

RESEARCH

Open Access



# Implementation and comparison of two concentration methods to detect and characterize bacteriophages and bacterial hosts from large drinking water samples

Mathilde Duvivier<sup>1,2</sup>, Bouziane Moumen<sup>1</sup>, Yann Héchard<sup>1</sup>, Laurent Moulin<sup>2</sup>, Vincent Delafont<sup>1</sup> and Sébastien Wurtzer<sup>2\*</sup>

## Abstract

Drinking water distribution systems (DWDS) are low biomass biomes harboring a large variety of microorganisms. Much of the attention has been focused on bacteria, whose diversity and abundance in DWDS were repeatedly shown to be influenced by abiotic factors such as pH, temperature, growth inhibitors and water sources. However, little is known about biotic factors, such as bacteriophage presence, even though they are known to be present in DWDS and to influence bacterial dynamics. While bacteriophage impact has been assessed in natural environments such as oceans, little is known about the way they shape DWDS bacterial communities. To fill this knowledge gap and accessing bacteriophage diversity from such low biomass environment, the present study aimed to propose and compare two methods based on ultrafiltration and adsorption/elution methods, already used for the concentration of bacteria and virus from water. To this end, both methods were compared with a weekly sample collection, for one month, on the DWDS of Paris, France. Metagenomic sequencing was performed on concentrated samples to investigate the presence and diversity of bacteriophages, using a coupling of complementary bioinformatic prediction tools. Though viral fractions represented a minority of recovered contigs (1.5 to 2.5%), most were associated with *Caudoviricetes* class. The predicted bacterial hosts matched with the observed bacterial diversity, highlighting the robustness of host prediction tool. A total of 437 putative phages were present in all samples, constituting a core phage diversity. Among those, 380 viral contigs contained sequences showing significant non-viral matches. We leveraged this information to further refine the inference of bioinformatics pairs of bacterial hosts and their phages. In conclusion, we propose a method to simultaneously concentrate bacteriophages with bacteria from low-biomass environment. Through metagenomics, this study showed that an optimized bioinformatic pipeline could provide an overview of DWDS phage diversity. Moreover, this method allowed to detect sequence similarities between phages and bacteria, suggesting potential genetic exchanges and providing clues for host spectrum. Altogether, this study highlights the tight interactions between

\*Correspondence:  
Sébastien Wurtzer  
sebastien.wurtzer@eaudeparis.fr

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

bacteria and bacteriophages in drinking water and the possibility to study both phages and potential hosts to better grasp their intricate interplay.

**Keywords** Drinking water, Bacteriophages, Metagenomics, Filtration, Diversity, Microbiome

## Introduction

Several steps of physical and chemical treatments allow water to be safe for human consumption, but many microorganisms, including bacteria, micro-eukaryotes and viruses are still present in drinking water. This observation implies that distribution networks supplying drinking water harbor a complex microbial ecosystem. Heterotrophic bacteria in drinking water may range from  $10^3$  to  $10^6$  cells/mL [1]. While such concentrations may not pose health problems per se, it is critical to ensure that the presence of opportunistic pathogens and overall bacterial growth within networks remain tightly controlled. Beyond the plant treatment step (using disinfection such as ultraviolet light or ozone), biological stability across drinking water distribution systems (DWDS) is maintained using disinfectant residuals such as free chlorine, chlorine dioxide or monochloramine. Nevertheless, these treatments cannot inactivate all microorganisms [2], but end up strongly influencing structure and composition of bacterial communities rather than obliterating it [3]. Numerous studies [4–10] have described bacteria and micro-eukaryotes [5] in DWDS over time and across different geographical areas such as Italy [6], Spain [7] and France [8, 9]. Most studies indicate that DWDS are dominated by *Pseudomonadota* with varying proportion between *Alphaproteobacteria*, *Betaproteobacteria*, *Deltaproteobacteria* and *Gamma-proteobacteria* [6–8, 10, 11]. Some studies showed that microbial diversity can vary among different drinking water treatment plants, influenced by network material and treatment methods within the same metropolitan area [8, 10]. Thus, it is likely that each DWDS harbors a unique microbiome.

Bacterial growth in drinking water is limited by several abiotic factors including nutrient availability, water temperature, pH or the presence of growth inhibitors [1]. Furthermore, the abundance variation of specific bacterial communities can be related to biofilm formation [6], water residence time in pipes [1] and flushing [10]. While numerous studies [1, 4, 6, 10] have investigated the impact of abiotic parameters on prokaryotic community composition, little is known regarding the impact of biotic factors. Among those, predation and parasitism are expected to be significant drivers shaping bacterial communities. This has been suggested for studies documenting protist predation in water, including DWDS [1]. However, studies exploring the impact of phages on bacteria in DWDS are still lacking. Indeed, like virtually all environments, DWDS constitute a niche for various

bacteriophages. Very few studies have investigated the drinking water virome, but it has been demonstrated the presence of bacteriophages affiliated with *Microviridae* family and *Caudoviricetes* class [12, 13]. Interactions between phages and bacteria have been broadly studied in marine environments [14], in gut [15] and in soils. Bacteriophages are ubiquitous and exhibit a high diversity of virion structures, chemical properties, genome structures and present multiple infection cycles: lytic, lysogenic and pseudolysogenic [16]. Phages are important drivers of bacterial communities through processes such as bacterial lysis, the redirection of bacterial metabolism, the promotion of horizontal gene transfer and the modulation of bacterial immune systems. Lytic phages can drive density and abundance of bacterial communities and can be involved in biogeochemical cycles such as sulfur, carbon and nitrogen cycles, as shown in oceans [17]. Lytic phages can also influence the mutation rate of bacteria [16, 18–20], contributing to their genomic evolution. In addition, temperate phages directly impact bacterial evolution when their genome integrates into the bacterial genome, possibly disrupting bacterial genes.

Historically, phages have been studied using culture methods, electron and fluorescence microscopy or molecular techniques such as PCR [21]. However, these methods do not provide a global view of phage diversity. This is particularly true for culture methods which require a permissive bacterial host, able to grow under laboratory conditions. Viral diversity in drinking water has been mainly studied for detecting human viruses such as enteroviruses and adenoviruses. The low concentrations of these viruses in drinking water have called for the development of methods for concentrating large volumes of water. Adsorption/elution (AE) filtration, also called VIRADEL method, and ultrafiltration have thus been developed and applied to various types of water to study and monitor eukaryotic virus diversity [22]. AE filtration is based on electrical charge, a positively or negatively charged membrane is used and microbial particles are then eluted with an alkaline buffer with beef extract, for most of the time [23, 24]. Conversely, the ultrafiltration method is not based on charge but on size exclusion employing several types of filter geometries such as hollow, spiral wound or tubular fibers [25] which display pores with a molecular weight cut-off. A backflush elution with an inorganic chemical solution [26] is then employed to recover microorganisms from water. The AE method has demonstrated its efficiency in recovering norovirus and coliphages from

seawater [27], polioviruses and enteroviruses from drinking and river waters [28]. The UF method can facilitate the simultaneous recovery of various microorganisms, including bacteria, viruses, and parasites, from drinking and source waters [24, 26, 29]. It appears that AE method is predominantly employed for virus concentration while UF method is utilized for concentrating diverse types of microorganisms, including viruses. Both methods enable the concentration of large water volumes which is critical to investigate low-biomass environments [26], such as DWDS. Although both techniques have been tested on various types of water, they have not been extensively applied to study non-targeted bacterial or phage diversity, suggesting potential for their application, coupled with metagenomics, in elucidating phage diversity in drinking water.

With the recent advances in sequencing technology applications, especially metagenomics and computational approaches, we can detect, discover and describe an increasing number of organisms and viruses. Nevertheless, bacteriophages remain challenging to study due to their high genomic diversity and the lack of universally conserved genes. Their discovery and description require more extensive efforts compared to bacteria for instance [30], including the use of multiple bioinformatics tools such as assembly, viral sequence prediction, removal of non-viral sequences, taxonomic assignment, and host prediction [21].

Our study aimed to address this gap by comparing UF and AE methods to recover sufficient phages and bacteria from tap water of Paris to be sequenced. Additionally, this study describes phage and bacterial diversities with a developed bioinformatics analysis pipeline including taxonomic assignment of phages and bacteria, host prediction and bioinformatics pair identification. The insights gained from this study will contribute to our understanding of phage diversity in low-biomass environments such as drinking water and their interactions with bacteria, including those that may pose a risk to human health.

