

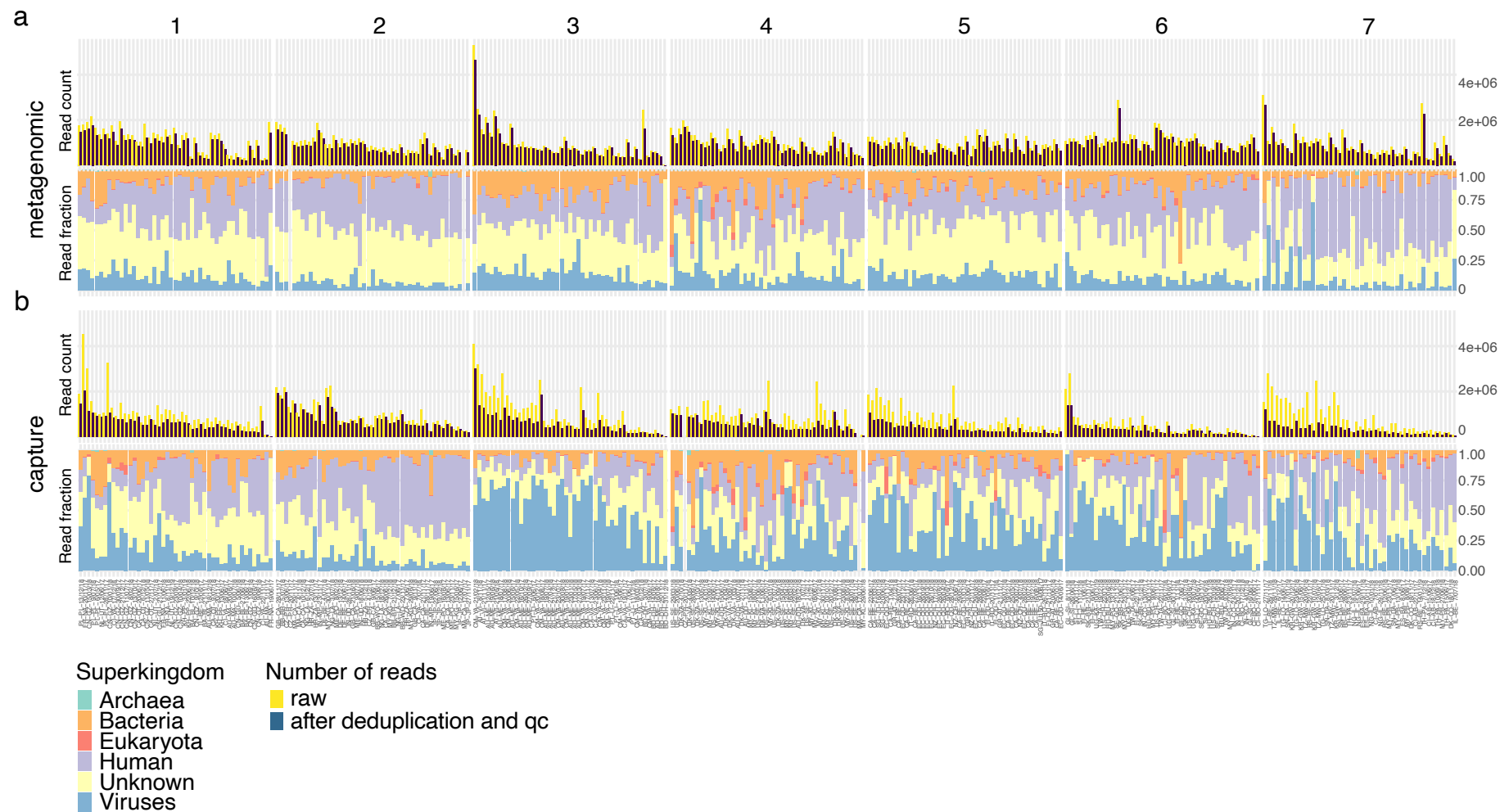
Supplementary Information for:

