


Article

Human Polyomaviruses (HPyV) in Wastewater and Environmental Samples from the Lisbon Metropolitan Area: Detection and Genetic Characterization of Viral Structural Protein-Coding Sequences

Ana Carolina Condez ¹, Mónica Nunes ^{2,3}, Andreia Filipa-Silva ^{2,3}, Inês Leonardo ^{2,3}
and Ricardo Parreira ^{1,4,*}

- ¹ Unidade de Microbiologia Médica, Instituto de Higiene e Medicina Tropical (IHMT), Universidade Nova de Lisboa (NOVA) Rua da Junqueira N°100, 1349-008 Lisboa, Portugal; a.condez@campus.fct.unl.pt
² iBET, Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2781-901 Oeiras, Portugal; mnunes@ibet.pt (M.N.); andreiasilva@ibet.pt (A.F.-S.); ines.leonardo@ibet.pt (I.L.)
³ ITQB NOVA, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal
⁴ Global Health and Tropical Medicine (GHTM) Research Centre, 1349-008 Lisboa, Portugal
* Correspondence: Ricardo@ihmt.unl.pt; Tel.: +351-21-3652600 (ext. 454)



Citation: Condez, A.C.; Nunes, M.; Filipa-Silva, A.; Leonardo, I.; Parreira, R. Human Polyomaviruses (HPyV) in Wastewater and Environmental Samples from the Lisbon Metropolitan Area: Detection and Genetic Characterization of Viral Structural Protein-Coding Sequences. *Pathogens* **2021**, *10*, 1309. <https://doi.org/10.3390/pathogens10101309>

Academic Editor: Martine Peeters

Received: 13 September 2021

Accepted: 8 October 2021

Published: 12 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Due to the lack of reliable epidemiological information regarding the geographic distribution and genetic diversity of human polyomaviruses (HPyV) in Portugal, we addressed these issues in this initial study by focusing on the Lisbon Metropolitan area, the most populated and culturally diverse hub in the country. The HPyV structural protein-coding sequence was partially amplified using two touch-down PCR multiplex protocols, starting from water samples, collected between 2018 and 2020, where viral genomes were detected. The obtained results disclosed the frequent detection of HPyV1, HPyV2, HPyV5, and HPyV6 in 35.3% ($n = 6$), 29.4% ($n = 5$), 47.1% ($n = 8$) and 29.4% ($n = 5$), respectively, of the water samples analyzed. The sequences assigned to a given viral species did not segregate to a single genotype, this being especially true for HPyV2 for which five genotypes (including a putative new genotype 9) could be identified. The phylogenetic trees obtained for HPyV5 and HPyV6 had less resolving power than those obtained for HPyV1/HPyV2, but both viruses were shown to be genetically diverse. This analysis emphasizes the epidemiological helpfulness of these detection/genetic characterization studies in addition to being relevant tools for assessment of human waste contamination.

Keywords: human polyomavirus; wastewater; environmental water; multiplex-PCR; phylogenetic analysis; Portugal

1. Introduction

Polyomaviruses are small (~50 nm) non-enveloped viruses with circular double-stranded genomes with approximately 5 kbp, that infect mammals (humans included) and birds [1]. The International Committee for the Taxonomy of Viruses places them within the *Polyomaviridae* family (Order *Sepolyvirales*), where they are distributed into six genera, encompassing a total of 112 viral species [2]. At least 12 of these species are considered bona fide human polyomaviruses (HPyV), and these have been placed in three genera (*Alpha-*, *Beta-*, and *Deltapolyomavirus*), defined as a function of phylogenetic relationships established based on the analysis of the large T antigen [2,3].

The high global seroprevalence regarding the great majority of HPyVs suggests they infect humans from an early age [4,5]. However, despite being true human pathogens, the diseases caused by HPyVs are usually associated with diverse stages of immunosuppression, either in senior individuals, those infected with the human immunodeficiency virus (HIV), or

undergoing chemotherapy, or in association with organ transplants [6,7]. Four of these viruses have been linked with diseases that may be serious/life-threatening, and which include hemorrhagic cystitis in patients who have undergone bone marrow transplantation (caused by HPyV1/BK polyomavirus), progressive multifocal leucoencephalopathy (PML) associated with AIDS (caused by HPyV2/JC polyomavirus), Merkel-cell carcinoma, an aggressive cell tumor caused by HPyV5 (or MC virus), or trichodysplasia spinulosa, a rare skin disease caused by HPyV8 (or TS polyomavirus) [8] (Supplementary Table S1).

Human polyomaviruses are frequently found in excreta as it has been repeatedly demonstrated since the isolation of the first (named HPyV1 or the BK polyomavirus) from urine in the early 1970s [9]. Indeed, it has become increasingly apparent that they are present in the environment all year-round, which supports their use as indicators of human waste contamination. Human health risk assessment regarding the presence of fecal contaminants in environmental waters has traditionally been based on bacterial counts and especially on *Bacteroides* sp., which outnumber conventional fecal indicator bacteria (FIB), such as coliforms and enterococci [10–12]. However, several human viruses, including adenoviruses, aichiviruses, and human polyomaviruses have been viewed as better options for human enteric contamination assessment due to their host-specificity, non-seasonal distribution, and high titers in both influent (i.e., untreated sewage) and environmental water samples [13].

The lack of consistent epidemiological data regarding the geographic distribution, and genetic diversity, of human polyomaviruses in Portugal, has prompted us to approach this study from a molecular epidemiology perspective. Indeed, molecular epidemiology studies based on the analysis of viruses (most frequently their genomes) frequently disclose genetic diversity and population fluctuations and allow an assessment of the viral factors that may be associated with different outcomes in virus–host interactions [14].

Due to logistic constraints associated with the COVID-19 pandemic, this exploratory study was carried out from the analysis of water samples that were obtained in the Lisbon metropolitan area (LMA), where northern and southern areas are separated by the Tagus River. For this purpose, two touch-down nested-PCR multiplex protocols were developed for the detection of HPyVs, not only from raw sewage but also from environmental waters (estuary/river/rivulet) samples. This was followed by their genetic analyses mostly focused on DNA sequencing of individual plasmid recombinant molecules, but further complemented with Next Generation Sequencing (NGS) data. When combined with phylogenetic analyses, our data demonstrated the circulation of diverse strains of HPyV1, HPyV2, HPyV5, and HPyV6.

2. Results

2.1. Primer Design and Touch-Down PCR Setup

The work carried out in this report was based on the detection and genetic analyses of partial HPyV genomic sequences amplified from both wastewater (raw sewage; $n = 9$) and environmental (estuary/river/rivulet; $n = 8$) samples ($n_{\text{total}} = 17$) collected at 14 sites in LMA (Figure 1), within a three-year window frame, from 2018 to 2020. Because of the exploratory nature of this analysis, the water samples used originated from a relatively small number of sites. Viral-like particles were concentrated from these samples using the skimmed milk flocculation protocol, and the HPyV structural protein-coding region was targeted for amplification. A preliminary analysis of the complete structural-protein coding regions using reference sequences downloaded from GenBank (Supplementary Figure S1) indicated that viral sequences were distributed into two large monophyletic clusters. One of these included HPyV3, HPyV4, HPyV6, HPyV7, and the so-called Quebec polyomavirus (QPyV) [15], while a larger cluster grouped HPyV1, HPyV2, HPyV5, HPyV8, HPyV9, HPyV10, HPyV11, HPyV12, HPyV13, and HPyV14. The observed distribution of HPyV sequences into two separate phylogenetic lineages (Lineage-A/L-A and Lineage-B/L-B) guided the design of amplification primers, using lineage-specific multiple sequence alignments (MSA). However, sequences from HPyV12, HPyV13, HPyV14, and QPyV were

not considered for primer design due to their low prevalence and/or uncertainty around their assignment as human-specific viruses [16–18].

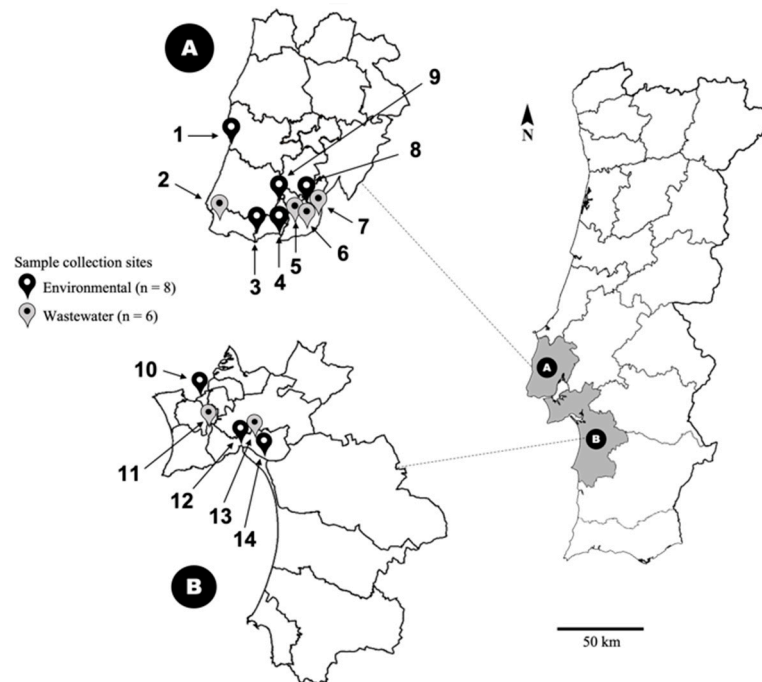


Figure 1. Geographic distribution of the wastewater and environmental sampling sites (numbered from 1 to 14). The areas indicated by (A) (sites 1 to 9) and (B) (sites 10 to 14) correspond to the two largest administrative districts of the LMA that are separated by the Tagus River and its estuary. The number of wastewater and environmental sampling sites is indicated between brackets.

The PCR amplification primers used during this study were defined using software tools and visual inspection of MSA of the structural-protein coding regions of the HPyV segregating into L-A and L-B (Table 1), and aimed to obtain segments with approximately similar, though not identical, sizes. These primers were then used to define two multiplex nested-PCR approaches for HPyV amplification, which were designated PCR-A and PCR-B, respectively for amplification of genomes clustering within L-A and L-B. As indicated in Table 1, each round of PCR-A included a total of 8 types of primers for the amplification of HPyV1/HPyV2, HPyV5, HPyV8/HPyV9, and HPyV10/HPyV11, while for PCR-B a smaller number of primers was used ($n = 4$) to target the amplification of HPyV3/HPyV4, HPyV6/HPyV7. Furthermore, when aligned to a reference HPyV sequence, the designed primers showed none to minimal ($n = 1$) mismatched to a homologous reference, but higher (from 2 to the total length of the primer) with a heterologous HPyV (Table 1).

