Environmental Research 231 (2023) 116040

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Mapping human pathogens in wastewater using a metatranscriptomic approach

João Carneiro [a,*], Francisco Pascoal [a,f], Miguel Semedo [a], Diogo Pratas [b,c,d], Maria Paola Tomasino [a], Adriana Rego [a], Maria de Fátima Carvalho [a,e], Ana Paula Mucha [a,f], Catarina Magalhães [a,f]

[a] CIIMAR, Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros Do Porto de Leixões, Av. General Norton de Matos, S/n, 4450-208, Matosinhos, Portugal
[b] IEETA - Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Portugal
[c] Department of Virology, University of Helsinki, Finland
[d] Department of Electronics Telecommunications and Informatics, University of Aveiro, Portugal
[e] School of Medicine and Biomedical Sciences (ICBAS), University of Porto, Portugal
[f] Department of Biology, Faculty of Sciences, University of Porto, Rua Do Campo Alegre S/n, 4169– 007, Porto, Portugal

ARTICLE INFO

Handling Editor: Aijie Wang

Keywords:
Human pathogens
Wastewater
Metatranscriptomics
Public health

ABSTRACT

The monitoring of cities' wastewaters for the detection of potentially pathogenic viruses and bacteria has been considered a priority during the COVID-19 pandemic to monitor public health in urban environments. The methodological approaches frequently used for this purpose include deoxyribonucleic acid (DNA)/Ribonucleic acid (RNA) isolation followed by quantitative polymerase chain reaction (qPCR) and reverse transcription (RT)–qPCR targeting pathogenic genes. More recently, the application of metatranscriptomic has opened opportunities to develop broad pathogenic monitoring workflows covering the entire pathogenic community within the sample. Nevertheless, the high amount of data generated in the process requires an appropriate analysis to detect the pathogenic community from the entire dataset. Here, an implementation of a bioinformatic workflow was developed to produce a map of the detected pathogenic bacteria and viruses in wastewater samples by analysing metatranscriptomic data. The main objectives of this work was the development of a computational methodology that can accurately detect both human pathogenic virus and bacteria in wastewater samples. This workflow can be easily reproducible with open-source software and uses efficient computational resources. The results showed that the used algorithms can predict potential human pathogens presence in the tested samples and that active forms of both bacteria and virus can be identified. By comparing the computational method implemented in this study to other state-of-the-art workflows, the implementation analysis was faster, while providing higher accuracy and sensitivity. Considering these results, the processes and methods to monitor wastewater for potential human pathogens can become faster and more accurate. The proposed workflow is available at https://github.com/waterpt/watermonitor and can be implemented in currently wastewater monitoring programs to ascertain the presence of potential human pathogenic species.

1. Introduction

In the advent of infectious disease outbreaks, identifying the infectious agents in wastewater can be helpful, but the computational tools and software to analyse extreme amounts of data are a bottleneck (Garner et al., 2021). Wastewater monitoring traditionally focuses on indicator species, with laboratorial methods well-tailored for faecal indicator bacteria. Several studies have focused on improving laboratorial methods for the identification of specific viral indicators (Crits-Christoph et al., 2021; Ekwanzala et al., 2021; Farkas et al., 2020; Sherchan et al., 2020; Tomasino et al., 2021a). The concerns related to the recovery of viral genetic material can be the result of low viral loads, which limits the use of next-generation sequencing (NGS) (Huang et al., 2019). Recently, there has been a growing focus on NGS-based approaches, which include marker gene amplicon sequencing, whole genome sequencing, shotgun sequencing of environmental DNA and

* Corresponding author.
E-mail addresses: joaomiguelsov@gmail.com, jcarneiro@ciimar.up.pt (J. Carneiro).

https://doi.org/10.1016/j.envres.2023.116040
Received 8 March 2023; Received in revised form 28 April 2023; Accepted 2 May 2023
Available online 5 May 2023

RNA (Garner et al., 2021). The first approach typically includes short-read amplicon sequencing, which limits pathogen taxonomy identification. Metagenomes allow the identification of the DNA of viruses and bacteria, but for the latter, the presence of virulence genes must be confirmed to identify pathogenic species because of the high complexity of samples. Moreover, molecular approaches and metagenomics can capture RNA viruses and may also detect residual DNA of non-living bacteria, but cannot detect infective virions (Bogler et al., 2020). Metatranscriptomics, on the other hand, can detect RNA viruses and only detects the presence of active (expressing) bacteria (Shakya et al., 2019). However, metatranscriptomics is challenging because of the variable and often very short half-life of RNA and the inhibition of RT−PCR by organic substances expected in wastewater samples (Farkas et al., 2020; Garner et al., 2021). For these reasons, some authors defend the use of DNA-based viral indicators (Farkas et al., 2020). Identifying viruses in wastewater can be inconsistent due to fast degradation of viral particles and variability in water volume. Multiple daily sampling events are important for monitoring the presence of the virus in the population (Ahmed et al., 2020; Foladori et al., 2020).

Improvements in detection and monitoring of microorganisms in wastewater using methods like qPCR and RT-qPCR have allowed for consistent analysis. Tracking the SARS-CoV-2 virus in wastewater can measure correlations with reported COVID-19 infections (Amereh et al., 2022; Cervantes-Avilés et al., 2021). Nevertheless, only with the use of cutting-edge omics technologies based in NGS-based approaches (e.g., metagenomics, metatranscriptomics and metaproteomics) can an accurate map of gene expression profiling in wastewater samples be obtained (Ekwanzala et al., 2021). Several software tools (Freitas et al., 2015; Menzel et al., 2016; Pratas et al., 2018; Tovo et al., 2020; Truong et al., 2015; Wood et al., 2019; Wood and Salzberg, 2014) have been developed to better understand the data retrieved from these monitoring processes and there has been a growth of metatranscriptomics projects in public repositories (Shakya et al., 2019). The procedures related to the computational detection of potential viral and bacterial pathogens on metatranscriptomic data present some bottlenecks. This situation is due to the high amount of time needed to analyse the generated data and the expensive software and computer workstations needed to compute thousands of transcriptomes. Additionally, the advanced technical expertise needed for data analysis may hinder its widespread application in wastewater treatment plant (WWTP) monitoring programs. Early warning system to identify and rapidly mitigate the spread of many pathogens, including norovirus, hepatitis viruses and salmonella, and more recently SARS-CoV-2, were routinely implemented by wastewater monitoring in many regions ("Wastewater monitoring comes of age," 2022). Here, a reproducible bioinformatics metatranscriptomic approach workflow, optimized for fast computations was built using a combination of free tools and in-house algorithms to map the taxonomic profiles of human pathogens of WWTP samples using transcriptomes as the raw data. The main raised question was: Is it possible to improve current computational metatranscriptomic approaches to detect potential human pathogens by circumventing their major drawbacks, which includes the incapacity to deal with large number of samples, low sensitivity and accuracy of the computational methods, and algorithms without optimization? A computational workflow will be implemented in this study to circumvent some of the limitations of other computational methods and improve the drawbacks of current metatranscriptomics approaches. This will be achieved by developing a computational workflow that can process each sample with: (1) higher accuracy and sensitivity to determine the taxonomic sample profile, (2) faster computations with improved performance, (3) a validation by several different algorithms using a statistical approach, (4) a reproducibility workflow, (5) free tools optimization that can be included in current early warning of pathogens monitoring wastewater programs.

2. Methods

2.1. Sample collection and RNA sequencing

A 24-hr composite influent wastewater sample was collected from the Sobreiras WWTP (Porto, Portugal) on May 28, 2020. The untreated wastewater sample was acidified to a pH of 3.5 using 2.0 N HCl, according to Ahmed et al. (Ahmed et al., 2020; Warish et al., 2015). Twenty milliliters of the acidified wastewater sample was immediately filtered through 3 μm + 0.45 μm pore size electronegative membranes with 90 mm diameter (SSWP04700 and HAWP04700; Merck Millipore). Immediately after filtration, the membranes were added to a 5-mL bead tube containing the microbial inactivation reagents and lysis solutions (PM1-RNeasy PowerMicrobiome Kit Compone–t - Qiagen, GMBH, Germany and β-mercaptoethanol, Sigma). Total RNA was extracted directly from the filters with the RNeasy PowerWater kit (Qiagen) for the cell lysis steps followed by the PowerMicrobiome kit (Qiagen) for the RNA extraction and purification steps, according to the manufacturer's protocol and previously described methodologies (Tomasino et al., 2021a, 2021b). In the final step of the extraction kit, RNA was eluted with 100 μL of elution buffer. Total RNA (151.2 ng/μL) was quantified by Nanodrop and stored at −80 °C before shipping for RNA-Seq. RNA libraries and sequencing were performed at BGI-Genomics by using their workflow. Ribosomal RNA was removed using the Ribo-Zero rRNA Removal Kit (BGI). RNA molecules were fragmented into small pieces, and first-strand complementary DNA (cDNA) was generated using random hexamer-primed reverse transcription, followed by second-strand cDNA synthesis with/without dUTP instead of dTTP. The synthesized cDNA was then subjected to end-repair and was 3′ adenylated. Adapters were ligated to the ends of these 3′ adenylated cDNA fragments and cDNA fragments amplified, and the PCR products were purified with Ampure XP Beads (AGENCOURT). The double-stranded PCR products were heat denatured and circularized by the splint oligo sequence. Single strand circle DNA (ssCir DNA) was generated as the final library that was amplified with phi 29 to make DNA nanoballs (DNBs). The DNBs were then loaded into the patterned nanoarray for PE100 (or PE150) sequencing on the DNBseq platform.

