



6-11-2018

## Detecting Viral Genomes in the Female Urinary Microbiome

Andrea Garretto  
*Loyola University Chicago*

Krystal Thomas-White  
*Loyola University Chicago*

Alan J. Wolfe  
*Loyola University Chicago*

Catherine Putonti  
*Loyola University Chicago*, [cputonti@luc.edu](mailto:cputonti@luc.edu)

Follow this and additional works at: [https://ecommons.luc.edu/bioinformatics\\_facpub](https://ecommons.luc.edu/bioinformatics_facpub)



Part of the [Bioinformatics Commons](#), and the [Biology Commons](#)

### Recommended Citation

Garretto, Andrea; Thomas-White, Krystal; Wolfe, Alan J.; and Putonti, Catherine. Detecting Viral Genomes in the Female Urinary Microbiome. *Journal of General Virology*, 99, : 1141-1146, 2018. Retrieved from Loyola eCommons, Bioinformatics Faculty Publications, <http://dx.doi.org/10.1099/jgv.0.001097>

This Article is brought to you for free and open access by the Faculty Publications and Other Works by Department at Loyola eCommons. It has been accepted for inclusion in Bioinformatics Faculty Publications by an authorized administrator of Loyola eCommons. For more information, please contact [ecommons@luc.edu](mailto:ecommons@luc.edu).



This work is licensed under a [Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 License](#).  
© The Authors 2018

1 **Detecting Viral Genomes in the Female Urinary Microbiome**

2

3 Andrea Garretto<sup>1</sup>, Krystal Thomas-White<sup>2,\*</sup>, Alan J. Wolfe<sup>2</sup>, Catherine Putonti<sup>1,2,3,4</sup>

4 1 Bioinformatics Program, Loyola University Chicago, Chicago, IL, USA

5 2. Department of Microbiology and Immunology, Loyola University Chicago, Maywood, IL, USA

6 3 Department of Biology, Loyola University Chicago, Chicago, IL, USA

7 4 Department of Computer Science, Loyola University Chicago, Chicago, IL, USA

8

9 \*Current address: 325 Sharon Park Dr, Suite 522, Menlo Park, CA, USA

10

11 Corresponding Author Details: Catherine Putonti. [cputonti@luc.edu](mailto:cputonti@luc.edu) 1-773-508-3277

12 Keywords: urinary microbiome; urinary virome; bacteriophage; JC polyomavirus

13 Subject Category: Prokaryotic Viruses and Animal Viruses

14

15 The GenBank SRA accession numbers for the 30 metagenomes produced here are as follows:

16 ERR926109 through ERR926123 and ERR926139 through ERR926153, under the BioProject Accession #  
17 PRJEB8104.

18

19 **Abstract**

20 Viruses are the most abundant component of the human microbiota. Recent evidence has uncovered a  
21 rich diversity of viruses within the female bladder, including both bacteriophages and eukaryotic viruses.  
22 We conducted whole genome sequencing of the bladder microbiome of 30 women: 10 asymptomatic  
23 “healthy” women and 20 women with overactive bladder. These metagenomes include sequences  
24 representative of human, bacterial, and viral DNA. This analysis, however, focused specifically on viral  
25 sequences. Using the bioinformatic tool virMine, we discovered sequence fragments, as well as  
26 complete genomes, of bacteriophages and the eukaryotic virus JC polyomavirus. The method employed  
27 here is a critical proof-of-concept: the genomes of viral populations within the low biomass bladder  
28 microbiota can be reconstructed through whole genome sequencing of the entire microbial community.

29

30

31 The old paradigm that the bladder is sterile results from the use of standard urine culture-dependent  
32 methods that are optimized for *E. coli* (1,2). However, there is definitive evidence that communities of  
33 bacteria exist within the bladder (3-6), as well as for associations between these bladder microbiota and  
34 urinary symptom levels, treatment response, and UTI risk (7-15). Furthermore, the bladder microbiota  
35 of individuals both with and without urinary symptoms includes viral species. The viruses isolated from  
36 urine include several viruses that infect eukaryotes (16-22), as well as those that infect bacteria  
37 (bacteriophages [phages]) (23-25). Metagenomic sequencing of the urinary virome, which detects  
38 eukaryotic viruses and phages in the lytic cycle, revealed an abundance of phages (26,27).

