BRIEF REPORT



Hepeliviruses in two waterbodies in Berlin, Germany

Roland Zell¹ · Marco Groth² · Lukas Selinka¹ · Hans-Christoph Selinka³

Received: 26 October 2022 / Accepted: 10 December 2022 / Published online: 25 December 2022 © The Author(s) 2022

Abstract

The order *Hepelivirales* comprises RNA viruses of four families (*Alphatetraviridae, Benyviridae, Hepeviridae,* and *Matonaviridae*). Sequencing of virus genomes from water samples from the Havel River and the Teltow Canal (Teltowkanal) in Berlin, Germany, revealed 25 almost complete and 68 partial genomes of viruses presumably belonging to the order *Hepelivirales*. Only one of these viruses exhibited a relationship to a known member of this order. The members of one virus clade have a polymerase with a permuted order of the conserved palm subdomain motifs resembling the polymerases of permutotetraviruses and birnaviruses. Overall, our study further demonstrates the diversity of hepeliviruses and indicates the enzootic prevalence of hepeliviruses in unknown hosts.

The development of next-generation sequencing techniques has enabled analysis of unculturable viruses from sources as diverse as water, soil, and samples of prokaryotic and eukarvotic organisms. As a result, sequence data for thousands of viruses have been obtained, but an immense amount of "viral dark matter" [14] remains to be investigated. Among the many virus groups that are now better understood because of the increase in sequence data are hepeviruses and related viruses of the order Hepelivirales. The present classification of the higher-order taxa of RNA viruses is based essentially on the similarity of the RNA-dependent RNA polymerase (RdRp), the only universal gene of all RNA viruses. According to Wolf and coworkers, the global RdRp phylogenetic tree consists of five major branches, with the members of the order Hepelivirales belonging to the "alphavirus supergroup" of branch 3 [19]. Except for the similarity of their RdRp, the members of the *Hepelivirales* share few commonalities regarding genome structure, capsid symmetry, or host range:

Handling Editor: Tim Skern.

Roland Zell roland.zell@med.uni-jena.de

- ¹ Section of Experimental Virology, Institute for Medical Microbiology, Jena University Hospital, Friedrich Schiller University, Jena, Germany
- ² CF DNA Sequencing, Leibniz Institute on Aging, Fritz Lipmann Institute, Jena, Germany
- ³ Section II 1.4 Microbiological Risks, Department of Environmental Hygiene, German Environment Agency, Berlin, Germany

Mono-, bi-, and multipartite viruses have been described. The lengths of genomic RNAs (gRNAs) vary from 6.7 kb to 9.8 kb for viruses with monopartite genomes and up to 16.8 kb in total for the multipartite benyviruses. gRNAs may be polyadenylated (Benyviridae, Hepeviridae, Matonaviridae) or not (Alphatetraviridae) and encode a large nonstructural polyprotein at their 5' end with an N-terminal N7-methyltransferase, a central helicase, and a C-terminal 'alpha-like' RdRp. A capsid protein (CP) is encoded either by a subgenomic mRNA (sgRNA) or by a second gRNA molecule (Benyviridae and Helicoverpa armigera stunt virus of the family Alphatetraviridae). Particles are either icosahedral, with a T = 3 (*Hepeviridae*) or T = 4 lattice (*Alphatetraviri*dae, Matonaviridae), or rod-like with helical nucleocapsids (Benyviridae). The CP of icosahedral viruses has a jellyroll fold. Only members of the family Matonaviridae have enveloped virions. The wide host range includes vertebrates (Hepeviridae, Matonaviridae), lepidopteran insects (Alphatetraviridae), and plants (Benyviridae).