2.2. Bioinformatics analysis of the metatranscriptomes

2.2.1. Taxonomic profiling

The bioinformatic workflow to detect viral and bacterial microorganisms was done using FALCON-meta (Pratas et al., 2018) and GOTTCHA2 (Freitas et al., 2015) in the collected wastewater sample (Fig. 1). FALCON-meta with optimized parameters was used to detect the presence of viruses and bacteria in the sample using metatranscriptomic data against an extensive database of complete bacterial and viral genomes from NCBI (database reference build in December 2020 using the toolkit for genomics and proteomics (GTO - https://github.com/cobilab/gto); SARS-CoV-2 virus was added to the FALCON-meta database). One sample with SARS-CoV-2 virus (Ricardo Jorge Institute sample SAMEA6844883-ERS4572485) was used as a positive control. To run FALCON-meta for the wastewater sample retrieved from the Sobreiras WWTP and for the control sample, an improved algorithm derived from Pratas et al., (2018) and available at https://github.com/waterpt/watermonitor, was used. FALCON-meta uses a cache-hash for the deepest context model, where the parameter *c* enables storing only the latest entries up to a certain number of hash collisions in memory. This model allows the use of deep context orders with very sparse representations while removing space constraints and enabling a constant maximum peak of RAM. Generally, increasing *c* renders higher precision at the cost of higher RAM.

A GOTTCHA2 analysis was performed, with a minimum coverage of 0.005, using the viral and bacterial database (Freitas et al., 2015) of complete reference genomes retrieved from NCBI. The workflow for GOTTCHA2 calculations was implemented using the Kbase platform

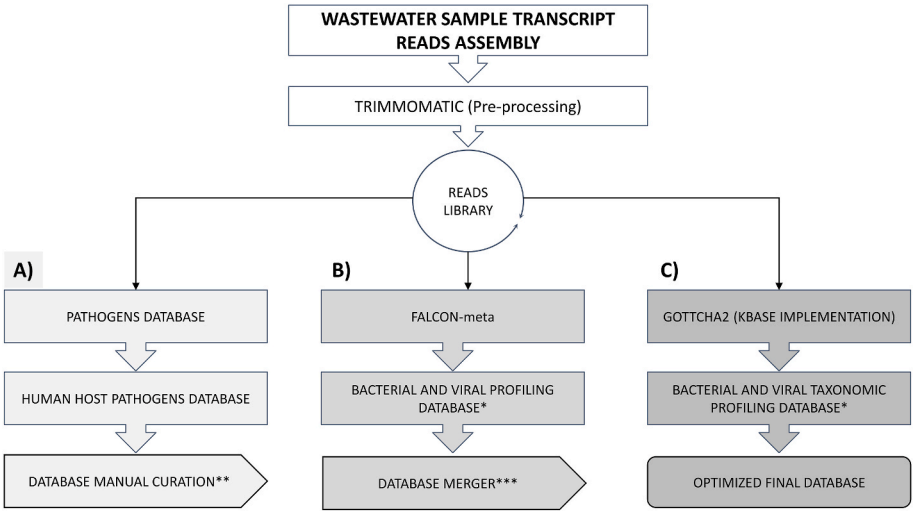

Fig. 1. Workflow of the methodology used to identify the species present in the wastewater sample.
*A positive control was used for SARS-CoV-2 virus identification.
** SARS-CoV-2 viruses was added to the database and records with ambiguous host names discarded.
*** The species was used as primary key to merge the databases that present equal species for each record.

(https://www.kbase.us/about/). Trimmomatic (Bolger et al., 2014) and FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) were used to perform quality control for the single-end reads and ran the GOTTCHA2 signature-based metagenomic taxonomic profiling tool with default parameters. The taxonomic profiling obtained with GOTTCHA2 and FALCON-meta was updated to include information on the species pathogenic to humans (Shaw et al., 2020). To visualize the potential active pathogenic bacteria and RNA viruses identified, the R package ggplot 2 and Circos Software (Krzywinski et al., 2009) were employed. An in-house algorithm was written to automatically plot the final human pathogenic dataset using Circos software for GOTTCHA2 and FALCON-meta results. The included R scripts and in-house Circos algorithm calculate the microbial diversity statistics for each metatranscriptome, create Circos plots based on the pathogenic organism, and create stacked bar graphs for each detected transcript, displaying the relative abundance and percentage of similarity. These visualizations illustrate all the species, with special relevance given to the pathogenic species (read count, relative abundance and percentage of similarity) identified in the final dataset.

The results of GOTTCHA2 and FALCON-meta were compiled in a final dataset using data mining through the R programming language. This dataset was built using the following procedure (Fig. 1).

1. Curation of the dataset for human bacterial and viral pathogens (HumanPathogensDB) using the data from Liam P. Shaw et al., 2020 (Fig. 1-A).
2. Database creation using R language with information that merges the results from GOTTCHA2 and FALCON-meta, considering the information in HumanPathogensDB. The species name was used as the primary key, and the species that were not present in the sample were removed (Fig. 1-B and 1-C).
3. Compilation of the final dataset combining the results for the relative abundance (GOTTCHA2 normalized abundance) and percentage of similarity (FALCON-Meta) of each human pathogenic species.
4. Exportation of the results to an optimized and curated Excel sheet that was statistically analysed using the occurrence of each taxonomic entity (species) in both datasets as criteria. The accuracy of each unique result was evaluated considering the species with higher values of relative abundance (>0.0001) and conservation scores (percentage of similarity >60%). Pearson's correlation, p values, and confidence intervals were calculated using nonparametric correlations that follow the z-approximation (Hollander et al., 2015).
5. The same analysis procedure was performed considering a simulated metatranscriptomic dataset (MT1) from the MOSCA software pipeline (Sequeira et al., 2019) with known pathogenic species and their relative abundance.

2.2.2. Functional annotation

The functional annotation of the metatranscriptome assembly was achieved by using RASTtk software (Brettin et al., 2015). The virulence genes in the sample were analysed to confirm the presence of putative human viral and bacterial pathogens. The trimmed single reads file was assembled with SPAdes (Bankevich et al., 2012). The parameters to run were the default considering the genetic code of most bacteria and viruses. The results were exported from the Kbase platform to a CSV format. R scripts and the in-house Circos algorithm were also implemented to retrieve the microbial diversity statistics of the annotated genes for each metatranscriptome sample.

3. Results and discussion

3.1. Metatranscriptome data

The metatranscriptomic files of the collected wastewater samples were analysed, and after the removal of low-quality sequences with trimmomatic, 125, 326, 964 reads were obtained from a total of 125, 874, 982 input reads. FASTQC retrieved 0 sequences flagged as poor quality with a sequence length between 36 and 100 nucleotides and 51% GC content. The control transcriptome that contained segments of the SARS-CoV-2 genome was positive for the presence of this virus when using GOTTCHA2 and FALCON-meta analysis. The results allowed to confirm that both software tools can detect SARS-CoV-2.

3.2. GOTTCHA2 detected pathogens

We determined the presence of potential human pathogens in the sample as calculated by GOTTCHA2. The identified human pathogens were represented by bacteria, with only one virus detected (*Mamastrovirus 1*). The species *Laribacter honkongensis* (relative abundance of 0.004), *Arcobacter butzleri*, *Streptococcus suis* and *Bacteroides uniformis* presented the highest relative abundance among the human pathogens detected, although with a low global read count among the total species detected (Fig. 2-a and Fig. 3, Table 1). The total mapped base pairs (bp) were in accordance with the relative abundance results, although the species *Arcobacter butzleri* mapped bp was lower than expected from the relative abundance results. Proteobacteria pathogenic species showed a higher number of reads (504,344 from a total of 4,451,520). The phylum Proteobacteria was the most represented among all detected species (n = 72), mostly with facultatively anaerobic metabolism species (Fig. 2-

