THE USE OF HIGH-THROUGHPUT SEQUENCING FOR DETECTION OF VIRAL SEQUENCES IN COMPLEX SAMPLES FROM ENVIRONMENT AND INDUSTRY

Katarina Bačnik

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Supervisor: Prof. Maja Ravnikar, PhD, National Institute of Biology, Ljubljana, Slovenia
Co-Supervisor: Denis Kutnjak, PhD, National Institute of Biology, Ljubljana, Slovenia

Evaluation Board:
Prof. Kristina Gruden, PhD, Chair, National Institute of Biology, Ljubljana, Slovenia
Prof. Tatjana Avšič Županc, PhD, Member, Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia
Prof. Philippe Roumagnac, PhD, Member, CIRAD, UMR PHIM, Montpellier, France

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL

Katarina Bačnik

THE USE OF HIGH-THROUGHPUT SEQUENCING FOR DETECTION OF VIRAL SEQUENCES IN COMPLEX SAMPLES FROM ENVIRONMENT AND INDUSTRY

Doctoral Dissertation

UPORABA VISOKOZMOGLJIVEGA SEKVENCIRANJA ZA ZAZNAVANJE VIRUSNIH ZAPOREDIJ V KOMPLEKSNIH VZORCIH IZ OKOLJA IN INDUSTRIJE

Doktorska disertacija

Supervisor: Prof. Maja Ravnikar, PhD

Co-Supervisor: Denis Kutnjak, PhD

Ljubljana, Slovenia, May 2022

Acknowledgments

This thesis is a result of great collaborative work of many colleagues and was supported by the funding of the Slovenian Research Agency (ARRS) through the PhD grant and core financing grants (P4-0407, P4-0165) and SPS BioPharm.Si project (OP20.00363), funded by the Ministry of Education, Science and Sport of the Republic of Slovenia and European Regional Fund.

First, I would like to thank my mentors, Maja Ravnikar and Denis Kutnjak, who guided my work throughout the past years and supported me with countless answers, great advice and their enthusiasm. Additionally, big thanks goes to Ion Gutierrez-Aguirre for teaching me lots of skills in and outside the lab and for many helpful discussions. I would also like to thank Anja Pecman for her friendly support with my first experiments and analysis and to Magda Tušek Žnidarič for her wisdom about electron microscopy and wider. Part of my thesis was done in collaboration with researchers from the University of Zagreb, therefore I would like to thank Sandra Hudina, Ana Bielen and Silvija Černi for their engaging work and friendly company. I also had the opportunity to collaborate with colleagues from Lek, and I would especially like to thank Matjaž Vogelsang and Nika Tuta for sharing with me their knowledge about cell cultures. Thanks also to Miguel Aranda from CEBAS-CSIC, Murcia in Spain for hosting me in his lab for a short-term scientific mission. I would also like to thank the members of the dissertation committee, Kristina Gruden, Tatjana Avšič Županc and Philippe Roumagnac.

Thanks to everyone else from the FITO group for such a friendly and stimulating working environment. Thinking back, I realize I was learning and working together with many of you. Thank you! I would especially like to thank Špela Prijatelj Novak and Vesna Dukić for their help with the test plants and Arijana Filipić, Živa Lengar, Tjaša Stare and Anna Coll for great office company and discussions over coffee.

Last but not least, thanks to my family, my friends and especially Vid for always supporting and encouraging me!

Abstract

Metagenomics based on high throughput sequencing (HTS) has opened a new era of discovery and genomic characterization of viruses associated with a given host or environment. The search for novel viral sequences provides an excellent opportunity to improve early detection of pathogens, and to predict viral hosts and environmental reservoirs before the occurrence of significant outbreaks. HTS-based studies of viruses are still challenging, especially in complex matrices, due to the high diversity of viruses, large fraction of unknown and yet to be discovered viruses and different properties of various sample matrices containing viruses. We used non-targeted metagenomics for virus discovery and diversity studies in complex sample matrices, addressing ecological, aetiological and biopharmaceutical perspectives of virome research.

We evaluated the composition of wastewater virome by applying an optimized virus concentration method followed by HTS and infectivity assays. We focused on plant viruses and detected representatives of 47 plant virus species, including emerging crop threats. We also demonstrated infectivity for pathogenic and economically relevant plant viruses from the genus *Tobamovirus* (family *Virgaviridae*), with pepper mild mottle virus (PMMoV) and tobacco mild green mosaic virus (TMGMV) remaining infective even after conventional wastewater treatment. These results are exposing the risks associated with the spread of viruses from the waste into the wider ecosystem and with the uncontrolled use of reclaimed water for irrigation. At the same time, the results are showing that wastewater can be used for early detection and monitoring of plant viruses.

We characterized the virome of signal crayfish (*Pacifastacus leniusculus*), an important freshwater invasive invertebrate species, and potential differences in viral composition and abundance along its invasion range. Study resulted in discovery of novel and divergent RNA viruses, including signal crayfish-associated reo-like, hepe-like, toti-like, and picorna-like viruses, their phylogenetic relationships and potential association with observed pathologies. Additionally the results enabled a better understanding of the potential risk of virus transmissions because of this invader's dispersal.

In biopharmaceutical industry, viral contamination is one of the major concerns for biological products. We employed the developed metagenomics approaches to search for adventitious viruses in animal cell lines used in biopharmaceutical industry. Study of production and parental Chinese hamster ovary (CHO) cell lines of diverse origin did not indicate the presence of adventitious viral agents, however we detected an expected background of virus-like nucleic acids in the samples, which originate from remains of expression vectors, endogenized viral elements and residuals of bacteriophages. This study serves as a baseline for further investigations of CHO cell lines for adventitious viruses using HTS.

Povzetek

Metagenomika, ki temelji na visokozmogljivem sekvenciranju (HTS), je omogočila novo obdobje odkrivanja in genomske karakterizacije virusov, povezanih z določenim gostiteljem ali okoljem. Iskanje novih virusnih sekvenc ponuja odlično priložnost za izboljšanje zgodnjega odkrivanja patogenov, za napovedovanje virusnih gostiteljev in okoljskih rezervoarjev pred pojavom pomembnih izbruhov bolezni. Študije virusov, ki temeljijo na HTS, so še vedno zahtevne, zlasti v kompleksnih matriksih, zaradi velike raznolikosti virusov, velikega deleža neznanih in še neodkritih virusov ter različnih lastnosti matriksov, ki vsebujejo viruse. Za odkrivanje virusov in študije raznolikosti v kompleksnih matriksih vzorcev smo uporabili netarčni metagenomski pristop, pri čemer smo obravnavali ekološke, etiološke in biofarmacevtske vidike raziskav viromov.

Ocenili smo sestavo viroma odpadne vode z uporabo optimizirane metode za koncentriranje virusov, ki sta ji sledila HTS in testi infektivnosti. Osredotočili smo se na virom rastlinskih virusov in zaznali predstavnike 47 vrst rastlinskih virusov, vključno z porajajočimi se rastlinskimi patogeni. Dokazali smo tudi infektivnost za patogene in ekonomsko pomembne rastlinske viruse iz rodu *Tobamovirus* (družina *Virgaviridae*), ki ostanejo infektivni tudi po konvencionalnem čiščenju odpadne vode. Ti rezultati izpostavljajo tveganja, povezana s širjenjem virusov z odpadno vodo v širši ekosistem in nenadzorovano uporabo reciklirane vode za namakanje. Hkrati rezultati kažejo, da je odpadno vodo mogoče uporabiti za zgodnje odkrivanje in spremljanje rastlinskih virusov.

Karakterizirali smo virom signalnega raka (*Pacifastacus leniusculus*), pomembne sladkovodne nevretenčarske invazivne vrste, vključno s potencialnimi razlikami v virusni sestavi in številčnosti vzdolž območja invazije. Študija je privedla do odkritja novih in divergentnih RNA virusov, vključno z reo, hepe, toti in picorna podobnimi virusi, povezanimi s signalnimi raki, njihovimi filogenetskimi odnosi in potencialno povezavo z opaženimi patologijami. Poleg tega so rezultati omogočili boljše razumevanje možnega tveganja za prenos virusa kot posledice širitve invazivne vrste.

V biofarmacevtski industriji je virusna kontaminacija ena glavnih skrbi za varnost bioloških zdravil. Za iskanje naključno prisotnih virusov v živalskih celičnih linijah, ki se uporabljajo v biofarmacevtski industriji, smo uporabili metagenomski pristop. Analiza vzorcev produkcijskih in starševskih CHO celičnih linij raznolikega izvora ni pokazala prisotnosti naključno prisotnih virusnih povzročiteljev, vendar smo v vzorcih zaznali pričakovano ozadje virusom podobnih nukleinskih kislin, ki izvirajo iz ostankov ekspresijskih vektorjev, endogeniziranih virusnih elementov in ostankov bakteriofagov. Ta študija služi kot izhodišče za nadaljnje preiskave CHO celičnih linij za iskanje naključno prisotnih virusov z uporabo HTS.

Contents

Abbreviations

HTS	...	high-throughput sequencing
DNA	...	deoxyribonucleic acid
RNA	...	ribonucleic acid
WWTP	...	wastewater treatment plant
PMMoV	...	pepper mild mottle virus
SARS-CoV-2	...	severe acute respiratory syndrome coronavirus 2
ss	...	single stranded
CHO	...	Chinese hamster ovary
PCR	...	polymerase chain reaction
RT-qPCR	...	reverse-transcription real-time polymerase chain reaction
CIM®	...	convective interaction media
ds	...	double stranded
TEM	...	transmission electronic microscopy
TMGMV	...	tobacco mild green mosaic virus
ToMV	...	tomato mosaic virus
CGMMV	...	cucumber green mild mottle virus
ToBRFV	...	tomato brown rugose fruit virus
totRNA	...	total RNA
RdRp	...	RNA-dependent RNA polymerase

Chapter 1

Introduction

1.1 High-Throughput Sequencing for Discovery of Viruses

Viruses are the most abundant and likely the most diverse source of genetic material on Earth, infecting all cellular organisms and even other viruses (Zhang et al., 2018). We know less about viruses than any group of organisms, with likely only a fraction of one percent of all viruses characterized (Geoghegan & Holmes, 2017).

The technology broadly referred to as next generation sequencing or high-throughput sequencing (HTS) uses various chemistries to obtain large amounts of data about the nucleic acid composition of a sample (Huang et al., 2019). It is possible to extract the total nucleic acids (either DNA or RNA, or both) from environmental samples, to prepare HTS libraries and finally to sequence these to identify all of the nucleic acids in the samples. Such an approach is also termed shotgun metagenomics, and it provides a relatively unbiased picture of the sequences of nucleic acids of the biological entities present in a sample. Due to high diversity of viruses and a lack of conserved gene among all viral species, discovery of novel or unexpected viral species, until recently, represented a complicated task. Consequently, shotgun metagenomics made a dramatic impact in the field of virology. Due to improvements of protocols, availability of HTS technologies and progress in bioinformatic analyses, it is now possible to explore all viruses associated with a given host or diverse environments (Zhang et al., 2018).

New applications became possible and viral metagenomics became a central and fundamental way to interrogate the viral world in many research fields, including untargeted discovery of new viruses with new taxonomic classifications providing a deeper and broader understanding of their diversity, population and genetic structure characterization and virus epidemiology studies (Roux et al., 2021). The search for novel viral sequences provides an excellent opportunity to improve early detection of pathogens, and to predict viral hosts and environmental reservoirs before the occurrence of significant outbreaks (Grubaugh et al., 2019), allowing risk reduction strategies for spillover events and diminishing the severity of emerging outbreaks (Santos et al., 2021).

Nevertheless, non-targeted metagenomics-based discovery and diversity studies of viruses are still challenging, especially in complex matrices, due to the high diversity of viruses, large fraction of unknown and yet to be discovered viruses and different properties of various sample matrices containing viruses.

1.1.1 High-throughput sequencing for detection of viruses in wastewater

Across the many different environmental waters, most extensively in seawater (Coutinho et al., 2017; Suttle, 2016) but also in freshwater systems (De Cárcer et al., 2015), viruses have been shown to be the most abundant and diverse biological entities (Suttle, 2016). Viruses have important ecological roles in these waters, by influencing global biochemical cycling and the composition dynamics of the biological communities (Fuhrman, 1999; Gazitúa et al., 2021). Among the viral nucleic acids detected in aqueous ecosystems, also some belong to known human, animal and plant pathogens (Fong & Lipp, 2005; Mehle et al., 2018), raising concerns regarding the possibility of waterborne infections.

The low concentration of virions in water samples makes them difficult to detect and requires the use of concentration methods during the sample preparation process. Ultrafiltration, polyethylene glycol precipitation, CsCl ultracentrifugation, skimmed milk flocculation, adsorption-elution methods and inclusion of a preamplification step within the library preparation are some of the options that can be used to increase the sensitivity of HTS when studying water viromes (Gutiérrez Aguirre et al., 2018; Hjelmsø et al., 2017; Rastrojo & Alcamí, 2017).

HTS-based methods allow us to reveal a hidden diversity of viral species in aquatic environments (Nataša Mehle et al., 2018; Zinger et al., 2012) and have recently exposed the composition of viromes from different water environments, including wastewater (Adriaenssens et al., 2021; Alhamlan et al., 2013; Aw et al., 2014; Brien et al., 2017; Gulino et al., 2020) and sewage samples (Cantalupo et al., 2011; Fernandez-cassi et al., 2018; Hjelmsø et al., 2017; Martínez-Puchol et al., 2020).

Wastewaters harbour a high diversity of viruses dominated by double-stranded DNA bacteriophages belonging to the order *Caudovirales* (Adriaenssens et al., 2021; Gulino et al., 2020). In the different stages of wastewater treatment plants (WWTPs), bacteriophages have been observed in concentration 10–1000 times higher than in natural aquatic environments, suggesting that wastewater is an important reservoir and source of bacteriophages (Parmar et al., 2018). Although not focusing on plant viruses, several studies primarily targeting phages or animal/human viruses have identified a range of known or novel plant viruses in wastewater and reclaimed water (Rosario, Nilsson, et al., 2009). Additionally, a number of plant viruses have been found in raw sewage (Cantalupo et al., 2011; Martínez-Puchol et al., 2020) and human feces (Shkoporov et al., 2018; Zhang et al., 2006); e.g., pepper mild mottle virus (PMMoV) can survive the transit through the human digestive tract (Zhang et al., 2006) and has been proposed as an indicator for fecal pollution in water sources (Rosario, Symonds, et al., 2009).

Many human viruses can be excreted in feces and urine (Xagoraraki & Brien, 2019), including enteric viruses (Wong et al., 2012) and respiratory viruses such as influenza (Heijnen & Medema, 2011) and coronaviruses (Jones et al., 2020). Therefore, wastewaters provide an insight into the diversity of viruses circulating in a given environment and reflect the infections that have been transmitted in the human population (Cantalupo et al., 2011; Xagoraraki & Brien, 2019). Wastewater-based epidemiology has the potential to predict critical locations and critical moments for viral disease onset (Xagoraraki & Brien, 2019). This approach gained significant attention and development in the COVID-19 epidemic worldwide (Daughton, 2020) representing an effective complementary tool for community-level infectious disease surveillance. In the frame of SARS-CoV-2 wastewater surveillance, HTS allows the identification of both known and novel mutations in the virus sequence and the definition of the relative abundance of individual virus variants in a

wastewater sample (Rios et al., 2021). Wastewater-based epidemiology approaches have previously not been applied for studies of plant pathogenic viruses.

1.1.2 High-throughput sequencing for detection of invertebrate viruses

Current knowledge of virus biodiversity is still biased and fragmentary, reflecting a focus on culturable or disease-causing agents (Junglen & Drosten, 2013; Shi et al., 2016a). Although invertebrates comprise the vast majority of the *Metazoa* (animals), knowledge of invertebrate viruses has been mostly limited to viral pathogens causing high mortalities and arboviruses (arthropod-borne viruses), which are vectored by arthropods and cause disease in human and other vertebrate species (Zhang et al., 2018). This view has started to change with metagenomics revealing remarkable levels of virus diversity in invertebrates (Chang et al., 2021; Li et al., 2015; Porter et al., 2019; Shi et al., 2016a; Wu et al., 2020), and, in this perspective, the detection of vertebrate disease-causing viruses became the exception rather than the rule (Zhang et al., 2018). By sampling a diverse range of invertebrate taxa (Shi et al., 2016a) and mining meta-transcriptomes (Chang et al., 2021), unprecedented levels of RNA virus genetic diversity have been revealed. In addition to revealing host viromes (Sadeghi et al., 2018), virus metagenomics provides insights into different aspects of virus evolution (Dolja & Koonin, 2018), including the role of invertebrates and their ecological interactions with other organisms in the evolution of RNA viruses (Chang et al., 2021).

For understanding the diversity of invertebrate viruses, arthropods, including crustaceans, may be of particular importance because they account for over 80 % of the total animal diversity (Zhang, 2013), are globally abundant and often live in extremely large and dense populations (Dolja & Koonin, 2018). Most of the knowledge on crustacean pathogenic viruses comes from the study of viral diseases in aquaculture (Dragičević, Bielen, et al., 2020). For instance, white spot syndrome virus causes the white spot disease that has mostly been described due to its symptoms in aquaculturally relevant species such as *Cherax quadricarinatus* (Sánchez-Paz, 2010), although it displays a low degree of host specificity and is therefore also an important pathogen in crayfish populations in the wild (Longshaw, 2016). Although viruses are the least studied group of crayfish pathogens (Dragičević, Bielen, et al., 2020), increasing availability of HTS technologies and metagenomics approaches is resulting in studies expanding viral diversity associated with crustacean species (Shi et al., 2016a) and comprehensive descriptions of known and novel viruses (Bateman & Stentiford, 2017). For instance, this approach was recently employed to discover bunya-like brown spot virus, a new negative single stranded (ss) RNA virus belonging to the *Phenuiviridae* family that was associated with a massive disease outbreak in the population of the endangered white-clawed crayfish *Austropotamobius pallipes* (Grandjean et al., 2019).

1.1.3 High-throughput sequencing for detection of adventitious viruses

The World Health Organization (WHO) defines adventitious agents as microorganisms that may have been unintentionally introduced into the manufacturing process of a biological medicinal product (WHO, 2013). Viral contamination is a major concern for biological products as viruses are often more difficult to detect than other microbial

contaminants and human pathogenic viruses can potentially replicate in mammalian cell cultures used in the production of biologicals (Merten, 2002). One of the most important cell lines used in the production of biopharmaceuticals are Chinese hamster ovary (CHO) cells which offer several advantages, such as ease of manipulation and a proven safety profile of products from these cells in humans (Berting et al., 2010). While adventitious agents, potentially introduced via source materials, environmental factors, or cell substrates, are rare, they can have important public health consequences, including the potential to interrupt product supply (McClenahan & Krause, 2019). The utilization of the current combination of in vitro, in vivo, and polymerase chain reaction (PCR) assays for the identification of adventitious viruses has a limited range of detection and requires prior knowledge of potential contaminants (Barone et al., 2020; Brussel et al., 2019). HTS on the other hand is able to detect all types of nucleic acid sequences in a sample and much like in environmental studies shotgun metagenomics has great potential to broaden the ability to detect adventitious agents in biologics (Khan et al., 2018).

Although currently recommended testing has a good record for demonstrating the absence of adventitious agents (Klug et al., 2017), HTS gained significant attention in the field of biologicals since a porcine circovirus 1 contamination has been detected in a licensed pediatric vaccine (Victoria et al., 2010) and a novel rhabdovirus has been discovered in the Sf9 insect cell line (Ma et al., 2014). These cases highlighted that the currently recommended adventitious virus detection assays, even with extensive testing, can fail to detect novel and even some known viruses (Khan et al., 2020). Therefore, HTS has been introduced as a supplemental test and an alternative method for adventitious virus testing (Brussel et al., 2019; Khan et al., 2018, 2020; McClenahan & Krause, 2019).

Similar to other molecular methods, rigorous handling procedures need to be used for HTS to avoid unintended introduction of viral sequences through sample handling and reagents, and appropriate negative controls should be included (Khan et al., 2020). HTS can potentially identify inert viral sequences leading to unnecessary, lengthy, and resource-consuming investigations, however in addition to nucleic acids detection, HTS of newly synthesized viral RNA has been used to demonstrate biological activity of the virus (Brussel et al., 2019; Cheval et al., 2019).

Although much progress has been made in HTS applications for adventitious virus detection (Charlebois et al., 2020) and regulatory authorities are increasingly recognizing the potential of HTS for the detection of a broad range of viruses, some gaps and challenges still need to be addressed. For instance, since HTS is constantly evolving, there is a need for publicly available viral and infected-cell standards for evaluating the sensitivity and specificity of HTS for standardization and validation of the methodology, including the technical and bioinformatics steps (Cleveland et al., 2020).

1.2 Scientific Problems and Aims of the Research

The purpose of applying HTS to samples analyzed in the framework of this doctoral thesis was to describe viral communities, known as viromes, to evaluate the diversity of viruses in a sample/environment, discover known and unknown viruses, potentially unraveling the etiology of the disease and more specifically to use it as a tool in the assessment of adventitious virus absence in biologicals.

1.2.1 Wastewater virome and the release of infective plant viruses from wastewater into the environment

HTS-based studies allow us to reveal the diversity of viral species in aquatic environments (Mehle et al., 2018) and have recently showed high diversity of viruses in wastewater samples, mostly focusing on the presence of bacteriophages and pathogenic human viruses in wastewater. However, as plants make up over 80 % of the biomass on Earth (Baron et al., 2018), plant viruses, shown to be able to infect plants through the water (Mehle et al., 2014), represent an important, but often overlooked group of pathogens, significantly affecting ecosystems and agricultural production worldwide (Nicaise, 2014). Plant viruses were first shown to be present in environmental waters in considerable amounts 30 years ago (Koenig, 1986). Since then, many questions have been raised concerning their survival, origin and spread by water, especially in the light of increased irrigation and use of hydroponic systems in agriculture (Mehle & Ravnikar, 2012). A number of plant viruses have been found in human feces and wastewater samples; e.g.PMMoV, which can survive the transit through the human digestive tract (Zhang et al., 2006) and is released into wastewater after the consumption of pepper products (Symonds et al., 2019). Therefore, plant viruses can potentially circulate between urban environment and croplands: from infected plants, through human and animal consumption and excreta, to wastewater, from where they could be transmitted back to plants through, e.g., irrigation with reclaimed wastewater. Metagenomic approaches can reveal a spectrum of viral nucleic acids in water samples; however, confirming the infectivity of the pathogenic viruses discovered through sequencing analysis is essential to assess the relevance of such findings. Although infectious plant viruses have been isolated from various types of water (e.g., rivers, ponds, irrigation and drainage canals) in different locations worldwide (Jeżewska et al., 2018; Koenig, 1986; Mehle et al., 2018), their survival and direct transmission through water is still largely unknown.

