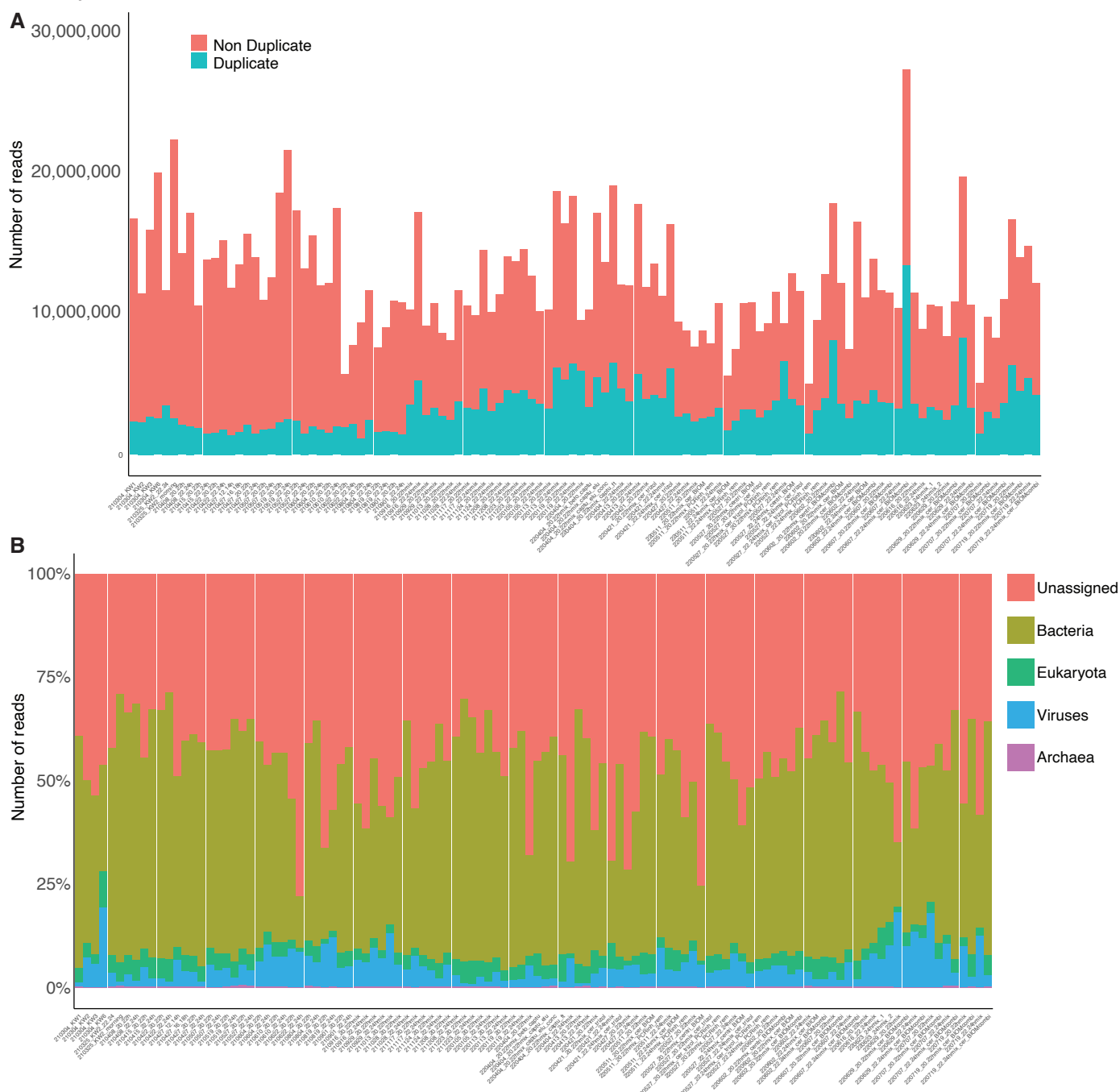


Figure S1, part 1



**Supplementary Figure S1. A**, number of raw reads per sample, with non duplicated reads in red, and duplicate reads (about 10-20% in most samples) in petrol. **B**, percentages of kaiju assignments to bacteria/eukaryota/viruses/archaea, as well as the percentage of reads that could not be assigned to a taxonomy. **C**, Principal component analysis (PCA) of all samples based on the 16082 taxonomies after initial filtering, with outlier samples labeled, and the top three contributors to PC1 and PC2, respectively, labelled with red arrows as well as taxonomy ID and name. **D**, barplot with the number of reads assigned to the outlying annotations (Methods). **E**, heatmap depicting abundance of the top 20 families for viruses, eukaryotes and bacteria, and 14 families for archaea in the total RNA sequencing data. Reads are shown as log10 transformed counts per million reads from the kaiju metagenomics pipeline, aggregated per month as indicated in the bottom. The bottom panel shows 0-1 normalized signal for four specific viruses. **F**, as in Fig. E but for virus families. **G**, As in E, but not aggregated per month (i.e. individual sequencing libraries). **H**, total number of reads mapping to astrovirus genomes from the individual Berlin samples. Shown is the totality of mapped reads when the raw sequencing reads files were mapped to all 643 individual astrovirus genomes (petrol), or when the raw sequencing reads files were mapped to the joint 9 astrovirus genomes shown in Fig. 1. **I**, heatmap depicting abundance of PMMV, crAssphage and astroviruses over time in the individual Berlin total RNA sequencing libraries. Reads are shown as log10 transformed mapped per million. **J**, heatmap depicting abundance of the astroviruses over time in the California samples. Reads are shown as log10 transformed mapped per million. **K**, phylogenetic tree of the genomic sequences used for mapping from taxonomy ID 568715, astrovirus MLB1. Indicated on the right is the percent identity of the sequences closest to the predominant sequence from MT766325.

