Complete Genome Sequences of BK Polyomavirus Strains from Two Patients with Urinary Tract Infection, Sequenced Using the Ion Torrent Platform

Graham Rose,[a] Kyriaki Ranellou,[b] Raju Misra,[a] Colin Crump,[b]
David Wooldridge,[a] Surendra Parmar,[c] Christopher Maddren,[d] Saheer Gharbia,[a]
Hamid Jalal[c]

Genomics Research Unit, Public Health England, London, United Kingdom[a]; University of Cambridge, Department of Virology, Cambridge, United Kingdom[b]; PHE, Public Health Laboratory, Clinical Microbiology and Public Health Laboratory, Addenbrooke's Hospital, Cambridge, United Kingdom[c]; Life Sciences Solutions, Thermo Fisher Scientific, Paisley, United Kingdom[d]

ABSTRACT BK polyomavirus is an important pathogen in kidney transplant patients. We report here two complete genome sequences, those of isolates CAMB-1035 and CAMB-1055, identified in two urine samples tested for urinary tract infection at a hospital in eastern England, United Kingdom. Variation and phylogenetic analyses indicate that both isolates belong to subtype Ib-1.

Address correspondence to Graham Rose, graham.rose@phe.gov.uk, or Kyriaki Ranellou, kyriaki.ranellou@addenbrookes.nhs.uk.

BK polyomavirus (BKPyV) is a nonenveloped double-stranded DNA (dsDNA) virus and ubiquitous human pathogen, with up to 90% of adults worldwide being seropositive (1, 2). Primarily, infection occurs during childhood, after which a latent asymptomatic infection is established in the urogenital tract (3). However, in immuno-compromised patients, such as bone marrow and renal transplant recipients, active infection is associated with a wide range of clinical manifestations, including serious damage to the bladder and kidneys (1, 3). Here, we report the complete genome sequences of BKPyV isolates from two patients, collected from residual urine samples destined for disposal after being sent for routine diagnosis at Public Health England (PHE) Cambridge for suspected urinary tract infection.

After recording available demographic and clinical data, samples were anonymized, and two of them, CAMB-1035 and CAMB-1055, were selected for genome sequencing after testing positive for BKPyV by an in-house quantitative PCR (qPCR) assay. DNA extraction and sequencing were performed as follows: nucleic acid was extracted using the PureLink viral RNA/DNA kit (Invitrogen) from 100 μl of urine, followed by OneStep PCR inhibitor removal column (Zymo Research). The qPCR indicated low viral concentrations (threshold cycle [C_T], 26.7 and 27.5); therefore, a PCR amplification step with a set of six custom BKPyV-specific primers covering the complete genome was used. Libraries were prepared using the Ion Xpress Plus fragment library kit (Life Technologies), per the manufacturer's instructions, and then size selected by 2% E-Gel SizeSelect (Invitrogen). Fragment size and concentration were measured by an Agilent Bioanalyzer and a Qubit fluorometer, respectively. Libraries were amplified on the One Touch 2 with the Ion PGM template OT2 400 kit (Life Technologies) and sequenced on an Ion Torrent PGM sequencer using the 318 Chip kit version 2 and Ion PGM Hi-Q sequencing kit (Life Technologies). Sequencing reads were trimmed by base quality and human host sequences filtered as described previously (4).

After assembly using metaSPAdes (version 3.7.1) (5), each BKPyV assembly consisted of a single circular contig which was split using Gepard version 1.40 (6) at the origin of replication based on the BKPyV-Dun reference (GenBank accession number

NC_001538), generating a 5,141-bp genome. Using a pipeline with BWA (7) and SAMtools (8), 98,095 reads (20.0 Mb) and 158,009 reads (30.8 Mb) mapped to the assembled genomes of CAMB-1035 and CAMB-1055, respectively, both with 100% genome coverage, and at average 3,878× and 5,995× coverages, respectively.

Nearly 300 BKPyV strains have now been completely sequenced. BLASTn analysis of the two genomes showed >97% sequence identity to BKPyV-Dun and a characteristic genome organization of six genes encoding regulatory and structural proteins separated by a noncoding region. The highest sequence identity (>99.8%) was to isolates SJH-LG-152 (GenBank accession number JN192431) and SJH-LG-306 (GenBank accession number JN192437). Globally, there are four BKPyV subtypes (I to IV), with subtype I divided into four subgroups (9). Phylogenetic reconstruction by maximum likelihood using 33 genomes representative of the main lineages (10) clustered both genomes within subgroup Ib-1, which is within the most prevalent subtype (11). After filtering and annotating variants compared to BKPyV-Dun (12, 13), 15/79 CAMB-1035 single-nucleotide polymorphisms (SNPs) and 18/85 CAMB-1055 SNPs were nonsynonymous, of which most fell within the structural region encoding the capsid proteins and agnoprotein (26 nonsynonymous SNPs [78.8%]).

Accession number(s). All sequences are publicly available and can be found within the EBI European Nucleotide Archive under the chromosome accession numbers LT934539 (CAMB-1035) and LT960370 (CAMB-1055) and ENA study accession number PRJEB22539. The version described in this paper is the first version.

REFERENCES

1. Hirsch HH, Steiger J. 2003. Polyomavirus BK. Lancet Infect Dis 3:611–623. https://doi.org/10.1016/S1473-3099(03)00770-9.
2. Egli A, Infanti L, Dumoulin A, Buser A, Samaridis J, Stebler C, Gosert R, Hirsch HH. 2009. Prevalence of polyomavirus BK and JC infection and replication in 400 healthy blood donors. J Infect Dis 199:837–846. https://doi.org/10.1086/597126.
3. Moens U, Krumbholz A, Ehlers B, Zell R, Johne R, Calvignac-Spencer S, Lauber C. 2017. Biology, evolution, and medical importance of polyomaviruses: an update. Infect Genet Evol 54:18–38. https://doi.org/10.1016/j.meegid.2017.06.011.
4. Rose G, Shaw AG, Sim K, Wooldridge DJ, Li MS, Gharbia S, Misra R, Kroll JS. 2017. Antibiotic resistance potential of the healthy preterm infant gut microbiome. PeerJ 5:e2928. https://doi.org/10.7717/peerj.2928.
5. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. MetaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.
6. Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics 23:1026–1028. https://doi.org/10.1093/bioinformatics/btm039.
7. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv arXiv:1303.3997 [q-bio.GN]. https://arxiv.org/abs/1303.3997.
8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
9. Krumbholz A, Bininda-Emonds ORP, Wutzler P, Zell R. 2008. Evolution of four BK virus subtypes. Infect Genet Evol 8:632–643. https://doi.org/10.1016/j.meegid.2008.05.006.
10. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.
11. Drew RJ, Walsh A, Laoi BN, Crowley B. 2012. Phylogenetic analysis of the complete genome of 11 BKV isolates obtained from allogenic stem cell transplant recipients in Ireland. J Med Virol 84:1037–1048. https://doi.org/10.1002/jmv.23240.
12. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330.
13. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) 6:80–92. https://doi.org/10.4161/fly.19695.