Table 1. Primers used for detection of human polyomaviruses using touchdown nested-PCRs, expected amplicon sizes and number of primer mismatches for each HPyV type.

Reaction	Primers (5'-3')	Number of Mismatches ^a											
		HPyV1	HPyV2	HPyV3	HPyV4	HPyV5	HPyV6	HPyV7	HPyV8	HPyV9	HPyV10	HPyV11	
PCR-A	1st round	HPyV1/2Fo: CTGCTCCTCAATGGATGTTGC (from 1489 to 1509) ^b	0	0	-	-	10	-	-	7	9	7	17
		HPyV1/2Ro: ATCATRITCTGGGTCCCCTGGAAG (from 2606 to 2584) ^b	0	0	-	-	8	-	-	4	7	7	8
		HPyV5Fo: GAAAATAGCTTGCTGCATTCTG (from 879 to 900) ^c	35	15	-	-	0	-	-	7	10	10	14
		HPyV5Ro: GGGCCCACTCCATTCTCATC (from 1979 to 1960) ^c	5	6	-	-	0	-	-	2	2	5	8
		HPyV8/9Fo: AGGAGGRGCAMATCAAAGAG (from 1198 to 1217) ^d	3	2	-	-	9	-	-	0	0	3	10
		HPyV8/9Ro: ATAAAYTCTGACTTCTTCMAC (from 2336 to 2316) ^d	6	5	-	-	4	-	-	0	0	10	8
		HPyV10/11Fo: CCTGGATAYAGACAMTTTSA (from 806 to 825) ^e	9	17	-	-	10	-	-	11	13	0	0
		HPyV10/11Ro: TTAMAGGATAAGGATTTCTVA (from 2334 to 2313) ^e	7	8	-	-	6	-	-	5	1	0	1
		Expected amplicon size (bp) ^f	1113	1094	-	-	1101	-	-	1139	1136	1529	1474
PCR-A	2nd round	HPyV1/2Fi: GTACGGGACTGTAACACCTGC (from 1526 to 1546) ^b	0	0	-	-	12	-	-	8	8	11	14
		HPyV1/2Ri: CCATACATAGGCTGCCCATC (from 2534 to 2515) ^b	0	0	-	-	7	-	-	7	7	3	7
		HPyV5Fi: CAATCAAACCTAGTGAATCTG (from 951 to 971) ^c	31	18	-	-	0	-	-	12	10	13	10
		HPyV5Ri: GGATCAGGACACCATACTTC (from 1859 to 1840) ^c	7	5	-	-	0	-	-	7	5	6	4
		HPyV8/9Fi: GGWTTGTATGGTGATATAAC (from 1250 to 1269) ^d	7	6	-	-	9	-	-	0	0	5	7
		HPyV8/9Ri: ATTAAARTAYCTAGGTAGGCCTCT (from 2192 to 2170) ^d	4	5	-	-	6	-	-	0	0	5	5
		HPyV10/11Fi: AGAGCTTTTTGGGARGCTKT (from 881 to 900) ^e	8	7	-	-	14	-	-	6	10	0	1
		HPyV10/11Ri: CCCAGGCCTCYACWGGATAR (from 2055 to 2036) ^e	3	6	-	-	5	-	-	4	5	0	1
		Expected amplicon size (bp) ^f	989	985	-	-	909	-	-	943	940	1175	1136

Table 1. Cont.

Reaction	Primers (5'-3')	Number of Mismatches ^a											
		HPyV1	HPyV2	HPyV3	HPyV4	HPyV5	HPyV6	HPyV7	HPyV8	HPyV9	HPyV10	HPyV11	
PCR-B	1st round	HPyV3/4Fo: GGACGTGTTC AATAGAATTGC (from 980 to 1000) ^g	-	-	0	0	-	10	8	-	-	-	-
		HPyV3/4Ro: CCAATGCCATTTTCATCCAA (from 2282 to 2263) ^g	-	-	0	0	-	5	4	-	-	-	-
		HPyV6/7Fo: GACTCGGCCCAAGARTTGG (from 708 to 726) ^h	-	-	9	13	-	0	0	-	-	-	-
		HPyV6/7Ro: GCACCTGTGGCTTCTGRGG (from 2220 to 2202) ^h	-	-	7	6	-	0	0	-	-	-	-
		Expected amplicon size (bp) ^f	-	-	1303	1318	-	1513	1513	-	-	-	-
PCR-B	2nd round	HPyV3/4Fi: CATCATATTACAATRCGGGG (from 1015 to 1034) ^g	-	-	0	0	-	8	7	-	-	-	-
		HPyV3/4Ri: GTTCCATTCTRTACAGCTC (from 1895 to 1876) ^g	-	-	0	0	-	7	5	-	-	-	-
		HPyV6/7Fi: TGGCACTTCAAYTGTGGTTG (from 738 to 757) ^h	-	-	27	18	-	0	0	-	-	-	-
		HPyV6/7Ri: WCCAATKACATCCAAGGGGC (from 1730 to 1711) ^h	-	-	16	5	-	0	0	-	-	-	-
		Expected amplicon size (bp) ^f	-	-	881	893	-	993	1002	-	-	-	-

K (G or T), M (A or C), R (A or G), S (C or G), W (A or T), Y (C or T), V (A, C or G). (HPyV1 = BK polyomavirus; HPyV2 = JC polyomavirus; HPyV3 = KI polyomavirus; HPyV4 = WU polyomavirus; HPyV5 = MC polyomavirus; HPyV8 = TS polyomavirus; HPyV11 = STL polyomavirus). ^a According to the GenBank reference sequences for each HPyV. The number of nucleotides in indels was counted as a mismatch.

^b Positions numbered according to the HPyV1 GenBank reference sequence NC_001538. ^c Positions numbered according to the HPyV5 GenBank reference sequence NC_010277. ^d Positions numbered according to the HPyV8 GenBank reference sequence NC_014361. ^e Positions numbered according to the HPyV10 GenBank reference sequence NC_018102. ^f Estimated according to the GenBank reference sequence for each HPyV. ^g Positions numbered according to the HPyV3 GenBank reference sequence NC_009238. ^h Positions numbered according to the HPyV6 GenBank reference sequence NC_014406.

Due to the large number of primers used, and the complexity of the template samples (which contained DNA from multiple sources), many possible competitor sequences may hybridize with the primers used. Therefore, a touch-down PCR protocol was chosen to increase both the specificity and sensitivity of the anticipated amplification [19]. In a preliminary experiment, we assessed the performance of a multiplex versus a singleplex approach using primers HPyV1/2Fo and HPyV1/2Ro (1st round) and HPyV1/2Fi and HPyV1/2Fi (2nd round) and a diluted extract of HPyV2 DNA (clinical isolate) as a positive control. Not surprisingly, the obtained results showed that despite using a nested-PCR approach, the expected 1 kb DNA (approximate size) fragment could be detected after the 1st round of amplification in the singleplex format, but only after the 2nd round in the multiplex format. This seems to indicate a loss in the detection sensitivity when multiple primers were combined in a single amplification reaction (Supplementary Figure S2).

The detection of polyomavirus DNA was deemed successful whenever an expected-sized amplicon was observed on agarose gels. For PCR-B, only the HPyV6/HPyV7 primer combination produced amplification specific products after the 2nd round, with no spurious amplicons ever been observed. On the contrary, PCR-A sometimes yielded non-specific amplification products that could not be eliminated with varying amounts of input DNA. Altogether, from a total of nine wastewater and eight environmental samples, HPyV DNA was detected from ten of these (58.8%). However, and as expected, HPyV DNA was consistently detected in sewage samples (100% amplification success) but not from environmental samples (12.5% success), most probably reflecting a considerable difference in viral titer between both sample types.

2.2. Analysis of PCR Amplification Results

Within the geographic range this study focused on, and considering the heterochronous sampling strategy used, the amplification products resulting from either PCR-A or PCR-B were anticipated not to conform to a given HPyV sequence type. Therefore, the genetic complexity of the obtained amplicons was investigated using an approach essentially based on DNA cloning and analysis of a total number of 73 sequences, most of which ($n = 68$) were obtained from the analysis of independent recombinant plasmids. Four of the total sequences were obtained as a result of NGS analysis of a pool of PCR products (see below), while the remainder corresponded to an amplicon obtained using only HPyV5-specific primers. The obtained sequences were subsequently genetically typed using BLASTn (all of them), and phylogenetic analysis (for those with sizes >650 nt). Altogether, the obtained results disclosed the presence of HPyV1, HPyV2, HPyV5 and HPyV6 in 35.3% ($n = 6$), 29.4% ($n = 5$), 47.1% ($n = 8$) and 29.4% ($n = 5$) respectively, considering the number of water samples analyzed (Figure 2). On the other hand, when the analysis was performed considering the number of obtained sequences, those corresponding to HPyV1, HPyV2, HPyV5, and HPyV6 accounted for 20.55% ($n = 15$), 21.92% ($n = 16$), 35.62% ($n = 26$), and 21.91% ($n = 16$), respectively (Table 2). The process of assigning a viral sequence to a given species (e.g., HPyV1, HPyV2, and so on) was supported by the high phylogenetic signal of the dataset used (which resolved 95.60% of the random sequence quartets using likelihood mapping). This distribution of the amplified sequences into four viral species was further confirmed by NGS sequencing using a pooled library of the 10 PCR-A and PCR-B amplicons obtained from seven samples (six wastewater/one environmental), collected in 2018, 2019, and 2020. From a total of 496,197 HPyV-specific short-sequence reads (accounting for 86.0% of the total number of reads analyzed), 148,851 (30.0%) were assigned to HPyV1, 111,781 (22.5%) to HPyV2, 201,255 (40.6%) to HPyV5, and 34,310 (6.9%) to HPyV6. High coverage and full-length sequencing of the PhiX174 used as a control supported the distribution of the HPyV sequences. Indeed, the qualitative distribution of viral types was the same as the one obtained for individual plasmid analysis. However, many factors may have contributed to the numerical difference in the distribution of the different HPyV types when the results of the two approaches are compared. Curiously, from the 69 sequences corresponding to either PCR amplicons or recombinant plasmids, and for which the sequence at the

amplicon ends could be inspected, nine (13.0%; $n = 4$ HPyV1; $n = 3$ HPyV2; $n = 2$ HPyV5) had been amplified with a priori unexpected primers combination, where one of them was heterologous when considered the type of the sequence in question (e.g., HPyV1 amplified with primer HPyV1/2Fi and HPyV8/9Ri).