39 Because the bladder microbiota exist at a substantially lower biomass (1,5,6) than many other human  
40 niches (e.g. the gut (28)), sequencing the bladder’s virome presents unique technical difficulties. From  
41 the gut, the viral biomass can be separated and the extracted DNA can be sequenced directly (29,30). In  
42 contrast, previous urine virome metagenomic studies have relied on DNA amplification prior to  
43 sequencing to increase DNA concentrations (26). These amplification methods, however, have well  
44 documented biases (31). As such, the complete diversity of the virome may not be captured.  
45 Alternatively, we hypothesized that the challenges of sequencing the bladder virome could be overcome  
46 bioinformatically. Bioinformatic approaches have successfully identified complete viral genomes from

47 bacterial metagenomes (e.g. 32). Moreover, complete viral genomes have been reconstructed from viral  
48 metagenomes containing significant quantities of non-viral (bacterial and eukaryotic) DNA (e.g. 33).  
49 Thus, we conducted whole genome sequencing of the bladder microbiota and examined the sequence  
50 data specifically for viral sequences. This approach has the potential to capture both lytic and lysogenic  
51 phage sequences present in the community.

52 In a previously published study (10), urine was collected aseptically via transurethral catheter from 10  
53 women without urinary symptoms (control) and 20 women with reported overactive bladder symptoms  
54 (OAB) and stored with the DNA preservative AssayAssure (Sierra Molecular) at -80°C. In the current  
55 study, 5 ml of each urine sample was thawed and the DNA was extracted, as described previously  
56 (10,34). Briefly, the urine was incubated in a lysis solution containing mutanolysin and lysozyme and the  
57 DNA extracted from the sample using the DNeasy blood and tissue kit (Qiagen, Valencia, CA), according  
58 to the manufacturer's instructions. The Illumina Nextera kit was used for whole genome library  
59 preparation with fragment sizes of 200-300 bp. Sequencing was conducted on the Illumina HiSeq 2500  
60 platform, producing paired-end 100bp x 2 reads. Human contaminating reads were filtered out by  
61 mapping to the Human reference genome (hg19) with bowtie2 (35). **Supplementary Table 1** lists the  
62 number of raw reads and filtered reads for each patient sample sequenced. Most reads produced  
63 represent bacterial and viral species; on average, only 5.3% of the reads mapped to the human  
64 reference genome sequence. Raw sequencing data are available from NCBI's SRA database, BioProject  
65 Accession # PRJEB8104. The accession numbers for each sample are listed in **Supplementary Table 1**.

66 **Supplemental Figure S1** outlines the analytic process. Each individual metagenome data set was  
67 assembled separately. Raw reads were first trimmed for quality using the tool sickle (36) and then  
68 assembled by SPAdes (v3.10.1) with the "meta" (metagenomic) option (37). There was only a weak  
69 correlation between the number of reads produced for a given sample and the number of contigs  
70 assembled from those reads ( $r=0.23$ ). Next, the virMine (38) tool was used to classify the contigs  
71 produced. Briefly, virMine first filtered out contigs less than 1000 bp in length; this length is a user-  
72 defined parameter and was selected to eliminate partial gene sequences and repetitive elements from  
73 downstream analyses. For the remaining contigs, open reading frames were predicted, translated, and  
74 compared to virMine's bacterial and viral protein sequence databases (RefSeq protein sequences).  
75 These comparisons enabled us to classify each contig as bacterial, viral, or unknown (exhibiting no  
76 similarity to bacterial or viral). Genome assembly and virMine statistics are listed in **Supplementary**

77 **Table 1.** The microbiomes were dominated by bacterial contigs (90% on average). The contigs classified  
78 as “unknown” were queried via megablast to the NCBI nr/nt database finding that the overwhelming  
79 majority were human in origin (results not shown). Thus, here we will focus on the 252 contigs from the  
80 30 metagenomes that were predicted to be viral.

81 Twenty-seven of the 30 bladder metagenomes examined included contigs predicted to be viral. To  
82 further evaluate these contigs, each was queried against the nr/nt database via the NCBI web interface  
83 using the megablast algorithm (**Supplementary Table 2**). In comparing the contigs to this database, eight  
84 samples were identified as containing sequences of human origin. The virMine software characterized  
85 these contigs as viral, as they did not resemble bacterial sequences and had moderate sequence  
86 similarity to a sequence in the viral database. The contigs within another seven samples were uniformly  
87 short (~1 kbp) and only exhibited sequence similarity to annotated transposases. Transposases, along  
88 with integrases, can be encoded by a phage to allow that phage to enter its lysogenic (latent) life cycle  
89 by inserting itself into the bacterial genome (the inserted phage genome is now called the “prophage”)  
90 (39). Thus, while these contigs suggest the presence of lysogenic phages within the bladder microbiota,  
91 they do not provide insight into the phage species. The remaining 12 metagenomes, however, had  
92 recognizable phage and/or eukaryotic virus sequences.

