Metatranscriptomic Compendium of 55900 RNA Viruses Deciphers Human Health Implication and Ecological roles of RNA Virome in Global Wastewater Treatment Plants

Ling Yuan[1,2], Feng Ju[2,3]

[1] College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China

[2] Environmental Microbiome and Biotechnology Laboratory (EMBLab), Key Laboratory of Coastal Environment and Resources Research of Zhejiang Province, School of Engineering, Westlake University, Hangzhou 310030, China

[3] Westlake Laboratory of Life Sciences and Biomedicine, School of Life Sciences, Westlake University, Hangzhou 310024, China

*** Correspondence to**

Dr. Feng Ju

Assistant Professor

Westlake University

Email: jufeng@westlake.edu.cn

Address: 600 Dunyu Road, Xihu District, Hangzhou 310030, China

Tel.: 571-87963205 (lab), 571-87380995 (office)

Fax: 0571-85271986

Abstract

RNA viruses play diverse functional roles in engineered ecosystems, influencing biotechnological process and serving as indicators of human health. Comprehending the ecological and health significance of RNA viruses in wastewater treatment plants (WWTPs) can derive their valuable implications on microbial community control and wastewater-based epidemiology. This study delved into an extensive analysis of RNA sequencing data, totally over 3.8 Tb, sourced from 557 metatranscriptomes across global WWTPs, to scrutinize the diversity, host associations, and auxiliary metabolic functions of RNA viruses. We identified 11414 RNA virus operational taxonomic units from the WWTPs, doubling the current known diversity of the RNA viruses in global engineered systems. Phylogenetic analysis of RNA-dependent RNA polymerases supported the establishment of the five established RNA virus phyla while also advocating for taxonomy revisions due to our discovery of novel clades. Notably, the RNA viral community within the WWTPs was predominated by prokaryotic viruses, encompassing both previously identified RNA phage lineages (*Leviviricetes* and *Cystoviridae*) and potential prokaryotic viruses from newly-identified clades. Detections of prevalent human RNA viruses such as *Astrovirus*, *Respirovirus*, *Rotavirus*, and *Norovirus*, alongside high-risk human RNA viruses like SARS-Cov-2 and Zika virus, highlighted the potential of leveraging wastewater-based surveillance for human health protection. Moreover, the presence of auxiliary metabolic genes encoded by RNA viruses suggested their involvement in diverse host metabolic processes, including enhancing translation efficiency, cellular respiration, nitrogen metabolism, and even antibiotic resistance. Collectively, our findings unveil the previously hidden diversity, health implications, and biochemical impacts of RNA

viruses within WWTPs, underscoring their multifaceted roles in engineered environmental systems.

Key Words: RNA virus, Wastewater treatment plant, Virus-host associations, Virus-encoded auxiliary metabolic gene, human health, metatranscriptomics.

Introduction

The diversity and ecological roles of RNA viruses are less understood compared with DNA viruses. Nevertheless, RNA viruses are abundant biological entities in natural ecosystems, infecting both eukaryotes and prokaryotes crucial for human health[1-3], host metabolism manipulation[4-6], and overall earth biodiversity[7-9]. Recent advancements in RNA virus exploration, particularly through RdRp detection in total RNA sequencing datasets, called metatranscriptomics, have significantly hastened their discovery[6-8, 10-14]. These culture-independent techniques allow for the investigation of RNA viruses across various ecosystems, shedding light not only on these causing disease in humans, livestock, and crops[12, 15]. The known diversity of RNA viruses has been continuously expanded in several recent studies targeting large-scale surveys of marine[6, 10, 11] and soil ecosystems[13, 16], confirming that RNA viruses are nonnegligible members of both aquatic and terrestrial ecosystems.

Wastewater treatment plants (WWTPs) represent one of the world's largest biotechnology applications and are vital to the functioning of modern urban societies[17]. WWTPs are engineered systems used for purifying wastewater based on artificially enhanced microbial metabolism aimed at removing organic pollutants and nutrients and act as the physical and ecological link between human populations and natural water environment[18]. Virus is also one of the important biological members in WWTPs, with the potential to alter microbial diversity[19] and can provide health information of the population[20]. We are now beginning to understand the diversity of DNA viruses[21, 22] and RNA viruses[12, 23] in WWTPs. The surveillance of human RNA viruses in wastewater put forward the feasibility of wastewater-based epidemiology and its application for human health protection[20, 24], as recently exemplified for early

diagnosis and warning of SARS-CoV-2. However, the lack of knowledge on the overall diversity, host interactions, and biochemical roles of the collection of RNA viruses (i.e., RNA virome) in WWTPs remain unclear.

To explore the RNA viruses in the global WWTPs, this study built the first large-scale catalog of RNA virus and constructed the WWTP RNA Viruses Database (WRVirDB) by analyzing the largest metatranscriptomic datasets of WWTPs to date (3.8 Tb). We first resolved the phylogenetic and taxonomic diversity of RNA viruses in the global WWTPs, which led to 11414 unique RNA viral "species". Then, through investigation of the host range or specific hosts for RNA viruses, the presence of RNA viruses potentially infecting microbes (both functional niches and bacterial pathogens) and human highlighted the dual nature of RNA viruses for human society. Finally, the potential auxiliary metabolic functions encoded by RNA viruses revealed the involvement of RNA viruses in host metabolism and biochemical cycling. Together, we leveraged these results to elaborate the hidden diversity of RNA viruses and their ecological roles in the global WWTPs.

Results

Metatranscriptomic compendium and cross-compartment distribution patterns of RNA viruses in the global WWTPs

To explore the diversity of RNA viruses in global WWTPs, we analyzed more than 3.8 Tb RNA sequencing data from 557 publicly available metatranscriptomes covering five different groups of compartment samples of the WWTPs (Fig. 1a), including influent (IN), denitrification sludge (DS), activated sludge (AS), effluent (EF), and aerobic digestion sludge (AD) samples. A total of 55900 RNA viral contigs were identified by iteratively hidden Markov model (HMM)-based searching of RdRp genes followed by the assessment of authenticity of RNA viruses (See Methods and Extended Data Fig. 1). The 55900 RNA viral contigs were further clustered into 11414 RNA virus operational taxonomic units (vOTUs) at 90% average nucleotide identity (ANI) over 80% alignment fraction of the shorter sequence (AF) to empirically evaluate RNA viruses at the "species" rank, including 9276 RNA vOTUs encoded "complete" (or near-complete) RdRp domains (Supplementary Table 2). The overall diversity of unique RNA vOTUs in the WWTPs showed no indication of reaching saturation (Fig. 1b), suggesting that additional samples are required to assemble a more comprehensive catalog of RNA viruses in WWTP systems (Fig. 1b).

Further comparison of α-diversity patterns (measured in Shannon's diversity index) of RNA vOTUs revealed a highly varied richness of RNA viruses in the WWTP samples, especially for influent samples (Fig. 1c). Non-Metric Multidimensional Scaling (NMDS) analysis based on vOTUs composition in all the metatranscriptomes suggested that sample types could well separate (Analysis of similarity, ANOSIM R^2 = 0.821, P = 0.001) the RNA viral community structure in the WWTPs (Fig. 1d). To

fairly assess RNA vOTUs in terms of diversity and abundance throughout the WWTPs, samples from four compartments (i.e., IN, DS, AS, and EF) of 12 WWTPs from Switzerland, which were spiked with mRNA internal standards to achieve absolute quantification estimation and sequenced at a similar depth (8.26 ± 1.90 G per sample), were selected and comparatively analyzed. The result showed that the influent samples showed significantly higher α-diversity of RNA vOTUs compared with other three types of compartment samples (Fig. 1e). The RNA viral communities were more similar within than between compartments (ANOSIM R^2 = 0.325, P = 0.001) (Fig. 1f and Fig. 1g). In terms of the estimation of absolute concentration of RNA viruses in the WWTPs, the combined absolute abundance of RNA vOTUs was significantly ($P < 0.00001$) higher by about two orders of magnitude in the sludge samples ($1.34\times10^{12} \pm 9.06\times10^{11}$ and $1.30\times10^{12} \pm 1.08\times10^{12}$ for DS and AS, respectively) compared to the wastewater samples ($6.95\times10^{9} \pm 2.67\times10^{9}$ and $1.79\times10^{10} \pm 1.81\times10^{10}$ for IN and EF, respectively) (Fig. 1h). Except pairwise relationship between DS and AS (ANOSIM $P > 0.05$), the RNA vOTUs composition in any two compartments of the WWTPs were significantly separated (ANOSIM $P < 0.01$). The separated RNA vOTUs distribution in global (Fig. 1d) and local (Fig. 1f) scale indicated that the wastewater treatment process exerted environmental filtering impacts on RNA viruses in the WWTPs, leading to significant variations in the RNA viral community structure across the different stages of wastewater treatment during pre-treatment, treatment and post-treatment. Among samples of 12 WWTPs in Switzerland, the RNA viral community structures of influent samples were most similar to each other (having the lowest Bary-Curtis dissimilarity rank values), and then the dissimilarity within groups increased as the wastewater passed through the denitrification and activated sludge tanks, finally reached the maximum dissimilarity

within the effluent sample group (Fig. 1g). The observed increasing dissimilarity throughout the WWTPs, as well as the larger variations of α-diversity and absolute abundance of RNA vOTUs in the effluent samples than influent samples, revealed the wastewater treatment can apply distinct and uneven effects to RNA virome.

Knowledge expansion of phylogenetic diversity and taxonomic novelty of RNA viruses in the global WWTPs

RdRp-based megataxon clustering and phylogeny analysis supported the five established RNA virus phyla approved by the International Committee on Taxonomy of Viruses (ICTV) (Fig. 2). The majority of the RNA vOTUs that encoded "complete" RdRp domains (8755, 94.4%) can be posited into already established phyla (Supplementary Figs. 1-5). Moreover, 6307 (68.0%) RNA vOTUs could be classified into known viral families (Fig. 3a). The RNA vOTUs were most taxonomically assigned to *Fiersviridae* (1772, 28.1%) and *Steitzviridae* (580, 9.2%), both of which were known families infecting prokaryotes, followed by *Botourmiaviridae*, which was known to infect plants or fungi (Fig. 3a). To assess our expansion of knowledge on the diversity of RNA viruses in engineered ecosystems, the RNA viral contigs identified in this study were clustered (90% ANI and 80% AF%) with the RNA viral contigs identified in the latest global RNA virome analysis[12]. The known diversity of unique RNA vOTUs in engineered systems was almost doubled by this study, from 6004 previously known RNA vOTUs to a total of 13159 RNA vOTUs Extended Data Fig. 2). Our WWTP RNA vOTUs were also found to replenish the global taxonomical diversity of RNA viruses to some extent. For example, *Leviviricetes* was expanded from 9295 known RNA vOTUs to 10710 RNA vOTUs (15.2% increase), and *Cystoviridae* was expanded from 267 known RNA vOTUs to 337 RNA vOTUs (19.1%

increase). These two groups were notably the only two groups known to infect prokaryotes so far. *Birnaviridae*, the viral family that was known to infect vertebrates or insects, was almost tripled from 43 known RNA vOTUs to 128 RNA vOTUs. The identified RNA viruses also suggested new clade establishment. Among the RdRps placed outside of known family clades, a large amount of WWTP RdRps (n = 1350) and reference RdRps could only classified into *Kitrinoviricota* phylum without further assigned taxonomic information (Supplementary Fig. 4). These *Kitrinoviricota* RdRps made up two RdRp megataxa (7 and 16, Fig. 2b). RNA viral orders such as *Patatavirales* and *Picornavirales* of *Pisuviricota* also witnessed diverse RdRp clades outside of the known families (Supplementary Fig. 5). Besides, there were 4 novel clades mainly consisting of WWTP RdRps (Supplementary Figs. 7-10) and were posited under *Negarnaviricota* (Supplementary Fig. 2) in the global RdRp phylogenetic tree. WWTP novel clade 1 and 2 seemed to be the probably evolutionary intermediates between *Insthoviricetes* and *Elliovirícetes* (Supplementary Figs. 7-8), and WWTP novel clade 3 and 4 seemed to be novel families of *Bunyavirales* (Supplementary Figs. 9-10). Together, these findings resolved the phylogenetic diversity of RNA viruses in the WWTPs and called for the establishments of new RNA virus taxonomy to support future wastewater-based surveillance.

Tracking host associations of RNA prokaryotic viruses in the global WWTPs

Elucidating the RNA virus-host associations in WWTPs is crucial for advancing effective management strategies in the context of public health and environmental protection. Through host information available for viruses from established orthornaviran taxonomy, 3337 RNA vOTUs encoded “complete” RdRp domains (36.0%) were putative prokaryotic RNA viruses (*Leviviricetes* n = 3245 and

Cystoviridae n = 92), 1630 (17.6%) RNA vOTUs were RNA viruses probably infecting plants or fungi (Fig. 3a and Supplementary Table 2). The patterns of relative abundance of RNA vOTUs were different across sample types and sampling countries. Putative prokaryotic RNA vOTUs accounted for more than 60% average abundance proportion of RNA viral community in each WWTP compartment group (i.e., IN, DS, AS, EF) in Switzerland (Extended Data Fig. 3). Putative plants or fungi-associated RNA vOTUs dominated 277 influent and 14 effluent samples from the USA (Fig. 3b). In the activated sludge samples (n = 61) from all the sampled countries, the putative prokaryotic RNA vOTUs was the most abundant category in the majority (n = 53) of samples with no geographical restrictions (Fig. 3b). Anaerobic digester sludge samples showed a distinct higher proportion of RNA vOTUs that putatively infect protozoa compared with other sample types (Fig. 3b). Among these RNA viruses intricately living in the WWTPs, putative prokaryotic RNA viruses and human RNA viruses were most concerned for their roles in wastewater engineering and human health, respectively.

In complementary to the taxonomy-instructed host ranges, CRISPR spacer match and Shine-Dalgarno (SD) sequences were also co-used to predict putative prokaryotic RNA viruses (See Methods). Based on CRISPR spacer match (0 or 1 mismatch) of the Type-III CRISPR systems which can target RNA and appear possibilities to defend against RNA bacteriophages[25], 13 RNA vOTUs were predicted to associate with prokaryotic genomes encoding type-III cas proteins (Supplementary Table 3), exhibiting intriguing dual patterns. On the one hand, some RNA vOTUs were found to associate with the same host bacteria. For example, three RNA vOTUs (KB8_39385, JI8_57570, and JI8_173841) were predicted to infect *Halopseudomonas formosensis* under phylum *Proteobacteria* (Fig. 4), which was recently reported with

potential to degrade poly(ester-urethane) coatings[26]. The RNA vOTU KB8_39385 was from *Fiersviridae* and JI8_57570 was associated with SD motifs, both were probable bona fide prokaryotic viruses. Likewise, another two RNA vOTUs, i.e., AD-AS_64110 and HP20_1992, were predicted to both infect *Mycobacterium tuberculosis*, one of the most notorious human pathogens blamed as the causative agent of tuberculosis (Fig. 4). On the other hand, one single RNA vOTU was revealed to potentially associate with multiple host bacteria. For example, the RNA vOTU FD-DS_117671 was predicted to infect two representative species under *Campylobacterota*. The first one was *Aliarcobacter cryaerophilus*, a globally emerging foodborne and zoonotic pathogen. The second one was linked to a bacterial contig assembled from WWTP metagenomes and taxonomically assigned to *Sulfurimonas*, which functions to reduce nitrate, oxidize sulfur and hydrogen (Fig. 4). Likewise, one RNA vOTU from *Botourmiaviridae* (JI2_39922) and one RNA vOTU from *Fierviridae* (IL21_2901) both showed the probable ability to infect two different species under *Firmicutes* (Fig. 4). Notably, RNA vOTU IL35_83533 recovered from the anaerobic digester sample was predicted to infect *Methanosarcina*, a known coccoid methanogen archaeon (Fig. 4). Although there was no confirmed RNA archaeal virus so far, the association between RNA vOTU and archaeal genome containing RT-encoding Type-III CRISPR systems reflected the potential history of archaeal infection of this *Picornavirales*-assigned novel RNA virus species.

Besides CRISPR-spacer match, identification of SD sequences indicated 949 RNA vOTUs to be potential prokaryotic viruses, which mostly came from *Leviviricetes* (n = 769) and *Cystoviridae* (n = 36). Both CRISPR-spacer match and SD-sequence-based approaches predicted some novel prokaryotic RNA viruses beyond the known clades. For example, 36 RNA vOTUs from *Kitrinoviricota* and 12 RNA vOTUs from

Howeltoviricetes were associated with SD sequences, suggesting that novel prokaryotic clades may be present in these lineages (Supplementary Table 2). *Partitiviridae* and *Picobirnaviridae* were previously known to infect eukaryotes but were found to associated with SD sequences recently, with 12 and 5 cases, respectively, were also identified in this study (Supplementary Table 2). Besides, some RNA vOTUs from *Botourmiaviridae* and *Mitoviridae* that associated with plants or fungi were predicted with the ability to infect bacteria by both CRISPR-spacer and SD-sequences-based approaches (Fig. 4 and Supplementary Table 2), which expanded the knowledge on the host ranges for the viral families. Thus, these results indicated that the taxonomic range of prokaryotic RNA viruses was broader than currently recognized.

Prevalence, concentration, and fate of human RNA virus in the global WWTPs

Based on whole genome homology match, 21 RNA vOTUs were identified as putative human RNA viruses, including 8 *Astrovirus*, 7 *Norovirus* and 3 *Rotavirus* (Fig. 5a). These putative human RNA vOTUs were all viruses that may cause gastroenteritis in humans. RNA vOTUs of *Norovirus* were mainly detected in influent samples, while RNA vOTUs of *Astrovirus* and *Rotavirus* were detected in all sample types (Fig. 5b and Supplementary Table 4). To detect human RNA viruses more sensitively, we additionally searched clean RNA reads from part of metatranscriptomes (n = 59, Switzerland n = 47 and USA n = 12) against NCBI Virus Database and detected some other human RNA viruses in addition to the RNA viruses described above. *Respirovirus* that may cause respiratory tract illness was detected in all the examined samples, and Zika virus was detected in most (n = 57) examined samples (Extended Data Fig. 4). Influenza virus was detected in 6 samples from

Switzerland, and Hepatitis virus C and E were also detected in some samples (Extended Data Fig. 4). SARS-CoV-2 was detected in 5 influent samples from the USA, and three other subtypes of human coronavirus 229E, 4408, and OC43 were also detected in some samples (Extended Data Fig. 4). Reads-based diversity analysis separated the community structure of human RNA viruses in Switzerland and the USA (Extended Data Fig. 5), revealing the difference of health status and prevalent viruses between the corresponding populations.

