

## Genetic diversity of human polyomavirus JCPyV in Southern California wastewater

Amana Rafique and Sunny C. Jiang

### ABSTRACT

JC Polyomavirus (JCPyV) has the potential to be used as a viral marker for human waste contamination because at least 40% of the human population excretes this virus through its urine. In addition, each of 6 known subtypes of JCPyV is associated with a specific human ethnicity group, which has allowed for tracing of human migration. This study aims to explore the possibility of linking the genetic diversity of JCPyV with source of human waste. Primary sewage effluent from Irvine Ranch Water District (IRWD) sewage treatment facility was collected and examined for the presence of JCPyV using polymerase chain reaction (PCR). JCPyV was detected by nested PCR using primers specific to the conservative regulatory region in 100% of the 13 samples collected monthly over a year. Only 5 samples were amplifiable by the primers specific to the diverse intergenic region of the virus. Sequence analyses of cloned amplicons from the intergenic region indicated that JCPyV matched subtypes of European, Asian, African and African-American origins. A genotype that is unique from previously identified clinical sample is also revealed. This research suggests the diversity of JCPyV subtypes can be used as a tool to trace the source of human waste contamination.

**Key words** | genetic diversity, JC polyomavirus, PCR, source identification

**Amana Rafique**  
**Sunny C. Jiang** (corresponding author)  
Department of Environmental Health  
Science & Policy,  
University of California, Irvine,  
Irvine 92697,  
California  
Tel.: 949-824-5527  
Fax: 949-824-2056  
E-mail: [sjiang@uci.edu](mailto:sjiang@uci.edu)

### ABBREVIATIONS

JCPyV	JC polyomavirus
IGR	Intergenic Region
IRWD	Irvine Ranch Water District
ORF	Open reading frame
PCR	Polymerase chain reaction
RR	Regulatory region

### INTRODUCTION

The polyomavirus, JC (JCPyV), is an opportunistic pathogen and responsible for progressive multifocal leukoencephalopathy (PML) in immunocompromised individuals (Hammarin *et al.* 1996), usually those with Acquired Immune Deficiency Syndrome (AIDS). Infected individuals

may have brain tissue damage which can lead to brain tumors (Enam *et al.* 2002). JCPyV contains a double-stranded circular DNA genome which is approximately 5.13 Kb in size and replicates bidirectionally. There is an early region (small and large T antigen), a late region (VP1, VP2, VP3, and agnoprotein), and a regulatory region (RR) containing the promoters, enhancers, and the replication origin (Frisque *et al.* 1984). The V-T intergenic region which consists of the region between VP1 and the large T antigen are variable between genotypes.

This virus is known to spread through close human contact, usually from parent to child, although recent studies imply that environment can also play a role for viral transmission (Bofill-Mas & Girones 2003). Studies conducted in Japan have shown that children and their parents within the same household share an identical strain of JCPyV (Kitamura *et al.* 1994). Infection generally occurs

doi: 10.2166/wh.2008.067

during childhood; the virus can remain persistent in renal cells and can continue to be excreted in urine for a life-long period (Padgett & Walker 1973). The infection remains asymptomatic throughout an infected individual's life, unless the immune system becomes compromised.

JCPyV has a wide distribution within the human population. Studies estimate that the prevalence of JCPyV within the human population is 75%, by examining adults who have antibodies for JCPyV (Padgett & Walker 1973). There are currently seven major genotypes of JCPyV, which correspond to three main areas of the world – Europe, Asia, and Africa: genotypes 1 and 4 are found in Europe, types 2 and 7 are in Asia, types 3 and 6 are in Africa, and type 5 is a recombinant type between type 2b and 6. A possible eighth genotype is being proposed for isolates found in Papua New Guinea (Jobe *et al.* 2001). The viral genome is inherently stable within the human body. Studies have shown using 18 full sequences of JCPyV that the virus has co-evolved with human migration. The analysis of the phylogenetic relationship of JCPyV (isolated from different groups of humans) confirms the anthropological evidence that all humans originated from Africa (Pavesi 2003).