**Unveiling the global urban virome through
wastewater metagenomics**

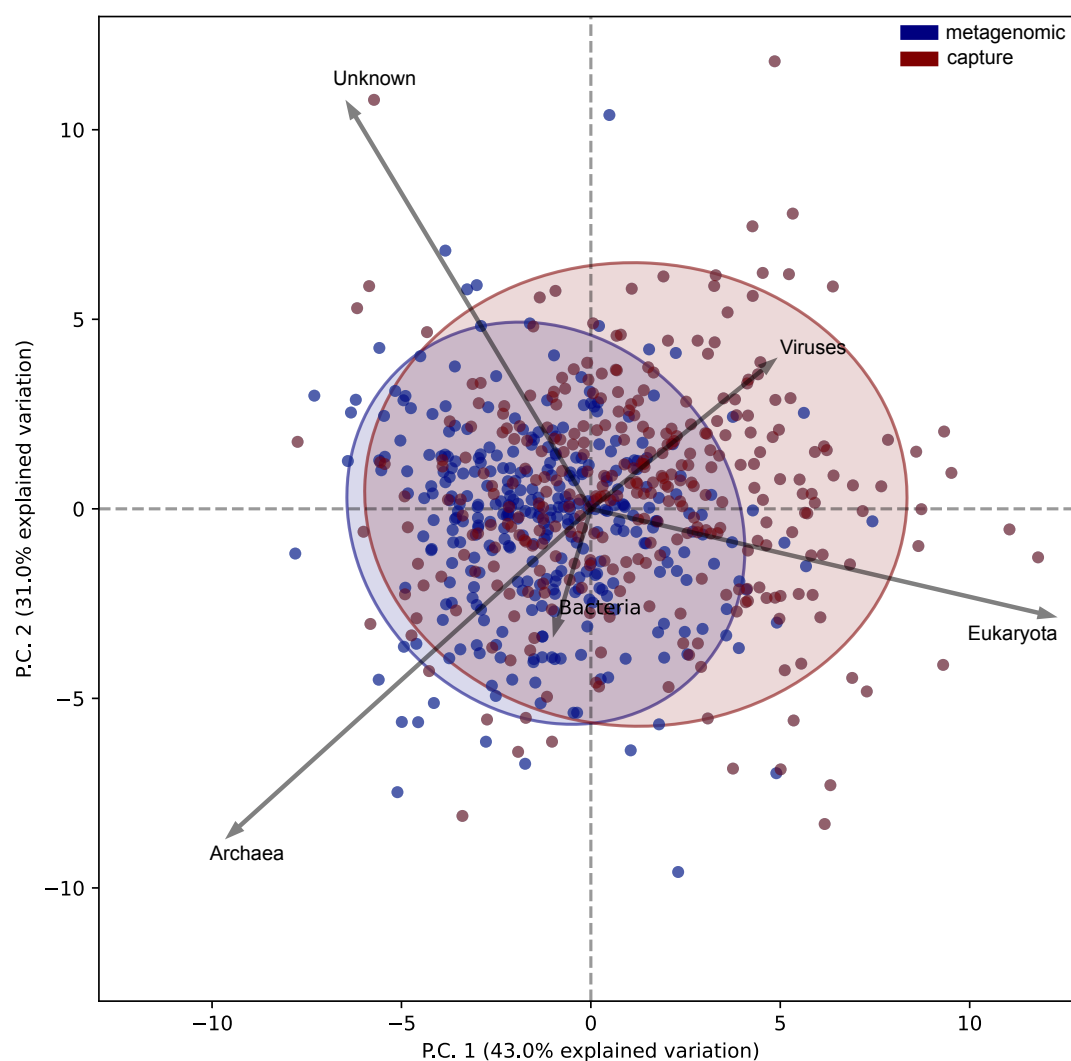
Nathalie Worp, David F. Nieuwenhuijse, Ray W. Izquierdo-Lara, Claudia M. E. Schapendonk, Christian Brinch, Emilie Egholm Bruun Jensen, Patrick Munk, Rene S. Hendriksen, Global Sewage Surveillance Consortium, Frank Aarestrup, Bas B. Oude Munnink, Marion P.G. Koopmans, Miranda de Graaf