Based on the absolute abundance estimation for human RNA viruses in WWTP samples from Switzerland, *Rotavirus* displayed the highest absolute concentration in the influent samples ($1.18 \times 10^{10} \pm 1.63 \times 10^{10}$ copies/L), followed by *Respirovirus* ($7.92 \times 10^{9} \pm 1.78 \times 10^{9}$ copies/L) and *Astrovirus* ($5.36 \times 10^{9} \pm 2.93 \times 10^{9}$ copies/L) (Supplementary Table 5). The frequently detected human RNA viruses in the influent samples, e.g., *Astrovirus*, *Husavirus*, *Norovirus*, *Respirovirus*, *Sapovirus*, and Zika virus, were significantly decreased but not completely eliminated in the effluent samples (Fig. 5b). Generally, as measured by copies per liter, the detected human RNA viruses were decreased from the influent to effluent samples by 2~3 orders of magnitude but were enriched in the sludge samples by one order of magnitude (Fig. 5b and Supplementary Table 5). Together, our results demonstrate that quantitative metatranscriptomic investigation of human RNA viruses in absolute amounts can well assist future wastewater-based surveillance for public health.

Ecological roles of RNA viruses in the WWTPs revealed by auxiliary metabolism

After gene function annotation, protein structure prediction and endogenous virus elements (EVE) evaluation, 183 RNA vOTU-encoding genes were identified as auxiliary metabolic genes (AMGs). The most common functional type of AMGs in

the RNA vOTUs was ribosomal protein (n = 22), including both prokaryotic ribosomal proteins (50S and 30S) and eukaryotic ribosomal proteins (60S and 40S) (Supplementary Table 6). The ribosomal protein-encoding RNA vOTUs can be detected in 28 out of 38 individual WWTPs, suggesting this was a common strategy for RNA viruses in the WWTPs. Six AMGs involved in cellular respiration and one AMG involved in pentose phosphate pathway, which were mainly encoded by *Leviviricetes*, showed potential involvement of RNA viruses in central carbon metabolism (Fig. 6). Three flagellin protein *Fla*A and one chemotaxis protein *Che*A suggested the potential viral assistance in cell motility (Supplementary Table 6). The nitrate reductase *nxr*A and nitrite reductase *nrf*A were two AMGs involved in nitrogen metabolism (Fig. 6). The RNA vOTU encoding *nxr*A was classified into *Cystoviridae* and thus seemed to be a prokaryotic RNA virus, while the RNA vOTU encoding *nrf*A was classified into *Picornavirales*, an order known to infect eukaryotes so far. It should be noted that 3 subtypes of antibiotic resistance genes (ARGs) against beta-lactam were identified in RNA vOTUs, including metallo-beta-lactamase, carbapenem resistance protein, and penicillin resistance protein (Fig. 6). The metallo-beta-lactamase was encoded by an RNA vOTU associated with SD motifs, and was only detected in the WWTP (TG) where it was identified. The carbapenem resistance protein was encoded by the RNA vOTU from *Bunyavirales* and was detected in 4 WWTPs (3 WWTPs from Switzerland and 1 WWTP from the USA). The penicillin resistance protein was encoded by the RNA vOTU from *Fierviridae* and seemed to be prevalent locally in Switzerland. It was detected in 8 out of 12 WWTPs from Switzerland. Another acriflavine resistance protein was found in a putative fungi-infecting RNA vOTU from *Narnaviridae* (Fig. 6), this gene may provide resistance to acriflavine, a chemical was used against African trypanosomiasis, and fungal

infections of fish. A series of other metabolic processes were also probably affected by RNA viruses, such as ATP-binding (n = 18), amino acid metabolism (n = 7), metabolism of cofactors and vitamins (n = 5), and iron-sulfur clustering (n = 4) (Supplementary Table 6). These identified AMGs may be responsible for diverse extra metabolic capabilities of RNA viruses, and most of the AMGs (n = 169) were first reported in the RNA viral genomes. The investigation of RNA virus-encoded AMGs offered a perspective into how viruses alter biochemical cycling in the WWTPs, considering their potential effects on host energy metabolism, genetic information processing, and antimicrobial resistance.

Discussion

This study constructed the largest catalog of RNA viruses in global WWTP systems to date, resulting in the establishment of the WWTP RNA Viruses Database (WRVirDB). This endeavor effectively doubled the known diversity of RNA viruses within engineering systems, and greatly advanced our comprehension of phylogeny, host range, and auxiliary metabolic capabilities of RNA viruses in WWTPs. The comprehensive depiction of RNA viral diversity and functionality uncovered the biochemical impacts and health implications of RNA viruses in the WWTPs, and highlighted the significance of RNA viruses as a previously overlooked but important ecological driver in global wastewater engineering systems.

The RdRp-based phylogeny well illustrated the major branches of RNA viruses and the position of WWTP RNA vOTUs on the global viral RdRp phylogenetic tree. The 5 major branches of RdRp domains in this study agreed with the previous study[11, 12, 27], *Lenarviricota* took the basal position, *Pisuviricota* and *Kitrinoviricota* were placed on the top branch (Fig. 2). Although the previous study[12] found that *Flasuviricetes* may

be an “offender” and was placed inside *Pisuviricota*, both megataxon clustering and phylogenetic analysis suggested *Flasuviricetes* should be affiliated into *Kitrinoviricota*, which was consistent with other previous studies[10, 11]. *Cystoviridae* was discovered to be a plausible “offender” again[12], the whole members from the megataxon of *Cystoviridae* were away from their respective phyla *Duplornaviricota* but posited inside *Pisuviricota*. Considering that about 95% of RNA vOTUs can be classified into established 5 phyla, the putative additional phylum was not discussed in this study. However, several clades were probably to generate putative additional classes (especially clades inside *Kitrinoviricota*), orders (for example, WWTP novel clade1, WWTP novel clade2 and clades inside *Pisoniviricetes* and *Stelpaviricetes*) and families (such as WWTP novel clade3, WWTP novel clade4, and clades inside *Bunyavirales*, *Ghabrivirales*, *Picornavirales*, and *Patatavirales*). Taking advantage of the explored divergent RNA viruses in the WWTPs, the global RdRp tree was improved and promoted.

To unravel the diversity of prokaryotic RNA viruses and potential associations between RNA viruses and both functional niches and bacterial human pathogens in the WWTPs, the prokaryotic RNA viruses was predicted were shown to be diverse and abundant in the WWTPs. The putative prokaryotic RNA viruses from known RNA phage clades (*Leviviricetes* and *Cystoviridae*) were found to contribute to more than a third of the richness diversity of RNA vOTUs in the WWTPs. The high diversity of putative prokaryotic RNA viruses was reasonable, as bacteria were the most numerous organisms living in the WWTPs. When compared the distribution of RNA viruses between WWTPs and other ecosystems[12], the RNA virome compositions were found to be typical in each ecosystem, possibly due to the variations in the host organism compositions. Putative prokaryotic viruses were abundant in the engineered

and soil samples, while the RNA viruses in the host-associated and natural water samples were more likely to be eukaryotes-associated (Extended Data Fig. 6). In addition to the putative prokaryotic viruses within known clades as discussed above, several novel putative bacterial and even archaeal viruses were predicted by CRISPR spacer match or SD sequences identification. Our WWTP RNA virome study and several recent studies[12, 13] both revealed that the diversity of RNA bacteriophage may be underestimated. Through the predicted specific RNA virus-bacteria host associations, some bacterial hosts exhibited potential pathogenicity and function in nutrients removal indicated by their taxonomy or genetic information (Supplementary Table 4). The unveiled genomic sequences and phylogenetic features of "desirable" (i.e., those infect human pathogen) and "undesirable" (i.e., those infect functional niches) prokaryotic RNA viruses contributed to future phage therapy research and facilitated the improvement of sewage management practices for RNA viruses as emerging contaminants.

Wastewater-based surveillance has the potential to effectively monitor transmissions of infectious diseases, especially when applied to track the trends of SARS-CoV-2 variants[20, 28]. As we passed through the outbreak period of SARS-CoV-2, wastewater surveillance coupled with designed sampling campaigns and advanced culture-independent technologies can continue to provide genomic detection and quantification information for human RNA viruses that can shed into sewage. Our detection of SARS-CoV-2 and other human RNA viruses that caused several kinds of illness, such as gastroenteritis, influenza, hepatitis, and zika fever in the WWTPs at either reads-level or assembly-level confirmed a wide range of detectable human viruses through metatranscriptomes. The assembly-based approach identified more complete human RNA viral genomes that assigned to 17 unique organisms, while

reads-based alignments identified a broad detection of 264 unique organisms of human RNA viruses in the global WWTPs. Additionally, the wastewater surveillance of human RNA viruses in spatial-temporal sample series and samples from different WWTP compartments can reflect the unique distribution, dynamic trends, and fate of human RNA viral community throughout the wastewater treatment process in different regions. The absolute quantification of human RNA viruses in some WWTP samples further assisted in accurate abundance estimation of human RNA viruses instead of relative abundance estimation which may lead to a biased abundance trend[29]. Our efforts exhibited the viability of wastewater-based surveillance using metatranscriptomes, enabling both detection and tracking of human RNA viruses. Actually, wastewater-based surveillance has been reported to offer less biased, more comprehensive, timely, and cost-effective monitoring of SARS-CoV-2 compared to clinical-based surveillance methods[20]. If future spatial–temporal viral genomic analyses of wastewater samples will be available, such surveillance could provide both objective health status of the population and early warning of infectious diseases locally and globally.

Viruses influence biogeochemical cycling not only through infecting and lysing their hosts but also by reprogramming the specific metabolisms of host cells[30]. Virus-encoded AMGs are widely discovered in dsDNA viruses and are ecologically important in biogeochemistry[31-33], whereas auxiliary metabolic capabilities are less well-studied in RNA viruses. Our exploration of gene functions of RNA vOTUs revealed that RNA viruses also encoded diverse kinds of AMGs, which may provide RNA viruses with evolutionary advantages during infection. RNA viruses can promote energy production of host cells by own-encoding cellular respiration proteins and ATP-binding/hydrolyzed proteins as examples. RNA viruses-encoded ribosomal

proteins, translation initiation factor, and elongation factors may facilitate the efficiency of translation. The extra produced energy and stimulated translations, as well as additional viral assistance in other basic metabolic processes (e.g., RNA vOTUs encoded several oxidoreductases, peptidases, and transferases) provided the potential for better RNA viral proliferations, which has been reported as strategies for DNA viruses[33, 34] but mostly first reported in the RNA viral genomes[6]. However, RNA viruses showed a limited participation in element cycling in WWTPs by related AMGs, only two AMGs involved in nitrogen metabolism were identified, without any sulfur metabolism proteins or phosphate regulation proteins, which was dissimilar with DNA viruses in WWTP systems[21]. DNA viruses has been regarded as important reservoirs of ARGs[35]. Surprisingly, RNA viruses were also found to encode clinically important beta-lactamase genes, suggesting their potential involvement in the process and even transmission of clinically important antibiotic resistance. Given that the AMGs were acquired by horizontal gene transfer events between viruses and host genomes during infection, the identification of AMGs not only indicated the probable biochemical roles of RNA viruses in the WWTPs, but also suggested the probable host of AMG-encoding RNA viruses. The types of ribosomal proteins were highly consistent with viral taxonomy-based host ranges, i.e., 40S and 60S ribosomal proteins were found in RNA vOTUs that may infect eukaryotes, and 30S and 50S ribosomal proteins were found in RNA vOTUs that may infect prokaryotes (Supplementary Table 2). In turn, the identified encoding patterns of AMGs can be additional clues to predict probable hosts for RNA viruses. For example, the RNA vOTUs encoded ARGs and bacterial motility proteins should be potential prokaryotic viruses.

In conclusion, this study not only extensively explored the most comprehensive

diversity of RNA viruses in the global WWTPs to date, but also represented major advances: (1) confirmed the high diversity and abundance of prokaryotic viruses within RNA virome in WWTP systems. (2) unravel the roadmap for the distribution of human RNA viruses in global WWTPs. (3) emphasize the diversity and biochemical functions introduced by RNA virus-encoded AMGs, particularly focusing on antibiotic resistance genes (ARGs). These major findings underscore the implications of RNA viruses for both human health and the wastewater treatment process. However, the global diversity of RNA viruses in WWTPs still needs to be further explored in the future for two reasons. First, the identification of RNA viruses in this study was based on the detection of RdRp, the universal marker gene of *Orthornavirae* kingdom. Thus, the non-RdRp coding fragments of RNA viruses would not be reported. Second, insufficient data volume of sequencing for identifying RNA viruses in WWTPs and unavoidable lost signals during bioinformatic assembly and annotation would affect the recovery of RNA viruses. Nonetheless, the explored genomic diversity, host associations and gene functions of RNA viruses in the WWTPs were meaningful and valuable data source for future studies and experiments. This study evidenced that RNA viruses play essential ecological roles and showed important health implications in WWTPs.

Methods

Metatranscriptome collection

To build the first catalogues of RNA virus of global WWTPs, a total of 557 available metatranscriptomes of WWTP samples were downloaded from the public databases including National Center for Biotechnology Information Sequence Read

Archive (NCBI SRA), Metagenomic for Rapid Annotations using Subsystems Technology (MG-RAST) and National Genomics Data Center (NGDC) to identify RNA viruses. These metatranscriptomic samples included 289 influent, 80 denitrification sludge, 61 activated sludge, 26 effluent, and 101 anaerobic digester sludge samples from 9 countries (Supplementary Table 1). The 47 samples from Switzerland were spiked with mRNA internal standards in the previous study[29], which enabled the absolute concentration estimation of RNA viruses in these samples. The detailed information for these public metatranscriptomes were listed in Supplementary Table 1.

Quality control, metatranscriptomic assembly and gene prediction

The raw sequencing reads were trimmed by Trimmomatic v0.39[36] with default parameters. MEGAHIT[37] has been regarded as a successful assembly tool for metatranscriptomes with efficient memory and computational time[10]. The trimmed reads of each metatranscriptome were assembled into contigs using MEGAHIT v1.1.3[37] with default parameters. Only contigs > l kbp were used for the downstream analysis. Genes were predicted from assembled contigs using Prodigal v2.6.3[38] with option “-p meta”.

Identification and evaluation of RNA viruses

The methodology of RNA virus identification referred to a previous study which systemically searched RNA viruses in the global ocean metatranscriptomes[10]. In detail, RNA viruses were identified by iterative HMM searches against virus RdRp HMM profiles (Extended Data Fig. 1). The initial 65 RdRp HMM profiles were downloaded from (doi:10.5281/zenodo.5731488)[10]. For every iteration, all the protein sequences

predicted from the assembled contigs, along with the known virus RdRp protein sequences from NCBI Genbank release 253 (n = 3694787), and recent RNA virus studies from ocean (n = 44828[10] and n = 4593[11]) and global metatranscriptomes from diverse ecosystems (n = 378253)[12] were searched against the RdRp HMM profiles by hmmsearch (v3.3.2)[39] with the option "-A", which supported to trim the portion aligned to the RdRp HMM profile. The hit protein sequences with a bit-score ≥ 30 and with ≥ 70% aligned portions to the best RdRp HMM profiles were retained. The aligned portions of hit protein sequences were clustered into updated RdRp HMM profiles by vFam pipeline[40] with default parameters. The RdRp profile HMMs were generated and updated for eight iterations to meet the near saturated number of identified virus RdRps. A total of 56086 putative virus RdRps and 455 RdRp HMM profiles were identified after eight search-and-updated iterations. The contigs encoded putative RdRps were regarded as putative RNA viral contigs.

To evaluate the authenticity of the putative RNA viruses, false-positive RNA viruses were filtered by 1) a competitive hmmsearch approach[10] and 2) mapping against corresponding metagenomic datasets. First, 56086 putative RdRps were searched against the database merged by PfamA (v33.0)[41] HMM profiles and 455 RdRp HMM profiles constructed in this study. If the putative RdRp showed a best match to a non-RdRp HMM profile, it was considered as a false-positive virus RdRp. All of the 56086 putative RdRps passed the competitive hmmsearch test, and they all showed a best match to the RdRp HMM profile. Then, based on the hypothesis that RNA viruses (under *Orthornavirae* kingdom) would not be present in the DNA fraction, the occurrence of putative RNA viral contigs in the metagenomes were examined and used to evaluate false-positive RNA viruses. We downloaded additional 134 metagenomes originated from the samples same as part of the WWTP

metatranscriptomes. The raw DNA sequences were mapped to the putative RNA viral contigs by Bowtie2 v2.3.5.1[42] in "--very-sensitive" mode. If >20% of the positions in the putative RNA viral contig were mapped by DNA reads in any metagenome, this RNA viral contig was considered as a false-positive RNA virus. In total, 165 putative RNA viral contigs were identified as false-positive RNA viruses and were discarded in this step.

To examine the completeness of virus RdRp domains, RNA viral contigs were translated into protein sequences by transeq (EMBOSS:6.6.0.0)[43] with all six frames and the standard translational code. The translated protein sequences were searched against the 455 RdRp HMM profiles once more, and the sequences with a bit-score ≥ 30 and with ≥ 90% aligned portions to the best RdRp HMM profiles were considered as RNA viruses containing "complete" RdRp domains. Together, 55900 RNA viral contigs were identified from 557 WWTP metatranscriptomes, including 44399 RNA viral contigs contained "complete" RdRp domains.

RNA vOTUs clustering

The RNA viral contigs were clustered into RNA vOTUs of 90% ANI across 80% of AF (parameters referred to reference[10], the scripts used for RNA vOTU clustering were downloaded from https://bitbucket.org/berkeleylab/checkv/src/master/scripts/). The clustered RNA vOTUs were pragmatically at approximately species level[10]. In total, 55900 RNA viral contigs were clustered into 11414 RNA vOTUs, including 9276 RNA vOTU representatives contained "complete" RdRp domains.