93 Two patient samples – OAB045 and OAB052 – contained numerous contigs with homologies to  
94 annotated phage genes, including genes annotated as encoding tail proteins, phage tail tape measure  
95 proteins, phage DNA packaging proteins, phage portal proteins, terminases, and capsid proteins.  
96 Furthermore, these contigs represented phage genome fragments, including several coding regions. For  
97 instance, in the OAB052 sample, a 4898 bp contig was identified, containing annotated regions for a  
98 phage terminase, phage portal protein, endopeptidase Clp, major capsid protein, phage DNA packaging  
99 protein, and two hypothetical proteins. This contig is homologous to a region within the 18.3 kbp  
100 putative prophage (determined via PHAST (40)) in the *Gardnerella vaginalis* HMP9231 genome. As such,  
101 it is unlikely that the contig identified here represents a complete, intact phage genome. Nevertheless, it  
102 may represent a *Gardnerella* prophage, which we previously showed to be prevalent within *Gardnerella*  
103 strains of the bladder (41). We next examined the contigs that were classified as bacterial by the virMine  
104 tool. Blast queries found significant homology (e-score=0) between the larger contigs within the OAB052  
105 metagenome and *G. vaginalis* genome records in GenBank. Thus, we hypothesize that the larger viral  
106 contigs detected within the OAB052 patient sample represent lysogenic phages. While here we have

107 presented the analysis of just one of these contigs, similar observations were made: viral sequences  
108 exhibited homologies to annotated prophages within bacterial species that were also found within the  
109 sample's metagenome.

110 Larger phage sequences were identified in three patient samples – OAB010, OAB018, and OAB039.

111 **Table 1** lists the contigs identified in each of these samples. While many of these larger phage sequences  
112 include novel genic content (i.e. low or no sequence homology to records in GenBank), each exhibited  
113 some homology to recognized prophage sequences within bacterial genomes (per PHAST (40)). The  
114 most similar phage species are listed in **Table 1**. Based upon the size of the assembled genome and the  
115 presence of “hallmark” viral genes (42), we were able to confidently predict the completeness of several  
116 of these assembled sequences. The phage sequences listed in **Table 1** were then annotated using the  
117 RAST server (43) (**Supplementary Table 3**). The genome map for the putative complete phage genome  
118 sequence within the OAB018 patient sample is shown in **Fig. 1** (generated using Geneious, Auckland,  
119 NZ). Phage sequences identified here are not necessarily unique to the microbiota of the urinary tract  
120 (**Supplementary Table 2**). For instance, the sequence of contig 28 from the OAB010 sample is 99%  
121 identical to a prophage found within a *Streptococcus agalactiae* strain isolated from a patient's blood  
122 sample (44), as well as from a strain isolated from a diseased tilapia (GenBank record CP016501). These  
123 larger sequences are informative both of the bioinformatic approach employed here and the samples  
124 themselves. First, complete (or near-complete) phage genomes can be reconstructed by sequencing  
125 bladder microbiome samples. Second, because we sequenced the bacterial and viral fractions together,  
126 it is possible to associate phages and their bacterial host. Last, we found evidence of related phages  
127 present in the bladder microbiota of different patients. For instance, the OAB018 and OAB039 patient  
128 samples both contain phage sequences similar to the *Lactobacillus*-infecting phages PLE2 and phi adh.  
129 These phages were first detected as prophages within the genomes of the probiotic strains *L. casei* BL23  
130 (45) and *L. gasseri* ADH, respectively. Further sequencing of the bladder microbiota is necessary to  
131 ascertain if these phage families are common constituents of the bladder virome.