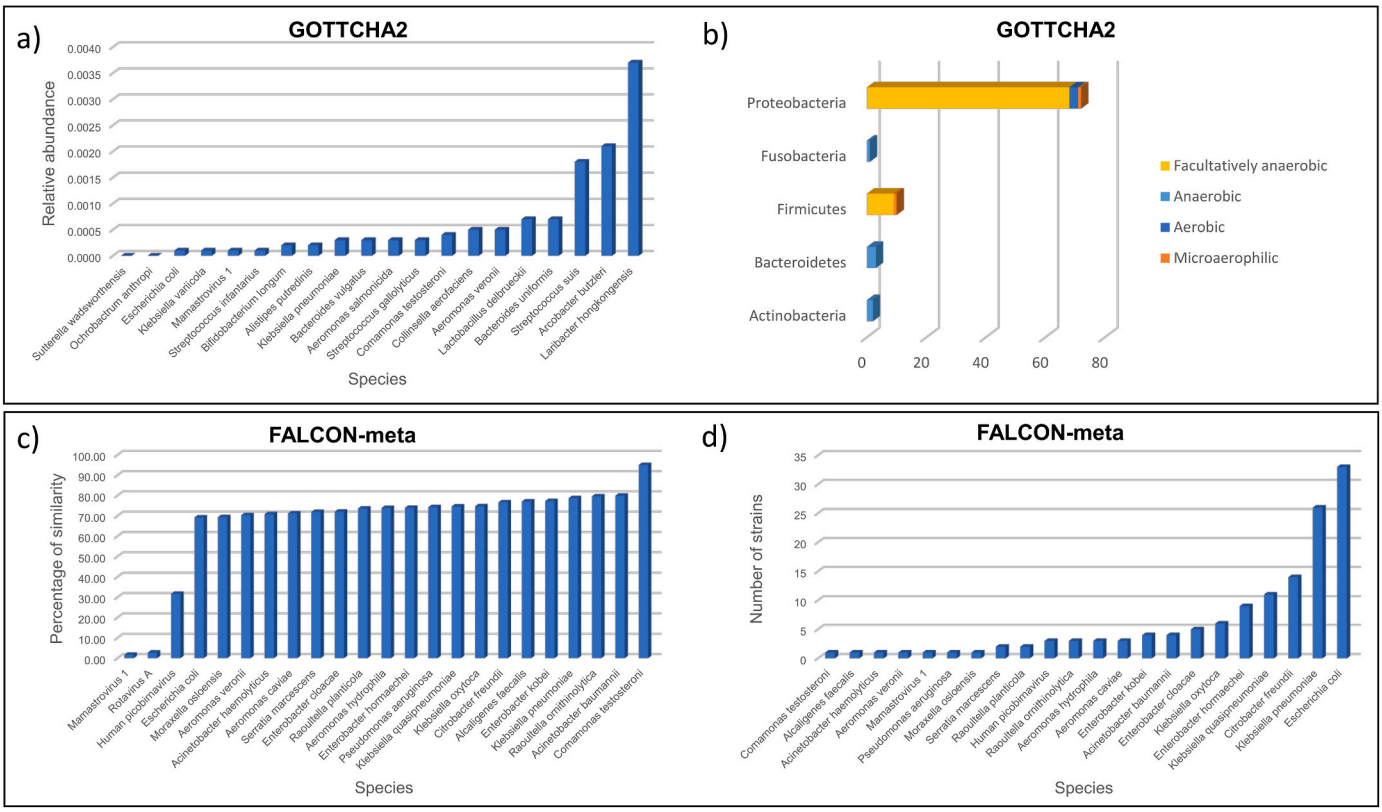

Fig. 2. Fig. 2: a) GOTTCHA2 sum of relative abundance of the Human pathogenic species present in the wastewater sample; b) GOTTCHA2 detected Phyla for the Human pathogenic species present in the wastewater sample, considering the metabolic dependence of oxygen; c) FALCON-meta percentage of similarity of the Human pathogenic species present in the wastewater sample; d) FALCON-meta number strains of the Human pathogenic species present in the wastewater sample.

b). The phyla Fusobacteria, Firmicutes, Bacteroidetes and Actinobacteria were also present in the analysed sample. The dataset also included mainly gram-negative pathogenic bacteria (n = 72) and only 12 g-positive bacterial species. The only virus identified with pathogenicity potential was *Mamastrovirus 1* (Fig. 3). The raw GOTTCHA2 values combined with the pathogen database are available in Supplementary Table S1.

3.3. FALCON-meta detected pathogens

To validate the potential human pathogenic species detected the FALCON-meta analysis of the metatranscriptomic data was performed. The results were in accordance with GOTTCHA2 (Fig. 4, Supplementary Table S2), but different strains inside each species were also detected (e.g., 32 strains for *Escherichia coli*). These strains were not detected in GOTTCHA2 calculations. The FALCON-meta analysis revealed that although some bacterial orders were detected in the GOTTCHA2 calculations, it is highly probable that they are not present in the sample (Fig. 5). The percentage of similarity in these cases was low (Fig. 6); consequently, the FALCON-meta data only included 4 orders from GOTTCHA2 detected species data. Furthermore, the species with pathogenicity potential with the highest similarity percentage was *Comamonas testoteroni* (more than 90%), while the other bacterial species had similarity percentages in the range of 75%–80% (Fig. 6). In contrast, the viral species had the lowest similarity percentage (lower than 10%), except for the record with reference AF246940.1, corresponding to the human picobirnavirus. All the detected bacteria were gram-negative (23 species). Within the 4 bacterial orders that were found in the GOTTCHA2 and FALCON-meta results data, several species were unique to the FALCON-meta data (Fig. 2-c and 2-d). For instance, the sample presented 11 potentially human pathogenic species from the order Enterobacterales, while the GOTTCHA2 results only presented 3 potentially Enterobacterales pathogenic species. Of these 11 species, *Raoultella ornithinolytica*, *Acinetobacter baumannii* and *Comamonas testosterone* presented the highest similarity percentages. The facultatively anaerobic bacteria from the genus *Klebsiella* (n = 43 strains from a total of 2061 strains), which included the species *Klebsiella pneumoniae*, showed an average percentage of similarity of 76%. The species *Escherichia coli* (n = 33 strains) was also detected in the sample with an average percentage of similarity of 69%.

Concerning the potential human pathogenic viral species present in the sample, the computational workflow detected human picobirnavirus, rotavirus A and *Mamastrovirus 1*. The average percentages of similarity (<32%), as calculated by FALCON-meta, were very low for Rotavirus A and *Mamastrovirus 1*, which suggests that these species are not actually present in the sample. Only human picobirnavirus was detected with a percentage of similarity of approximately 92% for one of the detected nucleotide segments.

3.4. Metatranscriptomic approach final dataset

To improve the accuracy of the computational metatranscriptomic approach a combination of the results from different algorithms was performed. The results from the combination of GOTTCHA2 and FALCON-meta analysis data detected a large number of viruses and bacteria. Nevertheless, the statistical analysis of the dataset showed that only a small number of potential pathogenic bacterial and viral species were present in the analysed sample. After using the filter criteria considering the relative abundance (>0.0001) and percentage of similarity (>60%), the final dataset revealed a total of 4 bacterial pathogen species, which included a total of 64 strains in the wastewater sample (Supplementary Table S3). The potential pathogenic bacterial species detected were *Escherichia coli, Comamonas testosteroni, Aeromonas veronii, and Klebsiella pneumoniae*. The species identified were affiliated

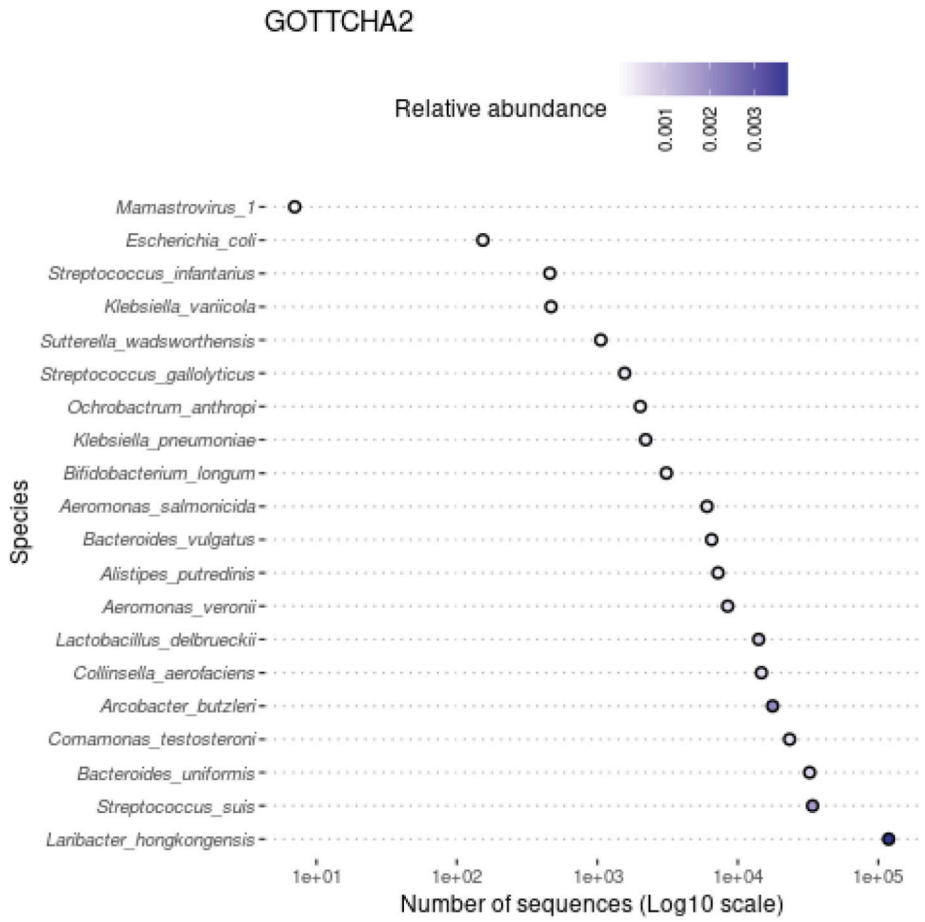

Fig. 3. GOTTCHA2 results indicating the potentially pathogenic species, based on a human pathogen database. The total number of sequences (in Log 10 scale) is filled with a gradient proportional to relative abundance.