RdRp-based taxonomy classification

The taxonomic megataxon of RNA vOTUs were determined by a network-based clustering method[10]. The public RdRp reference protein sequences that matched to RdRp HMM profiles after eight hmmsearch iterations were collected and pre-clustered at 50% amino-acid identity by Uclust[44] with options “usearch –cluster_fast -id 0.5 -sort length”, and resulted in 30504 public RdRp centroids. Then, 30504 public RdRp centroids along with 9276 WWTP RNA vOTUs containing “complete” RdRp domains were merged and used for clustering megataxon of RNA viruses. Pairwise comparisons of these merged RdRp sequences were conducted by blastp v2.10.0 with options “-gapopen 9 -gapextend 1 -word_size 3 -threshold 10”. E-values were extracted and negative-log10-transformed by “mcxload” in MCL v14-137[45] with options “--stream-mirror --stream-neg-log10 -stream-tf 'ceil(200)' ”. Transformed e-values were used for MCL[45] network clustering with option “-I 0.9” to establish RdRp megataxon (approximately at class to order level). In total, 22 RdRp megataxon were clustered and labeled by the public reference RdRp centroids with known taxonomic classification within the megataxon. The accuracy of the label for each megataxon was calculated as the number of public RdRp centroids affiliated with the taxonomy divided by the total number of known public centroids (those with labels at phylum level at least) in the megataxon. For example, megataxon 1 comprised 5676 RdRps in total, with 714 RdRps identified in this study and 4962 public RdRp centroids. Among the 4962 public RdRp centroids, 4482 RdRps could be taxonomically classified into phylum level or higher, 4257 RdRps could be taxonomically classified as *Picornavirales* (except for *Marnaviridae*) or Stelpviricetes, the accuracy of the label for this megataxon was calculated as 4257 divided by 4482. One megataxon (megataxon 22, n = 68) were mainly consisted of RdRps identified in this study (n = 56) with only two known reference RdRps. This megataxon was regarded as novel

megataxon and was not labeled.

RdRp-based phylogenetic analysis

The public RdRp centroids, WWTP RNA vOTUs containing "complete" RdRp domains, along with the group II introns and non-LTR retrotransposons were aligned and used to generate the approximate global tree for viral RdRps. In detail, the protein sequences of group II introns and non-LTR retrotransposons were downloaded from Genbank Release 253. The RdRps within megataxon 18 were affiliated with viral families of *Birnaviridae* and *Permutotetraviridae*, which encoded conserve motifs with the order permutation "C-A-B" instead of the canonical permutation "A-B-C". Such RdRp permutations can make RdRp sequences unalignable with those of the canonical configuration[11]. Therefore, the RdRp sequences of megataxon 18 (n = 388) were excluded from the analysis of global RdRp tree construction and analyzed individually. In total, public RdRp centroid sequences (n = 30217), group II introns (n = 950) and non-LTR retrotransposons (n = 83), and WWTP RNA vOTUs containing "complete" RdRp domains (except for RdRps of megataxon 18, n = 9175), were aligned by MUSCLE[46] under "-super5" mode. The aligned sequences were trimmed by Trimal[47] with sites having more than 20% gaps removed. The alignment of trimmed sequences was used to construct an approximate maximum likelihood tree by the FastTree[48] with options "-wag -gamma". The constructed tree was illustrated in iTOL[49]. The tree was rerooted between RdRps and RTs. The RdRp-based sub-trees for each RNA virus phylum (Supplementary Figs. 1-5) were pruned from the global tree. The family-level taxonomic classification for RdRps was assigned based on the placement of the sequences in the tree, requiring the RdRps fell within a clade to be assigned to a specific family. The family-level taxonomy classification should be

consistent in both methods of megataxon clustering and phylogenetic analysis.

Functional gene annotation of RNA vOTUs

The alternative translation code usage of RNA vOTUs was checked before functional annotations. The open reading frames (ORFs) of RNA vOTUs were predicted under all available codes with Prodigal v2.6.3[38]. The codon producing the longest average length of ORFs was chosen. Then, the functional annotation of predicted ORFs was determined by one iteration of hhblits[50] with options "-n 1 -e 0.001" using UniRef30_2023_02 database (https://wwwuser.gwdg.de/~compbiol/uniclust/2023_02/UniRef30_2023_02_hhsuite.tar.gz). Only hits with >95% probability score were used for functional annotation. The predicted ORFs were also searched against the NCBI nr database by blastp with a bitscore > 50 were retained.

The RNA virus-encoded genes were regarded as AMGs if they met the following requirements: (i) both hhblit topmatch in the Uniref database and blastp topmatch in the NCBI nr database were originated from cellular organism. (ii) the AMG sequences were not derived from potential endogenous viral elements (EVEs). The RdRp sequences from probable AMG-encoding RNA viral contigs were blastn against all cellular genomes (> 50kb) in the NCBI Refseq database, and the thresholds for predicting EVEs were set to 300 nucleic acids for alignment length and 1×10^{-20} for e-value. (iii) The predicted protein structure of each probable AMG was consistent with the homology-based annotation. The probable AMG sequences were annotated with Phyre2[51], and the thresholds were set to 30% for alignment coverage and 90% for confidence.

Host prediction of RNA vOTUs

Virus-host relationships were predicted based on four approaches. (i) sequences similarity to known RNA viruses, (ii) CRISPR spacer match, (iii) SD motif prediction, and (iv) taxonomy-based prior known host range.

For sequences similarity to known RNA viruses, the complete nucleotides of reference RNA viral genomes (within *Orthornavirae*) were downloaded from NCBI Virus database (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/) and used as a nucleotide database. The 11414 RNA vOTUs were searched against the nucleotide database by blastn. The RNA vOTU was considered to be the same species as the reference RNA virus with the thresholds of 90% blastn identity and 80% alignment length of the RNA vOTU. Then, the host of the RNA vOTU was also referred to the known host information of the matched reference RNA virus. Additionally, to further examine human RNA viruses more sensitively, 59 metatranscriptomes including 47 samples from Switzerland and 12 samples from USA were selected for reads-based searching. The clean RNA reads were blastn against complete nucleotides of human RNA viruses in NCBI Virus database. The reads matched to reference human RNA viral genomes with ≥ 90% identity along ≥ 80% reads length were considered as fragments of the related human RNA viral species. The number of matched reads were used for detect occurrence and abundance estimation of the human RNA viruses in the samples. Only human RNA viral organisms with a total read count > 10 in the samples were retained for analyses.

For CRISPR spacer match, RNA vOTUs were compared to prokaryotic CRISPR spacer sequences to predict virus-prokaryote associations. CRISPR spacers were predicted for all > 10 kb reference bacterial and archaeal genomes downloaded from

NCBI Refseq database (https://ftp.ncbi.nlm.nih.gov/genomes/refseq/) with CRT[52]. To link RNA vOTUs to predicted CRISPR spacers, only blastn-short hits with >94% identity and with 0 or 1 mismatch over the whole spacer were considered. Then, the existence, type and location of CRISPR systems in the potential host genomes were checked based on the GenPept files of reference hosts. We also used the same approach to compare RNA vOTUs to CRISPR spacers predicted from metagenomic contigs assembled from 134 corresponding WWTP metagenomes. The DNA contigs were assembled by WWTP metagenomes using metaSPAdes[53] with default parameters. The taxonomic classification of host DNA contigs was determined by Kaiju[54]. The ORFs from host DNA contigs were predicted by Prodigal v2.6.3[38] and annotated by hhblit[50] with options "-n 1 -e 0.001" using UniRef30_2023_02 database. Only hits with probability >95% were retained. The predicted vOTU-host pairs were considered if the host genome encoded Type III CRISPR systems.

For SD motif prediction, ORFs calling was run as described above (see Functional gene annotation of RNA vOTUs), and the information of "rbs_motif" for each predicted ORF was extracted from Prodigal output files. The canonical SD motifs were identified based on the reference list from the previous study[12]. Then, the vOTUs with ≥ 10 predicted ORFs were examined for the proportion of ORFs with SD motifs. One RNA vOTU was considered as probable prokaryotic RNA virus if ≥ 50% of the predicted ORFs were associated with an SD motif.

For taxonomy-based method, the taxonomic classification for RNA vOTUs (see RdRp-based taxonomy classification and RdRp-based phylogenetic analysis) were used to retrieve known host range of each family.

Quantitative profiling and diversity analysis of RNA vOTUs

To calculate relative abundance of RNA vOTUs in each sample, clean reads from each metatranscriptome were first trimmed off polyA and polyp stretches using bbduk v37.62 (https://jgi.doe.gov/data-andtools/bbtools/) with options "literal=AAAAAAAAAA,TTTTTTTTTT, hdist=2 k=10 minlength=30, ktrim=r restrictright=20; ktrim=l restrictleft=20" for three times. WWTP RNA viral contigs were also trimmed off polyA and polyT stretches. PolyA/T-trimmed clean reads in each metatranscriptome were individually mapped against polyA/T-trimmed RNA viral contigs using Bowtie2 v2.3.5.1[42] with options "--local -D 20 -R 3 -N 0 -L 16 -i S,1,0.50 --very-sensitive --non-deterministic -I 0 -X 2000"[10]. The relative abundance of RNA vOTUs in each sample was calculated by CoverM v0.6.1 (https://github.com/wwood/CoverM) with options "--min-read-percent-identity 90 --min-read-aligned-percent 75 -m trimmed_mean --min-covered-fraction 30", which recruited reads that mapped with ≥ 90% identity ≥ 75% of the read length and restricted RNA viral contigs should be covered ≥ 30% length horizontally by mapped reads. The calculated Tmean abundances for RNA viral contigs were normalized by the number of mapped reads as filtered by CoverM. The final relative abundance of RNA vOTUs were calculated by summing the normalized abundances of contigs belonging to the RNA vOTUs.

To estimate absolute abundance of RNA vOTUs in 47 metatranscriptomes from Switzerland, the statistical information of spiked mRNA internal standards as referred to reference[29] was used in this study. The absolute abundance of RNA vOTU in each sample was measured in both "copies/g-VSS" and "copies/L" rank using the equation 1 and 2, respectively.

$$\text{Absolute abundance (copies/g-VSS)} = \frac{N_{spiked\ RIS}}{m_{biomass}} \times \frac{N_{mapped\ reads}/L_{vOTU}}{N_{RIS\ reads}/L_{RIS}} \quad (1)$$

$$\text{Absolute abundance (copies/L)} = \frac{N_{spiked\ RIS}}{V_{sample}} \times \frac{N_{mapped\ reads}/L_{vOTU}}{N_{\ RIS\ reads}/L_{RIS}} \quad (2)$$

$N_{spiked\ RIS}$ was the copy number of spiked mRNA internal standards in the sample, $m_{biomass}$ was the mass of collected volatile suspended solids of the sample, $N_{mapped\ reads}$ was the number of reads mapped to the RNA vOTU in the metatranscriptome, L_{vOTU} was the length of the RNA vOTU, $N_{RIS\ reads}$ was the number of reads mapped to the mRNA internal standards in the metatranscriptome, L_{RIS} was the length of the mRNA internal standards.

Similarly, the above formula was also applied to estimate the absolute concentration of detected human RNA viruses in 47 samples from Switzerland at reads level, using the length of matched reference genome to replace the length of vOTU in the formula.

The α-diversity of RNA vOTUs in the samples was measured with the Shannon's index calculated by function "diversity" in "vegan" package[55] in R. The β-diversity of RNA vOTUs in the samples was measured using the Bray-Curtis dissimilarity method calculated by function "vegdist" in "vegan" package[55] in R, using the log-transformed relative abundance table of RNA vOTUs in the samples as input. Then, the Non-Metric Multidimensional Scaling (NMDS) analysis was performed using function "metaMDS" in R package "vegan"[55]. The analysis of Similarity (ANOSIM) was performed using function "anosim" in R package "vegan". The difference of Shannon index H and absolute concentration of RNA vOTUs between pairwise compartments from Switzerland was determined by the Mann−Whitney U test using the function

"wilcox.test" with option "paired = FALSE" in R.

Data availability

All the sequences of RNA vOTUs and RdRps identified in this study and the detail information of RNA vOTUs were available at https://github.com/emblab-westlake/WRVirDB. The accession number in a publicly available database will be provided before manuscript submission.

References

1. Wu, D., Wu, T., Liu, Q. & Yang, Z. The SARS-CoV-2 outbreak: What we know. *International Journal of Infectious Diseases* **94**, 44-48 (2020).

2. Nimgaonkar, I., Ding, Q., Schwartz, R.E. & Ploss, A. Hepatitis E virus: advances and challenges. *Nature Reviews Gastroenterology & Hepatology* **15**, 96-110 (2017).

3. Hall, A.J. et al. Norovirus Disease in the United States. *Emerging Infectious Diseases* **19**, 1198-1205 (2013).

4. Moreno-Altamirano, M.M.B., Kolstoe, S.E. & Sánchez-García, F.J. Virus Control of Cell Metabolism for Replication and Evasion of Host Immune Responses. *Frontiers in Cellular and Infection Microbiology* **9** (2019).

5. Gualdoni, G.A. et al. Rhinovirus induces an anabolic reprogramming in host cell metabolism essential for viral replication. *Proceedings of the National Academy of Sciences* **115** (2018).

6. Dominguez-Huerta G, Zayed A A, Wainaina J M & al., e. Diversity and ecological footprint of Global Ocean RNA viruses. *Science* **376**, 1202-1208 (2022).

7. Shi, M. et al. The evolutionary history of vertebrate RNA viruses. *Nature* **556**, 197-202 (2018).

8. Shi, M. et al. Redefining the invertebrate RNA virosphere. *Nature* **540**, 539-543 (2016).

9. Kamitani, M., Nagano, A.J., Honjo, M.N., Kudoh, H. & Kümmerli, R. RNA-Seq reveals virus–virus and virus–plant interactions in nature. *FEMS Microbiology Ecology* **92** (2016).

10. Zayed A A, Wainaina J M, al., e. & Sullivan, M.B. Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. *Science* **376**, 156-162 (2022).

11. Wolf, Y.I. et al. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat Microbiol* **5**, 1262-1270 (2020).

12. Neri, U. et al. Expansion of the global RNA virome reveals diverse clades of bacteriophages. *Cell* **185**, 4023-4037 e4018 (2022).

13. Chen, Y.-M. et al. RNA viromes from terrestrial sites across China expand environmental viral diversity. *Nature Microbiology* **7**, 1312-1323 (2022).

14. Callanan, J. et al. Expansion of known ssRNA phage genomes: From tens to over a thousand. *Science Advance* **6**, eaay5981 (2020).

15. Zhang, Y.-Z., Chen, Y.-M., Wang, W., Qin, X.-C. & Holmes, E.C. Expanding the RNA Virosphere by Unbiased Metagenomics. *Annual Review of Virology* **6**, 119-139 (2019).

16. Starr, E.P., Nuccio, E.E., Pett-Ridge, J., Banfield, J.F. & Firestone, M.K. Metatranscriptomic reconstruction reveals RNA viruses with the potential to shape carbon cycling in soil. *Proc Natl Acad Sci U S A* **116**, 25900-25908 (2019).

17. Zhang, Y., Wang, Y., Tang, M., Zhou, J. & Zhang, T. The microbial dark matter and "wanted list" in worldwide wastewater treatment plants. *Microbiome* **11** (2023).

18. Ju, F., Xia, Y., Guo, F., Wang, Z. & Zhang, T. Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed wastewater treatment plants. *Environ Microbiol* **16**, 2421-2432 (2014).

19. Chen, Y., Wang, Y., Paez-Espino, D., Polz, M.F. & Zhang, T. Prokaryotic viruses impact functional microorganisms in nutrient removal and carbon cycle in wastewater treatment plants. *Nat Commun* **12**, 5398 (2021).

20. Karthikeyan, S. et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature* **609**, 101-108 (2022).

21. Yuan, L. & Ju, F. Potential Auxiliary Metabolic Capabilities and Activities Reveal Biochemical Impacts of Viruses in Municipal Wastewater Treatment Plants. *Environmental Science & Technology* **57**, 5485-5498 (2023).

22. Shi, L.D. et al. A mixed blessing of viruses in wastewater treatment plants. *Water Res* **215**, 118237 (2022).

23. Rothman, J.A. et al. RNA Viromics of Southern California Wastewater and Detection of

SARS-CoV-2 Single-Nucleotide Variants. *Applied and Environmental Microbiology* **87** (2021).

24. Yousif, M. et al. SARS-CoV-2 genomic surveillance in wastewater as a model for monitoring evolution of endemic viruses. *Nature Communications* **14** (2023).

25. Makarova, K.S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* **18**, 67-83 (2019).

26. de Witt, J. et al. Biodegradation of poly(ester urethane) coatings by Halopseudomonas formosensis. *Microbial Biotechnology* **17** (2023).

27. Wolf, Y.I. et al. Origins and Evolution of the Global RNA Virome. *mBio* **9** (2018).

28. Hemalatha, M. et al. Surveillance of SARS-CoV-2 spread using wastewater-based epidemiology: Comprehensive study. *Science of The Total Environment* **768** (2021).

29. Ju, F. et al. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic exchange, and upregulated expression in the effluent microbiomes. *ISME J* **13**, 346-360 (2019).

30. Pratama, A.A. et al. Expanding standards in viromics: in silico evaluation of dsDNA viral genome identification, classification, and auxiliary metabolic gene curation. *PeerJ* **9**, e11447 (2021).

31. Roux, S. et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689-693 (2016).

32. Gazitua, M.C. et al. Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. *ISME J* **15**, 981-998 (2021).

33. Rosenwasser, S., Ziv, C., Creveld, S.G.V. & Vardi, A. Virocell Metabolism: Metabolic Innovations During Host-Virus Interactions in the Ocean. *Trends Microbiol* **24**, 821-832 (2016).

34. Hurwitz, B.L. & U'Ren, J.M. Viral metabolic reprogramming in marine ecosystems. *Curr Opin Microbiol* **31**, 161-168 (2016).

35. Debroas, D. & Siguret, C. Viruses as key reservoirs of antibiotic resistance genes in the environment. *ISME J* **13**, 2856-2867 (2019).

36. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

37. Li, D., Liu, C.M., Luo, R., Sadakane, K. & Lam, T.W. MEGAHIT: an ultra-fast single-node

solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674-1676 (2015).

38. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

39. Finn, R.D., Clements, J. & Eddy, S.R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29-37 (2011).

40. Tse, H., Skewes-Cox, P., Sharpton, T.J., Pollard, K.S. & DeRisi, J.L. Profile Hidden Markov Models for the Detection of Viruses within Metagenomic Sequence Data. *PLoS ONE* **9** (2014).

41. Finn, R.D. et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279-285 (2016).

42. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).

43. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open software suite[<emboss.pdf>. *Trends in genetics* **16**, 276-277 (2000).

44. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461 (2010).

45. Enright, A.J., VAN, D.S. & Ouzounis, C.A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**, 1575-1584 (2002).

46. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).

47. Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).

48. Price, M.N., Dehal, P.S. & Arkin, A.P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

49. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127-128 (2007).

50. Steinegger, M. et al. HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20** (2019).

51. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J. The Phyre2 web portal

for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845-858 (2015).

52. Bland, C. et al. CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**, 209 (2007).

53. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P.A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* **27**, 824-834 (2017).

54. Menzel, P., Ng, K.L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* **7** (2016).

55. Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**, 927-930 (2003).

Acknowledgements

We thank Mr. Guoqing Zhang and Ms. Xinyu Huang at Westlake University for their helpful suggestions and technical supports. We thank the Westlake University HPC Center for computation support.

Author contributions

F.J. designed and supervised the study. F.J. and L.Y. wrote the manuscript. L.Y. performed the bioinformatics analysis, statistical analysis, and data visualization

Competing interests

The authors declare no competing interests

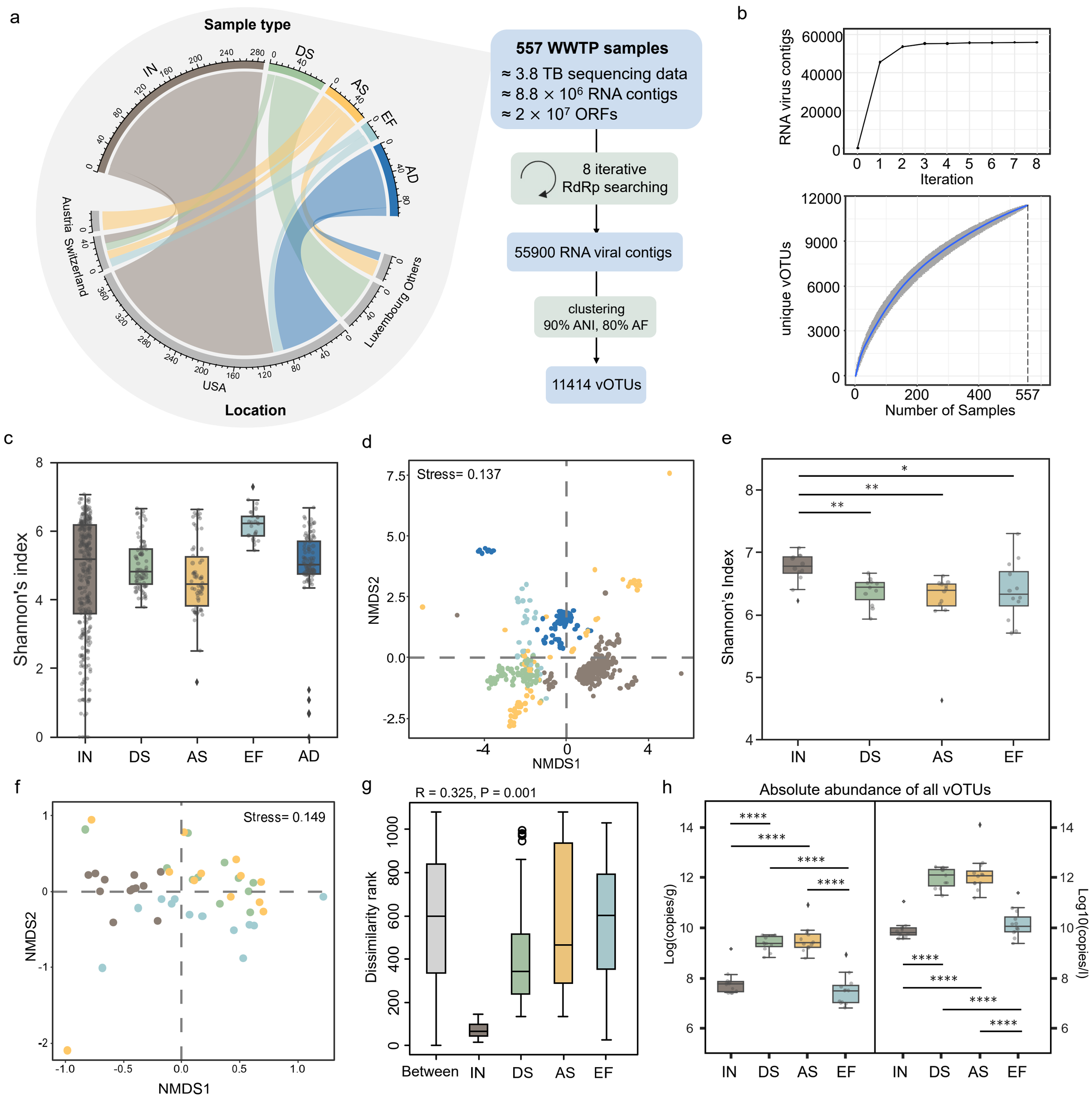


Fig. 1 Exploring RNA viruses in the global WWTP metatranscriptomes. a. Distribution of sample types and locations of 557 publicly available WWTP metatranscriptomes and the brief bioinformatic workflow for RNA vOTUs identification. b. Number of RNA viral contigs detected in the eight hmmsearch iterations (up) and accumulation curve of unique RNA vOTUs detected in the samples (down). c. α-diversity of RNA vOTUs measured in Shannon's index across 5 sample types in all WWTP samples. d. NMDS analysis of abundance distribution o f RNA vOTUs in all WWTP samples. e. α-diversity of RNA vOTUs measured in Shannon's index across 4 sample types in the 47 Swiss WWTP samples. f. NMDS analysis of abundance distribution of RNA vOTUs in 47 WWTP samples from Switzerland. g. ANOSIM analysis showed the dissimilarity rank of RNA vOTUs composition between and within different WWTP compartments from Switzerland. h. absolute abundance measured in copies per gram (left) and copies per liter (right) of all identified RNA vOTUs across 4 sample types in the 47 Swiss WWTP samples. Statistical significance was calculated by the Mann-Whitney U test and illustrated in e and h as follows: * for adjusted-P < 0.05, ** for adjusted-P < 0.01, and **** for adjusted-P < 0.00001. IN: influent, DS: denitrification sludge, AS: activated sludge, EF: effluent, AD: anaerobic digester sludge

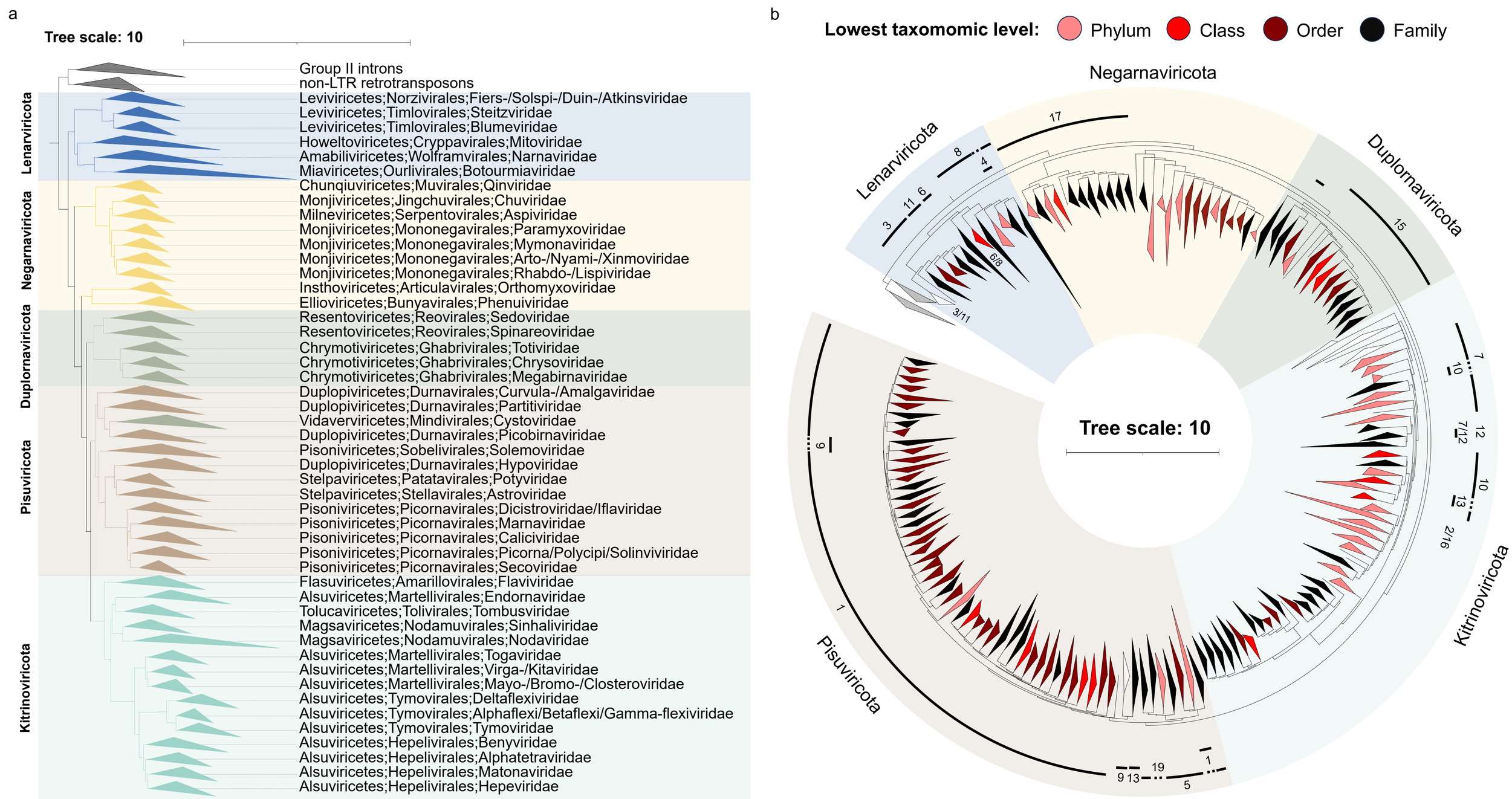


Fig. 2 Phylogenetic analysis and megataxa clustering of RdRps revealed the diversity of RNA viruses in the WWTPs. a. The RdRp-based phylogenetic tree at the family level. The tree consisted of reference RdRp centroids and RNA vOTUs encoded "complete" RdRp domains that can be classified into established RNA viral families. For viral families with multiple clades, only the clade with the largest number of members were retained in the tree. b. The RdRp-based phylogenetic tree consisted of all reference RdRp centroids and RNA vOTUs encoded "complete" RdRp domains. The color of each clade in the tree represented the lowest taxonomic level of this clade. The number next to each clade represented that the majority of its members belonged to the corresponding megataxon. The specific labels for each megataxon can be found in Table 1.

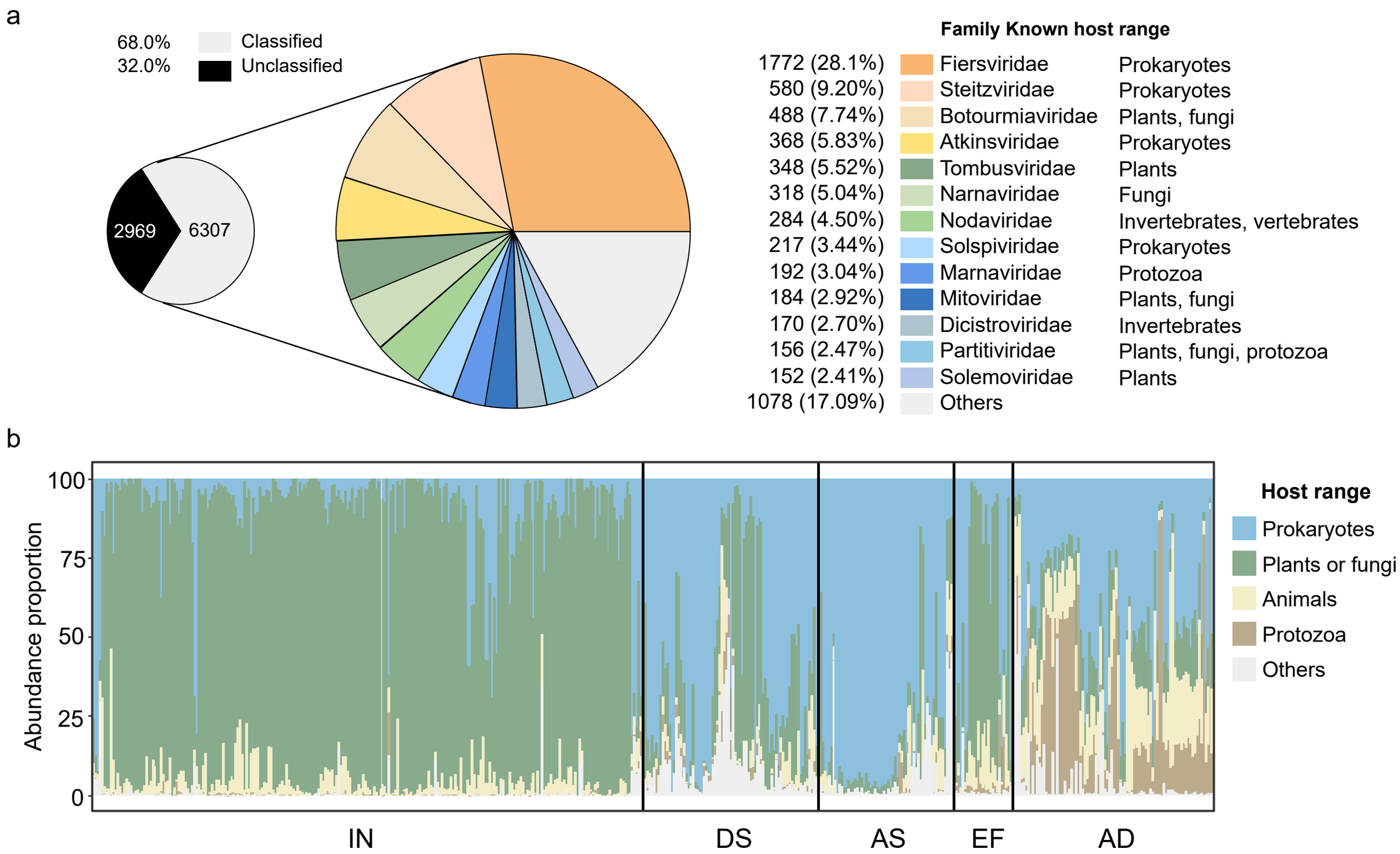


Fig. 3 Taxonomy-based diversity and host ranges of RNA viruses in the WWTPs. a. The richness percentage of RNA vOTUs at family level. Among 9276 RNA vOTUs encoded "complete" RdRp domains, 6307 RNA vOTUs can be classified into established viral families. The prior known host range of each viral family was shown on the right. b. The abundance proportion of RNA vOTUs in the WWTP samples grouped by their host ranges suggested by affiliated taxonomy. Potential RNA viruses infecting prokaryotes contained RNA vOTUs classified into *Leviviricetes* and *Cystoviridae*. Potential RNA viruses infecting plants or fungi contained RNA vOTUs classified into *Alphaflexiviridae*, *Amalgaviridae*, *Aspiviridae*, *Benyviridae*, *Betaflexiviridae*, *Botourmiaviridae*, *Bromoviridae*, *Closteroviridae*, *Curvulaviridae*, *Endornaviridae*, *Megabirnaviridae*, *Mitoviridae*, *Narnaviridae*, *Secoviridae*, *Solemoviridae*, *Tombusviridae*, *Tymoviridae*, and *Virgaviridae.* Potential RNA viruses infecting animals (vertebrates and invertebrates) contained RNA vOTUs classified into *Alphatetraviridae*, *Astroviridae*, *Birnaviridae*, *Caliciviridae*, *Dicistroviridae*, *Hepeviridae*, *Iflaviridae*, *Nodaviridae*, *Orthomyxoviridae*, *Permutotetraviridae*, *Phenuiviridae*, *Picobirnaviridae*, and *Picornaviridae*. Potential RNA viruses infecting protozoa contained RNA vOTUs classified into *Marnaviridae*, *Matonaviridae*, and *Mayoviridae*. The abundance proportion of RNA vOTUs targeting each host group was calculated by the abundance of RNA vOTUs classified into the corresponding taxonomic group divided by the total abundance of RNA vOTUs with a known host range (n = 6294). The RNA vOTUs without prior known host range were not included in this calculation. IN: influent, DS: denitrification sludge, AS: activated sludge, EF: effluent, AD: anaerobic digester sludge.