Figure S1, part 2

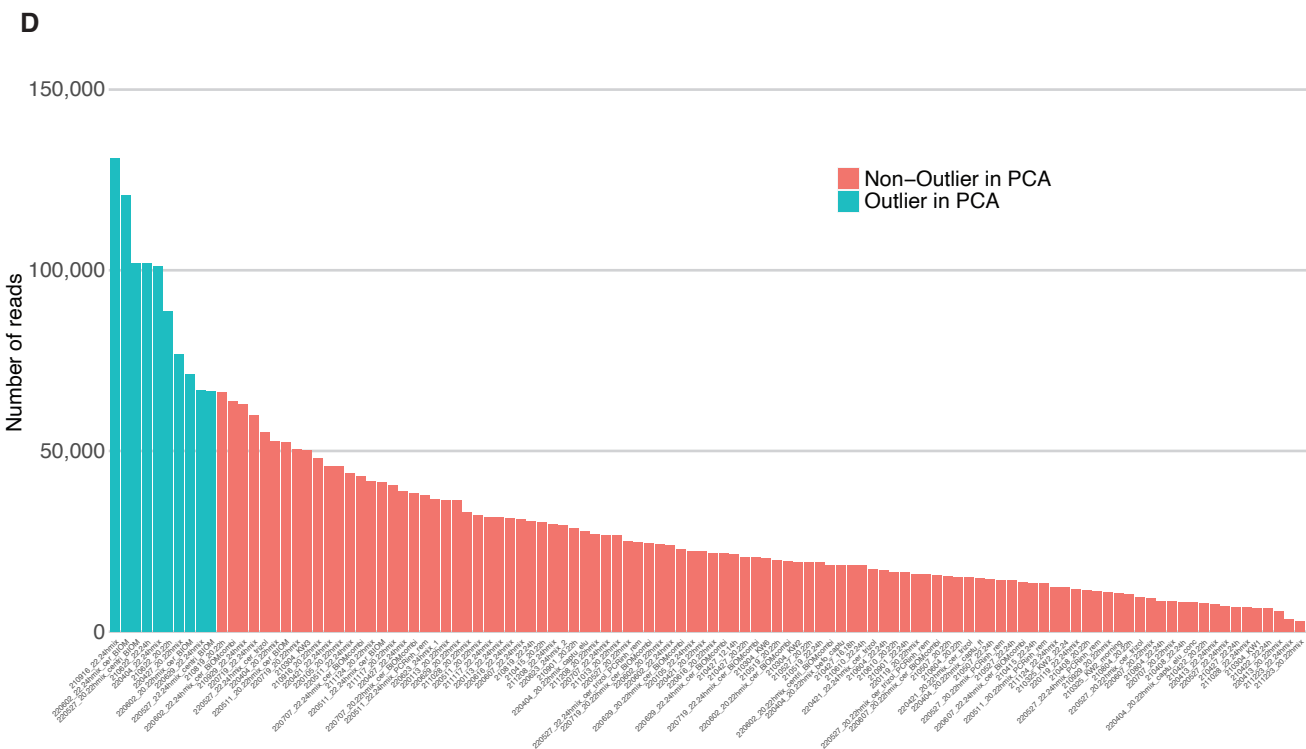
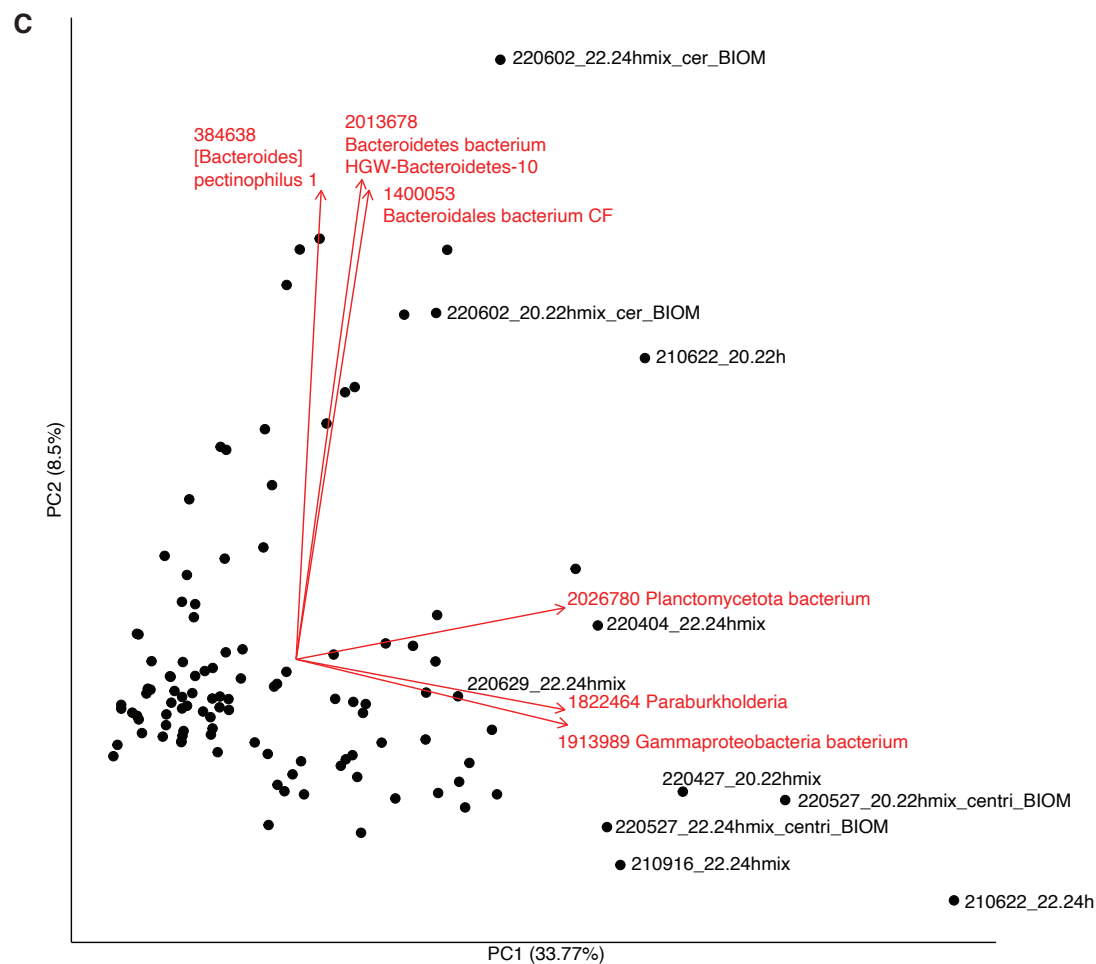
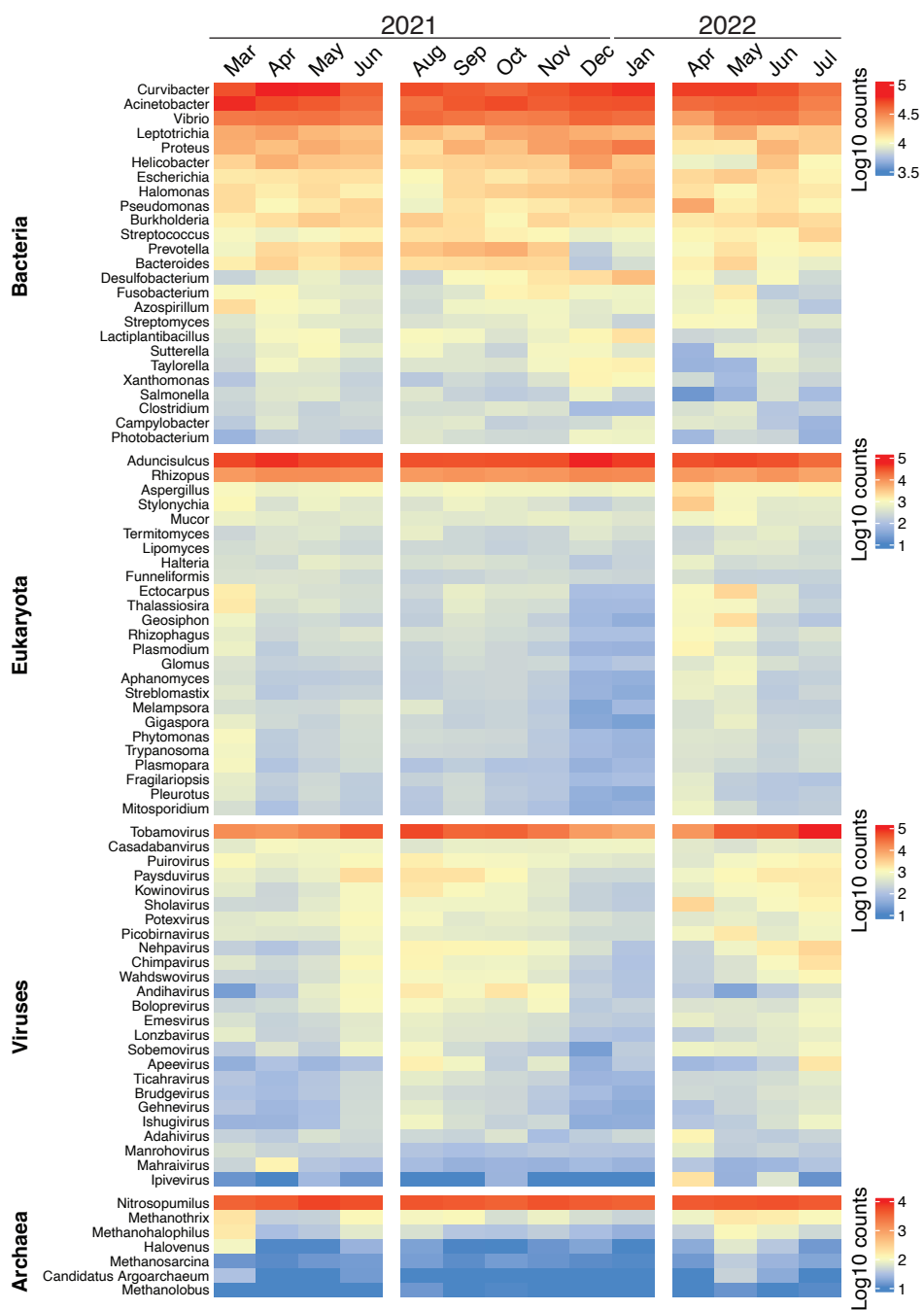