**Table 1** – Water samples used for this study, along with the associated concentration method

| Sample | Filtration method      | Filtered volume | Flow rate | Sample collection date |
|--------|------------------------|-----------------|-----------|------------------------|
| AE_W1  | Adsorption/<br>elution | 2417 L          | 75 to     | 06/03/2024             |
| AE_W2  |                        | 2204 L          | 110 L/h   | 06/11/2024             |
| AE_W3  |                        | 2692 L          |           | 06/17/2024             |
| AE_W4  |                        | 2567 L          |           | 06/24/2024             |
| UF_W1  | Ultrafiltration        | 300 L           | 15 L/h    | 06/04/2024             |
| UF_W2  |                        |                 |           | 06/13/2024             |
| UF_W3  |                        |                 |           | 06/18/2024             |
| UF_W4  |                        |                 |           | 06/25/2024             |

## Methods

### Drinking water sampling

Drinking water sampling was performed during June 2024 with a weekly sample collection from tap water in Ivry-sur-Seine laboratory of Eau de Paris, which is fed by Paris drinking water (Table 1). The concentration of residual disinfectants in Eau de Paris network did not exceed 0.1 mg/L in the point of distribution and the sampled tap is frequently used with a weekly purge, thus limiting the residence time to less than 24 h. Water was concentrated using two methods (Fig. 1): ultrafiltration (UF) and adsorption/elution (AE).

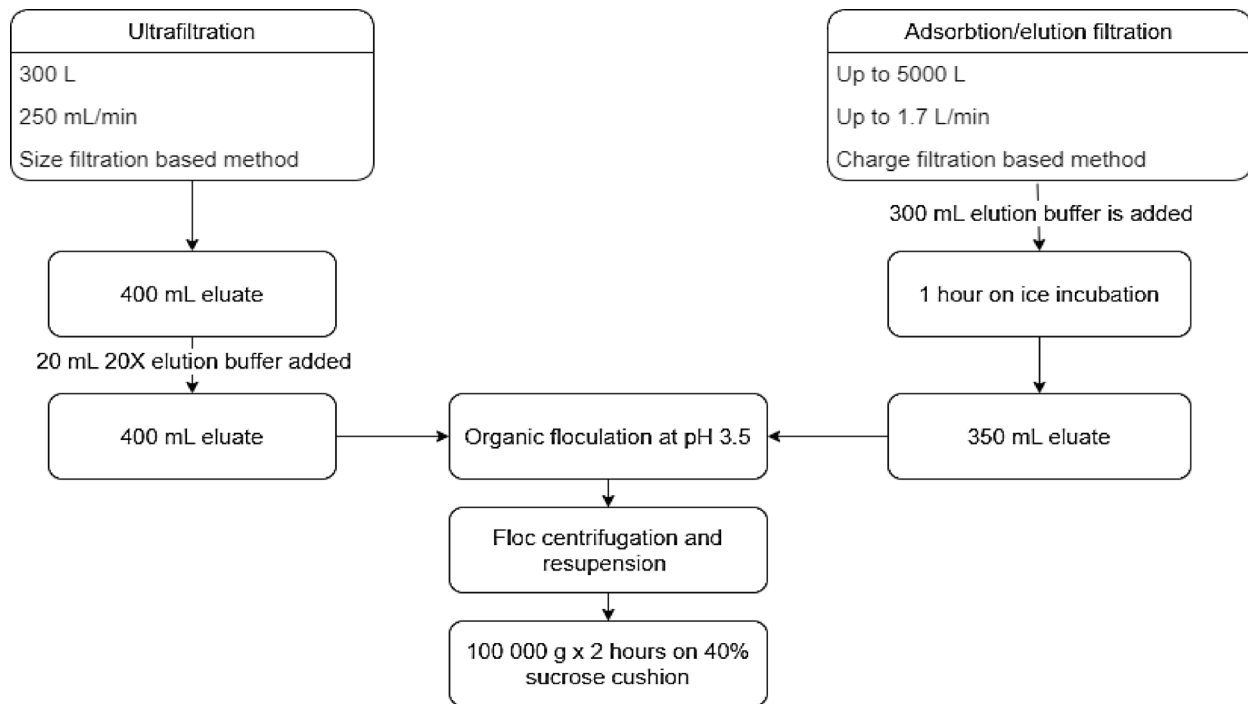
Ultrafiltrations were performed using InuvaiR180 cartridges (Fresenius, Germany). Firstly, the InuvaiR180 filter was primed with 0.05% STPP buffer (sodium triphosphate  $\text{Na}_5\text{P}_3\text{O}_{10}$ ) via a peristaltic pump. Drinking water was filtered for 17 h corresponding to *ca.* 300 L with a flow rate of 15 L/h. After sampling, the remaining drinking water inside, and outside hollow fibers was removed and the space between cartridge and fibers was filled with 0.5% STPP. Elution was then performed by backflushing 400 mL of 0.5% STPP. Finally, a 400 mL eluate was recovered. Twenty mL of an organic flocculation buffer (glycine 1 M – beef extract 20%) was then added to concentrate samples and pH was set to 3.5, allowing floc formation under gentle stirring for one hour.

Adsorption/elution was conducted using NanoCeram® filters (Argonide Corporation). The device was directly connected to a tap and drinking water filtered by monitoring flow rate and filtered volume. Filtrations were conducted during 23 to 33 h according to the flow rate comprised between 75 to 110 L/h. After reaching approximately 2500 L of filtered water, filtration was stopped, and the resident drinking water was removed from the cartridge with a pump. An elution buffer (glycine 0.05 M – STPP 0.1% – beef extract 1% – anti-foam 0.1% – Tween 80 0.1% – pH 9.5) was added to the filter and incubated on ice for one hour to desorb particles from the filter. Then, approximately 350 mL of eluate were recovered. Organic flocculation was performed by lowering pH eluates at 3.5 for one hour under gentle magnetic agitation.

After organic flocculation of UF and AE eluates, flocs were centrifuged at 4000*xg* for one hour and resuspended in 8 mL of  $\text{Na}_2\text{HPO}_4$  0.15 M pH 8.4. The resuspended floc was finally ultracentrifuged at 100 000*xg* for two hours on a 40% sucrose cushion. These filtration and concentration methods allowed to get final concentrated drinking water samples of approximately 1 mL.

### Nucleic acid extraction and Illumina sequencing

Final concentrated drinking water samples were further purified to remove beef extract and other organic particles. Six hundred  $\mu\text{L}$  of PM1 solution (Qiagen) and 6  $\mu\text{L}$  of  $\beta$ -mercaptoethanol were added to 200  $\mu\text{L}$  of concentrated



**Fig. 1** Experimental workflow developed for virus concentration

drinking water samples. Then, samples were thoroughly mixed with a high-speed benchtop homogenizer at 3 m/s for 5 min before adding 150  $\mu$ L of IRS solution (Qiagen) and incubated on ice for 5 min. Finally, samples were centrifuged at 16 000 g for 2 min to pellet organic particles. Four hundred  $\mu$ L of supernatant were collected to be extracted with the EZ1&2 Virus Mini Kit v2.0 (Qiagen) and EZ2 automatic extractor. DNA concentrations were approximately equal to 1 ng/ $\mu$ L (Additional data 1).

DNA libraries were constructed using NEBNext<sup>®</sup> Ultra<sup>™</sup> II DS DNA Library Prep Kit for Illumina<sup>®</sup> (New England Biolabs), according to the manufacturer's instructions. Sequencing was performed with the MiSeq System (Illumina) using MiSeq Reagent Kit v2 (2  $\times$  250 bp paired end).

#### Viral sequence prediction, taxonomic assignment and host prediction

Raw reads were trimmed and quality filtered using Fastp v0.23.2 [31]. Filtered reads from each sample were assembled using Megahit v1.2.9 [32]. Furthermore, because of low biomass in drinking water, a co-assembly was performed with all filtered reads of the 8 samples using Megahit v1.2.9 ( $-\text{min-count } 2$   $-\text{k-min } 21$   $-\text{k-max } 141$   $-\text{k-step } 2$   $-\text{min-contig-len } 200$ ). Co-assembling reads from multiple samples enhances the recovery of low-abundance metagenomes, reduces assembly fragmentation, and enables a more precise study of population dynamics, particularly when the microbial composition of the samples is unknown [33].