The first aim of this doctoral thesis was to evaluate the composition of wastewater virome with the focus on the presence of plant viruses. To address this aim, we studied the presence of plant viruses in influent and effluent of a WWTP, coupling different approaches of convective interaction media (CIM) monolithic chromatography concentration and HTS-based shotgun metagenomics.

Our second aim was to confirm the infectivity of plant viruses in wastewater influents and effluents to identify the potential of wastewater as a reservoir and transmission source for infective plant viruses. To address this aim, we mechanically inoculated test plants with concentrated influents and effluents, followed by sampling of leaves for further confirmation of virus infections, including the sequencing of small RNAs.

1.2.2 Signal crayfish (*Pacifastacus leniusculus*) virome along its invasion range

Viruses are the least studied group of crayfish pathogens (Dragičević, Bielen, et al., 2020), and relatively few have been formally characterized and classified (Bateman & Stentiford, 2017), despite their large potential impact on aquaculture and populations in the wild. Crayfish are keystone species of freshwater ecosystems and successful invasive species. Invasive crayfish species have often rapidly expanded their range and exerted a high number of documented negative impacts (Lodge et al., 2012; Twardochleb et al., 2013). They have a high potential for introducing and spreading of the emerging diseases and their pathogens have been identified as one of the most prominent factors contributing to

native species population declines (Crowl et al., 2008; Van der Veer & Nentwig, 2015). As the most successful crayfish invader and possible vector for infectious agents, signal crayfish (*Pacifastacus leniusculus)* is among the major drivers of the native crayfish species decline in Europe.

We have investigated the signal crayfish population in a recently invaded Korana river, Croatia, where signal crayfish range expansion is well monitored (Dragičević, Faller, et al., 2020; Hudina et al., 2013, 2017) and occurs in both upstream and downstream directions (Dragičević, Faller, et al., 2020). Host population density, with increased contact rates among individuals at highly dense populations, is an important factor affecting rates of virus transmission (Anderson & May, 1982; Costa et al., 2021). Recently established populations at invasion fronts have 6-11 times lower crayfish abundance in comparison to longer established populations in invasion cores, and also contain a co-occurring native congener (narrow-clawed crayfish, *Pontastacus leptodactylus*) (Dragičević, Faller, et al., 2020). Acute necrotizing hepatopancreatitis was recorded recently along the whole invasion range of the signal crayfish in the Korana River and is currently classified as idiopathic (Bekavac et al., 2022).

The third aim of this doctoral thesis was to characterize the virome of an important invasive invertebrate species to identify putative viral sequences in the signal crayfish hepatopancreas, describe the phylogenetic relationships of the novel signal crayfish-associated viruses and investigate their potential involvement in the observed necrotizing hepatopancreatitis. The signal crayfish individuals were sampled at four locations along its invasion range, including upstream and downstream invasion cores and invasion fronts. Signal crayfish samples were pooled by location and sequenced. We focused our bioinformatics analysis on the presence of viral sequences and subsequent characterization of viral genomes.

Our fourth aim was to elucidate the potential differences in viral composition and abundance of viral sequences along signal crayfish invasion range in the river. To address this aim we compared the patterns of reads abundance at different locations and calculated nucleotide diversities of the detected viral sequences along the invasion range. Potential differences and possible influence of different factors and processes on signal crayfish virome composition along its invasion range were discussed.

1.2.3 Metagenomic characterization of parental and production CHO cell lines for detection of adventitious viruses

Apart from detection of viruses in environmental samples and understudied hosts, HTS-based methods can also be used to detect the presence or confirm the absence of viruses in, e.g., final products or production cell lines, where viral contamination is a big concern. The need for advanced virus detection technologies is increasingly recognized for adventitious virus detection in biological products, where virus testing of raw materials and cells is essential for the safety of the final product (Khan et al., 2018). CHO cells are the most commonly used host cells by the recombinant-biologic industry, with published reports indicating that approximately 70 % of approved biological products are manufactured using CHO cells (Barone et al., 2020; Berting et al., 2010). As a consequence CHO cell lines have the highest occurrence of reported virus contaminations compared with other mammalian cell lines (Barone et al., 2020). HTS can potentially identify inert viral sequences leading to unnecessary analysis, therefore studies providing understanding of background of virus-like nucleic acids in CHO cell lines and their origin represent a necessary baseline for future investigations.

The fifth aim of this doctoral thesis was to characterize viral metagenome of selected production and parental CHO cell lines. To address this aim we tested different sample preparation approaches, where total RNA extracted from cells and nucleic acids isolated from either non-treated, ultra-centrifuged or filtrated growth media were used as inputs for HTS, followed by metagenomic analysis to search for adventitious virus sequences, including the overlapping virus-like sequences between parental and corresponding production cell line.

1.3 Research Hypotheses

1.1 High diversity and abundance of plant viruses is present in WWTP influents and effluents.
1.2 Wastewater contains infective plant viruses, which remain infective after conventional wastewater treatment.

2.1 Novel viral sequences are present in signal crayfish hepatopancreas tissue with potential involvement of specific virus in the observed necrotizing hepatopancreatitis.
2.2 Differences in diversity and abundance of novel signal crayfish-associated viral sequences can be observed along the signal crayfish invasion range.

3.1 Adventitious viruses and/or virus-like nucleic acids are present in production and parental CHO cell lines of diverse origin.

1.4 Publications Included and Candidate's Contributions

The focus of our first paper (Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment) was to study the presence, diversity and infectivity of plant viruses in wastewater influents and effluents. We detected representatives of 11 families of plant pathogenic ssRNA viruses, including emerging crop threats and demonstrated infectivity for pathogenic and economically relevant plant viruses from the genus *Tobamovirus* (family *Virgaviridae*). The PhD candidate performed the majority of lab experiments and conducted bioinformatics analysis of sequencing data. She also wrote the first draft of the manuscript.

In our second paper (Virome analysis of signal crayfish (*Pacifastacus leniusculus*) along its invasion range reveals diverse and divergent RNA viruses), we studied the virome of the signal crayfish, an important freshwater invasive invertebrate species, and potential differences in viral composition and abundance along its invasion range. Study resulted in the discovery of novel RNA viruses, including signal crayfish-associated reo-like, hepe-like, toti-like, and picorna-like viruses. We discussed the potential association with observed pathologies and possible influences of different factors and processes on signal crayfish virome composition along the invasion range, e.g., the differences in signal crayfish population density. The PhD candidate performed bioinformatics analysis to identify putative viral sequences, characterize novel viral genomes and their phylogenetic relationships. She also wrote the first draft of the manuscript.

The third paper (Metagenomic characterization of parental and production CHO cell lines for detection of adventitious viruses) describes the metagenomics investigation of one production and three parental CHO cell lines for the presence of adventitious viral agents.

The study did not indicate the presence of adventitious viral agents; however, it revealed an expected background of virus-like nucleic acids in the samples, which originate from the remains of expression vectors, endogenized viral elements and residuals of bacteriophages. The PhD candidate performed bioinformatics analysis to identify virus-like sequences in production and parental cell lines and wrote the first draft of the manuscript.

Chapter 2

Scientific Publications

2.1 Viromics and Infectivity Analysis Reveal the Release of Infective Plant Viruses from Wastewater into the Environment

Katarina Bačnik, Denis Kutnjak, Anja Pecman, Nataša Mehle, Magda Tušek Žnidarič, Ion Gutiérrez Aguirre, Maja Ravnikar

Water research, 2020, 177 (2020), 115628, https://doi.org/10.1016/j.watres.2020.115628

In this publication, we applied an optimized virus concentration method followed by HTS and infectivity assays to determine the abundance, diversity and biological relevance of plant viruses in wastewater influents and effluents. We detected representatives of 47 plant virus species, including emerging crop threats. We also demonstrated infectivity for pathogenic and economically relevant plant viruses from the genus *Tobamovirus* (family *Virgaviridae*), which remain infective even after conventional wastewater treatment. These results demonstrate the potential of metagenomics to capture the diversity of plant viruses circulating in the environment and expose the potential risk of the uncontrolled use of reclaimed water for irrigation.

The PhD candidate performed the concentration of wastewater samples, isolation of nucleic acids followed by reverse-transcription real-time polymerase chain reaction (RT-qPCR) analyses for targeted detection of viruses and HTS analysis. Further, she performed the mechanical inoculation of test plants followed by systematic sampling of leaves, isolation of nucleic acids, RT-qPCR analyses, transmission electronic microscopy (TEM) and HTS for virome analysis. She performed bioinformatics analyses of the obtained sequencing reads from wastewater samples and plant material samples including comparison of the data obtained using different concentration methods. She also wrote the first draft of the manuscript.

Water Research 177 (2020) 115628

Contents lists available at ScienceDirect

Water Research

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

Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment

Katarina Bačnik [a, b], Denis Kutnjak [a], Anja Pecman [a, b], Nataša Mehle [a], Magda Tušek Žnidarič [a], Ion Gutiérrez Aguirre [a], Maja Ravnikar [a, c, *]

[a] National Institute of Biology, Department of Biotechnology and Systems Biology, Večna pot 111, 1000, Ljubljana, Slovenia
[b] Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000, Ljubljana, Slovenia
[c] University of Nova Gorica, Vipavska cesta, 5000, Nova Gorica, Slovenia

ARTICLE INFO

Article history:
Received 22 November 2019
Received in revised form
13 February 2020
Accepted 14 February 2020
Available online 22 February 2020

Keywords:
Wastewater
Virome
High-throughput sequencing
Plant viruses
Tobamoviruses
Infectivity

ABSTRACT

Viruses represent one of the most important threats to agriculture. Several viral families include highly stable pathogens, which remain infective and can be transported long distances in water. The diversity of plant viruses in wastewater remains understudied; however, their potential impact is increasing with the increased irrigation usage of reclaimed wastewater. To determine the abundance, diversity and biological relevance of plant viruses in wastewater influents and effluents we applied an optimized virus concentration method followed by high-throughput sequencing and infectivity assays. We detected representatives of 47 plant virus species, including emerging crop threats. We also demonstrated infectivity for pathogenic and economically relevant plant viruses from the genus *Tobamovirus* (family *Virgaviridae*), which remain infective even after conventional wastewater treatment. These results demonstrate the potential of metagenomics to capture the diversity of plant viruses circulating in the environment and expose the potential risk of the uncontrolled use of reclaimed water for irrigation.

1. Introduction

Expanding human population, increasing urbanization, predicted global warming and agronomic exploitation of land place enormous pressure on supplies of drinking and irrigation water. Among microbiological water contaminants, viruses are a group of particular concern (Rosario et al., 2009a); although many of them are highly stable pathogens, they remain understudied. Up until now, most of the studies have focused on the presence of bacteriophages and pathogenic human viruses in wastewater. However, as plants make up over 80% of the biomass on Earth (Bar-On et al., 2018), plant viruses, shown to be able to infect plants through the water (Mehle et al., 2014), represent an important, but often overlooked group of pathogens, significantly affecting ecosystems and agricultural production worldwide (Nicaise, 2014). The financial impact of yield losses due to plant viruses in agriculture are globally estimated to cost 30 billion USD annually (Sastry and Zitter, 2014). Studies on water mediated plant virus transmission together with assessment of traditional wastewater treatment efficiency on plant virus inactivation are lacking, even though reclaimed water use for irrigation is being increasingly introduced worldwide (Pedrero et al., 2010; Thebo et al., 2017). Global estimates of the extent to which irrigated croplands are influenced by wastewater, both treated and untreated are 35.9 Mha (~17% of all irrigated croplands), of which the majority are located in countries where less than 75% of wastewater is treated (Thebo et al., 2017).

It was hypothesized that the surface wash out of infected decaying plant residues and the related soil surface layer can bring plant viruses into water (Mehle and Ravnikar, 2012) and they may be washed away by rainwater reaching drainage water and wastewater. Additionally, a number of plant viruses have been found in raw sewage (Cantalupo et al., 2011) and human feces (Shkoporov et al., 2018; Zhang et al., 2006); e.g., pepper mild mottle virus (PMMoV) can survive the transit through the human digestive tract (Zhang et al., 2006) and is released into wastewater after consumption of pepper products (Symonds et al., 2019). PMMoV is globally distributed and present in various water sources in greater abundance than human pathogenic viruses, without substantial seasonal fluctuation (Kitajima et al., 2018), thus, it has been proposed as a water pollution indicator (Kuroda et al., 2015; Rosario et al., 2009b).

* Corresponding author. National Institute of Biology, Department of Biotechnology and Systems Biology, Večna pot 111, 1000, Ljubljana, Slovenia.
E-mail address: maja.ravnikar@nib.si (M. Ravnikar).

https://doi.org/10.1016/j.watres.2020.115628

Abbreviations

PMMoV	pepper mild mottle virus
CIM®	convective interaction media
HTS	high-throughput sequencing
WWTP	wastewater treatment plant
QA	quaternary amine
CGMMV	cucumber green mild mottle virus
TEM	transmission electronic microscopy
sRNA	small RNA
SRA	short reads archive
TMGMV	tobacco mild green mosaic virus
ToMV	tomato mosaic virus
ToBRFV	tomato brown rugose fruit virus
CMV	cucumber mosaic virus
PVY	potato virus Y
PepMV	pepino mosaic virus
MNSV	melon necrotic spot virus

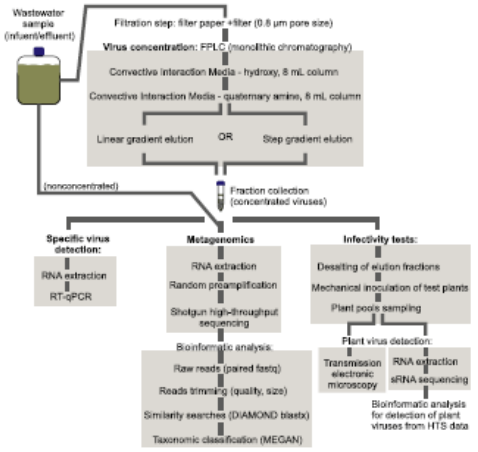


Fig. 1. Schematic representation of methods used for analysis of water virome and infectivity of viruses in water. It represents steps included in virus concentration, specific virus detection using quantitative reverse transcription PCR (RT-qPCR), sample preparation for HTS and consecutive bioinformatics analysis and sample preparation for mechanical inoculation of test plants and subsequent virus detection using transmission electronic microscopy and small RNA sequencing. FPLC – fast performance liquid chromatography.

The low concentration of virions in water samples makes them difficult to detect and requires the use of concentration methods during the sample preparation process. Different concentration approaches and sample preparation modifications can be used to increase the sensitivity of virus detection in wastewater (Haramoto et al., 2018; Hjelmsø et al., 2017). Convective interaction media (CIM®) monoliths have shown to be a fast and efficient tool for concentration of viral particles and free nucleic acids from high volume water samples, pointing to CIM monoliths as a promising tool for sample preparation in water virome studies (Gutiérrez Aguirre et al., 2018). High-throughput sequencing (HTS) based metagenomics studies have recently exposed the composition of viromes from different water environments, e.g., wastewater (Alhamlan et al., 2013; Aw et al., 2014; O'Brien et al., 2017; Tamaki et al., 2012), sewage (Cantalupo et al., 2011; Fernandez-Cassi et al., 2018; Hjelmsø et al., 2017; Ng et al., 2012), reclaimed water (Rosario et al., 2009a) and fresh water (Djikeng et al., 2009; Mohiuddin and Schellhorn, 2015).

Metagenomic approaches reveal an unprecedented spectrum of viral nucleic acids in water samples; however, confirming the infectivity of the viruses discovered through sequencing analysis is essential to assess the relevance of such findings. Although infectious plant viruses have been isolated from various types of water (e.g., rivers, ponds, irrigation and drainage canals) in different locations worldwide (Ježewska et al., 2018; Koenig, 1986; Mehle et al., 2018), their survival and direct transmission through water is still largely unknown. None of the studies to date have approached infectivity testing for plant viruses detected in wastewater, most probably due to the low viral concentrations in combination with the low sensitivity and time intensive test plant bioassays used for the confirmation of plant virus infectivity.

To identify the potential of reclaimed water as a reservoir and transmission source for infective plant viruses we focus this study on the presence of plant viruses in influent and effluent of a wastewater treatment plant (WWTP), coupling CIM monolithic chromatography concentration, HTS-based shotgun metagenomics and biological infectivity tests (Fig. 1). The combination of fast and efficient virus concentration and metagenomics enables us to study the population of plant viruses in wastewater, highlighting the presence of important plant pathogenic viruses, some of which were not previously detected in the region.

2. Materials and methods

2.1. Samples and description of WWTP

Water samples (5 L) of WWTP influents and effluents were sampled in summer 2016 and 2017. The selected WWTP located in central Slovenia (Central WWTP Domžale-Kamnik, Slovenia) is a conventional two-stage activated sludge plant, upgraded with four sequence batch reactors and additional anaerobic digestion of activated sludge. The capacity of the plant, designed for organic matter removal from wastewater, is 149 000 population equivalents, and an average daily inflow of approximately 20 000 m^3. The facility collects municipal wastewater, rainwater and industrial wastewater from five communities including households, farms, and industry. The recipient of treated water from the plant is the Kamniška Bistrica river.

2.2. Concentration of wastewater samples for viral detection

Before concentration, all samples were filtered through filter paper and cellulose acetate membranes with a pore size of 0.8 µm (Sartorius, Goettingen, Germany). The concentration of each sample was done using CIM quaternary amine (QA) 8 mL monolithic column (BIA separations, Ajdovščina, Slovenia) on a fast protein liquid chromatography system AKTA Purifier 100 (GE Healthcare, Chicago, IL, USA). Influent and effluent samples were loaded at 80 mL/min flow rate onto the CIM QA column. After the column loading and a washing step using 50 mM HEPES buffer, pH 7, we applied either (i) step gradient elution concentration approach (Supplementary Fig. 2a), where an abrupt change in concentration of salt in the elution buffer was used or ii) linear gradient elution concentration approach (Supplementary Fig. 2b), where a gradually increasing concentration of salt in the elution buffer was introduced. The protocol used for concentration of viruses was described

previously (Gutiérrez-Aguirre et al., 2011, 2009) and outlined in detail in Supplementary materials and methods 1.

2.3. Detection of viruses in wastewater samples using RT-qPCR

RT-qPCR was used to assess the performance of chromatographic concentration procedures by determining the presence and viral nucleic acids recoveries of two selected plant viruses (PMMoV and cucumber green mild mottle virus - CGMMV) for step gradient elution (Inf16-Step) and linear gradient elution (Inf17-Lin). Extraction of nucleic acids for RT-qPCR analysis was done following the CIM concentration from a 140 µL aliquot using a QIAamp viral RNA mini kit (QIAGEN, Germantown, MD, USA). RT-qPCR was performed using published assays for PMMoV (Haramoto et al., 2013) and CGMMV (Zhao et al., 2015) using AgPath-ID™ One-Step RT-qPCR Kit (Life Technologies, Carlsbad, CA, USA) on a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). The average Cq obtained for each elution fraction was normalized to the external luciferase control and used to estimate the virus recovery (%) in the fractions relative to the original sample via the standard curve approach as described in previous studies (Balasubramanian et al., 2016; Gutiérrez-Aguirre et al., 2009).

2.4. Metagenomic analysis of wastewater samples

For the metagenomic analysis of wastewater samples, nucleic acids were extracted using TRIzol LS (Life tehnologies, CA, USA), following the manufacturer's protocol with the addition of glycogen (100 µg) to 400 µL of water sample in the beginning of the extraction. Isolated nucleic acid concentration was quantified using Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and Qubit 4 flourometer (Thermo Fisher Scientific, MA, USA). Samples with concentration of RNA lower than 1 ng/µL after extraction were randomly preamplified (see Table 1) as described previously (Fernandez-Cassi et al., 2018; Wang et al., 2003). Briefly, RNA templates were reverse transcribed using SuperScript III (Life Technologies, CA, USA) and Primer A (5′-GTTTCCCAGTCACGATANNNN NNNNN′- 3), which contains a specific sequence followed by 9 random nucleotides for random priming. A second cDNA strand was constructed using Sequenase 2.0 (Thermo Fisher Scientific, MA, USA). To obtain sufficient DNA for library preparation, a PCR amplification step using Primer B (5′-GTTTCCCAGTCACGATA′-3) and AmpliTaqGold (Life Technologies, CA, USA) was performed. PCR products were cleaned and concentrated using Agencourt AMPure XP (Beckman Coulter, IN, USA).

Step gradient elution of samples from 2016 (influent and effluent) and 2017 (effluent) resulted in two elution fractions (Step-E1 and Step-E2), that were separately extracted and sequenced. For influent and effluent samples (year 2017), nucleic acids from the linear gradient elution fractions selected in three different parts of the gradient were separately extracted. Before sequencing, nucleic acids from the elution fractions were pooled together to three fraction groups (Lin^a, Lin^b, Lin^c). For the effluent sample Lin^a (F6–F8), Lin^b (F16–F18) and Lin^c (F35–F37) were sequenced. For the influent sample Lin^a (F6–F8), Lin^b (F16–F20) and Lin^c (F35–F36) were sequenced (shown in Supplementary Fig. 1). The negative control of isolation was spiked with luciferase RNA (Promega, WI, USA) and was also sent for sequencing as a control for contamination in the isolation process and sample cross-talk during sequencing.