Detection of JCPyV in human sewage has been shown in 98% of the samples tested in Europe, Africa, and Washington D.C. U.S.A. (Bofill-Mas *et al.* 2001). It is suggested that the detection of JCPyV in the environment could provide an indicator of human waste contamination over a wide geographic region (Bofill-Mas & Girones 2003; Albinana-Gimenez *et al.* 2006; McQuaig *et al.* 2006). Detection of JCPyV and BKPyV using general primers targeting the common region of both viruses has been shown in South Florida with 19 out of 20 environmental samples testing positive (McQuaig *et al.* 2006). More recently, Albinana-Gimenez *et al.* (2006) has applied a quantitative PCR method developed for clinical assay (Pal *et al.* 2006) for quantification of this virus in wastewater, river water and drinking water. All primers in these previous studies used general primers targeting at conservative region of the virus. In this study, we hope to explore the link between subtype of JCPyV and sources of pollution by characterizing the genetic diversity of this virus in southern California urban sewage. Here we report the diversity of JCPyV using primers specific to diverse intergenic region and identification of a previously unknown genotype.

## METHODS

### Sample sources

Primary sewage effluent was obtained on a monthly basis from the Irvine Ranch Water District (IRWD) in Orange County, California, during the period from May 2005 to May 2006. IRWD serves the city of Irvine and also parts of Costa Mesa, Lake Forest, Newport Beach, Tustin, and Orange in Southern California, with a metropolitan area of 85,019 acres and a population of 322,000 ([www.irwd.com](http://www.irwd.com)). A total of 13 monthly samples were used for this study.

### Viral concentration and DNA purification

Sixty ml of primary sewage was divided into six 10 ml aliquots and concentrated by ultracentrifugation for 1.5 hours at 41,000 rpm (L8-70M Ultracentrifuge, Beckman Coulter) using a SW 41 Ti swinging-bucket rotor. The pellets were re-suspended and combined from all six tubes to a final volume of approximately 500  $\mu$ L. An equal volume of chloroform was then added and vortexed. The aqueous phase was collected after centrifugation for viral nucleic acid purification and DNA extraction using the QIAamp Viral RNA Mini Kit, following the manufacturer's protocol (Qiagen Inc.)

### Enzymatic amplification

Two separate regions in JCPyV were amplified using PCR, the highly conserved Regulatory Region (RR) to verify the presence of the virus, and the variable V-T Intergenic Region (IGR) to identify the diversity of the virus among the sewage samples.

The RR's primers that were used are 5'-CCCTATT-CAGCACTTTGTCC-3' (JR1) located from 4992–5011 and 5'-CAAACCACTGTGTCTCTGTC-3' (JR2) located from 428–447. The nested PCR primers for the RR region were 5'-GGGAATTTCCCTGGCCTCCT-3' (JR3) located from 5060–5079 and 5'-ACTTTCACAGAAGCCTTACG-3' (JR4) located from 298–317 (Bofill-Mas *et al.* 2001). Amplifications were carried out in a 50- $\mu$ l reaction mixture containing the following: 1.3mM MgCl<sub>2</sub>, 0.2mM each

dNTPs, 12.5  $\mu$ M of each primer, 1.25 units of GoTaq<sup>®</sup> DNAPolymerase (Promega), and 2  $\mu$ l of viral DNA extraction. The PCR amplifications were performed at 94°C for 4 min for denaturation, followed by 29 cycles of denaturing at 92°C for 1 min, annealing 61°C for 1 min, and extension 72°C for 1.15 min, a final extension of 72°C for 4 min, and a hold step at 4°C.