After assembly, VirSorter2 v2.2.4 [34], geNomad [35] and VIBRANT v1.2.0 [36] were used with default parameters on all assemblies (including co-assembly) to predict viral sequences. VIBRANT and VirSorter2 tools were selected for the prediction of viral contigs due to their extensive use in viral metagenomics. In addition, we incorporated geNomad, a less commonly employed tool, owing to its capacity to identify plasmids and distinguish them from bacterial and viral sequences, thereby minimizing false positives. Predicted viral sequences from co-assembly were used for further analysis, but not contigs from individual assemblies. CheckV v1.0.1 were used to perform a quality-control of the contigs predicted as viral. However, we chose to keep all predicted viral contigs because of low biomass in drinking water and the possibility to eliminate highly divergent viruses.

All filtered reads of individual samples were mapped with Minimap2 v2.26-r1175 [37], one by one, to co-assembly to calculate relative abundance (Samtools v1.18 and Reads per kilo base of transcript per million mapped reads calculation method) of contigs across the 4 weeks sampling. Alongside, viral contigs were assigned with DemoVir and geNomad, hosts were predicted with iPhoP v1.3.3 [38] and bacterial contigs were taxonomically assigned with MMseqs2 v15.6f452 against the Genome Taxonomy Database v220 released 09-RS220 (on 24th April 2024).



### Alignment between viral and bacterial contigs

A BLAST database was created using makeblastdb, based on 437 contigs predicted as viral and present in all samples with a relative abundance greater than 0.00001%. Contigs that were not predicted as viral (mostly bacterial contigs) were compared with BLAST against our viral database. When the bacterial taxonomy of the bacterial contig matched the host prediction of the viral contig blast hit, the alignment was kept. Alignment configurations were studied with Dplyr package in RStudio.

## Results

### Complementarity of viral sequence prediction tools

To assess phage diversity in drinking water, weekly tap water samples were collected and processed using two concentration methods (Fig. 1). The aim was to identify the most effective recovery method, especially for DNA bacteriophages. A total of eight samples were sequenced (Table 1) using a metagenomic approach. Additionally, we used three viral prediction tools to improve viral sequence identification.

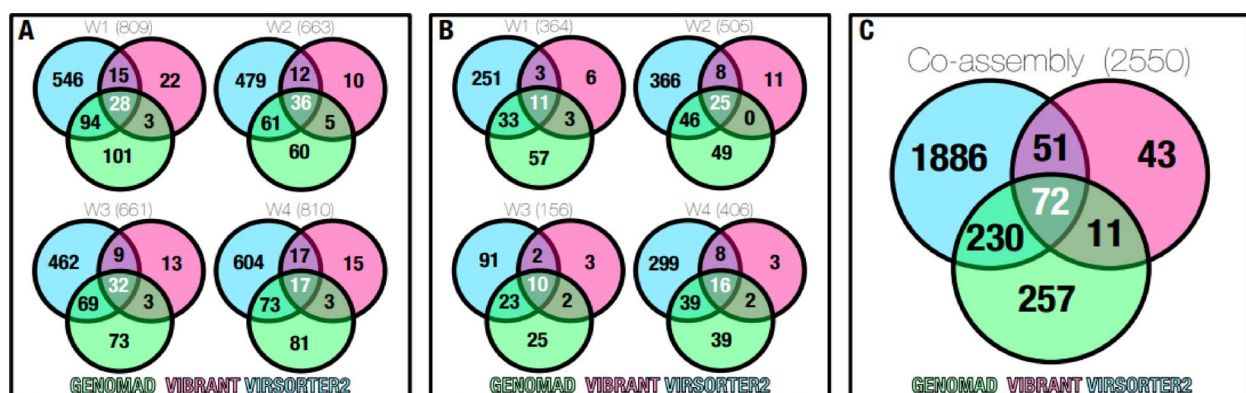
Raw reads count of all samples reached approximately 10 M for 2 Gbp (Additional data 1). After assembly, 14 726 to 86 273 contigs were obtained for each sample, with a N50 range comprised between 915 and 2945 and assembly sizes ranging from 12 to 75 Mbp (Additional data 1). Because of low biomass in drinking water and the large fraction of unknown bacteriophages preventing the accurate detection of these entities and their population dynamics, a co-assembly of all reads of the 8 samples was also performed. This co-assembly allowed to reconstitute 262 692 contigs corresponding to 210 Mbp with a N50 of 908 pb and 68 to 98% of mapped reads to co-assembly (Additional data 1). After the assembly step, viral contig prediction was performed using the combination of three tools: VirSorter2, VIBRANT and geNomad. The association of these three viral sequence prediction tools allowed improved detection of unknown bacteriophages,

while minimizing biases due to only one prediction tool use. The three tools used for viral prediction are complementary, as highlighted by the modest overlap of viral contigs identified by all three tools (Fig. 2). VirSorter2 predicted the largest number of viral contigs, followed by geNomad and VIBRANT, the latter being the more stringent tool for predicting viral sequences. The use of these three tools allowed to identify 2550 putative viral contigs from the co-assembly (Fig. 2C). All contigs predicted as viral from the co-assembly were kept to further analysis, not only common contigs predicted by the three tools.

### Comparison of both filtration and concentration methods

The prediction of 2550 viral contigs in the co-assembly (Fig. 2C) is less than the sum of the different predicted viral contigs from each individual assembly (Fig. 2A-B). This fact strongly suggests that multiple and same putative viral contigs are detected across different samples.

Only contigs from co-assembly were considered for further analysis. Relative abundances of all contigs from co-assembly were calculated: viral fraction represents 1.5 to 2.7% for the 8 samples (Additional data 1). However, it is difficult to tell which method is better to recover viruses, especially bacteriophages from drinking water, due to the heterogeneity of viral fraction between both methods. Moreover, Shannon indexes calculated on all contigs and on viral contigs showed that there was no important difference in diversity between AE and UF methods (Additional data 2). Nevertheless, a slight tendency to higher viral and global diversities for AE compared to UF can be observed (Additional data 2), probably due to the higher filtered volume in adsorption/elution method. Indeed, there were more contigs predicted as viral for AE samples (Additional data 1). However, based on filtered volumes, ultrafiltration allowed to significantly increase (Mann–Whitney test; p-value 0.0286) the recovery of putative viral contigs (Additional data 3) and seems to recover