We sent isolated RNA (or preamplification products) for reverse transcription, library preparation and sequencing to SeqMatic LLC (Fremont, CA, USA). Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA, USA) was used to construct the sequencing libraries, which were shotgun-sequenced using an Illumina MiSeq (Illumina, San Diego, CA, USA); resulting in 1 million to 3.4 million 2 × 250 bp reads per sample. The remains of sequencing adaptors and primer sequences in the case of preamplified samples were trimmed and resulting reads were further filtered by using quality filter (Limit = 0.01; no ambiguous nucleotide allowed) and by size (reads shorter than 25 bp were discarded) in CLC Genomics Workbench 11 (Qiagen Bioinformatics, Redwood City, CA, USA). In order to normalize samples according to the number of reads, random subsampling was done and all the samples were normalized to

Table 1
Metadata of wastewater samples analyzed by HTS with summary of reads' taxonomic classifications; overview of classified reads (%), reads classified as viral (%), viral reads classified as plant viruses (%), plant virus species richness and corresponding SRA acession numbers.

Sample name	Year of sampling	Sample type	Concentration approach	Random preampli-fication	Classified reads (%)	Reads classified as viral (%)	Viral reads classified as plant viruses (%)	Plant virus species richness	**Accession number (SRA*)**
Inf16-None	2016	influent	none	no	53.12	0.13	48.80	10	SRR9317870
Inf16-Step			step	no	44.88	5.61	1.73	17	SRR9317869 SRR9317872
Eff16-None	2016	effluent	none	no	38.26	0.53	1.22	8	SRR9317871
Eff16-Step			step	no	34.54	2.40	2.81	19	SRR9317874 SRR9317873
Inf17-Lin[a]	2017	influent	linear	yes	38.92	6.88	0.57	12	SRR9317876
Inf17-Lin[b]			linear	yes	29.77	7.65	11.47	22	SRR9317875
Inf17-Lin[c]			linear	no	49.00	3.47	3.96	16	SRR9317868
Eff17-None	2017	effluent	none	yes	79.85	0.72	36.60	8	SRR9317867
Eff17-Step			step	no	20.26	1.88	16.96	12	SRR9317880 SRR9317879
Eff17-Lin[a]			linear	yes	14.71	2.63	0.82	3	SRR9317878
Eff17-Lin[b]			linear	yes	23.48	6.30	44.53	11	SRR9317877
Eff17-Lin[c]			linear	yes	51.81	0.84	26.70	7	SRR9317882

*short reads archive (SRA).

926 418 reads, which was the minimum number of reads obtained among the samples. Trimmed, size and quality filtered normalized reads subsets were exported from CLC Genomics Workbench and compared for similarity with a complete NCBI nr database (June 2018), using DIAMOND (Buchfink et al., 2015) with default parameters. The results of the DIAMOND classifications were used as an input for the taxonomic classification of the reads using MEGAN (Metagenome Analyzer, version 6.12.0) (Huson et al., 2016) with the following parameters for the Lowest Common Ancestor algorithm: min score = 50.0, max expected = 0.01, top percent = 10.0, min support percent = 0, min support = 1, and LCA percent = 100.0. The obtained MEGAN outputs were used to present an overview of taxonomic classification of the sequencing reads. To compare viromes of differently concentrated samples, single-stranded (ss) RNA viral reads were exported as a BIOM file. Differential heat trees of pairwise comparisons between samples were constructed by Metacoder R package (Foster et al., 2017) where the relationships between viral taxons presented in the trees are inferred from NCBI taxonomy hierarchy. All plant infecting viruses whose genomes were detected in metagenomes of wastewater samples were listed in Fig. 3 and the corresponding viral species richness (number of observed viral species) was presented in Table 1. The reads classified as plant viruses were manually inspected and some reads were manually reassigned or left unclassified on the species level to ensure the correct classification of plant virus species. For this, we used an approach that combined BlastN (NCBI nt database, September 2018) and BlastX (NCBI nr database, September 2018) determination of sequence identity to the most similar sequences in the respective databases with detailed review of each identified taxon; reads with lower identity than defined by ICTV species demarcation criteria were classified only on the genus level.

2.5. Infectivity tests and detection of infective plant viruses

Nicotiana occidentalis and *Nicotiana benthamiana* plants, that are susceptible to infection by a wide range of viruses including most of the detected tobamoviruses (Wylie et al., 2015), were selected to test the infectivity of plant viruses after the concentration of wastewater samples using CIM monolithic chromatography. Before inoculation of test plants, selected elution fractions were desalted using PD-10 desalting columns (GE Healthcare, IL, USA) following the manufacturer's instructions and using inoculation buffer (20 mM sodium phosphate buffer, pH 7.6 with 2% PVP (M 10000) (Sigma, Tokyo, Japan)) as an equilibration buffer. Inoculation of test plants with influent sample included eight separate linear gradient elution fractions (F16–F23) and a negative control of inoculation, where plants were inoculated only with inoculation buffer. Inoculation of test plants with effluent sample included a step gradient elution fraction (Ef17-Step-E1) and a negative control of inoculation (Supplementary Fig. 1). All fractions that were used for inoculation of test plants originated from the same concentration experiment as the fractions that were sequenced to ensure the consistency and comparability with metagenomics analysis (Supplementary Fig. 1). Sampling, concentration process, desalting and inoculation of test plants were all done in the same day to avoid freezing of the samples and prevent the possible negative influence of concentrated substances on stability of viral particles.

We inoculated each selected elution fraction on eight plants of each tobacco species (*Nicotiana occidentalis* and *Nicotiana benthamiana*.). We mechanically inoculated two to three leaves per plant using carborundum (0.062 mm, VWR International, Radnor, PA, USA) that was put on each leaf and the elution fraction (approximately 50 μL per leave) was gently rubbed over it. Additional plants inoculated only with inoculation buffer served as negative controls. During the 4-week period following the mechanical inoculation, we observed plants for the appearance of symptomatic changes. We sampled plant material of non-inoculated leaves every week after inoculation for detection of viruses using transmission electronic microscopy (TEM) and small (s) RNA sequencing analysis.

We collected pooled samples of leaf tissue (non-inoculated leaves, approximately 1 cm^2 per leave) separately for different test plants inoculated with different CIM fractions, but exhibiting similar disease symptoms. We also sampled non-inoculated leaves of asymptomatic plants (see Supplementary Fig. 1 for details). Plant material samples were further used to isolate total RNA using TRIzol reagent (Thermo Fisher Scientific, MA, USA). As a control we included a sample of *Phaseolus vulgaris* plant material infected with an endornavirus (kindly provided by Mike Rott, Canadian Food Inspection Agency - Dartmouth, Canada) in the sequencing experiment.

sRNA deep sequencing was used as a generic detection method, which enable detection of all the different types of viruses (different RNA and DNA viruses) infecting plants at the same time (Pecman et al., 2017). Total RNA was sent to SeqMatic LLC (Fremont, CA, USA) for sRNA library preparation (TailorMix miRNA Sample Preparation Kit, SeqMatic LLC, USA) and sequencing. The samples were multiplexed in one run of a NextSeq (Illumina, CA, USA) in 1 × 100 bp mode. Reads were further analyzed using a bioinformatics pipeline for virus detection from HTS data as described in a previous study (Pecman et al., 2017). Shortly, trimmed and size-selected reads were first mapped to the NCBI Viral RefSeq database, containing representatives of all viral genomes with completely sequenced genomes. Results of the mapping were manually inspected. In parallel de novo assembly of trimmed and size-selected reads was performed. Assembled contigs were compared for similarity against all viral sequences deposited in the NCBI GenBank nt database using BLASTn and BLASTx. The presence of suspected viral sequences was confirmed and complete consensus genome sequence was generated by mapping the reads to the complete viral genome sequences of the most similar viral isolates from the NCBI GenBank database.

2.6. TEM analysis of viruses

A sample (25 μL) of the chromatographic fractions was applied to formvar-coated (Agar Scientific, Stansted, UK), carbon-stabilized copper grids that were negatively stained using a 1% (w/v) aqueous solution of uranyl acetate (SPI Supplies, West Chester, PA, USA). Selected elution fractions from year 2 (influent linear gradient fraction 17 and 18) were ultracentrifugated using Airfuge (Beckman Coulter Inc., Brea, CA, USA) onto TEM grids before negative staining as described previously (Hammond et al., 1981). All samples were examined with TEM (Philips CM 100, Netherlands) as described in a previous study (Ruščić et al., 2015).

Plant material of randomly selected plants inoculated with wastewater samples exhibiting distinctive symptoms was also examined using TEM. Homogenates of plant material (20 μl) were applied to Formvar-coated, carbon-stabilized copper grids and negatively stained using a 1% (w/v) aqueous solution of uranyl acetate (SPI Supplies, PA, USA), followed by visualization by TEM.

3. Results

To examine the virome composition of wastewater, we determined the metagenomes of the influent and effluent samples of a WWTP with special emphasis on plant viruses. Viruses were first concentrated using CIM® monolithic chromatography using either: i) step gradient elution concentration approach (Supplementary Fig. 2a) or ii) linear gradient elution concentration approach

(Supplementary Fig. 2b). After linear gradient elution we selected, sequenced and separately analyzed three fraction groups based on the abundance of PMMoV and CGMMV, determined by RT-qPCR (Supplementary Fig. 2b, Supplementary Table 1). Selected fraction groups cover the initial (Lin^{a}), middle (Lin^{b}) and final (Lin^{c}) part of the elution gradient.

3.1. Overview of the highly diverse wastewater virome reveals understudied group of plant viruses

To provide a trustworthy representation of viral species (Rodriguez-Brazzarola et al., 2018) we analyzed HTS data obtained by sequencing wastewater samples (Table 1) using reads-based protein similarity searches. Taxonomic classification showed that a substantial fraction of reads in each metagenome (20–85%) had no matches to known organisms (Table 1, Fig. 2a). This is similar to previous studies, in which the majority of the sequences in wastewater samples had no similarity to known genes or proteins (Aw et al., 2014; Bibby and Peccia, 2013). A large fraction of reads (10–50%) in influent and effluent samples belonged to bacteria (Fig. 2a). A smaller fraction of reads was classified as eukaryotic organisms and archaea (Fig. 2a). Eukaryotic reads were abundantly present in non-concentrated samples (5–13%) and were classified mostly as genomes of plants, aquatic arthropods, rodents and enteric parasites (data not shown). In the same non-concentrated influent and effluent samples the percentage of viral reads was less than 1% (0.1–0.7%), however, it increased to 7.7% in CIM concentrated samples (Table 1, Fig. 2a).

By analyzing metagenomes of the 12 samples (Table 1) we detected members of 56 different viral families, including those with double-stranded (ds)DNA, ssDNA, dsRNA and ssRNA genomes. (Fig. 2b). Examination of wastewater samples by TEM also revealed a diversity of virion morphologies with filamentous viral particles and phage-like virions (Supplementary Fig. 3). As previously shown, the viromes of WWTP influents and effluents contain high abundance of dsDNA bacteriophages. CrAssphage sequences, originally identified from metagenomic analysis of human feces (Dutilh et al., 2014), were abundantly detected in influent samples (Supplementary Table 3). Gut-associated Bacteroides phage B124-14, that can be used to distinguish human gut contaminated metagenomes (Ogilvie et al., 2018) and few reads of circular (Rep)-encoding single-stranded DNA (CRESS DNA) viruses, recently reported in wastewater effluents (Rosario et al., 2019) were also detected in a majority of the samples (Supplementary Table 3).

Important human pathogens such as norovirus, sapovirus, enterovirus, rotavirus A, hepatitis E virus and JC polyomavirus A were also detected in wastewater viromes (Supplementary Table 3). Pathogenic viruses of other hosts such as sequences of members of *Picornavirales* order belonging to *Dicistroviridae* family were detected mainly in influent samples. We also detected bee viruses belonging to *Iflaviridae* family and *Picornavirales* order in both influent and effluent samples (Supplementary Table 3). In influent (Inf17-Lin^{b}) and effluent (Ef17-Lin^{b}) linear gradient elution fractions, we detected representatives of more than 20 families of ssRNA viruses.

Plant viral genomes mostly consist of ssRNA (~75% of all plant viruses) (Hull and Bustamante, 1998) and were therefore mostly detected in analyzed wastewater samples as sequences classified as ssRNA viruses, with exception of the dsRNA pepper cryptic virus (Fig. 3). The most commonly found viruses recovered before and after wastewater treatment belonged to the family *Virgaviridae*, with all of the reads corresponding to the *Tobamovirus* genus (Fig. 2c) most abundantly represented by sequences of PMMoV, tobacco mild green mosaic virus (TMGMV), tomato mosaic virus (ToMV) and CGMMV (Fig. 3). We also detected sequences of tomato

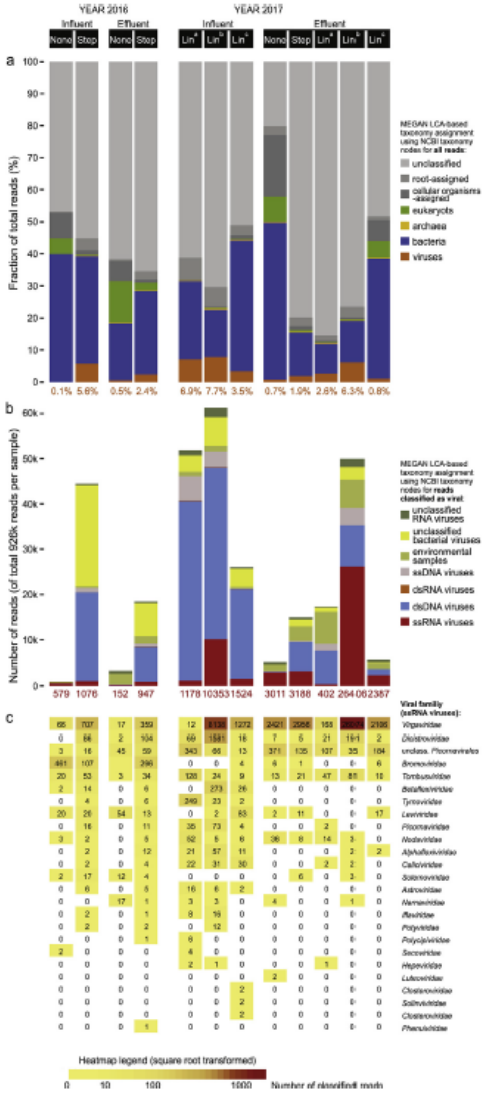


Fig. 2. Taxonomic classification of the sequencing reads from non-concentrated and differently concentrated influent and effluent wastewater samples from two consecutive years (explained in Table 1). (a) Relative amount of the reads classified on the level of domains (the % below the columns represent the fractions of reads classified as viral in corresponding samples). (b) A more detailed overview of the absolute number of viral reads classified on the level of viral groups according to NCBI taxonomy nomenclature (the numbers below the columns represent absolute number of reads classified as ssRNA viruses in corresponding samples). (c) Heatmap for families of positive-sense ssRNA viruses and corresponding number of reads belonging to respective families detected in each sample. The numbers within each cell represent number of reads per normalized sample.

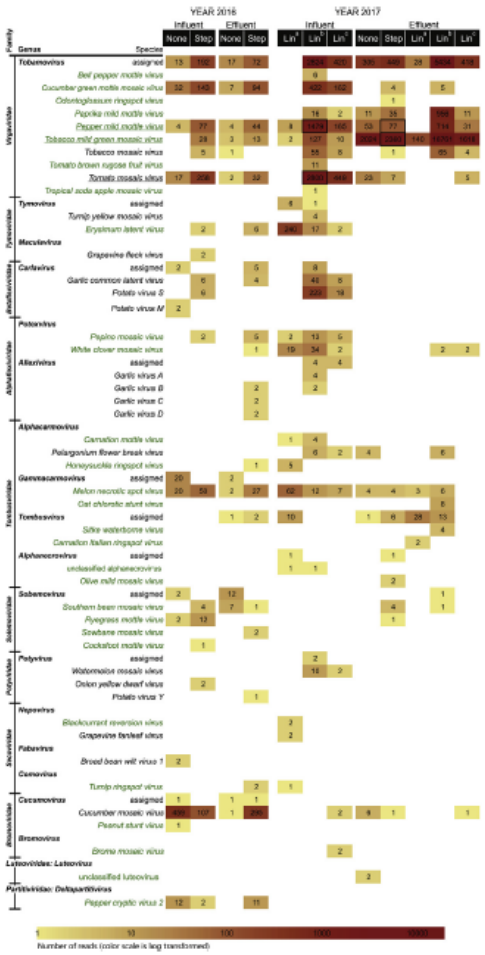


Fig. 3. A heatmap showing the plant viral species and corresponding number of their reads detected in corresponding wastewaters samples. Reads were taxonomically classified using Diamond-MEGAN pipeline and further manually curated. Reads classified on genus level (assigned) represent: (i) sequences of different species from respective genera not classified to the species level with used classification pipeline, (ii) manually curated assignments for sequences with lower species identity than stated in species demarcation criteria (potential novel viral species). Viruses with confirmed infectivity in this study are underlined and viruses that were not reported in Slovenia before are colored green.

brown rugose fruit virus (ToBRFV), which was not found in the country before and is an emerging tobamovirus that has recently brought concern to plant health authorities worldwide. Representatives of less abundant families such as *Potyviridae*, *Betaflexiviridae* and *Tymoviridae* were mostly present in the influent samples and were rarely detected after the wastewater treatment in the effluent samples (Fig. 2c). Sequences belonging to *Bromoviridae* family were detected abundantly in samples from 2016 and only few reads were detected in samples from 2017 (Fig. 2c). Whilst tobamoviruses were numerically dominant plant viruses in the samples (Fig. 2c), the list (Fig. 3) of detected potentially pathogenic plant viruses includes 47 species from 19 genera including the economically important viruses, such as cucumber mosaic virus (CMV), potato virus Y (PVY) and pepino mosaic virus (PepMV). With in-depth analysis of reads assigned to plant viral families, sequences with low similarity to known plant viruses were also detected, that could only be classified at the genus level and probably represent previously undescribed viral species (Fig. 3).

3.2. Optimized viral concentration enables more detailed insight into diversity of plant viruses

While HTS is a powerful tool for uncovering viruses in the environment, sample and molecular processing steps can alter the distribution of viral groups in the concentrated samples. This is why we have compared non-concentrated and differently concentrated wastewater samples. We analyzed fractions obtained by either step gradient elution or linear gradient elution with RT-qPCR for presence and relative concentrations of two tobamoviruses (PMMoV, CGMMV), calculated their recoveries (Supplementary Fig. 2) and compared the virome composition of differently concentrated samples. We were able to efficiently concentrate PMMoV and CGMMV for approximately two orders of magnitude according to RT-qPCR results. (Supplementary Fig. 2, Supplementary Table 1, Supplementary Table 2).

We observed a higher fraction of viral reads in both step/linear gradient concentrated influent and effluent samples compared to non-concentrated ones (Table 1, Fig. 2a). The enrichment for viral sequences was most noticeable in Lin^b linear gradient samples (Table 1, Fig. 2a). The fractions at the peak of absorbance (Lin^b) have very high recoveries according to RT-qPCR for selected tobamoviruses (CGMMV and PMMoV) and contain the highest number of reads belonging to ssRNA viruses (Fig. 2b: Inf17- Lin^b, Ef17- Lin^b), mostly due to the high abundance of viral reads classified as tobamoviruses (Fig. 2c).

Next, we aimed to get a detailed insight into the plant virus abundance and diversity differences between different concentrations of the same sample. Thus, we visualized pairwise comparisons of taxa abundances for ssRNA viral reads of differently concentrated effluent sample (from year 2017) as differential abundance heat trees (Fig. 4). Overall enrichment of *Virgaviridae* family in Ef17-Lin^b effluent sample was pronounced (Fig. 4). In summary, we detected higher number of plant viral species in concentrated influent and effluent samples compared to non-concentrated ones, especially in Lin^b fractions (Table 1). Both concentration approaches (step and linear) allowed us to focus on the diversity of ssRNA viruses, with the linear gradient having the potential to study specific viral groups, namely tobamoviruses, in more detail.

3.3. Plant viruses in wastewater are infective and survive wastewater treatment

To determine if the plant viruses detected by HTS have a potential to be transmitted to plants by wastewater, we performed biological infectivity tests. Inoculation with non-concentrated wastewater in a preliminary experiment resulted in no detectable infection (results not shown). Therefore, in subsequent experiments we mechanically inoculated test plants using fractions concentrated by either linear (influent) or step (effluent) gradient elution and monitored the plants for symptom development. Using

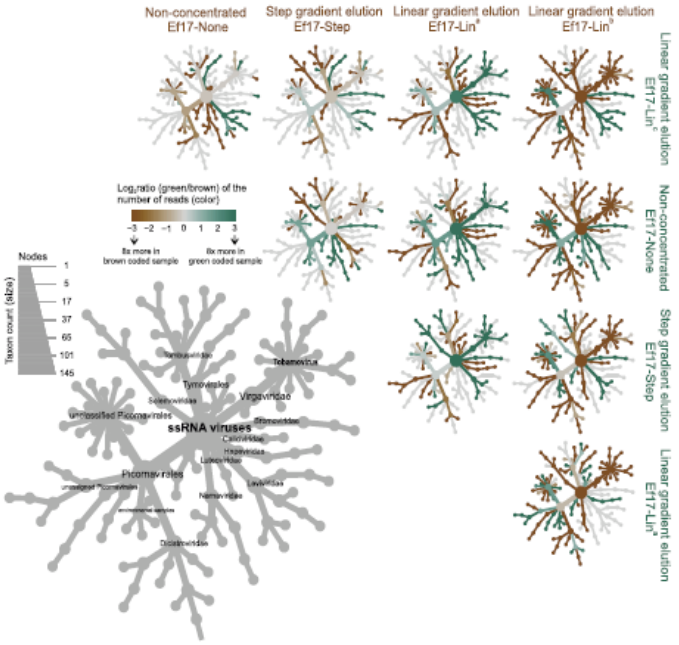


Fig. 4. Pairwise comparisons between ssRNA viromes of differently concentrated wastewater effluent samples from year 2017 highlight the enrichment for tobamoviral reads in linear gradient fraction group b (Lin[b]). Enlarged grey tree in the lower left corner shows the taxonomic groups represented in the ssRNA viromes of all the normalized samples, where node sizes correspond to the number of taxa based on all samples together. Comparisons between ssRNA viromes of differently concentrated samples are shown for all the pairs of samples as small heat trees above the diagonal. The differences in abundance of the taxa for the two compared samples are shown in color: brown color designates higher abundance of taxa in brown coded sample (listed in the horizontal axis) and green color designates higher abundance in green coded sample (listed in the vertical axis). The color of each taxon represents the log-2 ratio of reads abundance of the two different samples compared.

sRNA HTS (a non-targeted plant virus detection approach) we detected high number of reads assigned to tobamoviruses in pools of symptomatic and also asymptomatic plants (Table 2). Complete genome sequences of detected viruses were covered by sRNA reads and consensus viral genomes of infecting viruses have been reconstructed.