In addition to the classified members of the order *Hepe-livirales*, a great number of distantly related, unclassified hepe-like viruses (hepeliviruses) have been described. Among these viruses are the bastroviruses from humans, other mammals, and fish [1, 10, 17] as well as so-called 'bastro-like' viruses and many other hepe-like viruses detected in mammals, lower vertebrates, invertebrates, plants, sewage, and environmental water samples, e.g., [4, 5, 12, 13, 15–18, 20]. The genomes of these viruses vary from 5 to 12.5 kb in size and contain 2-8 open reading frames (ORFs)

indicating that the 'hepe-like' viruses likely comprise several new virus families.

Considering the information given above on viruses in the order *Hepelivirales* and their reservoirs, especially in sewage and environmental water samples, the aim of this project was to analyse the virome of two waterbodies in Berlin, Germany, to reveal the spectrum of viruses and to characterise the viruses detected by phylogenetic analysis. The present study describes 25 almost complete and 68 partial hepelivirus genomes. These viruses were designated Havel hepe-like virus (HHLV) 1 to 38 and Teltowkanal hepe-like virus (TkHLV) 1 to 46. Further, we describe the partial genomes of six astro-like viruses, named Havel astrolike virus (HALV) 1 to 3, and Teltowkanal astro-like virus (TkALV) 1 to 3. None of the viruses belong to a known family of the order Hepelivirales or Stellavirales except for TkHLV-14. Whereas TkHLV-14 has the P70-like CP of omegatetraviruses, the genome sequences of the remaining viruses display similarity to hepe-like and astro-like viruses from crustacea, aquatic arthropods, and unspecified lophotrochozoa, which is compatible with their detection in environmental water samples.

The methods of sample collection, virus enrichment, RNA preparation, and sequencing have been described previously [21]. Briefly, 50-L freshwater samples were collected from the Teltow Canal (Teltowkanal) and the Havel River in Berlin, Germany (for sampling dates and sampling site coordinates, see Table 1). Virus particles were concentrated by glass wool filtration and eluted from

R. Zell	et	al	
---------	----	----	--

the column with buffer containing 3% beef extract and 50 mM glycine, pH 9.5. After adjusting its pH to 7, the eluate (180 ml) was filtered (0.45 µm). Virus particles were sedimented by ultracentrifugation $(100,000 \times g \text{ for } 2.5 \text{ h})$ at 4°C). Single-end and paired-end sequencing, respectively, was done using a HiSeq 2500 System (Table 1). Sequence data were extracted in FastQ format employing the bcl2fastq tool of Illumina, followed by adapter and quality trimming with Cutadapt (parameters: -q 10 -m 30 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAG TGT; -A only with PE data) and removal of amplification duplicons. Assembly was performed with the clc_assembler (parameters: -p fb ss 0850) and metaSPAdes using standard parameters (only for PE data; for parameters and detailed information, see Table 1). A protein database created with all NCBI GenBank entries for the Taxonomy ID 10239 (search term "viruses[organism]") was screened with DIAMOND [3] and BLAST+ v2.6.0 to identify hepelivirus candidates. The final genome sequences were curated by manual assembly of overlapping contigs if necessary. Sequence alignments were made in MEGA version X [8] and adjusted manually. Codon-adjusted nucleotide sequence alignments were used for maximum-likelihood tree inference with IQ-TREE 2.1.3 for Windows [9]. Optimal substitution models were selected using ModelFinder, implemented in IQ-TREE. Branch support was assessed with 50,000 ultrafast bootstrap replications using UFBoot2, also implemented in IQ-TREE [7].