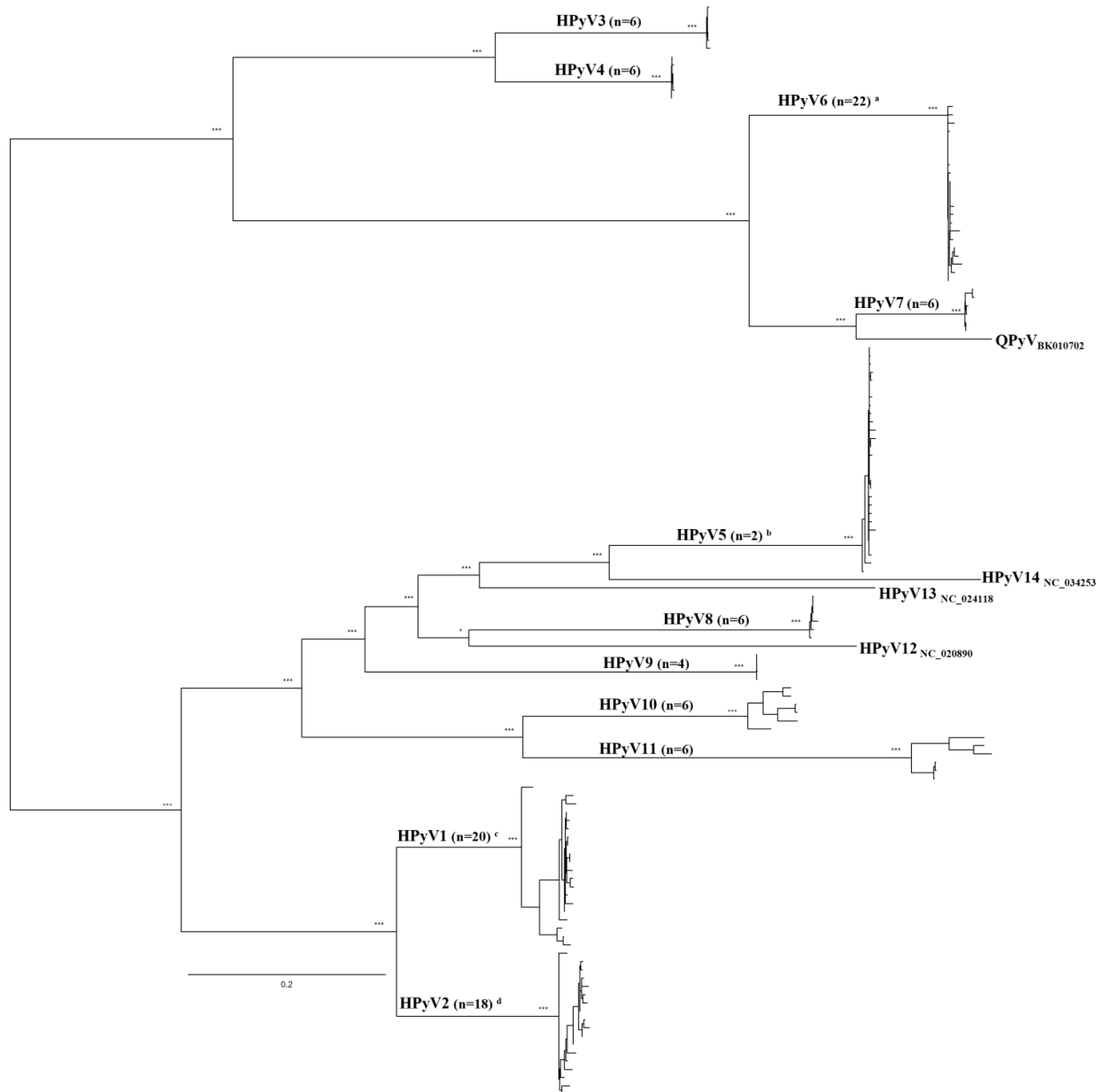


Figure 2. Phylogenetic analysis of HPyV sequences (maximum likelihood) considering a structural protein-coding genome fragment with a minimum of 650 nt. At specific branches, the identity of the HPyV monophyletic clusters is indicated. The number of sequences in each of these is indicated in brackets. The ^a ($n = 16$), ^b ($n = 23$), ^c ($n = 14$), and ^d ($n = 12$) in superscript refer to the number of sequences obtained in this study. For HPyV12, 13, 14, and the QPyV the reference sequences are indicated as isolated branches, where the subscripts indicate their respective accession numbers. At the main tree branches, the number of “*” indicates the support revealed by the different phylogenetic reconstruction methods used, assuming as relevant aLRT and bootstrap values $\geq 75\%$, as well as posterior probability values ≥ 0.80 when a Bayesian approach was further used to confirm the proposed phylogenetic clustering. The bar indicates the average number of substitutions per site.

Table 2. Distribution of the HPyV sequences analyzed in this work into different genetic types.

HPyV	Genotype/Subtype/Subgroup (% Total; % per Virus Species)	Accession Number
PyV1	Ia (n = 2/2.74% total; 13.33% HPyV1)	LC636348 ^{a=5,c} , LC636350
	Ib1 (n = 11/15.07% total; 73.33% HPyV1)	LC636351, LC636353, LC636359, LC636362, LC636363, LC636364, LC636374, LC636378, LC636381, LC636385, LC636400
	III (n = 1/1.37% total; 6.67% HPyV1)	LC636340
	I-like (n = 1/1.37% total; 6.67% HPyV1)	LC636403
HPyV2	1A (n = 3/4.11% total; 18.75% HPyV2)	LC636370, LC636395, LC636401
	1B (n = 3/4.11% total; 18.75% HPyV2)	LC636349 ^{a=5,c} , LC636360 ^{a=13,c} , LC636404
	2A2 (n = 3/4.11% total; 18.75% HPyV2)	LC636358, LC636379, LC636380
	2B (n = 1/1.37% total; 6.25% HPyV2)	LC636396
	3A (n = 2/2.74% total; 12.50% HPyV2)	LC636377, LC636399 ^{a=11,c}
	4 (n = 3/4.11% total; 18.75% HPyV2)	LC636357, LC636365 ^{a=13,c} , LC636394
	9 (n = 1/1.37% total; 6.25% HPyV2)	LC636376
HPyV5	Africa (n = 8/10.96% total; 30.77% HPyV5)	LC636333, LC636335, LC636352, LC636355, LC636356, LC636384, LC636387 ^{a=5,c} , LC636398
	Europe/North America (n = 17/23.29% total; 65.38% HPyV5)	LC636341, LC636342, LC636343, LC636347, LC636354, LC636369, LC636373, LC636375, LC636386, LC636388 ^{a=5,c} , LC636389, LC636390, LC636391 ^{a=5,c} , LC636392, LC636393, LC636397, LC636402
	Eu/NAm/Af (n = 1/1.37% total; 3.85% HPyV5)	LC636361
HPyV6	n.a. ^b (n = 16/21.91% total; 100.00% HPyV6)	LC636334, LC636336, LC636337, LC636338, LC636339, LC636344, LC636345, LC636346, LC636366, LC636367, LC636368, LC636371, LC636372, LC636382, LC636383, LC636405

^a The collection site of the sequences that were not phylogenetically analyzed is marked using the same annotation as the one used in the trees. ^b Not applicable. ^c Sequences typed using only BLASTn results.

For a selection of DNA extracts prepared from two wastewater and two environmental samples from which no HPyV had been found using the multiplex approach, a singleplex amplification assay was conducted to investigate whether the amplification performance would be impacted (Supplementary Table S1). Hence, from the totality of the analyzed samples, only one of them persisted as negative for every reaction while for the remainder, the individual amplification reactions revealed the detection of HPyV1 and/or HPyV2, HPyV5, or HPyV6 DNA. Although no other HPyV was identified, the detection of at least one of the previously detected HPyV in three of the four samples seems to indicate that the use of multiple pairs of possibly competing primers in the multiplex format reduces the

sensitivity of the amplification. Indeed, for at least one of the samples, and as the use of control DNA had already shown, the singleplex format may reveal amplification products in the first round of amplification, whereas these are only seen in the second round in the multiplex amplification format. Finally, while the obtained sequences allowed for their genetic characterization using multiple tools, the size of their corresponding amplicons, always above 1 kbp especially in the first amplification round, might have compromised the efficiency of detection of some viral sequence types compared to others.

Although the number of sequences obtained for each virus type was not the same, their analysis revealed considerable genetic heterogeneity for each virus type. Furthermore, the construction of species-specific phylogenetic trees was characterized by the asymmetrically different phylogenetic signals of the HPyV1 and HPyV2 datasets (90.22% and 89.62%) versus the HPyV5 and HPyV6 datasets (47.56% and 43.22%), which indicated that for the latter, the clear-cut definition of genetic lineages may not always be devoid of ambiguities and should be verified with different genetic analyses tools.

2.3. Genetic Characterization of Viral Sequences

Phylogenetic analyses carried out using a combination of maximum likelihood and a Bayesian approach (Supplementary Figure S3) showed that HPyV1 segregated into four major lineages, while for HPyV2, seven major clusters, including Portuguese sequences, were identified (Table 2). For both HPyV1 and HPyV2, one of the sequences failed to conform to a previously identified genetic subtype (LC636403 and LC636376, respectively). Therefore, for HPyV1, LC636403 has been characterized as a subtype I-like sequence, while for HPyV2, the singularity of one of LC636376, which was further confirmed by the construction of NNn and by PCOORD (Supplementary Figure S4), lead us to suggest it might correspond to a previously undescribed genotype 9. Overall, for HPyV1, Ib1 was the most frequently identified viral subtype encompassing over 73% of the HPyV1 sequences, while for HPyV2, the most diverse of the virus types identified, an even number of sequences were ascribed to genotype 1 (subtype A and B), 2 (subgroup A2), and 4 (Table 2). For HPyV5 most of the sequences (>65.3%) fell into the so-called Europe/North American cluster, as defined by Martel-Jantin et al. [20]. Due to a comparatively sparser representation of HPyV6 sequences in the public databases, the topological stability of the obtained trees did not support the unambiguous identification of individual genetic lineages (when compared to HPyV1, HPyV2, and HPyV5 trees). However, while the phylogenetic tree suggested a considerable heterogeneity of the 16 sequences obtained, this heterogeneity is somewhat abated in the corresponding NNn, and especially by PCOORD analysis, in which the great majority of the obtained sequences clustered together (with a near-neighboring LC636346), with only one of them (LC636383) segregating away from all the others (Supplementary Figure S4). Finally, the more divergent or single-segregating branches in the phylogenetic trees were investigated by recombinant detection tools, but in no case was there any evidence that they might correspond to intragenic mosaic sequences.