Four weeks after inoculation with influent linear gradient elution fractions from year 2017 (see Supplementary Fig. 1), some *N. benthamiana* plants showed reduction in size, curling and mottling of the leaves (Fig. 5a). sRNA sequencing of the pooled plant samples with similar symptoms (Table 2) identified PMMoV in the pool. *N. occidentalis* plants inoculated with the same fractions showed two different types of symptomatic changes. Again, we confirmed PMMoV infection in pool of *N. occidentalis* samples with the symptoms of curling and mottling (Fig. 5b, Table 2). Moreover, in the pool of *N. occidentalis* plants exhibiting curling and necrotic spots (Fig. 5c, Table 2) we detected mixed infection with PMMoV and ToMV. Finally, also in the pool of *N. occidentalis* plants not showing any distinct disease symptoms at the time of sampling we detected PMMoV (Table 2).

To assess the effect of the wastewater treatment on plant virus infectivity, we also inoculated test plants with the effluent of the WWTP from year 2017, which is released directly into the river. For mechanical inoculation we used step gradient elution fraction E1 (for details, see Supplementary Fig. 1). *N. benthamiana* plants again showed disease symptoms with reduction of the size of the plant and curling, which were linked with PMMoV infection (Fig. 5d, Table 2) using sRNA sequencing of pooled symptomatic plant material. Additionally, a pool of *N. occidentalis* plants with mosaic symptoms (Fig. 5e) as well as pool of asymptomatic *N. benthamiana* plants were shown to be infected with TMGMV (Table 2). All of the influent and effluent fractions that successfully infected test plants contained high number of reads of the detected infective viruses (Fig. 3). TEM examination of symptomatic plant material samples revealed the presence of filamentous virus particles of about 300 nm in length, typical for tobamoviruses (Fig. 5f and g).

4. Discussion

In this study, we explore the diversity of viruses in wastewater, focusing on highly economically relevant ssRNA plant viruses. We described the presence of such viruses in influent and effluent samples of WWTP and confirmed the infectivity of some plant viruses, thus proving their biological significance in wastewater. These results implicate the flux of infective pathogenic plant viruses from anthropogenic environments into environmental waters and open important questions about the role of wastewater in the spread of viral diseases of plants.

In this study we detected members of 11 families of plant pathogenic ssRNA viruses (Fig. 3). Tobamoviruses, which were the

Table 2
Viruses detected using sRNA sequencing in plant tissue pools after inoculation of test plants with wastewaters CIM concentrates from year 2017 and summary of mapping reads to corresponding consensus viral genome sequences.

Plant sample pools	Test plants	Symptoms	Viruses detected	Reads mapped to corresponding viral genome sequence	Reads mapped to corresponding viral genome sequence (%)	Average depth of coverage for corresponding viral species	GenBank accession number (complete genome)	SRA accession number (raw data)
INFLUENT								
No_INF-1	N. occidentalis	curling, mottling	PMMoV	1374 679	16.86	4630.64	MN267898	SRR9319298
No_INF-2	*N. occidentalis*	curling, necrotic	ToMV	1270 247	13.90	4320.37	MN267904	SRR9319297
		curling, necrotic	PMMoV	122 901	1.56	414.2	MN267899	
No_INF-3	*N. occidentalis*	asymptomatic	PMMoV	285 047	5.98	958.18	MN267900	SRR9319300
Nb_INF-4	*N. benthamiana*	curling, mottling	PMMoV	1057 897	13.95	3574.51	MN267901	SRR9319299
Nb_INF-5	*N. benthamiana*	asymptomatic	/	/	/	/		SRR9319304
EFFLUENT								
No_EFF-1	*N. occidentalis*	mosaic	TMGMV	2214 61	28.55	7460.34	MN267903	SRR9319303
No_EFF-2	*N. occidentalis*	asymptomatic	TMGMV	1208 726	15.80	4063.91	MN267902	SRR9319296
Nb_EFF-3	*N. benthamiana*	curling	PMMoV	1 804 063	23.81	6076.79	MN267897	SRR9319305
Nb_EFF-4	*N. benthamiana*	asymptomatic	/	/	/	/		SRR9319302

most abundant plant viruses detected in analyzed samples, are considered a major risk to a range of agriculturally important plant species belonging to the *Solanaceae*, *Cucurbitaceae* and other plant families. Their sequences were previously detected in different environmental waters including ballast water (Kim et al., 2015), irrigation systems (Boben et al., 2007), drinking water (Haramoto et al., 2013) and raw and urban sewage (Cantalupo et al., 2011; Fernandez-Cassi et al., 2018). Viruses from this group are known to have extremely stable virions that can resist high temperatures (Bawden and Pirie, 1959; Mutombo et al., 1992). Tobacco, tomato, and pepper are known hosts for ToMV, TMGMV and PMMoV (Smith, 2017), all three found to be present in wastewater as infective particles in this study. We detected reads belonging to CGMMV, which is the most economically important cucurbit-infecting tobamovirus and currently considered a significant threat to the production of cucumber, melon, watermelon, gherkin, and pumpkin (Smith, 2017). We also detected sequences of ToBRFV, first identified in tomatoes in Jordan in 2015 (Mansour and Falk, 2016). Outbreaks of this emerging virus are causing major concerns for growers of tomato and pepper in different countries worldwide (EPPO, 2019a). The virus is easily transmitted, which together with the lack of ToBRFV resistant tomato varieties generates a significant threat for tomato production worldwide (Luria et al., 2017). The detection of ToBRFV sequences in wastewater in Slovenia, a country where it has not been reported yet, thus raises questions about its origin, possibility of unnoticed presence and risks of its water-mediated transmission.

From the *Bromoviridae* family, CMV sequences were detected in the majority of the samples. CMV is known to infect over 1200 plant species including important vegetable crops and ornamentals (Zitter and Murphey, 2009). In the past, it was isolated from river water (Mehle and Ravnikar, 2012; Piazzolla, 1986). We detected sequences of melon necrotic spot virus (MNSV) from *Tombusviridae* family in all analyzed wastewater samples. MNSV, affecting species of the *Cucurbitaceae* family, is transmitted mainly by attaching itself to the outer layer of the aquatic zoospores of the fungus *Olpidium bornovanus* (Gosalvez et al., 2003). Sequences of potato viruses such as carlaviruses (potato virus S and potato virus M) and potyvirus (PVY), which are on the list of quarantine pests in some countries (EPPO, 2019b) were also detected. PVY is the most important viral pathogen in potato worldwide, with different strains of PVY causing diseases in tobacco, potato, pepper and tomato (Tsedaley, 2015). PepMV (*Potexvirus*, *Alphaflexiviridae*) sequences were found in some influent and effluent samples. PepMV causes great concern in the greenhouse tomato industry and is considered as a pest recommended for regulation as quarantine pest (EPPO, 2019b). Experiments using a hydroponic system showed that PepMV and PVY can be released from plant roots into the nutrient solution and can infect healthy plants through their roots (Mehle et al., 2014).

Recent studies demonstrated the usage of wastewater metagenomics for detection of viral outbreaks in human populations (Fernandez-Cassi et al., 2018; Xagoraraki and Brien, 2019). In the present study we confirmed previous observations, that wastewaters metagenomics can also provide a comprehensive view into the presence and abundance of plant viruses, possibly reflecting the diversity of local plants as well as the plants consumed by local residents and animals (Ng et al., 2012). We detected nucleic acids of many viruses, which were never reported in Slovenia before (Fig. 3) suggesting that wastewater might provide a pathway for the dissemination and gradual globalization of plant viruses, however the number of sequences detected for specific viruses varied from very low to high. While metagenomics studies of plant viruses in wastewater collected from different sources in the region provide an insight into the diversity of plant viruses circulating in a given environment, including emerging ones, it needs to be followed by systematic tests for specific viruses, especially in terms of their infectivity.

The availability of an effective concentration method is crucial for metagenomic detection of viral pathogens for aquatic surveillance purposes (Fernandez-Cassi et al., 2018). The CIM monolithic gradient elution enables both concentration and fractionation of the wastewater virome, resulting in tobamovirus reads being most abundant in specific fraction groups (Inf17-Lin^{b}, Ef17-Lin^{b}) (Figs. 3c and 4). The ability of the CIM chromatographic sample preparation step to increase viral concentration and at the same time to remove impurities present in the sample (Rupar et al., 2013) likely contributed to the success of infectivity tests, confirming biological significance of viral sequences detected in HTS based surveillance of wastewater.

In the past, infectious plant viruses have been isolated from various types of environmental water in different locations around the world (Koenig, 1986; Mehle et al., 2018; Mehle and Ravnikar, 2012), including irrigation and drainage canals (Jeżewska et al., 2018). However, infectivity of plant viruses was never studied in

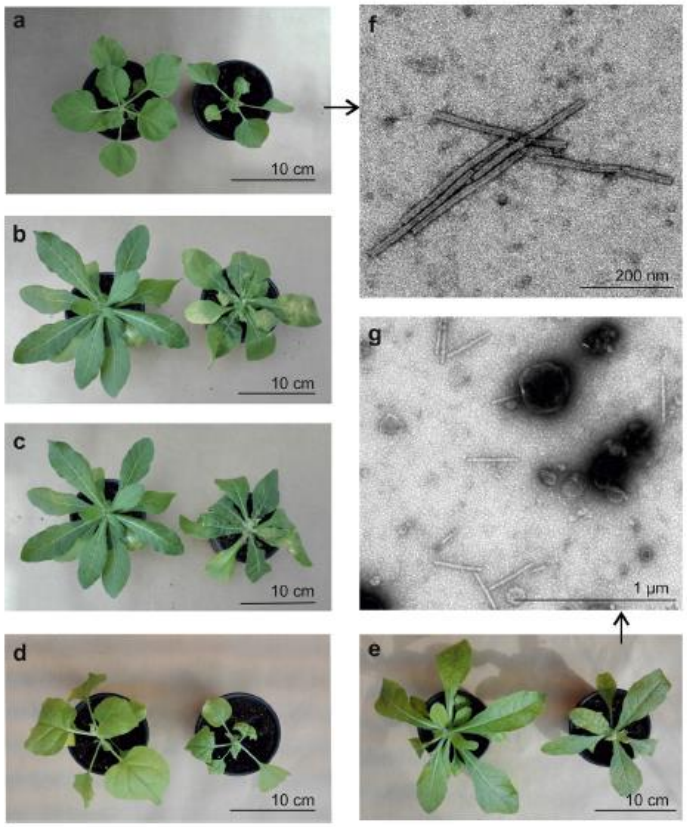


Fig. 5. Disease symptoms on inoculated test plants (right plant) after mechanical inoculation together with corresponding negative controls of inoculation (left plant) and viral particles in symptomatic plant material visualized by TEM. (a) PMMoV infected *N. benthamiana* inoculated with influent sample – fraction 20 (comprised in Inf17- Lin^{b}), (b) PMMoV infected *N. occidentalis* inoculated with influent sample – fraction 18 (comprised in Inf17- Lin^{b}), (c) ToMV infected *N. occidentalis* inoculated with influent sample – fraction 16 (comprised in Inf17- Lin^{b}), (d) PMMoV infected *N. benthamiana* inoculated with effluent sample – fraction E1 (comprised in Ef17-Step), (e) TMGMV infected *N. occidentalis* inoculated with effluent sample fraction E1 (comprised in Ef17-Step). (f) TEM micrograph of PMMoV infected *N. benthamiana* inoculated with influent sample, (g) TEM micrograph of TMGMV infected *N. occidentalis* inoculated with effluent sample.

influents and effluents of WWTP, even though using reclaimed wastewater for irrigation purposes is becoming a widespread practice (Pedrero et al., 2010). The combination of CIM concentration of wastewater samples and mechanical inoculation of selected test plants used here enabled us to observe symptomatic changes in selected test plants and confirm the infectivity of viruses detected by the metagenomic analysis of both treated and untreated wastewater. Using sRNA sequencing that enables generic detection of RNA snd DNA viruses and is widely used in plant virus discovery (Pecman et al., 2017; Roossinck et al., 2015), we were able to detect three infective tobamoviruses: PMMoV, ToMV and TMGMV (Table 2).

High diversity of plant viral sequences and confirmed infectivity of some important pathogenic plant viruses in both treated and untreated wastewater call for the broad consideration of the use of reclaimed water for irrigation purposes, since it poses a risk of disease spread to irrigated plants and plants grown in hydroponic systems. This consideration is especially relevant for highly stable and mechanically transmissible viruses, such as the ones detected in this study. Although virus removal in water was studied and can be achieved by careful selection of treatment processes, such as coagulation-sedimentation, ozonation, microfiltration (Canh et al., 2019; Kato et al., 2018) there are still needs for the development of new cost effective and environmentally friendly technologies for efficient virus inactivation.

This study suggests that plant viruses, especially tobamoviruses, can potentially circulate between urban environment and croplands: from infected plants, through human and animal consumption and excreta, to wastewater, from where they could be transmitted back to plants through, e.g., irrigation with reclaimed wastewater. The release of infective plant viruses into environment through wastewater might have consequences in rapid

transmission of pathogenic viruses to new areas, which cannot be reached by other vectors. Hypothetically, through such process the global incidence of diseases and their impact on the ecosystem can increase more rapidly for a broad range of crops (Mehle et al., 2018). Wastewater-based epidemiology has the potential to predict critical locations and critical moments for viral disease onset (Xagoraraki and Brien, 2019), however it has not yet been applied to predict and prevent plant viral disease outbreaks. The results and methodology used in this study provides a baseline for monitoring reclaimed water for the presence of pathogenic viruses and to monitor and improve wastewater treatment processes for efficient removal or inactivation of viruses.

5. Conclusions

- High-throughput sequencing analysis of wastewater metagenomes confirmed a high diversity and abundance of plant viruses in wastewater treatment plant influents and effluents. Highly stable tobamoviruses were the dominant representatives of plant viruses in wastewater.
- Nucleic acids of plant pathogenic viruses not previously detected in the region were present in both untreated and treated wastewater, including emerging and quarantine viruses. This indicates the applicability of wastewater-based epidemiology for early detection and monitoring of economically important plant viruses, however, detection of nucleic acids needs to be coupled with specific follow-up studies, especially concerning virus infectivity.
- Wastewater contained infective plant viruses, which remained infective after conventional wastewater treatment. This indicates the flux of plant viruses from the urban environment to the ecosystem. The presence of infective plant viruses in wastewater can also have consequences in rapid transmission of pathogenic viruses to new areas, especially when wastewater or reclaimed water is used for irrigation.

Data availability

All sequencing data that support the findings of this study are linked under the accession number PRJNA549409 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/).

Complete consensus viral genome sequences generated in this study have been deposited in GenBank under the accession numbers MN267897-MN267904.

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.

Acknowledgements

This work was supported by Slovenian research agency (P4-0407, P4-0165 and young researchers' grant awarded to K.B.). We would like to thank employees of Central wastewater treatment plant Domžale-Kamnik for providing the wastewater samples and Dr. Carolyn Malmstrom and Dr. Neil Boonham for reviewing an early version of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.watres.2020.115628.

References

Alhamlan, F.S., Ederer, M.M., Brown, C.J., Coats, E.R., Crawford, R.L., 2013. Metagenomics-based analysis of viral communities in dairy lagoon wastewater. J. Microbiol. Methods 92, 183–188. https://doi.org/10.1016/j.mimet.2012.11.016.

Aw, T.G., Howe, A., Rose, J.B., 2014. Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. J. Virol. Methods 210, 15–21. https://doi.org/10.1016/j.jviromet.2014.09.017.

Balasubramanian, M.N., Rački, N., Gonçalves, J., Kovač, K., Žnidarič, M.T., Turk, V., Ravnikar, M., Gutiérrez-Aguirre, I., 2016. Enhanced detection of pathogenic enteric viruses in coastal marine environment by concentration using methacrylate monolithic chromatographic supports paired with quantitative PCR. Water Res. 106, 405–414. https://doi.org/10.1016/j.watres.2016.10.020.

Bar-On, Y.M., Phillips, R., Milo, R., 2018. The biomass distribution on Earth. Proc. Natl. Acad. Sci. Unit. States Am. 115, 6506–6511. https://doi.org/10.1073/pnas.1711842115.

Bawden, F.C., Pirie, N.W., 1959. The infectivity and inactivation of nucleic acid preparations from tobacco mosaic virus. J. Gen. Microbiol. 21, 438–456.

Bibby, K., Peccia, J., 2013. Identification of viral pathogen diversity in sewage sludge by metagenome analysis. Environ. Sci. Technol. 47, 1945–1951. https://doi.org/10.1021/es305181x.

Boben, J., Kramberger, P., Petrovič, N., Cankar, K., Peterka, M., Štrancar, A., Ravnikar, M., 2007. Detection and quantification of Tomato mosaic virus in irrigation waters. Eur. J. Plant Pathol. 118, 59–71. https://doi.org/10.1007/s10658-007-9112-1.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12. https://doi.org/10.1038/nmeth.3176.

Canh, V.D., Hiroaki, F., Katayama, H., 2019. Removal of pepper mild mottle virus by full-scale micro filtration and slow sand filtration plants. Npj Clean Water 1–7. https://doi.org/10.1038/s41545-019-0042-1.

Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix, R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw sewage harbors diverse viral populations. mBio 2, 1–11. https://doi.org/10.1128/mBio.00180-11.

Djikeng, A., Kuzmickas, R., Anderson, N.G., Spiro, D.J., 2009. Metagenomic analysis of RNA viruses in a fresh water lake. PloS One 4, 1–5. https://doi.org/10.1371/journal.pone.0007264.

Dutilh, B.E., Cassman, N., Mcnair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., Felts, B., Dinsdale, E.A., Mokili, J.L., Edwards, R.A., 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat. Commun. 5, 1–11. https://doi.org/10.1038/ncomms5498.

EPPO, 2019a. EPPO Reporting Service 2019, p. 1.

EPPO, 2019b. EPPO A2 List of Pests Recommended for Regulation as Quarantine Pests.

Fernandez-Cassi, X., Timoneda, N., Martinez-Puchol, S., Rusinol, M., Rodriguez-Manzano, J., Figuerola, N., Bofill-Mas, S., Abril, J.F., Girones, R., 2018. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. Sci. Total Environ. 618, 870–880. https://doi.org/10.1016/j.scitotenv.2017.08.249.

Foster, Z.S.L., Sharpton, T., Grunwald, N.J., 2017. MetacodeR: an R package for manipulation and heat tree visualization of community taxonomic data from metabarcoding. PLoS Comput. Biol. 13, 1–15. https://doi.org/10.1371/journal.pcbi.1005404.

Gosalvez, B., Navarro, J.A., Lorca, A., Botella, F., Sánchez-Pina, M.A., Pallas, V., 2003. Detection of Melon necrotic spot virus in water samples and melon plants by molecular methods. J. Virol. Methods 113, 87–93. https://doi.org/10.1016/S0166-0934(03)00224-6.

Gutiérrez-Aguirre, I., Banjac, M., Steyer, A., Poljšak-Prijatelj, M., Peterka, M., Štrancar, A., Ravnikar, M., 2009. Concentrating rotaviruses from water samples using monolithic chromatographic supports. J. Chromatogr. A 1216, 2700–2704. https://doi.org/10.1016/j.chroma.2008.10.106.

Gutiérrez-Aguirre, I., Steyer, A., Banjac, M., Kramberger, P., Poljšak-Prijatelj, M., Ravnikar, M., 2011. On-site reverse transcription-quantitative polymerase chain reaction detection of rotaviruses concentrated from environmental water samples using methacrylate monolithic supports. J. Chromatogr. A 1218, 2368–2373. https://doi.org/10.1016/j.chroma.2010.10.048.

Gutiérrez Aguirre, I., Kutnjak, D., Rački, N., Rupar, M., Ravnikar, M., 2018. Monolith chromatography as sample preparation step. Viral Metagenomics: Methods and Protocols. Springer Science Business Media, pp. 63–75. https://doi.org/10.1007/978-1-4939-7683-6_6.

Hammond, G.W., Hazelton, P.R., Chuang, I., Klisko, B., 1981. Improved detection of viruses by electron microscopy after direct ultracentrifuge preparation of specimens. J. Clin. Microbiol. 14, 210–221.

Haramoto, E., Kitajima, M., Hata, A., Torrey, J.R., Masago, Y., Sano, D., Katayama, H., 2018. A review on recent progress in the detection methods and prevalence of human enteric viruses in water. Water Res. 135, 168–186. https://doi.org/10.1016/j.watres.2018.02.004.

Haramoto, E., Kitajima, M., Kishida, N., Konno, Y., Katayama, H., Asami, M., Akiba, M., 2013. Occurrence of pepper mild mottle virus in drinking water sources in Japan. Appl. Environ. Microbiol. 79, 7413–7418. https://doi.org/10.1128/AEM.02354-13.