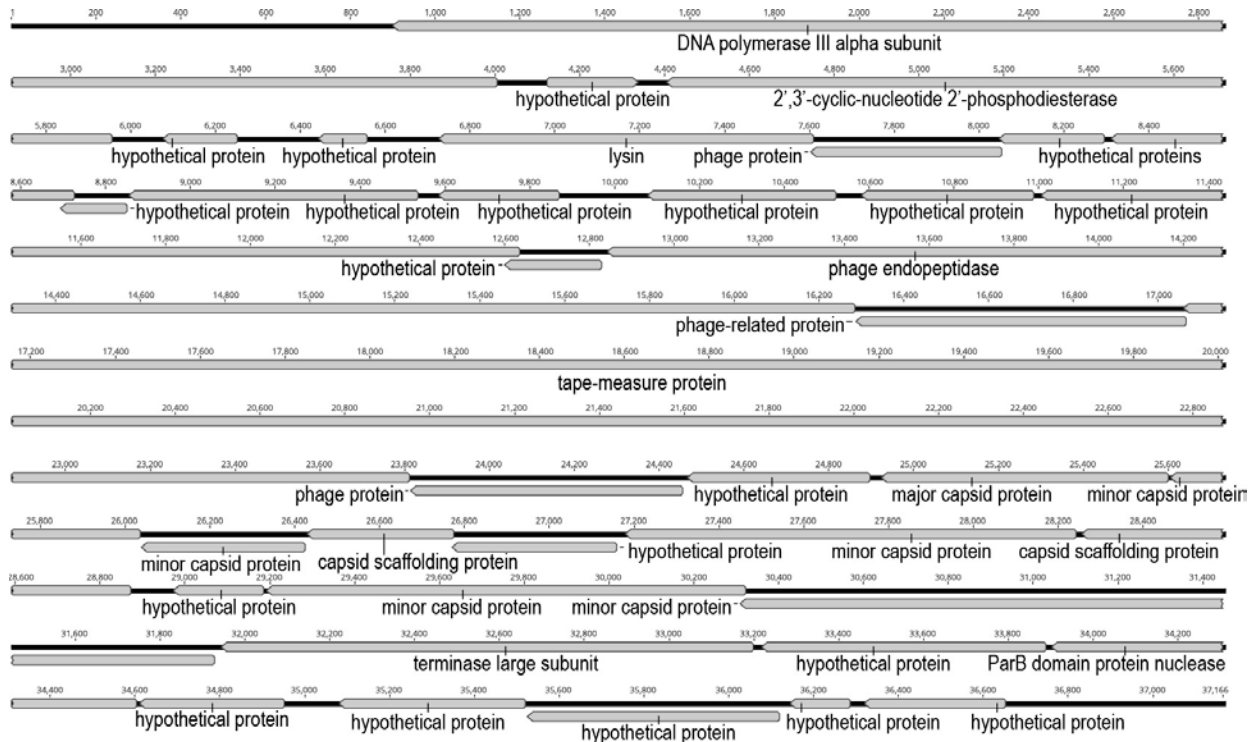
132 Five patient samples, OAB021, OAB026, OAB032, OAB042, and OAB045, contained recognizable  
133 complete genomes for the human polyomavirus JC (JCV). Furthermore, a partial genome sequence, 1023  
134 bp, was retrieved from patient sample OAB025. JCV is a circular double-stranded DNA virus (~5130 bp)  
135 and naturally occurs in the urine of healthy individuals. A previous study found that up to 80% of adults  
136 excreted JCV in their urine (46). Furthermore, JCV quasispecies have been detected in healthy

137 individuals (47). JCV, however, was not detected within any of the 10 asymptomatic “healthy”  
 138 individuals (controls) included in this study. While JCV infection has been associated with progressive  
 139 multifocal leukoencephalopathy, a fatal neurological disorder (48), JCV within individuals with  
 140 overactive bladder has yet to be studied. The prevalence of JCV within these five samples varied. Raw  
 141 reads were mapped to the RefSeq for the species (GenBank Accession: NC\_001699) using Bowtie 2 (v.  
 142 2.2.6) (35) revealing coverage of the JCV genome ranging from 12x to 726.9x. Coverage correlated with  
 143 the % reads in the sample corresponding to the JCV genome ( $r^2=0.9570$ ). JCV was most abundant in  
 144 patient samples OAB042 and OAB045, in which 4.4% and 3.2%, respectively, of the total reads  
 145 generated were classified as JCV.

Sample	Contig #	Length (kbp)	Coverage	Bacterial Blast Homology (sequence ID/ query coverage)	Most Similar Phage (length)
OAB010	28	17.5	16.56	<i>S. agalactiae</i> (99%/ 100%)	phiCT453B (36.7 kbp)
	31	8.1	11.61	<i>S. agalactiae</i> (95%/ 99%)	phiCT453B (36.7 kbp)
	39	3.4	14.21	<i>S. agalactiae</i> (100%/ 100%)	phiARI0923 (33.5 kbp)
OAB018	28	37.1	9.54	<i>L. helveticus</i> (87%/ 71%)	phig1e (42.3 kbp)
	49	26.8	10.89	<i>L. helveticus</i> (85%/ 15%)	phig1e (42.3 kbp)
	66	17.8	7.30	<i>L. allii</i> (72%/ 3%)	PLE2 (35.1 kbp)
	148	7.6	6.96	<i>L. helveticus</i> (76%/ 25%)	phi adh (43.8kbp)
OAB039	55	13.6	18.08	<i>L. allii</i> (72%/ 4%)	PLE2 (35.1 kbp)
	79	8.5	23.09	<i>L. gasseri</i> (67%/ 57%)	phi adh (43.8kbp)

146 **Table 1.** Putative complete/near-complete phage genomes identified within bladder microbiome  
 147 samples. Most similar phage sequences were determined using PHAST (40).