For most monophyletic clusters identified in the obtained phylogenetic trees (Supplementary Figure S3), the sequences associated with them were obtained from samples collected in the same year, either in the same or neighboring sampling points. When multiple sequences were obtained from the same sampling site, they did not necessarily cluster together but rather associated with different references along the branches of their respective phylogenetic trees (Supplementary Figure S3).

3. Discussion

In this study, we addressed the detection and the genetic diversity of HPyV circulating in the LMA area using wastewater and environmental water samples. Since Lisbon is the capital city of Portugal and the country's major economic, social, political, and cultural activity hub, its metropolitan naturally hosts residents from very different cultural and ethnic backgrounds. In addition, some of the samples analyzed in the course of this work had been collected before the onset of the COVID-19 pandemic, when the area was one

of Europe's most valued tourist spots, visited by thousands of tourists annually. When considered together, these conditions were expected to translate into a non-homogeneous distribution of HPyVs.

The experimental approach undertaken in this study was based on the design of two-touchdown nested-amplification protocols aiming at the detection of part of the viral structural protein-coding region of a broad-spectrum of HPyVs genomes. While the taxonomic differentiation of HPyVs is, by norm, based on the analysis of alignments of the primary sequence of the large T antigen [3], different studies have also disclosed the genetic resolution capability of the structural protein coding section of the HPyV genome [21–25]. The analysis of this region has been the base for the definition of multiple viral genotypes and subtypes [14], especially for the most systematically identified viruses, that include HPyV1 (or the BK polyomavirus), HPyV2 (or the JC polyomavirus), and HPyV5 (or the Merkel cell carcinoma polyomavirus). These expectations were further confirmed when the analysis of the phylogenetic signal of a dataset containing many representative reference sequences from all viral species (in addition to those obtained in this study) was found to be very high, with 95.60% of the randomly sampled sequence quartets being totally resolved (by likelihood mapping analysis).

Strategies of HPyV detection and analysis as the one here described have been previously used for environmental surveillance of HPyVs, repeatedly demonstrating the usefulness of human polyomaviruses as viral indicators of human waste contamination [13,26–33]. Indeed, their detection allows for the so-called microbial source tracking [34] which corresponds to an effort to track the most frequent sources of fecal contamination (human or animal) in environmental waters for human health risk assessment. While it has relied mostly on the use of *Bacteroides* spp. bacteria as indicators of fecal contamination [10], not only are they not exclusive human bacteria, but they may also multiply in the environment (especially in warmer climates), both of which are undesirable features for a reliable indicator of human-associated fecal pollution [35].

When combined with multiple approaches of DNA sequencing and their analysis, our approach allowed the investigation of the molecular epidemiology of HPyVs in LMA, which, to our knowledge, is here addressed for the first time. Its implementation called for the design of pairs of primers that were combined into two PCR protocols for targeted amplification of two groups of HPyV, supported by a preliminary phylogenetic analysis of the structural protein-coding section of viral reference sequences. On the one hand, these two groups included HPyV1, HPyV2, HPyV5, HPyV8, HPyV9, HPyV10, and HPyV11, while on the other HPyV3, HPyV4, HPyV6, and HPyV7. While a similar strategy had been previously used [33], the strategy here described differs from the latter as it does not allow for preliminary identification of the viral origin for the obtained amplicons as a function of their size. Indeed, the 2nd round amplification products had a priori similar sizes ranging from 881 bp (HPyV3) to 1175 bp (HPyV10), therefore falling into a relatively narrow size range (although considerably larger than the ones studied by Torres et al. [33]). When combined with complementary genetic analyses techniques (including phylogenetic, network, and PCOORD analyses), the experimental approach used in this study supported the detection and unambiguous characterization of different HPyV sequence types. While the sequencing effort was mostly focused on the analysis of independent recombinant DNA molecules with cloned HPyV inserts, it was further supported by an assessment of the genetic diversity of the HPyV types that could be detected by the amplification protocols used, via NGS analysis of a pooled sample of purified PCR fragments.

Whereas different studies have disclosed high rates of HPyV detection, even in environmental samples [26,30,36,37], the detection efficiency of the protocol here used was very high (100%) for wastewater, but considerably lower (12.5%) for environmental samples. Although the ability to detect HPyV genomes with high efficiency from samples collected from environments where these viruses are present in high titers (such as untreated sewage) has been testified in numerous studies carried out over a wide geographic range [30,32,33,38–40], curiously, the detection of HPyV genomes in samples collected in

western Europe have disclosed lower values (50–75%) even in wastewater samples [26,39]. These different studies most probably reflect differences in virus concentration and PCR amplification, though fluctuations of viral titers in these samples, especially in environmental samples, such as those collected in running rivers/rivulets or estuaries, cannot be discarded. Indeed, the environmental samples analyzed during this study were collected during winter, where the rainwater input is highest, or were under the diluting effect of tides, which contribute to bringing the HPyV titer down.

On the other hand, since HPyV are among the most frequently used human fecal contamination viral indicators [13,31,36,38,41], their presence in raw sewage was expected and could be confirmed in all the wastewater samples used. The most frequently found HPyVs in wastewater are HPyV1 (or BKPyV), HPyV2 (JCPyV), and HPyV5 (MCPyV) [14], and all of them were detected in this study, with HPyV5 accounting for a majority (35.62%) of the obtained DNA sequences. In addition, our analysis also revealed the presence of HPyV6 in wastewater as previously detected [42,43]. Whether the quantitative values calculated for the HPyV types detected truly reflect their distribution in the samples analyzed is open to discussion. Certainly, studies that depend on the efficiency of an *in vitro* amplification step, may be further biased by subsequent genetic bottlenecks associated with DNA cloning in the experimental workflow, which may further include a non-equivalent and finite analysis of independent clones obtained from each sample. Under these scenarios, the obtained results do not necessarily convey a quantitatively correct idea of the abundance of the different HPyV in each sample. Even the qualitative information obtained may be impacted by the amplification strategy used. Indeed, viruses other than the ones here described (e.g., HPyV3/KIPyV, HPyV10/MWPyV, and HPyV7) have been previously detected in wastewater and/or have been described to be excreted by the urinary or fecal route or due to skin peeling [44–48]. Therefore, their absence from our analysis may reflect (i) their absence, (ii) presence in low titer, or (iii) our inability to detect them. In any case, the independent recombinant plasmid vs. NGS data analysis indicated that the cloning strategy used did not introduce a major skew in our ability to identify HPyV sequences, previously amplified by PCR.

The so-called HPyV1 (also known as BKPyV) has been classified into four different genotypes (I–IV) according to the analysis of the genetic variability of the VP1 coding region [21]. Up to the present day, the majority of the HPyV1 sequences obtained belonged to genotype I, while genotypes II and IV were also detected. Genotype I is distributed worldwide, but its Ib1 and Ib2 subtypes have been associated with South-East Asians and Europeans/West Asians, respectively [49]. The frequent detection of genotype I HPyV1 is followed by that of genotype IV, where the six subtypes (IVa1, IVa2, IVb1, IVb2, IVc1, and IVc2) are prevalent in East Asia, whereas IVc2 is also being found in Europe and Northeast Asia [50]. On a global scale, and in contrast with genotypes I and IV, sequences assigned to genotypes II and III have been more scarcely detected, have a similar distribution, and appear more common in Africans [49]. Although genotype II has been more frequently associated with North/East Africans, their distributions seem to be similar, whereas genotype III has been found mostly in Africans [14].

For HPyV1, phylogenetic and/or sequence similarity searches (using BLASTn) disclosed the distribution of the obtained sequences into genotypes I and III [49]. In genotype I, although most sequences were associated with subtype Ib1, subtype Ia sequences were also detected. Furthermore, a genotype I-like sequence was singled out by phylogenetic analysis (LC636403), but this sequence was associated with genotype I by PCOORD and NNn suggested it most probably corresponds to an Ib1 sequence. Therefore, finding a majority of HPyV1 sequences associated with subtype Ib1 is congruent with the origin of most of the population in LMA. However, one of the sequences clustered among those in genotype III, which is not unexpected due to the extensive number of individuals of African origin living in the regions.

On the other hand, several algorithms have been used for genotyping HPyV2 (also known as JCPyV), allowing its assignment into several different genotypes that, as with

HPyV1 genotypes, are usually associated with specific human populations and geographical areas [51–53]. Two nomenclatures have been proposed for HPyV2 genotypes, with the one proposed and reviewed by Stoner et al. being the one used in this work [53]. Genotype 1 (further divided into 1A and 1B) is mostly associated with Europeans and North Americans, with genotype 4 registering a similar distribution pattern [52]. On the other hand, genotype 2 has revealed a higher degree of genetic complexity and has, therefore, been classified in five different subtypes: 2A, 2B, 2C, 2D, and 2E. Subtypes 2A and 2C form a paraphyletic group and encompass viral strains linked to Eastern Asians and Native Americans. Subtype 2B is usually associated with Europeans and Eastern Asians, whereas subtype 2D (which is subdivided into 2D1, 2D2, and 2D3) is commonly distributed in Asia [54–56]. Lastly, 2E is the subtype that predominates in Oceania [57]. Genotype 3 has been associated with some African and Asian populations, with subtype 3A being most prevalent in South, West, and Central Asia, scarcely in southern Europe [58–60], and almost the entirety of the African continent, except for the South, where subtype 3B predominates [54,58]. Likewise, genotype 6 is linked to African populations, is more commonly found in Central and Western Africa [58]. Genotype 7 is classified into subtypes 7A, 7B (encompassing 7B1 and 7B2), and 7C (further divided into 7C1 and 7C2) and is predominantly found in Asia, with subtype 7C also being found in Mauritius [54]. Genotype 8 is the most prevalent in the Western Pacific region, where both 8A and 8B dominate in Polynesia and Malesia, with the last one also being present in Papua New Guinea [57]. Lastly, genotype Eu-c appears to be only present in Northeast Siberia and Japan [52].