Hjelmsø, M.H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M., Elsässer, D., Aarestrup, F.M., Löfström, C., Bofill-Mas, S., Abril, J.F., Girones, R., Schultz, A.C., 2017. Evaluation of methods for the concentration and

extraction of viruses from sewage in the context of metagenomic sequencing. PloS One 12, 1–17. https://doi.org/10.1371/journal.pone.0170199.
Hull, R., Bustamante, P., 1998. Plant virus gene expression strategies. EJB Electron. J. Biotechnol. 1, 1–19.
Huson, D.H., Beier, S., Flade, I., Górska, A., El-hadidi, M., 2016. MEGAN Community edition - interactive exploration and analysis of larges-scale microbiome sequencing data. PLoS Comput. Biol. 12, 1–12. https://doi.org/10.1371/journal.pcbi.1004957.
Jeżewska, M., Trzmiel, K., Zarzyńska-nowak, A., 2018. Detection of infectious tobamoviruses in irrigation and drainage canals in Greater Poland. J. Plant Prot. 58, 202–205. https://doi.org/10.24425/119126.
Kato, R., Asami, T., Utagawa, E., Furumai, H., Katayama, H., 2018. Pepper mild mottle virus as a process indicator at drinking water treatment plants employing coagulation-sedimentation, rapid sand filtration, ozonation, and biological activated carbon treatments in Japan. Water Res. 132, 61–70. https://doi.org/10.1016/j.watres.2017.12.068.
Kim, Y., Aw, T.G., Teal, T.K., Rose, J.B., 2015. Metagenomic investigation of viral communities in ballast water. Environ. Sci. Technol. 49, 8396–8407. https://doi.org/10.1021/acs.est.5b01633.
Kitajima, M., Sassi, H.P., Torrey, J.R., 2018. Pepper mild mottle virus as a water quality indicator. npj Clean Water 1, 1–19. https://doi.org/10.1038/s41545-018-0019-5.
Koenig, R., 1986. Plant viruses in rivers and lakes. Adv. Virus Res. 31, 321–333.
Kuroda, K., Nakada, N., Hanamoto, S., Inaba, M., Katayama, H., Thuan, A., Thi, T., Nga, V., Oguma, K., Hayashi, T., Takizawa, S., 2015. Pepper mild mottle virus as an indicator and a tracer of fecal pollution in water environments : comparative evaluation with wastewater-tracer pharmaceuticals in Hanoi, Vietnam. Sci. Total Environ. 506–507, 287–298. https://doi.org/10.1016/j.scitotenv.2014.11.021.
Luria, N., Smith, E., Reingold, V., Bekelman, I., Lapidot, M., Levin, I., Elad, N., Tam, Y., Sela, N., Abu-ras, A., Ezra, N., Haberman, A., Yitzhak, L., Lachman, O., Dombrovsky, A., 2017. A new Israeli tobamovirus isolate infects tomato plants harboring Tm-22 resistance genes. PloS One 12, 1–19. https://doi.org/10.1371/journal.pone.0170429.
Mansour, N.S.A., Falk, M.C.B.W., 2016. A new tobamovirus infecting tomato crops in Jordan. Arch. Virol. 161, 503–506. https://doi.org/10.1007/s00705-015-2677-7.
Mehle, N., Gutiérrez-Aguirre, I., Kutnjak, D., Ravnikar, M., 2018. Water-mediated transmission of plant, animal, and human viruses. In: Advances in Virus Research, pp. 85–128.
Mehle, N., Gutiérrez-Aguirre, I., Prezelj, N., Delić, D., Vidic, U., Ravnikar, M., 2014. Survival and transmission of potato virus Y, pepino mosaic virus, and potato spindle tuber viroid in water. Appl. Environ. Microbiol. 80, 1455–1462. https://doi.org/10.1128/AEM.03349-13.
Mehle, N., Ravnikar, M., 2012. Plant viruses in aqueous environment - survival, water mediated transmission and detection. Water Res. 46, 4902–4917. https://doi.org/10.1016/j.watres.2012.07.027.
Mohiuddin, M., Schellhorn, H.E., 2015. Spatial and temporal dynamics of virus occurrence in two freshwater lakes captured through metagenomic analysis. Front. Microbiol. 6, 1–13. https://doi.org/10.3389/fmicb.2015.00960.
Mutombo, K., Michels, B., Ott, H., Cerf, R., Witz, J., 1992. Scanning calorimetric studies of the stability of tobacco mosaic virus and aggregates of its coat protein. Eur. Biophys. J. 83, 77–83.
Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B.S., Wommack, K.E., Delwart, E., 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J. Virol. 86, 12161–12175. https://doi.org/10.1128/JVI.00869-12.
Nicaise, V., 2014. Crop immunity against viruses : outcomes and future challenges. Front. Plant Sci. 5, 1–18. https://doi.org/10.3389/fpls.2014.00660.
O'Brien, E., Munir, M., Marsh, T., Heran, M., Lesage, G., Tarabara, V.V., Xagoraraki, I., 2017. Diversity of DNA viruses in effluents of membrane bioreactors in traverse city, MI (USA) and La grande motte (France). Water Res. 111, 338–345. https://doi.org/10.1016/j.watres.2017.01.014.
Ogilvie, L.A., Nzakizwanayo, J., Guppy, F.M., Dedi, C., Diston, D., Taylor, H., Ebdon, J., Jones, B.V., 2018. Resolution of habitat-associated ecogenomic signatures in bacteriophage genomes and application to microbial source tracking. ISME J. 12, 942–958. https://doi.org/10.1038/s41396-017-0015-7.
Pecman, A., Kutnjak, D., Gutiérrez-Aguirre, I., Adams, I., Fox, A., Boonham, N., Ravnikar, M., 2017. Next generation sequencing for detection and discovery of plant viruses and viroids: comparison of two approaches. Front. Microbiol. 8, 1–10. https://doi.org/10.3389/fmicb.2017.01998.
Pedrero, F., Kalavrouziotis, I., Alarcón, J.J., Koukoulakis, P., Asano, T., 2010. Use of treated municipal wastewater in irrigated agriculture-Review of some practices in Spain and Greece. Agric. Water Manag. 97, 1233–1241. https://doi.org/10.1016/j.agwat.2010.03.003.
Piazzolla, P., 1986. Presence of plant viruses in some rivers of southern Italy. J. Phytopathol. 116, 244–246.
Rodriguez-Brazzarola, P., Perez-Wohlfeil, E., Diaz-del-Pino, S., 2018. Analyzing the differences between reads and contigs when performing a taxonomic assignment comparison in metagenomics. Bioinforma. Biomed. Eng. 10813, 450–460. https://doi.org/10.1007/978-3-319-78723-7.
Roossinck, M.J., Martin, D.P., Roumagnac, P., 2015. Plant virus Metagenomics : advances in virus discovery. Phytophatology 105, 716–727. https://doi.org/10.1094/PHYTO-12-14-0356-RVW.
Rosario, K., Morrison, C.M., Mettel, K.A., Betancourt, Q., 2019. Novel circular rep-encoding single-stranded DNA viruses detected in treated wastewater 9–10. https://doi.org/10.1128/MRA.00318-19.
Rosario, K., Nilsson, C., Lim, Y.W., Ruan, Y., Breitbart, M., 2009a. Metagenomic analysis of viruses in reclaimed water. Environ. Microbiol. 11, 2806–2820. https://doi.org/10.1111/j.1462-2920.2009.01964.x.
Rosario, K., Symonds, E.M., Sinigalliano, C., Stewart, J., Breitbart, M., 2009b. Pepper mild mottle virus as an indicator of fecal pollution. Appl. Environ. Microbiol. 75, 7261–7267. https://doi.org/10.1128/AEM.00410-09.
Rupar, M., Ravnikar, M., Tušek-Žnidarič, M., Kramberger, P., Glais, L., Gutiérrez-Aguirre, I., 2013. Fast purification of the filamentous Potato virus Y using monolithic chromatographic supports. J. Chromatogr. A 1272, 33–40. https://doi.org/10.1016/j.chroma.2012.11.058.
Ruščič, J., Gutiérrez-Aguirre, I., Tušek Žnidarič, M., Kolundžija, S., Slana, A., Barut, M., Ravnikar, M., Krajačič, M., 2015. A new application of monolithic supports: the separation of viruses from one another. J. Chromatogr. A 1388, 69–78. https://doi.org/10.1016/j.chroma.2015.01.097.
Sastry and Zitter, 2014. Management of virus and viroid diseases of crops in the tropics. In: Plant Virus and Viroid Diseases in the Tropics. Springer Science Business Media, pp. 149–480.
Shkoporov, A.N., Ryan, F.J., Draper, L.A., Forde, A., Stockdale, S.R., Daly, K.M., Mcdonnell, S.A., Nolan, J.A., Sutton, T.D.S., Dalmasso, M., Mccann, A., Ross, R.P., Hill, C., 2018. Reproducible protocols for metagenomic analysis of human faecal phageomes. Microbiome 6, 1–17. https://doi.org/10.1186/s40168-018-0446-z.
Smith, A.D., 2017. Seed transmission of Tobamoviruses: aspects of global disease distribution. In: Advances in Seed Biology, pp. 235–260. https://doi.org/10.5772/32009.
Symonds, E.M., Rosario, K., Breitbart, M., 2019. Pepper mild mottle virus : agricultural menace turned effective tool for microbial water quality monitoring and assessing (waste) water treatment technologies. PLoS Pathog. 15, 1–7. https://doi.org/10.1371/journal.ppat.1007639.
Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P.Y., Yasunaga, T., Kamagata, Y., Liu, W.T., 2012. Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. Environ. Microbiol. 14, 441–452. https://doi.org/10.1111/j.1462-2920.2011.02630.x.
Thebo, A.L., Drechsel, P., Lambin, E.F., Nelson, K.L., 2017. A global, spatially-explicit assessment of irrigated croplands influenced by urban wastewater flows. Environ. Res. Lett. 12 https://doi.org/10.1088/1748-9326/aa75d1.
Tsedaley, B., 2015. A review paper on Potato virus Y (PVY) biology , economic importance and its managements. J. Biol. Agric. Healthc. 5, 110–127.
Wang, D., Urisman, A., Liu, Y., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R., Hickenbotham, M., Magrini, V., Eldred, J., Latreille, J.P., Wilson, R.K., Ganem, D., Derisi, J.L., 2003. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol. 1 https://doi.org/10.1371/journal.pbio.0000002.
Wylie, S.J., Zhang, C., Long, V., Roossinck, M.J., Koh, S.H., 2015. Differential responses to virus challenge of laboratory and wild accessions of australian species of Nicotiana, and comparative analysis of RDR1 gene sequences. PloS One 10, 1–22.
Xagoraraki, I., Brien, E.O., 2019. Wastewater-based epidemiology for early detection of viral outbreaks. In: Women in Water Quality, pp. 75–97. https://doi.org/10.1007/978-3-030-17819-2_5.
Zhang, T., Breitbart, M., Lee, W.H., Run, J.Q., Wei, C.L., Soh, S.W.L., Hibberd, M.L., Liu, E.T., Rohwer, F., Ruan, Y., 2006. RNA viral community in human feces: prevalence of plant pathogenic viruses. PLoS Biol. 4, 108–118. https://doi.org/10.1371/journal.pbio.0040003.
Zhao, X., Zhou, Q., Zhang, L., Yan, W., Sun, N., Liang, X., Deng, C., 2015. Comparison of three magnetic-bead-based RNA extraction methods for detection of cucumber green mottle mosaic virus by real-time RT-PCR. Arch. Virol. 160, 1791–1796. https://doi.org/10.1007/s00705-015-2444-9.
Zitter, T.A., Murphey, J.F., 2009. Cucumber mosaic virus. Plant Heal. Instr 1–11. https://doi.org/10.1094/PHI-I-2009-0518-01.

2.2 Virome Analysis of Signal Crayfish (*Pacifastacus leniusculus*) Along Its Invasion Range Reveals Diverse and Divergent RNA Viruses

Katarina Bačnik, Denis Kutnjak, Silvija Černi, Ana Bielen, Sandra Hudina

Viruses, 2021, 13(11), 2259; https://doi.org/10.3390/v13112259

In this publication we used HTS of ribosomal RNA-depleted total RNA isolated from the crayfish hepatopancreas and subsequent sequence data analysis to identify novel and divergent RNA viruses, including signal crayfish-associated reo-like, hepe-like, toti-like, and picorna-like viruses, phylogenetically related to viruses previously associated with crustacean hosts. The patterns of reads abundance and calculated nucleotide diversities of the detected viral sequences varied along the invasion range. This could indicate the possible influences of different factors and processes on signal crayfish virome composition: e.g., the differences in signal crayfish population density, the non-random dispersal of host individuals from the core to the invasion fronts, and the transfer of viruses from the native co-occurring and phylogenetically related crayfish species. The study reveals a high, previously undiscovered diversity of divergent RNA viruses associated with signal crayfish, and sets foundations for understanding the potential risk of virus transmissions as a result of this invader's dispersal.

The PhD candidate performed bioinformatics analysis to identify putative viral sequences, and to characterize viral genomes of novel and divergent RNA viruses. She performed phylogenetic analyses of RNA-dependent RNA polymerase (RdRp) sequences, determined the nucleotide diversities of populations of selected signal crayfish-associated viruses and designed RT-PCR primer pairs to confirm their presence in the samples. She also wrote the first draft of the manuscript.

viruses

MDPI

Article

Virome Analysis of Signal Crayfish (*Pacifastacus leniusculus*) along Its Invasion Range Reveals Diverse and Divergent RNA Viruses

Katarina Bačnik [1,2], Denis Kutnjak [1,*], Silvija Černi [3], Ana Bielen [4] and Sandra Hudina [3,*]

[1] Department of Biotechnology and Systems Biology, National Institute of Biology, 1000 Ljubljana, Slovenia; katarina.bacnik@nib.si
[2] Jozef Stefan International Postgraduate School, 1000 Ljubljana, Slovenia
[3] Department of Biology, Faculty of Science, University of Zagreb, 10000 Zagreb, Croatia; silvija.cerni@biol.pmf.hr
[4] Department of Biochemical Engineering, Faculty of Food Technology and Biotechnology, University of Zagreb, 10000 Zagreb, Croatia; abielen@pbf.hr
* Correspondence: denis.kutnjak@nib.si (D.K.); sandra.hudina@biol.pmf.hr (S.H.)

Abstract: Crayfish are a keystone species of freshwater ecosystems and a successful invasive species. However, their pathogens, including viruses, remain understudied. The aim of this study was to analyze the virome of the invasive signal crayfish (*Pacifastacus leniusculus*) and to elucidate the potential differences in viral composition and abundance along its invasion range in the Korana River, Croatia. By the high-throughput sequencing of ribosomal RNA, depleted total RNA isolated from the crayfish hepatopancreas, and subsequent sequence data analysis, we identified novel and divergent RNA viruses, including signal crayfish-associated reo-like, hepe-like, toti-like, and picorna-like viruses, phylogenetically related to viruses previously associated with crustacean hosts. The patterns of reads abundance and calculated nucleotide diversities of the detected viral sequences varied along the invasion range. This could indicate the possible influence of different factors and processes on signal crayfish virome composition: e.g., the differences in signal crayfish population density, the non-random dispersal of host individuals from the core to the invasion fronts, and the transfer of viruses from the native co-occurring and phylogenetically related crayfish species. The study reveals a high, previously undiscovered diversity of divergent RNA viruses associated with signal crayfish, and sets foundations for understanding the potential risk of virus transmissions as a result of this invader's dispersal.

Keywords: signal crayfish virome; RNA viruses; invasive alien species; invasion range; high-throughput sequencing

Citation: Bačnik, K.; Kutnjak, D.; Černi, S.; Bielen, A.; Hudina, S. Virome Analysis of Signal Crayfish (*Pacifastacus leniusculus*) along Its Invasion Range Reveals Diverse and Divergent RNA Viruses. *Viruses* **2021**, *13*, 2259. https://doi.org/10.3390/v13112259

Academic Editors: Leyi Wang and Ganwu Li

Received: 1 October 2021
Accepted: 4 November 2021
Published: 11 November 2021

1. Introduction

Until recently, knowledge of invertebrate viruses was limited mostly to viral pathogens causing high mortalities and arboviruses (arthropod-borne viruses), which are vectored by arthropods and cause disease in humans and other vertebrate species [1]. This view has changed, with metagenomics revealing remarkable levels of RNA virus diversity in invertebrates [2–5], and, in this perspective, the detection of invertebrate disease-causing viruses became the exception rather than the rule [1]. In addition to transforming our understanding of virus diversity, virus metagenomics provides insights into different aspects of virus evolution [6], including the role of invertebrates and their ecological interactions with other organisms in the evolution of RNA viruses [5].

For understanding the diversity of invertebrate viruses, arthropods, including crustaceans, may be of particular importance, as they account for over 80% of the total animal diversity [7], are globally abundant, and often live in extremely large and dense populations [6].

Crayfish, belonging to a diverse order of decapod crustaceans, are a keystone species of freshwater ecosystems, and their populations can reach high densities. They are an important component of freshwater food webs and ecosystem engineers due to their larger size and longer life span compared to other benthic macroinvertebrates, as well as due to their bioturbation and burrowing activity [8]. Crayfish are also among the most successful invasive alien species (IAS) in freshwater ecosystems [9]. Throughout history, crayfish have been frequently translocated for ornamental or aquaculture purposes [10] and, after being intentionally or unintentionally released into the wild, have often rapidly expanded their range and exerted a high number of documented negative impacts [10,11]. On the other hand, native crayfish populations in the wild are increasingly endangered on the global scale [9,10,12,13], and invasive crayfish and their pathogens have been identified as one of the most prominent factors contributing to the decline of native species' populations [14,15]. Some of the pathogens introduced by invasive crayfish are significantly more virulent towards the new (i.e., native) hosts than toward IAS hosts [16,17]. Thus, invasive crayfish species have a high potential for introducing and spreading the emerging diseases that may incur significant economic and ecological losses [18,19]. However, the existing literature on crayfish diseases is biased mostly towards the crayfish plague pathogen, the oomycete *Aphanomyces astaci*, while other disease agents, including fungi, bacteria, and viruses, are significantly understudied [20].

Most of the knowledge on crustacean pathogenic viruses comes from the study of viral diseases in aquaculture [20]. For instance, white spot syndrome virus causes the white spot disease that has mostly been described due to its symptoms in aquaculturally relevant species such as *Cherax quadricarinatus* [21], although it displays a low degree of host specificity, and is therefore also an important pathogen in crayfish populations in the wild [22]. Despite their large potential impact on aquaculture and populations in the wild, viruses are the least studied group of crayfish pathogens [20], and relatively few have been formally characterized and classified [23]. However, with the increasing availability of high-throughput sequencing (HTS) technologies and metagenomics approaches, which can reveal all of the viruses (i.e., virome) associated with an individual or an environmental sample, comprehensive descriptions and a revolution in our understanding of the viral diversity of crustacean species are occurring [23]. For instance, this approach was recently employed to discover bunya-like brown spot virus, a new negative, single stranded RNA virus belonging to the *Phenuiviridae* family that was associated with a massive disease outbreak in the population of the endangered white-clawed crayfish *Austropotamobius pallipes* [24].

The aim of our study was to analyze the virome of the signal crayfish, *Pacifastacus leniusculus*, (Dana, 1852), one of the most successful freshwater invaders in Europe. The signal crayfish is present in over 29 EU countries [25], reaches very high dispersal rates within European watercourses [26], and is listed as IAS of Union concern (EU Regulation No. 1143/2014 on invasive alien species). We have investigated the signal crayfish population in a recently invaded Korana river, Croatia, where signal crayfish range expansion is well monitored [27–29], and occurs in both upstream and downstream directions [27]. Recently established populations at invasion fronts have 6–11 times lower crayfish abundance in comparison to longer established populations in invasion cores, and also contain co-occurring native congener (narrow-clawed crayfish, *Pontastacus leptodactylus*, Eschscholtz, 1823) [27]. Host population density, with increased contact rates among individuals in highly dense populations, is an important factor affecting rates of virus transmission [30,31].

In this study, we sampled the signal crayfish individuals at four locations along its invasion range, including upstream and downstream invasion cores and invasion fronts. We sequenced and analyzed signal crayfish samples to: (1) identify putative viral sequences in the signal crayfish hepatopancreas, as this is the tissue most often associated with viral infections in freshwater crayfish [20]; (2) describe the phylogenetic relationships of the most abundant novel signal crayfish-associated viruses; (3) discuss the potential differences in virome composition and the abundance of viral sequences along the invasion range. Our

results reveal a high, previously undiscovered diversity of viruses in signal crayfish, with putative signal crayfish-associated reo-like, hepe-like, toti-like, and picorna-like viruses, representing divergent sequences most similar to the viruses previously associated with crayfish hosts.

2. Materials and Methods

2.1. Study Area and Sample Collection

Research was carried out in the continental part of Croatia, in the Korana River. Here, the signal crayfish was illegally introduced in the lower section of the river, and is spreading both upstream and downstream [28], with its invasion range currently stretching along 33 km [27]. Significant differences in crayfish abundance occur along the signal crayfish invasion range, i.e., between the longer established populations in the center of its distribution (invasion core) compared to the recently established populations at distribution edges (invasion fronts). Invasion front locations have up to 11 times lower crayfish abundance compared to the invasion core [27]. Also, at invasion fronts, interspecific populations of the signal crayfish and the native narrow-clawed crayfish (*Pantastacus leptodactylus*, Eschscholtz, 1823) are present, with 2.5–12.9 higher abundances of the native *P. leptodactylus* compared to the signal crayfish [27]. Populations in the invasion cores comprise only signal crayfish, which have gradually displaced the native *P. leptodactylus* [27,29], as depicted in Figure 1.

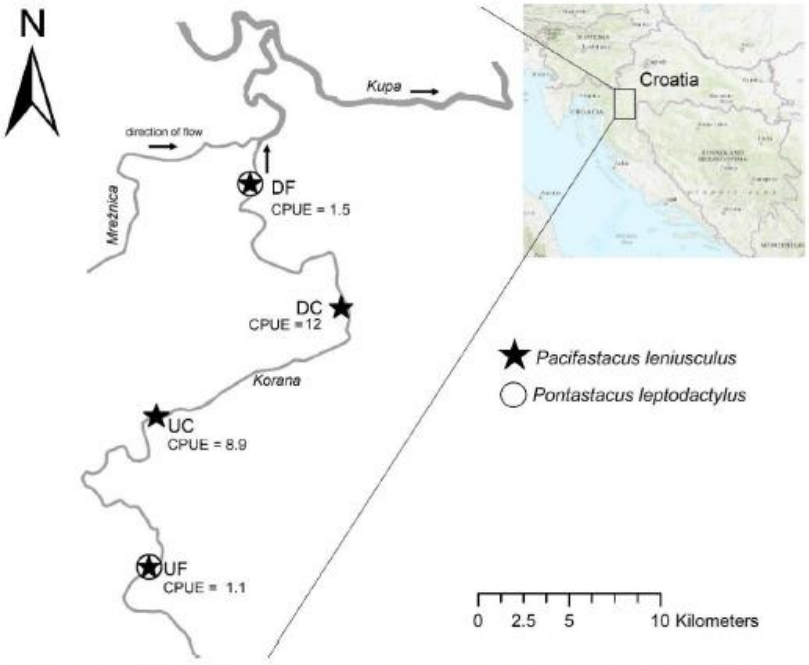


Figure 1. Position of sampling locations and differences in crayfish (*Pacifastacus leniusculus* and *Pontastacus leptodactylus*) presence and abundance (CPUE; catch per unit effort) along the signal crayfish invasion range in the Korana River in 2018. Sampling was performed at both upstream (UF) and downstream (DF) invasion fronts and upstream (UC) and downstream (DC) invasion cores.