148



149

150 **Fig. 1.** Genome map for the 37.1 kbp contig 28 from the OAB018 patient sample.

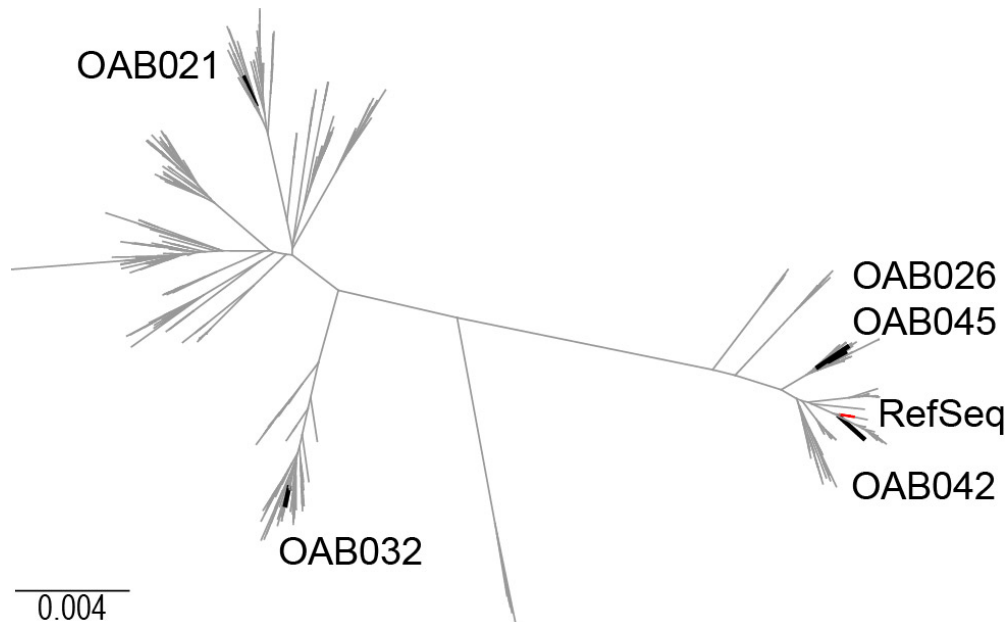
151

152 Previous research has identified subtypes of JCV and found that these subtypes can correspond with  
153 different human population groups (49). Thus, we next determined the subtypes of the five JCV  
154 complete genomes from the bladder microbiome samples by comparing their genomes to 605 publicly  
155 available genomes representative of the diversity of the species (**Supplementary Table 2**). The  
156 sequences were aligned using MUSCLE through Geneious; the alignments were trimmed, removing the  
157 tandem repeats (as their placement at the 5' or 3' end of the genome sequence varied among the  
158 genome sequence records), and a phylogenetic tree was inferred using FastTree (50) (**Fig. 2**). Clades  
159 were labeled according to their documented genotype, determined from the literature (49) and from  
160 GenBank records. Genotype classifications rely on coding sequence variation, most notably the VP1  
161 capsid coding sequence (51). This tree aids in gaining greater insight into the JCV genomes detected  
162 within the patient samples. The JCV strains identified in patient samples OAB026 and OAB045 are  
163 representative of subtype 1, genotype 1B (exhibiting greatest sequence similarity to isolates from  
164 individuals of German heritage (49)). The JCV virus from patient sample OAB042 is also categorized as  
165 subtype 1 (genotype 1A) via sequence homology (50). Subtype 1 is relatively common in the United  
166 States and Europe (52) and these three patients self-reported as "White/ Non-Hispanic." The JCV strains



167 identified in patient samples OAB032 and OAB021 are classified as belonging to genotype 3A (prevalent  
168 in Africa and southwestern Asia) and 2A (prevalent amongst individuals of Japanese and Native  
169 American decent), respectively, based upon their nearest neighbors and placement within the  
170 phylogenetic tree (**Fig. 2**) (49,53). However, the self-reported ethnicities of these patients are  
171 incongruent with the ethnicities typically associated with these subtypes; patient OAB032 self-reported  
172 as “White/ Hispanic” and patient OAB021 self-reported as “Black/ Non-Hispanic.” As the majority of  
173 sequencing and genotyping studies of JCV have been largely restricted to individuals with or without  
174 neurological diseases, our findings here prompt further investigation of the presence and genotypes of  
175 JCV in individuals with and without lower urinary tract symptoms to ascertain if JCV plays any role in  
176 urinary tract symptoms or disease.

177



178

179 **Fig. 2.** Phylogenetic tree for 610 complete genomes of JCV, including strains isolated in this study (tree  
180 branches shown in black and labeled) and the reference sequence (NC\_001699) for the species (shown  
181 in red).