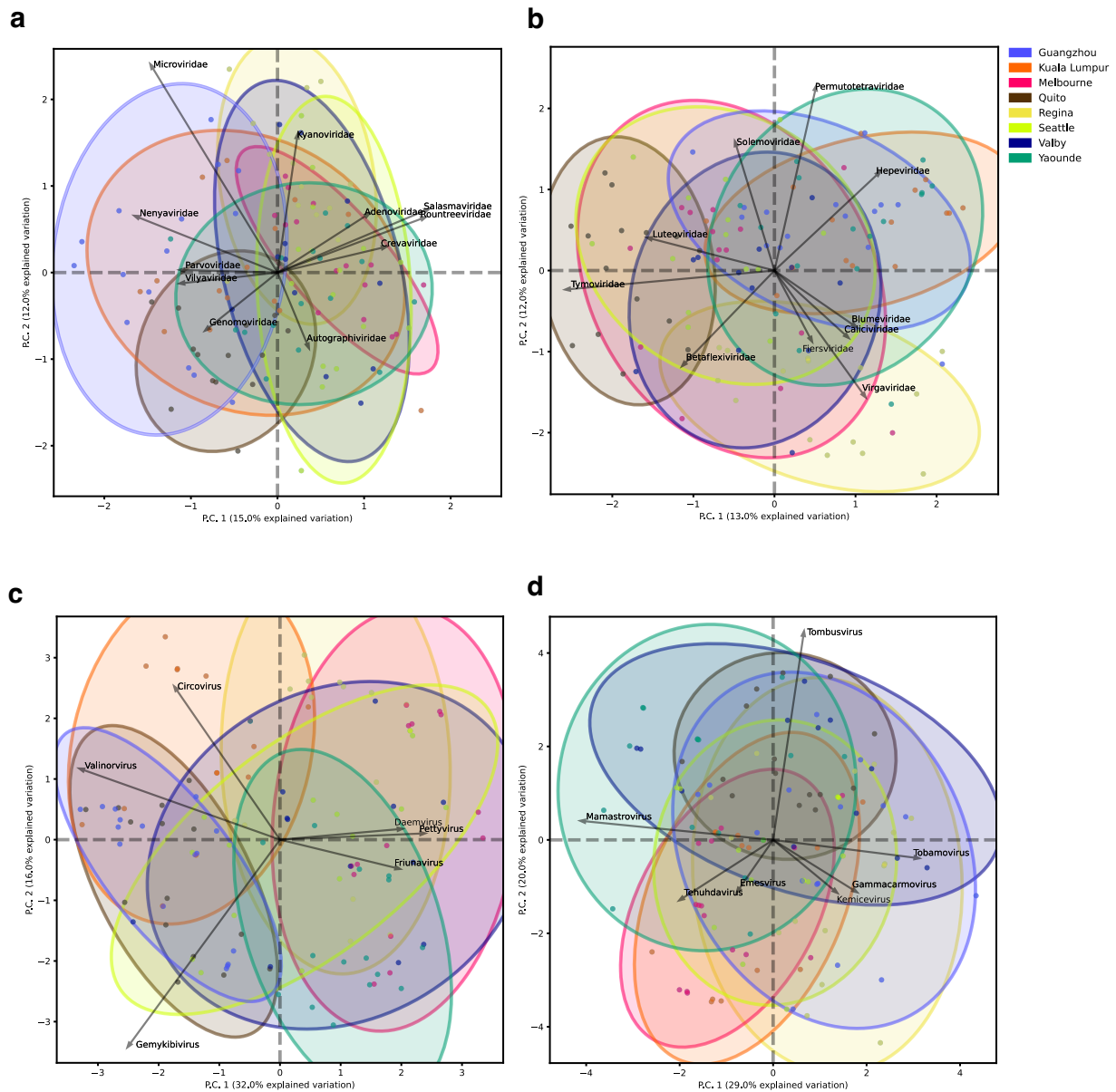
Supplementary Fig. 1. GastroCap probeset design. The total nucleotide space allocated per virus included in the probe design is shown in the segments. From the innermost to the outermost layers the arrangement progresses from viral families to genera and species. Source data are provided as a Source Data file.