In contrast to HPyV1 data, for which most of the viral sequences obtained (73.33%) could be assigned to a single subtype (Ib1), a much more heterogeneous distribution of sequence subtypes was found when we considered HPyV2. Indeed, genotype I (subtypes 1A and 1B Europeans and North Americans), genotype 2 (subtypes 2A2 Eastern Asians and Native Americans and 2B Europeans and Eastern Asians), along with genotype 3 (subtype A South, West, and Central Asia and almost the entirety of the African continent), genotype 4 and a putatively new genotype 9 were also identified. The genetic singularity of the latter was not only suggested by their isolated position in the phylogenetic trees, but also by both NNN and PCOORD analyses. Once again, due to the heterogeneity of the population living in the area where this study was focused, finding a putative new HPyV2 genotype was not unexpected, and formally opens the possibility that other yet undescribed viral variants remain to be characterized.

Based on the analysis of a fragment composed of the large-T antigen coding region and VP1's coding region, HPyV5 has been divided into five genotypes, each with a strong geographical association: Europe/North America, Africa, Asia, Oceania, and South America [14,20]. Somewhat surprisingly, HPyV5 sequences comprised 35.62% of the total sequences analyzed in this study with most of them (65.38%) clustering among the so-called "Europe/North America" genotype. However, 30.77% of the HPyV5 sequences were also associated with the African genotype. Given the structure of the population living in/visiting Lisbon and its surroundings, the observed distribution was not unexpected. However, one sequence (LC636361) once again failed to cluster inside any of the proposed genetic clusters and radiated away from the African and Europe/North America, and was, therefore, assigned as a member of a broader Europe/North American/African "supercluster". Its genetic singularity was also confirmed by the NNN analysis (where it could not be associated with any of the groups of references), while it was considered a divergent member of the Europe/North America cluster by PCOORD analysis. Whether this sequence again indicates the existence of a new genetic lineage or its allocation to a specific genetic group in the phylogenetic tree, the result of a low phylogenetic signal of the dataset analyzed remains to be determined in future studies. In any case, additional analyses based on a larger number of samples from different locations most certainly will clarify our understanding of the geographic distribution of this group of polyomaviruses. The latter could also be improved with the analysis of larger, and more polymorphic sequence fragments, or better representative sequence datasets with high phylogenetic-associated signals.

Besides HPyV5, other less frequent HPyVs in waters samples include HPyV6. Due to the still low number of available HPyV6 sequences in databases, studies regarding the molecular epidemiology of this virus remain scarce. Even so, a couple of previous studies have reported the identification of two main clades when analyzing whole-genome sequences and/or partial sequences of HPyV6 [14]. The first group is comprised of HPyV6 sequences from various geographical locations (mostly from North America and Europe, but also including those from Asia and Oceania), while the second group mainly includes sequences of Asian origin. Furthermore, another clustering phenomenon appears to occur inside the first group, where HPyV6 sequences form subgroups according to their geographical origin [14]. However, due to the lack of information available in databases concerning the late region of the HPyV6 genome, we were unable to assess the occurrence of this distribution in this study. Curiously, our analysis revealed an a priori unexpectedly high number of sequences assigned to HPyV6, suggesting that their excretion route is either the feces or urine, with skin-peeling also being a valid hypothesis. Due to the low topological stability of all the phylogenetic trees of HPyV6 sequences, it was not possible to assign them to specific genetic types, although seven of them, obtained from samples collected in 2018 and 2020 at two different collection points, did share a common ancestry, and were assigned to a tree clade also including a sequence from Australia (Supplementary Figure S3). Once again, one sequence (LC636383) diverged from all others in the phylogenetic tree and was associated with different references by NNN and PCOORD. Therefore, as diverse HPyV6 seem to be frequent in LMA, subsequent analyses of these viruses will elucidate their genetic structure.

The use of wide-range viral detection protocols, especially when they explore multiplex amplification strategies, needs to consider the possibility of false-negative detection results, especially when viruses are present in low numbers. Indeed, and as our results have shown, the combination of multiple amplification primers in a single tube may reduce the efficiency of viral detection, especially if single-step (i.e., non-nested) PCR protocols are used. Most protocols reveal the pervasive presence of at least some types of HPyV (including HPyV1, HPyV2, and HPyV5) in wastewater samples where genetic analyses can disclose their genetic diversity, especially in areas where the sewage processing plants received the discharge of sewage effluents from millions of a heterogeneous population. Such is the case of LMA, which has been until recently a major tourist hub and gives residence to native as well as non-native Caucasians (especially from Brazilian and Eastern European ascent), Asian (especially Chinese), and African ancestry (most frequently from the former Portuguese colonies, that include Cape-Verde, Angola, Mozambique, and Guinea-Bissau). Therefore, not only does the detection of HPyV may indicate human waste contamination, the analysis of these viruses (through the study of their genomes) gives important information regarding their molecular epidemiological distribution patterns in the general population. Despite the possible limitations of studies such as the one here described, this study confirms that (i) environmental viral analysis is a powerful tool and that (ii) further methodology developments are still required to fully exploit this research ground.

4. Materials and Methods

4.1. Virus Concentration by Skimmed-Milk Flocculation and DNA Extraction

All the samples analyzed, corresponding to either solid particle-free wastewater (raw sewage influent) or environmental water samples were collected during October 2018, April 2019, July 2020, and October–November in 2020 in LMA (Figure 1). All samples were brought to the laboratory within <6 h of collection and were immediately stored at 4 °C and processed within 48 h.

Viral-like particles (VLP) were concentrated from 1 L of influent or 10 L of environmental water samples using organic flocculation with skimmed milk, essentially as previously described by Hjelmsø et al. [61] and Calgua et al. [62]. In brief, skimmed milk (Difco, Franklin Lakes, NJ, USA) and Paragon Scientific™ Synthetic Sea Water (Thermo Fisher Scientific, Waltham, MA, USA) were added to a final concentration of 1% (*w/v*) and 3.2%

(*w/w*), respectively. These mixtures were acidified to pH 3.5 with 1 M HCl, and then continuously mixed for 8 h at room temperature. The obtained flocculants sedimented overnight at room temperature, after which the supernatant was carefully removed with a vacuum pump (to avoid disturbing the sediment) and a residual volume of approximately 500 mL was centrifuged at $5500\times g$, for 45 min, at 4 °C. The supernatant was subsequently removed, and the obtained sediment was resuspended in 8 mL of 1/2 phosphate buffer (0.2 M Na₂HPO₄, 0.2 M NaH₂PO₄, pH 7.5).

All the samples analyzed were collected during October 2018, April 2019, July 2020, and October–November in 2020 in LMA (Figure 1). From each collection point, one sample was retrieved, corresponding to either solid particle-free wastewater (raw sewage influent) or environmental water samples, except for collection points 5 and 6, from which 3 and 2 samples were collected, respectively. Total DNA was extracted from 17 samples of VLP-enriched sediment using QIAamp[®] DNA Mini Kit (Qiagen, Germantown, MD, USA) and eluted in 100 mL of AE buffer.

4.2. Primer Design and Touch-Down Multiplex-PCR Set-Up

The compilation of the different nucleotide (nt) and amino acids (aa) sequences used in this work was based on the selection of structural protein-coding (VP1-2) and whole-genomic data available in GenBank on 31 April 2020. These were either directly identified via their accession numbers or indirectly, singled-out because of similarity searches using BLASTn or using NCBI-Virus (<https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/>, accessed on 31 April 2020).

Primers were designed using a combined approach that included visual inspection of multiple alignments of nt sequences performed using the iterative G-INS-I method as implemented in MAFFT vs. 7 [63], and Primer Design-M [64]. A total of 24 primers were designed (Table 1) to allow the amplification of 11 HPyV sequences in two parallel sets of two nested touch-down multiplexed-amplification PCR protocols. In one case, PCR-A was set up to allow the amplification of HPyV1, 2, 5, 8, 9, 10, 11 using 4 pairs of primers in both the 1st and 2nd PCR rounds. On the other hand, PCR-B targeted the amplification of HPyV3, 4, 6, and 7 using 2 pairs of primers in both the 1st and 2nd PCR rounds. The thermal profiles of the two touchdown amplification protocols were the same for the 1st and 2nd PCR rounds. For the first round of PCR-A, the cycling conditions were 3 min at 95 °C, 10 cycles of 30 s at 95 °C, 30 s at 55 °C (with a decrease of 1 °C per cycle), 1 min and 15 s at 72 °C, 30 cycles of 30 s at 95 °C, 30 s at 45 °C, 1 min and 15 s at 72 °C and a final extension step of 7 min at 72 °C. The same thermal profile was used for the second round, except for the annealing temperature starting with 56 °C and stabilizing at 46 °C after the first 10 cycles. The same applies to PCR-B, with the cycling conditions used in the amplification protocol for both rounds being the same, with the annealing temperature for the first round starting at 63 °C/59 °C, reaching 53 °C/49 °C after the first 10 cycles, for the first and second round, respectively.

The different primers were designed so as not to include more than three degenerate positions, a minimal length of 19 nt, and a T_m not below 45 °C. All the primers were designed to allow the amplification, using nested-PCR protocols, of approximately overlapping and similar-sized DNA fragments (from 881 to 1175 nt) in the 2nd round of amplification.

4.3. DNA Cloning and Sequencing

PCR amplicons were either purified from agarose gels using the Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, USA) and directly sequenced or cloned in a plasmid vector using either the CloneJET PCR Cloning Kit (Thermo Fisher Scientific, USA) or the NZY-A PCR Cloning Kit (NZYtech, Lisboa, Portugal), and NovaBlue (Novagen, Wisconsin, USA) competent cells, prepared essentially as described by Chung et al. [65]. Individual clones were selected, and their plasmid content was extracted (classical alkaline lysis miniprep) and analyzed. Recombinant plasmids containing inserts of exogenous DNA were sequenced using Sanger's method (STABVIDA, Caparica, Portugal). Depending on the

efficiency of clone recovery, a minimum of five and a maximum of 14 recombinant plasmid molecules were sequenced for each sample. In addition, a pooled sample containing 10 purified PCR fragments obtained using the PCR-A and PCR-B protocols from DNA extracted from six wastewater and one environmental sample, with a 5% spike of PhiX174 DNA, was analyzed by NGS. This was carried out using a DNA library of single indexing paired-end (150 bp \times 2) fragments prepared using the NEBNext Ultra II Library Prep Kit (NEB, Ipswich, MA, USA), and analyzed on an MiSeq Illumina sequencer.