182

183 Here, we have shown that challenges in isolating viral species from the low biomass bladder microbiome  
184 can be circumvented via bioinformatic classification tools; whole genome, as well as partial genome,  
185 sequences can be reconstructed from complex samples. While the sheer size of bacterial genomes lends  
186 to greater representation in whole genome sequencing data, viral genomes were detected without

187 amplification within 27 of the 30 urinary samples examined here. This further supports prior estimates  
188 of the abundance of viruses within the bladder microbiota (25,26). Moreover, as our results show, our  
189 strategy can detect both lysogenic and lytic phages, as well as eukaryotic viruses.

#### 190 **Abbreviations:**

191 OAB=overactive bladder

192 JCV=Human polyomavirus JC

#### 193 **References**

- 194 1. **Price TK, Dune T, Hilt EE, Thomas-White KJ, Kliethermes S et al.** The clinical urine culture: enhanced  
195 techniques improve detection of clinically relevant microorganisms. *J Clin Microbiol* 2016;54:1216–1222.
- 196 2. **Kass EH.** Pyelonephritis and bacteriuria. A major problem in preventive medicine. *Ann. Intern Med*  
197 1962;56:46–53.
- 198 3. **Wolfe AJ, Toh E, Shibata N, Rong R, Kenton K et al.** Evidence of uncultivated bacteria in the adult  
199 female bladder. *J Clin Microbiol* 2012;50:1376–1383.
- 200 4. **Fouts DE, Pieper R, Szpakowski S, Pohl H, Knoblach S et al.** Integrated next-generation sequencing of  
201 16S rDNA and metaproteomics differentiate the healthy urine microbiome from asymptomatic  
202 bacteriuria in neuropathic bladder associated with spinal cord injury. *J Transl Med* 2012;10:174.
- 203 5. **Khasriya R, Sathiananthamoorthy S, Ismail S, Kelsey M, Wilson M et al.** Spectrum of bacterial  
204 colonization associated with urothelial cells from patients with chronic lower urinary tract symptoms. *J*  
205 *Clin Microbiol* 2013;51:2054–2062.
- 206 6. **Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ et al.** Urine is not sterile: use of enhanced  
207 urine culture techniques to detect resident bacterial flora in the adult female bladder. *J Clin Microbiol*  
208 2014;52:871–876.
- 209 7. **Thomas-White KJ, Kliethermes S, Rickey L, Lukacz ES, Richter HE et al.** Evaluation of the urinary  
210 microbiota of women with uncomplicated stress urinary incontinence. *Am J Obstet Gynecol*  
211 2017;216:55.e1-55.e16.
- 212 8. **Brubaker L, Nager CW, Richter HE, Visco A, Nygaard I et al.** Urinary bacteria in adult women with  
213 urgency urinary incontinence. *Int Urogynecology J* 2014;25:1179–1184.
- 214 9. **Pearce MM, Zilliox MJ, Rosenfeld AB, Thomas-White KJ, Richter HE et al.** The female urinary  
215 microbiome in urgency urinary incontinence. *Am J Obstet Gynecol* 2015;213:347.e1-11.
- 216 10. **Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K et al.** The female urinary  
217 microbiome: a comparison of women with and without urgency urinary incontinence. *mBio*  
218 2014;5:e01283-1214.
- 219 11. **Thomas-White KJ, Hilt EE, Fok C, Pearce MM, Mueller ER et al.** Incontinence medication response  
220 relates to the female urinary microbiota. *Int Urogynecology J* 2016;27:723–733 (2016).