Figure S1, part 3

## E



**F**

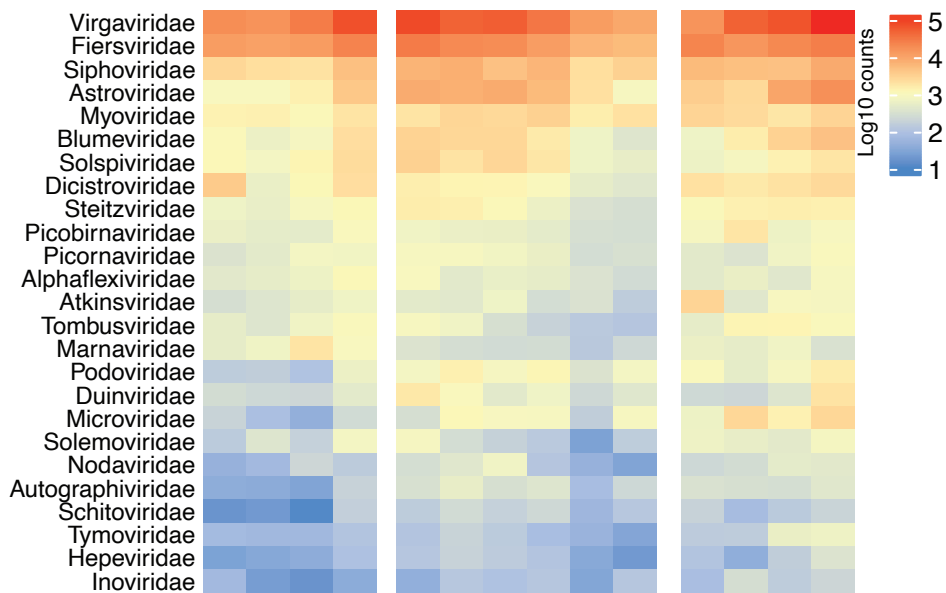


Figure S1, part 4

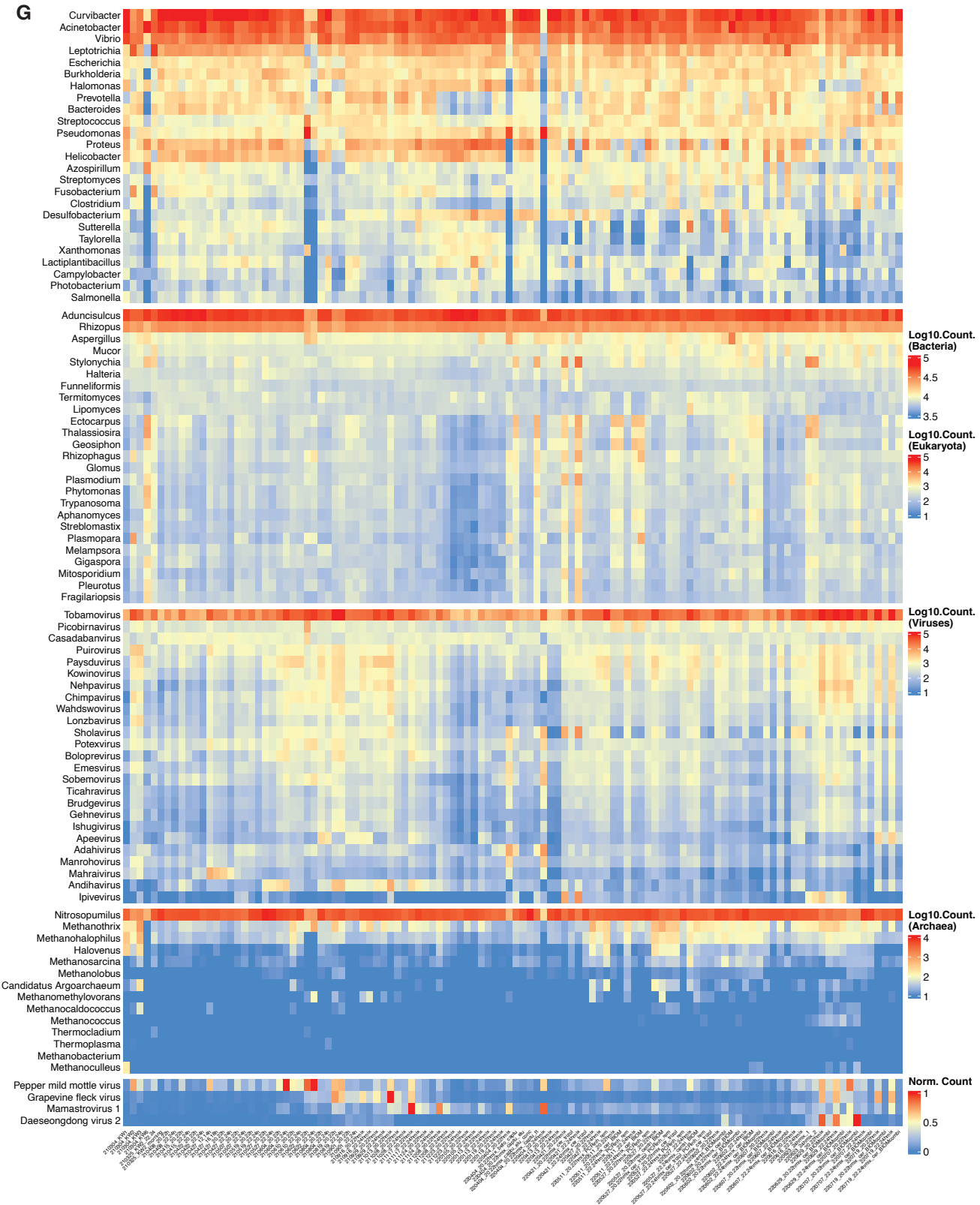
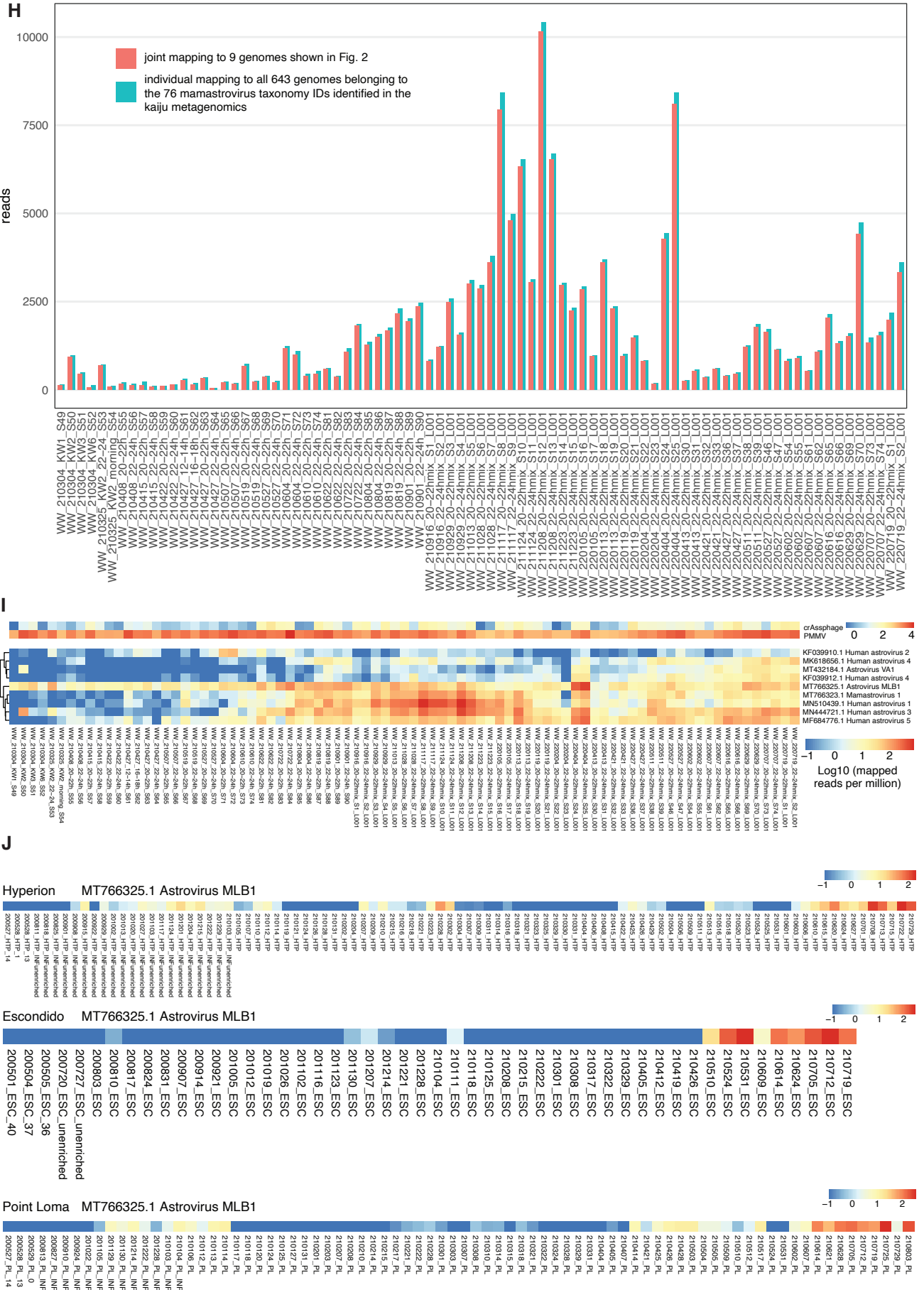




Figure S1, part 5



**K**

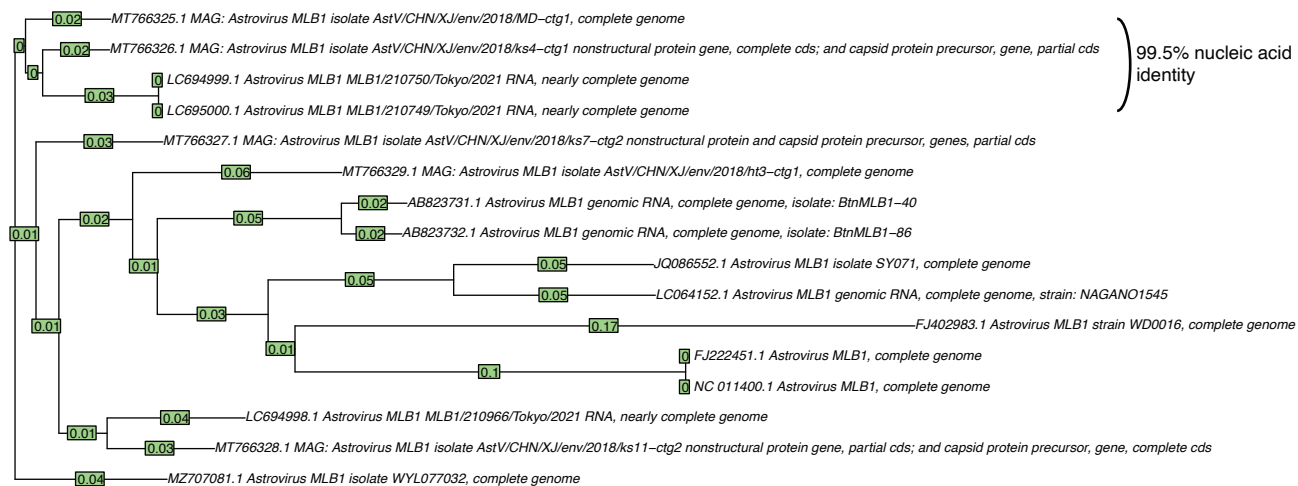
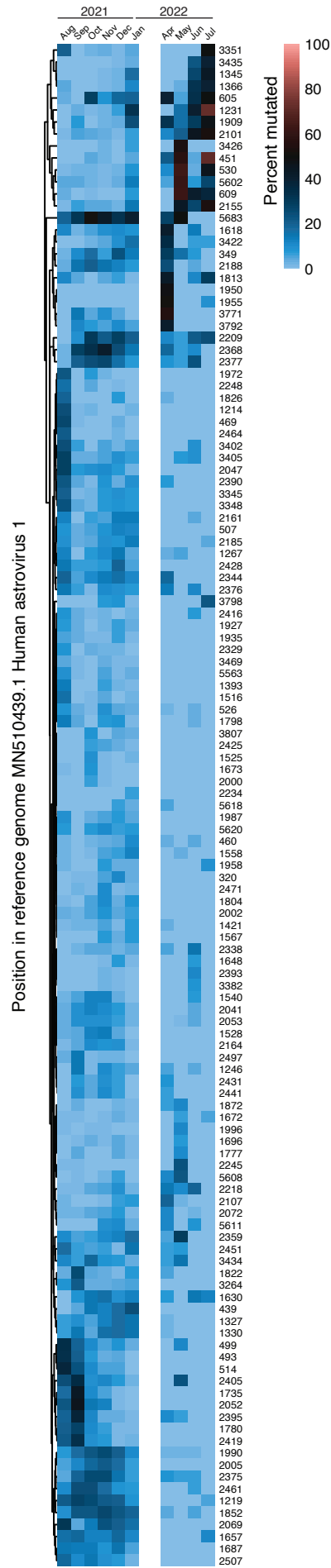