Table 1
GOTTCHA2 relative abundance (>0.00004), read count, mapped base pairs and base pair mismatch of human viral and bacterial pathogen species detected in the analysed wastewater sample.

Species	RELATIVE ABUNDANCE	TOTAL_BP_MAPPED	AVERAGE of TOTAL_BP_MISMATCH	AVERAGE of READ_COUNT
Laribacter hongkongensis	0.00373	8,727,449	80,874	118,511
Arcobacter butzleri	0.00211	970,374	10,441	17,710
Streptococcus suis	0.00177	2408773	61,726	34,107
Bacteroides uniformis	0.00069	2476792	16,163	32,494
Lactobacillus delbrueckii	0.00068	1045996	9305	14,080
Collinsella aerofaciens	0.00055	1124284	21,490	14,737
Aeromonas veronii	0.00045	491,222	8340	8473
Comamonas testosteroni	0.00041	1464441	22,275	23,404
Streptococcus gallolyticus	0.00031	117,383	1841	1566
Bacteroides vulgatus	0.00028	411,059	2289	6509
Klebsiella pneumoniae	0.00027	113,803	570	2211
Aeromonas salmonicida	0.00025	448,160	5963	6032
Alistipes putredinis	0.00024	531,016	2088	7236
Bifidobacterium longum	0.00017	207,064	1013	3097
Klebsiella variicola	0.00011	24,498	186	468
Mamastrovirus 1	0.00008	516	20	7
Streptococcus infantarius	0.00007	25,782	172	459
Escherichia coli	0.00006	9792	41	153
Sutterella wadsworthensis	0.00004	90,624	419	1058
Ochrobactrum anthropi	0.00004	154,480	1424	2024

with different taxonomic groups, which included the orders Aeromonadales, Burkholderiales and Enterobacterales. Durnavirales and Stellavirales virus orders were also detected (human picobirnavirus and mamastrovirus). The average percentage of similarity for the bacterial strains detected was above 69%. The bacterial species with the highest average conservation score was *Comamonas testosteroni* (95.07%). Human picobirnavirus was the only viral species that had a high probability of being present in the sample due to the higher values of relative abundance (0.00011) and percentage of similarity (91.89%). The final database showed a linear positive correlation (Fig. 7, Pear'on's r correlation coefficient = 0.342, $p < 0.001$) between the relative abundance and the percentage of similarity when considering the different species

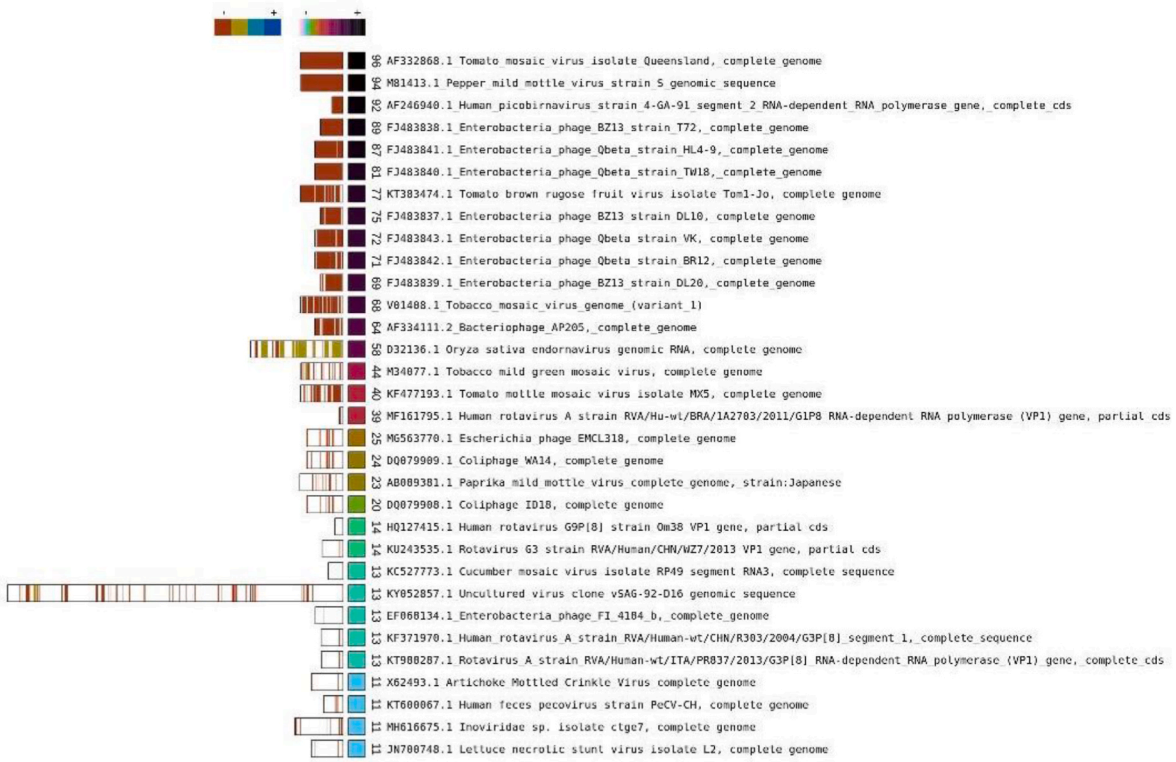

Fig. 4. FALCON-meta results showing all species with the highest percentage of similarity, including the pathogenic species. The size and percentage of similarity of human viral and bacterial pathogens detected in the sample are shown.

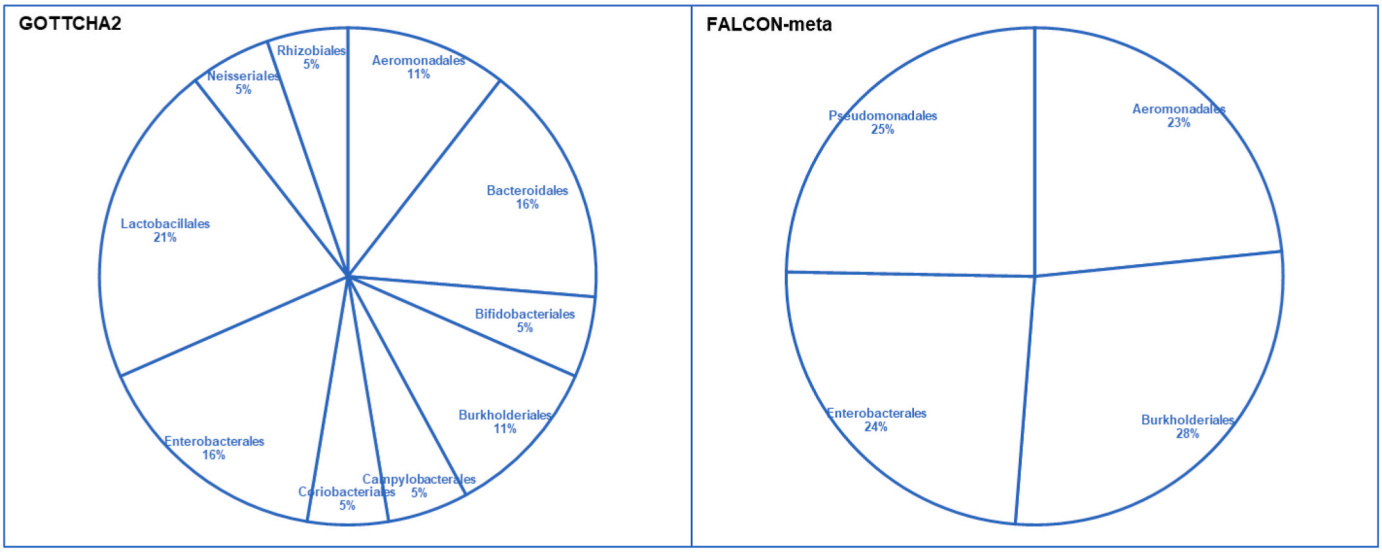

Fig. 5. Percentage of human pathogenic bacteria present in the analysed wastewater sample, considering the taxonomic rank order, as calculated by the GOTTCHA2 and FALCON-meta algorithms.

strains.

To understand which species were present in the studied sample a deep analysis was performed and discussed. Species such as *Escherichia coli* were detected in the sample, as observed in other data from wastewater pathogen detection methods (Ramírez-Castillo et al., 2015). From all *Escherichia* coli strains with a high probability of occurrence in the analysed sample, the *Escherichia coli* O157:H7 strain was detected, which can cause disease in humans by producing Shiga-like toxins 1 and 2 (Fijalkowski et al., 2014).

Comamonas testosteroni was also identified in the sample. This microorganism was reported to have caused some human infections, although with low virulence (Tiwari and Nanda, 2019). Some strains can be used for bioremediation processes due to their ability to degrade various organic pollutants (Li et al., 2017). *Comamonas testosteroni* strains have been isolated from diverse environments, in accordance with the obtained results (Ma et al., 2009). The strain T5-67, identified in the sample, was associated with the horizontal spread of integrons within the aerobic biofilm bacterial community (Huyan et al., 2020), which can have important implications for wastewater treatment.