Sample ID	MR137-16	20161205	MR233-17	MR644-18		
Waterbody	Teltow Canal	Teltow Canal	Teltow Canal	Havel River		
Sampling site	Berlin, Bäkebrücke	Berlin, Bäkebrücke	Berlin, Bäkebrücke	Berlin, Heerstraße		
Sampling site coordinates	52°26′03″N 13°18′57″E	52°26′03″N 13°18′57″E	52°26′03″N 13°18′57″E	52°30′46″N 13°12′14″E		
Sampling date	13 July 2016	17 Aug 2016	18 July 2017	28 June 2018		
Sample size	50 L	50 L	50 L	50 L		
RNA preparation kit	NucliSENS miniMAG (Biomérieux)	NucliSENS miniMAG (Biomérieux)	QIAamp Viral RNA mini kit (QIAGEN)	QIAamp Viral RNA mini kit (QIAGEN)		
Sequencing method	HiSeq 2500 50 bp, SE , HT	HiSeq 2500 100 bp, SE , RR	HiSeq 2500 150 bp, PE , RR	HiSeq 2500 150 bp, PE , RR		
Reads	62,220,180*	55,164,912*	70,018,635 [§]	51,902,006 [§]		
Assembly method	clc/-	clc/-	clc/metaSPAdes	clc/metaSPAdes		
Contigs/Scaffolds (>200 nt)	7,952/-	178,946/-	537,529/1,314,849	162,082/388,367		
Contigs assigned by DIAMON	ID to					
Viruses	867	17,816	41,408/66,371	8,810/16,922		
Hepelivirales	7	18	91/98	34/50		
Stellavirales	7	26	113/159	76/96		

 Table 1
 Sample information

Abbreviations: SE, single-end sequencing; PE, paired-end sequencing; HT, HiSeq 2500 high-throughput mode; RR, HiSeq 2500 Rapid Run mode; clc, clc assembler v5.2.1; metaSPAdes, metaSPAdes assembler v3.15.3 (Python: v3.9.11)

*SE reads after adapter trimming

[§]PE read pairs after adapter trimming and removal of duplicons

Illumina sequencing results varied depending on the yields of virus enrichment and the sequencing method (Table 1). DIAMOND analysis suggested the presence of many scaffolds/contigs belonging to the orders Hepelivirales and Stellavirales. However, the vast majority of scaffolds/ contigs with hepe-like sequences were misassigned to the order Stellavirales, and close to 100% of all assignments to lower ranks (family, genus, species) could not be verified by BLAST. Using BLAST, 93 partial or full-length genome sequences of hepe-like viruses were identified, with sizes ranging from 970 nt to 9268 nt, plus six partial astro-like virus genome sequences with lengths from 539 nt to 6815 nt (GenBank accession nos. OP699055 to OP699153; Supplementary Table S1). The genome layout of 21 (almost) complete hepe-like virus genomes is presented in Supplementary Figure S1. Interestingly, many viruses were present in two or more samples with almost identical sequences, indicating that a very similar hepeli virome is present in both waterbodies and in three consecutive years (Supplementary Table S1). Most of the contigs smaller 1 kb were not further characterized.

As a large nonstructural polyprotein with methyltransferase, helicase, and RdRp domains is a common feature of hepeliviruses, phylogenetic analysis of these three domains was conducted. The sequences of 57 of the 93 novel hepeliviruses were suitable for these analyses and included. The sequence alignments contained reference sequences of members of the four *Hepelivirales* families, bastroviruses, and numerous unclassified hepe-like viruses. In addition, the RdRp tree contained representative sequences of mamastroviruses, avastroviruses, and unclassified astro-like viruses as well as the HALV and TkALV sequences (Fig. 1 and Supplementary Fig. S2).

The tree topologies in the three phylogenetic analyses demonstrate several robust clades in addition to the families Alphatetraviridae, Benyviridae, Hepeviridae, Matonaviridae, and Astroviridae. The first clade includes human and animal bastroviruses, which always clustered in two subclades. The second clade, the so-called 'bastro-like' viruses, which include HHLV-6, -10, -15, and -16 and TkHLV-28, -29, -30, -31, and -38 from our water samples, were always located apart from the bastroviruses but close to the alphatetraviruses (Fig. 1). The genomes of bastro-like viruses contain two ORFs. In addition to the phylogenetic results and the fact that they have a different host, which is likely a non-vertebrate, 'bastro-like' viruses are distinguished from the human and animal bastroviruses by the presence of a non-homologous CP. This is of particular interest, as the CP of human and animal bastroviruses exhibits a striking similarity to the astrovirus CP, whereas the CP of 'bastro-like' viruses does not (Supplementary Fig. 3). A third clade is comprised of up to 11 viruses of arthropods, HHLV-8, and TkHLV-22, -32,