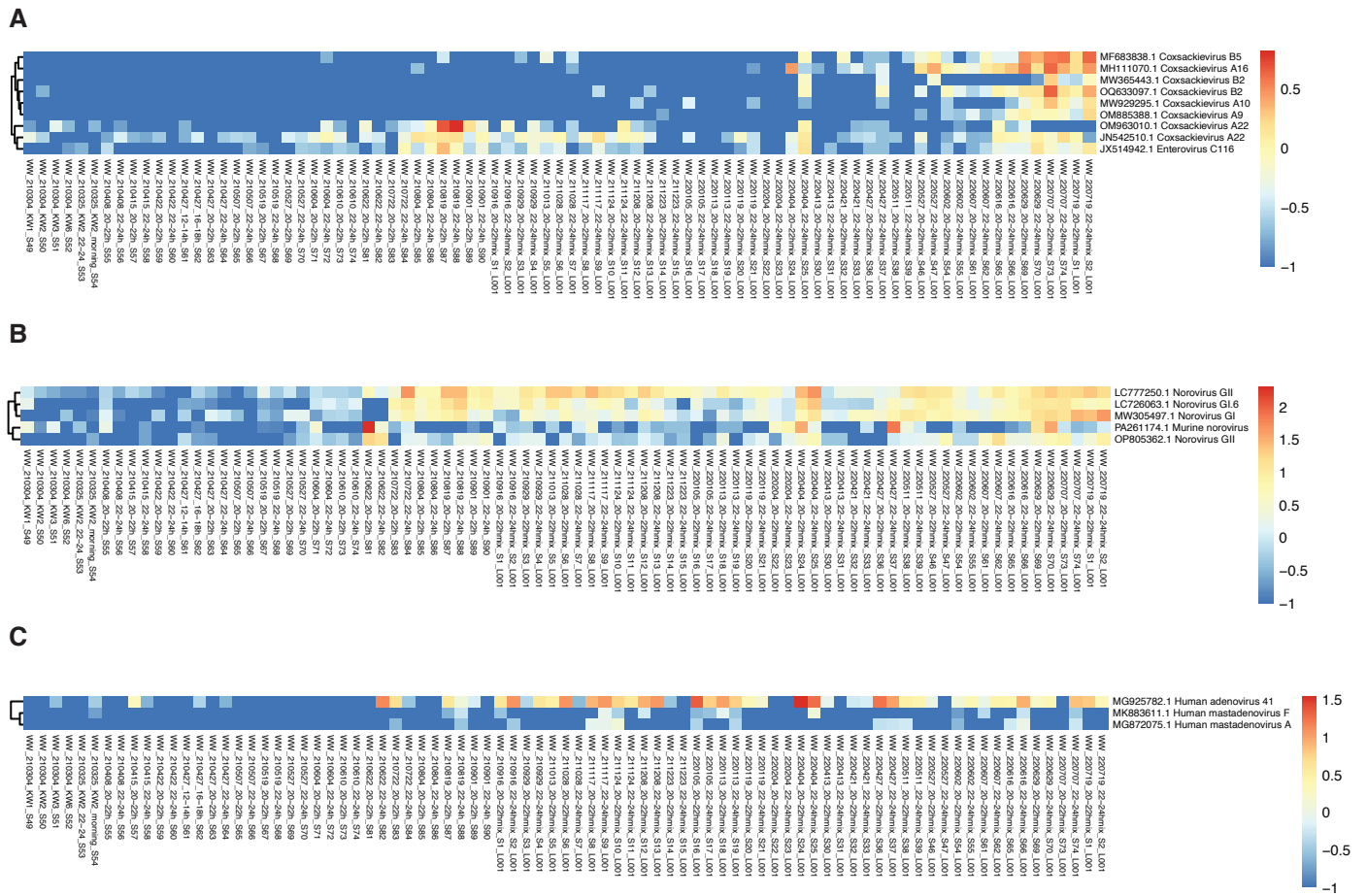
Figure S2

A



**Supplementary Figure S2. Profiling of individual nucleotide changes in sewage samples.** A, sewage sequencing data from the Berlin samples was merged by month as indicated, and the percent mutation per position indicated on the right was calculated in respect to the reference sequence (MN510439.1 human astrovirus 1). The color scale is indicated on the right of the figure.

Figure S3, part 1



**Supplementary Figure S3. A-C**, heatmap depicting abundance of enteroviruses (A), noroviruses (B) and adenoviruses (C) over time in the Berlin RNA samples. Reads are shown as log<sub>10</sub> transformed mapped per million. **D**, heatmap depicting abundance of selected viruses from the xHyb enrichment. Reads are shown as log<sub>10</sub> transformed mapped per million, aggregated per month as indicated in the bottom, and summed over all parts for segmented genomes. **E**, for selected viruses from the German clinical virology network, shown are log<sub>10</sub> transformed percent test positivity per month. **F**, direct comparison for four viruses, with for each virus the log<sub>10</sub> transformed number of reads mapped per million from wastewater on top (labelled wastewater), and log<sub>10</sub> transformed percent test positivity (bottom, labelled clinical). **G**, upper part: Ct values of RT-qPCR assays on selected target transcripts before and after ElementZero enrichment. Lower part, sequencing reads in absolute and relative numbers for SARS-CoV-2 and TBRV before and after enrichment. **H**, RNA-sequencing coverage profiles before and after ElementZero enrichment. For both SARS-CoV-2 and TBRV, shown are, for a selected sample, from top to bottom, the coverage before and after enrichment, and the position of the enrichment probes.

Figure S3, part 2

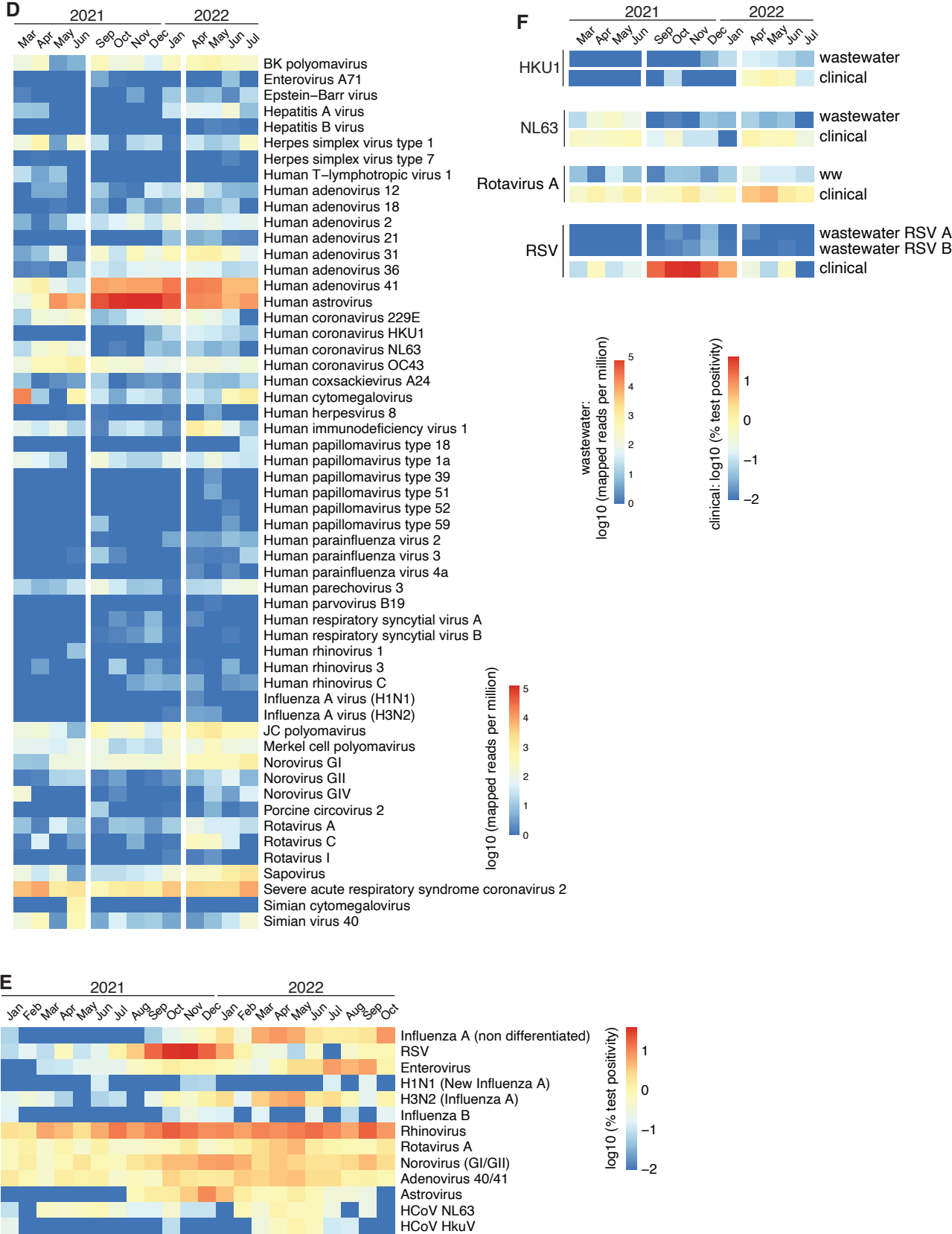


Figure S3, part 3

G RT-qPCR, Ct values

		PMMV	PMMV	E	E	N1	N1	N2	N2	IMP-Orf1b	IMP-Orf1b	ToBRFV	ToBRFV	
ElementZero eluate	SARS-CoV-2 input	WW0119 20-22h mix	25,3	25,3	33,1	33,2	33,4	34,7	33,2	35,3	34,3	33,8	20,9	20,8
		WW0119 22-24h mix	24,4	24,4	33,4	33,2	32,8	33,2	33,8	33,6	33,4	34,3	23,8	24,0
		WW0204 22-24h mix	26,1	26,2	34,4	nd	nd	34,5	nd	35,2	34,6	34,8	22,9	22,7
	SARS-CoV-2	WW0119_20-22h mix	nd	35,6	nd	35,8	nd	nd	nd	36,5	nd	nd	nd	nd
		WW0119 22-24h mix	35,3	35,4	35,3	nd	nd	36,6	nd	nd	35,5	nd	36,3	nd
		WW0204 22-24h mix	35,2	36,3	nd	37,0	nd	37,1	nd	nd	nd	nd	nd	nd
	ToBRFV	WW0119 20-22h mix	35,3	35,7	nd	nd	nd	nd	nd	nd	nd	nd	23,7	23,7
		WW0119 22-24h mix	34,2	36,3	nd	nd	nd	nd	nd	nd	nd	nd	22,3	22,8
		WW0204 22-24h mix	34,4	36,9	nd	nd	nd	nd	nd	nd	nd	nd	24,9	25,3

sequencing

		aligning reads		aligning reads, per million			
		SARS-CoV-2	ToBRFV	SARS-CoV-2	ToBRFV		
ElementZero eluate	SARS-CoV-2 input	WW0119 20-22 mix	16634889	12	66331	1	3987
		WW0119 22-24 mix	18592825	9	102505	0	5513
		WW0204 22-24 mix	6265867	9	38831	1	6197
	SARS-CoV-2	WW0119 20-22h mix	6621905	2927	93	442	14
		WW0119 22-24h mix	8730960	3020	331	346	38
		WW0204 22-24h mix	7443317	2729	27	367	4
	ToBRFV	WW0119 20-22h mix	5731701	0	1653218	0	288434
		WW0119 22-24h mix	10717491	1	743102	0	69335
		WW0204 22-24h mix	7734757	2	483384	0	62495