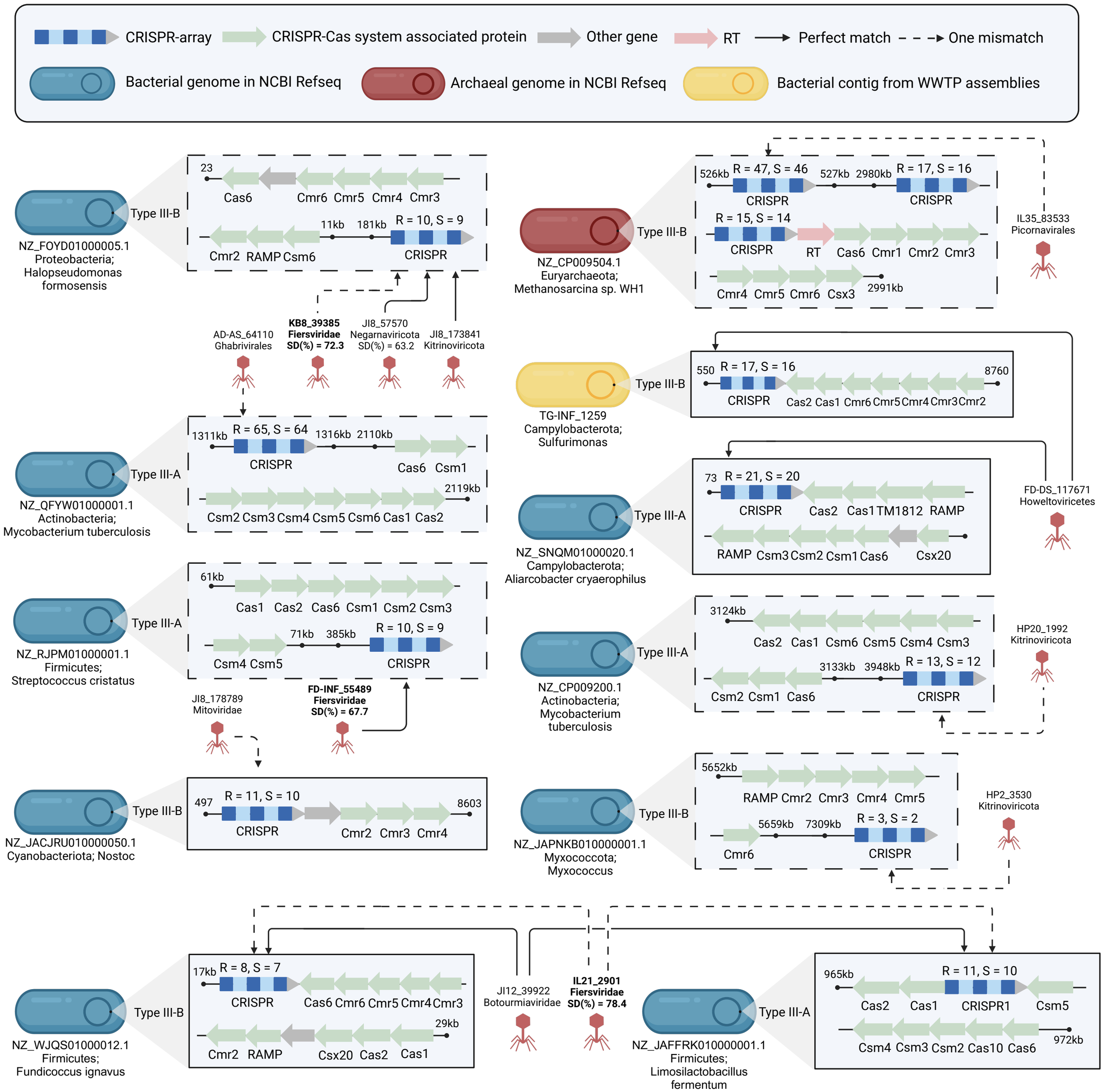


Fig. 4 RNA virus-prokaryotes host associations suggested by CRISPR spacer match. CRISPR spacer blastn-short alignments were shown by arrows from the virus to prokaryotic CRISPR spacer (solid for perfect match and dashed for 1 mismatch). The frame for host genome was illustrated in the solid style if the matched spacer was adjacent to the type-Ⅲ CRISPR system and was dashed if spacer was not adjacent. The specific type of annotated CRISPR system was labeled for each potential host genome. Coding regions point right for (+) and left for (−) sense. The taxonomy for microbial hosts and RNA viruses were shown. The labels for RNA vOTUs within known clades infected prokaryotes (*Leviviricetes* and *Cystoviridae*) were shown in bold font. The percentage of ORFs in the RNA vOTU associated with SD sequences was indicated in the figure if the percentage > 50%. The genomes colored in blue or red were reference bacterial or archaeal genomes in NCBI Refseq database, respectively. The bacterial genome colored in yellow was contig assembled in the WWTP metagenomes. The taxonomy classification of assembled contig was determined by Kaiju.

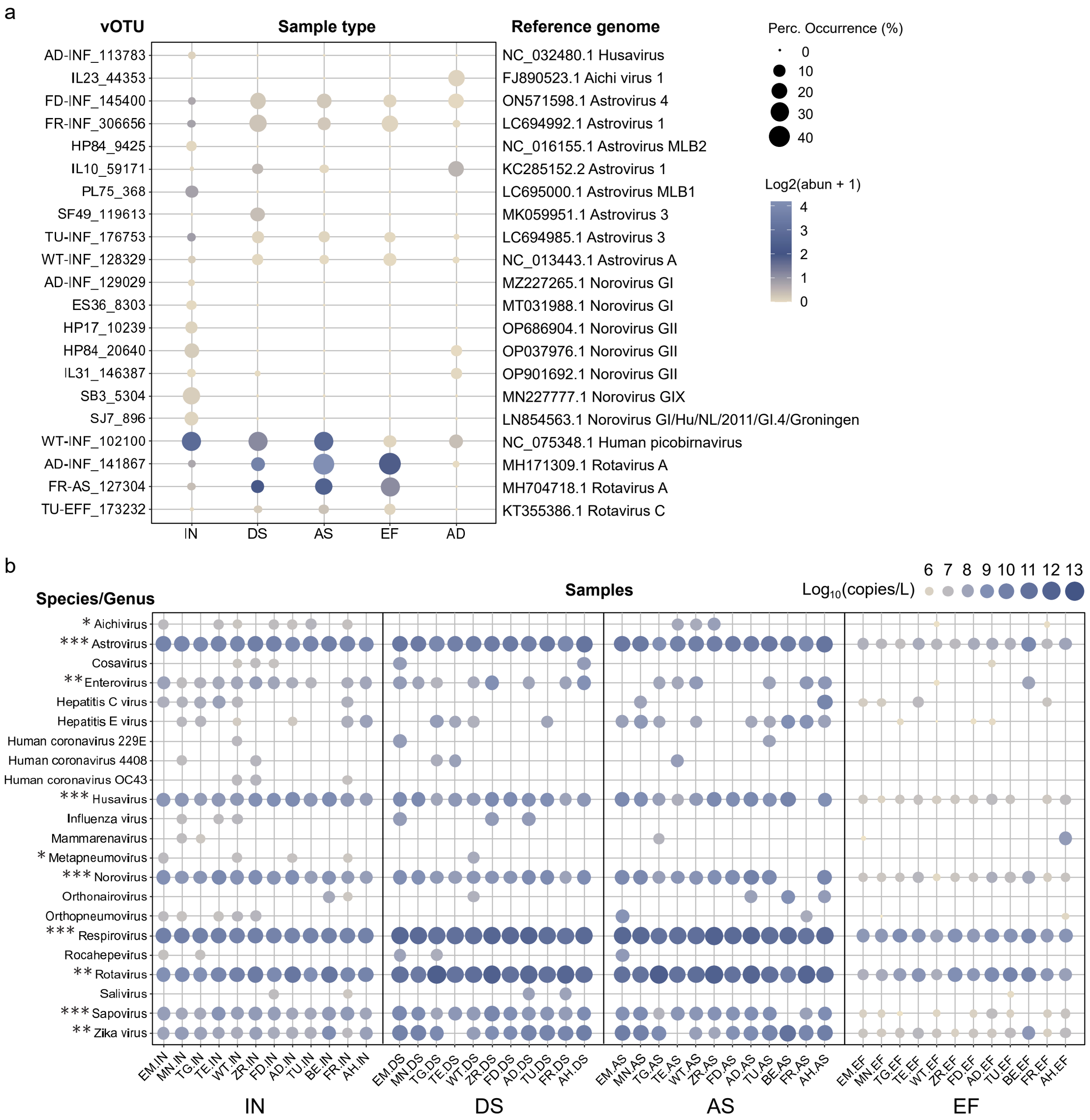


Fig. 5 Distribution of human RNA viruses in different stages in the WWTPs. a. 21 RNA vOTUs were predicted to be human RNA virus by blastn similarity (identity > 90% and coverage > 80% against reference genome in NCBI Virus). The matched reference genomes were labeled in right. The size of bubble represented the percentage of occurrence of the RNA vOTUs and the color represented the relative abundance of the RNA vOTUs. b. The estimated absolute concentration (measured in copies per liter) of human RNA viruses in the 47 metatranscriptomes from Switzerland based on reads-level analyses.

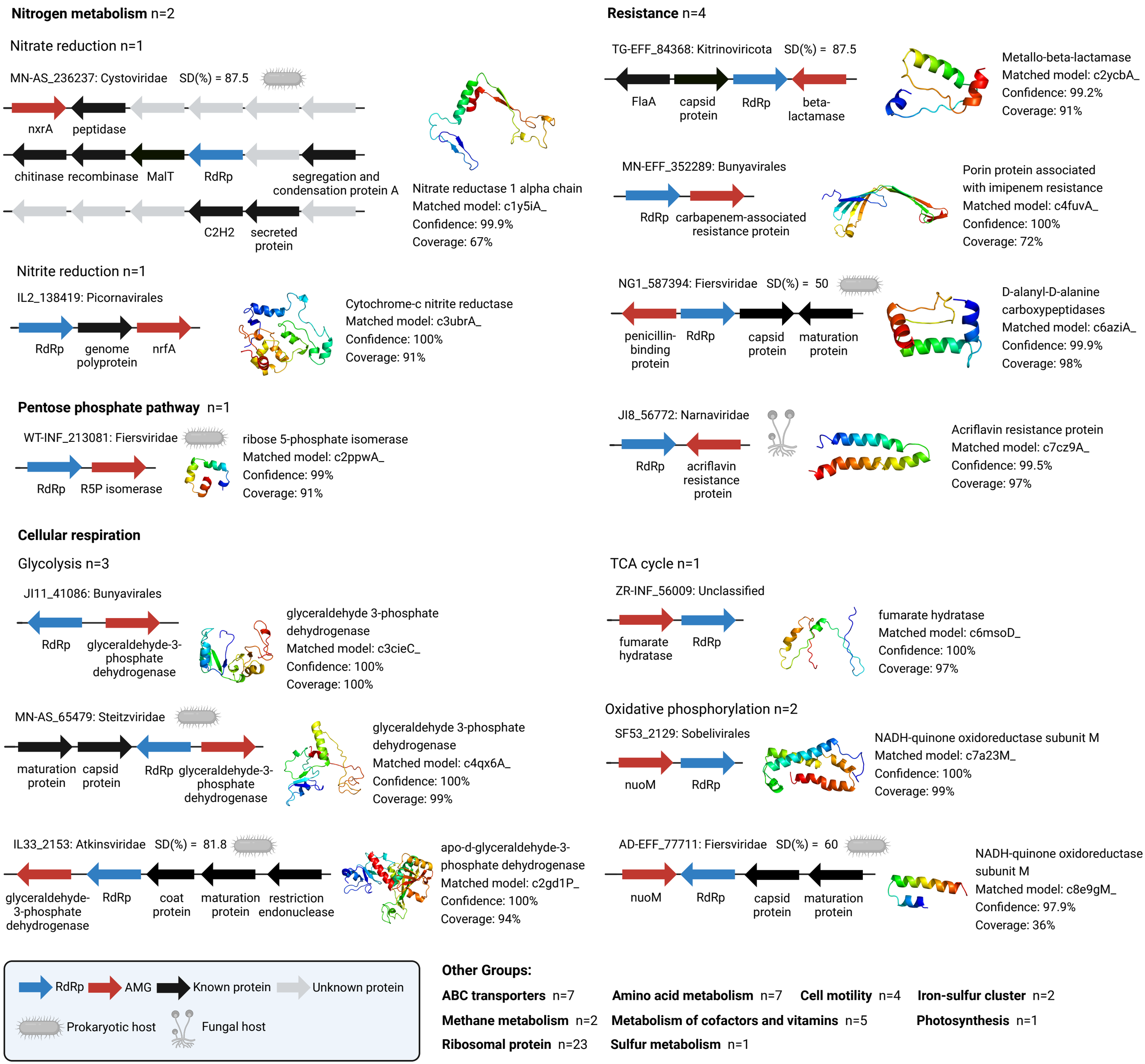
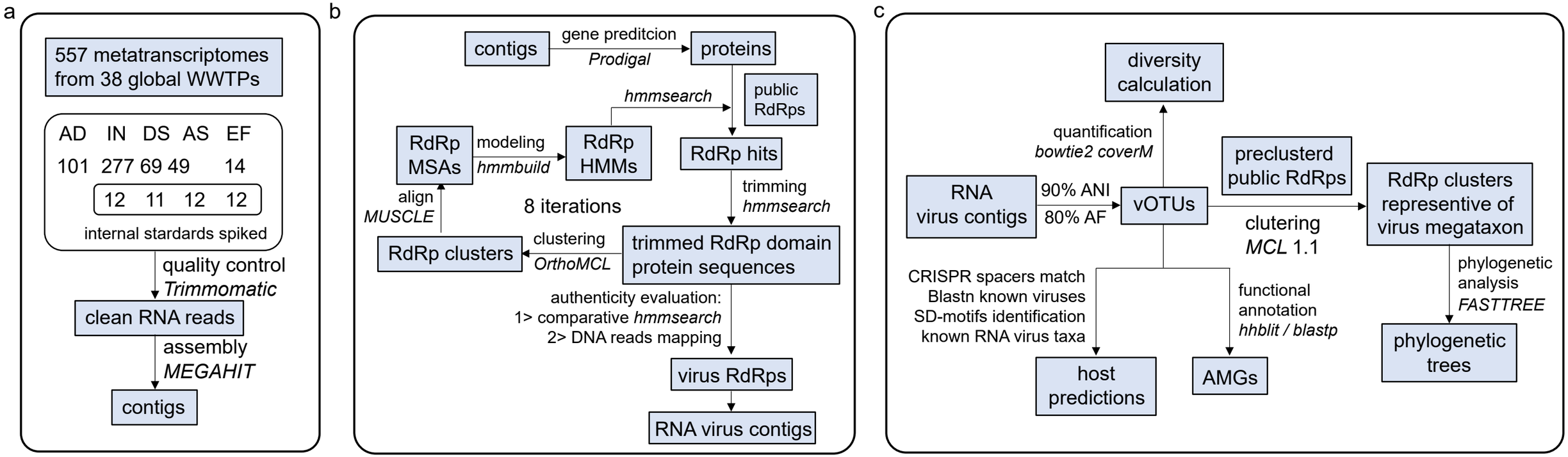


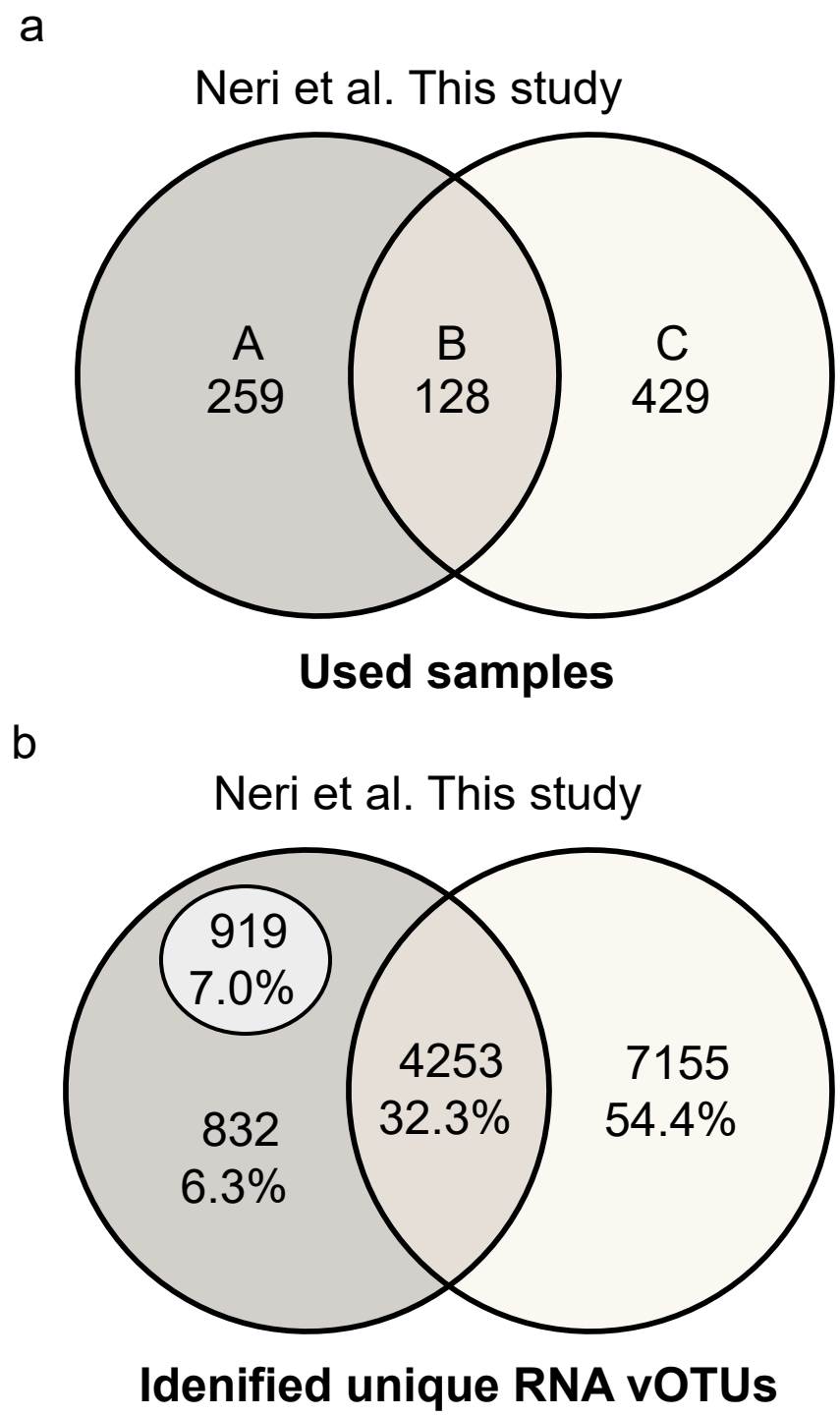
Fig. 6 The AMGs encoded by RNA vOTUs revealed involvement of RNA viruses in various metabolic processes. The AMGs encoded by RNA vOTUs were marked in red. Coding regions point right for (+) and left for (−) sense. The predicted 3D structure of AMGs and the matched reference protein model by Phyre2 were listed beside the gene flow. The percentage of ORFs in the RNA vOTU associated with SD sequences was indicated in the figure if the percentage > 50%. The known hosts were shown beside the label of RNA vOTU based on its affiliated taxonomy.