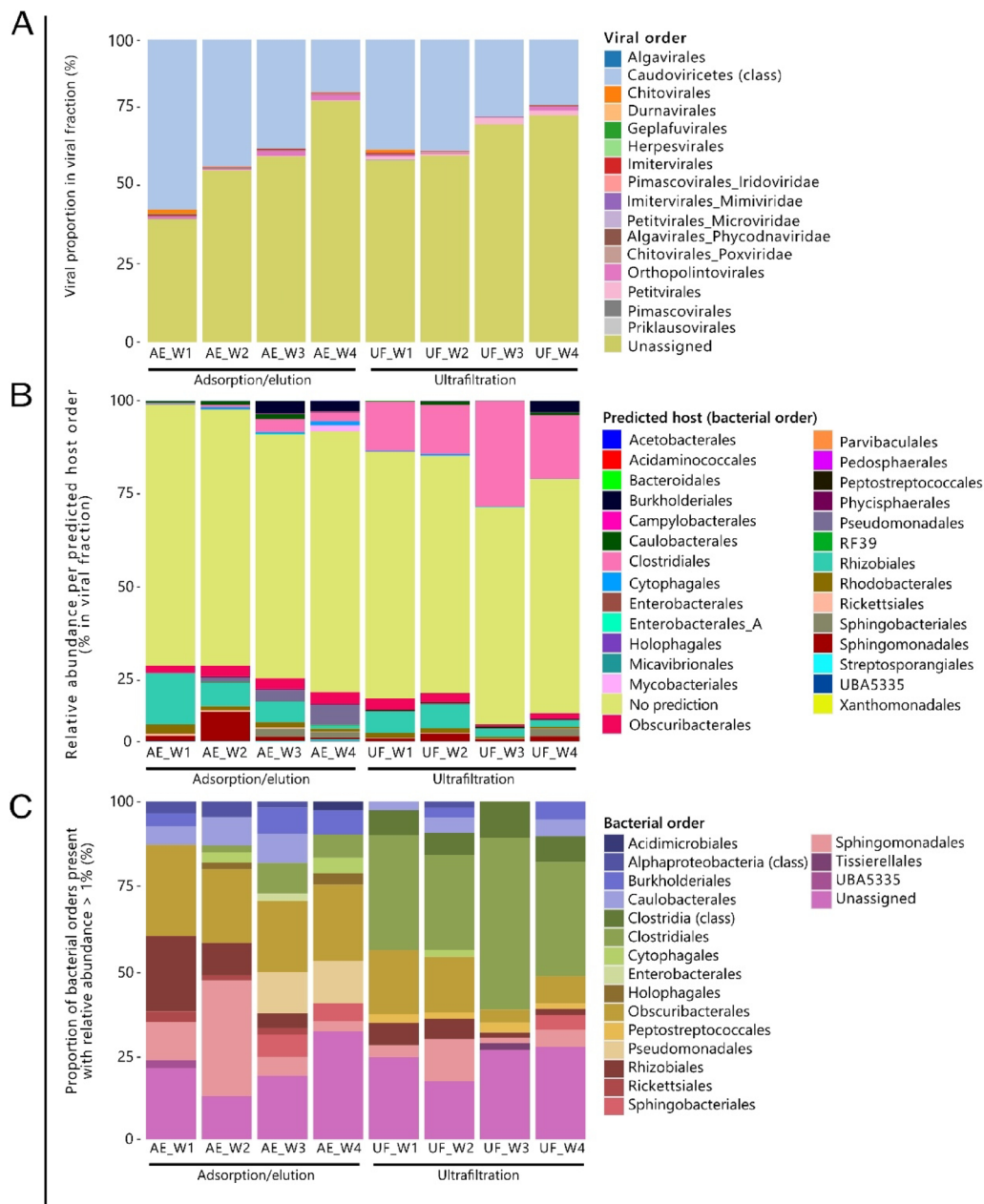
**Fig. 2** Complementarity of VirSorter2, geNomad and VIBRANT for predicting viral contigs. Viral contig prediction repartition between samples collected with **a** AE method **b** UF method for 4 weeks (W1, W2, W3 and W4). Viral prediction on **c** co-assembly of the 8 samples collected with AE and UF methods

higher diversity (viral or global) at equal volumes, compared to AE.

**Sample diversity**

After viral sequence prediction, a taxonomic assignment was performed for putative viral sequences with DemoVir and geNomad. Alongside, non-viral contigs were also taxonomically assigned with MMseqs2 and compared to host prediction of the sequences predicted as viral.

With both methods, bacteriophages from *Caudoviricetes* class were the most detected apart from an important fraction of unassigned sequences (Fig. 3A). Bacteriophages from *Petitvirales* order and *Microviridae* family were also detected but with a significantly smaller abundance than *Caudoviricetes* bacteriophages. No important differences of the viral diversity were observed between both filtration and concentration methods for taxonomically assigned contigs. Several differences were observed for host prediction between methods but not according



**Fig. 3** Most of assigned viral contigs are associated with *Caudoviricetes* class and with the most abundant bacterial orders. **a** Viral contig taxonomic assignment using DemoVir and geNomad **b** Host prediction of viral contigs using iPHoP **c** Bacterial contig taxonomic assignment using MMseqs2 for contigs with a relative abundance greater than 1%

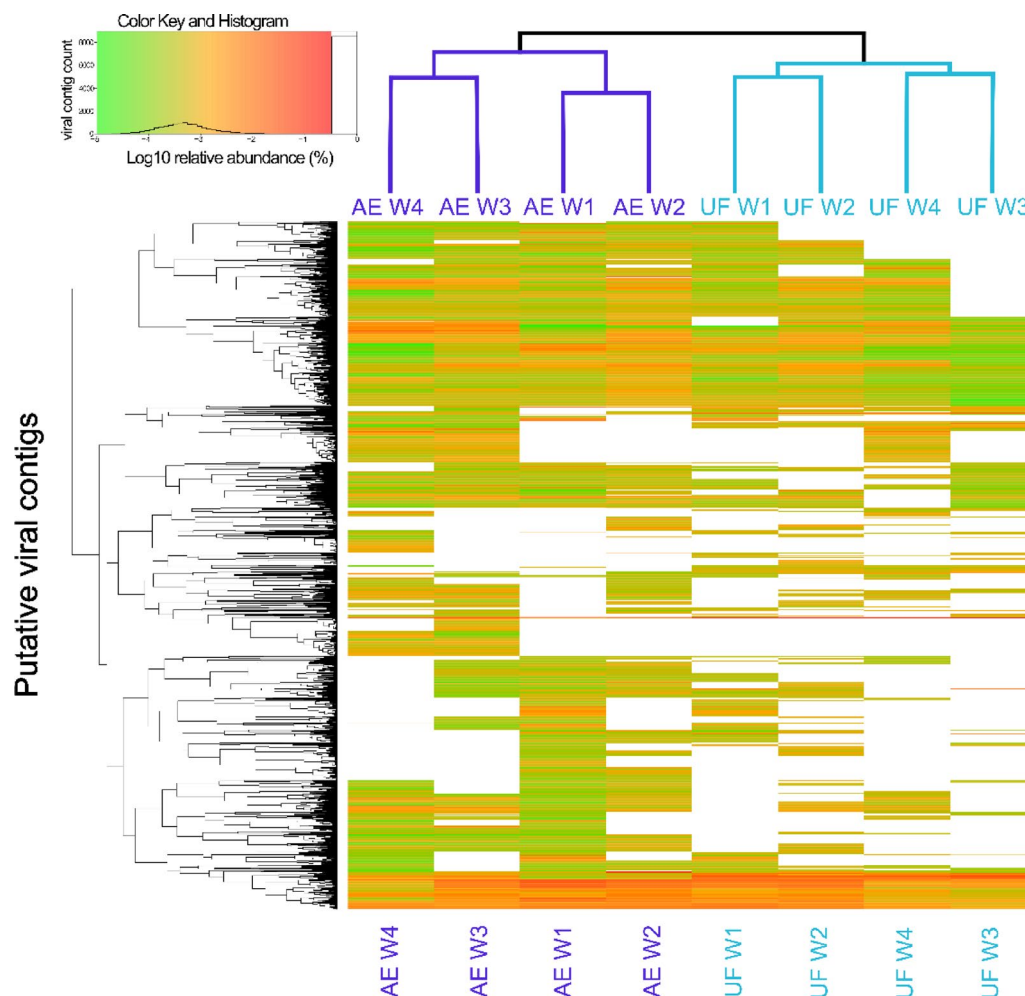
to sampling dates. Hosts belonging to *Pseudomonadales* were predicted with AE method only (Fig. 3B), while *Clostridiales* hosts were predicted far more for UF samples. Bacteria from *Rhizobiales*, *Obscuribacterales*, *Sphingomonadales*, *Sphingobacteriales* and *Burkholderiales* orders were also predicted as hosts for both methods. Regardless, a large proportion of viral contigs were not associated with any host (approximately 75%), for all samples. It is also important to specify that unassigned viral sequences can be associated with a host. On the other hand, some assigned viral sequences were not associated with any host.

Taxonomic microbial assignment of all non-viral contigs highlighted several bacterial orders present in all samples with a relative abundance greater than 1%: *Sphingomonadales*, *Sphingobacteriales*, *Rhizobiales*, *Obscuribacterales*, *Burkholderiales*, *Caulobacterales*, *Clostridiales* and *Pseudomonadales* (Fig. 3C). These bacterial orders, detected with taxonomic assignment of non-viral contigs, were also the most predicted hosts of viral contigs. This suggests that an important fraction of

phages was associated with the predominant bacterial orders. In other words, most predicted hosts were also detected in all samples. Moreover, *Clostridiales* order was much more present for UF samples, suggesting that AE method did not efficiently recover bacteria from *Clostridiales* order. This difference was correlated with the host prediction difference between both filtration and concentration methods and suggests that bacteriophages present in UF samples and associated with *Clostridiales* are intrabacterial and thus probably prophages. Predicted hosts were found in the 8 samples, highlighting a consistency between bacterial taxonomy (Fig. 3C) and host prediction (Fig. 3B).