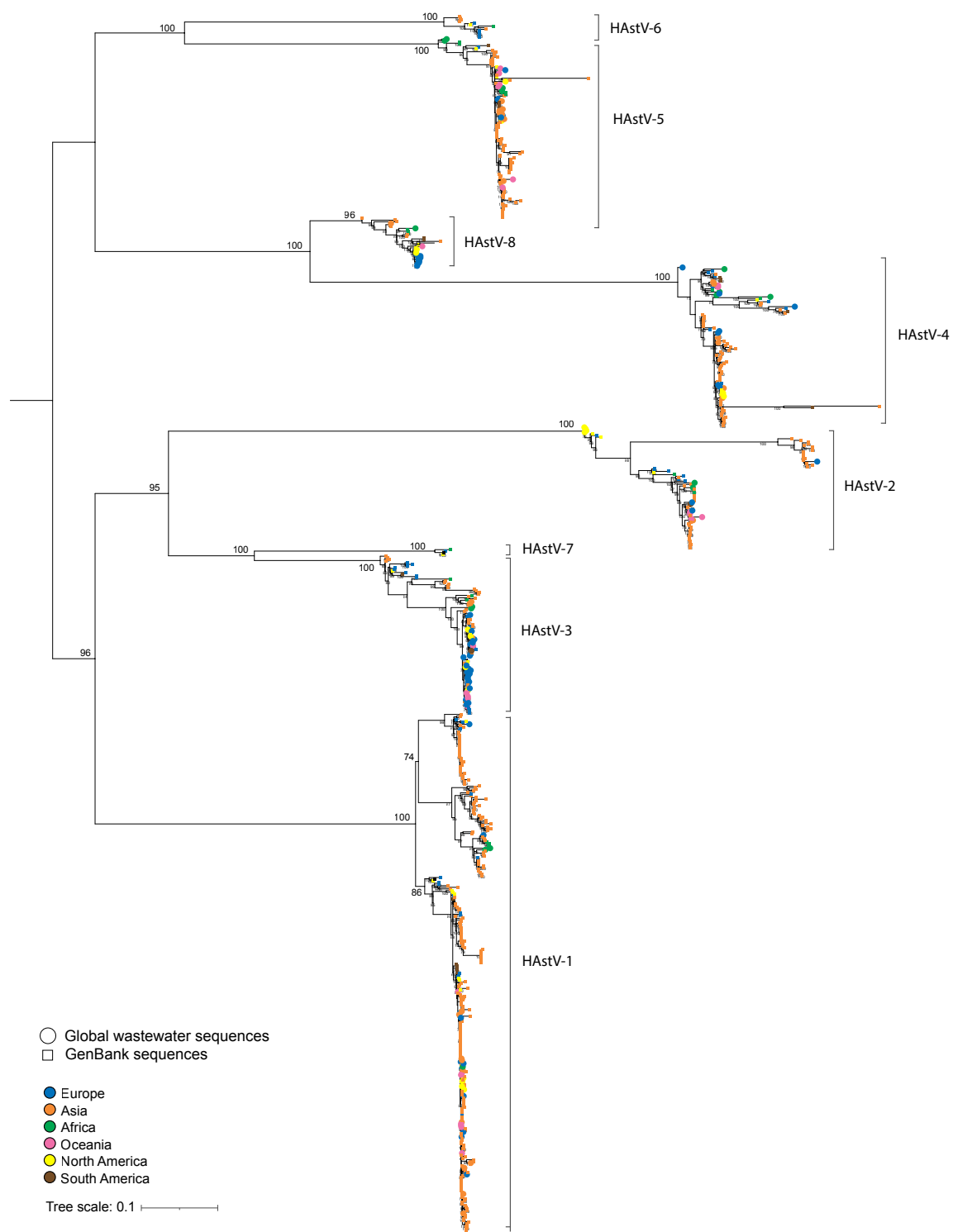
Supplementary Fig. 2. Raw read statistics and superkingdom read fractions for the metagenomic (a) and capture (b) sequence dataset. In each section, bar plots show the number of reads per sample (raw and after deduplication + quality control steps). Stacked bar plots display relative read abundance per sample at superkingdom level. Each facet represents samples sequenced within a specific batch (1-7).