Crayfish were collected in the period of increased crayfish activity (early autumn 2018) using baited LiNi traps [32]. A total of 110 signal crayfish individuals were captured from four invasion range endpoints (Figure 1), which were previously [27] identified as: the upstream invasion front (UF), upstream invasion core (UC), downstream invasion core (DC), and downstream invasion front (DF). For each location, we calculated the catch per unit effort (CPUE) that equals the number of crayfish captured per each trap per trapping night, and is a frequently used measure of relative crayfish abundance (Figure 1). The captured indigenous crayfish species were returned to the same location where they were caught, while signal crayfish individuals from all four sampling locations were taken on ice to the laboratory where hepatopancreas samples were randomly collected from 25 individuals from each location. For each individual, the complete organ was removed from the body, placed in a sterile Petri dish, and carefully chopped into smaller pieces using a sterile scalpel. The cut pieces of the organ were stored at −80 °C in an RNA stabilizing agent (RNA later; Sigma Aldrich, MO, USA) until RNA extraction.

2.2. RNA Extraction and Sequencing

The total RNA was extracted from hepatopancreas tissue using the RNeasy Lipid Tissue Mini Kit (Qiagen CA, USA). Before RNA extraction, equal amounts of tissue (30 mg) of each of the 25 individuals collected at the same location were pooled together, resulting in four composite pools corresponding to four sampling locations. After homogenizing tissue in liquid nitrogen, RNA extraction, followed by on-column DNase digestion step (RNase-Free DNase Set, Qiagen), was performed as recommended by the manufacturer. The eluted RNA from all four samples was sent for ribosomal RNA depletion (Illumina Ribo-Zero Plus rRNA Depletion Kit), sequencing library preparation (NEBNext Ultra RNA Library Prep Kit), and shotgun sequencing (Illumina HiSeqX, 2 × 150 bp) to CD Genomics, USA.

2.3. Bioinformatic Analysis for Detection of Viral Sequences

Sequencing datasets were quality checked, filtered, and analyzed as described below. The overview of the bioinformatic analyses performed is schematized in Supplementary material, Figure S1. The remains of the sequencing adaptors were trimmed, and resulting reads were filtered by using a quality filter (Limit = 0.01; no ambiguous nucleotide allowed) and by size (reads shorter than 25 bp were discarded) in CLC Genomics Workbench 20 (Qiagen Bioinformatics, CA, USA). Trimmed, size, and quality filtered reads were exported from the CLC Genomics Workbench and compared for similarity with a complete NCBI nr database (2020) using Diamond blastx [33] with default parameters, followed by the taxonomic classification of the reads and visualization of the results using Megan (Metagenome Analyzer, version 6.20.19, Tübingen, Germany) [34].

To identify longer virus-like sequences, reads were used for de novo assembly using SPAdes (version 3.14.0, St. Petersburg, Russia) [35], and the de novo constructed contigs were queried for similarity on protein level using Diamond blastx [33], followed by the taxonomic classification of the reads and visualization of the results using Megan [34]. The contig sequences that were classified as viral in all four analyzed samples were imported into the CLC Genomics Workbench, where additional steps of de novo assembly of viral contigs from all four samples was performed, resulting in a further assembly of overlapping sequences from different samples. The contigs that were not classified as viral by Diamond-Megan analysis were further analyzed using conserved domain search against Pfam (v32) database, and additional virus-like sequences detected using this analysis were added to the list of virus-like sequences from Diamond-Megan analysis.

The contigs that were identified as virus-like sequences according to the described analysis (Figure S1) were further investigated using blastn (NCBI-nt, April 2021) analysis to identify any possible misclassified sequences. After the additional manual inspection of all the similarity searches, we classified virus-like sequences into different categories (Table S2). Non-viral hits, such as host sequences misclassified as viral sequences (similar

to viruses from *Nudiviridae* family) and possible endogenized virus-like sequences (e.g., members of *Retroviridae* and *Adintoviridae* families), as well as viral sequences belonging to bacteriophages (e.g., members of *Caudovirales* order, *Leviviridae* family), plant infecting viruses (e.g., members of *Solemoviridae* and *Bromoviridae* families), and sequences identified as laboratory contamination (e.g., *Cryphonectria hypovirus*) were excluded from further analysis. The final selection of virus-like contigs included only sequences of potential invertebrate viruses longer than 300 nts. The differences between the abundance of reads corresponding to selected virus-like contigs detected in individual samples from different locations were identified by mapping the reads to virus-like contig sequences in CLC Genomics Workbench 20 (length fraction = 0.90, similarity fraction = 0.90), and the percentage of mapped reads was calculated (Table S1). In order to normalize samples according to the number of reads before mapping them to contig sequences, random subsampling was done and all samples were normalized to 97,563,944 reads, which was the minimum number of reads obtained among the samples. Decimal logarithm transformed average read coverage values were visualized as a heatmap, and used for clustering the locations and viruses using the heatmap.2 function in a gplots package in R 4.0.5 [36]

All sequencing data that support the findings of this study are linked under the accession number PRJNA754774 in the NCBI BioProject database. The sequences of selected viral contigs identified in this study are available in Table S2 and have been deposited in the GenBank under the accession numbers OK317706-OK317734.

2.4. PCR Detection of Selected Viral Contigs in Hepatopancreas Samples

Based on the presence of RNA-dependent RNA polymerase (RdRp) domain, contig length, and average read coverage, we selected four viral contigs for further analysis; virus-like contig 4 (signal crayfish-associated reo-like virus 1), virus-like contig 139 (signal crayfish-associated hepe-like virus 1), virus-like contig 1 (signal crayfish-associated picorna-like virus 1), and virus-like contig 141 (signal crayfish-associated toti-like virus 1). To confirm their presence in the samples, primer pairs (Table S3) amplifying the fragments of newly detected sequences were constructed using Primer-BLAST [37]. cDNA was synthesized from 2 µL of total RNA in a reaction volume of 20 µL by random priming using a High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, CA, USA) as described by the manufacturer, using the following reaction conditions: 25 °C for 10 min, 37 °C for 120 min, and 85 °C for 5 min. Target fragments were amplified in a total volume of 25 µL using 1 µL of cDNA template, 2 mM $MgCl_2$, 0.2 mM dNTP mixture, 0.5 µM each primer, and 0.0625 U of G2 polymerase (GoTaq Flexi DNA Polymerase, Promega, WI, USA) and following cycling conditions: 95 °C for 2 min, 40 cycles of 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 30 s; followed by 5 min of final extension at 72 °C. Amplicons were detected after electrophoresis in 1% agarose gel stained with SYBR Safe DNA Gel Stain (Invitrogen, CA, USA).

2.5. Phylogenetic Analyses

For selected virus-like contigs 4, 139, 1, and 141, the most similar homologous sequences with corresponding RdRp sequences were retrieved from the NCBI using blastp search against NCBI nr database (May 2021), and used in the phylogenetic analysis. The RdRp protein sequences of recognized viral species belonging to reo-like, hepe-like, picorna-like and toti-like virus groups selected according to previous studies [3] were additionally included in the phylogenetic analysis from the GenBank database. RdRp amino acid sequences were aligned using the L-INS-I algorithm of the MAFFT program [38], and alignments were trimmed manually and with TrimAI (automated 1 mode) [39]. Alignments were visualized in MEGAX [40]. Sequences included into the alignments with corresponding GenBank accession numbers are given in Alignments S1–S4. Maximum likelihood trees were inferred using PhyML (v. 3.00, Montpellier, France) [41] available in the Phylemon 2.0 web server [42] employing the Le Gascuel (LG + I + G) amino acid replacement model, selected by IQTree ModelFinder [43], using 1000 bootstrap replicates.

2.6. Analysis of Nucleotide Diversities

To compare the possible differences in the diversities of populations of selected viruses detected at all sampling locations, we determined the nucleotide diversities of populations of three signal crayfish-associated viruses, which were present at all sampling locations with an average read coverage higher than 40. For these analyses, read data was additionally filtered. First, reads with low nucleotide quality (<20) were discarded using a FASTQ Quality Filter (FASTX-toolkit). Reads from individual samples were then mapped to selected viral contigs (length fraction = 0.90, similarity fraction = 0.90). Before mapping, the read datasets were subsampled to amount for the same read coverage of individual contig sequence in different samples (higher read coverage was normalized by using a lower number of reads per sample). Single-nucleotide polymorphism (SNP) calling was performed using a Low Frequency Variant Detection tool in the CLC Genomics Workbench 20 (required significance = 1%, minimum coverage = 10, minimum count = 2, minimum frequency = 1%). SNP tables were used as an input for SNPGenie software [44], which was used to calculate diversity indices for each viral contig—location combination. Calculated nucleotide diversity (π) values, which represent the mean number of pairwise differences per nucleotide in a population of mapped reads, were plotted and compared according to the location.

3. Results and Discussion

3.1. Novel Viral Sequences Identified in Hepatopancreas of Signal Crayfish

3.1.1. Overview of the Newly Identified Viral Sequences in Hepatopancreas of Signal Crayfish

To identify viral sequences, we performed sequencing of ribosomal RNA depleted total RNA resulting in high numbers (100–164 M) of reads per sample (Table 1). First, in a reads-based protein similarity search, a small fraction of reads (at most, 0.018%) was classified as viral sequences (Table 1). Read sequences similar to different RNA viruses (classified as Riboviria) were the most abundant viral sequences in all four samples. Our analysis of de novo assembled contigs, presented in Table S2 together with blast resulting similarities and average read coverage values, supported the observed presence of sequences similar to different unclassified RNA viruses. Results revealed the presence of diverse virus-like sequences, with selected virus-like contigs (longer than 300 nts) representing sequences associated with invertebrate hosts shown in Table 2.

Table 1. Taxonomic classification of reads obtained from different locations (UF—upstream front, UC—upstream core, DC—downstream core, DF—downstream front).

Sample Name	Raw Reads	Reads after Trimming	Accession Number (SRA *)	% of Classified Reads (Diamond)	% of Reads Class. as Viral (Diamond)	% of Reads Class. as *Riboviria* (Diamond)	No. of Viral Read Class. as *Riboviria* (Diamond)
UF	164,043,698	160,643,914	SAMN20800941	15.47	0.00159	0.00094	1511
UC	149,357,758	146,166,391	SAMN20800942	19.06	0.00895	0.00816	11,929
DC	100,131,420	97,563,944	SAMN20800943	23.24	0.01737	0.01620	15,802
DF	136,134,398	132,904,626	SAMN20800944	15.29	0.00961	0.00821	10,905

* short reads archive (SRA).

Table 2. Selected putative invertebrate virus-like contigs (>300 nts) representing signal crayfish-associated viruses identified in this study with their contig length (nts) and closest match in the blastx (NCBI-nr) analysis.

Contig Name	Virus Name	GenBank Accesion	Contig Length	Blastx Results		
				Closest Protein Hit Name and Species	e-Value	Identities *
Virus-like contig 4	Signal crayfish associated reo-like virus 1	OK317706	4234	RdRp [Cherax quadricarinatus reovirus]	2.74×10^{-20}	106/422 (25%)
Virus-like contig 139	Signal crayfish associated hepe-like virus 1	OK317707	10,400	hypothetical protein [Beihai hepe-like virus 4]	2.64×10^{-117}	213/480 (44%)
Virus-like contig 141	Signal crayfish associated toti-like virus 1	OK317708	8576	hypothetical protein 4 [Wenzhou crab virus 5]	$<1.00 \times 10^{-250}$	353/869 (41%)
Virus-like contig 1	Signal crayfish associated picorna-like virus 1	OK317711	4587	hypothetical protein [Beihai picorna-like virus 99]	1.28×10^{-86}	324/1251 (26%)
Virus-like contig 10	Signal crayfish associated picorna-like virus 2	OK317712	531	hypothetical protein [Beihai picorna-like virus 99]	2.75×10^{-39}	79/175 (45%)
Virus-like contig 11	Signal crayfish associated picorna-like virus 3	OK317713	320	hypothetical protein [Beihai picorna-like virus 99]	4.97×10^{-51}	80/105 (76%)
Virus-like contig 15	Signal crayfish associated picorna-like virus 4	OK317714	898	hypothetical protein 1 [Changjiang picorna-like virus 6]	1.67×10^{-37}	85/151 (56%)
Virus-like contig 9	Signal crayfish associated picorna-like virus 5	OK317715	771	hypothetical protein [Beihai picorna-like virus 99] Y	1.30×10^{-96}	170/257 (66%)
Virus-like contig 5	Signal crayfish associated picorna-like virus 6	OK317716	714	hypothetical protein [Wenzhou picorna-like virus 38]	1.56×10^{-19}	62/152 (41%)
Virus-like contig 13	Signal crayfish associated picorna-like virus 7	OK317717	1894	hypothetical protein [Beihai sesarmid crab virus 2]	9.08×10^{-99}	226/642 (35%)
Virus-like contig 66	Signal crayfish associated tombus-like virus 1	OK317718	4504	replicase [Caledonia beadlet anemone tombus-like virus 1]	$<1.00 \times 10^{-250}$	352/862 (41%)
Virus-like contig 84	Signal crayfish associated tombus-like virus 2	OK317719	2981	replicase [Caledonia beadlet anemone tombus-like virus 1]	3.00×10^{-114}	246/628 (39%)
Virus-like contig 55	Signal crayfish associated tombus-like virus 3	OK317720	1425	RdRp [Riboviria sp.]	2.00×10^{-92}	150/299 (50%)
Virus-like contig 35	Signal crayfish associated tombus-like virus 4	OK317721	665	RdRp [Riboviria sp.]	4.00×10^{-32}	76/161 (47%)
Virus-like contig 24	Signal crayfish associated tombus-like virus 5	OK317722	538	hypothetical protein 1 [Hubei tombus-like virus 16]	3.98×10^{-04}	43/99 (43%)
Virus-like contig 169	Signal crayfish associated tombus-like virus 6	OK317723	301	hypothetical protein 2 [Hubei unio douglasiae virus 2]	2.00×10^{-42}	69/82 (84%)
Virus-like contig 140	Signal crayfish associated chu-like virus 1	OK317724	2216	hypothetical protein 2 [Beihai hermit crab virus 3]	5.71×10^{-56}	161/588 (27%)
Virus-like contig 65	Signal crayfish associated chu-like virus 2	OK317725	746	hypothetical protein 2 [Beihai hermit crab virus 3]	6.45×10^{-35}	69/167 (41%)
Virus-like contig 7	Signal crayfish associated chu-like virus 3	OK317726	493	RdRp [Beihai hermit crab virus 3]	3.98×10^{-48}	44/54 (81%)
Virus-like contig 145	Signal crayfish associated chu-like virus 4	OK317727	1009	RdRp [Beihai hermit crab virus 3]	$1.63E \times 10^{-60}$	120/219 (55%)
Virus-like contig 146	Signal crayfish associated chu-like virus 5	OK317728	418	RdRp [Beihai hermit crab virus 3]	$1.00E \times 10^{-27}$	54/98 (55%)
Virus-like contig 222	Signal crayfish associated chu-like virus 6	OK317729	356	RdRp [Beihai hermit crab virus 3]	$2.00E \times 10^{-45}$	80/111 (72%)
Virus-like contig 27	Signal crayfish associated partiti-like virus 1	OK317730	1188	RdRp [Caledonia partiti-like virus]	$1.00E \times 10^{-142}$	221/394 (56%)

Table 2. *Cont.*

Contig Name	Virus Name	GenBank Accesion	Contig Length	Blastx Results		
				Closest Protein Hit Name and Species	e-Value	Identities *
Virus-like contig 116	Signal crayfish associated picorna-like virus 8	OK317731	734	polyprotein [Picornaviridae sp.]	$3.70E \times 10^{-10}$	42/88 (48%)
Virus-like contig 2	Signal crayfish associated picorna-like virus 9	OK317732	412	hypothetical protein 1 [Picornavirales sp.]	$1.05E \times 10^{-24}$	45/100 (45%)
Virus-like contig 36	Signal crayfish associated picorna-like virus 10	OK317733	457	RdRp [Picornavirales sp.]	4.32×10^{-16}	60/102 (58%)
Virus-like contig 83	Signal crayfish associated sobemo-like virus 1	OK317734	431	hypothetical protein [Hubei sobemo-like virus 43]	1.00×10^{-50}	84/143 (59%)
Virus-like contig 147	Signal crayfish associated sobemo-like virus 2	OK317709	512	hypothetical protein 1 [Beihai sobemo-like virus 17]	$1.00E \times 10^{-10}$	44/128 (34%)
Virus-like contig 30	Signal crayfish associated narna-like virus 1	OK317710	378	RdRp [Beihai narna-like virus 18]	5.60×10^{-18}	48/128 (38%)

* number of matching amino acids/number of total amino acids of the closest protein hit (percentage of amino acid identity).

Sequences belonging to potential invertebrate viruses exhibited relatively low levels of similarity to known viral sequences, with sequence identities ranging from 25% to 84% according to blastx results. Many virus-like contigs representing novel, highly divergent RNA viral sequences were most similar to reo-like, hepe-like, toti-like, picorna-like, tombus-like, chu-like, and partiti-like viruses. A small fraction of viral reads and the dominance of previously unknown RNA viruses in aquatic arthropod hosts samples is in line with previous studies [3,5,45]. Many of the virus-like sequences identified in this study had the highest identities with unclassified RNA viruses discovered as part of a large-scale metagenomics study of invertebrate viromes [3], where similarities with viruses described in different crustacean hosts were also observed.

Some viral contig sequences (virus-like contigs 139, 141) discovered in this study likely correspond to full-length genomes, given that the length and genomic organization is similar to the annotated closest hits from the blastx analysis. High average read coverage values were obtained for some virus-like contigs (Figure 2), which may further support the assumption that they represent crayfish infecting viruses. However, confidently assigning host associations based on metagenomics studies and bioinformatic analyses remains a challenge [1], and therefore, some virus-like contigs identified here may be associated with other organisms that are present within the studied host [46].

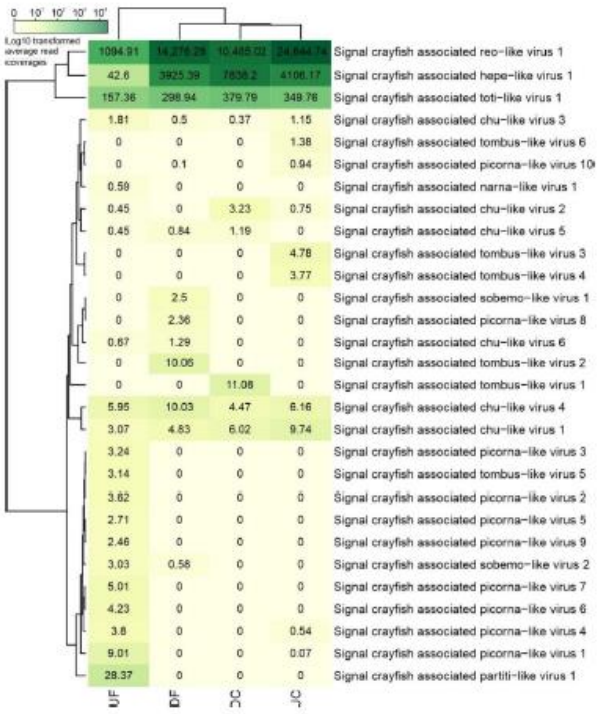


Figure 2. Heatmap showing average read coverage resulting from mapping of normalized read datasets from individual locations to selected putative invertebrate virus-like contigs together with the dendrograms clustering the sampling locations (top) and viral contigs (left) according to average read coverage values. Heatmap and dendrograms were constructed using decimal logarithm transformed values, however, non-transformed average read coverage numbers are plotted on the heatmap.

3.1.2. Phylogenetic Relationships and Genome Organization of Selected Newly Identified Viral Sequences

Based on the presence of the RdRp domain and contig length, we selected four distinct viral contigs for further analyses. They represent putative novel virus species belonging to reo-like (virus-like contig 4), hepe-like (virus-like contig 139), toti-like (virus-like contig 141), and picorna-like (virus-like contig 1) virus clades. These viral sequences showed similarities to viruses previously described in crustacean hosts, suggesting that they were likely infecting signal crayfish rather than being associated with the environment or co-infecting microorganisms. The presence of signal crayfish-associated reo-like virus 1, hepe-like virus 1, and toti-like virus 1 sequences was confirmed in all four analyzed samples by obtaining the products of the expected size after RT-PCR (Figure 3), while the presence of signal crayfish-associated picorna-like virus 1 was confirmed in the UF sample only, in line with the sequencing results (Figure 2).

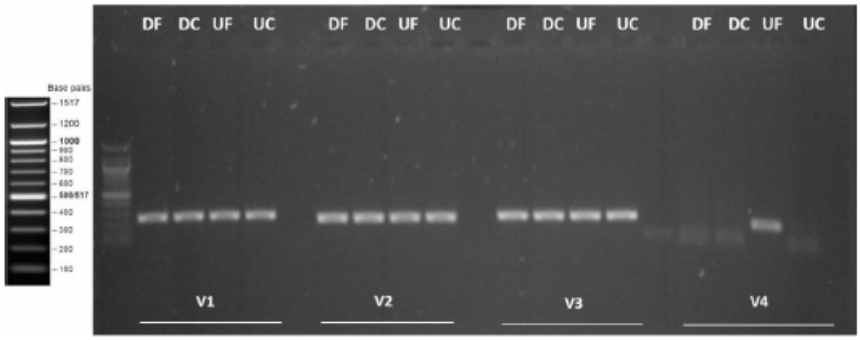


Figure 3. Gel electrophoresis of PCR amplicons originating from newly identified signal crayfish-associated viruses. Four primer pairs were used for amplifying app. 200 nts long regions of signal crayfish-associated reo-like virus 1 (V1), signal crayfish-associated hepe-like virus 1 (V2), signal crayfish-associated toti-like virus 1 (V3), and signal crayfish-associated picorna-like virus 1 (V4) genomes. Samples from different locations (UF—upstream front, UC—upstream core, DC—downstream core, DF—downstream front) were analyzed. Unmarked wells represent no template controls of each amplification reaction.