4.4. Nucleotide Sequence Analyses

Sanger sequence analysis was carried out using Bioedit 7.2 [66] while NGS low-quality filtering, adapter trimming, contig assembly, and taxonomic identification were carried out using the Genome Detective software [67].

Five nucleotide sequence datasets were analyzed during this study. One of these datasets included representative ($n = 1$ to 6) nt sequences of reference HPyV genomes as well as 65 sequences obtained in this study with a minimum length of 650 nt. Considering the preliminary taxonomic identification results for the sequences obtained during this study (using BLASTn/MegaBlast option), four HPyV-specific datasets (HPyV1, HPyV2, HPyV5, and HPyV6) were prepared. For those viruses for which a complete genomic sequence was available, protein datasets (including non-structural and structural protein) were also assembled.

Multiple alignments of nt sequences were performed using the iterative G-INS-I method as implemented in MAFFT vs. 7 [63], followed by their edition using GBlocks [68], allowing for less strict flanking positions. The multiple sequence alignments were systematically verified to ensure the correct alignment of homologous codons using BioEdit 7.2 [66].

The evolutionary information contained in each aligned dataset (phylogenetic signal) was assessed by likelihood mapping [69] using TREE-PUZZLE v5.3 [70]. Phylogenetic reconstructions were performed based on the maximum likelihood (ML) optimization criterion using the GTR + Γ + I as the best data-fitting substitution model, as suggested by IQ-TREE [71]. The stability of the obtained tree topologies was assessed both by bootstrapping with 1000 re-samplings of the original sequence data, and with an approximate likelihood ratio test (aLRT) with 1000 data iterations, carried out using IQ-TREE running on an Ubuntu server.

On the other hand, phylogeographic analyses carried out using a Bayesian statistical framework took advantage of the BEAST v1.10 software package [72]. All these phylogenetic reconstructions were carried out assuming a relaxed uncorrelated lognormal molecular clock model [73] as indicated by the ML Clock Test implemented in MEGA X, allowing for the accommodation of among-lineage rate variation, the GTR + Γ + I model, and a smooth skyline demographic prior [74]. Two independent Markov chain Monte-Carlo (MCMC) runs were performed using BEAST v1.10 until 1×10^8 states were sampled, and 10% of which were discarded as burn-in. Chain convergence was verified using the Tracer software v1.7.1 (<http://beast.bio.ed.ac.uk/tracer>, accessed on 31 April 2020), which was also used to check for an adequate effective sample size higher than 200 after the removal of burn-in. The trees were logged on every 10,000th MCMC step, and the tree distribution was summarized using TreeAnnotator software v1.8.3 as a maximum clade credibility (MCC) tree, using median heights as the node heights in the tree. The FigTree v1.4.2 software was used to visualize the phylogenetic trees (<http://tree.bio.ed.ac.uk/software/figtree/>, accessed on 31 April 2020).

NeighborNet networks (NNn) were constructed using distance matrixes (corrected with the HKY model) using Splits Tree 4 software [75]. Principal coordinate analysis was performed using PCOORD (available at <https://www.hiv.lanl.gov/content/sequence/PCOORD/PCOORD.html>, accessed on 31 April 2020). Database sequence similarity searches were carried out with BLASTn (MegaBLAST option). The analyzing of putative genetic recombination events was carried out using RDP4 [76].

The nucleotide sequences obtained in the course of this work were deposited at the DDBJ database under accession numbers LC636333 to LC636405.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/pathogens10101309/s1>, Figure S1: Analysis of the phylogenetic relationships of 15 HPyV lineages (maximum likelihood) using the complete structural protein coding-sequence of references downloaded from the public genomic databases. Each sequence is identified by its accession number, HPyV type, country of origin, and, when possible, viral genotype and/or subtype. At the main tree branches ** indicates the support revealed by relevant aLRT and bootstrap values ($\geq 75\%$). The bar indicates the average number of substitutions per site. Figure S2: Analysis of the sensitivity/specificity of amplification using the PCR-A and B touchdown protocols. MWM indicates the NZYtech molecular weight marker VI (NZYtech, Portugal). Lanes 1 and 6: 1st and 2nd round negative controls for PCR-A; lanes 2 and 7: analysis of the 1st (lane 2) and 2nd round (lane 7) amplification reactions of PCR-A using a positive control [a HPyV2 DNA extract (diluted 1:10,000) was used in the 1st-round (lane 2) and a 1:50 dilution of the 1st-round product as input in the 2nd-round (lane 7)]; lanes 3 and 8: 1st (lane 3) and 2nd (lane 8) round of a singleplex reaction using primers HPyV1/2Fo and HPyV1/2Ro or HPyV1/2Fi and HPyV1/2Fi, for the 1st and 2nd round, respectively. The input DNA was the same as the one mentioned for the reactions in lanes 2 and 7; lanes 4 and 9: negative control (1st and 2nd-round, respectively) for PCR-B; lanes 5 and 10: analysis of 1st (lane 5) and 2nd round (lane 10) amplification reactions of PCR-B using as input for the 1st round a diluted (1:10,000) HPyV2 DNA extract and as 1:50 dilution of the product obtain in the 1st round as input in the 2nd round (lane 10). The “*” indicates the expected band with approximately 1100 bp (1st round) and 990 (2nd round). Figure S3: Phylogenetic analysis of HPyV1 (a), HPyV2 (b), HPyV5 (c), and HPyV6 (d) sequences (maximum likelihood) considering a structural protein-coding genome region. At specific branches, the identity of the HPyV is indicated by their accession number and country of origin. For the analysis of HPyV1, HPyV2 and HPyV5, the different viral genotypes/subtypes, as defined previously [14,20], are indicated. The sequences obtained during this study are indicated in boldface. In the latter, the numbers indicated in superscript refer to the identity of the sample collection sites, as they are indicated in Figure S1. The consensus sequence generated by NGS is indicated by the arrow. At the main tree branches, the number of “*” indicates the support revealed by the different phylogenetic reconstruction methods used, assuming as relevant aLRT and bootstrap values $\geq 75\%$, as well as posterior probability values ≥ 0.80 when a Bayesian approach was further used to confirm the proposed phylogenetic clustering. The bars indicate the average number of substitutions per site. Figure S4: NeighborNet networks (A) and PCOORD (B) analyses of HPyV1 (A1/B1), HPyV2 (A2/B2), HPyV5 (A3/B3), and HPyV6 (A4/B4) sequences. The sequences obtained in the course of this study are indicated by their accession number. In the PCOORD graphs, the two first dimensions accounts for 50.30% (B1), 47.02% (B2), 73.95% (B3), and 40.20% (B4) if the total nucleotide differences, while the first 10 axes cover 90.60% (B1), 78.78% (B2), 90.88% (B3), and 67.75% (B4) of the dataset sequence variation. Table S1: Human polyomaviruses classification, isolation origin, and associated human diseases. Table S2: Analysis of the performance of the primers used for amplification of HPyV sequences in a singleplex vs. multiplex format.

Author Contributions: Conceptualization, M.N. and R.P.; water collection and sample processing, A.C.C., M.N. and R.P.; funding acquisition, M.N. and R.P. molecular cloning and bioinformatic analyses, A.C.C. and R.P.; next-generation sequencing analysis, A.F.-S., I.L. and M.N.; manuscript writing and correction, A.C.C., A.F.-S., I.L., M.N. and R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work received financial support from the Global Health and Tropical Medicine Center (which is funded through FCT contract UID/Multi/04413/2013), as well as from the AgriWWater Project (PTDC/CTA-AMB/29586/2017), and iNOVA4Health (UID/Multi/04462/2019).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article or Supplementary Material.

Acknowledgments: The authors would like to thank Águas de Portugal for granting us access to the plant-grounds, and their help in collecting raw sewage. We would also like to thank Ana Miguel

