). Robust biologic and statistical support was not always available for the proposed subgroups. For example, Krumbholz et al. analyzed 60 partial VP1 sequences (nucleotides 1163 to 1913) and obtained bootstrap values of only 19 to 22% for separation of subgroups Ia, Ib, and Ic (19). In this study, we have reviewed all available BKV whole-genome sequences and reappraised BKV genotyping schema using phylogenetic as well as single nucleotide polymorphism (SNP) analysis.

## MATERIALS AND METHODS

**Retrieval of published sequences.** Publicly available whole-genome sequences of 178 BKV strains were retrieved from GenBank according to the published literature. If the whole-genome sequences of several strains were identical, only one whole-genome sequence was retained. All data were collected before 1 July 2008. The accession numbers of these 162 unique sequences are V01109, V01108, M23122, EF376992, DQ989813, DQ989812, DQ989811, DQ989810,

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

<sup>∇</sup> Published ahead of print on 24 December 2008.

DQ989809, DQ989808, DQ989807, DQ989806, DQ989805, DQ989804, DQ989803, DQ989802, DQ989801, DQ989800, DQ989799, DQ989798, DQ989797, DQ989796, DQ989795, DQ989794, DQ305492, AY628238, AY628237, AY628236, AY628235, AY628234, AY628233, AY628232, AY628231, AY628230, AY628229, AY628228, AY628227, AY628226, AY628225, AY628224, AB301103, AB301101, AB301100, AB301097, AB301096, AB301095, AB301094, AB301093, AB301092, AB301091, AB301090, AB301089, AB301088, AB301087, AB301086, AB298947, AB298946, AB298945, AB298942, AB298941, AB269869, AB269868, AB269867, AB269866, AB269865, AB269864, AB269863, AB269862, AB269861, AB269860, AB269859, AB269858, AB269857, AB269856, AB269855, AB269854, AB269853, AB269852, AB269851, AB269850, AB269849, AB269848, AB269847, AB269846, AB269845, AB269844, AB269843, AB269842, AB269841, AB269840, AB269838, AB269837, AB269836, AB269834, AB269832, AB269831, AB269830, AB269829, AB269828, AB269827, AB269826, AB269825, AB269824, AB263938, AB263936, AB263935, AB263934, AB263932, AB263931, AB263930, AB263929, AB263928, AB263927, AB263926, AB263925, AB263924, AB263923, AB263922, AB263921, AB263920, AB263919, AB263918, AB263917, AB263916, AB263915, AB263914, AB263913, AB263912, AB260033, AB260032, AB260031, AB260030, AB260029, AB260028, AB217921, AB217920, AB217919, AB217918, AB217917, AB213487, AB211391, AB211390, AB211389, AB211388, AB211387, AB211386, AB211385, AB211384, AB211383, AB211382, AB211381, AB211379, AB211378, AB211377, AB211376, AB211375, AB211374, AB211373, AB211372, AB211371, AB211370, and AB211369. Sequences were renamed to incorporate the viral subtype as determined by the submitting investigator. In the older publications, assignment of subtype was based on restriction fragment polymorphism analysis or the presence of specific nucleotides at defined locations in the viral genome (15, 17). In more recent publications, subtype assignment was based on clustering of sequences with historically defined strains by phylogenetic methods. In all, 105 subtype I strains, 4 subtype II strains, 2 subtype III strains, and 51 subtype IV strains were analyzed.

**Phylogenetic analyses.** Analysis was carried out for whole-genome sequences as well as the coding sequences derived from the agnogene, VP1, VP2, VP3, small t antigen gene, and LTA gene. The intron of LTA was not included. Sequence alignments were performed with ClustalW (37) at the EMBL-EBI website (<http://www.ebi.ac.uk/clustalw/>) (4) using default parameters, followed by manual adjustment using known landmarks in the viral genome. Sequences were numbered using BKV Dunlop as the reference strain (accession no. V01108) following the system of Seif et al., in which nucleotide position 1 is the NCCR position next to the start codon of LTA (31). Neighbor-joining (NJ) (30) trees were constructed with MEGA version 4.1 (20). Divergences were estimated by Kimura's two-parameter method. All phylogenetic trees were visualized using MEGA 4.1 Tree explorer (20). A bootstrap test with 1,000 replicates was used to estimate the confidence of the branching patterns of the trees (9).

**Analysis of gene polymorphisms.** Consensus sequences for all subtype- and subgroup-specific sequences were obtained using the following stringent consensus generation criteria implemented in Sequencher 4.0 (Gene Codes Corporation, Ann Arbor, MI). (i) Where there was just one sequence, the consensus was N (rule 1). (ii) Where there were between two and four sequences, if all agreed, the consensus was the sequence contribution; any unconfirmed position was N (rule 2). (iii) Where there were between five and seven sequences, if all or all but one agreed, the consensus was the sequence contribution; any unconfirmed position was N (rule 3). (iv) Where there were eight or more sequences, if all or all but one or two agreed, the consensus was the sequence contribution; any unconfirmed position was N (rule 4).

Based on the number of available subtype- and subgroup-specific sequences, rule 2 was applied for subtype II/III consensus sequences and rule 4 for subtype I and IV consensus sequences. Thus, up to two disagreements (including singletons which can potentially represent an unfixed random mutation or a PCR artifact) were ignored in generating a subtype- or subgroup-specific consensus sequence by rule 4. Consensus sequences were then aligned and evaluated for the presence of subtype- and subgroup-specific polymorphic sites, which were defined as those where at least two alternate nucleotides were present. Single-nucleotide differences comparing a string of any two closely aligned sequences were referred to as SNPs. The term "informative SNP" was applied to a nucleotide change which could assist in separating one viral subtype or subgroup from another. All subtype or subgroup assignments based on the consensus were manually verified against relevant available genome sequences.

**Mapping of gene polymorphisms to VP1 and LTA proteins.** BKV VP1 (Swiss-Prot P03088) and LTA (P03070) were modeled with MODELLER9v1 (23), using the available tridimensional structures of simian virus 40 (SV40) VP1 (Protein Data Bank code 1SVA) and LTA (Protein Data Bank code 1SVL) (11), respectively. The identity between template and target sequences varied from 85% (VP1) to 77% (J domain of LTA). The original model was refined by

performing a moderate number (500, as 10 rounds of 50 steps) of adopted basis Newton-Raphson minimization steps in CHARMm (3). The model was then analyzed with PROCHECK (21), which detected no geometrical errors.

**Analysis of genetic recombination.** Subtype- and subgroup-specific SNPs were extracted, aligned, and examined visually for interstrain recombination. Additionally, phylogenetic trees made using whole-genome sequences were compared for incongruencies with respect to trees made with gene specific DNA sequences. The extent of putative recombination among sequences was further examined in diversity plots (22). The observed sequence differences between the Dunlop reference strain (accession no. V01108) and other viral strains were calculated for windows of 500 sites and moved in steps of 50 nucleotides.

## RESULTS

**Phylogenetic analyses of whole-genome sequences.** Using whole-genome sequences and the NJ Kimura two-parameter method, 162 unique viral strains were clustered into previously defined subtypes I, II, III, and IV with high bootstrap values (Fig. 1). Subtypes II and III were more closely related to each other than subtypes I and IV. Subtypes V and VI, previously defined by us on the basis of more limited sequence data, are best considered to be subgroups of subtype I and clustered with Ib2 and Ib1, respectively. Subgroups of subtype I are shown to be definite phylogenetic entities separated by bootstrap values of 100%, which are much better than those previously obtained using more limited sequence information based on the VP1 gene (33). Bootstrap values for subgroups of subtype IV ranged from 68 to 87%. The phylogenetic analyses performed are summarized in Fig. 1 to 3 and Table 1.

**Phylogenetic analyses of VP1 sequences.** All major subtypes seen on whole-genome analyses were confirmed in phylogenetic trees constructed with the complete VP1 gene (Fig. 2). However, lower bootstrap separation values were obtained for subtype II (60% for VP1 versus 100% for the whole genome), Ib (55% for VP1 versus 87% for the whole genome), IVa (34% versus 83%), and IVc (21% versus 68%). Whereas subgroups IVb1 and IVb2 clustered together in the whole-genome tree, IVb2 strains separated out in the VP1 trees and clustered together with IVc strains (Fig. 2). Thus, to completely define these subtypes and subgroups, genetic information extending beyond the VP1 region was needed. Distinctions between major subtypes and subgroups blurred further if phylogenetic trees were made using the 327-bp sequence (nucleotides 1630 to 1956, Dunlop numbering) that has been used in some publications on BKV genotyping. In particular, it was no longer possible to distinguish subgroups of subtype IV from each other. Thus, with a rapidly expanding database of DNA sequences, a genotyping schema based entirely on VP1 can no longer adequately capture the genetic diversity of BKV.