#### Core phageome and bioinformatics pairs

To get a global view of whether viral diversity was influenced by sampling methods or sampling dates, a heatmap-dendrogram was constructed on contigs predicted as viral (2550 contigs), based on their relative abundances in the samples (Fig. 4). Samples clearly clustered according to the method used and not the sampling date,



**Fig. 4** A “core phageome” is observed among both methods and all samples, including 437 viral contigs

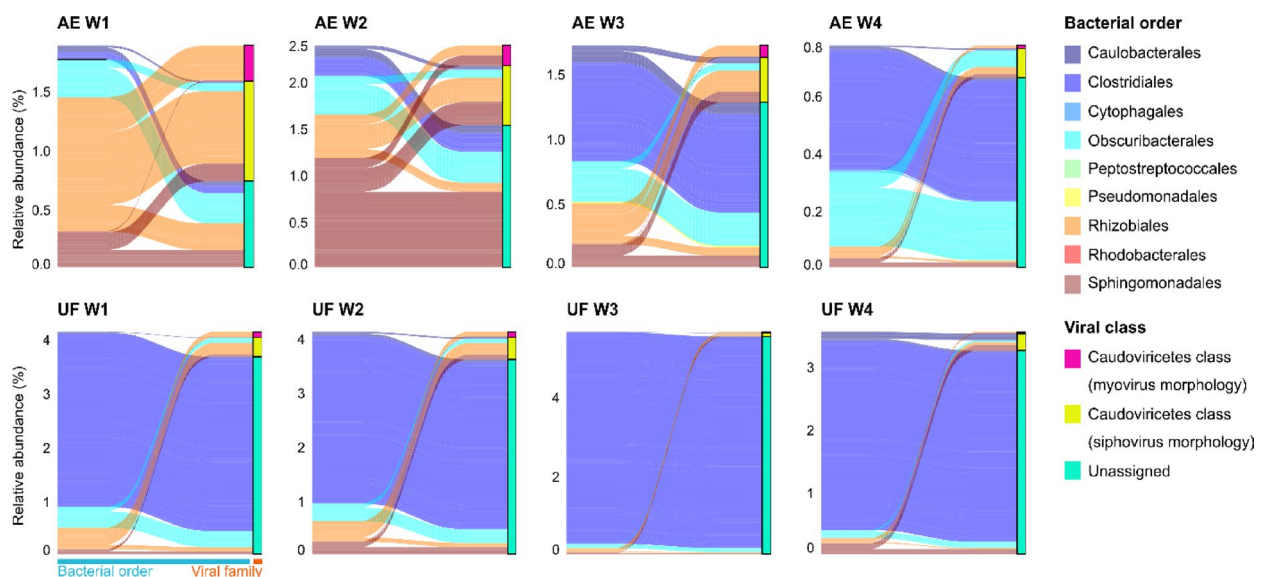
suggesting the former has the strongest impact on viral community composition. A “core phageome” was identified, defined here by viral contigs found in all samples with a relative abundance greater than 0.00001%, including 437 contigs on the 2550 contigs predicted as viral. However, considering each method separately, 1023 and 511 putative viral contigs are found in all AE and UF samples respectively, corresponding to approximately 2% of relative abundance between both methods (Additional data 4).

Because many host prediction tools are based on sequence homology and genetic exchanges occurring between bacteriophages and bacteria, we thought that some viral contigs could be aligned to other non-viral contigs in our dataset. A database with the 437 viral contigs of the core phageome were created and a Megablast of all contigs which have not been predicted as viral, was performed against the core phageome database. A total of 25 935 contigs, assigned to bacteria or unknown, were aligned to 380 viral contigs. We also found that 915 bacterial contigs were aligned to 108 viral contigs of the core phageome with a match between bacterial taxonomy and host prediction. Therefore, host prediction of viral contig matched the taxonomy of the first bacterial contig blast hit. Viral contigs belonging to *Caudoviricetes* class were associated with *Caulobacteriales*, *Clostridiales*, *Obscuribacteriales*, *Pseudomonadales*, *Rhizobiales*, *Rhodobacteriales* and *Sphingomonadales* bacterial orders for all samples (Fig. 5). However, the fraction of bacterial contigs associated with viral contigs was bigger for UF samples (relative abundances from 3.5 to 5%) than for AE samples (from 0.8 to 2.5%).

We next investigated these 915 alignments between bacterial and viral contigs (Fig. 6) according to bacterial contig length, viral contig length, alignment length and sequence homology to better describe the possibility of getting bioinformatics pairs and highlighting eventual genetic exchanges between bacteriophages and bacteria. Alignments between viral and bacterial contigs exhibited multiple configurations. Alignments can be at extremities of bacterial contigs/viral contigs or inside bacterial and viral contigs. The large part of alignments appeared to be inside bacterial and viral contigs. Moreover, alignments showed various length and sequence homology. On 915 alignments of core phageome, 340 are less than 100 bp, 323 smaller than 300 bp, 170 smaller than 600 bp, 43 less than 1000 bp and 39 are greater than 1000 bp. The majority appears to be associated with small bacterial contig length (less than 1500 bp) but do not seem to be correlated with viral contig length. Furthermore, it seemed that smaller the alignment, the better its sequence homology (95 to 100%) between viral and bacterial contigs. If we take a deeper look at long alignments (greater than 300 pb), totaling 252 alignments, 42 are associated with bacterial contigs greater than 5000 bp with sequence homologies comprised between 72.8 and 99%. These alignments could be susceptible to contain coding sequences.

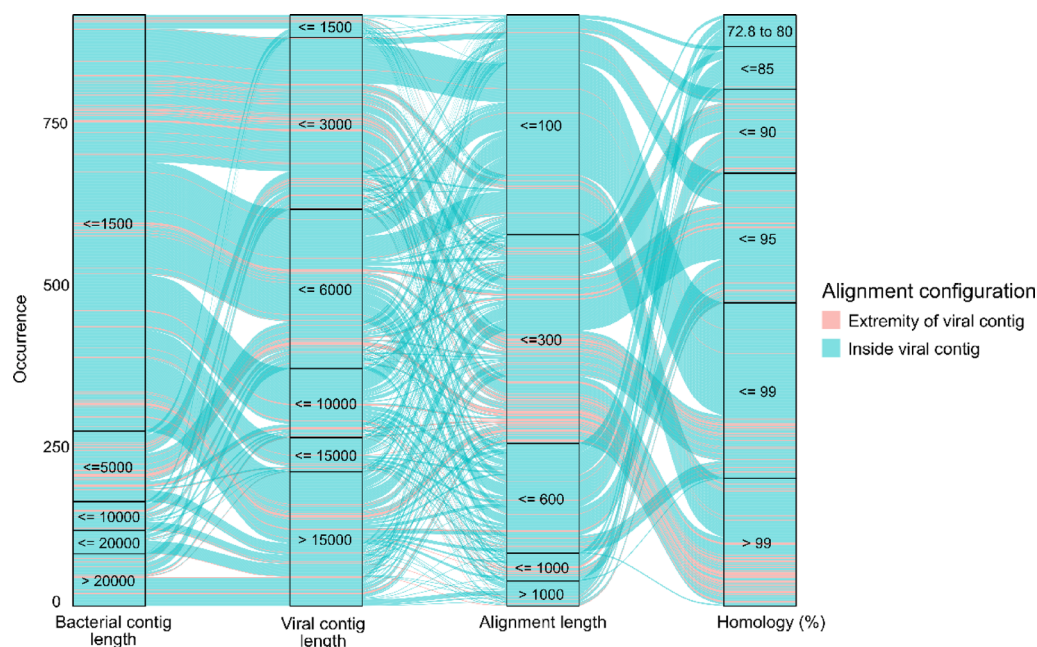
## Discussion

This study aimed to compare two filtration and concentration methods to recover bacteria and bacteriophages from a tap fed with drinking water of Paris and describe bacteriophage diversity across a one-month sampling period by Illumina sequencing.