Previous studies have already detected *Aeromonas veronii* in

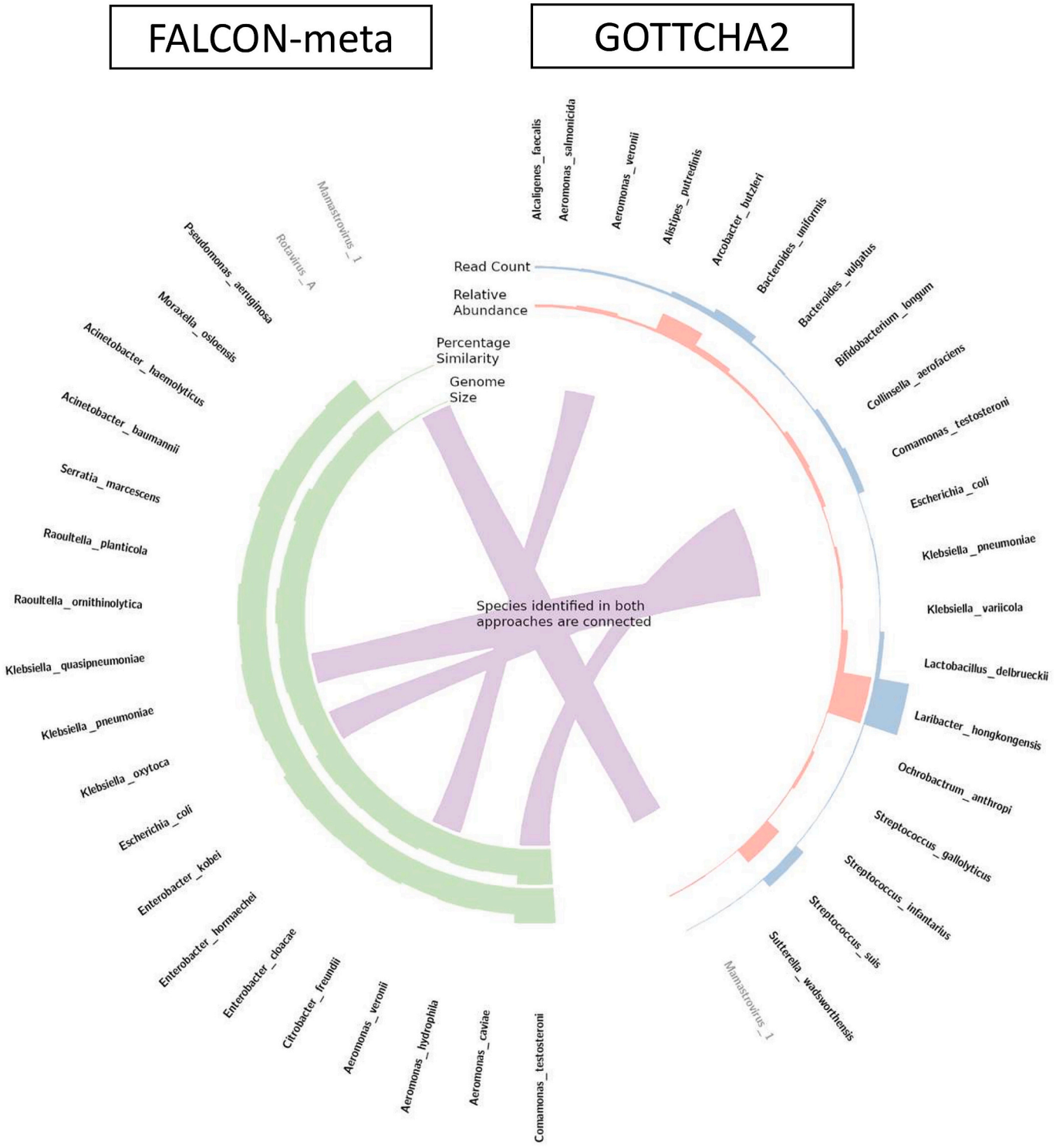

Fig. 6. Circular figure illustrating the potentially pathogenic species identified with GOTTCHA2 and FALCON-meta. The shared species are illustrated with inner ribbons. For the FALCON-meta data, the genome size and percentage of similarity are illustrated. For the GOTTCHA2 data, the read count and relative abundance are illustrated.

wastewater (Skwor et al., 2020). In this context, the *Aeromonas veronii* strain WP8–W19-CRE-03 identified in the analysed sample is antibiotic resistant and potentially highly pathogenic due to the multiple antibiotic resistance proteins identified in the sample (Supplementary Table S4), including the tetracycline resistance regulatory protein. The results findings related to antibiotic resistance in this sample are crucial considering that even treated wastewater can become a reservoir of these resistant bacterial strains (Figueira et al., 2011; Skwor et al., 2020), with implications for public health. Public health measures can be advised considering that some Aeromonas spp. Strains cause several types of diseases, including intestinal, blood, skin and soft tissue and trauma-related infections (Figueira et al., 2011; Lamy et al., 2009).

Some types of wastewaters were already associated with hotspots of antibiotic resistant bacteria (ARB), including the *Klebsiella pneumoniae*

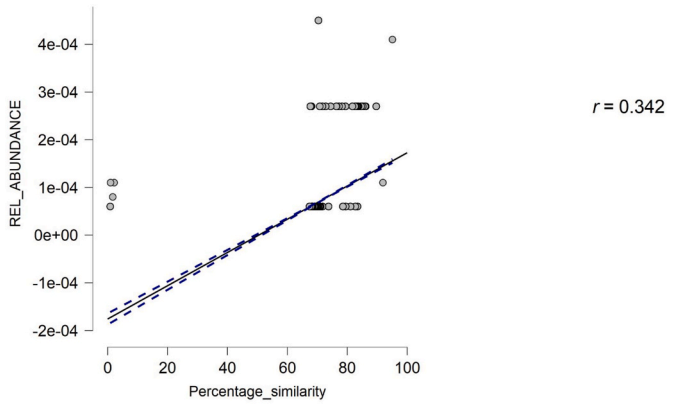

Fig. 7. Scatterplot with correlation between GOTTCHA2 relative abundance (REL_ABUNDANCE) and the FALCON-meta percentage of similarity of the detected species. Pearson's r (correlation coefficient) = 0.342, $p < 0.001$. The dotted line represents the confidence intervals (95%) as calculated by JASP.

detected in the analysed sample (Gatica et al., 2016; Kumar et al., 2020; Popa et al., 2021; Rozman et al., 2020). Considering that the bacteria *Klebsiella pneumoniae* can cause high morbidity and mortality rates due to human infections (Bassetti et al., 2018), its detection is mandatory in light of putative high antibiotic resistance. The tetracycline resistance genes were detected and can be associated with the *Klebsiella pneumoniae strains*. These strains can survive in different wastewater environments, and studies have demonstrated that after chlorine treatment, wastewater samples can present 80% tetracycline resistance genes (Popa et al., 2021).

The opportunistic enteric pathogen human picobirnavirus was identified in the wastewater sample. This type of virus was observed in other studies, including the human picobirnavirus strain 4-GA-91 (Bhattacharya et al., 2007; Ghosh and Malik, 2021; Symonds et al., 2009; Zhang et al., 2015). Human gastroenteritis is often associated with this type of virus (Malik et al., 2014), and thus, the identification of picobirnavirus should be evaluated in an epidemiological context. This matter is of importance for both raw wastewater samples and final effluent samples (Symonds et al., 2009) allowing the implementation of directed and fast public health measures if needed. The interpretation of these results should also be made at light of new existing hypothesis that propose that picobirnavirus is not an animal infectious virus but rather they may infect evolutionarily microorganisms that live and thrive in the gastrointestinal tract (Wang, 2022).

3.5. RASTtk virulence genes detection

In order to ascertain if the detected potential pathogenic species still have the capacity to infect human host an analysis of the virulence genes in the sample was performed. The results of the RASTtk (Supplementary Table S4) showed that the putative active forms of RNA detected in the wastewater sample were forms of the RNA-directed RNA polymerase beta chain. There was also the presence of genes that encode different protein forms of *thioredoxin*, Large subunit (LSU) ribosomal protein L10p (P0) and *rubrerythrin*, with a high gene count. These proteins are related to cell division initiation-related clusters, ribosomal proteins, single-copy ribosome LSU, bacterial bacterioferritin and proteins with encapsulation of dye de-colourising peroxidase or ferritin-like protein oligomers. Considering the bacterial genes associated with pathogenicity, the computational method identified the *Bacillus subtilis* spore coat *staphylococcal* pathogenicity islands (SaPI) (n = 29), the guanosine monophosphate synthetase (GMP) SaPI (n = 17), and the heat shock dnaK gene cluster extended SaPI trans-translation by stalled ribosomes. The tetracycline resistance regulatory protein (TetR) (n = 2) was detected in the sample (Figueira et al., 2011; Igbinosa and Okoh, 2012), which putatively is associated with *Aeromonas veronii* and other antibiotic-resistant bacteria.