**Phylogenetic analyses of LTA sequences.** In general, phylogenetic trees based on LTA sequences validated BKV major subtypes and subtypes identified by trees using whole-genome or VP1 sequences (Fig. 3). In fact, clade separation was better than that achieved by VP1 trees. Thus, compared to VP1 trees, LTA trees were associated with higher bootstrap separation values for subtype II (100% versus 60%), subgroup Ib1 (100% versus 71%), and IVa (80% versus 34%) (Table 1). Whereas, IVb2 strains clustered together with IVc strains with a low bootstrap value (22%) in VP1 trees, subgroups IVb1 and IVb2 clustered together in the LTA trees (bootstrap value of 56%). These differences reflect the fact that LTA is a larger gene with more informative sites available than the VP1 gene (Fig. 3). In



FIG. 1. Phylogenetic tree constructed by the NJ method using BKV whole-genome sequences.

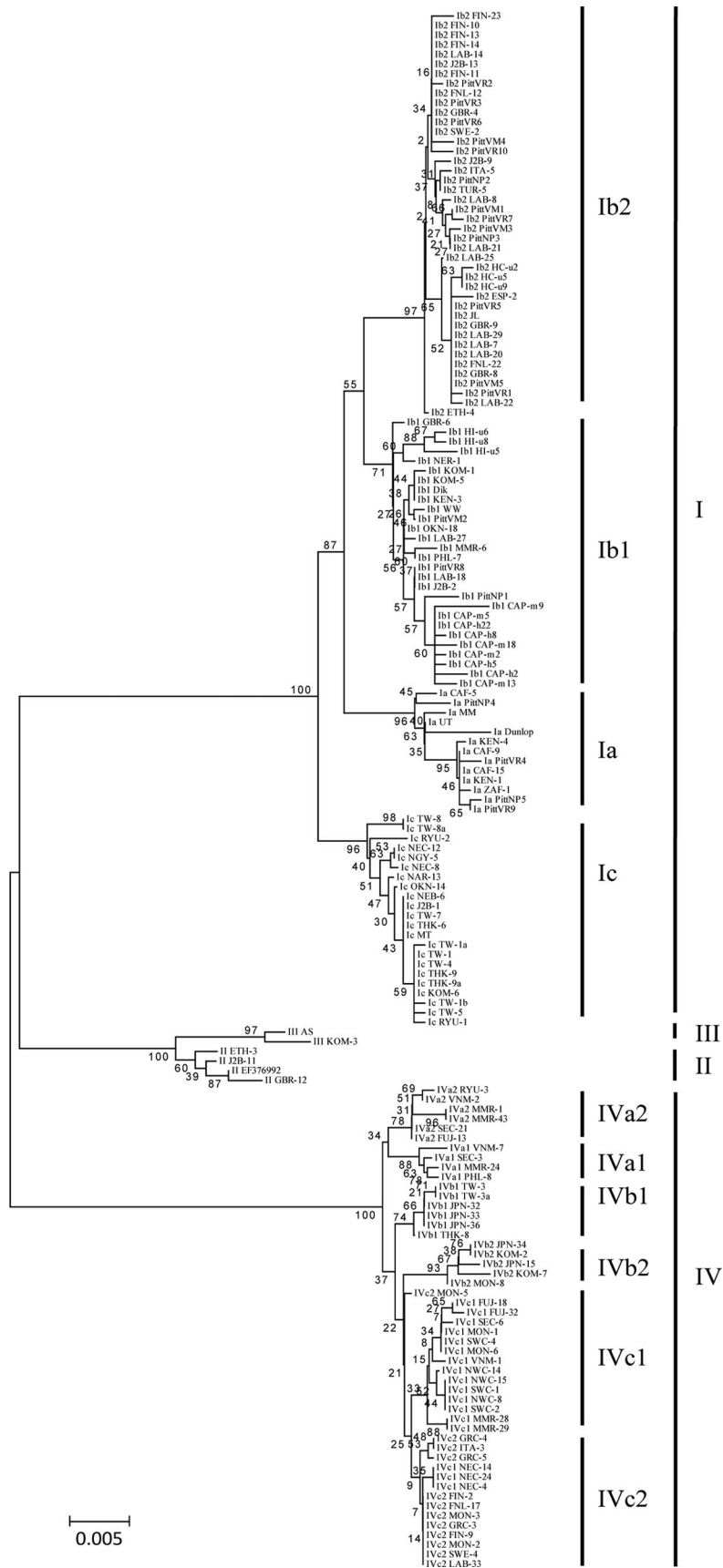


FIG. 2. Phylogenetic tree constructed by the NJ method using BKV VP1 gene sequences.

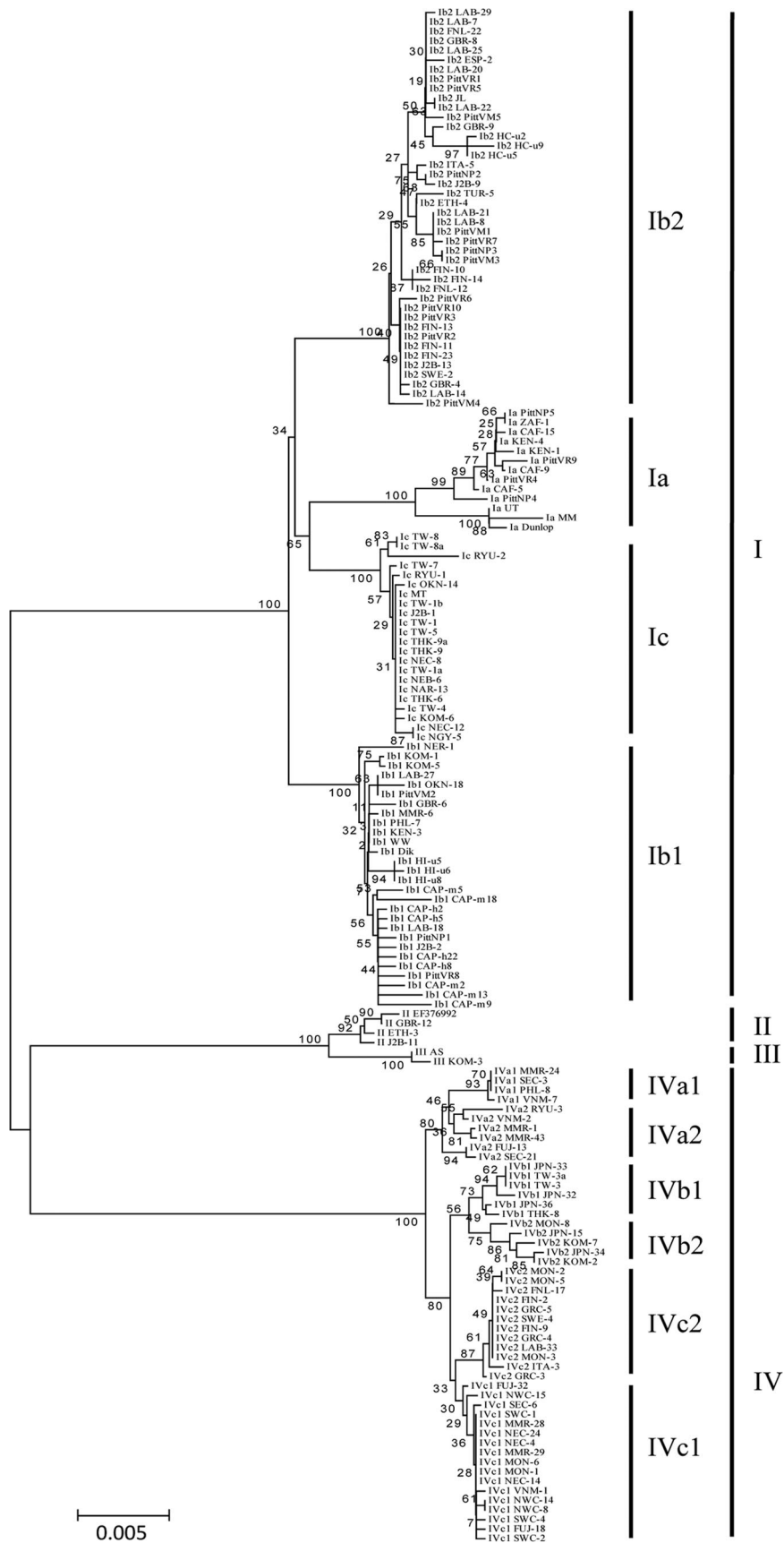


FIG. 3. Phylogenetic tree constructed by the NJ method using BKV LTA gene sequences.