Matos (Laboratório Environment, Re-action, Separation and Thermodynamics da Faculdade de Farmácia da Universidade de Coimbra) for providing a HPyV2 DNA extract.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. John, R.; Buck, C.B.; Allander, T.; Atwood, W.J.; Garcea, R.L.; Imperiale, M.J.; Major, E.O.; Ramqvist, T.; Norkin, L.C. Taxonomical developments in the family *Polyomaviridae*. *Arch. Virol.* **2011**, *156*, 1627–1634. [[CrossRef](#)]
2. Moens, U.; Calvignac-Spencer, S.; Lauber, C.; Ramqvist, T.; Feltkamp, M.C.; Daugherty, M.D.; Verschoor, E.J.; Ehlers, B. Report on the Taxonomy of the Polyomaviridae. Virus Taxonomy: 2020 Release. 2020. Available online: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5656788/> (accessed on 31 April 2020).
3. Calvignac-Spencer, S.; Feltkamp, M.C.W.; Daugherty, M.D.; Moens, U.; Ramqvist, T.; John, R.; Ehlers, B. A taxonomy update for the family *Polyomaviridae*. *Arch. Virol.* **2016**, *161*, 1739–1750. [[CrossRef](#)] [[PubMed](#)]
4. Kean, J.M.; Rao, S.; Wang, M.; Garcea, R.L. Seroepidemiology of human polyomaviruses. *PLoS Pathog.* **2009**, *5*, e1000363. [[CrossRef](#)]
5. Van Der Meijden, E.; Bialasiewicz, S.; Rockett, R.J.; Tozer, S.J.; Sloots, T.P.; Feltkamp, M.C.W. Different serologic behavior of MCPyV, TSPyV, HPyV6, HPyV7 and HPyV9 polyomaviruses found on the skin. *PLoS ONE* **2013**, *8*, e81078. [[CrossRef](#)]
6. Rinaldo, C.H.; Tylden, G.D.; Sharma, B.N. The human polyomavirus BK (BKPv): Virological background and clinical implications. *Apmis* **2013**, *121*, 728–745. [[CrossRef](#)]
7. Wiedinger, K.; Bitsakis, C.; Chang, S. Reactivation of human polyomaviruses in immunocompromised states. *J. Neurovirol.* **2014**, *20*, 1–8. [[CrossRef](#)] [[PubMed](#)]
8. Dalianis, T.; Hirsch, H.H. Human polyomaviruses in disease and cancer. *Virology* **2013**, *437*, 63–72. [[CrossRef](#)] [[PubMed](#)]
9. Gardner, S.D.; Field, A.M.; Coleman, D.V. New human papovavirus (B.K.) isolated from urine after renal transplantation. *Lancet* **1970**, *297*, 1253–1257. [[CrossRef](#)]
10. Fiksdal, L.; Maki, J.S.; LaCroix, S.J.; Staley, J.T. Survival and detection of *Bacteroides* spp., prospective indicator bacteria. *Appl. Environ. Microbiol.* **1985**, *49*, 148–150. [[CrossRef](#)] [[PubMed](#)]
11. Harwood, V.J.; Levine, A.D.; Scott, T.M.; Chivukula, V.; Lukasik, J.; Farrah, S.R.; Rose, J.B. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. *Appl. Environ. Microbiol.* **2005**, *71*, 3163–3170. [[CrossRef](#)]
12. McQuaig, S.; Griffith, J.; Harwood, V.J. Association of fecal indicator bacteria with human viruses and microbial source tracking markers at coastal beaches impacted by nonpoint source pollution. *Appl. Environ. Microbiol.* **2012**, *78*, 6423–6432. [[CrossRef](#)]
13. Rachmadi, A.T.; Torrey, J.R.; Kitajima, M. Human polyomavirus: Advantages and limitations as a human-specific viral marker in aquatic environments. *Water Res.* **2016**, *105*, 456–469. [[CrossRef](#)] [[PubMed](#)]
14. Torres, C. Evolution and molecular epidemiology of polyomaviruses. *Infect. Genet. Evol.* **2020**, *79*, 104150. [[CrossRef](#)] [[PubMed](#)]
15. Ondov, B.D.; Starrett, G.J.; Sappington, A.; Kostic, A.; Koren, S.; Buck, C.B.; Phillippy, A.M. Mash Screen: High-throughput sequence containment estimation for genome discovery. *Genome Biol.* **2019**, *20*, 232. [[CrossRef](#)] [[PubMed](#)]
16. Korup, S.; Rietscher, J.; Calvignac-Spencer, S.; Trusch, F.; Hofmann, J.; Moens, U.; Sauer, I.; Voigt, S.; Schmuck, R.; Ehlers, B. Identification of a novel human polyomavirus in organs of the gastrointestinal tract. *PLoS ONE* **2013**, *8*, e58021. [[CrossRef](#)]
17. Fahsbender, E.; Altan, E.; Estrada, M.; Seguin, M.A.; Young, P.; Leutenegger, C.M.; Delwart, E. Lyon-IARC polyomavirus DNA in feces of diarrheic cats. *Microbiol. Resour. Announc.* **2019**, *8*, 54–56. [[CrossRef](#)]
18. Prezioso, C.; Van Ghelue, M.; Pietropaolo, V.; Moens, U. Detection of quebec polyomavirus DNA in samples from different patient groups. *Microorganisms* **2021**, *9*, 1082. [[CrossRef](#)]
19. Korbie, D.J.; Mattick, J.S. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nat. Protoc.* **2008**, *3*, 1452–1456. [[CrossRef](#)]
20. Martel-Jantin, C.; Filippone, C.; Tortevoeye, P.; Afonso, P.V.; Betsem, E.; Descorps-Declere, S.; Nicol, J.T.J.; Touzé, A.; Coursaget, P.; Crouzat, M.; et al. Molecular epidemiology of merkel cell polyomavirus: Evidence for geographically related variant genotypes. *J. Clin. Microbiol.* **2014**, *52*, 1687–1690. [[CrossRef](#)]
21. Jin, L.; Gibson, P.E.; Booth, J.C.; Clewley, J.P. Genomic typing of BK virus in clinical specimens by direct sequencing of polymerase chain reaction products. *J. Med. Virol.* **1993**, *41*, 11–17. [[CrossRef](#)]
22. Morel, V.; Martin, E.; François, C.; Helle, F.; Faucher, J.; Mourez, T.; Choukroun, G.; Duverlie, G.; Castelain, S.; Brochot, E. A simple and reliable strategy for BK virus subtyping and subgrouping. *J. Clin. Microbiol.* **2017**, *55*, 1177–1185. [[CrossRef](#)] [[PubMed](#)]
23. Auit, G.S.; Stoner, G.L. Two major types of JC virus defined in progressive multifocal leukoencephalopathy brain by early and late coding region DNA sequences. *J. Gen. Virol.* **1992**, 2669–2678.
24. Agostini, H.T.; Yanagihara, R.; Davis, V.; Ryschkewitsch, C.F.; Stoner, G.L. Asian genotypes of JC virus in Native Americans and in a Pacific Island population: Markers of viral evolution and human migration. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 14542–14546. [[CrossRef](#)] [[PubMed](#)]
25. Baez, C.F.; Diaz, N.C.; Venceslau, M.T.; Luz, F.B.; Guimarães, M.A.A.M.; Zalis, M.G.; Varella, R.B. Phylogenetic and structural analysis of merkel cell polyomavirus VP1 in Brazilian samples. *Virus Res.* **2016**, *221*, 1–7. [[CrossRef](#)] [[PubMed](#)]
26. Iaconelli, M.; Petricca, S.; Libera, S.D.; Di Bonito, P.; La Rosa, G. First detection of human papillomaviruses and human polyomaviruses in river waters in Italy. *Food Environ. Virol.* **2015**, *7*, 309–315. [[CrossRef](#)]

27. Kokkinos, P.; Ziros, P.; Meri, D.; Filippidou, S.; Kolla, S.; Galanis, A.; Vantarakis, A. Environmental surveillance. An additional/alternative approach for virological surveillance in Greece? *Int. J. Environ. Res. Public Health* **2011**, *8*, 1914–1922. [[CrossRef](#)]
28. Kirs, M.; Caffaro-filho, R.A.; Wong, M.; Harwood, V.J.; Moravcik, P.; Fujioka, S. Human-associated *Bacteroides* spp. and human polyomaviruses as microbial source tracking markers in Hawaii. *Appl. Environ. Microbiol.* **2017**, *82*, 6757–6767. [[CrossRef](#)] [[PubMed](#)]
29. Calgua, B.; Fumian, T.; Rusiñol, M.; Rodriguez-Manzano, J.; Mbayed, V.A.; Bofill-Mas, S.; Miagostovich, M.; Girones, R. Detection and quantification of classic and emerging viruses by skimmed-milk flocculation and PCR in river water from two geographical areas. *Water Res.* **2013**, *47*, 2797–2810. [[CrossRef](#)]
30. Fumian, T.M.; Guimarães, F.R.; Vaz, B.J.P.; Da Silva, M.T.T.; Muylaert, F.F.; Bofill-Mas, S.; Gironés, R.; Leite, J.P.G.; Miagostovich, M.P. Molecular detection, quantification and characterization of human polyomavirus JC from waste water in Rio de Janeiro, Brazil. *J. Water Health* **2010**, *8*, 438–445. [[CrossRef](#)]
31. Rusiñol, M.; Fernandez-Cassi, X.; Hundesa, A.; Vieira, C.; Kern, A.; Eriksson, I.; Ziros, P.; Kay, D.; Miagostovich, M.; Vargha, M.; et al. Application of human and animal viral microbial source tracking tools in fresh and marine waters from five different geographical areas. *Water Res.* **2014**, *59*, 119–129. [[CrossRef](#)]
32. McQuaig, S.M.; Scott, T.M.; Lukasik, J.O.; Paul, J.H.; Harwood, V.J. Quantification of human polyomaviruses JC virus and BK Virus by TaqMan quantitative PCR and comparison to other water quality indicators in water and fecal samples. *Appl. Environ. Microbiol.* **2009**, *75*, 3379–3388. [[CrossRef](#)]
33. Torres, C.; Barrios, M.E.; Cammarata, R.V.; Cisterna, D.M.; Estrada, T.; Martini Novas, S.; Cahn, P.; Blanco Fernández, M.D.; Mbayed, V.A. High diversity of human polyomaviruses in environmental and clinical samples in Argentina: Detection of JC, BK, Merkel-cell, Malawi, and human 6 and 7 polyomaviruses. *Sci. Total Environ.* **2016**, *542*, 192–202. [[CrossRef](#)]
34. Harwood, V.J.; Staley, C.; Badgley, B.D.; Borges, K.; Korajkic, A. Microbial source tracking markers for detection of fecal contamination in environmental waters: Relationships between pathogens and human health outcomes. *FEMS Microbiol. Rev.* **2014**, *38*, 1–40. [[CrossRef](#)] [[PubMed](#)]
35. Rochelle-Newall, E.; Nguyen, T.M.H.; Le, T.P.Q.; Sengtaheuanghoung, O.; Ribolzi, O. A short review of fecal indicator bacteria in tropical aquatic ecosystems: Knowledge gaps and future directions. *Front. Microbiol.* **2015**, *6*, 308. [[CrossRef](#)] [[PubMed](#)]
36. McQuaig, S.M.; Scott, T.M.; Harwood, V.J.; Farrah, S.R.; Lukasik, J.O. Detection of human-derived fecal pollution in environmental waters by use of a PCR-based human polyomavirus assay. *Appl. Environ. Microbiol.* **2006**, *72*, 7567–7574. [[CrossRef](#)] [[PubMed](#)]
37. Bofill-Mas, S.; Pina, S.; Girones, R. Documenting the epidemiologic patterns of polyomaviruses in human populations by studying their presence in urban sewage. *Appl. Environ. Microbiol.* **2000**, *66*, 238–245. [[CrossRef](#)]
38. Ahmed, W.; Wan, C.; Goonetilleke, A.; Gardner, T. Evaluating sewage-associated JCV and BKV polyomaviruses for sourcing human fecal pollution in a coastal river in Southeast Queensland, Australia. *J. Environ. Qual.* **2010**, *39*, 1743–1750. [[CrossRef](#)]
39. Jurzik, L.; Hamza, I.A.; Wilhelm, M. Investigating the Reduction of Human Adenovirus (HAdV) and Human Polyomavirus (HPyV) in a Sewage Treatment Plant with a Polishing Pond as a Tertiary Treatment. *Water. Air. Soil Pollut.* **2015**, *226*, 284. [[CrossRef](#)]
40. Hewitt, J.; Greening, G.E.; Leonard, M.; Lewis, G.D. Evaluation of human adenovirus and human polyomavirus as indicators of human sewage contamination in the aquatic environment. *Water Res.* **2013**, *47*, 6750–6761. [[CrossRef](#)] [[PubMed](#)]
41. Bofill-Mas, S.; Albinana-Gimenez, N.; Clemente-Casares, P.; Hundesa, A.; Rodriguez-Manzano, J.; Allard, A.; Calvo, M.; Girones, R. Quantification and stability of human adenoviruses and polyomavirus JCPyV in wastewater matrices. *Appl. Environ. Microbiol.* **2006**, *72*, 7894–7896. [[CrossRef](#)]
42. Cantalupo, P.G.; Calgua, B.; Zhao, G.; Hundesa, A.; Wier, A.D.; Katz, J.P.; Grabe, M.; Hendrix, R.W.; Girones, R.; Wang, D.; et al. Raw sewage harbors diverse viral populations. *MBio* **2011**, *2*, e00180-11. [[CrossRef](#)]
43. Torres, C.; Barrios, M.E.; Cammarata, R.V.; Victoria, M.; Fernandez-Cassi, X.; Bofill-Mas, S.; Colina, R.; Blanco Fernández, M.D.; Mbayed, V.A. Phylogenetics of Merkel-cell polyomavirus and human polyomavirus 6: A long-term history with humans. *Mol. Phylogenet. Evol.* **2018**, *126*, 210–220. [[CrossRef](#)] [[PubMed](#)]
44. Rockett, R.J.; Sloots, T.P.; Bowes, S.; O'Neill, N.; Ye, S.; Robson, J.; Whiley, D.M.; Lambert, S.B.; Wang, D.; Nissen, M.D.; et al. Detection of Novel Polyomaviruses, TSPyV, HPyV6, HPyV7, HPyV9 and MWPyV in Feces, Urine, Blood, Respiratory Swabs and Cerebrospinal Fluid. *PLoS ONE* **2013**, *8*, e62764. [[CrossRef](#)] [[PubMed](#)]
45. Bofill-Mas, S.; Rodriguez-Manzano, J.; Calgua, B.; Carratala, A.; Girones, R. Newly described human polyomaviruses Merkel Cell, KI and WU are present in urban sewage and may represent potential environmental contaminants. *Virol. J.* **2010**, *7*, 141. [[CrossRef](#)]
46. Babakir-Mina, M.; Ciccozzi, M.; Alteri, C.; Polchi, P.; Picardi, A.; Greco, F.; Lucarelli, G.; Arcese, W.; Perno, C.F.; Perno, M. Excretion of the novel polyomaviruses KI and WU in the stool of patients with hematological disorders. *Antivir. Ther.* **2006**, *55*, 52–55. [[CrossRef](#)]
47. Siebrasse, E.A.; Bauer, I.; Holtz, L.R.; Le, B.; Lassa-claxton, S.; Canter, C.; Hmiel, P.; Shenoy, S.; Sweet, S.; Turmelle, Y.; et al. Human polyomaviruses in children undergoing transplantation, United States, 2008–2010. *Emerg. Infect. Dis.* **2012**, *18*, 1676–1679. [[CrossRef](#)]
48. Pinheiro, M.S.; Mendes, G.S.; Santos, N. Human polyomaviruses 10 and 11 in faecal samples from Brazilian children. *Braz. J. Microbiol.* **2020**, *51*, 585–591. [[CrossRef](#)]