The IGR's primers that were used are 5'-TTTTGGGA-CACTAACAGGAGG-3' (P1) located from 2107 to 2127 and 5'-AGCAGAAGACTCTGGACATGG-3' (P2) located from 2762 to 2742 (Kunitake *et al.* 1995). Amplifications were carried out in a 50- $\mu$ l reaction mixture using the same chemistry as for RR region PCR. The PCR amplifications were performed at 94°C for 4 min for denaturation, followed by 50 cycles of denaturing at 94°C for 1.5 min, annealing 55°C for 1.5 min, and extension 72°C for 2.5 min, a final extension of 72°C for 4 min, and a hold step at 4°C. A second set of nested primers for IGR designed by the Bofill-Mas group in Spain (Bofill-Mas *et al.* 2000) were also tested. However, these primer sets yield non-specific amplification for sewage samples used in this study. Two different amplicons were also observed using positive control JCPyV mad-1 strain (data not shown). Trouble shooting of amplification conditions and communication with the group in Spain who originally designed and tested the primer set did not resolve the non-specific amplification issue. Thus, these nested primer sets were not used for identification of the diversity of IGR sequences from sewage samples. All PCR products were analyzed on a 1% ethidium bromide stained agarose gel.

### Cloning and sequencing

PCR amplicons from the IG region were purified using the QIAamp Quick PCR Purification Kit and cloned into the pGEM-T easy vector (Promega), following the manufacturer's recommended protocol. White colonies were randomly selected; recombinant plasmids were extracted using QIAprep Spin Miniprep Kit (Qiagen Inc.) following the manufacturer's protocol. M13 primer was used for sequencing using the ABI PRISM Big Dye version 3.0 sequencing chemistry (Applied Biosystems, Foster City, CA, USA) by either slab gel (ABI Prism 377) or capillary (ABI Prism 3100) electrophoresis.

### Phylogenetic analysis

Fourteen retrieved sequences were analysed using the Lasergene software program (DNASTar, Inc.; Madison, WI, USA). Open reading frames (ORFs) were detected by using ATG as the start codon; TGA, TAG, and TAA as the stop codons; and 25 codons which is approximately 75 basepairs as the minimum ORF size. The sequences were searched against the NCBI GenBank database for both blastn and blastp. Alignment of sequences was performed by Clustal W using MegAlign (DNASTar, Inc.). The phylogenetic tree was constructed by the neighbor-joined (NJ) method and was viewed through MegAlign. Reference sequences which were used for the phylogenetic tree comparison are listed in Table 1.

### Nucleotide sequence accession numbers

The sequences reported in this paper have been deposited in the GenBank database under accession numbers EF369493, EF369494, EF369495, EF369496, EF369497, EF369498, EF369499, EF369500, EF369501, EF369502, EF369503, EF369504, EF369505, EF369506 for sequences 1 to 14, respectively.

**Table 1** | Reference strains used for phylogenetic analysis

Genotype	GenBank accession number
Type 1 (Europe)	
1A (Positive Control)	J02226
1A	AF015526
1B	AF015227
Type 2 (Asia)	
2A	AF015529
2B	AF015532
2C	AF015534
2D	AF015536
Type 3 (Africa)	
3A	U73500
3B	U73501
Type 4 (Europe and African American)	AF015528
Type 5 (minor European)	AF015684
Type 6 (West Africa)	AF015537
Type 7 (Southeast Asia)	U61771

## RESULTS

### Prevalence of JCPyV in urban sewage

All 13 samples were positive for RR of JCPyV by nested-PCR amplification. The occurrence of JCPyV was 100% in all the Southern California sewage samples tested, which add to the data on the high prevalence of JCPyV in sewage samples from Europe, Africa and east coast U.S. (Bofill-Mas & Girones 2003; Albinana-Gimenez *et al.* 2006; McQuaig *et al.* 2006). However, only 5 of the 13 samples were amplifiable for IGR of JCPyV.

### Nucleotide sequence analysis

Blast search results indicated eleven of the fourteen sequences were aligned with gene coding for JCPyV VP1 capsid protein. Neighbor-join tree of these sequences is shown in Figure 1. Sequences 2, 3, 6, and 8 retrieved from sewage are related to JCPyV Type 1 strain, which belong to the European clade. Sequence 11 is related to JCPyV Type 2, which is the Asian clade. Sequence 14 is related to Type 3,

which is the African clade. Sequences 10, 12, and 13 are related to type 4, which is the European and African-American clade. Sequence 9 forms its own branch suggesting a previously unidentified type of strain. However, additional sequences should be retrieved from sewage samples before the confirmation of a new type within California.