In our study, signal crayfish-associated reo-like virus 1 (virus-like contig 4, 4234 nts) had the highest average read coverage in all samples (Figure 2). A variety of reoviruses have been described previously in crustacean hosts [23]. Using blastx, we observed a distant similarity to the RdRp partial coding sequence of Cherax quadricarinatus reovirus, which was recently associated with hepatopancreatic samples in redclaw crayfish (*Cherax quadricarinatus*) from Australia [47]. This relationship was also confirmed by phylogenetic analysis (Figure 4). Cherax quadricarinatus reovirus causes the necrosis of hepatopancreocytes and inflammatory cells in hepatopancreatic tubules [48]. Notably, we have recorded a similar condition, the acute necrotizing hepatopancreatitis, along the whole invasion range of the signal crayfish in the Korana River, which is currently classified as idiopathic [49]. Thus, future studies should examine the recorded signal crayfish-associated reo-like virus 1 as a potential causative agent of this disease.

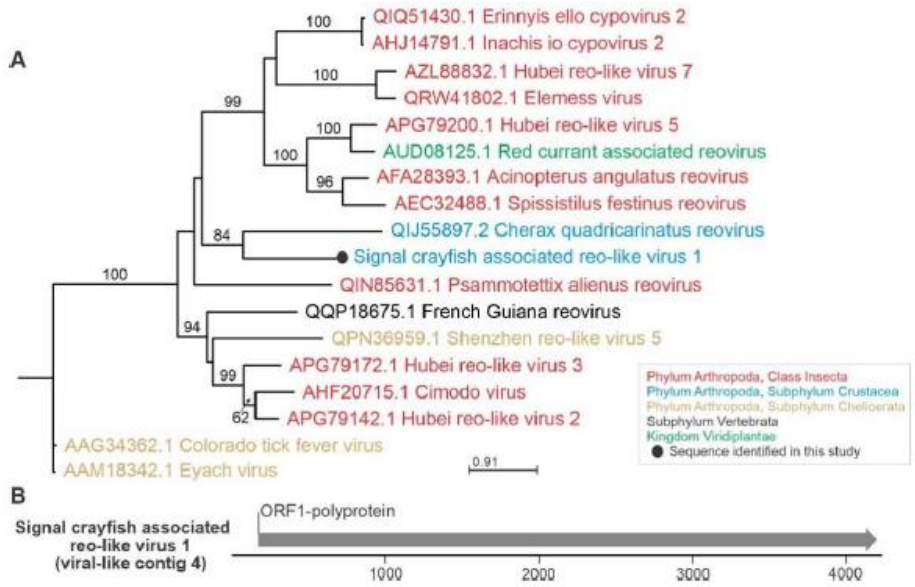


Figure 4. Phylogenetic relationships and genome organization of signal crayfish-associated reo-like virus 1. (**A**) Phylogenetic tree built with maximum likelihood approach, based on the alignment of the conserved segment of RNA-dependent RNA polymerase domain of signal crayfish-associated reo-like virus 1 (virus-like contig 4) and representative selected sequences of phylogenetically related viruses. The numbers on the branches represent bootstrap support values (>50% shown), the branch length represents the average number of amino acid substitutions per site (**B**) The predicted partial genome organization of the novel signal crayfish-associated reo-like virus 1, with positions and length of the open reading frames (ORF) indicated with corresponding arrow length.

Signal crayfish-associated hepe-like virus 1 (virus-like contig 139, 10,400 nts), detected in all samples, is represented by near complete genome sequence with four open reading frames (ORF), including ORF1—nonstructural polyprotein with RdRp domain (Figure 5). In blastx analysis, this contig had the highest (44%) amino acid identity to sequence of Beihai hepe-like virus 4, identified from the blue swimmer crab (deposited sequence is 11,635 nts long) and octopus (deposited sequence is 12,648 nts long) in China [3]. Consistent with the blastx search, phylogenetic analysis showed clustering of the detected virus sequence with Beihai hepe-like virus 4 and other hepe-like viruses from different crustacean hosts (Figure 5). Aside from reports in large-scale metagenomics studies [3], novel hepe-like virus has recently been characterized from the cephalothoraxes of the economically important giant freshwater prawn *Macrobrachium rosenbergii* with growth retardation, however, it could not be determined as the causative agent of the disease [50].

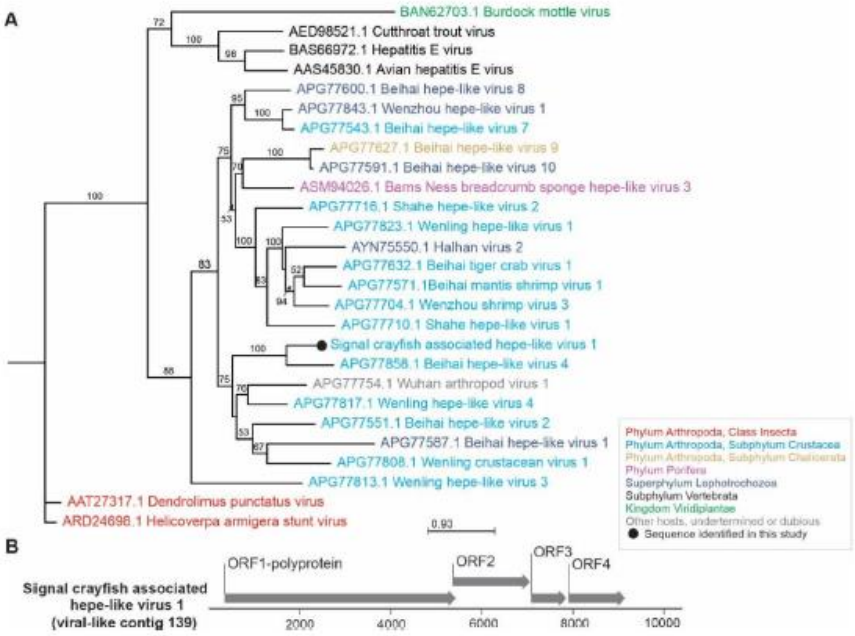


Figure 5. Phylogenetic relationships and genome organization of signal crayfish-associated hepe-like virus 1. (**A**) Phylogenetic tree built with maximum likelihood approach, based on the alignment of the conserved segment of RNA-dependent RNA polymerase domain of signal crayfish-associated hepe-like virus 1 (virus-like contig 139) and representative selected sequences of phylogenetically related viruses. The numbers on the branches represent bootstrap support values (>50% shown), the branch length represents the average number of amino acid substitutions per site. (**B**) The predicted genome organization of the novel signal crayfish-associated hepe-like virus 1, with positions and lengths of the open reading frames (ORF) indicated with corresponding arrow lengths.

The signal crayfish-associated toti-like virus 1 (virus-like contig 141, 8576 nts), containing four ORFs, was clustered with the Wenzhou crab virus 5 (deposited sequence is 8691 nts long) in a phylogenetic analysis (Figure 6). It had a 41% amino acid identity in blastx analysis with the same virus, which was associated with the *Charybdis* crab host [3] and was detected in all of our samples. Similarities to other viruses, especially toti-like viruses, from different insect hosts were also detected using blastx. Other viruses from toti-like virus clade, with no blastx similarity to sequence identified in this study, and not included in phylogenetic analysis, are known to be associated with crayfish; infectious myonecrosis virus (IMNV), the causative agent of infectious myonecrosis in the Pacific white shrimp (*Litopenaeus vannamei*) [51], and the Cherax giardiavirus-like virus (CGV) in freshwater crayfish (*Cherax quadricarinatus*) [52]. However, IMNV targets mesoderm-derived tissues, like muscles, connective tissue, and hemocytes, and does not replicate in enteric tissues like hepatopancreas [53]. CGV was, however, reported by histology in hepatopancreas, similar to the toti-like virus identified in this study, but it has been neither sequenced nor isolated [52,53].

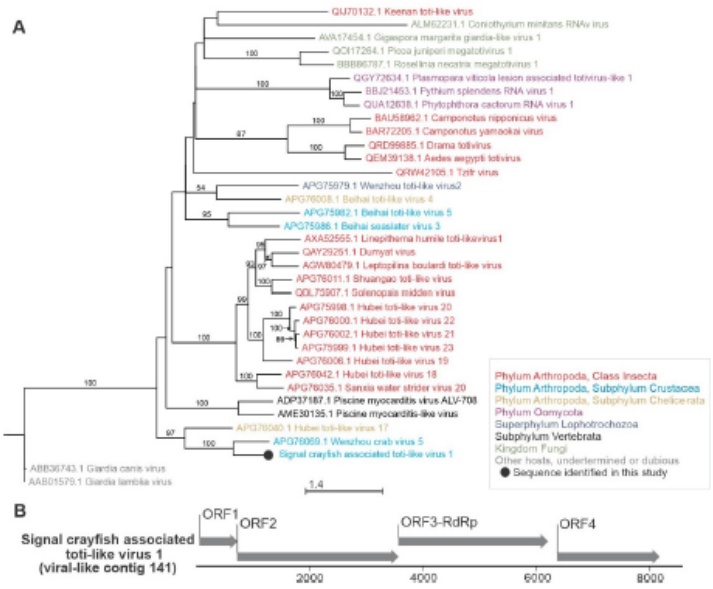


Figure 6. Phylogenetic relationships and genome organization of signal crayfish-associated toti-like virus 1. (**A**) Phylogenetic tree built with maximum likelihood approach, based on the alignment of the conserved segment of RNA-dependent RNA polymerase domain of signal crayfish-associated toti-like virus 1 (virus-like contig 141) and representative selected sequences of phylogenetically related viruses. The numbers on the branches represent bootstrap support values (>50% shown), the branch length represents the average number of amino acid substitutions per site. (**B**) The predicted genome organization of the novel signal crayfish-associated toti-like virus 1 with positions and lengths of the open reading frames (ORF) indicated with corresponding arrow lengths.

Distinct contigs similar to previously described picorna-like viruses, a loosely defined broad group of viruses, were detected mostly in UF, with the highest blastx similarities to different crustacean picorna-like viruses (Table 2), such as Changjiang picorna-like virus 6, Beihai picorna-like virus 99, and Wenzhou picorna-like virus 38 [3]. The longest (4587 nts) picorna-like virus contig (virus-like contig 1, signal crayfish-associated picorna-like virus 1) with the highest read coverage had 26% amino acid identity in the blastx analysis with the Beihai picorna-like virus 99 (deposited sequence is 6503 nts long) isolated from a hermit crab in China [3]. Based on phylogenetic analysis and genome organization (Figure 7), with ORF1 nonstructural polyprotein comprising RdRp domain, signal crayfish-associated picorna-like virus 1 represents a single-sequence branch placed in the cluster that contains viruses of different invertebrate hosts, including crustacean viruses. In the same cluster, Taura syndrome virus is a known disease-causing crustacean virus that has spread rapidly throughout shrimp farming regions globally [23], and was reported to induce significant changes in the hepatopancreas transcriptome of the Pacific white shrimp [54]. Besides most similar viruses included in our phylogenetic analysis, mud crab dicistrovirus has been described previously as pathogenic to the mud crab (*Scylla paramamosain*) [55]. Other sequences distantly similar to picorna-like viruses (virus-like contigs 10, 11, 15, 9, 5, 13, 116, 2, 36) were also identified with low average read coverage, and could not be further assembled into longer contigs.

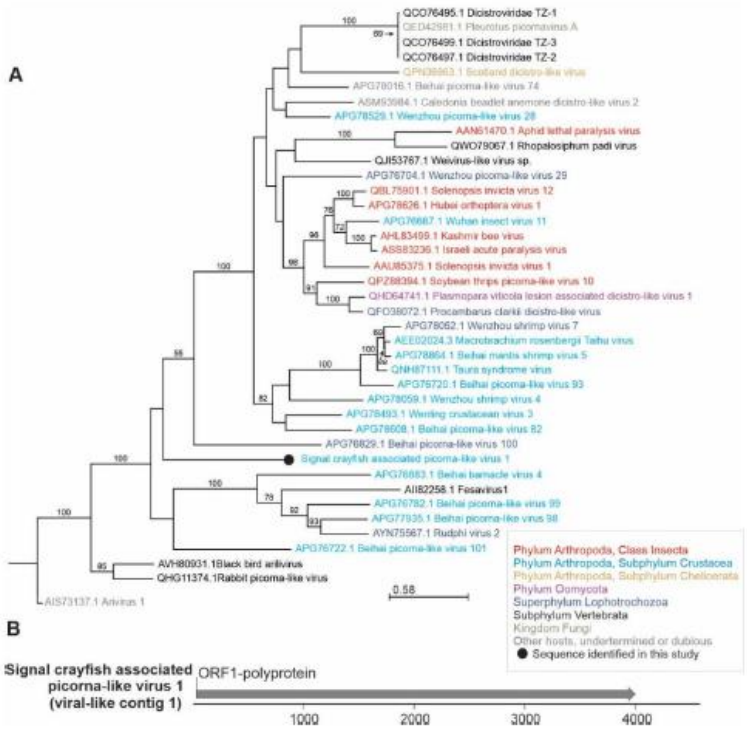


Figure 7. Phylogenetic relationships and genome organization of signal crayfish-associated picorna-like virus 1. (**A**) Phylogenetic tree built with maximum likelihood approach, based on the alignment of the conserved segment of RNA-dependent RNA polymerase domain of signal crayfish-associated picorna-like virus 1 (virus-like contig 1) and representative selected sequences of phylogenetically related viruses. The numbers on the branches represent bootstrap support values (>50% shown), the branch length represents the average number of amino acid substitutions per site. (**B**) The predicted partial genome organization of the novel signal crayfish-associated picorna-like virus 1, with positions and length of the open reading frame (ORF) indicated with corresponding arrow length.

Further viral contigs potentially associated with signal crayfish, often detected in only one sample and with low average read coverage, belonged to tombus-like, chu-like, partiti-like, sobemo-like, and narna-like virus groups. Sequences similar to tombus-like viruses were previously associated with plant hosts, however, numerous tombus-related viruses have recently been described in non-plant hosts, such as marine and terrestrial invertebrates. Based on the phylogenetic analyses, it was suggested that this group of viruses was primarily associated with invertebrates, and was later horizontally transferred from aquatic invertebrates to plant hosts [3,6]. Apart from HTS based studies, tombus-like viruses have not been widely studied as infectious agents in crustacean hosts, except for the mud crab tombus-like virus that was associated with the sleeping disease in crabs, but only as a part of co-infection with mud crab dicistrovirus [56]. In this study, distinct tombus-like virus sequences were detected in signal crayfish hepatopancreas. Signal crayfish-associated tombus-like virus 1 (virus-like contig 66, 4504 nts) with a near complete genome sequence found in the DC location was most similar to the Caledonia beadlet anemone tombus-like virus 1 (41% identity on amino acid level) isolated from the anemone

species *Actinia equine* [57]. Signal crayfish-associated tombus-like virus 2 (virus-like contig 84, 2981 nts) was detected in the DF location and was also most similar to Caledonia beadlet anemone tombus-like virus 1 (39% identity on amino acid level) [57]. These two tombus-like virus contigs had an average read coverage higher than ten, in contrast to lower read coverage of additional tombus-like virus contigs (virus-like contigs 55, 35, 169) (1425, 665, 301 nts respectively) detected in the UC location with the highest blastx-based identities with different invertebrate and plant infecting tombus-like viruses. Low read coverage tombus-like contigs may represent plant-infecting viruses, derived from food sources, while those with higher read coverage are more likely to be true crayfish viruses.

At all locations, different contigs similar to the Beihai hermit crab virus 3 [3] were identified. Virus-like contig 140 (signal crayfish-associated chu-like virus 1, 2216 nts) was the longest, and had a 27% amino acid sequence identity with hypothetical protein 2. Virus-like contig 145 (signal crayfish-associated chu-like virus 4, 1009 bp) had a 30% amino acid sequence identity to RdRp. The Beihai hermit crab virus 3 belongs to the *Mivirus* genus from the recently discovered negative-strand RNA *Chuviridae* family, characterized by HTS analysis only [3,31,58]. Diverse chuvirus-derived endogenous viral elements were also detected in mosquito genomes [59]. Additional contigs (virus-like contigs 65, 7, 146, 222) exhibited similarity with the Beihai hermit crab virus 3 hypothetical protein 2 or RdRp, but could not be further assembled into longer contigs, and did not contain full-length RdRp domain. In the UF location we have also detected a signal crayfish-associated partiti-like virus 1 (virus-like contig 27, 1188 nts) with a 56% amino acid identity to Caledonia partiti-like virus isolated from the breadcrumb sponge (*Halichondria panacea*) [57]. According to the blastx results, signal crayfish-associated partiti-like virus 1 was also similar to other crustacean-associated viruses (Wenling partiti-like virus 13, Wenling partiti-like virus 11) [3] and may with relatively high average read coverage (28) represent a crayfish-infecting virus.

3.2. Abundance of Viral Sequences and Nucleotide Diversity of Selected Virus Populations along the Invasion Range

We have compared the abundance of viral sequences as well as the nucleotide diversity of selected virus populations at different sampling locations in order to obtain insights into the possible patterns of variations along the invasion range. Based on the observed patterns, we could postulate hypotheses about how the observed patterns could be connected with ecological conditions at different sample sites, which represent a baseline for future research work on the virus ecology of crayfish and/or invasive alien species.

In our study, an abundance of virus-like reads, distinct virus-like contigs, and sequences mapping to signal crayfish-associated virus-like contigs varied between locations along the invasion range, with virus-like reads constituting between 0.0016% of the total reads in UF and 0.0174% in DC (Table 1). When comparing the two upstream and two downstream samples, the ones at the invasion cores had higher numbers of virus-like reads detected in comparison to those from the corresponding front sites. In addition, higher read coverage of viral sequences belonging to two out of three of the most represented newly identified signal crayfish-associated viruses was observed in the invasion cores compared to invasion fronts (Figure 2). At the same time, higher crayfish abundances were observed at the invasion cores (Figure 1).

The differences in signal crayfish abundance along the invasion range could influence the abundance of virus sequences detected, according to the classical epidemiological theory that links larger populations with higher contact rates with an increased likelihood of viral transmission [60–62]. The UF location differed the most from other locations in several aspects. The lowest percentage of virus-like reads was detected, and also lower reads coverages for signal crayfish-associated reo-like virus 1 (virus-like contig 4), crayfish-associated hepe-like virus 1 (virus-like contig 139), and toti-like virus 1 (virus-like contig 141) were found for this locality (Figure 2). Also, the lowest crayfish abundance was estimated at this locality, which probably resulted in lower encounter rates between individuals. Additionally, in the UF, we have recorded a lower severity of the acute necrotizing hepatopancreatitis [49]. This correlates with above mentioned lower read

abundance at this location, and implies that some of the detected putative novel viruses in this study may be associated with the observed pathological changes in hepatopancreas. To test this hypothesis, further studies would be needed to associate individual crayfish samples of a known health status with the presence of putative novel viruses discovered in this study.

Despite the lower total number of virus-like reads in the UF sample, the highest number of distinct putative virus-like contigs similar to unclassified RNA viruses (Table S2) was found here, and picorna-like virus contigs were found predominantly in this sample. According to previous research [63–65], a decrease in viral diversification is expected in less dense groups, such as invasion fronts. In contrast, we observed the highest number of distinct viral contigs in the UF location, a site with the lowest crayfish abundance. However, at invasion fronts, the population of the native *P. leptodactylus* is also present with 2.5–12.9 higher abundances than the signal crayfish [27], potentially contributing to the inter-specific transmission of viruses. The observed differences in the number of total virus-like reads between the samples could also partially result from methodological limitations, such as difficulties in detecting highly divergent viral sequences that are considerably different from the reference genomes of known viruses [66].

For the three viruses, which were present at all four locations with relatively high average read coverage, we determined the diversities of their populations by calculating nucleotide diversities. This enabled us to estimate how the diversity of each discovered viral species population might be changing along its host invasion range. For all analyzed viruses, we observed differences in nucleotide diversity between populations at the invasion cores and invasion fronts. We observed two contrasting patterns. For signal crayfish-associated reo-like virus 1, the nucleotide diversity was decreased at the invasion fronts, while for signal crayfish-associated hepe-like virus 1 and toti-like virus 1, the nucleotide diversity was higher at the invasion fronts (Figure 8).

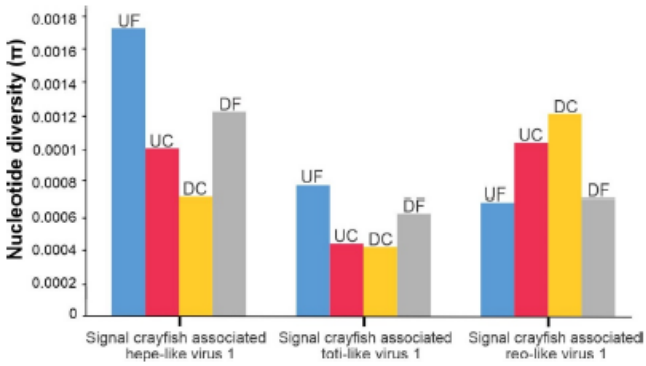


Figure 8. Nucleotide diversity values (π) calculated for signal crayfish-associated hepe-like virus 1, signal crayfish-associated toti-like virus 1, and signal crayfish-associated reo-like virus 1 at different locations (UF—upstream front, UC—upstream core, DC—downstream core, DF—downstream front) along the invasion range of the signal crayfish in the Korana River, Croatia.