nd = not detected

H

Sample 220119 22-24h mix

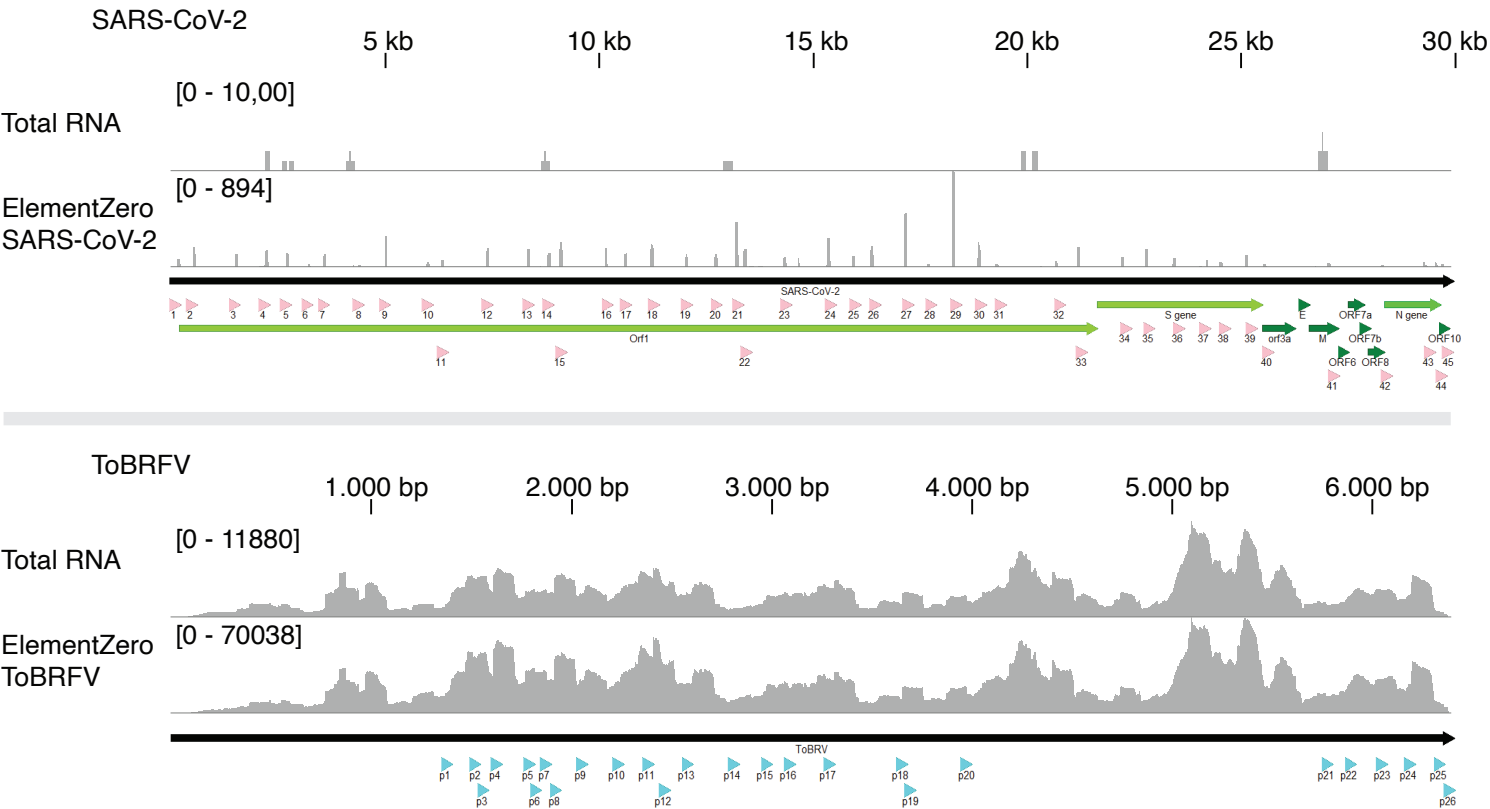
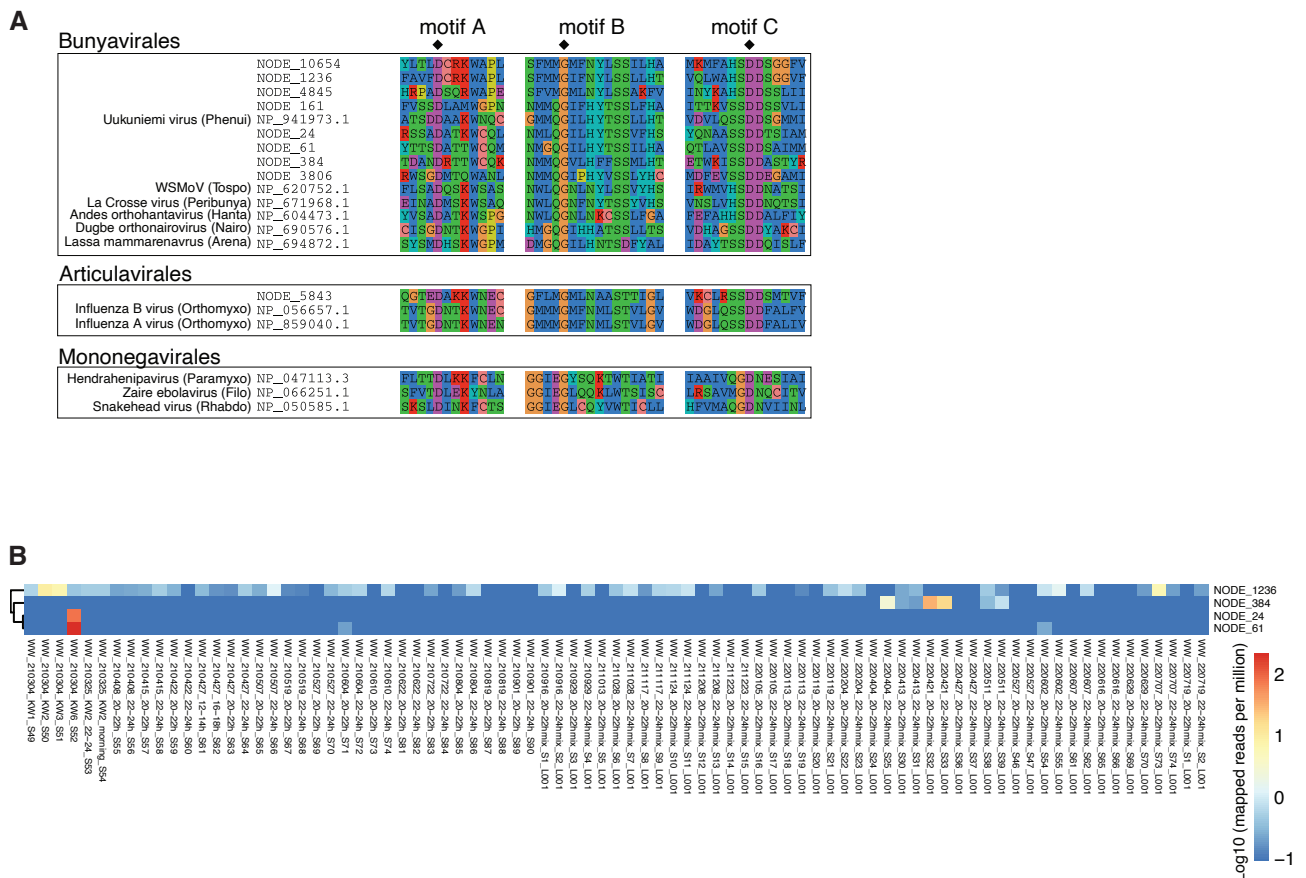