and -33. These viruses share one characteristic feature: an RdRp with permuted palm motifs. Whereas all positive-stranded RNA viruses with a canonical RdRp have seven conserved structural motifs (consisting of β-strands and α -helices) in the order G-F-A-B-C-D-E [11], the 11 hepe-like viruses of this clade exhibit the permuted order -C-A-B- of their active site motifs (Fig. 2 and Supplementary Fig. S5). The genomes of Beihai barnacle virus 1, Changjiang hepe-like virus 1, and Beijing sediment hepelike virus 1 encode four hypothetical proteins in addition to the polyprotein and have sizes of 10-11 kb, which is considerable larger than the genomes of the monopartite viruses of the acknowledged Hepelivirales families. This genome layout, however, is not conserved. TkHLV-32 and -33 have only three ORFs, with ORF1 and 2 corresponding to ORFs 1 and 4 of the former viruses.

Nine hepeliviruses were found to have peptidase-like CPs (Supplementary Fig. S4), and six possessed a third ORF encoding a protein with a Zn-binding RING/Ubox domain (data not shown).

We expected to trace human and animal enteric viruses, as all of the samples were collected in the metropolitan area of Berlin, Germany, and the two waterbodies are linked and receive the discharge of a local wastewater treatment plant and the drain water effluents of the city of Berlin after periods of heavy rainfall. Although such viruses are detectable in Havel River and Teltow Canal samples by specific PCR methods throughout the year [2 and unpublished results], we failed to identify sequence reads of known pathogenic noroviruses (Caliciviridae), enteroviruses (Picornaviridae), astroviruses (Astroviridae), or hepatitis A or hepatitis E viruses (Picornaviridae, Hepeviridae) in our metagenomes. Instead, we were able to demonstrate the presence of many unclassified hepe-like viruses and a few astroviruses. As shown recently for the picorna-like viruses of the Havel River [21], this result indicates the presence of a vast number of uncultured viruses compared to the still moderate number of viruses in the known virosphere. One notable group of unclassified hepe-like viruses are the bastroviruses, which can be differentiated in phylogenetic analysis into human bastroviruses with three ORFs and animal bastroviruses with two ORFs (Fig. 1). These viruses likely have vertebrate hosts. Another characteristic feature of bastroviruses is a CP with striking similarity to the astrovirus CP (Supplementary Fig. S3). The significance of this similarity, as well as the prevalence of bastroviruses in vertebrates remains to be elucidated. One member of this group, the Guangdong fish caecilians hepevirus, however, exhibits more variation in the CP gene than in the domains of the nonstructural protein (Supplementary Figs. S2 and S3). Overall, the four Hepelivirales families, the bastroviruses, and the 'bastro-like' viruses each have unique structural proteins with little or no similarity to one other.



Fig. 1 Phylogenetic analysis of the viral methyltransferase (**A**), helicase (**B**), and RdRp domains (**C**) of hepeliviruses. Blue boxes indicate reference virus families; yellow, brown, and green boxes denote unclassified bastroviruses, 'bastro-like' viruses, and hepe-like viruses with permuted RdRp. Sequences of Havel hepe-like viruses (HHLV; printed in red), Teltowkanal hepe-like viruses (TkHLV; printed in green), unclassified hepe-like viruses (printed in black) and reference strains of the families *Alphatetraviridae*, *Benyviridae*, *Hepeviridae*, *Matonaviridae*, and *Astroviridae* (printed in blue) were aligned in MEGA and used for tree inference with IQ-TREE 2. Optimal substi-

Another clade of hepe-like viruses (Fig. 1) is comprised of the so-called 'bastro-like' viruses. They have rather small genomes (<7 kb) that are significantly different from those of the bastroviruses. Some bastro-like viruses have been detected in insectivore bats and insects (*Culex* mosquitoes, wasps), while others have been found in aquatic organisms (sponges, bivalves, fish). The presence of bastro-like viruses in environmental water samples from the Havel River and Teltow Canal is compatible with the assumption that these viruses have aquatic hosts.