**Fig. 5** Contigs belonging to *Caulobacteriales*, *Clostridiales*, *Obscuribacteriales*, *Pseudomonadales*, *Rhizobiales*, *Rhodobacteriales* and *Sphingomonadales* can be associated with *Caudoviricetes* class





**Fig. 6** Majority of alignments between bacterial and viral contigs are localized inside viral contigs. Configuration of 915 alignments between bacterial and viral contigs

#### Ultrafiltration is more suitable for simultaneous recovery of bacteria and viruses

Comparison of global diversity did not show an important difference between AE and UF methods. As a matter of fact, both filtration methods are in theory, according to manufacturer's specifications, able to adsorb any microorganisms [39]. The comparison of Shannon indexes suggests that, at equal volumes of drinking water filtered, UF is sufficient to get approximately the same Shannon index (Additional data 2). However, AE filtration allows to filter up to 5000 L of drinking water, which could be interesting for rare event discovery compared to UF filtration allowing drinking water filtration up to 300 L. In addition, no important viral diversity difference (considering Shannon indexes on putative viral contigs) was seen between both methods. In comparing AE and UF approaches, a distinction was observed in the number of putative viral contigs with AE samples exhibiting a higher count, though without an increase of the viral fraction (Additional data 1). Despite this discrepancy, the UF method, based on filtered volume, proved significantly more effective in capturing viral entities than AE approach (Additional data 3). According to viral contigs and their abundances, samples clustering occurred by method rather than by sampling date (Fig. 4), suggesting that distinct viruses or different abundance of same viruses are recovered between both approaches. Moreover, when considering each method independently, 1023 and 511 putative viral contigs were respectively found in all AE and UF samples, with 437 shared putative viral sequences (Additional data 4). This implies that

some detected viruses are method-specific. Moreover, the higher number of viral contigs observed in AE samples, despite their comparable abundance to UF samples, supports the potential advantage of AE method to detect rare events.

Furthermore, host prediction of viral contigs between concentration methods exhibited one difference: much more *Clostridiales* hosts were predicted for UF samples than for AE samples which is consistent with bacterial taxonomy for contigs with a relative abundance greater than 1%. Bacteria belonging to *Clostridiales* order were much more detected in UF samples. This strongly suggests that AE method does not allow to efficiently recover bacteria from *Clostridiales* order. Methods used for virus concentration are different and come with their own biases. Positively charged NanoCeram® filters have been shown to be efficient for recovering various viruses from different types of water such as seawater [27], freshwater [23, 28, 40, 41], drinking water [28] and wastewater. However, little is known about their capacity to recover bacteria, and these studies mostly focused on human viruses. Elution of particles from these filters is based on electrical surface charge and pH. Thus, it is likely that our method for particle desorption is not optimized for recovering bacteria from *Clostridiales* order. Furthermore, regarding the Shannon index for AE samples, it is likely that an important fraction of bacteria stayed within the nano fibers and were not eluted. It is also possible that some viruses could stay trapped. Indeed, recovery efficiency can be influenced by water matrix, elution buffer and by the type of microorganisms [24]. Ultrafiltration

efficiency, on the contrary, has been studied for a larger range of microorganisms including viruses and bacteria, from different types of water. The critical step, as AE method, is elution and the ability to recover microorganisms from the membrane. The buffer elution used for UF elution varies between studies, some uses Tween 80, beef extract, glycine, polyphosphate or others [42]. It appears that recovery efficiency depends on the buffer used but also on the microorganism that is looked for. This suggests, if all microorganisms are looked for, that UF elution buffer, should be as non-selective as possible. Additionally, organic flocculation on a UF eluate has not been documented. Concentration of UF eluates employs, most of the time, ultrafiltration based on centrifugation [24–26, 41, 43] with or without polyethylene glycol precipitation [29]. To our knowledge, no study using the same protocol of UF concentration was found in scientific literature. It is the first time that organic flocculation with ultracentrifugation is used to perform secondary and tertiary concentrations. We chose organic flocculation rather than ultrafiltration centrifugation because of UF eluate volume (400 mL) and the presence of particles leading to a difficulty for eluate concentration (clogging of ultrafilter). It is worth noting that organic flocculation shows different efficiencies between viruses due to physico-chemical parameters [25], therefore diversity and abundance of microorganisms and bacteriophages could be biased.

It is difficult to tell which method is better to recover bacteriophages or bacteria, but UF method seems to be more adequate unless rare events are looked for. Assessing recovery rates and differences between these two concentration methods based on metagenomics data present considerable challenges. Consequently, when focusing on specific microorganisms, a more appropriate approach would involve quantitative evaluation, such as quantitative PCR. However, in the context of studying a complex environment harboring extensive microbial diversity, estimating and comparing recovery rates between the two methods appears exceedingly difficult, if not impossible. For this reason, we opted to compare both methods based on putative viral contig count, viral fraction abundance, and the detection of different microorganisms. In conclusion, it is difficult to compare our results with existing published methods and concentration, elution and final concentration are three factors to consider. Nonetheless, UF elution is less selective than AE method and appears to allow recovery of a wider range of microorganisms. Choice of the method will depend on what we are looking for. For this study, bacteria and phages have been studied together, for this purpose, UF method appears to be a better choice.

### Host prediction of viral sequences is consistent with bacterial orders detected in samples

Methods used for this study allowed to detect bacteriophages and bacteria from large volumes of tap water fed by drinking water of Paris. Bacterial taxonomy for contigs with a relative abundance greater than 1% shows that *Alphaproteobacteria*, *Cyanobacteria*, *Bacteroidota*, *Baccillota* were the bacterial classes that were the most detected in our samples with *Sphingomonadales*, *Sphingobacteriales*, *Rhizobiales*, *Obscuribacterales*, *Burkholderiales*, *Caulobacterales*, *Clostridiales* and *Pseudomonadales* orders being the most present. These bacterial classes were also found by Perrin *et al.*, over a one-year period in drinking water of Paris in reservoirs and from taps, with a domination by *Pseudomonadota* and *Hyphomicrobium* and *Phreatobacter* genus [8]. In other studies, *Pseudomonadota* and *Cyanobacteria* were also largely found in DWDS [7, 8, 10, 11, 44]. However, repartition between *Alphaproteobacteria*, *Betaproteobacteria*, *Deltaproteobacteria* and *Gammaproteobacteria* varied between studies and between sample localization, in the case of different water sample localizations for a same study [10]. Though, these studies did not explore phage diversity and tried to explain bacterial community variations through temporality, localization, water source and abiotic factors such as pH, temperature, chlorine concentration or flushing, but not bacteriophages which are important drivers of bacterial communities [16]. As said, bacterial community structure and diversity are different according to multiple parameters and it appears difficult to identify a core microbiome [45]. This suggests that each DWDS has its own bacteriophage structure and diversity as well. Understand how bacteriophages can impact bacterial evolution and community variation in drinking water will contribute to its quality monitoring.

Here, we demonstrate the possibility to detect, taxonomically characterize and predict hosts of bacteriophages in drinking water. Bacteriophages are considered as the most present microbiological entities on earth, but their study is quite difficult due to the lack of universally conserved gene, the mosaic nature of their genomes [46] and the over-representation of *Caudoviricetes*, mostly siphoviruses in databases. Furthermore, highly divergent viruses may not have any similarity with known viruses [30] and the frequent discovery of novel viruses leads to recurrent modifications of taxonomic classification of bacteriophages [47]. Additionally, bioinformatics tools are not always up to date and may lead to misclassification. Despite the difficulty of taxonomically assigning bacteriophages, it is possible to predict their host even with no taxonomic assignment. Host prediction of our viral contigs, based on sequence similarity between phages and bacteria, is consistent with the taxonomic assignment of bacteria: most predicted hosts are found in

our metagenomics dataset. The host prediction tool algorithm is an important criterion to be considered. Besides, IPHoP tool [38], like other viral host prediction tools, uses the ability of bacteriophages to incorporate bacterial DNA [48]. Based on this premise, bacteriophages lacking bacterial DNA cannot be associated with any host. Some tools do not rely on alignments between phages and bacteria, but they are considered less accurate than alignment-based approaches [48].

However, our study is based on a one-month sampling which is insufficient to get a global view of bacterial and phage communities and the potential negative or positive correlation between them. Bacterial orders detected are also predicted as hosts for contigs considered as viral even though a large part is not associated with any host. This strongly suggests that when a bacterium is present, a corresponding phage is also present.

#### **Genetic homology between bacteria and bacteriophages allows potential bioinformatics pairs identification and suggests genetic exchanges**

Interest in bacteriophage diversity and their functional role in ecosystem is quite new and tedious, but a growing body of studies investigate these microbiological entities. Thus, viral metagenomics (viromics) is an emerging discipline with methods and standards in constant evolution [49]. Recent advances in metagenomics, computational approaches and bioinformatics tools will allow to unravel bacteriophage diversity and their role in ecosystems but there is a need of standardization and guidelines.