Mega-taxa	Total leaves	WWTP RdRp centroids	Known public centroids	Megataxa label	Label accuracy
1	5676	714	4482	Pisuviricota;Pisoniviricetes;Picornavirales;No-Marna/Stelpaviricetes	94.98
2	5153	863	3807	Kitrinoviricota;Alsuviricetes/Unknown Class(es)	99.32
3	4300	2531	1697	Lenarviricota;Leviviricetes;Norzivirales	99.71
4	2843	496	2148	Lenarviricota;Miaviricetes	97.58
5	2464	428	1563	Pisuviricota;Duplopiviricetes/Pisoniviricetes;Sobelivirales	91.36
6	2131	382	1332	Lenarviricota;Howeltoviricetes	99.40
7	2105	561	1420	Kitrinoviricota;Unknown Class(es)	95.92
8	2049	365	902	Lenarviricota;Amabiliviricetes	99.00
9	1919	207	1597	Pisuviricota;Pisoniviricetes;Marnaviridae	92.86
10	1900	444	1374	Kitrinoviricota;Toluca-/Flasuviricetes	96.51
11	1777	717	1036	Lenarviricota;Leviviricetes;Timlovirales	88.13
12	1689	289	1340	Kitrinoviricota;Magsaviricetes	99.18
13	1331	178	1067	Pisuviricota;Pisoniviricetes;Sobelivirales;Solemo-/Alverna-/Barnaviridae	88.10
14	1299	443	531	Negarnaviricota;Ellio-/Insthoviricetes	98.49
15	1108	237	747	Duplornaviricota;Chrymotiviricetes	99.87
16	724	86	550	Kitrinoviricota;Unknown class(es)	100
17	554	52	421	Negarnaviricota;Monji-/Chunqiu-/Milne-/Yunchangviricetes	96.91
18	388	101	220	Permuto-like/Birnaviridae	98.18
19	199	93	93	Duplornaviricota;Vidaverviricetes;Cystoviridae	98.92
20	102	33	35	Duplornaviricota;Resentoviricetes	100
21	68	56	1	New-Megataxon	NA

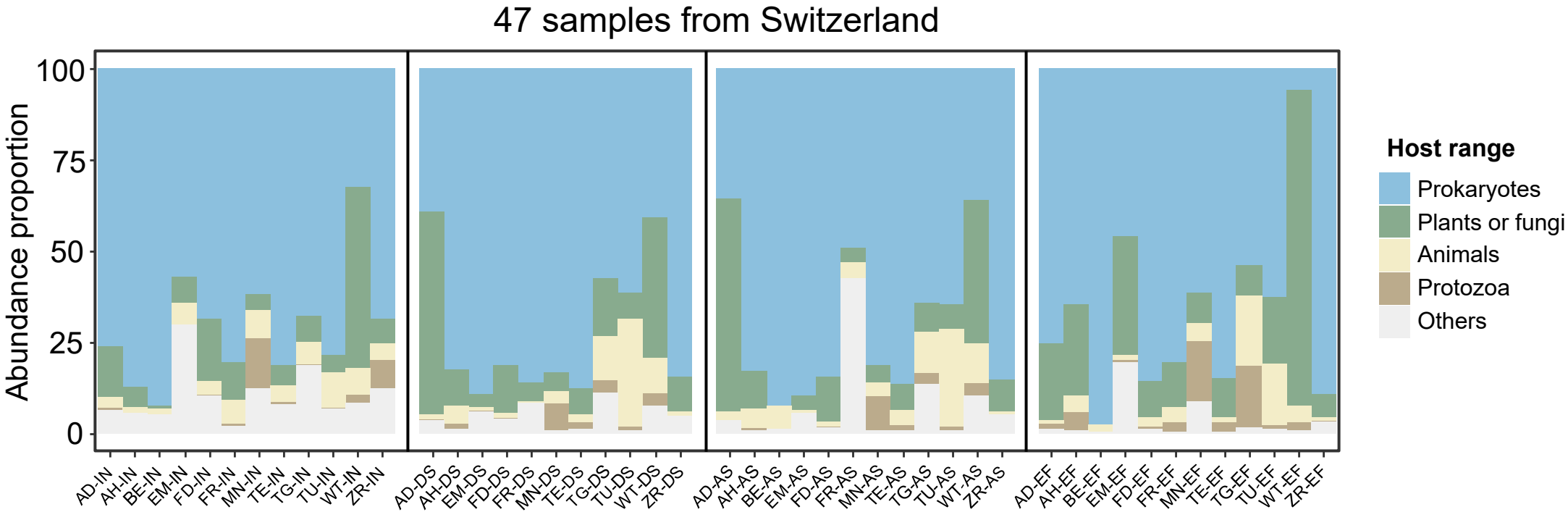
Table. 1 Megataxon of viral RdRps clustered by WWTP RdRps and public reference RdRp centroids. The labels of 21 megataxa of RNA viruses based on RdRp clustering using mcl (I = 0.9). The accuracy of the label for each megataxon was calculated as the number of public RdRp centroids affiliated with the taxonomy divided by the total number of known public centroids (those with labels at phylum level at least) in the megataxon.



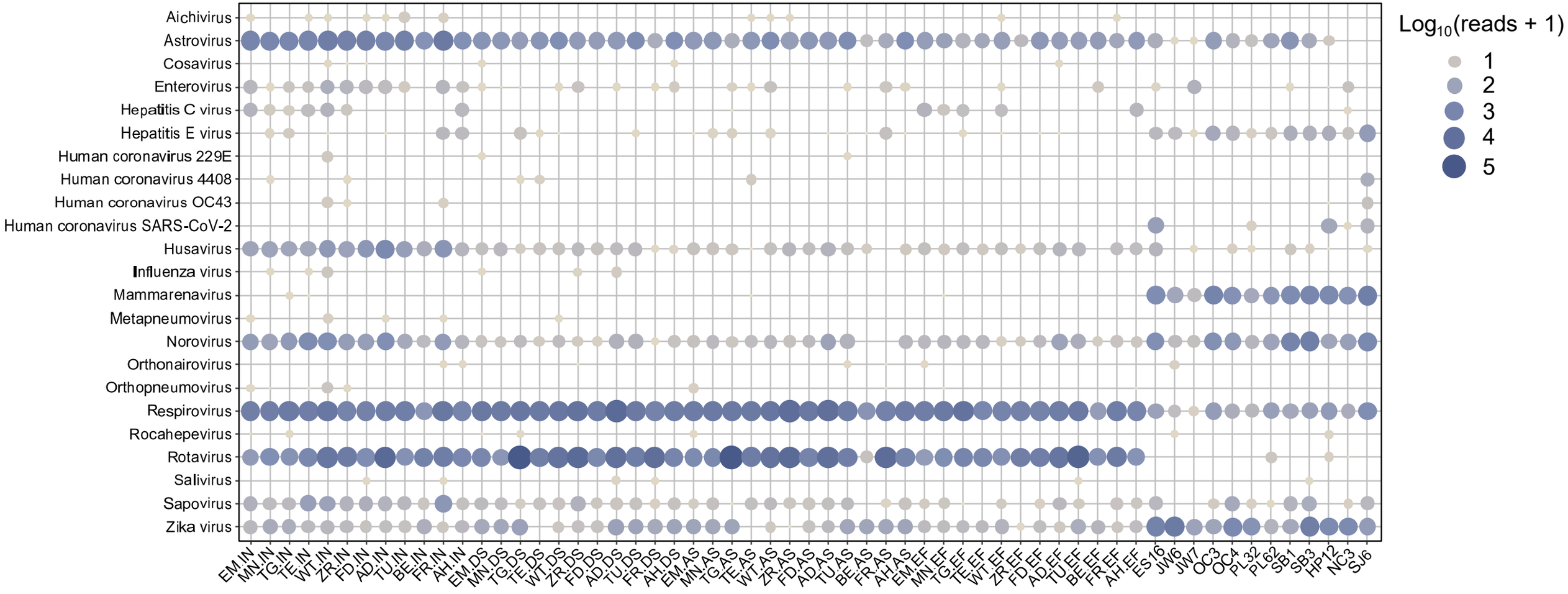
Extended Data Fig. 1 The bioinformatic pipeline for studying RNA viruses in the WWTPs in this study. a. A total of 557 metatranscriptomes from 38 global WWTPs were downloaded from public databases and used for RNA virus exploration. Among the collected metatranscriptomes, 47 samples from Switzerland were spiked with mRNA internal standards and thus allowed us to estimate the absolute concentration of RNA viruses in these samples. The metatranscriptomes were first quality controlled and then assembled into contigs. b. After gene prediction, putative viral RdRps were identified by an iterative hmmsearch approach. The authenticity of identified RdRps was assessed by a comparative hmmsearch of putative viral RdRps against RdRp and non-RdRp profiles as well as checking the occurrence of putative viral RdRp-encoding sequences in the metagenomes. c. The RNA vOTUs were clustered at 90% average nucleotide identity (ANI) over 80% alignment fraction of the shorter sequence (AF). Then, to systemically investigate the diversity and ecological roles of RNA viruses in the WWTPs, a series of bioinformatic analyses including taxonomic classification, phylogenetic analyses, host prediction and AMG annotation were conducted for RNA vOTUs



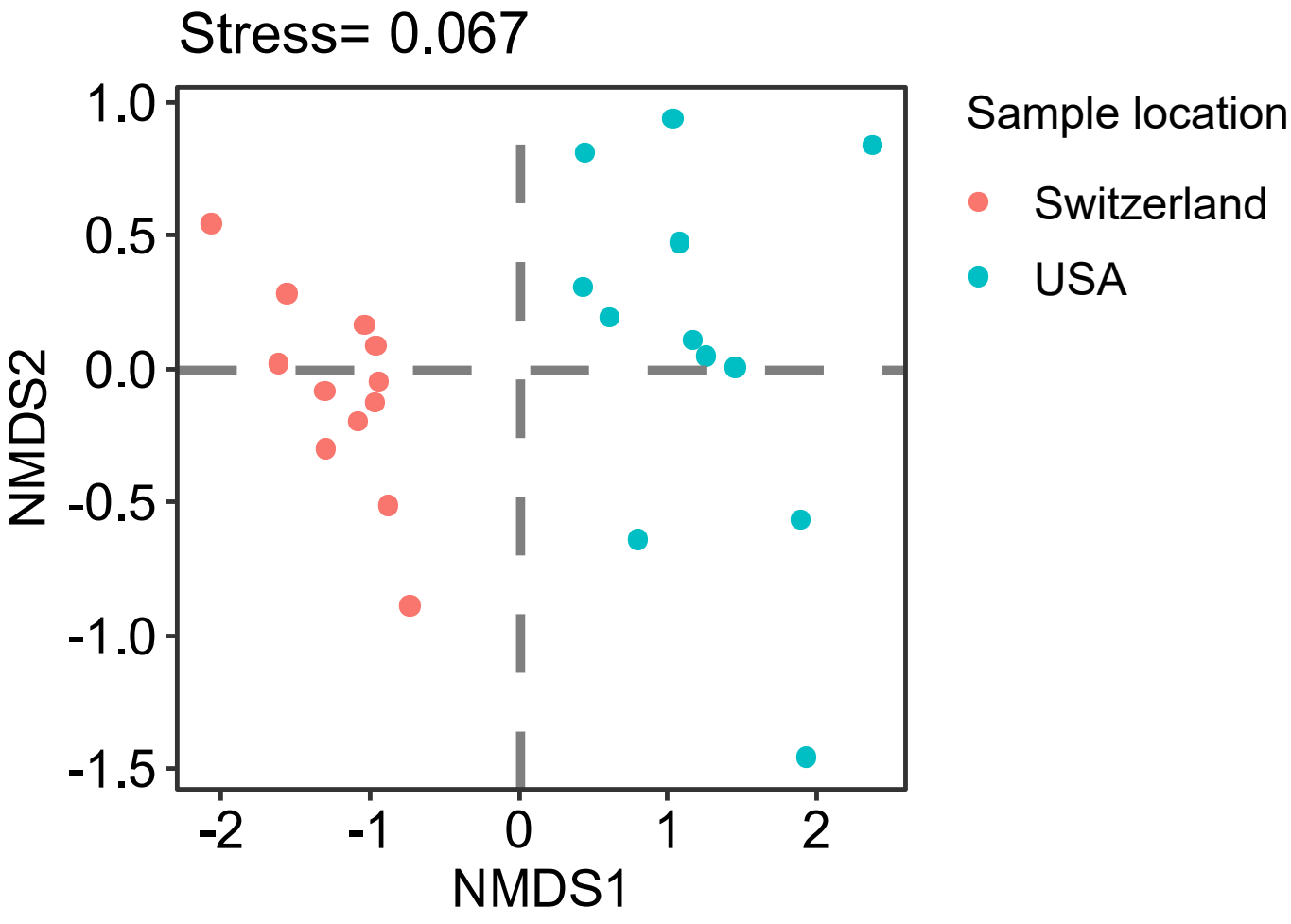
Extended Data Fig. 2 Comparison of selected samples and unique vOTUs recovery between our study and (Neri, Wolf et al. 2022) a. The comparison of collected engineered metatranscriptomes between two studies. Neri et al. searched RNA viruses in thousands of metatranscriptomes from diverse ecosystems which included 387 engineered metatranscriptomes. Our study used a part of the metatranscriptomes from full-scale WWTPs same as Neri et al. (samples n = 128), excluding the samples from lab-scale bioreactors or biogas samples (samples n = 259), and included additional 429 metatranscriptomes from other public datasets. b. To compare the unique vOTUs recovery between two studies, we merged and clustered all the RNA viral contigs from two studies at 90% ANI and 80% AF, which resulted in 13159 RNA vOTUs in total. Among the 13159 clustered RNA vOTUs, 4253 RNA vOTUs occurred both in Neri et al. and this study, 919 RNA vOTUs only occurred in the samples used in Neri et al. but not used in this study (samples n = 259), 832 RNA vOTUs occurred in the samples used in both study but failed to be identified in this study may be due to the different viral identification methods, and an additional 7155 RNA vOTUs were only identified in our study. Given that Neri et al. was the first omic-based study for RNA viruses which largely included hundreds of engineered samples, we compared the recovered diversity of RNA viruses between Neri et al. and our study, and found that we doubled the known diversity of RNA viruses in engineered systems here, from 6004 to 13159 RNA vOTUs



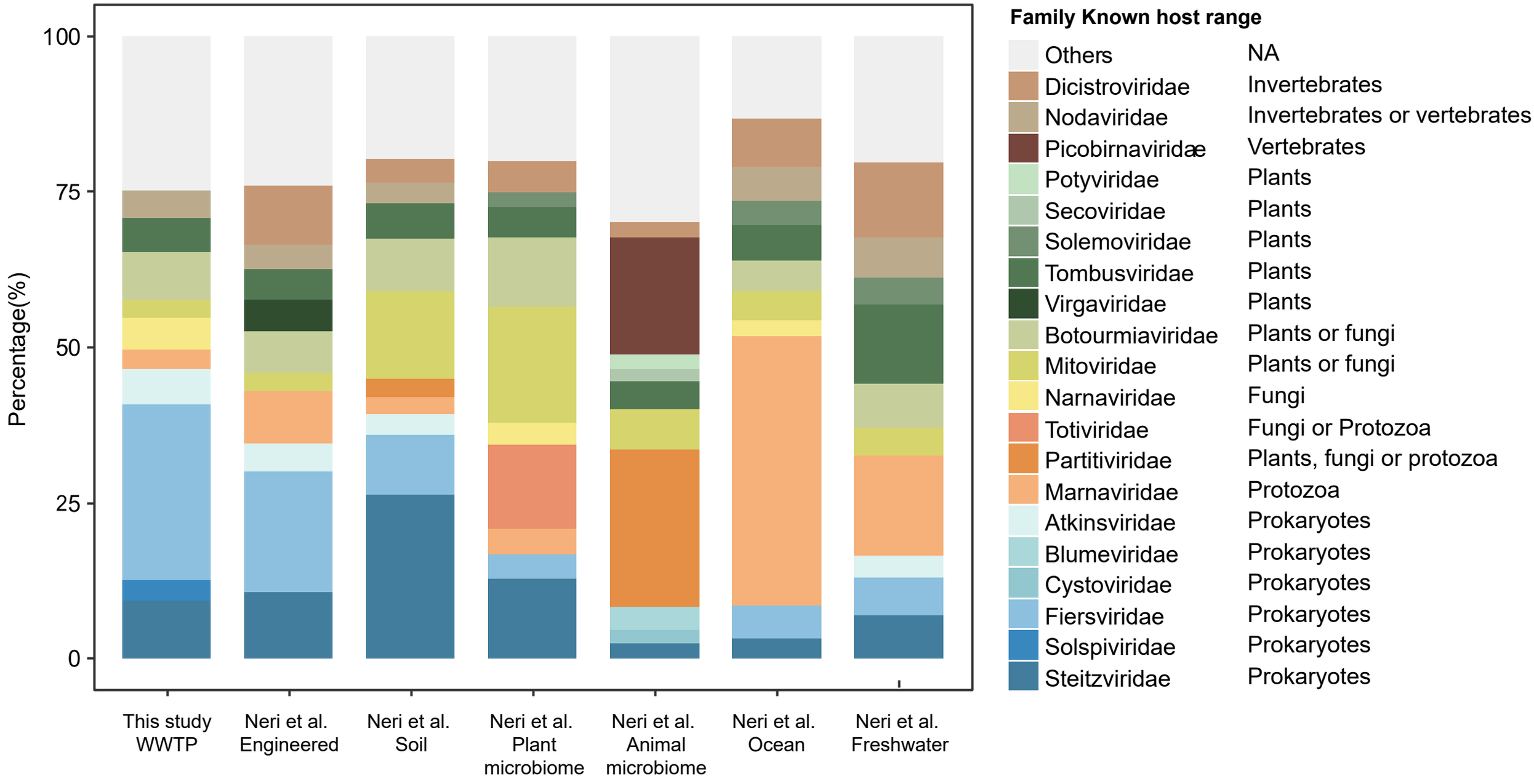
Extended Data Fig. 3 The abundance proportion of RNA vOTUs in the 47 samples from Switzerland grouped by their host ranges suggested by affiliated taxonomy. Potential RNA viruses infecting prokaryotes contained RNA vOTUs classified into *Leviviricetes* and *Cystoviridae*. Potential RNA viruses infecting plants or fungi contained RNA vOTUs classified into *Alphaflexiviridae*, *Amalgaviridae*, *Aspiviridae*, *Benyviridae*, *Betaflexiviridae*, *Botourmiaviridae*, *Bromoviridae*, *Closteroviridae*, *Curvulaviridae*, *Endornaviridae*, *Megabirnaviridae*, *Mitoviridae*, *Narnaviridae*, *Secoviridae*, *Solemoviridae*, *Tombusviridae*, *Tymoviridae*, and *Virgaviridae*. Potential RNA viruses infecting animals (vertebrates and invertebrates) contained RNA vOTUs classified into *Alphatetraviridae*, *Astroviridae*, *Birnaviridae*, *Caliciviridae*, *Dicistroviridae*, *Hepeviridae*, *Iflaviridae*, *Nodaviridae*, *Orthomyxoviridae*, *Permutotetraviridae*, *Phenuiviridae*, *Picobirnaviridae*, and *Picornaviridae*. Potential RNA viruses infecting protozoa contained RNA vOTUs classified into *Marnaviridae*, *Matonaviridae*, and *Mayoviridae*. The abundance proportion of RNA vOTUs targeting each host group was calculated by the abundance of RNA vOTUs classified into the corresponding taxonomic group divided by the total abundance of RNA vOTUs with a known host range (n = 6294). The RNA vOTUs without prior known host range were not included in this calculation.