A third interesting clade includes viruses with rather large genomes (up to 11 kb). These viruses have been detected in mantis flies, crayfish, and barnacles and in various

tution models: TVM+F+R6 (A) and (B), and TVMe+R7 (C). The trees in panels A and B were arbitrarily rooted with members of the family *Matonaviridae*, and in panel (C), with members of the family *Astroviridae* and astro-like viruses. Scale bars indicate substitutions per site. Teltowkanal hepe-like viruses identified in consecutive samples are indicated by a diamond (\blacklozenge). Triangles (\blacktriangle) and dots (\bigcirc) indicate viruses from this study with almost complete and partial genome sequences, respectively. Details of the trees are presented in Supplementary Figure S2

environmental samples, including our Havel River and Teltow Canal water samples. As shown in Fig. 2, the RdRp of these viruses has a permuted order of the conserved palm subdomain motifs. A permuted RdRp has been observed only in birnaviruses [11] and two permutotetraviruses (Thosea asigna virus and Euprosterna elaeasa virus) [6]. In addition to the permuted order of the RdRp palm subdomain motifs, the polymerase sequences of the members of the families *Birnaviridae* and *Permutotetraviridae* differ significantly from those of the alphavirus supergroup, which includes the hepeliviruses. To our knowledge, this study is the first description of members of the alphavirus supergroup with a permuted RdRp.



Fig. 2 Schematic presentation of RNA-dependent RNA polymerases (RdRp) with the canonical and permuted order of the conserved palm motifs. Blue boxes represent motifs A to E of the palm subdomain,

and brown boxes represent motifs G and F of the finger subdomain. DxxxxD, GxxxTxxxN, and GDD indicate conserved amino acids of the RdRp active site

One obstacle to metagenomic studies of environmental water samples is the lack of information on the virus hosts. Even virus detection in faecal samples or the gills and guts of fish, arthropods, or molluscs may indicate dietary uptake or accumulation by filtration rather than infection. However, the increasing number of hepeliviruses associated with non-vertebrates suggests that they are the actual hosts. Further, the similarity of sequences from various sources and locations suggests a similar host range and a global distribution of hepeliviruses in terrestrial, marine, and freshwater ecosystems, which clearly deserves a more extensive investigation regarding prevalence and host range. Recurrent detection of identical viruses in samples from two linked waterbodies sampled in consecutive years, as shown in the present study, suggests an enzootic prevalence of hepeliviruses in many unknown hosts.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00705-022-05688-0.

Author contributions Conceptualization and study design: Roland Zell, Hans-Christoph Selinka. Sampling, transport, large-scale virus enrichment: Hans-Christoph Selinka. RNA preparation: Roland Zell. Sequencing and sequence data processing: Marco Groth. Data curation, bioinformatic analysis, and phylogenetic analysis: Marco Groth, Lukas Selinka, Roland Zell. Manuscript preparation: Roland Zell, Hans-Christoph Selinka. All authors have read and approved the final version of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The BioProject was submitted to the BioProject Database: BioProject ID PRJNA889576. Sequence reads were submitted to the Short Read Archive of NCBI: SRR2189516, SRR2189517, SRR2189518, SRR2189519.BioSamples were submitted to BioSample Database: acc. nos. SAMN31248992, SAMN31248993, SAMN31248994, SAMN31248995. Release date: Sept. 12th, 2022.Sequence data were submitted to GenBank: acc. nos. OP699055 - OP699153. We received a statement by GenBank staff that sequences will be released on Dec. 28th, 2022.