TABLE 1. Efficiency of bootstrap phylogenetic trees constructed by different BKV regions

Subtype or subgroup comparison <sup>a</sup>	No. of strains	WGS (1–5153)		VP1 (1564–2652)		Jin, 327 bp (1630–1956)		LTA gene (2722–5153)		LTA 2nd exon (2722–4566)		LTA, 325 bp (3148–3472)	
		BS (%) <sup>b</sup>	No. (%) of strains correctly assigned/total	BS (%) <sup>b</sup>	No. (%) of strains correctly assigned/total	BS (%) <sup>b</sup>	No. (%) of strains correctly assigned/total	BS (%) <sup>b</sup>	No. (%) of strains correctly assigned/total	BS (%) <sup>b</sup>	No. (%) of strains correctly assigned/total	BS (%) <sup>b</sup>	No. (%) of strains correctly assigned/total
<b>Subtypes</b>													
I vs II/III/IV	105	100	105/105 (100)	100	105/105 (100)	100	105/105 (100)	100	105/105 (100)	100	105/105 (100)	99	105/105 (100)
II vs III	4	100	4/4 (100)	60	4/4 (100)	39	4/4 (66.7)	92	4/4 (100)	85	4/4 (100)	64	2/4 (50)
III vs II	2	100	2/2 (100)	97	2/2 (100)	94	2/2 (100)	100	2/2 (100)	100	2/2 (100)	89	2/2 (100)
IV vs II/III	51	100	51/51 (100)	100	51/51 (100)	99	51/51 (100)	100	51/51 (100)	100	51/51 (100)	100	51/51 (100)
<b>Subgroups</b>													
Ib2 vs Ib1/Ia/Ic	42	100	42/42 (100)	97	42/42 (100)	48	42/42 (100)	100	42/42 (100)	99	42/42 (100)	57	42/42 (100)
Ib1 vs Ib2/Ia/Ic	28	100	28/28 (100)	71	28/28 (100)	34	27/28 (96.43)	100	28/28 (100)	99	28/28 (100)	97	28/28 (100)
Ia vs Ib1/Ib2/Ic	13	100	13/13 (100)	96	13/13 (100)	28	13/13 (100)	100	13/13 (100)	99	13/13 (100)	96	13/13 (100)
Ic vs Ib1/Ib2/Ia	22	100	22/22 (100)	96	22/22 (100)	47	20/22 (90.91)	100	22/22 (100)	100	22/22 (100)	90	22/22 (100)
IVa vs IVb/IVc	10	83	10/10 (100)	34	10/10 (100)	NA <sup>d</sup>	0/10 (0)	80	10/10 (100)	79	10/10 (100)	39	10/10 (100)
IVb vs IVa/IVc	11	87	11/11 (100)	NA <sup>c</sup>	6/11 (54.55)	NA <sup>d</sup>	0/11 (0)	56	11/11 (100)	62	11/11 (100)	NA <sup>e</sup>	0/11 (0)
IVc vs IVa/IVb	30	68	30/30 (100)	21	30/30 (100)	NA <sup>d</sup>	0/30 (0)	33	30/30 (100)	29	30/30 (100)	NA <sup>e</sup>	0/30 (0)

<sup>a</sup> The subtype and subgroup assigned from all strains in different trees were compared with the WGS tree which was assumed to be 100% correct.

<sup>b</sup> BS, bootstrap value.

<sup>c</sup> Not applicable (NA) because although six subgroup IVb1 strains were clustered together and five subgroup IVb2 strains were also clustered together, subgroup IVb2 clustered within the IVc subgroup.

<sup>d</sup> Not applicable because all subgroup IVa, IVb, and IVc strains can't be separated from each other.

<sup>e</sup> Not applicable because subgroup IVb and IVc strains were clustered together and can't be separated from each other.

an effort to reduce the amount of sequence information required to reliably classify BKV strains, phylogenetic trees were also constructed with the 1,845-bp sequence of the LTA second exon (nucleotides 2722 to 4,566) and a 325-bp sequence (nucleotides 3148 to 3472) arbitrarily chosen to be of approximately the same size as the VP1 sequence used in many publications on BKV genotyping. The second exon sequence was virtually as effective as the entire LTA sequence in separating out major clades (Table 1). The 325-bp sequence allowed separation of subtypes I, III, and IV with bootstrap values ranging from 89 to 100%. However, there was a fall in bootstrap values associated with subgroups Ib2 (57%) and IVa (39%). Subtype II and subgroups IVb and IVc could not be clearly resolved. Thus, BKV phylogeny can not be reliably determined by sequencing short amplicons that are typical of current molecular diagnostic assays used in clinical practice.

**Phylogenetic analysis of other BKV genes.** Agnogene sequences are short, with few informative sites, and could not be used to separate the major BKV subtypes (Table 2). VP2 gene region sequences could resolve the subtypes but not the subgroups. The VP3 gene is located within VP2 gene and contains most of its informative sites. The portion of VP2 gene that extends beyond VP3 contains five subtype-specific SNPs (see below). Phylogenetic trees based on small-T-antigen sequences can divide BKV into major subtypes and all subgroups of subtype I but cannot resolve subgroups of subtype IV (data not shown).

**Use of SNPs for determination of BKV subtypes and subgroups.** DNA sequencing followed by construction of phylogenetic trees is not a practical method for BKV typing in a routine diagnostic laboratory. Therefore, we analyzed all whole-genome sequences for SNPs that might uniquely identify major BKV subtypes and subgroups. The locations of all subtype-informative SNPs are given in Table 2. Table 3 lists subgroup-informative SNPs within specific subtypes. Tables 4 and 5 specify the actual nucleotides and amino acids that define these SNPs. Phylogeneti-

cally informative SNPs are available in all genes; the largest number belongs to LTA, followed by VP1 and VP2/VP3. The LTA region contains 23 subtype I-, 2 subtype II-, 13 subtype III-, and 34 subtype IV-specific SNPs. Specifically, a 321-bp region spanning nucleotides 3150 to 3470 (Dunlop numbering) contains several SNPs that are unique to subtype I, while others are subtype IV specific. SNPs at positions 1367, 2112, 2370, 2550, 3634, and 4339 (Dunlop numbering) are informative for multiple subtypes and may be regarded as "super" SNPs. A minimal SNP set encompassing LTA positions 3634, 3772, 3934, and 4339 can be used to distinguish between the four major subtypes of BKV. Within the VP1 region, 14 subtype I-, 7 subtype III-, and 24 subtype IV-specific SNPs are found, but none specific for subtype II or subgroups of subtype IV are present. Even if whole-genome sequence data are considered, there are only five subtype II-specific SNPs available at this time. Two of these SNPs are located in the LTA region, and one each is located in the small t, VP2, and agnogene regions. The majority of the SNPs were synonymous. Notable exceptions included four SNPs in VP1 (L1766Q, L1767Q, D1787A, and S1793D), of which two (positions 1787 and 1793) localized to surface loop structures. Additionally, SNPs at positions 1769 and 1770 resulted in a conservative change (lysine to arginine) in the 69th amino acid of the VP1 gene, a site predicted to interact with the cellular receptor for BKV. Nonconservative amino acid changes affecting functional domains in the LTA include those localized to the origin binding domain (Q4298L, Q4299L, and H4080Y) and the helicase domain (T3035A and T3036Q).

**Recombination analyses.** In general, VP1 and LTA sequences generated phylogenetic trees similar to those obtained with whole-genome sequences. Absence of mosaicism in the diversity plots also suggested that major interstrain genetic recombination events are not a common event for BKV. Based on SNP analysis, a few potentially recombinant sequences were noted, and two of these are described as follows:

(i) A subtype II sequence, GBR-12 (AB263920), has three

TABLE 2. SNP positions informative for the BKV genotype

Gene (positions)	Subtype(s)	No. of SNPs	SNP position(s) (Dunlop numbering)
Agnogene (388–588)	II	1	427
VP2/VP3 (624–1679)	I	17	848, 1067, 1154, 1217, 1274, 1284, 1304, 1316, 1342, 1347, 1361, 1364, 1367, 1400, 1422, 1425, 1514
	II and III	10	734, 781, 815, 1064, 1133, 1172, 1193, 1199, 1223, 1367
	II	1	1409
	III	1	1247
	IV	8	986, 1187, 1322, 1337, 1367, 1389, 1427, 1429
VP1 (1564–2652)	I	14	1760, 1787, 1793, 1848, 1912, 1978, 2073, 2112, 2237, 2325, 2370, 2544, 2550, 2583
	II and III	11	1824, 1858, 1971, 2086, 2112, 2142, 2274, 2370, 2391, 2510, 2550
	III	7	1766, 1767, 1768, 1770, 1857, 2259, 2559
	IV	24	1704, 1769, 1770, 1784, 1854, 1869, 1938, 1965, 2007, 2013, 2034, 2067, 2109, 2112, 2184, 2199, 2235, 2274, 2370, 2406, 2413, 2457, 2541, 2550
LTA (2722–4566, 4911–5153)	I	23	2802, 3025, 3035, 3133, 3229, 3232, 3421, 3430, 3450, 3451, 3478, 3579, 3634, 3868, 3979, 4024, 4069, 4076, 4105, 4153, 4339, 4444, 5139
	II and III	28	2788, 2791, 2802, 2809, 3039, 3079, 3081, 3139, 3250, 3303, 3481, 3501, 3511, 3570, 3577, 3589, 3634, 4090, 4330, 4339, 4483, 5141, 5142
	II	2	3934, 4270
	III	13	2808, 3070, 3124, 3193, 3409, 3772, 3877
	IV	34	2920, 3036, 3058, 3121, 3157, 3195, 3202, 3232, 3265, 3376, 3400, 3469, 3580, 3634, 3757, 3772, 3781, 3829, 3859, 3871, 3916, 3942, 3955, 3985, 4021, 4080, 4081, 4298, 4299, 4339, 4459, 4462, 4944, 5061
Small t antigen (4911–4635 [partial])	I	2	4836, 4806
	II and III	7	4878, 4877, 4830, 4764, 4734, 4704, 4692
	II	1	4726
	IV	1	4797