- 221 12. **Nienhouse V, Gao X, Dong Q, Nelson DE, Toh E et al.** Interplay between bladder microbiota and  
222 urinary antimicrobial peptides: mechanisms for human urinary tract infection risk and symptom severity.  
223 *PloS One* 2014;9:e114185.
- 224 13. **Karstens L, Asquith M, Davin S, Stauffer P, Fair D et al.** Does the urinary microbiome play a role in  
225 urgency urinary incontinence and its severity? *Front Cell Infect Microbiol* 2016;6:78.
- 226 14. **Thomas-White KJ, Lin H, Gao X, Fok C, Mueller ER et al.** Urinary symptoms and associated urinary  
227 microbes in urogynecologic surgical patients. 2018a. In review.
- 228 15. **Thomas-White KJ, Gao X, Lin H, Fok C, Ghanayem K et al.** Urinary microbes and post-operative  
229 urinary tract infection risk in urogynecologic surgical patients. 2018b. In review.
- 230 16. **Iwasawa A, Kumamoto Y, Maruta H, Fukushima M, Tsukamoto T et al.** Presence of human  
231 papillomavirus 6/11 DNA in condyloma acuminatum of the urinary bladder. *Urol Int* 1992;48:235–8.
- 232 17. **Echavarria M, Forman M, Ticehurst J, Dumler JS, Charache P.** PCR method for detection of  
233 adenovirus in urine of healthy and human immunodeficiency virus-infected individuals. *J Clin Microbiol*  
234 1998;36:3323–6.
- 235 18. **Karim RZ, Rose BR, Brammah S, Scolyer RA.** Condylomata acuminata of the urinary bladder with  
236 HPV 11. *Pathology* 2005;37:176–8.
- 237 19. **Burián Z, Szabó H, Székely G, Gyurkovits K, Pankovics P et al.** Detection and follow-up of torque  
238 teno midi virus (“small anelloviruses”) in nasopharyngeal aspirates and three other human body fluids in  
239 children. *Arch Virol* 2011;156:1537–41.
- 240 20. **Hirsch HH, Kardas P, Kranz D, Leboeuf C.** The human JC polyomavirus (JCPyV): virological  
241 background and clinical implications. *APMIS* 2013;121:685–727.
- 242 21. **Rinaldo CH, Tylden GD, Sharma BN.** The human polyomavirus BK (BKPyV): virological background  
243 and clinical implications. *APMIS* 2013;121:728–45.
- 244 22. **Assetta B, Atwood WJ.** The biology of JC polyomavirus. *Biol Chem* 2017;398:839–55.
- 245 23. **Brown-Jaque M, Muniesa M, Navarro F.** Bacteriophages in clinical samples can interfere with  
246 microbiological diagnostic tools. *Sci Rep* 2016;6:33000.
- 247 24. **Malki K, Sible E, Cooper A, Garretto A, Bruder K et al.** Seven bacteriophages isolated from the  
248 female urinary microbiota. *Genome Announc* 2016;4:e01003-16.
- 249 25. **Miller-Ensminger T, Garretto A, Brenner J, Thomas-White K, Zambom A et al.** Bacteriophages of the  
250 urinary microbiome. *J Bacteriol* 2018;200:e00738-17.
- 251 26. **Santiago-Rodriguez TM, Ly M, Bonilla N, Pride DT.** The human urine virome in association with  
252 urinary tract infections. *Front Microbiol* 2015;6:14.
- 253 27. **Rani A, Ranjan R, McGee HS, Metwally A, Hajjiri Z et al.** A diverse virome in kidney transplant  
254 patients contains multiple viral subtypes with distinct polymorphisms. *Sci Rep* 2016;6:33327.

- 255 28. **Sender R, Fuchs S, Milo R.** Revised estimates for the number of human and bacteria cells in the  
256 body. *PLoS Biol* 2016;14:e1002533.
- 257 29. **Minot S, Sinha R, Chen J, Li H, Keilbaugh SA et al.** The human gut virome: Inter-individual variation  
258 and dynamic response to diet. *Genom Res* 2011;21:1616-1625.
- 259 30. **Marique P, Bolduc B, Walk ST, van der Oost J, de Vos WM et al.** Healthy human gut phageome. *Proc*  
260 *Natl Acad Sci USA* 2016;113:10400-10405.
- 261 31. **Yilmaz S, Allgaier M, Hugenholtz P.** Multiple displacement amplification compromises quantitative  
262 analysis of metagenomes. *Nat Methods* 2010;7:943-4.
- 263 32. **Dutilh B, Cassman N, McNair K, Sanchez S, Silva G et al.** A highly abundant bacteriophage  
264 discovered in the unknown sequences of human faecal metagenomes. *Nature Commun* 2014;5:4498.
- 265 33. **Mokili JL, Dutilh BE, Lim YW, Schneider BS, Taylor T et al.** Identification of a novel human  
266 papillomavirus by metagenomic analysis of samples from patients with febrile respiratory illness. *PLoS*  
267 *One* 2013;8:e58404.
- 268 34. **Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ.** Evaluation of methods for the extraction and  
269 purification of DNA from the human microbiome. *PLoS One* 2012;7:e33865.
- 270 35. **Langmead B, Salzberg SL.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012;9:357-  
271 359.
- 272 36. **Joshi NA, Fass JN.** Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files.  
273 2011. <https://github.com/najoshi/sickle>.
- 274 37. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al.** SPAdes: A new genome assembly  
275 algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455-477.
- 276 38. **Garretto A, Hatzopoulos T, Kalesinskas L, Putonti C.** virMine: Automated detection and annotation  
277 of complete viral genomes from complex metagenomic samples. 2017.  
278 <https://github.com/thatzopoulos/virMine>.
- 279 39. **Krupovic M, Forterre P.** Single-stranded DNA viruses employ a variety of mechanisms for integration  
280 into host genomes. *Ann NY Acad Sci* 2015;1341:41-53.
- 281 40. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** PHAST: A fast phage search tool. *Nucl Acids Res*  
282 2011;39:W347-52.
- 283 41. **Malki K, Shapiro JW, Price TK, Hilt EE, Thomas-White K et al.** Genomes of *Gardnerella* strains reveal  
284 an abundance of prophages within the bladder microbiome. *PLoS One* 2016;11:e0166757.
- 285 42. **Koonin EV, Senkevich TG, Dolja VV.** The ancient Virus World and evolution of cells. *Biol Direct*  
286 2006;1:29.
- 287 43. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T et al.** The RAST Server: Rapid annotations using  
288 subsystems technology. *BMC Genomics* 2008;9:75.