Declarations

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

Ethical approval This article does not contain studies with humans or animals.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Bauermann FV, Hause B, Buysse AR, Joshi LR, Diel DG (2019) Identification and genetic characterization of a porcine hepe-astrovirus (bastrovirus) in the United States. Arch Virol 164:2321–2326
- Beyer S, Szewzyk R, Gnirss R, Johne R, Selinka H-C (2020) Detection and characterization of hepatitis-E virus genotype 3 in wastewater and urban surface waters in Germany. Food Environ. Virol. 12:137–147. https://doi.org/10.1007/s12560-020-09424-2
- Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nat Meth 12:59–60

- 4. dos Anjos K, Nagata T, Melo FL (2017) Complete genome sequence of a novel bastrovirus isolated from raw sewage. Genome Announc 5(40):e01010-e1017
- French R, Charon J, Lay CL, Muller C, Holmes EC (2022) Human land use impacts viral diversity and abundance in a New Zealand river. Virus Evol 8:veac032
- Gorbalenya AE, Pringle FM, Zeddam JL, Luke BT, Cameron CE, Kalmakoff J, Hanzlik TN, Gordon KHJ, Ward VK (2002) The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. J Mol Biol 324:47–62
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol Biol Evol 32:268–274
- Oude Munnink BB, Cotten M, Canuti M, Deijs M, Jebbink MF, van Hemert FJ, Phan MVT, Bakker M, Farsani SMJ, Kellam P, van der Hoek L (2016) A novel astrovirus-like RNA virus detected in human stool. Virus Evol 2(1):vew005
- Pan J, Vakharia VN, Tao YJ (2007) The structure of a birnavirus polymerase reveals a distinct active site topology. Proc Natl Acad Sci USA 104:7385–7390
- 12. Remnant EJ, Baty JW, Bulgarella M, Dobelmann J, Quinn O, Gruber MAM, Lester PJ (2021) A diverse viral community from predatory wasps in their native and invaded range, with a new virus infectious to honey bees. Viruses 13(8):1431
- Reuter G, Boros A, Toth Z, Kapusinszky B, Delwart E, Pankovics P (2018) Detection of a novel RNA virus with hepatitis E viruslike non structural genome organization in amphibian, agile frog (*Rana dalmatina*) tadpoles. Infect Genet Evol 65:112–116

- 14. Roux S, Hallam SJ, Woyke T, Sullivan MB (2015) Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. Elife 4:e08490
- Sadeghi M, Altan E, Deng X, Barker CM, Fang Y, Coffey LL, Delwart E (2018) Virome of >12 thousand *Culex* mosquitoes from throughout California. Virology 523:74–88
- Sanborn MA, McGuckin Wuertz K, Kim HC, Yang Y, Li T, Pollett SD, Jarman RG, Maljkovic Berry I, Klein TA, Hang J (2021) Metagenomic analysis reveals *Culex* mosquito virome diversity and Japanese encephalitis genotype V in the Republic of Korea. Mol Ecol 30(21):5470–5487
- Shi M, Lin XD, Chen X, Tian JH, Chen LJ, Li K, Wang W, Eden JS, Shen JJ, Liu L, Holmes EC, Zhang YZ (2018) The evolutionary history of vertebrate RNA viruses. Nature 556:197–202
- Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang YZ (2016) Redefining the invertebrate RNA virosphere. Nature 540:539–543
- Wolf YI, Kazlauskas D, Iranzo J, Lucia-Sanz A, Kuhn JH, Kropovic M, Dolja VV, Koonin EV (2018) Origins and evolution of the global RNA virome. MBio 9:e02329-18
- Wu N, Zhang P, Liu W, Wang X (2018) Sogatella furcifera hepelike virus: first member of a novel *Hepeviridae* clade identified in an insect. Virus Res 250:81–86
- 21. Zell R, Groth M, Selinka L, Selinka HC (2022) Picorna-like viruses of the Havel river, Germany. Front Microbiol 13:865287

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.