TABLE 3. Genomic positions of SNPs informative for BKV subgroups

Gene (positions)	Subgroup	No. of SNPs	SNP position(s) (Dunlop numbering)	
Agnogene (388–588)	Ib2	1	427	
VP2/VP3 (624–1679)	Ia	4	1023, 1146, 1169, 1272	
	Ib1	2	1091, 1322	
	Ib2	1	1022	
	Ic	1	1166	
	IVb	1	1343	
VP1 (1564–2652)	Ia	1	1989	
	Ib1	1	2076	
	Ib2	3	1575, 1908, 2127	
	Ic	2	1989, 1992	
	IVc	1	1977	
LTA (2722–4566, 4911–5153)	Ia	11	2908, 3172, 3190, 3424, 3562, 3709, 3844, 4075, 5055, 5076	
	Ib1	3	3652, 4417, 4435	
	Ib2	4	3654, 3673, 3749, 5103	
	Ic	7	2761, 3079, 3100, 3238, 3315, 3523, 4947	
	IVa	3	3454, 3919	
	IVb	1	3535	
	IVc	1	4525	
	Small t antigen (4911–4635 [partial])	None		

SNPs (C1701T, G1723A, and A/C1726G) which make this portion of the BKV genomic sequence different from the remaining known subtype II (ETH-3 [AB263916], J/1025/05 [EF376992], J2B-11 [AB301101]) and subtype III (AS [M23122 and KOM-3 [AB211386]) strains but identical to 51 known subtype IV strains. The concatenated SNP sequence TAG may represent a recombination between subtype II and subtype IV strains.

(ii) A subtype II sequence, J2B-11 (AB301101), has two SNPs (A2325G and C2337T) which make this portion of the BKV genomic sequence different from the remaining known subtype II strains (ETH-3 [AB263916], J/1025/05 [EF376992], and GBR-12 [AB263920]) but identical to both known subtype III strains (AS [M23122] and KOM-3 [AB211386]) and 51 known subtype IV strains. In this instance, the two SNP positions could potentially be explained as a recombination event between subtype II and subtype III/IV strains. However, one cannot exclude the possibility of two independent mutations resulting in the observed difference.

**DISCUSSION**

The first genotyping scheme for BKV was described by Jin et al. in 1993 (15, 16). These investigators defined four subtypes using restriction fragment length polymorphism and DNA sequencing spanning a short region of the VP1 gene (nucleotides 1744 to 1812). The four major subtypes broadly correlated with serotypes characterized earlier by Knowles et al. (18). In 2002, Stoner et al. suggested that subtype I could be divided into subgroups Ia and Ib (35), but it was not clear if these represented biologically relevant categories, since the defining nucleotide

TABLE 4. Subtype-specific BKV SNPs

Position <sup>a</sup>	BKV Dunlop strain sequence <sup>b</sup>	Nucleotide homology with Dunlop strain sequence <sup>c</sup>									Specificity <sup>d</sup>	
		Ia	Ib1	Ib2	Ic	II	III	IVa	IVb	IVc		
Agnogene												
427 (1)	GTT (V)	...	...	C.. (L)	...	A.. (I)	...	..N (X)	...	...	...	II
VP2												
734 (3)	GCT (A)	...	...	...	...	..C	..C	...	...	...	...	II, III
781 (2)	AGT (S)	...	...	...	...	.C (T)	.C (T)	...	...	...	...	II, III
815 (3)	ACT (T)	...	...	...	...	..A	..A	...	...	...	...	II, III
848 (3)	CCT (P)	...	...	...	...	..A	..A	..A	..A	..A	..A	I
986 (3)	GCT (A)	...	...	...	...	...	...	..A	..A	..A	..A	IV
1064 (3)	TAC (Y)	...	...	...	...	..T	..T	...	...	...	...	II, III
1067 (3)	CTT (L)	...	...	...	...	..A	..A	..A	..A	..A	..A	I
1133 (3)	AGG (R)	...	...	...	...	..A	..A	...	...	...	...	II, III
1154 (3)	ACC (T)	...	...	...	...	..T	..T	..T	..T	..N (X)	..N (X)	I
1172 (3)	AGA (R)	...	...	...	...	..G	..G	...	...	...	...	II, III
1187 (3)	TTT (F)	...	...	...	...	...	...	..C	..C	..C	..C	IV
1193 (3)	AGA (R)	...	...	...	...	..G	..G	...	...	...	...	II, III
1199 (3)	TCC (S)	...	...	...	...	..T	..T	...	...	...	...	II, III
1217 (3)	GAG (E)	...	...	...	...	..A	..A	..A	..A	..A	..A	I
1223 (3)	ACT (T)	...	...	...	...	..C	..C	...	...	...	...	II, III
1247 (3)	CCT (P)	...	...	...	...	...	..C	...	...	...	...	III
1274 (3)	CAA (Q)	...	G.. (E)	G.. (E)	G.. (E)	G.T (D)	G.T (D)	G.T (D)	G.T (D)	G.T (D)	G.T (D)	I
1284 (1)	GAT (D)	...	...	...	...	A.. (N)	A.. (N)	A.. (N)	A.. (N)	A.. (N)	A.. (N)	I
1304 (3)	CCC (P)	...	...	...	...	..T	..T	..T	..T	..T	..T	I
1316 (3)	AGA (R)	...	...	...	...	..G	..G	..G	..G	..G	..G	I
1322 (3)	GTA (V)	...	..G	...	...	...	...	..T	..T	..T	..T	IV
1337 (3)	GGT (G)	...	...	...	...	...	...	..A	..A	..A	..A	IV
1342 (2)	CGT (R)	...	...	...	...	..A (H)	..A (H)	..AG (Q)	..AA (Q)	..AG (Q)	..AG (Q)	I
1347 (1)	CAT (H)	...	...	...	...	A.. (N)	A.. (N)	A.. (N)	A.. (N)	A.. (N)	A.. (N)	I
1361 (3)	ACT (T)	...	...	...	...	..C	..C	..C	..C	..C	..C	I
1364 (3)	TAT (Y)	...	...	...	...	..C	..C	..C	..C	..C	..C	I
1367 (3)	AGT (S)	...	...	...	...	..C	..C	..A (R)	..A (R)	..A (R)	..A (R)	I, II, III, IV
1389 (1)	GAA (E)	...	...	...	...	...	...	C.. (Q)	C.. (Q)	C.. (Q)	C.. (Q)	IV
1400 (3)	ACA (T)	...	...	...	...	..C	..C	..C	..C	..C	..C	I
1409 (3)	ATG (M)	...	...	...	...	..A (I)	..A (I)	...	...	...	...	II
1422 (1)	CAA (Q)	...	...	...	...	A.G (K)	A.G (K)	A.G (K)	A.G (K)	A.. (K)	A.. (K)	I
1425 (1)	CAA (Q)	...	...	...	...	G.. (E)	G.. (E)	G.G (E)	G.G (E)	G.G (E)	G.G (E)	I
1427 (3)	CAA (Q)	...	...	...	...	G.. (E)	G.. (E)	G.G (E)	G.G (E)	G.G (E)	G.G (E)	IV
1429 (2)	AGT (S)	...	..C (T)	..C (T)	...	...	...	..A (N)	..A (N)	..A (N)	..A (N)	IV
1514 (3)	TTA (L)	...	...	...	...	..G	..G	..G	..G	..G	..G	I
VP1												
1704 (3)	GAG (E)	...	...	...	...	...	...	..A	..A	..A	..A	IV
1760 (2)	TTT (F)	...	...	...	...	..A (Y)	..A (Y)	..A (Y)	..AN (X)	..A (Y)	..A (Y)	I
1766 (2)	CTA (L)	...	...	...	...	...	..AG (Q)	...	...	...	...	III
1767 (3)	CTA (L)	...	...	...	...	...	..AG (Q)	...	...	...	...	III
1768 (1)	AAG (K)	...	...	...	...	...	C.C (H)	..GA (R)	..GA (R)	..GA (R)	..GA (R)	III
1769 (2)	AAG (K)	...	...	...	...	...	C.C (H)	..GA (R)	..GA (R)	..GA (R)	..GA (R)	IV
1770 (3)	AAG (K)	...	...	...	...	...	C.C (H)	..GA (R)	..GA (R)	..GA (R)	..GA (R)	III, IV
1784 (2)	AAT (N)	...	...	...	...	...	...	..C (T)	..C (T)	..C (T)	..C (T)	IV
1787 (2)	GAC (D)	...	...	...	...	..C (A)	..C (A)	..C (A)	..C (A)	..C (A)	..C (A)	I
1793 (2)	AGC (S)	...	...	...	...	GA. (D)	GAN (X)	GAN (X)	GA. (D)	GAN (X)	GAN (X)	I
1824 (3)	CCC (P)	...	...	...	...	..T	..T	...	...	...	...	II, III
1848 (3)	CCC (P)	...	...	...	...	..A	..A	..A	..A	..A	..A	I
1854 (3)	CCC (P)	...	...	...	...	...	...	..T	..T	..T	..T	IV
1857 (3)	AAT (N)	...	...	...	...	...	..C	...	...	...	...	III
1858 (1)	TTA (L)	...	...	...	..G	C..	C..	...	..G	..G	..G	II, III
1869 (3)	GAC (D)	...	...	...	...	...	...	..T	..T	..T	..T	IV
1912 (1)	CAA (Q)	...	...	...	...	A.. (K)	A.. (K)	A.. (K)	A.. (K)	A.. (K)	A.. (K)	I
1938 (3)	AGC (S)	...	...	...	...	...	...	..T	..T	..T	..T	IV
1965 (3)	CAA (Q)	...	...	...	...	...	...	..G	..G	..G	..G	IV
1971 (3)	GTG (V)	...	...	...	..A	..T	..T	..A	..A	..A	..A	II, III
1978 (1)	CAT (H)	...	...	...	...	A.. (N)	A.. (N)	A.. (N)	A.. (N)	A.. (N)	A.. (N)	I
2007 (3)	AGT (S)	...	...	...	...	...	...	..C	..C	..C	..C	IV
2013 (3)	TTC (F)	...	...	...	...	...	...	..T	..T	..T	..T	IV
2034 (3)	GGA (G)	...	...	...	...	...	...	..G	..G	..G	..G	IV