Supplementary Fig. 3. PCA clustering of the urban wastewater compositions from 62 cities, depicting metagenomic (blue) and capture samples (red) at the taxonomic superkingdom level. The principal components were derived from read counts subjected to a CLR transformation.



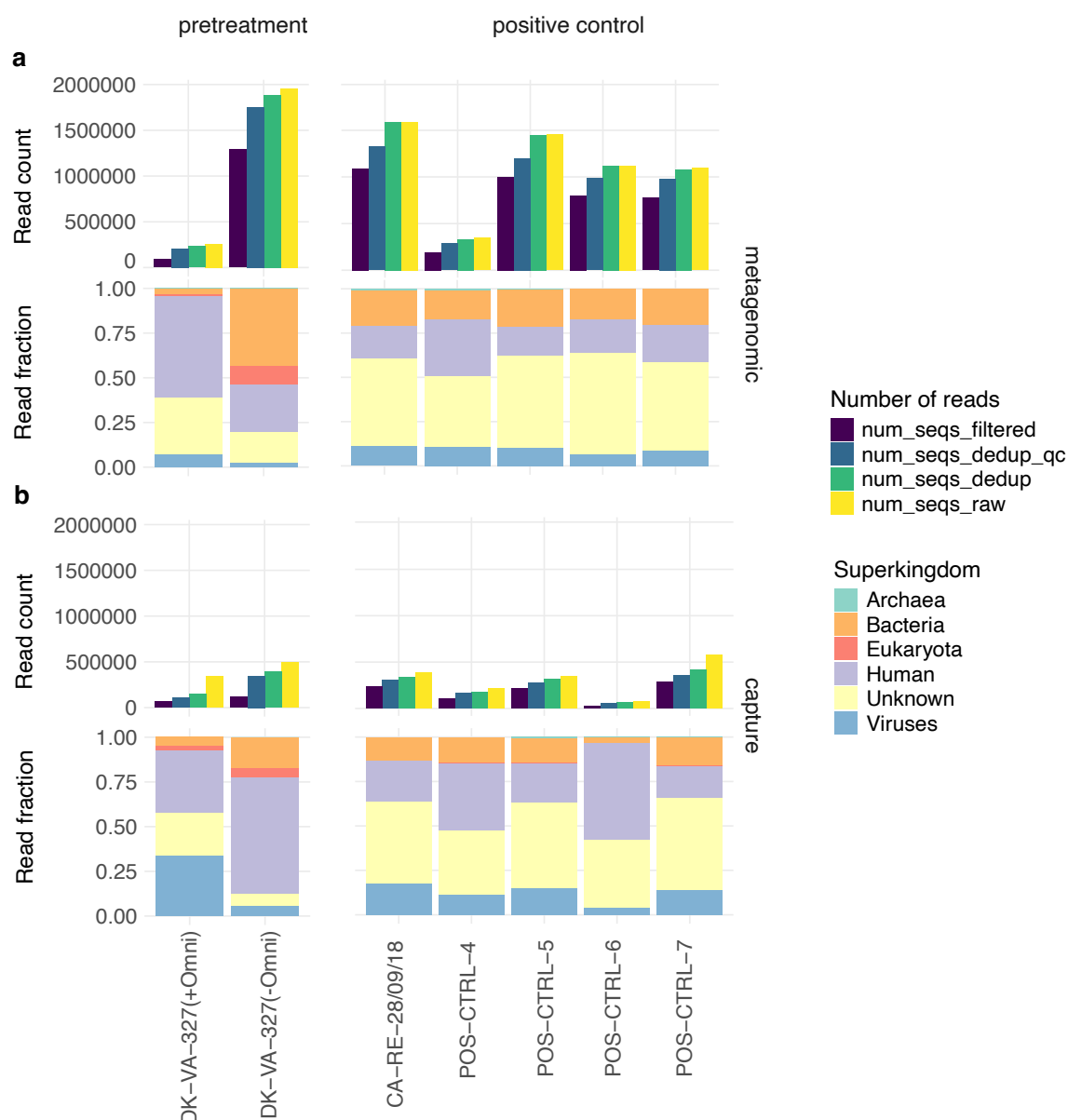
Supplementary Fig. 7. PCA clustering of the DNA and RNA virome compositions in urban wastewater samples from 8 cities of the shotgun metagenomic sequence dataset. The virome compositions of the longitudinal samples are shown separately for DNA viruses (a, c) and RNA viruses (b, d), and coloured by city. Taxonomic resolution is shown at the family level (a, b) and genus level (c, d). Principal components were derived from genome size-adjusted read counts subjected to a CLR transformation. Arrows represent viral families or genera, with their direction indicating their contribution to the principal components and their length indicating the strength of their contribution to the variance in virome composition.