Extended Data Fig. 4 Reads-based detection of human RNA viruses in the WWTPs. The bubble plot showed the number of reads aligned to human RNA viruses from NCBI virus database (species or organisms) with 90% identity at 80% reads length in the 59 metatranscriptomes from Switzerland or USA.



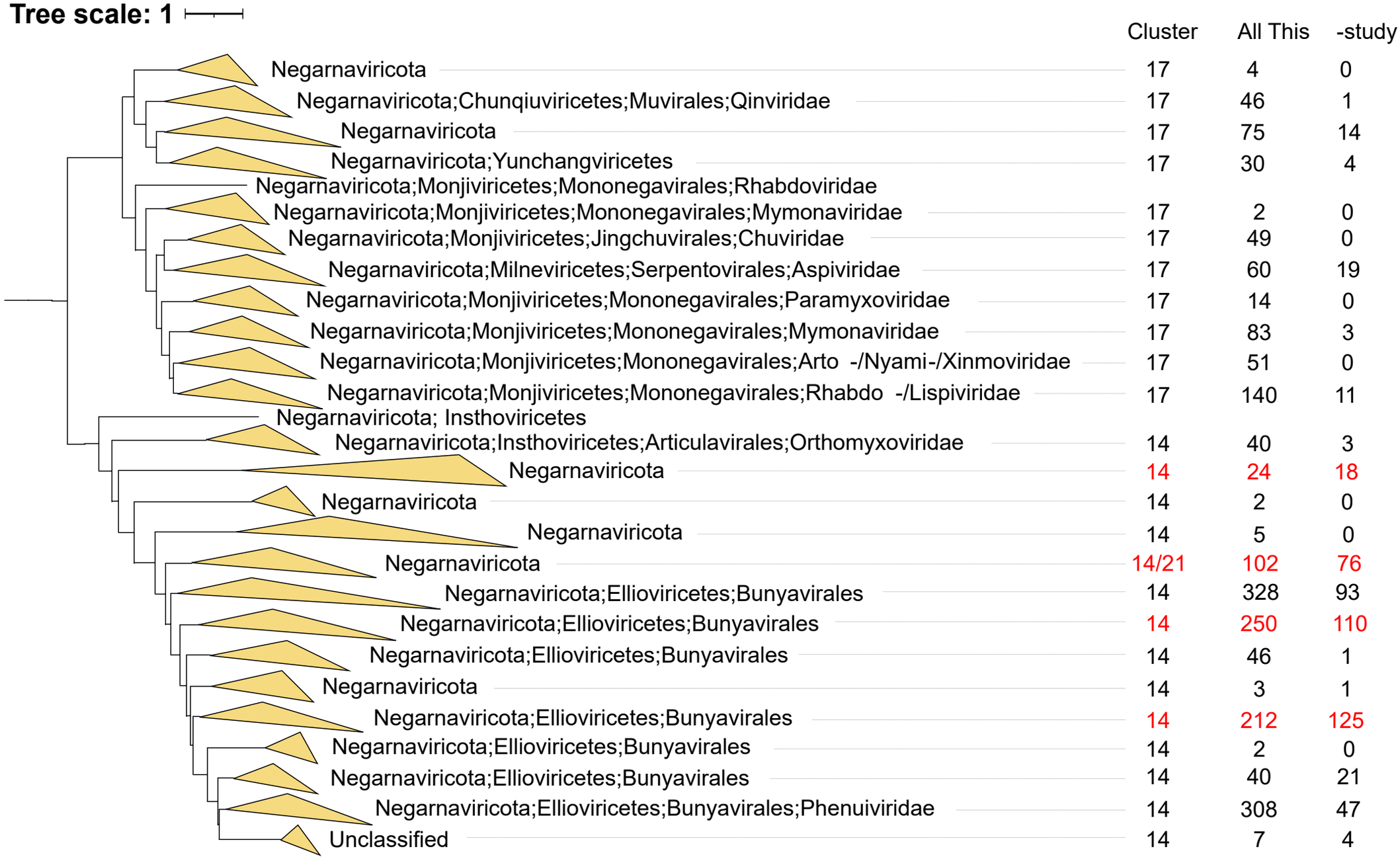
Extended Data Fig. 5 NMDS analysis of reads distribution of human RNA viruses in 12 influent samples from Switzerland and 12 influent samples from USA.



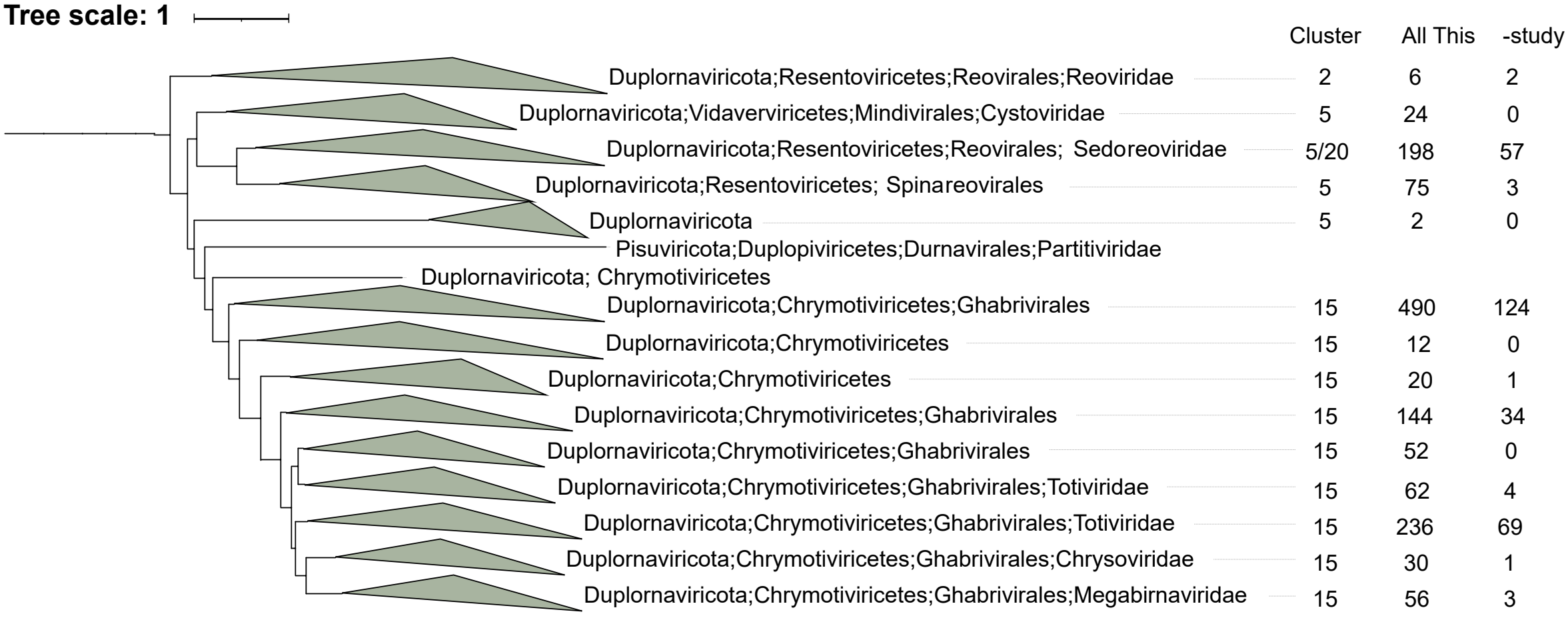
Extended Data Fig. 6 Comparison of abundant RNA viral families and their host ranges between this study and other ecosystems surveyed by Neri et al.

Lenarviricota

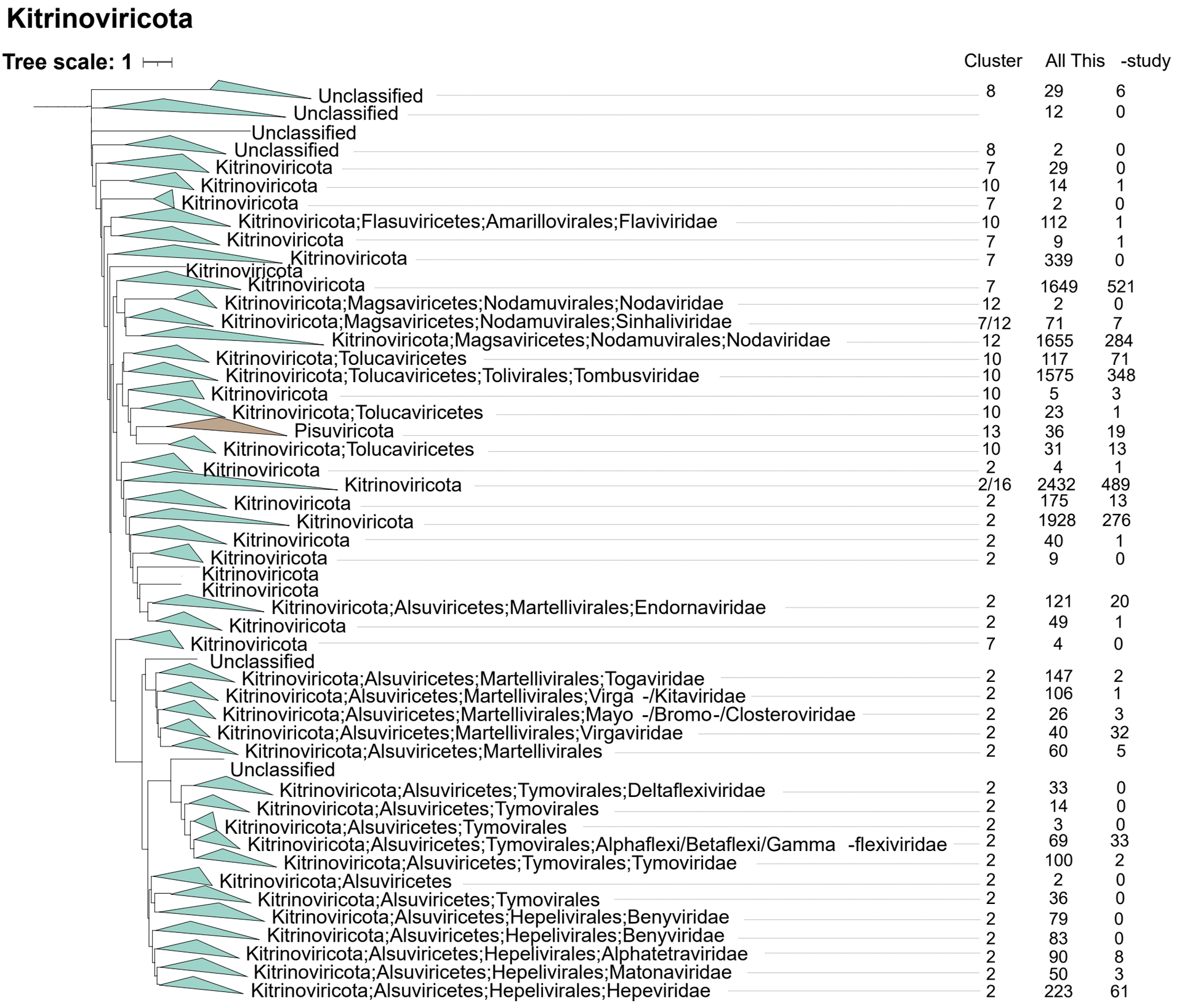
Supplementary Fig. 1 RdRp -based phylogenetic tree of *Lenarviricota*. The number beside each clade represented the affiliated megataxa, the number of all members in this clade and the number of RdRps identified in this study.



Supplementary Fig. 2 RdRp -based phylogenetic tree of *Negarnaviricota*. The number beside each clade represented the affiliated megataxa, the number of all members in this clade and the number of RdRps identified in this study. The clades labeled in red text were four novel clades mainly consisted of RdRps identified in this study.

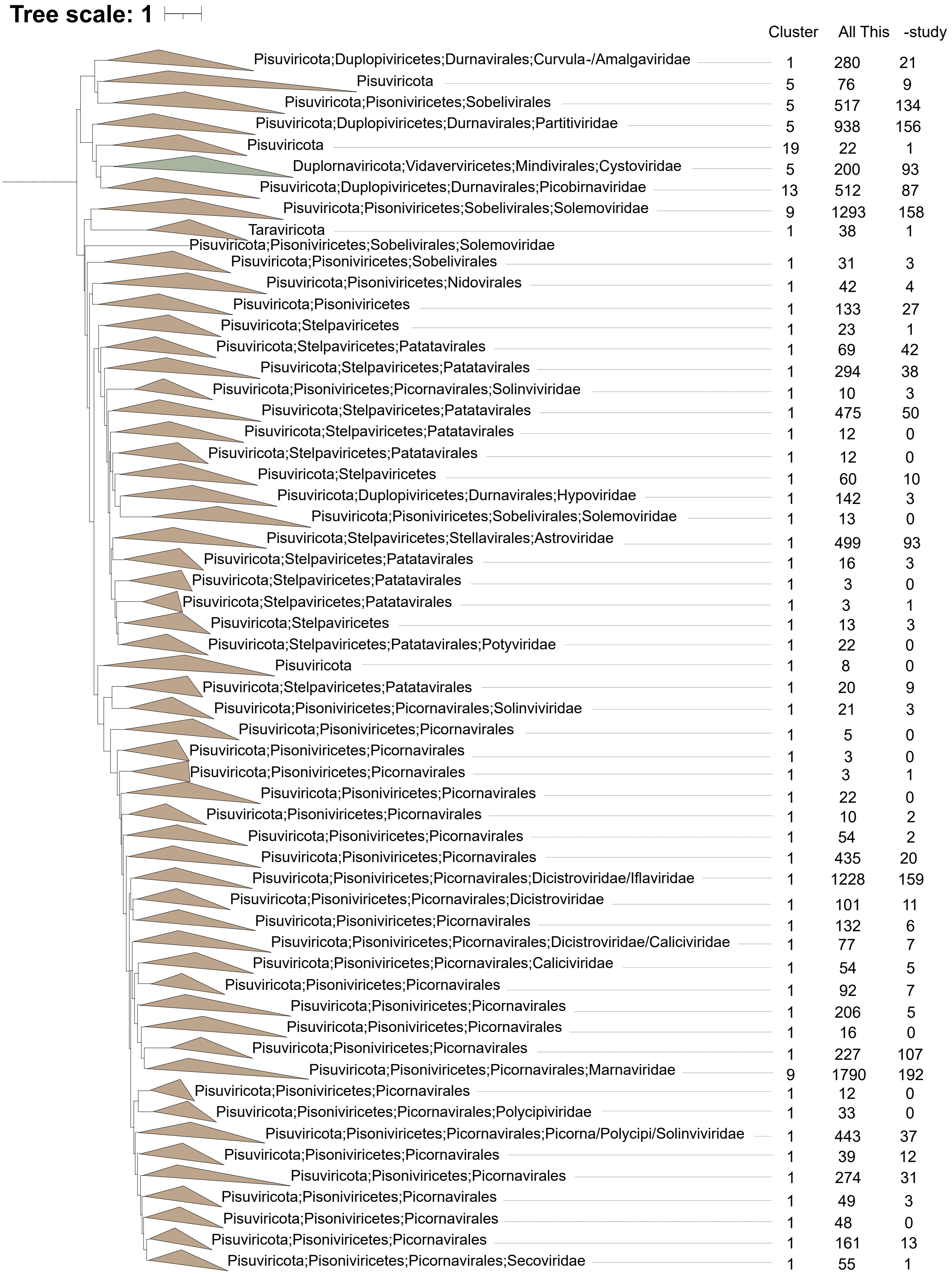


Supplementary Fig. 3 RdRp -based phylogenetic tree of *Duplornaviricota*. The number beside each clade represented the affiliated megataxa, the number of all members in this clade and the number of RdRps identified in this study.

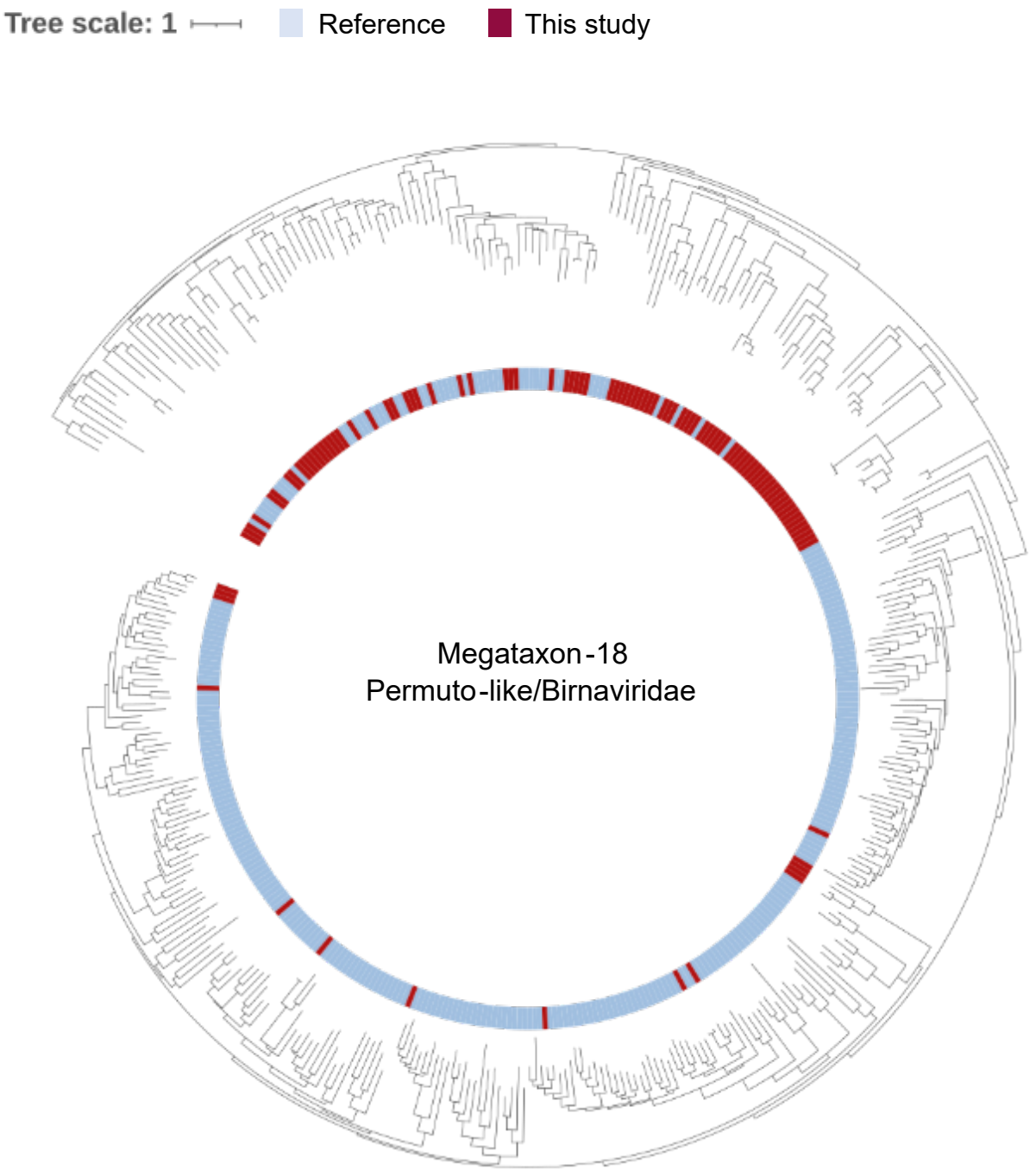


Supplementary Fig. 4 RdRp -based phylogenetic tree of *Kitrinoviricota*. The number beside each clade represented the affiliated megataxa, the number of all members in this clade and the number of RdRps identified in this study.