Continued on following page



TABLE 4—Continued

Position <sup>a</sup>	BKV Dunlop strain sequence <sup>b</sup>	Nucleotide homology with Dunlop strain sequence <sup>c</sup>									Specificity <sup>d</sup>
		Ia	Ib1	Ib2	Ic	II	III	IVa	IVb	IVc	
2067 (3)	AAT (N)	...	...	...	...	...	...	..C	..C	..C	IV
2073 (3)	AGG (R)	...	...	...	...	..A	..A	..A	..A	..A	I
2086 (1)	GAT (D)	...	..N (X)	..A (E)	..A (E)	C.A (Q)	C.A (Q)	..A (E)	..A (E)	..A (E)	II, III
2109 (3)	AAC (N)	...	...	...	...	...	...	..T	..T	..T	IV
2112 (3)	CCA (P)	...	...	...	...	..T	..T	..C	..C	..C	I, II, III, IV
2142 (3)	GAC (D)	...	...	...	...	..T	..T	...	...	...	II, III
2184 (3)	GAG (E)	...	...	...	...	...	...	..A	..A	..A	IV
2199 (3)	GAT (D)	...	...	...	...	...	...	..C	..C	..C	IV
2235 (3)	ACT (T)	...	...	...	...	...	...	..A	..A	..A	IV
2237 (2)	TTC (F)	...	...	...	..G (L)	..A. (Y)	..A. (Y)	..A. (Y)	..A. (Y)	..A. (Y)	I
2259 (3)	CCC (P)	...	...	...	...	...	..T	...	...	...	III
2274 (3)	GTG (V)	...	...	...	...	..T	..T	..A	..A	..A	II, III, IV
2325 (3)	CTT (L)	...	...	...	...	..N (X)	..G	..G	..G	..G	I
2370 (3)	GGC (G)	...	...	...	...	..G	..G	..A	..A	..A	I, II, III, IV
2391 (3)	GGA (G)	...	...	...	...	..G	..G	...	...	...	II, III
2406 (3)	AGA (R)	...	...	...	...	...	...	..G	..G	..G	IV
2413 (1)	GCA (A)	...	...	...	...	...	...	C.. (P)	C.. (P)	C.. (P)	IV
2457 (3)	AAT (N)	...	...	...	...	...	...	..C	..C	..C	IV
2510 (2)	AGA (R)	...	...	...	...	..A. (K)	..A. (K)	...	...	...	II, III
2541 (3)	GAA (E)	...	...	...	...	...	...	..G	..G	..G	IV
2544 (3)	TCC (S)	...	...	...	...	..T	..T	..T	..T	..T	I
2550 (3)	GTA (V)	...	...	...	...	..T	..T	..G	..G	..G	I, II, III, IV
2559 (3)	GTT (V)	...	..N (X)	..N (X)	..C	..G	..A	..C	..C	..C	III
2583 (3)	AGA (R)	...	...	..N. (X)	..A. (K)	CAG (Q)	CAG (Q)	CAG (Q)	NAG (X)	CAG (Q)	I
LTA gene											
5142 (3)	GTT (V)	..N (X)	..G	..G	..G	..C	..C	..G	..G	..G	II, III
5141 (1)	CTT (L)	...	...	...	...	T.A	T.A	..A	..A	..A	II, III
5139 (3)	CTT (L)	...	...	...	...	T.A	T.A	..A	..A	..A	I
5061 (3)	AGA (R)	...	...	...	...	...	...	..G	..G	..G	IV
4944 (3)	CAT (H)	...	...	...	...	...	...	..C	..C	..C	IV
4483 (3)	GAA (E)	...	...	...	...	..G	..G	...	...	...	II, III
4462 (3)	GAA (E)	...	...	...	...	...	...	..G	..G	..G	IV
4459 (3)	GAA (E)	...	...	...	...	...	...	..G	..G	..G	IV
4444 (3)	TCT (S)	...	...	...	...	..C	..C	..C	..C	..C	I
4339 (3)	ACC (T)	...	...	...	...	..A	..A	..T	..T	..T	I, II, III, IV
4330 (3)	TGC (C)	...	...	...	...	..T	..T	...	...	...	II, III
4315 (3)	ACT (T)	...	...	...	...	..C	..C	..N (X)	...	...	II, III, IVa1
4299 (1)	CAA (Q)	...	...	...	...	...	...	TT. (L)	TT. (L)	TT. (L)	IV
4298 (2)	CAA (Q)	...	...	...	...	...	...	TT. (L)	TT. (L)	TT. (L)	IV
4270 (3)	AAA (K)	...	...	...	...	..G	...	...	...	...	II
4153 (3)	ACC (T)	...	...	...	...	..T	..T	..T	..T	..T	I
4105 (3)	TAT (Y)	...	...	...	...	..C	..C	..C	..C	..C	I
4090 (3)	AGA (R)	...	...	...	...	..G	..G	...	...	...	II, III
4081 (3)	TAC (Y)	...	...	...	...	...	...	..T	..T	..T	IV
4080 (1)	CAT (H)	...	...	...	...	...	...	T.. (Y)	T.. (Y)	T.. (Y)	IV
4076 (2)	ACT (T)	...	..A	..A	..A	..TA (I)	..TA (I)	..TA (I)	..TA (I)	..TA (I)	I
4069 (3)	GAA (E)	...	...	...	...	..G	..G	..G	..G	..G	I
4024 (3)	GAA (E)	...	...	...	...	..G	..G	..G	..G	..G	I
4021 (3)	GAG (E)	...	...	...	...	...	...	..A	..A	..A	IV
3985 (3)	ATT (I)	...	...	...	...	...	...	..A	..A	..A	IV
3979 (3)	GAG (E)	...	...	...	...	..A	..A	..A	..A	..A	I
3955 (3)	GAG (E)	...	...	...	...	...	...	..A	..A	..A	IV
3942 (1)	TTA (L)	...	...	...	...	...	...	C..	C..	C..	IV
3934 (3)	GGT (G)	...	...	...	...	..A	...	...	...	...	II
3916 (3)	CAA (Q)	...	...	...	...	...	...	..G	..G	..G	IV
3877 (3)	GAC (D)	...	...	...	...	...	..T	...	...	...	III
3871 (3)	CCT (P)	...	...	...	...	...	...	..C	..C	..C	IV
3868 (3)	TAT (Y)	...	...	...	...	..C	..C	..C	..C	..C	I
3859 (3)	AAG (K)	...	...	...	...	...	...	..A	..A	..A	IV
3829 (3)	ATT (I)	...	..C. (T)	..C. (T)	..C. (T)	...	...	..CC (T)	..CC (T)	..CC (T)	IV
3781 (3)	GTA (V)	...	...	...	...	...	...	..G	..G	..G	IV
3772 (3)	GTT (V)	...	...	...	...	...	..A	..G	..G	..G	III, IV
3757 (3)	AGA (R)	...	...	...	...	...	...	..G	..G	..G	IV
3634 (3)	GGT (G)	...	...	...	...	..G	..G	..A	..A	..A	I, II, III, IV