### Amino acid sequence analysis

Each nucleotide sequence obtained was searched for ORF and confirms that the strain coding for proteins used in JCPyV. Approximately, 75% of the known ORF matched VP1 protein in JCPyV. The other 25% matched Large T antigen. No single sequence obtained coded for both the Large T antigen and VP1. The percentage similarity between the sequences varies from 59 to 100%.

## DISCUSSION

### Prevalence of JCPyV in urban sewage

The 100% detection rate for RR of JCPyV in a Southern California sewage facility is not surprising although a

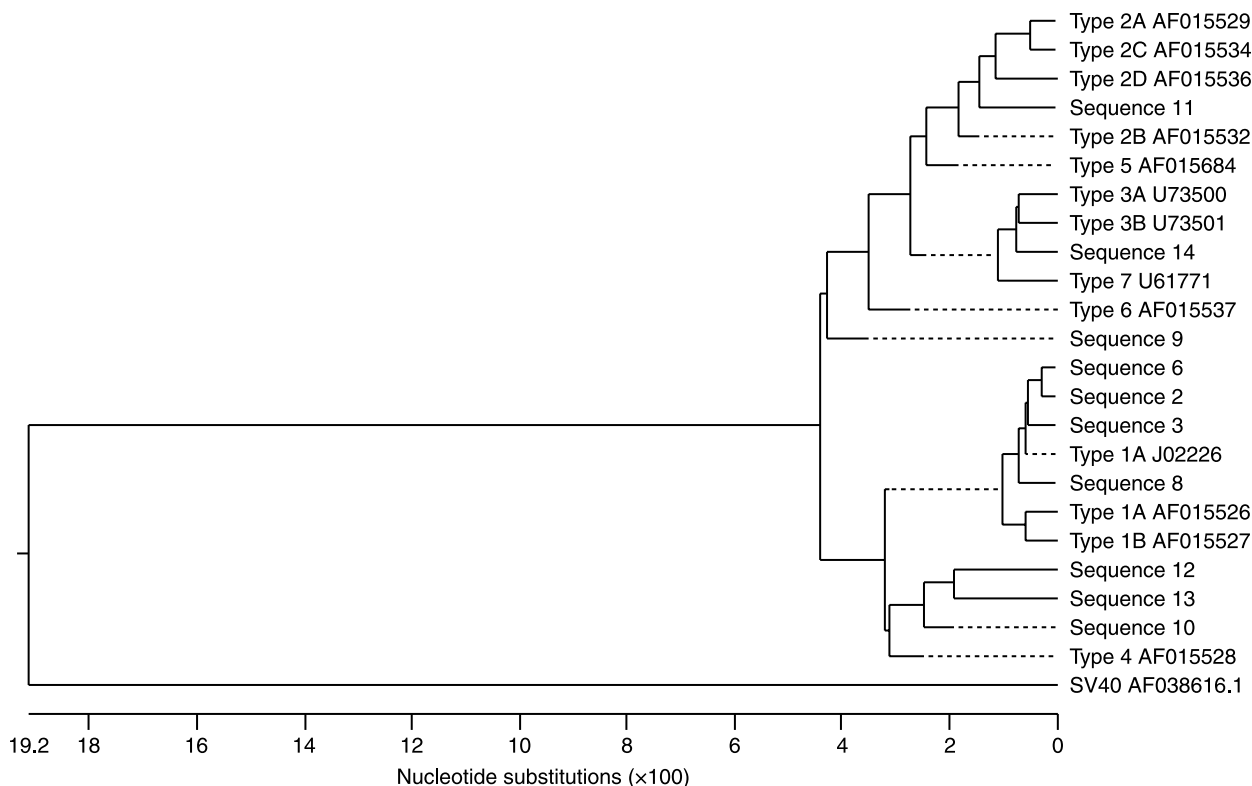


Figure 1 | Phylogenetic analysis of nucleic acid sequences retrieved from sewage samples and reference strains for the VP1 capsid protein gene.

sewage sample in this region has not been previously tested. Together with previous studies from other regions of the world (Bofill-Mas & Girones 2003; Albinana-Gimenez *et al.* 2006; McQuaig *et al.* 2006), this research confirms that JCPyV may be used as a viral marker for human waste contamination in the environment. Application of human viral marker may provide better indication of human specific waste than using fecal indicator bacteria. Other human viruses, i.e. adenoviruses, have also been suggested as a microbial source tracking tool for human waste identification (Pina *et al.* 1998; Noble *et al.* 2003) because of the host specific nature of viruses.