Bacteriophages are important drivers of bacterial communities, and indirectly, are involved in many ecosystems in food chains. They influence biogeochemical cycles, shape microbial communities and can drive bacterial evolution and metabolism [16]. To assess IPHoP tool (host prediction tool) robustness and possible genetic exchange between bacteria and phages, we aligned non-viral contigs to contigs predicted as viral and present in all samples ("core phageome"). These results demonstrate the robustness of IPHoP tool to predict bacterial hosts, the possibility of bioinformatics pairs identification, despite the difficulty to isolate bacterium-phage couples in laboratory and the eventuality of genetic exchanges. Indeed, the concentration of microorganisms in drinking water involves numerous steps, some of them are likely to have a negative impact on viral infectivity and the permissiveness of bacteria. Alignment configurations varied between small to bigger alignments with different ranges of sequence homology. Small alignments (less than 100 bp) may represent CRISPR arrays and exact matches between bacterial and viral contigs can suggest integration sites, CRISPR spacers, regions of genetic homology or integrated prophages [50]. For bigger alignments, genetic exchanges between phages and bacteria

may be a hypothesis. Indeed, phages, lytic or not, are able to acquire and keep bacterial genes. If these genes are beneficial for the virus, they can be maintained by natural selection in the bacteriophage genome. An example of this phenomenon is auxiliary metabolic genes in cyanophages involved in antioxidation, carbon metabolism, cell protection, cycling of nutrients or photosynthesis [14]. Therefore, to investigate the impact of phage community on bacterial populations in drinking water, it is insufficient to only study their diversity and their predicted hosts. Indeed, seeking AMG or integration sites of phages will give a better insight into ecological impacts of bacteriophages for a given environment. Furthermore, genetic recombinations between phages can occur during co-infection via recombinases, genes encoding proteins involved in non-homologous end-joining and transposases [51]. Thus, genes acquired by a given phage might come from another virus rather than a bacterium.

Here, we demonstrated the presence of genetic homology between phages and bacteria in a metagenomic dataset, suggesting genetic exchanges between these populations. These genetic homologies can be explained by multiple phenomena, and a deeper investigation is needed. First, search and annotation of protein coding sequences must be conducted and compared to functional annotation of bacterial contigs to investigate potential genetic exchanges between phages and bacteria. Secondly, nucleotide sequences and functional annotation of viral contigs must be compared to detect potential genetic exchanges between phages. These two points might be helpful to get a better understanding of interactions between phages and bacteria via genetic exchanges study. However, this type of investigation requires high quality metagenome-assembled genomes with low fragmentation. Thus, it might be difficult to perform a functional study of low biomass environment such as drinking water and will require important sequencing depth.

#### **Conclusion**

This present study showed the efficacy of ultrafiltration followed by organic flocculation and ultracentrifugation to recover simultaneously bacteria and viruses from drinking water. This method could limit biases of studying separately bacteria and bacteriophages of the same sample. Moreover, metagenomics sequencing, compared to metabarcoding, gives a better insight of interactions between these microbiological entities allowing better description of a given ecological niche. Here, we showed that most predicted hosts of viral contigs were also detected in drinking water samples, suggesting a tight interaction between bacteriophages and their hosts in DWDS. Moreover, bacterial contig alignments to putative viral sequences showed that important sequence homologies can be found, allowing the detection of

putative bioinformatics couples. This could be an alternative when isolation of bacterium-bacteriophage pairs in laboratory conditions is not successful. Furthermore, long alignments suggest eventual genetic exchanges between phages and bacteria. However, more investigations are needed such as prediction and annotation of open reading frames. Altogether, our proposed approach will enable full scale monitoring of viral and prokaryotic communities in DWDS, contributing to a better understanding of their intricate interplay.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40793-025-00818-y>.

Additional file 1 (DOCX 91 kb)

## Acknowledgements

Not applicable.

## Author contributions

Project conceptualization—L.M, S.W, Y.H, V.D, Project management—L.M, S.W, Y.H, V.D, Data acquisition and curation—M.D, Computational analysis – M.D, V.D, B.M, Original draft writing—M.D, Revision and editions—L.M, S.W, Y.H, V.D, B.M

## Funding

This work was funded by Eau de Paris.

## Availability of data and materials

Sequence data that support the findings of this study have been deposited in the Sequence Read Archive database with the primary accession code PRJNA1224805.

## Declarations

## Competing interests

The authors declare that they have no competing interests.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

All authors have consented for publication.

## Author details

<sup>1</sup>Laboratoire Ecologie Et Biologie Des Interactions, UMR 7267, Université de Poitiers, CNRS, Poitiers, France

<sup>2</sup>Direction de La Recherche, du Développement Et de La Qualité de L'Eau, Eau de Paris, 33 Avenue Jean Jaurès, 94200 Ivry-Sur-Seine, France

Received: 19 September 2025 / Accepted: 4 November 2025

Published online: 11 December 2025

## Bibliography

- Prest EI, Hammes F, van Loosdrecht MCM, Vrouwenvelder JS. Biological stability of drinking water: Controlling factors, methods, and challenges. *Frontiers Media S A*. 2016. <https://doi.org/10.3389/fmicb.2016.00045>.
- Wang HB, et al. Risks, characteristics, and control strategies of disinfection-residual-bacteria (DRB) from the perspective of microbial community structure. *Water Res*. 2021. <https://doi.org/10.1016/j.watres.2021.117606>.
- Gillespie S, et al. Assessing microbiological water quality in drinking water distribution systems with disinfectant residual using flow cytometry. *Water Res*. 2014;65:224–34. <https://doi.org/10.1016/j.watres.2014.07.029>.
- Abkar L, Moghaddam HS, Fowler SJ. Microbial ecology of drinking water from source to tap. *Sci Total Environ*. 2024. <https://doi.org/10.1016/j.scitotenv.2023.168077>.
- Gabrielli M, et al. Identifying Eukaryotes and Factors Influencing Their Biogeography in Drinking Water Metagenomes. *Environ Sci Technol*. 2023;57(9):3645–60. <https://doi.org/10.1021/acs.est.2c09010>.
- Bruno A, Sandionigi A, Bernasconi M, Panio A, Labra M, Casiraghi M. Changes in the drinking water microbiome: Effects of water treatments along the flow of two drinking water treatment plants in a urbanized area, milan (Italy). *Front Microbiol*. 2018;9:1–12. <https://doi.org/10.3389/fmicb.2018.02557>.
- Del Olmo G, et al. The microbial ecology of a Mediterranean chlorinated drinking water distribution systems in the city of Valencia (Spain). *Sci Total Environ*. 2021. <https://doi.org/10.1016/j.scitotenv.2020.142016>.
- Perrin Y, Bouchon D, Delafont V, Moulin L, Héchar Y. Microbiome of drinking water: a full-scale spatio-temporal study to monitor water quality in the Paris distribution system. *Water Res*. 2019;149:375–85. <https://doi.org/10.1016/j.watres.2018.11.013>.
- Delafont V, Brouke A, Bouchon D, Moulin L, Héchar Y. Microbiome of free-living amoebae isolated from drinking water. *Water Res*. 2013;47(19):6958–65. <https://doi.org/10.1016/j.watres.2013.07.047>.
- Ei-Chakhtoura J, Saikaly PE, Van Loosdrecht MCM, Vrouwenvelder JS. Impact of distribution and network flushing on the drinking water microbiome. *Front Microbiol*. 2018. <https://doi.org/10.3389/fmicb.2018.02205>.
- Proctor CR, Hammes F. Drinking water microbiology-from measurement to management. *Curr Opin Biotechnol*. 2015. <https://doi.org/10.1016/j.copbio.2014.12.014>.
- Hegarty B, Dai Z, Raskin L, Pinto A, Wigginton K, Duhaime M. A snapshot of the global drinking water virome: diversity and metabolic potential vary with residual disinfectant use. *Water Res*. 2022. <https://doi.org/10.1016/j.watres.2022.118484>.
- Huang D, et al. The association of prokaryotic antiviral systems and symbiotic phage communities in drinking water microbiomes. *ISME Commun*. 2023. <https://doi.org/10.1038/s43705-023-00249-1>.
- Breitbart M, Bonnain C, Malki K, Sawaya NA, Nature Publishing Group. Phage puppet masters of the marine microbial realm. *Nat Microbiol*. 2018. <https://doi.org/10.1038/s41564-018-0166-y>.
- Mirzaei MK, Maurice CF, Nature Publishing Group. Ménéage à trois in the human gut: interactions between host, bacteria and phages. *Nat Rev Microbiol*. 2017. <https://doi.org/10.1038/nrmicro.2017.30>.
- Chevallereau A, Pons BJ, van Houte S, Westra ER. Interactions between bacterial and phage communities in natural environments. *Nat Rev Microbiol*. 2022. <https://doi.org/10.1038/s41579-021-00602-y>.
- Roux S, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*. 2016;537(7622):689–93. <https://doi.org/10.1038/nature19366>.
- Pal C, Macià MD, Oliver A, Schachar I, Buckling A. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature*. 2007;450(7172):1079–81. <https://doi.org/10.1038/nature06350>.
- Betts A, Gray C, Zelek M, MacLean RC, King KC. High parasite diversity accelerates host adaptation and diversification. *Science*. 1979;360(6391):907–11. <https://doi.org/10.1126/science.aam9974>.
- Wielgoss S, Bergmiller T, Bischofberger AM, Hall AR. Adaptation to parasites and costs of parasite resistance in mutator and nonmutator bacteria. *Mol Biol Evol*. 2016;33(3):770–82. <https://doi.org/10.1093/molbev/msv270>.
- Santiago-Rodríguez TM, Hollister EB, Frontiers Media S.A. Unraveling the viral dark matter through viral metagenomics. *Front Immunol*. 2022. <https://doi.org/10.3389/fimmu.2022.1005107>.
- Chaqroun A, et al. Assessing infectivity of emerging enveloped viruses in wastewater and sewage sludge: relevance and procedures. *Sci Total Environ*. 2024. <https://doi.org/10.1016/j.scitotenv.2024.173648>.
- Hamza IA, Jurzik L, Stang A, Sure K, Überla K, Wilhelm M. Detection of human viruses in rivers of a densely-populated area in Germany using a virus adsorption elution method optimized for PCR analyses. *Water Res*. 2009;43(10):2657–68. <https://doi.org/10.1016/j.watres.2009.03.020>.
- Taligrot H, Wurtzer S, Monnot M, Moulin L, Moulin P. Implementation of a sensitive method to assess high virus retention performance of low-pressure reverse osmosis process. *Food Environ Virol*. 2024;16(1):97–108. <https://doi.org/10.1007/s12560-023-09570-3>.