Figure S4, part 1

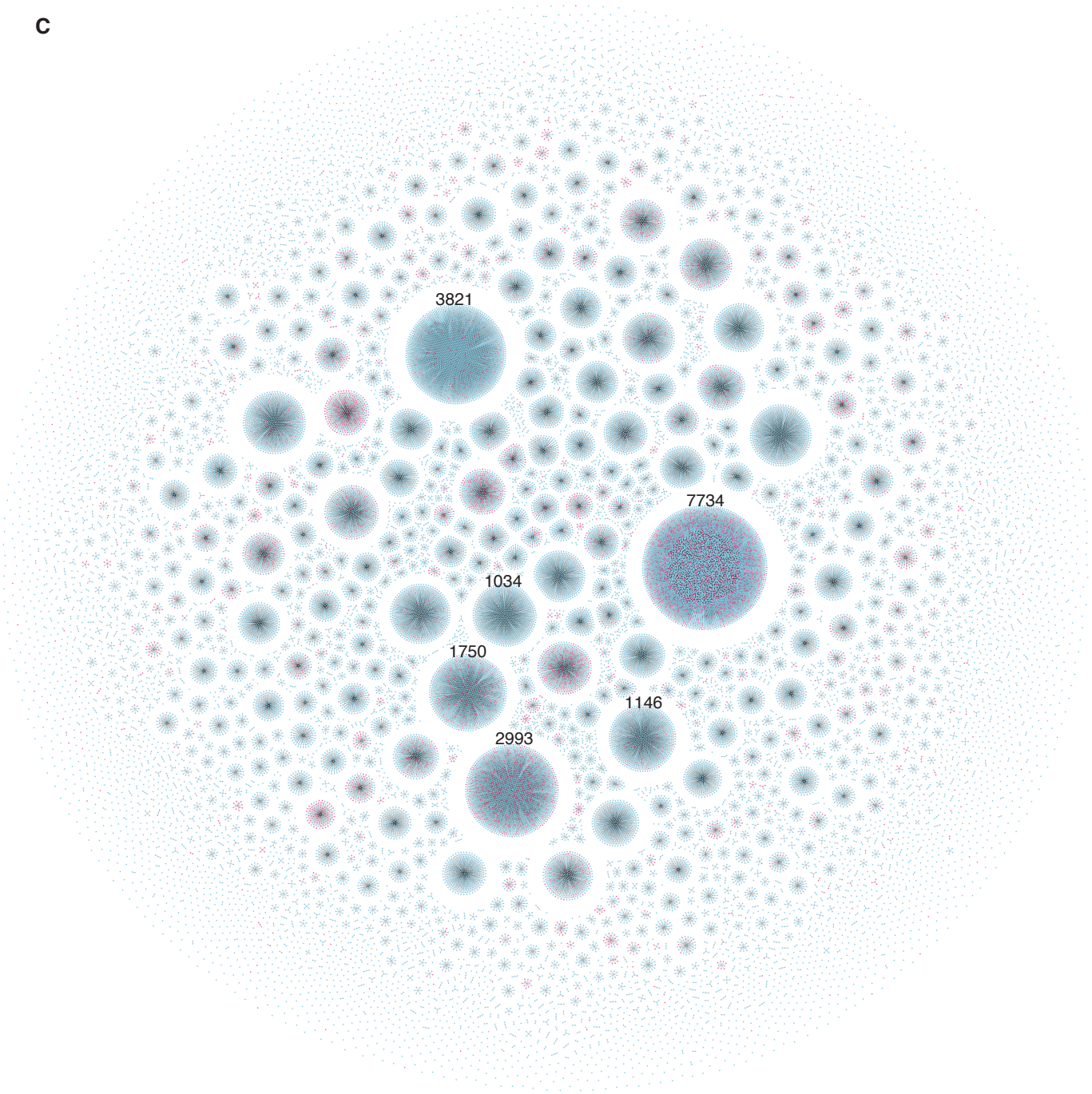


**Supplementary Figure S4. A,** Multiple sequence alignment of the region around most conserved motifs A, B and C of the viral RdRp for a selected set of novel and reference viruses. Two catalytic residues of motifs A and C and one residue of motif B involved in nucleotide selection are marked by diamond symbols. **B,** abundance of four novel bunyaviruses in the RNA sample time course. **C,** Known RdRp sequences were taken from the metagenomics virus discovery studies of Edgar et al., Zayed et al. and Neri et al. For each study, the *Lenarviricota* RdRp sequences were retrieved and clustered at 90% amino acid sequence identity. The three sequence sets were then joined and clustered again at 90% amino acid sequence identity to remove possible redundancy between studies. This formed the “known” set of RdRp sequences. The *Lenarviricota* RdRps from the wastewater RNA data were clustered with the known set to keep only those that have less than 90% amino acid sequence identity to any known RdRp, forming the “novel” set. For both sets, Palmscan was used to extract the well-conserved motif ABC region of the RdRp; RdRps without full motif ABC were discarded. This resulted in a total of 90376 RdRps of which 81312 are known and 9064 are novel. Finally, the total RdRp set was clustered at 50% amino acid sequence identity and this clustering was visualized using the igraph library in Python. Nodes representing known RdRps are colored in light blue, novel RdRps in red. Clusters with more than 1000 sequences are labelled with the number of sequences in the cluster.



Figure S4, part 2

C





**A**

June 2022 July 2022

May 25 2nd 7th 16th 23rd 29th 7th 19th

**Bacteria**

Acinetobacter  
Aliarcobacter  
Streptococcus  
Bacteroides  
Clostridium  
Aeromonas  
Acidovorax  
Pseudomonas  
Prevotella  
Staphylococcus  
Flavobacterium  
Chryseobacterium  
Bacillus  
Enterococcus  
Uruburuella  
Moraxella

Log10 counts

**Eukaryota**

Entamoeba  
Aspergillus  
Phytophthora  
Pythium  
Symbiodinium  
Perkinsus  
Tritrichomonas  
Streptomastix  
Piromyces  
Aduncisulcus  
Aphanomyces  
Rhizophagus  
Histomonas  
Trichomonas

Log10 counts

**Viruses**

Plateaulakevirus  
Shenzhenvirus  
Exceevirus  
Burzaovirus  
Sugarlandvirus  
Ambidensovirus  
Microvirus  
Bdellovirovirus  
Ripduovirus  
Kuravirus  
Friunavirus  
Pettyvirus  
Skunavirus  
Daemvirus  
Slopekvirus  
Pacinivirus  
Obolensvirus  
Fremauxvirus  
Audreyjarvirus  
Lahexavirus  
Nipunavirus  
Panexvirus  
Etemovirus  
Gemykibivirus  
Trinavirus

Log10 counts

**Archaea**

Methanotherix  
Methanosarcina  
Methanomethylivorans  
Can. Methanofastidiosum  
Methanobacterium  
Methanobrevibacter  
Methanospirillum  
Methanolobus  
Methanoculleus  
Methanohalophilus  
Methanococcoides  
Methanosphaera  
Methanohalobium  
Can. Burarchaeum  
Acidiplasma  
Methanococcus  
Can. Prometheoarchaeum

Log10 counts

**B**

Heatmap showing Log10 (mapped reads per million) for various adenovirus sequences across different samples. The color scale ranges from -1 (blue) to 3 (red).

Log10 (mapped reads per million)

Legend:

- MG925782.1 Human adenovirus 41
- MG872075.1 Human adenovirus 31
- GU191019.1 Human adenovirus 18
- DL073929.1 unidentified adenovirus
- MK905734.1 adeno-associated virus
- OP161130.1 adeno-associated virus 2
- MA630849.1 adeno-associated virus
- MZ668395.1 adeno-associated virus
- MK238400.1 crAssphage

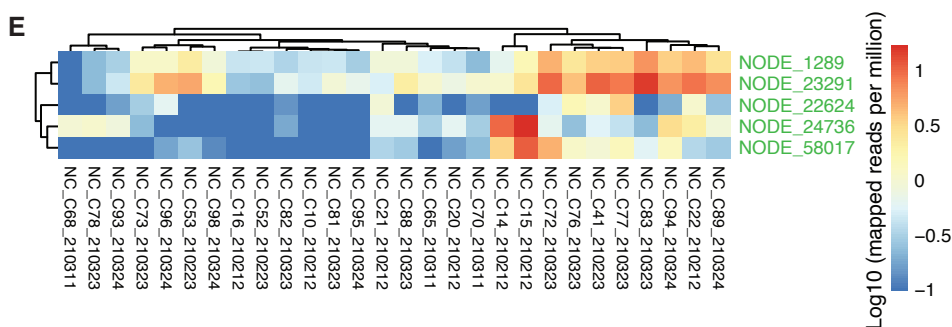
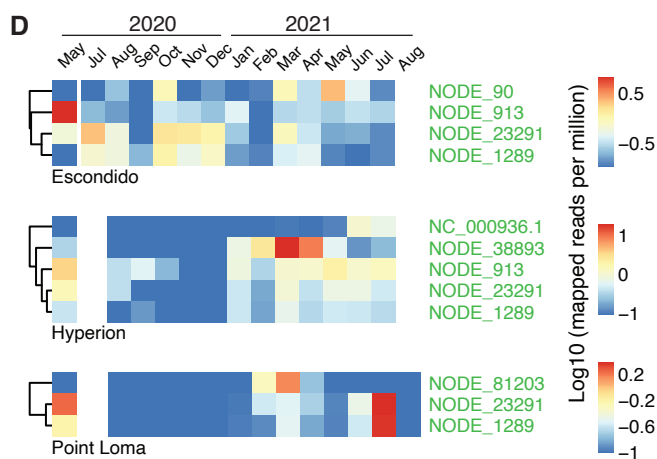
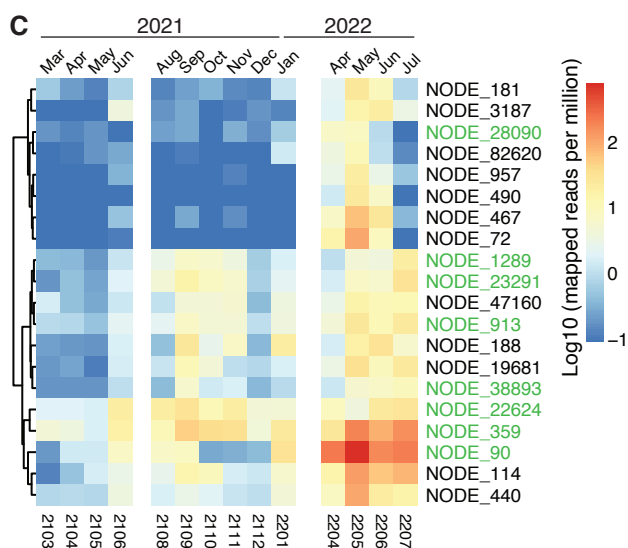
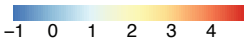
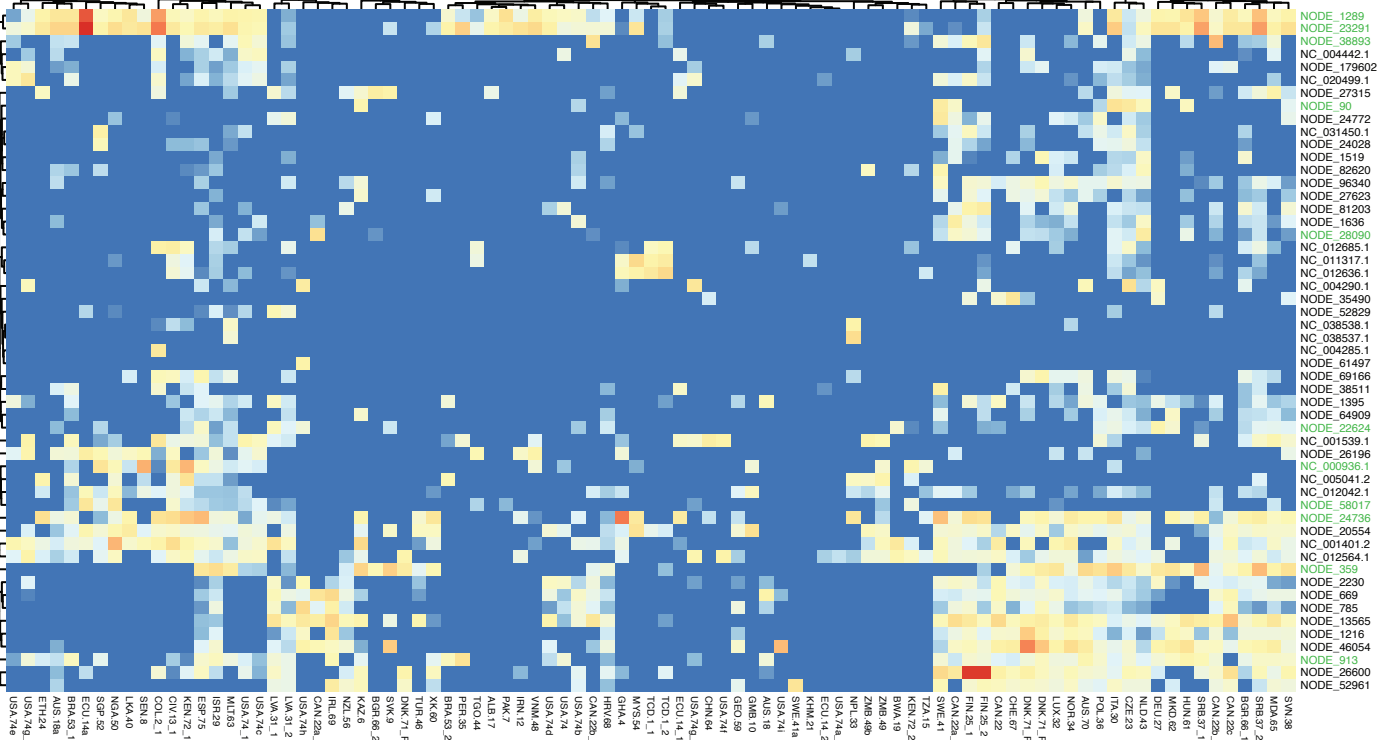