In spite of the prevalence of JCPyV in sewage samples as indicated by the positive detection of RR region, only a fraction of the samples were PCR positive for IGR by a single step PCR. In attempt to retrieve additional IGR sequences from Southern California sewage, nested primers designed by the Bofill-Mas group in Spain (Bofill-Mas *et al.* 2000) were tested. However, no additional IGR sequences were obtained using these new set of primers either. This result may be due to the presence of PCR inhibitors in sewage samples which give false negative results. The target IGR amplicon is longer than RR amplicon which may be more sensitive to inhibition. The PCR efficient for IGR is also lower as observed by multiple amplicons using the positive control strain Mad-1. We observed improved PCR detection after chloroform extraction in some samples. However, major portion of the samples were still not amplifiable. Alternatively, the negative result may be due to a greater diversity of IGR than we previously expected. The primers designed for clinical samples may not be suited for amplification of IGR from sewage. A previously unknown genotype of JCPyV was identified in this study. Future research may confirm the greater diversity of JCPyV in environments using degenerate primers for this region.

### Diversity of JCPyV in Southern California sewage

The diversity of the sequences retrieved from sewage is hypothesized to reflect the population diversity in the geographic area of Southern California. The majority of the sequences are clade with a European type strain. The strains of JCPyV in North and South America have not been studied extensively, as in Asia and Europe. This study offers

a glimpse of the different subtypes which may be present in North America. It is likely additional types other than classical types of JCV may be present in the Southern California region because of the highly mixed ethnicity and mixed-race marriages. The sequence 9 retrieved from sewage is distinguishable from known type strains. This is not too surprising because a potential eighth subtype is being explored in Papua New Guinea. The relevance of assessing the different subtypes in the environment will allow for additional information to be obtained on human waste pollution. This assay may be useful for assaying the source of fecal pollution from leakage of individual household septic tank because the subtype of the virus is associated with a specific ethnicity group. A database of diverse types of JCPyV in a geographical region will potentially allow for human sewage to be categorized.

### CONCLUSIONS

JCPyV is excreted by enough of the population to be detected in sewage at a very high frequency (Padgett & Walker 1973). In this study, JCPyV was detected in all primary sewage effluent samples from a southern California sewage treatment facility adding to the growing body of information from other regions of the world. The ability of detecting JCPyV in urine within human sewage provides one more tool to locate possible sources of human contamination within the environment.

Understanding the genetic variability of JCPyV may lead to finding the exact source of the pollution, and in turn allowing for better monitoring and remediation. Future research will focus on assaying the diversity of JCPyV within small waste treatment facilities, i.e. septic tanks, which may be more problematic than large sewage treatment plants. The RR may be used to indicate the presence or absence of JCPyV within an environmental sample, while the diverse type of IGR within sewage samples may pin point the source of contamination. Therefore, the ability to detect JCPyV and its stability in sewage make it a potential marker of human waste contamination within the environment.

### ACKNOWLEDGEMENTS

Funding for this research is provided by the Newkirk Center for Science and Society. Special thanks to Weiping Chu, at



the University of California, Irvine for technical assistance. Also thanks to the UCI Urban Water Research Center and the Center for Virus Research, for administrative support.