Continued on following page

TABLE 4—Continued

Position <sup>a</sup>	BKV Dunlop strain sequence <sup>b</sup>	Nucleotide homology with Dunlop strain sequence <sup>c</sup>									Specificity <sup>d</sup>	
		Ia	Ib1	Ib2	Ic	II	III	IVa	IVb	IVc		
3589 (3)	ATA (I)	...	...	...	...	..T	..T	...	...	...	...	II, III
3580 (3)	TTT (F)	...	...	...	...	...	...	..C	..C	..C	..C	IV
3579 (1)	TTG (L)	...	...	...	...	C.T	C.T	C..	C..	C..	C..	I
3577 (3)	TTG (L)	...	...	...	...	C.T	C.T	C..	C..	C..	C..	II, III
3570 (1)	ATT (I)	...	...	...	...	G.. (V)	G.. (V)	...	...	...	...	II, III
3511 (3)	GGA (G)	...	...	...	...	..C	..C	...	...	...	...	II, III
3501 (1)	CTA (L)	...	...	...	...	T..	T..	...	...	...	...	II, III
3481 (3)	GAT (D)	...	...	...	...	..C	..C	...	...	...	...	II, III
3478 (3)	TTG (L)	...	...	...	...	..A	..A	..A	..A	..A	..A	I
3469 (3)	GGT (G)	...	...	...	...	...	...	..C	..C	..C	..C	IV
3454 (3)	GTA (V)	...	...	...	...	..N (X)	...	..T	..N (X)	..G	..G	IVa
3451 (3)	AAC (N)	...	...	...	...	..T	..T	..T	..T	..T	..T	I
3450 (1)	CTA (L)	...	...	...	...	T..	T..	T..	T..	T..	T..	I
3430 (3)	ACC (T)	...	...	...	...	..T	..T	..T	..T	..T	..T	I
3421 (3)	CTA (L)	...	...	...	...	..G	..G	..G	..G	..G	..G	I
3409 (3)	ATA (I)	...	...	...	...	...	..T	...	...	...	...	III
3400 (3)	TAC (Y)	...	...	...	...	...	...	..T	..T	..T	..T	IV
3376 (3)	AAA (K)	...	...	...	...	...	...	..G	..G	..G	..G	IV
3303 (1)	TTA (L)	...	...	...	...	C..	C..	...	...	...	...	II, III
3265 (3)	CAT (H)	...	...	...	...	...	...	..C	..C	..C	..C	IV
3250 (3)	ACC (T)	...	...	...	...	..A	..A	...	...	...	...	II, III
3232 (3)	GGC (G)	...	...	...	...	..N (X)	..G	..T	..T	..T	..T	I, IV
3229 (3)	TTG (L)	...	C..	C..	...	..A	..A	..A	..A	..A	..A	I
3202 (3)	CCT (P)	...	...	...	...	...	...	..C	..C	..C	..C	IV
3195 (1)	CTG (L)	...	...	...	...	...	..T	T..	T..	T..	T..	IV
3193 (3)	CTG (L)	...	...	...	...	...	..T	T..	T..	T..	T..	III
3157 (3)	CCC (P)	...	...	...	...	...	...	..T	..T	..T	..T	IV
3139 (3)	AAA (K)	...	...	...	...	..G	..G	...	...	...	...	II, III
3135 (1)	TTA (L)	...	...	...	...	..G	C.G	C.G	C.G	C.G	C.G	I, II/III, IV
3133 (3)	TTA (L)	...	...	...	...	..G	C.G	C.G	C.G	C.G	C.G	I
3124 (3)	TCA (S)	...	...	...	...	...	..T	...	...	...	...	III
3121 (3)	GAG (E)	...	...	...	...	...	...	..A	..A	..A	..A	IV
3081 (1)	TTG (L)	...	...	...	..A	C.T	C.T	...	...	...	...	II, III
3079 (3)	TTG (L)	...	...	...	..A	C.T	C.T	...	...	...	...	Ic, II, III
3070 (3)	CTG (L)	...	...	...	...	...	..A	...	...	...	...	III
3058 (3)	TTT (F)	...	...	...	...	...	...	..C	..C	..C	..C	IV
3039 (1)	GCA (A)	...	...	...	...	T.. (S)	T.. (S)	...	...	...	...	II, III
3036 (1)	ACT (T)	..N (X)	..A	..A	..A	..AA (K)	..AA (K)	CAA (Q)	CAA (Q)	CAA (Q)	CAA (Q)	IV
3035 (2)	ACT (T)	..N (X)	..A	..A	..A	..AA (K)	..AA (K)	CAA (Q)	CAA (Q)	CAA (Q)	CAA (Q)	I
3025 (3)	CAA (Q)	...	...	...	...	..G	..G	..G	..G	..G	..G	I
2920 (3)	ATT (I)	..N (X)	..C	..N (X)	..C	...	...	..A	..A	..A	..A	IV
2809 (3)	TCC (S)	...	...	...	...	..A	..A	...	...	...	...	II, III
2808 (1)	CAA (Q)	...	...	...	...	...	G.. (E)	...	...	...	...	III
2802 (1)	TCA (S)	...	..N. (X)	...	...	— (-) <sup>f</sup>	— (-) <sup>f</sup>	GT. (V)	CT. (L)	CT. (L)	CT. (L)	I,II, III
2791 (3)	CAT (H)	...	...	...	...	..C	..C	...	...	...	...	II, III
2788 (3)	AGT (S)	...	...	...	...	..C	..C	...	...	...	...	II, III
Small t antigen gene <sup>e</sup>												
4878 (3)	ACC (T)	...	...	...	...	..T	..T	...	...	...	...	II, III
4877 (1)	CTG (L)	...	...	...	...	T..	T..	...	...	...	...	II, III
4836 (3)	TCT (S)	...	...	...	...	..A	..A	..A	..A	..A	..A	I
4830 (3)	CAC (H)	...	...	...	...	..T	..T	...	...	...	...	II, III
4806 (3)	CTT (L)	...	...	...	...	..A	..A	..A	..A	..A	..A	I
4797 (3)	AGG (R)	...	...	...	...	...	...	..A	..A	..A	..A	IV
4764 (3)	CCC (P)	...	...	...	...	..A	..A	...	...	...	...	II, III
4734 (3)	GAC (D)	...	...	...	...	..T	..T	...	...	...	...	II, III
4726 (2)	ACA (T)	...	...	...	...	..G. (R)	...	...	...	...	...	II
4704 (3)	CTA (L)	...	...	...	...	..C	..C	...	...	...	...	II, III
4692 (3)	ACT (T)	...	...	...	...	..C	..C	...	...	...	...	II, III

<sup>a</sup> Nucleotide position using the Seif convention as applied to the BKV Dunlop strain (accession no. V01108). The number in parentheses indicates the position of this nucleotide in the corresponding codon.

<sup>b</sup> Nucleotides corresponding to the BKV Dunlop strain at the specified position. The coded amino acid is indicated in parentheses.

<sup>c</sup> Alphanumeric designations indicate BKV subtypes. Dots indicate nucleotide homology with Dunlop strain sequence. Any change in coded amino acid is indicated in parentheses.

<sup>d</sup> Genotype assignment corresponding to the SNP in question.

<sup>e</sup> Nucleotides (Dunlop numbering) 4911 to 4635, which are unique to the small t antigen gene.

<sup>f</sup> The solid dash indicates a nucleotide deletion.