Figure S5, part 2

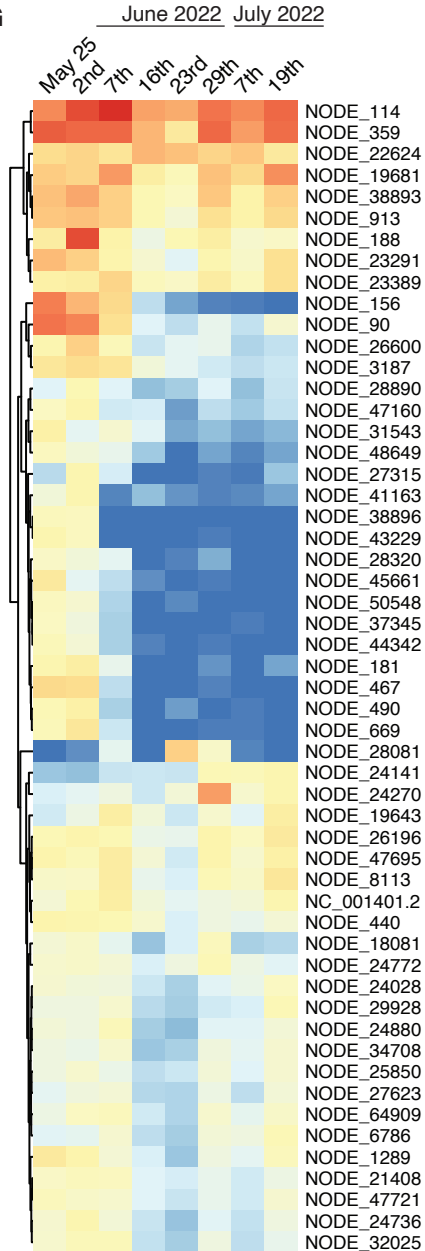
Log10 (mapped reads per million)



F



G



**A**

Figure S6, part 1

**Supplementary Figure S6. A**, Phylogenetic tree of the protein sequences of a random subset of 400 from the more than 40000 TnpBs in the NCBI protein database, and the 298 TnpB sequences, detected in the assembled sequence contigs from the both total RNA and DNA sequencing data, which have less than 90% identity to the closest relative in the NCBI protein database. The database sequences are colored in black and contain accession number and species. The wastewater sequences are colored in red and contain the contig name, the closest relative in the database, and the percent identity. The selected six clusters are labelled along with the average identities of the sequences in the cluster. In addition, the cluster of sequences coming from *E. coli*, is marked. About one fourth of TnpB sequences in the database come from *E. coli*. **B**, phylogenetic tree of TnpB DNA sequences. Sequence contigs were selected which extended more than 200 bp both upstream and downstream of the TnpB gene. Then, the DNA sequences of the TnpB genes together with 200bp both upstream and downstream were used for clustering. Sequences are colored by the cluster identified from **A**.

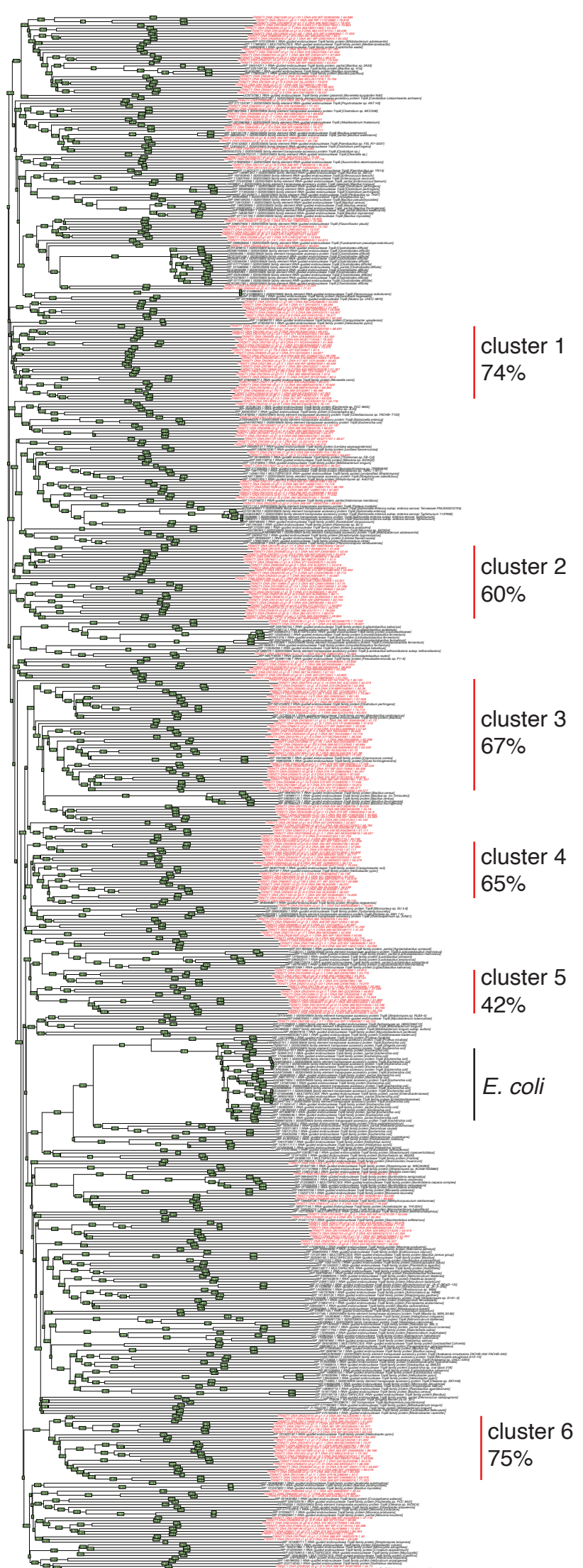
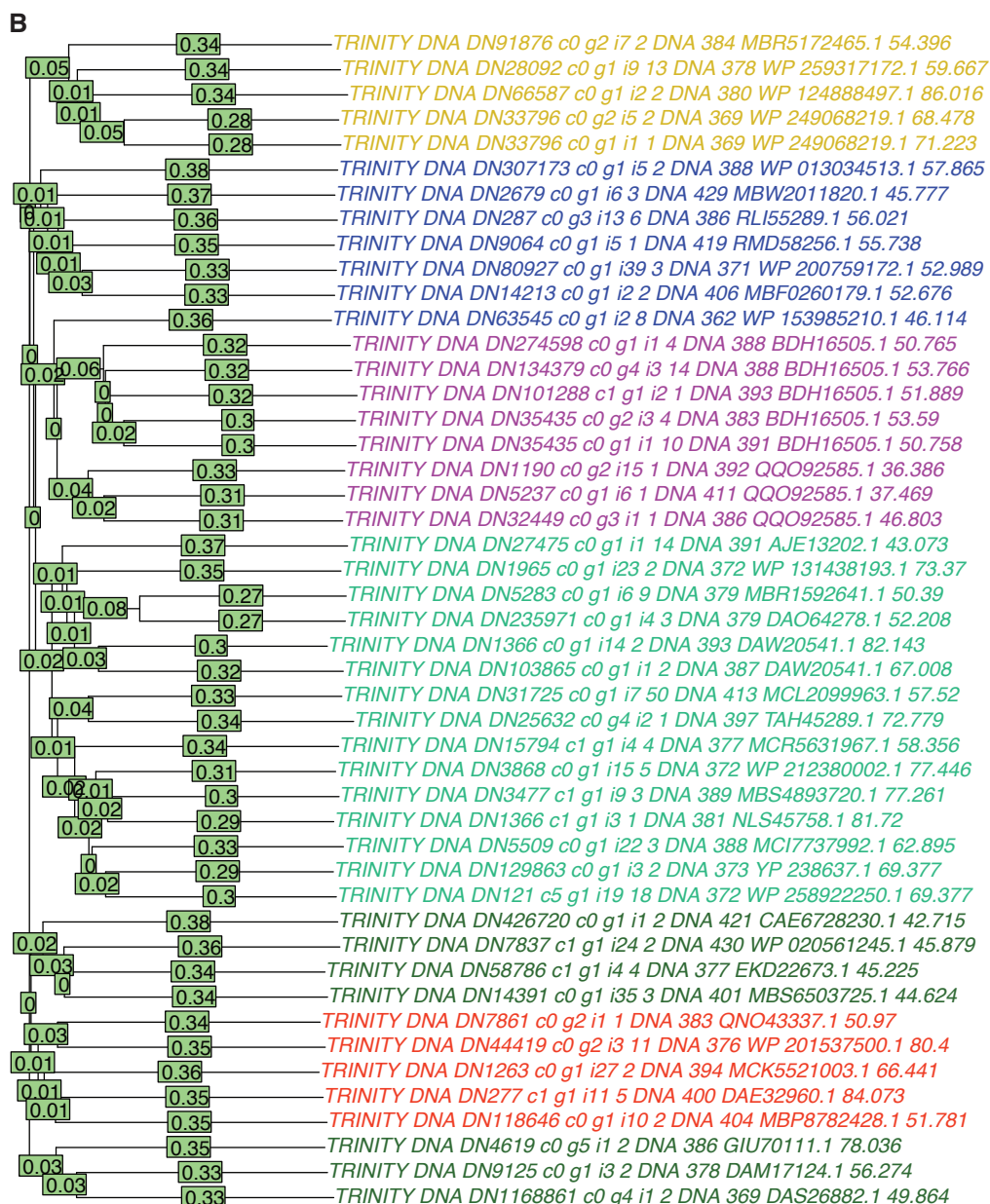
**B**