Considering the final functional annotation of all the viruses, the putative virion core protein (lumpy skin disease virus) was the only viral protein identified in the sample. There were no relevant results for other RNA forms and protein genes considering the viruses.

3.6. Simulated dataset control

The results of the simulated dataset retrieved from the MOSCA software pipeline revealed that the methodology developed in this study can accurately detect 58% of the bacterial and viral microorganism of the MT1 simulated metatranscriptomic file (Table 2). The values for relative abundance and percentage of identity for the simulated dataset were also calculated by the implemented methodology. Considering the obtained values, the species not detected (false negatives) in our workflow were a consequence of the accuracy test performed for each unique result. The evaluated results for the simulated dataset eliminated some species with lower values of relative abundance (<0.0001) and conservation scores (percentage of similarity <60%).

3.7. Example of use

The procedure to analyse a sample with the computational metatranscriptomic workflow is straightforward and is explained as follows. The R language script code and FALCON-meta algorithm are available at https://github.com/waterpt/watermonitor. The MT1 simulated data from the MOSCA pipeline in the KBase public workflow was also included in the Github project. The following steps should be taken to analyse the samples.

1 Create a profile account at the KBase (https://www.kbase.us/) online platform to run the GOTTCHA2 software. The workflow to run GOTTCHA2 for the simulated database from MOSCA can be replicated using information at https://narrative.kbase.us/narrative/128450.

Table 2
Results for the MT1 simulated dataset from the MOSCA software pipeline.

Taxonomy	Simulated for MT1 (%)	Identified in MT1 MOSCA (%)	Detected in Workflow	Relative Abundance
Acinetobacter	0	2.585	Yes	0.00002
Aeromonas hydrophila	0	0.534	No	0
Bacteroides thetaiotaomicron	0	1.193	Yes	0.02859
Chloroflexus	0	0.283	No	0
Clostridium botulinum	0	0.892	No	0
Desulfovibrio	7.074	6.394	Yes	0.00003
Desulfumoronadaceae	3.459	1.237	No	0
Escherichia	0	8.088	No	0
Eukaryota	0	0.045	No	0
Geobacter	1.415	1.167	Yes	0.00002
Methanobacteriales	0	0	Yes	0.12525
Methanomicrobiaceae	10.07	2.098	Yes	0.14647
Methanosarcina	30.94	12.74	Yes	0.14898
Methanospirillum	6.564	0.783	Yes	0.00146
Methanothrix	20.59	5.73	Yes	0.24237
Peptococcaceae	6.288	5.853	No	0
Pseudomonas	0	4.384	Yes	0.00034
Spirochaetia	0	4.557	Yes	0.00002
Staphylococcus	0	17	No	0
Synergistaceae	0	0.707	No	0
Syntrophaceae	3.616	2.525	Yes	0
Syntrophobacteraceae	1.018	0.147	Yes	0.00139
Syntrophomonadaceae	4.717	9.536	No	0
Viruses	0	0.024	Yes	0.48681

2 Install the FALCON-meta algorithm (https://github.com/cobilab/falcon) using ANACONDA (https://www.anaconda.com/, available for all operating system platforms). In Windows, the Linux subsystem must be installed. All software is open source.
3 Run the samples with GOTTCHA2 and FALCON-meta and save the results. FALCON-meta command should be:"./FALCON -v –F -t 15 -l 47 -x output_file.txt transcriptome_example.fasta your_reference_database.fasta".
4 To merge the results from both datasets with the pathogen reference file (vertebrates_pathogens.csv), the MERGE_TABLES_GOTTCHA2-FALCON script should be run.
5 One possible illustration (Fig. 6) is generated using circos software (circos.ca). To produce the example in Fig. 6, run the script circos.conf after producing all the necessary files in the "watermonitor/Circos" directory. To produce those files, the circos_pre_processing R script, available in the "watermonitor/Circos" directory, should be used. This step is an example tailored for the data presented here, and the user should adapt it to its own data. Additional visualization examples are made available in the R script.
6 The CIRCOS R script should be run to generate the graphic representation of the detected putative pathogenic strains.

3.8. Performance tests and algorithm improvements

The accuracy of the new FALCON-meta tool was evaluated using different c parameters (c = 30, c = 40, c = 50, c = 60). These results showed that FALCON-meta performance, even with the lowest memory usage, using an Intel i7 CPU, 16 GB RAM, and 512 SSD workstation, allowed the improvement of the results of GOTTCHA2 (Freitas et al., 2015). FALCON-Meta was tested with a limited number of CPUs and low RAM size. These experiments suggest that a minimum of 8 CPUs/8 GB of RAM are needed to efficiently process large sequence files from NCBI's nonredundant bacteria database (FASTA file with approximately 700 GB) and virus database.

Considering the comparison with other tools, two features of these metatranscriptomic approach methodology are important: the automatic identification of potential pathogenic microorganisms (e.g., virus, bacteria) and the possibility of differentiating species strains (Table 3). The differentiation of strains is of major relevance since only with this information the prediction of the pathogenicity of the microorganisms present in the sample is possible.

3.9. Computational metatranscriptomic approach limitations

Considering the final analysis, only one putative pathogenic virus was detected, and the number of detected viruses was lower than the bacterial species, which can be explained by our filtering criteria. Other reasons were outlined before to explain this difference between the detected virus and bacteria (Shakya et al., 2019). Additionally, the detection of bacteria and viruses can also be influenced by the methods used for nucleic acid extraction and sequencing. In this context, other studies implemented viral metagenomics separately from bacteria metagenomics (Petrovich et al., 2020). Considering this, the metatranscriptomics workflow implemented here should not be used for routine identification of viruses in wastewater until further optimization of both the sequencing procedure and the statistical validation using more samples. Finally, since the main results were obtained from a single sample, interpretation of the conclusions should be careful, although the workflow was validated with simulated data and controls. The comparison of the detected human pathogens should be done with other studies when the computational metatranscriptomic approach developed in this study can be tested as part of different monitoring systems.

3.10. Computational metatranscriptomic approach future perspectives

The data obtained using this computational metatranscriptomic approach can also have direct implications in the development of vaccines and therapeutic approaches for human pathogens, since early detection of new strains of human pathogenic bacteria and viruses is possible ("Wastewater monitoring comes of age," 2022). Furthermore, this type of analysis can complement clinical surveillance during human pathogens outbreaks showing a comprehensive view of infection burden and transmission and information on variants that are circulating in a community (Diamond et al., 2022). This high-resolution wastewater data can also be combined with information from ecosystems maps describing the distribution of habitats and species, including humans, to calculate where the impacts of wastewater pressures are highest and by this way establish conservation efforts (Tuholske et al., 2021).

However, this computational methodology also imposes some discussion about the ethical and privacy concerns (Jacobs et al., 2021). Currently, the fast analysis of this huge amount of data in almost real-time allows the identification and understanding of the population viral spread and disease trends. Over this, advances in the high-capacity computing resources, machine learning, as well as improved analytical chemistry techniques (Baum et al., 2021), can putatively allow the deeper knowledge of the transcriptomes and genomes present in different types of wastewater samples. Considering this, strategic definition of the objectives of human pathogens detection by monitoring programs should be transparent by clearly explaining the future use of the recovered information from analysed samples. Measures to protect the storage of this kind of information should also be implemented by using encryption tools and a feasible data management plan.

4. Conclusion

The computational metatranscriptomic approach implemented in this study allowed the identification of potential human pathogenic bacterial and viral species in a wastewater sample by cross validating the metatranscriptomic analysis with a database of reference human pathogens. The developed approach improved previous used methodologies (Sequeira et al., 2019; Westreich et al., 2018) considering that: (1) this computational workflow was built using freely available tools, (2) the computational tools can process the sample more rapidly and accurately, (3) the computations are reproducible, (4) the final detected human pathogens are validated by several different algorithms using statistical methods, (5) the implementation in current workflows that monitor the presence of pathogens in urban wastewater is straightforward. The presented workflow follows the best practices for metatranscriptomic analysis, including the pre-processing of the FASTQ read files and statistical validation of the results using two different tools. The features implemented in the developed workflow represent an improvement to other tools used in metatranscriptomic analysis, including the identification of different strains and the prediction of species with putative pathogenicity. The detected pathogen species can be used to ascertain some specific metabolic pathways linked to the putative active forms of RNA detected in environmental samples. The results are even more striking considering that the methodology was able to detect several multi resistant bacterial proteins associated with some bacterial strains, which can be relevant for the detection of sources of multidrug resistance, including ARBs, in wastewater. Finally, the developed workflow has a high potential for human pathogens detection, but this computational metatranscriptomic approach should not be used routinely to identify the presence of virus until further optimization and validation with several wastewater samples.

Credit author statement

João Carneiro, Francisco Pascoal, Miguel Semedo, Diogo Pratas: Conceptualization, Methodology, Software, Data curation, Writing – original draft. Maria F. Carvalho, Catarina Magalhães, Ana P. Mucha: Experimental plan. João Carneiro, Francisco Pascoal, Diogo Pratas: Visualization, Investigation. João Carneiro, Catarina Magalhães:

Table 3
Comparative description of the previous metatranscriptomic analysis workflows and the computational methodology implemented in this study.