TABLE 5. Subgroup-specific BKV SNPs

Position <sup>a</sup>	BKV Dunlop strain sequence <sup>b</sup>	Nucleotides corresponding to BKV Dunlop strain at specified position <sup>c</sup>									Subgroup <sup>d</sup>
		Ia	Ib1	Ib2	Ic	II	III	IVa	IVb	IVc	
Agnogene 427 (1)	GTT (V)	...	...	C.. (L)	...	A.. (I)	...	..N (X)	...	...	Ib2
VP2/VP3											
1022 (3)	ATT (I)	...	...	..A	...	...	...	...	...	...	Ib2
1023 (1)	CTG (L)	...	T..	T..	T..	T..	T..	T..	T..	T..	Ia
1091 (3)	TCT (S)	...	..C	...	...	...	...	...	...	...	Ib1
1146 (1)	TCT (S)	...	G.. (A)	G.. (A)	G.. (A)	G.. (A)	G.. (A)	G.. (A)	G.. (A)	G.. (A)	Ia
1166 (3)	TTG (L)	...	...	...	..A	...	...	...	...	...	Ic
1169 (3)	CAG (Q)	...	..A	..A	..A	..A	..A	..A	..A	..A	Ia
1272 (1)	CAA (Q)	...	G.. (E)	G.. (E)	G.. (E)	G.T (D)	G.T (D)	G.T (D)	G.T (D)	G.T (D)	Ia
1322 (3)	GTA (V)	...	..G	...	...	...	...	..T	..T	..T	Ib1
1343 (3)	CGT (R)	...	...	...	...	..A (H)	..A (H)	..AG (Q)	..AA (Q)	..AG (Q)	IVb
VP1											
1575 (3)	ACC (T)	...	...	..A	...	...	...	...	...	...	Ib2
1908 (3)	ACT (T)	...	...	..A	...	...	...	...	...	...	Ib2
1977 (3)	GAG (E)	...	...	...	...	...	...	..N (X)	..A	..A	IVc
1989 (3)	GGA (G)	...	..T	..T	..G	..C	..T	..C	..C	..C	Ia, Ic
1992 (3)	AAA (K)	...	...	...	..G	...	...	...	...	...	Ic
2076 (3)	TCA (S)	A.. (T)	A..C (T)	A.. (T)	A.. (T)	A.. (T)	A.. (T)	A.. (T)	A.. (T)	A.. (T)	Ib1
2127 (3)	CAG (Q)	...	...	..A	...	...	...	...	...	...	Ib2
LTA gene											
5103 (3)	TTA (L)	...	...	..G	...	...	...	...	...	...	Ib2
5076 (3)	AAT (N)	...	..C	..C	..C	..C	..C	..C	..C	..C	Ia
5055 (3)	GCT (A)	...	..C	..C	..C	..C	..C	..C	..C	..C	Ia
4947 (3)	GCT (A)	...	...	...	..C	...	...	...	...	...	Ic
4525 (3)	AGT (S)	...	..C	..N (X)	...	..C	..C	..C	..N (X)	..G (R)	IVc
4435 (3)	TCA (S)	...	..C	...	...	...	...	..N (X)	...	...	Ib1
4417 (3)	AAA (K)	...	..G	...	...	...	...	...	...	...	Ib1
4075 (3)	ACT (T)	...	..A	..A	..A	..TA (I)	..TA (I)	..TA (I)	..TA (I)	..TA (I)	Ia
3919 (3)	TTT (F)	...	...	...	...	...	..C	...	...	...	IVa
3844 (3)	CAC (H)	...	..T	..T	..T	..T	..T	..T	..T	..T	Ia
3749 (2)	ACC (T)	...	...	..G (S)	...	...	...	...	...	...	Ib2
3709 (3)	TTC (F)	...	..T	..T	..T	..T	..T	..T	..T	..T	Ia
3673 (3)	GGA (G)	...	...	..G	...	...	...	...	...	...	Ib2
3654 (1)	CTA (L)	...	..G	T..	...	...	...	...	...	...	Ib2
3652 (3)	CTA (L)	...	..G	T..	...	...	...	...	...	...	Ib1
3562 (3)	TTC (F)	...	..T	..T	..T	..T	..T	..T	..T	..T	Ia
3535 (3)	TTA (L)	...	...	...	...	...	...	..G	...	...	IVb
3523 (3)	CCC (P)	...	...	...	..T	..A	..A	..A	..A	..A	Ic
3454 (3)	GTA (V)	...	...	...	...	..N (X)	...	..T	..N (X)	..G	IVa
3424 (3)	GAG (E)	...	..A	..A	..A	..A	..A	..A	..A	..A	Ia
3315 (1)	TTG (L)	...	...	...	C.C	...	...	..N (X)	..A	..A	Ic
3238 (3)	CCA (P)	...	...	...	..T	...	...	...	...	...	Ic
3190 (3)	CAA (Q)	...	..G	..G	..G	..G	..G	..G	..G	..G	Ia
3172 (3)	CAA (Q)	...	..G	..G	..G	..G	..G	..G	..G	..G	Ia
3100 (3)	ATT (I)	...	...	...	..C	..A	..A	..A	..A	..A	Ic
3079 (3)	TTG (L)	...	...	...	..A	C.T	C.T	...	...	...	Ic
2908 (3)	GAG (E)	...	..A	..A	..A	..A	..A	..A	..A	..A	Ia
2761 (3)	TTT (F)	...	...	...	..C	...	...	...	...	...	Ic

<sup>a</sup> Nucleotide position using the Seif convention as applied to the BKV Dunlop strain (accession no. V01108). The number in parentheses indicates the position of this nucleotide in the corresponding codon.

<sup>b</sup> Nucleotides corresponding to those in the BKV Dunlop strain at the specified position. The coded amino acid is indicated in parentheses.

<sup>c</sup> Alphanumeric designations indicate BKV subtypes. Dots indicate nucleotide homology with the Dunlop strain sequence. Any change in coded amino acid is indicated in parentheses.

<sup>d</sup> Genotype assignment corresponding to the SNP in question.

changes resulted in no predicted change in coded amino acids. Takasaka et al. used 287-bp VP1 sequences from 45 kidney and 31 bone marrow transplant recipients to make phylogenetic trees, which separated subtype I into subgroups Ia, Ib, and Ic, with bootstrap values of 63 to 86% (36). In all of the aforementioned studies, the BKV genotyping was based on VP1 sequences.

A genotyping scheme based on LTA sequences should be considered for the following reasons. (i) It is a larger protein with more informative sites, even though overall nucleotide variability is slightly lower than that in the VP1 region (33). (ii) DNA sequencing using clinical samples on archived suboptimally preserved samples can be compromised by DNA degra-

dation. Under this circumstance, it is advantageous to be able to determine viral subtype using primers directed against more than one genomic area. (iii) Correlations between subtype and clinical syndromes become difficult when sequence data for the VP1 area are not available. This is particularly true of DNA sequence data generated by clinical labs that use PCR assays targeted to LTA. Likewise, it is difficult to ascertain the role of viral subtypes in literature focusing on the potential role of BKV in the pathogenesis of human neoplasms such as carcinoma of the prostate and posttransplant lymphoproliferative disease (5, 6, 29).

The utility of LTA for genotyping was illustrated for polyomavirus SV40 by Forsman et al., who showed that phylogenies based on SV40 T-antigen sequences are congruent with phylogenies based on whole-genome sequences (10). The same observation was subsequently made for BKV (33). Nonetheless, attempts to use nucleotides within the LTA for polyomavirus BKV genotyping have been limited to case reports of virus strains associated with AIDS-associated nephropathy and meningoencephalitis complicating chronic lymphocytic leukemia (34, 35). Our phylogenetic analysis of all BKV whole-genome sequences described to date confirms the feasibility of LTA-based genotyping. If the whole LTA gene or the 2nd exon sequences are used, subtypes I and IV can be separated by excellent bootstrap values. Indeed, even a partial 325-bp LTA sequence (nucleotides 3148 to 3472, Dunlop numbering) allows separation of I, III, and IV, with bootstrap values ranging from 90 to 100%. However, there is a fall in bootstrap values associated with subgroups Ib2 (56%) and IVa (39%). Subtype II and subgroups IVb and IVc cannot be clearly resolved by this partial sequence. In general, the complete repertoire of BKV subtypes and subgroups known currently cannot be represented by sequences derived from short amplicons that are typical of PCR assays used in clinical diagnostics.

Appropriately chosen SNP assays can obviate the need to sequence the entire LTA and provide an alternative way for BKV genotyping. Our analysis demonstrates that LTA harbors SNPs capable of distinguishing all major subtypes and subgroups. The SNPs identified by us include many previously reported SNPs but not all of them. The differences reflect methodological issues with respect to consensus calling. We have analyzed only the coding area, whereas other studies have also listed SNPs in the intergenic area between agnoprotein and VP2 or between VP1 and LTA (33). A consensus sequence can be affected not only by the consensus generation algorithm, but also by the sequences analyzed. For example, Jin et al. considered an SNP at position 1803 to be present based on analysis of 33 partial VP1 sequences (15). However, this position was invariant in the data set of 162 whole-genome sequences analyzed by us. Likewise, an SNP is considered to be present at TW-3 position 1840 (Dunlop no. 1959) based on alternate nucleotides C and G according to Nishimoto's analysis (25). However, using rule 4 of our consensus calling algorithm, the subtype IV stringent consensus call at this position was G, since only two out of 51 BKV subtype IV strains (JPN-34 and KOM-2) had variant nucleotides at this position.