49. Zheng, H.Y.; Nishimoto, Y.; Chen, Q.; Hasegawa, M.; Zhong, S.; Ikegaya, H.; Ohno, N.; Sugimoto, C.; Takasaka, T.; Kitamura, T.; et al. Relationships between BK virus lineages and human populations. *Microbes Infect.* **2007**, *9*, 204–213. [[CrossRef](#)] [[PubMed](#)]
50. Nishimoto, Y.; Zheng, H.Y.; Zhong, S.; Ikegaya, H.; Chen, Q.; Sugimoto, C.; Kitamura, T.; Yogo, Y. An Asian origin for subtype IV BK virus based on phylogenetic analysis. *J. Mol. Evol.* **2007**, *65*, 103–111. [[CrossRef](#)]
51. Sugimoto, C.; Kitamura, T.; Guo, J.; Al-ahdal, M.N.; Shchelkunov, S.N.; Otova, B.; Ondrejka, P.; Chollet, J.-Y.; El-Safi, S.; Ettayebi, M.; et al. Typing of urinary JC virus DNA offers a novel means of tracing human migrations. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 9191–9196. [[CrossRef](#)]
52. Yogo, Y.; Sugimoto, C.; Zheng, H.Y.; Ikegaya, H.; Takasaka, T.; Kitamura, T. JC virus genotyping offers a new paradigm in the study of human populations. *Rev. Med. Virol.* **2004**, *14*, 179–191. [[CrossRef](#)] [[PubMed](#)]
53. Stoner, G.L.; Jobes, D.V.; Fernandez, M.; Agostini, H.T.; Chima, S.C.; Ryschkewitsch, C.F. JC virus as a marker of human migration to the Americas. *Microbes Infect.* **2000**, *2*, 1905–1911. [[CrossRef](#)]
54. Saruwatari, L.; Sugimoto, C.; Kitamura, T.; Ohno, N.; Sakai, E.; Shresta, P.; Hoa, B.K.; Phi, P.T.P.; An, H.P.H.; Tuyet, N.T.A.; et al. Asian domains of four major genotypes of JC virus, Af2, B1-b, CY and SC. *Arch. Virol.* **2002**, *147*, 1–10. [[CrossRef](#)] [[PubMed](#)]
55. Agostini, T.; Shishido-hara, Y.; Baumhefner, R.W.; Singer, E.J.; Ryschkewitsch, C.F.; Stoner, G.L. JC virus Type 2: Definition of subtypes based on DNA sequence analysis of ten complete genomes. *J. Gen. Virol.* **1998**, *79*, 1143–1151. [[CrossRef](#)]
56. Cui, X.; Wang, J.C.; Deckhut, A.; Joseph, B.C.; Eberwein, P.; Cubitt, C.L.; Ryschkewitsch, C.F.; Agostini, H.T.; Stoner, G.L. Chinese Strains (Type 7) of JC Virus Are Afro-Asiatic in Origin But Are Phylogenetically Distinct from the Mongolian and Indian Strains (Type 2D) and the Korean and Japanese Strains (Type 2A)*. *J. Mol. Evol.* **2004**, *58*, 568–583. [[CrossRef](#)] [[PubMed](#)]
57. Jobes, D.V.; Friedlaender, J.S.; Mgone, C.S.; Agostini, H.T.; Koki, G. 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.* **2001**, *146*, 2097–2113. [[CrossRef](#)] [[PubMed](#)]
58. Takasaka, T.; Kitamura, T.; Sugimoto, C.; Guo, J.; Zheng, H.Y.; Yogo, Y. Phylogenetic analysis of major African genotype (Af2) of JC virus: Implications for origin and dispersals of modern Africans. *Am. J. Phys. Anthropol.* **2006**, *129*, 465–472. [[CrossRef](#)]
59. Agostini, H.T.; Deckhut, A.; Jobes, D.V.; Girones, R.; Schlunck, G.; Prost, M.G.; Frias, C.; Pérez-Trallero, E.; Ryschkewitsch, C.F.; Stoner, G.L. Genotypes of JC virus in East, Central and Southwest Europe. *J. Gen. Virol.* **2001**, *82*, 1221–1331. [[CrossRef](#)]
60. Pagani, E.; Delbue, S.; Mancuso, R.; Borghi, E.; Tarantini, L.; Ferrante, P. Molecular analysis of JC virus genotypes circulating among the Italian healthy population. *J. Neurovirol.* **2003**, *9*, 559–566. [[CrossRef](#)]
61. Hjelmsø, M.H.; Hellmér, M.; Fernandez-Cassi, X.; Timoneda, N.; Lukjancenka, O.; Seidel, M.; Elsässer, D.; Aarestrup, F.M.; Löfström, C.; Bofill-Mas, S.; et al. Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. *PLoS ONE* **2017**, *12*, e0170199. [[CrossRef](#)]
62. Calgua, B.; Mengewein, A.; Grunert, A.; Bofill-Mas, S.; Clemente-Casares, P.; Hundesa, A.; Wyn-Jones, A.P.; López-Pila, J.M.; Girones, R. Development and application of a one-step low cost procedure to concentrate viruses from seawater samples. *J. Virol. Methods* **2008**, *153*, 79–83. [[CrossRef](#)] [[PubMed](#)]
63. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)] [[PubMed](#)]
64. Yoon, H.; Leitner, T. PrimerDesign-M: A multiple-alignment based multiple-primer design tool for walking across variable genomes. *Bioinformatics* **2015**, *31*, 1472–1474. [[CrossRef](#)]
65. Chung, C.T.; Niemela, S.L.; Miller, R.H. One-step preparation of competent *Escherichia coli*: Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 2172–2175. [[CrossRef](#)] [[PubMed](#)]
66. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
67. Vilsker, M.; Moosa, Y.; Nooij, S.; Fonseca, V.; Ghysens, Y.; Dumon, K.; Pauwels, R.; Alcántara, L.C.; Vanden Eynden, E.; Vandamme, A.M.; et al. Genome Detective: An automated system for virus identification from high-throughput sequencing data. *Bioinformatics* **2019**, *35*, 871–873. [[CrossRef](#)]
68. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [[CrossRef](#)]
69. Strimmer, K.; von Haeseler, A. Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 6815–6819. [[CrossRef](#)]
70. Schmidt, H.A.; Strimmer, K.; Vingron, M.; von Haeseler, A. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **2002**, *18*, 502–504. [[CrossRef](#)]
71. Trifinopoulos, J.; Nguyen, L.; Von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, 232–235. [[CrossRef](#)]
72. Suchard, M.A.; Lemey, P.; Baele, G.; Ayres, D.L.; Drummond, A.J.; Rambaut, A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **2018**, *4*, vey016. [[CrossRef](#)] [[PubMed](#)]
73. Ho, S.Y.W.; Phillips, M.J.; Drummond, A.J.; Cooper, A. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Mol. Biol. Evol.* **2005**, *22*, 1355–1363. [[CrossRef](#)]
74. Drummond, A.J.; Rambaut, a.; Shapiro, B.; Pybus, O.G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **2005**, *22*, 1185–1192. [[CrossRef](#)] [[PubMed](#)]

-
75. Huson, D.H.; Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **2006**, *23*, 254–267. [[CrossRef](#)] [[PubMed](#)]
 76. Martin, D.P.; Murrell, B.; Golden, M.; Khoosal, A.; Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **2015**, *1*, vev003. [[CrossRef](#)]