## REFERENCES

- Albinana-Gimenez, N., Clemente-Casares, P., Bofill-Mas, S., Hundesa, A., Ribas, F. & Girones, R. 2006 Distribution of human polyomaviruses, adenoviruses, and hepatitis E virus in the environment and in a drinking-water treatment plant. *Environ. Sci. Technol.* **40**(23), 7416–7422.
- Bofill-Mas, S. & Girones, R. 2005 Role of the environment in the transmission of JC virus. *J. Neurovirol.* **9** (Suppl 1), 54–58.
- Bofill-Mas, S., Pina, S. & Girones, R. 2000 Documenting the epidemiologic patterns of polyomaviruses in human populations by studying their presence in urban sewage. *Appl. Environ. Microbiol.* **66**(1), 238–245.
- Bofill-Mas, S., Formiga-Cruz, M., Clemente-Casares, P., Calafell, F. & Girones, R. 2001 Potential transmission of human polyomaviruses through the gastrointestinal tract after exposure to virions or viral DNA. *J. Virol.* **75**(21), 10290–10299.
- Enam, S., Del Valle, L., Lara, C., Gan, D. D., Ortiz-Hidalgo, C., Palazzo, J. P. & Khalili, K. 2002 Association of human polyomavirus JCV with colon cancer: evidence for interaction of viral T-antigen and beta-catenin. *Cancer Res.* **62**(23), 7093–7101.
- Frisque, R. J., Bream, G. L. & Cannella, M. T. 1984 Human polyomavirus JC virus genome. *J. Virol.* **51**(2), 458–469.
- Hammarin, A. L., Bogdanovic, G., Svedhem, V., Pirskanen, R., Morfeldt, L. & Grandien, M. 1996 Analysis of PCR as a tool for detection of JC virus DNA in cerebrospinal fluid for diagnosis of progressive multifocal leukoencephalopathy. *J. Clin. Microbiol.* **34**(12), 2929–2932.
- Jobe, D. V., Friedlaender, J. S., Mgone, C. S., Agostini, H. T., Koki, G., Yanagihara, R., Ng, T. C. N., Chima, S. C., Ryschkewitsch, C. F. & Stoner, G. L. 2001 New JC virus (JCV) genotypes from Papua New Guinea and micronesia (type 8 and type 2E) and evolutionary analysis of 32 complete JCV genomes. *Arch. Virol.* **146**(11), 2097–2113.
- Kitamura, T., Kunitake, T., Guo, J., Tominaga, T., Kawabe, K. & Yogo, Y. 1994 Transmission of the human polyomavirus JC virus occurs both within the family and outside the family. *J. Clin. Microbiol.* **32**(10), 2359–2363.
- Kunitake, T., Kitamura, T., Guo, J., Taguchi, F., Kawabe, K. & Yogo, Y. 1995 Parent-to-child transmission is relatively common in the spread of the human polyomavirus JC virus. *J. Clin. Microbiol.* **33**(6), 1448–1451.
- McQuaig, S. M., Scott, T. M., Harwood, V. J., Farrah, S. R. & Lukasik, J. O. 2006 Detection of human-derived fecal pollution in environmental waters by use of a PCR-based human polyomavirus assay. *Appl. Environ. Microbiol.* **72**(12), 7567–7574.
- Noble, R. T., Weisberg, S. B., Leecaster, M. K., McGee, C. D., Ritter, K., Walker, K. O. & Vainik, P. M. 2003 Comparison of beach bacterial water quality indicator measurement methods. *Environ. Monit. Assess.* **81**(1–3), 301–312.
- Padgett, B. L. & Walker, D. L. 1973 Prevalence of antibodies in human sera against JC virus, an isolate from a case of progressive multifocal leukoencephalopathy. *J. Infect. Dis.* **127**(4), 467–470.
- Pal, A., Sirota, L., Maudru, T., Peden, K. & Lewis, A. M. 2006 Real-time, quantitative PCR assays for the detection of virus-specific DNA in samples with mixed populations of polyomaviruses. *J. Virological Methods* **135**(1), 32–42.
- Pavesi, A. 2003 African origin of polyomavirus JC and implications for prehistoric human migrations. *J. Mol. Evol.* **56**(5), 564–572.
- Pina, S., Puig, M., Lucena, F., Jofre, J. & Girones, R. 1998 Viral pollution in the environment and in shellfish: human adenovirus detection by PCR as an index of human viruses. *Appl. Environ. Microbiol.* **64**(9), 3376–3382.

First received 24 August 2007; accepted in revised form 14 October 2007. Available online March 2008