Figure S6, part 2



cluster1  
cluster2  
cluster3  
cluster4  
cluster5  
cluster6

## Supplemental Note: Enrichment using the VirBaits panel

Two processed wastewater samples (20-22 and 22-0) were additionally handled with another enrichment protocol. Samples were frozen in liquid nitrogen and disintegrated using the cryoPREP impactor (Covaris, Brighton, UK) as described (Wylezich et al. 2018). RNA was extracted using the RNAdvance tissue kit (Beckman Coulter, Germany). Complementary DNA (cDNA) was prepared from 82 and 169 ng total RNA of 20-22 and 22-0, respectively, using the SuperScript IV first-strand cDNA synthesis system (Invitrogen, Germany) and the NEBNext Ultra II nondirectional RNA second strand synthesis module (New England Biolabs, Germany). Sequencing libraries for generic metagenomics sequencing (Ion Torrent) were prepared from cDNA as detailed described (Wylezich et al. 2018) with some modifications (Pfaff et al. 2022). A part of the libraries was sequenced on an Ion Torrent S5XL instrument using Ion 530 chips and chemistry for 400-bp reads (Thermo Fisher Scientific, Germany). Another part (14 µl) of the libraries was treated with a VirBaits 2.0 panel (extended VirBaits panel, Wylezich et al. 2021) based on myBaits (Daicel Arbor Biosciences (Ann Arbor, USA) with a hybridization temperature of 65°C for 24 h according to the standard protocol (myBaits manual v.5.00, September 2020) and sequenced. VirBaits 2.0 contains oligonucleotide baits (80-nt) for different epizootic and zoonotic viruses as given in Wylezich et al. (2021, VirBaits panel) and supplemented with baits for Hepatovirus A, Hepatitis B virus, Hepacivirus C, Orthohepevirus A, infectious pancreatic necrosis virus, Zika virus, Usutu virus, Japanese encephalitis virus, Yellow Fever virus, Kyasanur forest disease virus, Tick-borne encephalitis virus, Dengue virus, Enterovirus C, Measles virus, Mumps virus, Salmonid novirhabdovirus, viral hemorrhagic septicemia virus, Marburg, Sindbis virus, Chikungunya virus, Rubella virus, Rustrela virus, La Crosse virus, Lassa mammarenavirus, Orthohantaviruses, Betacoronaviruses (VirBaits 2.0 one health panel). Datasets of both approaches (generic and VirBaits 2.0 treated) were analyzed using the RIEMS metagenomics analysis pipeline to taxonomically assign all individual read sequences of the obtained datasets (Scheuch et al. 2015).

As typical for libraries from low-input material and enriched samples, obtained datasets were quite small (Table).

Table: Overview of samples sequenced with generic libraries (original) and VirBaits-treated libraries (enriched) and dataset size.

Sample	20-22		22-0	
Approach	generic	VirBaits 2.0	generic	VirBaits 2.0
Library number	L5724	L5838	L5725	L5839
Reads obtained	57,369	1,052	82,004	662

Regarding virus reads detected with the generic approach, the majority of reads belongs to viruses that are associated with plants (e.g. Virgaviridae) and fungi but also viruses of invertebrate hosts (protists, arthropods) and bacteriophages. The amount of reads for that groups is given in the Figure. These viruses are depleted after the VirBaits-treatment since no target baits for them are included in the bait panel which may be beneficial when focussing on clinically relevant pathogens. The same is true for Astroviridae, a group that was quite abundant in generic datasets but was not enriched due to missing baits specific for them. In contrast, reads belonging to the Picornaviridae are enriched for both samples since the VirBaits 2.0 panel contains baits for Foot-and-mouth disease virus and Enterovirus A, B and C, namely Human hepatitis A virus, Swine vesicular disease virus, and Human poliovirus, respectively.

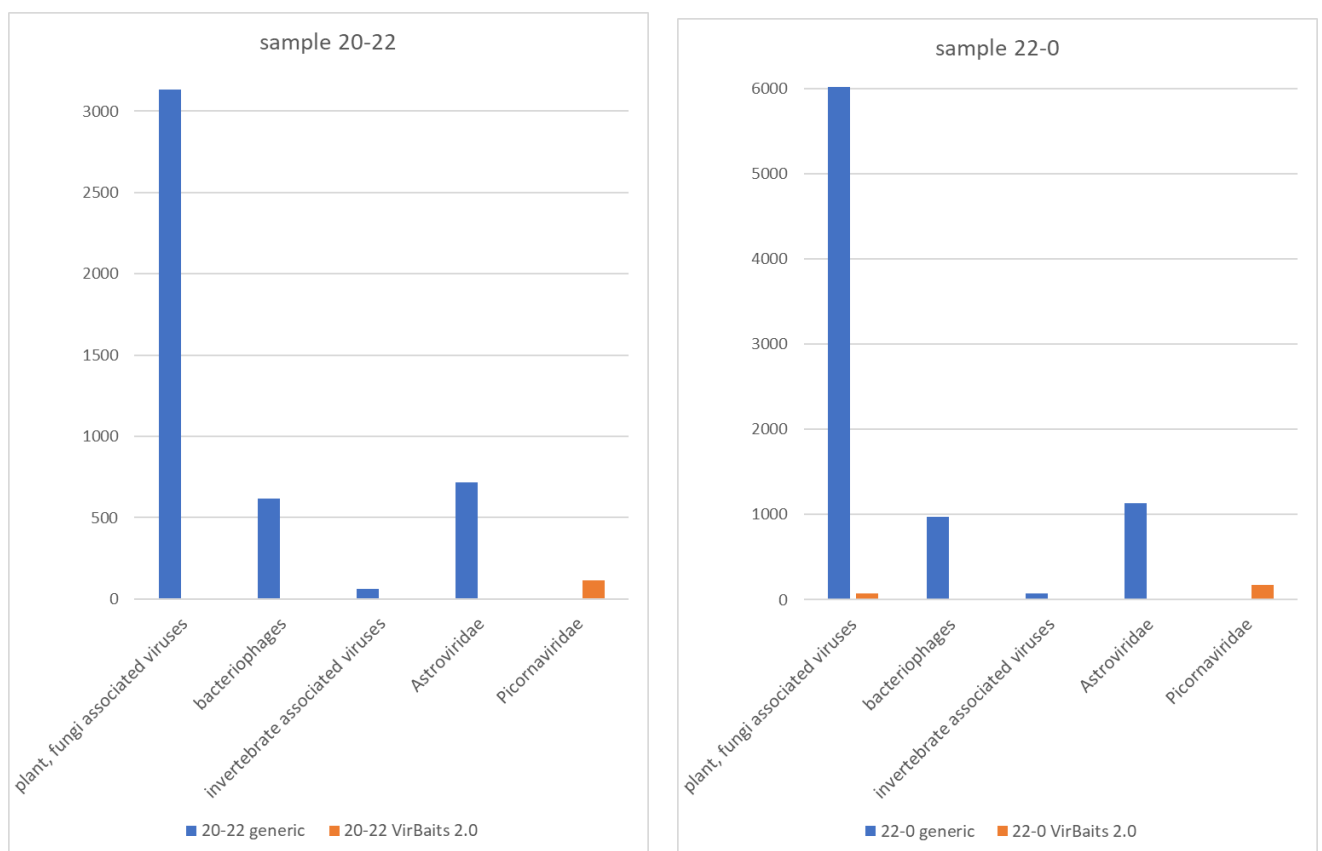


Figure: Overview of the amount of reads for the different virus groups detected in datasets before and after enrichment with the VirBaits 2.0 panel.

In conclusion, the VirBaits approach using myBaits custom capture kits is in principal well-suited to enrich reads for desired taxonomic groups especially viruses that are typically represented by only a few reads in complex datasets as seen here for Picornaviridae. The applied VirBaits 2.0 panel, however, is not optimal composed and adapted to wastewater samples since the panel do not cover all respiratory and gastrointestinal viruses (e.g. baits for Adenoviridae, Astroviridae, Caliciviridae, and Rotavirus are missing). Thus, for application with wastewater samples, a custom-designed wastewater myBaits panel needs to be created. This might be advantageous regarding cost reduction through smart panel design



considering the quite low identity that is prerequired between baits and target viruses. With the VirBaits panel based on RNA baits it is possible to enrich far related sequences and to obtain full genomes for virus taxa that are only about 72-74% related to the oligonucleotide bait sequences, and at least detection is possible with low identities of 36-46% (Wylezich et al. 2021). As also confirmed by Furtwängler et al. (2020), baits consisting of RNA perform better than the ones consisting of DNA. In addition, with the generic metagenomics sequencing is it possible to detect novel viruses also in not-enriched datasets as already shown (Bennett et al. 2020).

#### References:

- Bennett AJ, Paskey AC, Ebinger A, Pfaff F, Priemer G, Höper D, Breithaupt A, Heuser E, Ulrich RG, Kuhn JH, Bishop-Lilly KA, Beer M, Goldberg TL. Relatives of rubella virus in diverse mammals. *Nature*. 2020 Oct;586(7829):424-428. doi: 10.1038/s41586-020-2812-9. Epub 2020 Oct 7. Erratum in: *Nature*. 2020; 588(7836):E2. PMID: 33029010; PMCID: PMC7572621.
- Furtwängler A, Neukamm J, Böhme L, Reiter E, Vollstedt M, Arora N, Singh P, Cole ST, Knauf S, Calvignac-Spencer S, Krause-Kyora B, Krause J, Schuenemann VJ, Herbig A. Comparison of target enrichment strategies for ancient pathogen DNA. *Biotechniques*. 2020; 69(6):455-459. doi: 10.2144/btn-2020-0100.
- Pfaff F, Breithaupt A, Rubbenstroth D, Nippert S, Baumbach C, Gerst S, Langner C, Wylezich C, Ebinger A, Höper D, Ulrich RG, Beer M. Revisiting Rustrela Virus: New Cases of Encephalitis and a Solution to the Capsid Enigma. *Microbiol Spectr*. 2022; 10(2):e0010322. doi: 10.1128/spectrum.00103-22.
- Scheuch M, Höper D, Beer M. RIEMS: a software pipeline for sensitive and comprehensive taxonomic classification of reads from metagenomics datasets. *BMC Bioinformatics*. 2015; 16(1):69. doi: 10.1186/s12859-015-0503-6.
- Wylezich C, Papa A, Beer M, Höper D. A Versatile Sample Processing Workflow for Metagenomic Pathogen Detection. *Sci Rep*. 2018; 8(1):13108. doi: 10.1038/s41598-018-31496-1.
- Wylezich C, Calvelage S, Schlottau K, Ziegler U, Pohlmann A, Höper D, Beer M. Next-generation diagnostics: virus capture facilitates a sensitive viral diagnosis for epizootic and zoonotic pathogens including SARS-CoV-2. *Microbiome*. 2021; 9(1):51. doi: 10.1186/s40168-020-00973-z.