	Metatrans	COMAN	FMAP	SAMSA2	HUMAnN2	Carneiro J et al. WORKFLOW
Preprocessing	✓	✓	✓	✓	✗	✓
Taxonomic profiling	✓	✓	✗	✓	✓	✓
Identify Pathogens	✗	✗	✗	✗	✗	✓
Identify strains	✗	✗	✗	✗	✗	✓
Pathway analysis	✓	✓	✓	✗	✓	✗
Open Source	✓	✗	✓	✗	✓	✓
Summary report	✗	✗	✗	✗	✗	✓
Implemented partially in KBASE	✗	✗	✗	✗	✗	✓

Supervision. João Carneiro, Diogo Pratas, Adriana Rego: Validation, Data curation, Software. All authors: Writing- Reviewing and Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have shared a link to all data in the paper.

Acknowledgements

We would like to highlight the strong support of Águas e Energia do Porto and the municipality of Porto, which from the early stages of the COVID-19 pandemic outbreak foresaw the importance of the water systems in this particular subject.

This research was supported by Portuguese national funds through the Foundation for Science and Technology (FCT) within the scope of UIDB/04423/2020, UIDP/04423/2020, and UIDB/04565/2020.

The Foundation for Science and Technology (FCT) funded this study through grant 2020.03139. CEECIND to CM and CEECIND/02968/2017 to MFC and two PhD scholarships to AR (SFRH/BD/140567/2018) and FP (2020.04453. BD). JC also acknowledges the FCT funding for his research contract at Interdisciplinary Centre of Marine and Environmental Research CIIMAR, established under the transitional rule of Decree Law 57/2016, amended by Law 57/2017. D.P. is funded by national funds through FCT under the Institutional Call to Scientific Employment Stimulus (reference CEECINST/00026/2018). This work was supported by INCD funded by FCT and FEDER under project 2021.09782.CPCA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2023.116040.

References

Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J.W., Choi, P.M., Kitajima, M., Simpson, S.L., Li, J., Tscharke, B., Verhagen, R., Smith, W.J.M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K.V., Mueller, J.F., 2020. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2020.138764.

Amereh, F., Jahangiri-rad, M., Mohseni-Bandpei, A., Mohebbi, S.R., Asadzadeh-Aghdaei, H., Dabiri, H., Eslami, A., Roostaei, K., Aali, R., Hamian, P., Rafiee, M., 2022. Association of SARS-CoV-2 presence in sewage with public adherence to precautionary measures and reported COVID-19 prevalence in Tehran. Sci. Total Environ. 812 https://doi.org/10.1016/j.scitotenv.2021.152597.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 https://doi.org/10.1089/cmb.2012.0021.

Bassetti, M., Righi, E., Carnelutti, A., Graziano, E., Russo, A., 2018. Multidrug-resistant klebsiella pneumoniae: challenges for treatment, prevention and infection control. Expert Rev. Anti Infect. Ther. https://doi.org/10.1080/14787210.2018.1522249.

Baum, Z.J., Yu, X., Ayala, P.Y., Zhao, Y., Watkins, S.P., Zhou, Q., 2021. Artificial intelligence in chemistry: current trends and future directions. J. Chem. Inf. Model. https://doi.org/10.1021/acs.jcim.1c00619.

Bhattacharya, R., Sahoo, G.C., Nayak, M.K., Rajendran, K., Dutta, P., Mitra, U., Bhattacharya, M.K., Naik, T.N., Bhattacharya, S.K., Krishnan, T., 2007. Detection of Genogroup I and II human picobirnaviruses showing small genomic RNA profile causing acute watery diarrhoea among children in Kolkata, India. Infect. Genet. Evol. 7 https://doi.org/10.1016/j.meegid.2006.09.005.

Bogler, A., Packman, A., Furman, A., Gross, A., Kushmaro, A., Ronen, A., Dagot, C., Hill, C., Vaizel-Ohayon, D., Morgenroth, E., Bertuzzo, E., Wells, G., Kiperwas, H.R., Horn, H., Negev, I., Zucker, I., Bar-Or, I., Moran-Gilad, J., Balcazar, J.L., Bibby, K., Elimelech, M., Weisbrod, N., Nir, O., Sued, O., Gillor, O., Alvarez, P.J., Crameri, S., Arnon, S., Walker, S., Yaron, S., Nguyen, T.H., Berchenko, Y., Hu, Y., Ronen, Z., Bar-