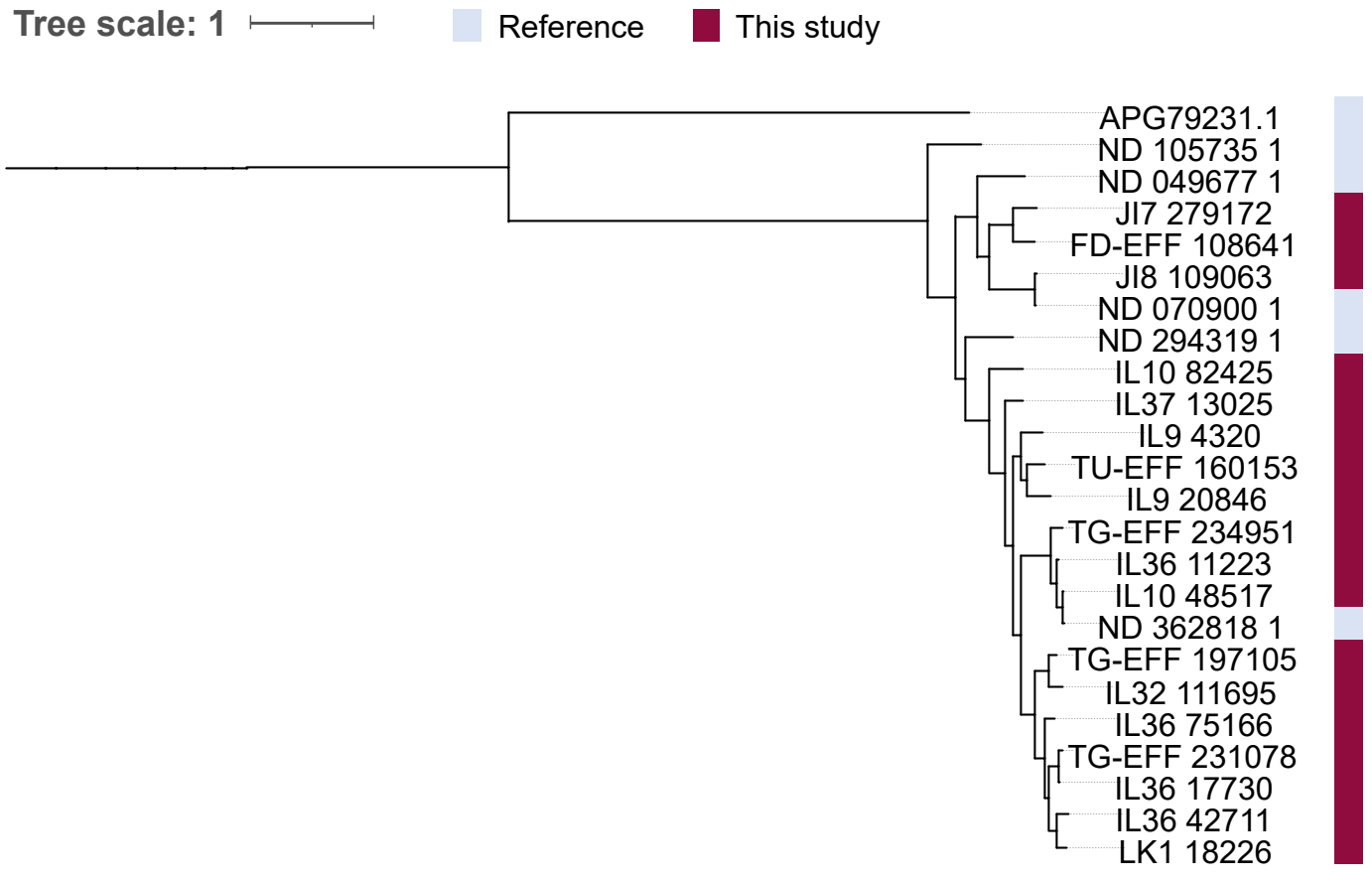
Pisuviricota



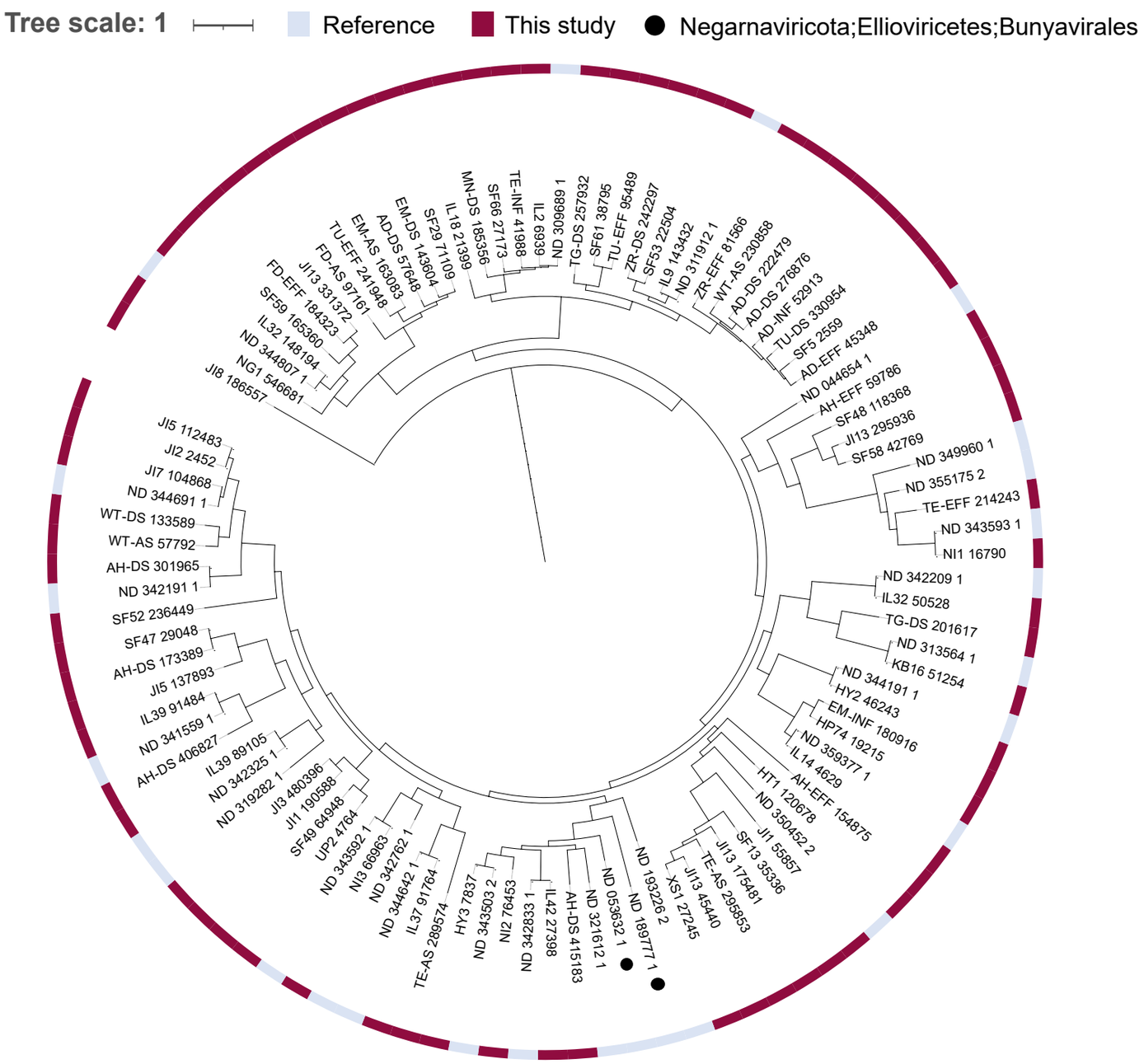
Supplementary Fig. 5 RdRp-based phylogenetic tree of *Pisuviricota*. The number beside each clade represented the affiliated megataxa, the number of all members in this clade and the number of RdRps identified in this study.



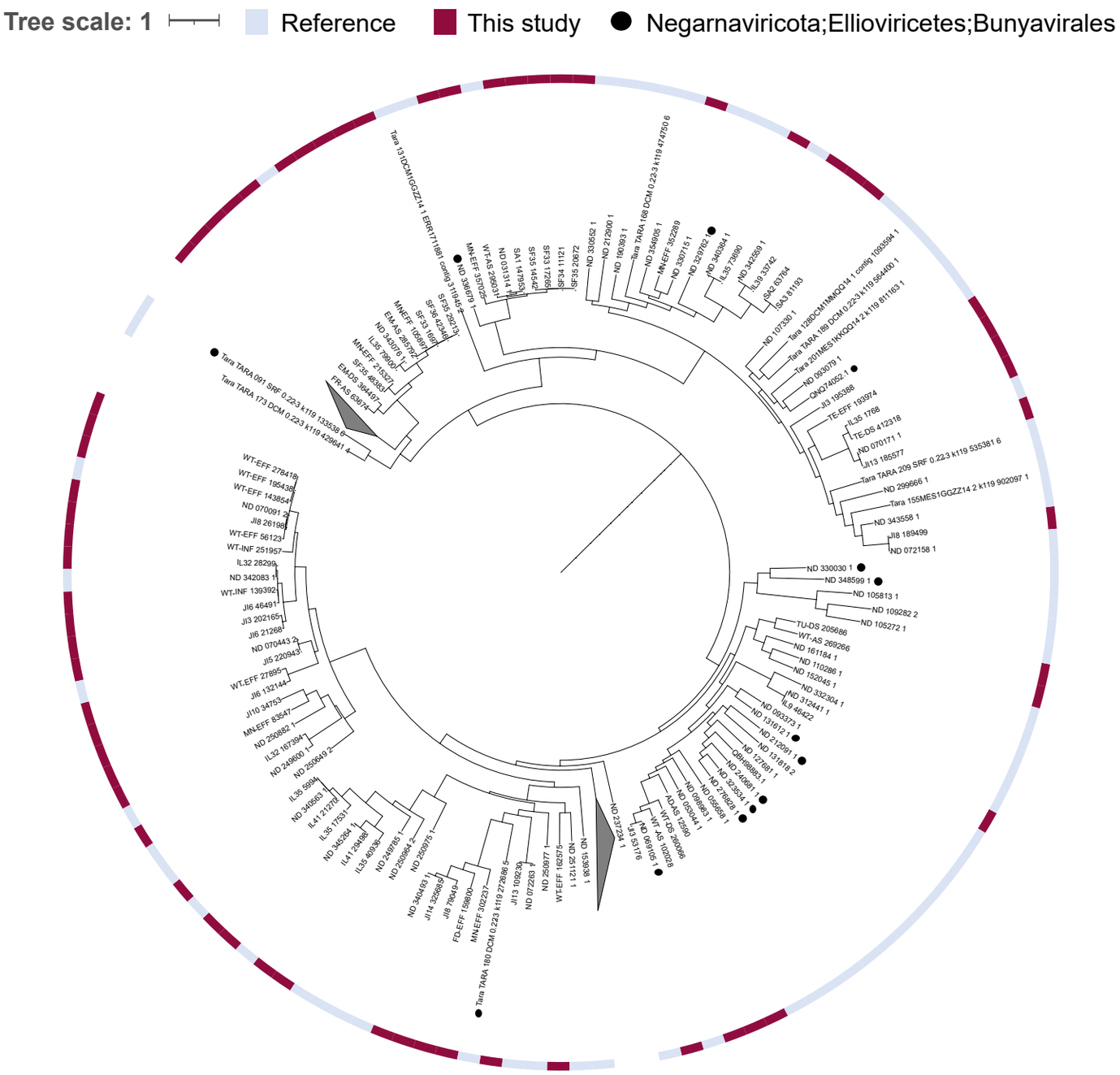
Supplementary Fig. 6 RdRp-based phylogenetic tree of *Permutoviridae* and *Birnaviridae*. The phylogenetic tree was constructed by RdRps clustered into megataxon 18. The tree was individually constructed and not included in the global RdRp tree because of the permutated conserve motifs.



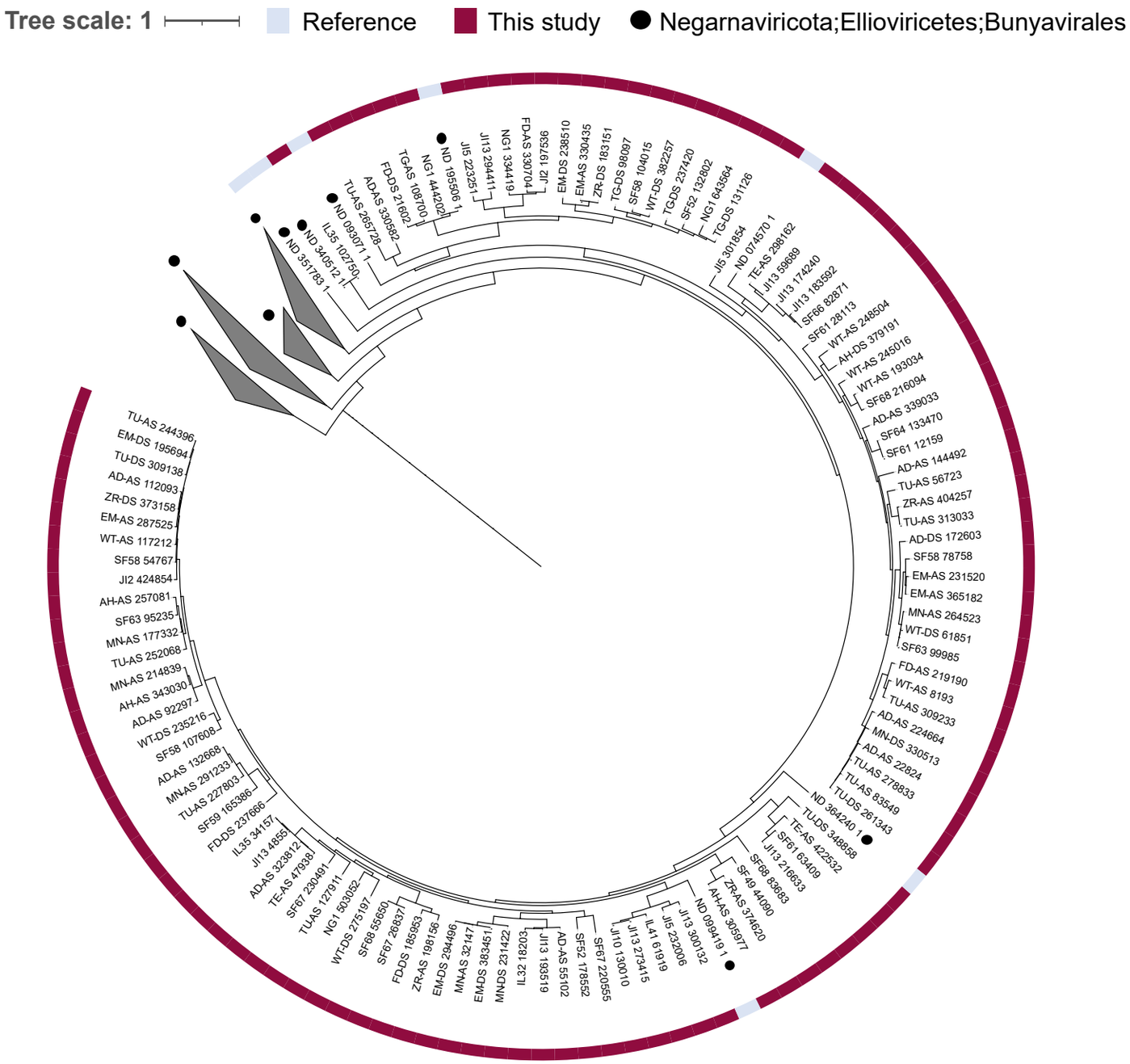
Supplementary Fig. 7 WWTP novel clade1 identified in this study. The clade was mainly consisted of WWTP RdRps or unknown reference RdRp centroids. The clade was under phylum *Negarnaviricota*.



Supplementary Fig. 8 WWTP novel clade2 identified in this study. The clade was mainly consisted of WWTP RdRps or unknown reference RdRp centroids. The clade was under phylum *Negarnaviricota*.



Supplementary Fig. 9 WWTP novel clade3 identified in this study. The clade was mainly consisted of WWTP RdRps or unknown reference RdRp centroids. The clade was under phylum *Negarnaviricota*.



Supplementary Fig. 10 WWTP novel clade4 identified in this study. The clade was mainly consisted of WWTP RdRps or unknown reference RdRp centroids. The clade was under phylum Negarnaviricota.