25. Ikner LA, Gerba CP, Bright KR. Concentration and recovery of viruses from water: a comprehensive review. *Food Environ Virol.* 2012. <https://doi.org/10.1007/s12560-012-9080-2>.
26. Polaczyk AL, et al. Ultrafiltration-based techniques for rapid and simultaneous concentration of multiple microbe classes from 100-L tap water samples. *J Microbiol Methods.* 2008;73(2):92–9. <https://doi.org/10.1016/j.mimet.2008.02.014>.
27. Gibbons CD, Rodríguez RA, Tallon L, Sobsey MD. Evaluation of positively charged alumina nanofibre cartridge filters for the primary concentration of noroviruses, adenoviruses and male-specific coliphages from seawater. *J Appl Microbiol.* 2010;109(2):635–41. <https://doi.org/10.1111/j.1365-2672.2010.04691.x>.
28. Karim MR, Rhodes ER, Brinkman N, Wymer L, Fout GS. New electropositive filter for concentrating enteroviruses and noroviruses from large volumes of water. *Appl Environ Microbiol.* 2009;75(8):2393–9. <https://doi.org/10.1128/AEM.00922-08>.
29. Kahler AM, Johnson TB, Hahn D, Narayanan J, Derado G, Hill VR. Evaluation of an ultrafiltration-based procedure for simultaneous recovery of diverse microbes in source waters. *Water.* 2015;7(3):1202–16. <https://doi.org/10.3390/w7031202>.
30. Krishnamurthy SR, Wang D. Origins and challenges of viral dark matter. *Viruses.* 2017. <https://doi.org/10.1016/j.virusres.2017.02.002>.
31. Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta.* 2023. <https://doi.org/10.1002/imt2.107>.
32. Li D, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods.* 2016. <https://doi.org/10.1016/j.jymeth.2016.02.020>.
33. Riley R, et al. Terabase-scale coassembly of a tropical soil microbiome. *Microbiol Spectr.* 2023. <https://doi.org/10.1128/spectrum.00200-23>.
34. Guo J, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome.* 2021. <https://doi.org/10.1186/s40168-020-00990-y>.
35. Camargo AP, et al. Identification of mobile genetic elements with geNomad. *Nat Biotechnol.* 2023. <https://doi.org/10.1038/s41587-023-01953-y>.
36. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome.* 2020. <https://doi.org/10.1186/s40168-020-00867-0>.
37. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094–100. <https://doi.org/10.1093/bioinformatics/bty191>.
38. Roux S, et al. iPHoP: an integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol.* 2023. <https://doi.org/10.1371/journal.pbio.3002083>.
39. Lyndon B, “Bio-Medical Removing Pathogens Using Nano-Ceramic-Fiber Filters Filters remove greater than 99.9999 percent of viruses and bacteria from wastewater,” 2005.
40. Lee H, et al. Evaluation of electropositive filtration for recovering norovirus in water. *J Water Health.* 2011;9(1):27–36. <https://doi.org/10.2166/wh.2010.190>.
41. Hill VR, Polaczyk AL, Kahler AM, Cromeans TL, Hahn D, Amburgey JE. Comparison of hollow-fiber ultrafiltration to the USEPA VIRADEL technique and USEPA method 1623. *J Environ Qual.* 2009;38(2):822–5. <https://doi.org/10.2134/jeq2008.0152>.
42. Hill VR, et al. Development of a rapid method for simultaneous recovery of diverse microbes in drinking water by ultrafiltration with sodium polyphosphate and surfactants. *Appl Environ Microbiol.* 2005;71(11):6878–84. <https://doi.org/10.1128/AEM.71.11.6878-6884.2005>.
43. Francy DS, et al. Comparison of filters for concentrating microbial indicators and pathogens in lake water samples. *Appl Environ Microbiol.* 2013;79(4):1342–52. <https://doi.org/10.1128/AEM.03117-12>.
44. Simmonds P, et al. Changes to virus taxonomy and the ICTV Statutes ratified by the International Committee on Taxonomy of Viruses (2024). *Arch Virol.* 2024;169(11):236. <https://doi.org/10.1007/s00705-024-06143-y>.
45. Bruno A, Agostinetto G, Fumagalli S, Ghisleni G, Sandionigi A. It's a long way to the tap: microbiome and DNA-based omics at the core of drinking water quality. *Int J Environ Res Public Health.* 2022. <https://doi.org/10.3390/ijerph19137940>.
46. Dion MB, Oechslin F, Moineau S. Phage diversity, genomics and phylogeny. *Nat Rev Microbiol.* 2020. <https://doi.org/10.1038/s41579-019-0311-5>.
47. Valencia-Toxqui G, Ramsey J. How to introduce a new bacteriophage on the block: a short guide to phage classification. *J Virol.* 2024. <https://doi.org/10.1128/jvi.01821-23>.
48. Coclet C, Roux S. Global overview and major challenges of host prediction methods for uncultivated phages. *Curr Opin Virol.* 2021. <https://doi.org/10.1016/j.coviro.2021.05.003>.
49. Pratama AA, et al. Expanding standards in viromics: in silico evaluation of dsDNA viral genome identification, classification, and auxiliary metabolic gene curation. *PeerJ.* 2021. <https://doi.org/10.7717/peerj.11447>.
50. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol Rev.* 2016;40(2):258–72. <https://doi.org/10.1093/femsre/fuv048>.
51. de Moura Sousa JA, Pfeifer E, Touchon M, Rocha EPC. Causes and consequences of bacteriophage diversification via genetic exchanges across lifestyles and bacterial taxa. *Mol Biol Evol.* 2021;38(6):2497–512. <https://doi.org/10.1093/molbev/msab044>.

## Publisher's Note

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