It is important to consider the possibility of genetic recombination while evaluating apparent phylogenetic relationships between viral strains. Recombination is an important mechanism of immune escape and development of drug resistance in

rapidly evolving viruses. In general, circular viruses are less prone to recombination than linear ones like human immunodeficiency virus. Nevertheless, the BKV NCCR is a well-known site for extensive genetic recombination (24, 35). Within the BKV coding area, we found only limited and inconclusive evidence of recombination. Recombination events that disrupt open reading frames of critical viral proteins can result in a nonviable virus and are not expected to be present in clinical material. Recombinants can also get quickly purged by natural selection if they result in lower viral fitness or replicative capacity. The absence of major recombination events in the BKV genome does not exclude the possibility that the gene polymorphism hot spots seen in both VP1 and LTA sites are the result of more local recombination that did not leave behind any detectable traces. Local recombination events can be indistinguishable from multiple mutations. Alternately, the aforementioned hot spots could be the result of small areas of nonreciprocal interstrain or intrastrain genetic exchange during meiosis referred to as "conversion." Recombination can also be an in vitro artifact, particularly during amplification of samples with mixed infections due to multiple viral strains.

#### ACKNOWLEDGMENTS

M.B. acknowledges the support of Pittsburgh Molecular Libraries Screening Center and the Thomas E. Starzl Transplantation Postdoctoral Fellowship in Transplantation Biology. This work was supported by NIH grants RO1 AI51227 and AI63360 to P.R.

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Institute of Allergy and Infectious Diseases.

#### REFERENCES

- Baksh, F. K., S. D. Finkelstein, P. A. Swalsky, G. L. Stoner, C. F. Ryschewitsch, and P. Randhawa. 2001. Molecular genotyping of BK and JC viruses in human polyomavirus-associated interstitial nephritis after renal transplantation. *Am. J. Kidney Dis.* **38**:354–365.
- Bauer, P. H., R. T. Bronson, S. C. Fung, R. Freund, T. Stehle, S. C. Harrison, and T. L. Benjamin. 1995. Genetic and structural analysis of a virulence determinant in polyomavirus VP1. *J. Virol.* **69**:7925–7931.
- Brooks, B., R. Broucoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**:187–217.
- Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**:3497–3500.
- Das, D., R. B. Shah, and M. J. Imperiale. 2004. Detection and expression of human BK virus sequences in neoplastic prostate tissues. *Oncogene* **23**:7031–7046.
- Das, D., K. Wojno, and M. J. Imperiale. 2008. BK virus as a cofactor in the etiology of prostate cancer in its early stages. *J. Virol.* **82**:2705–2714.
- 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, NY.
- Eckner, R., J. W. Ludlow, N. L. Lill, E. Oldread, Z. Arany, N. Modjtahedi, J. A. DeCaprio, D. M. Livingston, and J. A. Morgan. 1996. Association of p300 and CBP with simian virus 40 large T antigen. *Mol. Cell. Biol.* **16**:3454–3464.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- Forsman, Z. H., J. A. Lednický, G. E. Fox, R. C. Willson, Z. S. White, S. J. Halvorson, C. Wong, A. M. Lewis, Jr., and J. S. Butel. 2004. Phylogenetic analysis of polyomavirus simian virus 40 from monkeys and humans reveals genetic variation. *J. Virol.* **78**:9306–9316.
- Gai, D. H., R. Zhao, D. W. Li, C. V. Finkelstein, and X. S. Chen. 2004. Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**:47–60.
- Gomez-Lorenzo, M. G., M. Valle, J. Frank, C. Gruss, C. O. S. Sorzano, X. S. Chen, L. E. Donate, and J. M. Carazo. 2003. Large T antigen on the simian virus 40 origin of replication: a 3D snapshot prior to DNA replication. *EMBO J.* **22**:6205–6213.

13. Hoffman, N. G., L. Cook, E. E. Atienza, A. P. Limaye, and K. R. Jerome. 2008. Marked variability of BKV virus load measurement using quantitative real-time PCR among commonly used assays. *J. Clin. Microbiol.* **46**:2671–2680.
14. Ikegaya, H., P. J. Saukko, R. Tertti, K. P. Metsarinne, M. J. Carr, B. Crowley, K. Sakurada, H. Y. Zheng, T. Kitamura, and Y. Yogo. 2006. Identification of a genomic subgroup of BK polyomavirus spread in European populations. *J. Gen. Virol.* **87**:3201–3208.
15. Jin, L. 1993. Rapid genomic typing of BK virus directly from clinical specimens. *Mol. Cell. Probes* **7**:331–334.
16. Jin, L., and P. E. Gibson. 1996. Genomic function and variation of human polyomavirus BK (BKV). *Rev. Med. Virol.* **6**:201–214.
17. 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.
18. 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.
19. Krumbholz, A., R. Zell, R. Egerer, A. Sauerbrei, A. Helming, B. Gruhn, and P. Wutzler. 2006. Prevalence of BK virus subtype I in Germany. *J. Med. Virol.* **78**:1588–1598.
20. Kumar, S., M. Nei, J. Dudley, and K. Tamura. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **9**:299–306.
21. Laskowski, R. A., M. W. MacArthur, and D. S. Moss. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**:283–291.
22. Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **73**:152–160.
23. Marti-Renom, M. A., A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali. 2000. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**:291–325.
24. Monini, P., A. Rotola, D. Di Luca, L. De Lellis, E. Chiari, A. Corallini, and E. Cassai. 1995. DNA rearrangements impairing BK virus productive infection in urinary tract tumors. *Virology* **214**:273–279.
25. Nishimoto, Y., H. Y. Zheng, S. Zhong, H. Ikegaya, Q. Chen, C. Sugimoto, T. Kitamura, and Y. Yogo. 2007. An Asian origin for subtype IV BK virus based on phylogenetic analysis. *J. Mol. Evol.* **65**:103–111.
26. Nukuzuma, S., T. Takasaka, H.-Y. Zheng, S. Zhong, Q. Chen, T. Kitamura, and Y. Yogo. 2006. Subtype I BK polyomavirus strains grow more efficiently in human renal epithelial cells than subtype IV strains. *J. Gen. Virol.* **87**:1893–1901.
27. Randhawa, P. S., A. Vats, D. Zygmunt, P. A. Swalsky, V. Scantlebury, R. Shapiro, and S. Finkelstein. 2002. Quantitation of viral DNA in renal allograft tissue from patients with BK virus nephropathy. *Transplantation* **74**:485–488.
28. Roy, R., P. Trowbridge, Z. Yang, J. J. Champoux, and D. T. Simmons. 2003. The cap region of topoisomerase I binds to sites near both ends of simian virus 40 T antigen. *J. Virol.* **77**:9809–9816.
29. Rubio, L., F. J. Vera-Sempere, M. J. Moreno-Baylach, A. Garcia, I. Zamora, and J. Simon. 2005. LT, VP1 and TCR-BKV sequence analysis in a patient with post-transplant BKV nephropathy associated with EBV-related PTLD. *Pediatr. Nephrol.* **20**:1506–1509.
30. Saitou, N., and M. Nei. 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
31. Seif, I., G. Khoury, and R. Dhar. 1979. The genome of human papovavirus BKV. *Cell* **18**:963–977.
32. Shah, K. V. 1995. Polyomaviruses, p. 2027–2043. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), *Fields virology*. Lippincott-Raven, Philadelphia, PA.
33. Sharma, P. M., G. Gupta, A. Vats, R. Shapiro, and P. Randhawa. 2006. Phylogenetic analysis of polyomavirus BK sequences. *J. Virol.* **80**:8869–8879.
34. Smith, R. D., J. H. Galla, K. Skahan, P. Anderson, C. C. Linnemann, Jr., G. S. Ault, C. F. Ryschkewitsch, and G. L. Stoner. 1998. Tubulointerstitial nephritis due to a mutant polyomavirus BK virus strain, BKV(Cin), causing end-stage renal disease. *J. Clin. Microbiol.* **36**:1660–1665.
35. 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.
36. 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.
37. 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.
38. Valls, E., X. de la Cruz, and M. A. Martinez-Balbas. 2003. The SV40 T antigen modulates CBP histone acetyltransferase activity. *Nucleic Acids Res.* **31**:3114–3122.