Supplementary Fig. 8. Phylogenetic analysis of Human astrovirus (HAstV) type 1-8 of global wastewater and publicly available reference sequences. Maximum likelihood phylogenetic tree of HAstV type 1-8 based on partial ORF2 gene (capsid) sequences. Sequences obtained from GenBank are indicated in coloured squares, while sequences derived from wastewater with a minimum length of 500 bp are denoted in coloured dots. Colours represent the continent of origin (Europe, Asia, Africa, Oceania, North America, South America). Bootstrap values >70 are shown.

Supplementary Table 1. Viral contigs annotated with antimicrobial resistance genes (ARGs) in urban wastewater samples. Contig and gene lengths are given in base pairs.

Contig ID	Location	Sampling date	Contig length	Taxon name	Gene length	Annotation description	Accession(s)
35_B4_capture_N8_L28719	Seattle, USA	20/06/2018	28719	Caudoviricetes	882	23S rRNA (adenine(2503)-C(8))-methyltransferase Cfr	PF04055; TIGR04432; K06941; COG0820
35_B4_metagenomics_N13_L17881	Seattle, USA	20/06/2018	17881	Caudoviricetes	882	23S rRNA (adenine(2503)-C(8))-methyltransferase Cfr	PF04055; TIGR04432; K06941; COG0820
46_B6_metagenomics_N54_L3778	Chapel Hill, USA	02/11/2017	3778	Caudoviricetes	432	Rifampin ADP-ribosyl transferase	PF12120; K19062



Supplementary Fig. 9. Read statistics and superkingdom-level composition for pretreatment and positive control samples in the metagenomic (a) and capture (b) datasets. In each panel, bar plots show the number of reads per sample at various processing stages (raw, deduplicated, quality-controlled, and filtered) and stacked bar plots show the relative read abundances across superkingdoms. Samples DK-VA-327 in (a) and (b) are technical replicates of the same sample with and without OmniCleave treatment. Positive control samples (POS-CTRL 4–7) are technical replicates of CA-RE-28/09/18.