- 289 44. **Teatero S, McGeer A, Li A, Gomes J, Seah C *et al.*** Population structure and antimicrobial resistance  
290 of invasive serotype IV group B *Streptococcus*, Toronto, Ontario, Canada. *Emerg Infect Dis* 2015;21:585-  
291 91.
- 292 45. **Dieterle ME, Martin JF, Durán R, Nemirovsky SI, Rivas CS *et al.*** Characterization of prophages  
293 containing “evolved” Dit/Tal modules in the genome of *Lactobacillus casei* BL23. *Appl Microbiol*  
294 *Biotechnol* 2016;100:9201-9215.
- 295 46. **Chang H, Wang M, Tsai RT, Lin HS, Huan JS *et al.*** High incidence of JC viruria in JC-seropositive older  
296 individuals. *J Neurovirol* 2002;8:447–451.
- 297 47. **Van Loy T, Thys K, Tritsmans L, Stuyver LJ.** Quasispecies analysis of JC virus DNA present in urine of  
298 healthy subjects. *PLoS One* 2013;8:e70950.
- 299 48. **Khalili K, Gordon J, White MK.** The polyomavirus, JCV and its involvement in human disease. *Adv Exp*  
300 *Med Biol* 2006;577:274-87.
- 301 49. **Shackelton LA, Rambaut A, Pybus OG, Holmes EC.** JC virus evolution and its association with human  
302 populations. *J Virol* 2006;80:9928-33.
- 303 50. **Price M, Dehal P, Arkin A.** FastTree 2 - approximately maximum-likelihood trees for large  
304 alignments. *PLoS One* 2010;5:e9490.
- 305 51. **Agostini HT, Ryschkewitsch CF, Stoner GL.** Genotype profile of human polyomavirus JC excreted in  
306 urine of immunocompetent individuals. *J Clin Microbiol* 1996;34:159-64.
- 307 52. **Reid CE, Li H, Sur G, Carmillo P, Bushnell S *et al.*** Sequencing and analysis of JC virus DNA from  
308 natalizumab-treated PML patients. *J Infect Dis* 2011;204:237-44.
- 309 53. **Saruwatari L, Sugimoto C, Kitamura T, Ohno N, Sakai E *et al.*** Asian domains of four major genotypes  
310 of JC virus, Af2, B1-b, CY and SC. *Arch Virol* 2002;147:1-10.

311

## 312 **Funding Information**

313 This work was supported by the NIH (R01 DK104718 to A.J.W) and NSF (ABI 1149387 and 1661357 to  
314 C.P.). A.G. is supported by the Carbon Research Fellowship at Loyola University Chicago and the CREU  
315 research fellowship (CRA).

316

## 317 **Acknowledgements**

318 For prior patient recruitment, we want to acknowledge the Loyola Urinary Education and Research  
319 Collaborative (LUEREC), specifically Mary Tulke RN and Drs. Linda Brubaker, Elizabeth Mueller, Cynthia  
320 Brincat, Susanne Taegel, and Tanaka Dune and the patients who provided the samples for this study.

321

322

323 **Conflicts of Interest**

324 The authors declare that there are no conflicts of interest.

325

326 **Ethical Statement**

327 N/A

328

329

330 **SUPPLEMENTARY LEGENDS:**

331 **Supplementary Table 1:** GenBank Accession number, genome assembly statistics, and virMine analysis  
332 statistics for the 30 metagenomes examined.

333 **Supplementary Table 2:** Results for contigs predicted to be viral queried via BLAST against the nr/nt  
334 database via the NCBI web interface.

335 **Supplementary Table 3:** Annotations for putative complete/near-complete phage genomes listed in  
336 Table 1.

337 **Supplementary Figure 1:** Workflow for bioinformatic analysis.

338