Zeev, E., 2020. Rethinking wastewater risks and monitoring in light of the COVID-19 pandemic. Nat. Sustain. https://doi.org/10.1038/s41893-020-00605-2.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. https://doi.org/10.1093/bioinformatics/btu170.
Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomason, J.A., Stevens, R., Vonstein, V., Wattam, A.R., Xia, F., 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5 https://doi.org/10.1038/srep08365.
Cervantes-Avilés, P., Moreno-Andrade, I., Carrillo-Reyes, J., 2021. Approaches applied to detect SARS-CoV-2 in wastewater and perspectives post-COVID-19. J. Water Process Eng. https://doi.org/10.1016/j.jwpe.2021.101947.
Crits-Christoph, A., Kantor, R.S., Olm, M.R., Whitney, O.N., Al-Shayeb, B., Lou, Y.C., Flamholz, A., Kennedy, L.C., Greenwald, H., Hinkle, A., Hetzel, J., Spitzer, S., Koble, J., Tan, A., Hyde, F., Schroth, G., Kuersten, S., Banfield, J.F., Nelson, K.L., 2021. Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. mBio 12. https://doi.org/10.1128/mBio.02703-20.
Diamond, M.B., Keshaviah, A., Bento, A.I., Conroy-Ben, O., Driver, E.M., Ensor, K.B., Halden, R.U., Hopkins, L.P., Kuhn, K.G., Moe, C.L., Rouchka, E.C., Smith, T., Stevenson, B.S., Susswein, Z., Vogel, J.R., Wolfe, M.K., Stadler, L.B., Scarpino, S.V., 2022. Wastewater surveillance of pathogens can inform public health responses. Nat. Med. https://doi.org/10.1038/s41591-022-01940-x.
Ekwanzala, M.D., Budeli, P., Unuofin, J.O., 2021. Application of metatranscriptomics in wastewater treatment processes. In: Wastewater Treatment. https://doi.org/10.1016/b978-0-12-821881-5.00008-8.
Farkas, K., Walker, D.I., Adriaenssens, E.M., McDonald, J.E., Hillary, L.S., Malham, S.K., Jones, D.L., 2020. Viral indicators for tracking domestic wastewater contamination in the aquatic environment. Water Res. https://doi.org/10.1016/j.watres.2020.115926.
Figueira, V., Vaz-Moreira, I., Silva, M., Manaia, C.M., 2011. Diversity and antibiotic resistance of Aeromonas spp. in drinking and waste water treatment plants. Water Res. 45 https://doi.org/10.1016/j.watres.2011.08.021.
Fijalkowski, K.L., Kacprzak, M.J., Rorat, A., 2014. Occurrence changes of Escherichia coli (including O157:H7 serotype) in wastewater and sewage sludge by quantitation method of (EMA) real time-PCR. Desalination Water Treat. 52 https://doi.org/10.1080/19443994.2014.887499.
Foladori, P., Cutrupi, F., Segata, N., Manara, S., Pinto, F., Malpei, F., Bruni, L., La Rosa, G., 2020. SARS-CoV-2 from faeces to wastewater treatment: what do we know? A review. Sci. Total Environ. 743 https://doi.org/10.1016/j.scitotenv.2020.140444.
Freitas, T.A.K., Li, P.E., Scholz, M.B., Chain, P.S.G., 2015. Accurate read-based metagenome characterization using a hierarchical suite of unique signatures. Nucleic Acids Res. https://doi.org/10.1093/nar/gkv180.
Garner, E., Davis, B.C., Milligan, E., Blair, M.F., Keenum, I., Maile-Moskowitz, A., Pan, J., Gnegy, M., Liguori, K., Gupta, S., Prussin, A.J., Marr, L.C., Heath, L.S., Vikesland, P. J., Zhang, L., Pruden, A., 2021. Next generation sequencing approaches to evaluate water and wastewater quality. Water Res. https://doi.org/10.1016/j.watres.2021.116907.
Gatica, J., Kaplan, E., Cytryn, E., 2016. Antibiotic resistance elements in wastewater treatment plants: scope and potential impacts. Handb. Environ. Chem. 44 https://doi.org/10.1007/698-2015-361.
Ghosh, S., Malik, Y.S., 2021. The true host/s of picobirnaviruses. Front. Vet. Sci. https://doi.org/10.3389/fvets.2020.615293.
Hollander, M., Wolfe, D.A., Chicken, E., 2015. Nonparametric Statistical Methods, Nonparametric Statistical Methods. https://doi.org/10.1002/9781119196037.
Huang, B., Jennsion, A., Whiley, D., McMahon, J., Hewitson, G., Graham, R., De Jong, A., Warrilow, D., 2019. Illumina sequencing of clinical samples for virus detection in a public health laboratory. Sci. Rep. 9 https://doi.org/10.1038/s41598-019-41830-w.
Huyan, J., Tian, Z., Zhang, Y., Zhang, H., Shi, Y., Gillings, M.R., Yang, M., 2020. Dynamics of class 1 integrons in aerobic biofilm reactors spiked with antibiotics. Environ. Int. 140 https://doi.org/10.1016/j.envint.2020.105816.
Igbinosa, I.H., Okoh, A.I., 2012. Antibiotic susceptibility profile of aeromonas species isolated from wastewater treatment plant. Sci. World J. https://doi.org/10.1100/2012/764563, 2012.
Jacobs, D., McDaniel, T., Varsani, A., Halden, R.U., Forrest, S., Lee, H., 2021. Wastewater monitoring raises privacy and ethical considerations. IEEE Transactions on Technology and Society 2. https://doi.org/10.1109/tts.2021.3073886.
Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19 https://doi.org/10.1101/gr.092759.109.
Kumar, M., Ram, B., Sewwandi, H., Sulfikar, Honda, R., Chaminda, T., 2020. Treatment enhances the prevalence of antibiotic-resistant bacteria and antibiotic resistance genes in the wastewater of Sri Lanka, and India. Environ. Res. 183 https://doi.org/10.1016/j.envres.2020.109179.
Lamy, B., Kodjo, A., Laurent, F., 2009. Prospective nationwide study of Aeromonas infections in France. J. Clin. Microbiol. 47 https://doi.org/10.1128/JCM.00155-09.
Li, J., Luo, F., Chu, D., Xuan, H., Dai, X., 2017. Complete degradation of dimethyl phthalate by a Comamonas testosterone strain. J. Basic Microbiol. 57 https://doi.org/10.1002/jobm.201700296.
Ma, Y.F., Zhang, Y., Zhang, J.Y., Chen, D.W., Zhu, Y., Zheng, H., Wang, S.Y., Jiang, C.Y., Zhao, G.P., Liu, S.J., 2009. The complete genome of Comamonas testosteroni reveals its genetic adaptations to changing environments. Appl. Environ. Microbiol. 75 https://doi.org/10.1128/AEM.00933-09.
Malik, Y.S., Kumar, N., Sharma, K., Dhama, K., Shabbir, M.Z., Ganesh, B., Kobayashi, N., Banyai, K., 2014. Epidemiology, phylogeny, and evolution of emerging enteric picobirnaviruses of animal origin and their relationship to human strains. BioMed Res. Int. https://doi.org/10.1155/2014/780752.
Menzel, P., Ng, K.L., Krogh, A., 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat. Commun. 7 https://doi.org/10.1038/ncomms11257.
Petrovich, M.L., Zilberman, A., Kaplan, A., Eliraz, G.R., Wang, Y., Langenfeld, K., Duhaime, M., Wigginton, K., Poretsky, R., Avisar, D., Wells, G.F., 2020. Microbial and viral communities and their antibiotic resistance genes throughout a hospital wastewater treatment system. Front. Microbiol. 11 https://doi.org/10.3389/fmicb.2020.00153.
Popa, L.I., Gheorghe, I., Barbu, I.C., Surleac, M., Paraschiv, S., Măruţescu, L., Popa, M., Pîrcălăbioru, G.G., Talapan, D., Niţă, M., Streinu-Cercel, Anca, Streinu-Cercel, Adrian, Oţelea, D., Chifiriuc, M.C., 2021. Multidrug resistant Klebsiella pneumoniae ST101 clone survival chain from inpatients to hospital effluent after chlorine treatment. Front. Microbiol. 11 https://doi.org/10.3389/fmicb.2020.610296.
Pratas, D., Hosseini, M., Grilo, G., Pinho, A.J., Silva, R.M., Caetano, T., Carneiro, J., Pereira, F., 2018. Metagenomic composition analysis of an ancient sequenced polar bear jawbone from Svalbard. Genes 9. https://doi.org/10.3390/genes9090445.
Ramírez-Castillo, F.Y., Loera-Muro, A., Jacques, M., Garneau, P., Avelar-González, F.J., Harel, J., Guerrero-Barrera, A.L., 2015. Waterborne Pathogens: Detection Methods and Challenges. Pathogens. https://doi.org/10.3390/pathogens4020307.
Rozman, U., Duh, D., Cimerman, M., Turk, S.Š., 2020. Hospital wastewater effluent: hot spot for antibiotic resistant bacteria. J. Water, Sanit. Hyg. Dev. 10 https://doi.org/10.2166/washdev.2020.086.
Sequeira, J.C., Rocha, M., Madalena Alves, M., Salvador, A.F., 2019. MOSCA: an automated pipeline for integrated metagenomics and metatranscriptomics data analysis. In: Advances in Intelligent Systems and Computing. https://doi.org/10.1007/978-3-319-98702-6_22.
Shakya, M., Lo, C.C., Chain, P.S.G., 2019. Advances and challenges in metatranscriptomic analysis. Front. Genet. https://doi.org/10.3389/fgene.2019.00904.
Shaw, L.P., Wang, A.D., Dylus, D., Meier, M., Pogacnik, G., Dessimoz, C., Balloux, F., 2020. The phylogenetic range of bacterial and viral pathogens of vertebrates. Mol. Ecol. https://doi.org/10.1111/mec.15463.
Sherchan, S.P., Shahin, S., Ward, L.M., Tandukar, S., Aw, T.G., Schmitz, B., Ahmed, W., Kitajima, M., 2020. First Detection of SARS-CoV-2 RNA in Wastewater in North America: A Study in Louisiana, USA. Science of the Total Environment. https://doi.org/10.1016/j.scitotenv.2020.140621.
Skwor, T., Stringer, S., Haggerty, J., Johnson, J., Duhr, S., Johnson, M., Seckinger, M., Stemme, M., 2020. Prevalence of potentially pathogenic antibiotic-resistant aeromonas spp. in treated urban wastewater effluents versus recipient riverine populations: a 3-year comparative study. Appl. Environ. Microbiol. 86 https://doi.org/10.1128/AEM.02053-19.
Symonds, E.M., Griffin, D.W., Breitbart, M., 2009. Eukaryotic viruses in wastewater samples from the United States. Appl. Environ. Microbiol. 75 https://doi.org/10.1128/AEM.01899-08.
Tiwari, S., Nanda, M., 2019. Bacteremia caused by Comamonas testosteroni an unusual pathogen. J Lab Physicians 11. https://doi.org/10.4103/jlp.jlp_116_18.
Tomasino, M.P., Semedo, M., Vieira e Moreira, P., Ferraz, E., Rocha, A., Carvalho, M.F., Magalhães, C., Mucha, A.P., 2021a. SARS-CoV-2 RNA detected in urban wastewater from Porto, Portugal: method optimization and continuous 25-week monitoring. Sci. Total Environ. 792 https://doi.org/10.1016/j.scitotenv.2021.148467.
Tomasino, M.P., Semedo, M., Vieira, P., Ferraz, E., Rocha, A., Carvalho, M.F., Magalhães, C., Mucha, A.P., 2021b. Continuous Monitoring of SARS-CoV-2 RNA in Urban Wastewater from Porto, Portugal: Sampling and Analysis Protocols, 2021. medRxiv. https://doi.org/10.1101/2021.04.06.21254994, 04.06.21254994.
Tovo, A., Menzel, P., Krogh, A., Cosentino Lagomarsino, M., Suweis, S., 2020. Taxonomic classification method for metagenomics based on core protein families with Core-Kaiju. Nucleic Acids Res. 48 https://doi.org/10.1093/nar/gkaa568.
Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., Segata, N., 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat. Methods. https://doi.org/10.1038/nmeth.3589.
Tuholske, C., Halpern, B.S., Blasco, G., Villasenor, J.C., Frazier, M., Caylor, K., 2021. Mapping global inputs and impacts from of human sewage in coastal ecosystems. PLoS One 16. https://doi.org/10.1371/journal.pone.0258898.
Wang, D., 2022. The enigma of picobirnaviruses: viruses of animals, fungi, or bacteria? Curr Opin Virol. https://doi.org/10.1016/j.coviro.2022.101232.
Warish, A., Triplett, C., Gomi, R., Gyawali, P., Hodgers, L., Toze, S., 2015. Assessment of genetic markers for tracking the sources of human wastewater associated Escherichia coli in environmental waters. Environ. Sci. Technol. 49 https://doi.org/10.1021/acs.est.5b02163.
Wastewater monitoring comes of age, 2022. Nat Microbiol. https://doi.org/10.1038/s41564-022-01201-0.
Westreich, S.T., Treiber, M.L., Mills, D.A., Korf, I., Lemay, D.G., 2018. SAMSA2: a standalone metatranscriptome analysis pipeline. BMC Bioinf. 19 https://doi.org/10.1186/s12859-018-2189-z.
Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biol. 20 https://doi.org/10.1186/s13059-019-1891-0.
Wood, D.E., Salzberg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15 https://doi.org/10.1186/gb-2014-15-3-r46.
Zhang, S., Bai, R., Feng, R., Zhang, H., Liu, L., 2015. Detection and evolutionary analysis of picobirnaviruses in treated wastewater. Microb. Biotechnol. 8 https://doi.org/10.1111/1751-7915.12239.