Observed differences in nucleotide diversities of viral populations of the three investigated viruses from samples collected along the signal crayfish invasion range might result from different processes, and several hypotheses can be put forward to explain the observed patterns. For instance, higher nucleotide diversity might indicate a higher prevalence of a particular virus, e.g., reo-like virus in the core sites. However, since we were analyzing composite hepatopancreas samples, we are limited in our estimations of the prevalence of individual viruses. Furthermore, we speculate that the estimated nucleotide diversity (π) values potentially reflect transmission dynamics of analyzed viruses, with a

higher core nucleotide diversity of signal crayfish-associated reo-like virus 1 suggesting the transmission of the virus from longer established core populations with high signal crayfish abundance to recently established front populations of low signal crayfish abundance. The signal crayfish-associated reo-like virus 1 might be an endemic virus of signal crayfish, introduced to the new habitat along with its host, which could be confirmed/rejected by detecting the signal crayfish-associated reo-like virus 1 sequence in signal crayfish from other geographic regions. On the contrary, a higher nucleotide diversity of signal crayfish-associated toti-like virus 1 and hepe-like virus 1 in samples from invasion fronts could potentially reflect the introduction of these viruses from other host populations (e.g., from the native *P. leptodactylus* with higher density at the fronts) to populations of low signal crayfish abundance at invasion fronts. Successful inter-specific transmission relies on physical opportunities for transmission, such as occupying the same environment at the same time [67]. At invasion cores, the signal crayfish have displaced the native host, which would, under the proposed scenario, explain the decrease in nucleotide diversity for signal crayfish-associated toti-like virus 1 and hepe-like virus 1. However, such host associations should be further investigated, since very few of the crustacean viruses have been tested for infectivity in species other than the original host [68]; this hypothesis could be tested by analyzing native crayfish species for the presence of signal crayfish-associated toti-like virus 1 and hepe-like virus 1. Another possible hypothesis for the observed patterns might be a non-random dispersal of the host individuals during the invasion process, as demonstrated for a number of invasive species (i.e., [69]).

Finally, the results regarding the comparison of different locations can be confounded by other factors, which we were not able to elucidate using our experimental design (sequencing of samples pooled by location), and further analysis of individual samples would be needed to advance the understanding and test the postulated hypotheses.

4. Conclusions

- We reported, for the first time, the virome of signal crayfish hepatopancreas tissue and found a high diversity of novel divergent viral sequences most similar to different unclassified RNA viruses.
- We identified putative novel RNA viruses, including near complete genome sequence of signal crayfish-associated hepe-like virus 1 and toti-like virus 1, and the partial genomes of signal crayfish-associated reo-like virus 1 and picorna-like viruses. We identified additional tombus-like, partiti-like, and chu-like virus sequences potentially representing novel crayfish viruses. This pioneer study represents a baseline for the future research of a signal crayfish virome, e.g., to confirm the association of novel viruses with signal crayfish host, and to investigate their potential involvement in the observed necrotizing hepatopancreatitis.
- We speculate that the differences in the signal crayfish population density along the invasion range, non-random dispersal, and possibilities of inter-specific viral transmissions may have an effect on the diversity and abundance of signal crayfish-associated viral sequences. Different hypotheses can be postulated to explain these patterns, and this study represents a baseline for the further research of virus transmission dynamics as a result of the invader's fast dispersal, including inter-species transmission between the signal crayfish as an invader and *P. leptodactylus* as a co-occurring and phylogenetically related native species.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/v13112259/s1, Figure S1. Schematic representation of different steps of the bioinformatic analysis of sequencing reads (black arrows) and contigs (green arrows) to identify virus-like sequences in the signal crayfish hepatopancreas samples; Table S1. Selected putative invertebrate virus-like contigs (>300 nts) representing signal crayfish associated viruses identified in this study with their contig length (nts) and percentage of mapped reads resulting from mapping of reads from individual samples to virus-like contigs; Table S2. (A) Virus-like contigs from different locations (UF—upstream

front, UC—upstream core, DC—downstream core, DF—downstream front) identified in Diamond analysis together with average coverage values, Diamond classification, blastn similarity search results (NCBI-nt, April 2021) and sequences of individual virus-like contigs. (B) virus-like contigs from different locations (UF—upstream front, UC—upstream core, DC—downstream core, DF—downstream front) identified using pfam domain search, where the remaining contigs not classified as viral by Diamond (963721) were translated and compared with the entire Pfam database, together with average coverage values, Pfam search results and blastn (NCBI-nt, April 2021) and blastx (NCBI-nr, June 2021) similarity search results; Table S3. Sequences of primers used for PCR amplification of selected virus-like contigs representing signal crayfish associated viruses identified in this stud; Alignment S1. Sequence alignments (.fasta) with virus names and corresponding GenBank accession numbers of viruses used for phylogenetic analysis of signal crayfish associated reo-like virus 1; Alignment S2. Sequence alignments (.fasta) with virus names and corresponding GenBank accession numbers of viruses used for phylogenetic analysis of signal crayfish associated hepe-like virus 1; Alignment S3. Sequence alignments (.fasta) with virus names and corresponding GenBank accession numbers of viruses used for phylogenetic analysis of signal crayfish associated toti-like virus 1; Alignment S4. Sequence alignments (.fasta) with virus names and corresponding GenBank accession numbers of viruses used for phylogenetic analysis of signal crayfish associated picorna-like virus 1.

Author Contributions: Conceptualization, S.H. and S.Č.; methodology, D.K., K.B., S.Č. and S.H.; formal analysis, K.B.; investigation, S.H., A.B. and S.Č.; data curation, K.B.; writing, K.B. and S.H.; original draft preparation, K.B. and S.H.; writing—review and editing, K.B., A.B., D.K., S.H. and S.Č.; visualization, K.B., D.K., S.H. and S.Č.; supervision, D.K. and S.H.; project administration, S.H. and D.K.; funding acquisition, S.H. and D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the Croatian Science Foundation grant (HRZZ UIP 2017-05-1720) and Slovenian research agency (research core funding No P4-0407 and program for young researchers awarded to K.B.).

Institutional Review Board Statement: Ethical approval was not required for this study since no institutional, national, or EU ethical guidelines exist for crayfish. However, the dissection procedure (rapid cut of individuals' nerve cord from the thorax to the end of abdomen) was performed according to available guidelines for the humane killing of crayfish [70].

Informed Consent Statement: Not applicable.

Data Availability Statement: All sequencing data that support the findings of this study are linked under the accession number PRJNA754774 in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/sra, accessed on 9 November 2021). The sequences of selected viral contigs identified in this study are available in Table S2 and have been deposited in the GenBank under the accession numbers OK317706-OK317734.

Acknowledgments: We thank Paula Dragičević, Ivana Maguire, Dora Pavić, and Leona Lovrenčić for their help in crayfish sampling, and to Paula Dragičević for her help with RNA extraction.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, Y.Z.; Shi, M.; Holmes, E.C. Using Metagenomics to Characterize an Expanding Virosphere. *Cell* **2018**, *172*, 1168–1172. [CrossRef]
2. Li, C.X.; Shi, M.; Tian, J.H.; Lin, X.D.; Kang, Y.J.; Chen, L.J.; Qin, X.C.; Xu, J.; Holmes, E.C.; Zhang, Y.Z. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife* **2015**, *2015*, 1–26. [CrossRef]
3. Shi, M.; Lin, X.D.; Tian, J.H.; Chen, L.J.; Chen, X.; Li, C.X.; Qin, X.C.; Li, J.; Cao, J.P.; Eden, J.S.; et al. Redefining the invertebrate RNA virosphere. *Nature* **2016**, *540*, 539–543. [CrossRef]
4. Wu, H.; Pang, R.; Cheng, T.; Xue, L.; Zeng, H.; Lei, T.; Chen, M.; Wu, S.; Ding, Y.; Zhang, J.; et al. Abundant and Diverse RNA Viruses in Insects Revealed by RNA-Seq Analysis: Ecological and Evolutionary Implications. *Msystems* **2020**, *5*, e00039-20. [CrossRef] [PubMed]
5. Chang, T.; Hirai, J.; Hunt, B.P.V.; Suttle, C.A. Arthropods and the evolution of RNA viruses. *bioRxiv* **2021**. [CrossRef]
6. Dolja, V.V.; Koonin, E.V. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res.* **2018**, *244*, 36–52. [CrossRef]

7. Zhang, Z.-Q. Phylum Arthropoda. In *Animal Biodiversity: An Outline of Higher-Level Classification and Survey of Taxonomic Richness*; Zootaxa: Auckland, New Zealand, 2013; Volume 3703, p. 17.
8. Reynolds, J.; Souty-Grosset, C.; Richardson, A. Ecological roles of crayfish in freshwater and terrestrial habitats. *Freshw. Crayfish* **2013**, *19*, 197–218. [CrossRef]
9. Holdich, D.M.; Reynolds, J.D.; Souty-Grosset, C.; Sibley, P.J. A review of the ever increasing threat to European crayfish from non-indigenous crayfish species. *Knowl. Manag. Aquat. Ecosyst.* **2009**, *2009*, 394–395. [CrossRef]
10. Lodge, D.M.; Deines, A.; Gherardi, F.; Yeo, D.C.J.; Arcella, T.; Baldridge, A.K.; Barnes, M.A.; Lindsay Chadderton, W.; Feder, J.L.; Gantz, C.A.; et al. Global introductions of crayfishes: Evaluating the impact of species invasions on ecosystem services. *Annu. Rev. Ecol. Evol. Syst.* **2012**, *43*, 449–472. [CrossRef]
11. Twardochleb, L.A.; Olden, J.D.; Larson, E.R. A global meta-analysis of the ecological impacts of nonnative crayfish. *Freshw. Sci.* **2013**, *32*, 1367–1382. [CrossRef]
12. Souty-Grosset, C.; Grandjean, F.; Gouin, N. Conservation and management of native crayfish populations. *Freshw. Crayfish* **2004**, *14*, 1–20.
13. Richman, N.I.; Böhm, M.; Adams, S.B.; Alvarez, F.; Bergey, E.A.; Bunn, J.J.S.; Burnham, Q.; Cordeiro, J.; Coughran, J.; Crandall, K.A.; et al. Multiple drivers of decline in the global status of freshwater crayfish (Decapoda: Astacidea). *Philos. Trans. R. Soc. B Biol. Sci.* **2015**, *370*, 1–11. [CrossRef] [PubMed]
14. Crowl, T.A.; Crist, T.O.; Parmenter, R.R.; Belovsky, G.; Lugo, A.E. The spread of invasive species and infectious disease as drivers of ecosystem change. *Front. Ecol. Environ.* **2008**, *6*, 238–246. [CrossRef]
15. Van der Veer, G.; Nentwig, W. Environmental and economic impact assessment of alien and invasive fish species in Europe using the generic impact scoring system. *Ecol. Freshw. Fish* **2015**, *24*, 646–656. [CrossRef]
16. Unestam, T.; Weiss, D.W. The host-parasite relationship between freshwater crayfish and the crayfish disease fungus Aphanomyces astaci: Responses to infection by a susceptible and a resistant species. *J. Gen. Microbiol.* **1970**, *60*, 77–90. [CrossRef] [PubMed]
17. Martín-Torrijos, L.; Martínez-Ríos, M.; Casabella-Herrero, G.; Adams, S.B.; Jackson, C.R.; Diéguez-Uribeondo, J. Tracing the origin of the crayfish plague pathogen, Aphanomyces astaci, to the Southeastern United States. *Sci. Rep.* **2021**, *11*, 1–11. [CrossRef]
18. Zhang, Q.Y.; Gui, J.F. Virus genomes and virus-host interactions in aquaculture animals. *Sci. China Life Sci.* **2015**, *58*, 156–169. [CrossRef]
19. Geoghegan, J.L.; Di Giallonardo, F.; Cousins, K.; Shi, M.; Williamson, J.E.; Holmes, E.C. Hidden diversity and evolution of viruses in market fish. *Virus Evol.* **2018**, *4*, 1–11. [CrossRef]
20. Dragičević, P.; Bielen, A.; Petrić, I.; Hudina, S. Microbial pathogens of freshwater crayfish: A critical review and systematization of the existing data with directions for future research. *J. Fish Dis.* **2020**, *44*, 221–247. [CrossRef]
21. Sánchez-Paz, A. White spot syndrome virus: An overview on an emergent concern. *Vet. Res.* **2010**, *41*, 43. [CrossRef]
22. Longshaw, M. Parasites, Commensals, Pathogens and Diseases of Crayfish. In *Biology and Ecology of Crayfish*; Stebbing, M., Longshaw, P., Eds.; Taylor & Francis Group, LLC: Boca Raton, FL, USA, 2016; pp. 171–250.
23. Bateman, K.S.; Stentiford, G.D. A taxonomic review of viruses infecting crustaceans with an emphasis on wild hosts. *J. Invertebr. Pathol.* **2017**, *147*, 86–110. [CrossRef]
24. Grandjean, F.; Gilbert, C.; Razafimafondy, F.; Vucić, M.; Delaunay, C.; Gindre, P.; Bouchard, J.; Raimond, M.; Moumen, B. A new bunya-like virus associated with mass mortality of white-clawed crayfish in the wild. *Virology* **2019**, *533*, 115–124. [CrossRef] [PubMed]
25. Kouba, A.; Petrusek, A.; Kozák, P. Continental-wide distribution of crayfish species in Europe: Update and maps. *Knowl. Manag. Aquat. Ecosyst.* **2014**, *5*, 413. [CrossRef]
26. Hudina, S.; Faller, M.; Lucić, A.; Klobučar, G.; Maguire, I. Distribution and dispersal of two invasive crayfish species in the Drava River basin, Croatia. *Knowl. Manag. Aquat. Ecosyst.* **2009**, *9*, 394–395. [CrossRef]
27. Dragičević, P.; Faller, M.; Kutleša, P.; Hudina, S. Update on the signal crayfish, pacifastacus leniusculus (Dana, 1852) range expansion in Croatia: A 10-year report. *BioInvasions Rec.* **2020**, *9*, 793–807. [CrossRef]
28. Hudina, S.; Žganec, K.; Lucić, A.; Trgovčić, K.; Maguire, I. Recent invasion of the karstic river systems in croatia through illegal introductions of the signal crayfish. *Freshw. Crayfish* **2013**, *19*, 21–27. [CrossRef]
29. Hudina, S.; Kutleša, P.; Trgovčić, K.; Duplić, A. Dynamics of range expansion of the signal crayfish (Pacifastacus leniusculus) in a recently invaded region in Croatia. *Aquat. Invasions* **2017**, *12*, 67–75. [CrossRef]
30. Anderson, R.M.; May, R.M. Coevolution of Hosts and Parasites. *Parasitology* **1982**, *85*, 411–426. [CrossRef]
31. Costa, V.A.; Mifsud, J.C.O.; Gilligan, D.; Williamson, J.E.; Holmes, E.C.; Geoghegan, J.L. Metagenomic sequencing reveals a lack of virus exchange between native and invasive freshwater fish across the Murray-Darling Basin, Australia. *Virus Evol.* **2021**, *7*, 1–15. [CrossRef] [PubMed]
32. Westman, K. A new folding trap model which prevents crayfish from escaping. *Freshw. Crayfish* **1978**, *4*, 235–242.
33. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [CrossRef] [PubMed]
34. Huson, D.H.; Beier, S.; Flade, I.; Gorska, A.; El-Hadidi, M.; Mitra, S.; Ruscheweyh, H.J.; Tappu, R. MEGAN Community Edition—Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput. Biol.* **2016**, *12*, e1004957. [CrossRef]

35. 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.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]
36. R Core Team. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2021. Available online: https://www.R-project.org (accessed on 9 November 2021).
37. Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T.L. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform.* **2012**, *13*, 134. [CrossRef] [PubMed]
38. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
39. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef]
40. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
41. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [CrossRef]
42. Sánchez, R.; Serra, F.; Tárraga, J.; Medina, I.; Carbonell, J.; Pulido, L.; De María, A.; Capella-Gutíerrez, S.; Huerta-Cepas, J.; Gabaldón, T.; et al. Phylemon 2.0: A suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Res.* **2011**, *39*, 470–474. [CrossRef]
43. Nguyen, L.T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef]
44. Nelson, C.W.; Moncla, L.H.; Hughes, A.L. SNPGenie: Estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics* **2015**, *31*, 3709–3711. [CrossRef] [PubMed]
45. Wolf, Y.I.; Silas, S.; Wang, Y.; Wu, S.; Bocek, M.; Kazlauskas, D.; Krupovic, M.; Fire, A.; Dolja, V.V.; Koonin, E.V. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nat. Microbiol.* **2020**, *5*, 1262–1270. [CrossRef] [PubMed]
46. Xu, C.; Sun, X.; Taylor, A.; Jiao, C.; Xu, Y.; Cai, X.; Wang, X.; Ge, C.; Pan, G.; Wang, Q.; et al. Diversity, Distribution, and Evolution of Tomato Viruses in China Uncovered by Small RNA Sequencing. *J. Virol.* **2017**, *91*, 1–14. [CrossRef] [PubMed]
47. Hayakijkosol, O.; Jaroenram, W.; Owens, L.; Elliman, J. Reverse transcription polymerase chain reaction (RT-PCR) detection for Australian Cherax reovirus from redclaw crayfish (Cherax quadricarinatus). *Aquaculture* **2021**, *530*, 735881. [CrossRef]
48. Hayakijkosol, O.; Owens, L. Investigation into the pathogenicity of reovirus to juvenile Cherax quadricarinatus. *Aquaculture* **2011**, *316*, 1–5. [CrossRef]
49. Bekavac, A.; Dragičević, P.; Dragun, Z.; Beck, A.; Maguire, I.; Ivanković, D.; Fiket, Ž.; Gračan, R.; Hudina, S. Disturbance in invasion? Idiopathic necrotizing hepatopancreatitis in the signal crayfish Pacifastacus leniusculus (Dana, 1852) in Croatia. *J. Fish Dis.* **2021**, 1–16. [CrossRef]
50. Dong, X.; Hu, T.; Liu, Q.; Li, C.; Sun, Y.; Wang, Y.; Shi, W.; Zhao, Q.; Huang, J. A Novel Hepe-Like Virus from Farmed Giant Freshwater Prawn Macrobrachium rosenbergii. *Viruses* **2020**, *12*, 323. [CrossRef]
51. Lightner, D.V.; Pantoja, C.R.; Poulos, B.T.; Tang, K.F.J.; Redman, R.M.; Andrade, T.P.; Bonami, J.R. University Infectious Myonecrosis- New Disease in Pacific White Shrimp. *Glob. Aquac. Advocate* **2004**, *7*, 85.
52. Edgerton, B.; Owens, L.; Glasson, B.; De Beer, S. Description of a small dsRNA virus from freshwater crayfish Cherax quadricarinatus. *Dis. Aquat. Organ.* **1994**, *18*, 63–69. [CrossRef]
53. Sunarto, A.; Naim, S. Totiviruses of Crustaceans. In *Aquaculture Virology*; Elsevier Inc.: Amsterdam, The Netherlands, 2016; pp. 425–439.
54. Zeng, D.; Chen, X.; Xie, D.; Zhao, Y.; Yang, C.; Li, Y.; Ma, N.; Peng, M.; Yang, Q.; Liao, Z.; et al. Transcriptome Analysis of Pacific White Shrimp (Litopenaeus vannamei) Hepatopancreas in Response to Taura Syndrome Virus (TSV) Experimental Infection. *PLoS ONE* **2013**, *8*, e57515. [CrossRef]
55. Guo, Z.X.; He, J.G.; Xu, H.D.; Weng, S.P. Pathogenicity and complete genome sequence analysis of the mud crab dicistrovirus-1. *Virus Res.* **2013**, *171*, 8–14. [CrossRef] [PubMed]
56. Gao, Y.; Liu, S.; Huang, J.; Wang, Q.; Li, K.; He, J.; He, J.; Weng, S.; Zhang, Q. Cryo-electron Microscopy Structures of Novel Viruses from Mud Crab Scylla paramamosain with Multiple Infections. *J. Virol.* **2019**, *93*, e02255-18. [CrossRef] [PubMed]
57. Waldron, F.M.; Stone, G.N.; Obbard, D.J. Metagenomic sequencing suggests a diversity of RNA interference-like responses to viruses across multicellular eukaryotes. *PLoS Genet.* **2020**, *16*, e1007533. [CrossRef]
58. Argenta, F.F.; Hepojoki, J.; Smura, T.; Szirovicza, L.; Hammerschmitt, M.E.; Driemeier, D.; Kipar, A.; Hetzel, U. Identification of Reptarenaviruses, Hartmaniviruses, and a Novel Chuvirus in Captive Native Brazilian Boa Constrictors with Boid Inclusion Body Disease. *J. Virol.* **2020**, *94*, e00001-20. [CrossRef] [PubMed]
59. Dezordi, F.Z.; dos Santos Vasconcelos, C.R.; Rezende, A.M.; Wallau, G.L. In and Outs of Chuviridae Endogenous Viral Elements: Origin of a Potentially New Retrovirus and Signature of Ancient and Ongoing Arms Race in Mosquito Genomes. *Front. Genet.* **2020**, *11*, 1291. [CrossRef] [PubMed]
60. Anderson, R.M.; May, R.M. Population biology of infectious diseases: Part I. *Nature* **1979**, *280*, 361–367. [CrossRef]
61. Heesterbeek, J.A.P. A brief history of R0 and a recipe for its calculation. *Acta Biotheor.* **2002**, *50*, 189–204. [CrossRef]

62. Lloyd-Smith, J.O.; Schreiber, S.J.; Kopp, P.E.; Getz, W.M. Superspreading and the effect of individual variation on disease emergence. *Nature* **2005**, *438*, 355–359. [CrossRef]
63. Lindenfors, P.; Nunn, C.L.; Jones, K.E.; Cunningham, A.A.; Sechrest, W.; Gittleman, J.L. Parasite species richness in carnivores: Effects of host body mass, latitude, geographical range and population density. *Glob. Ecol. Biogeogr.* **2007**, *16*, 496–509. [CrossRef]
64. Gay, N.; Olival, K.J.; Bumrungsri, S.; Siriaroonrat, B.; Bourgarel, M.; Morand, S. Parasite and viral species richness of Southeast Asian bats: Fragmentation of area distribution matters. *Int. J. Parasitol. Parasites Wildl.* **2014**, *3*, 161–170. [CrossRef]
65. Webber, Q.M.R.; Fletcher, Q.E.; Willis, C.K.R. Viral Richness is Positively Related to Group Size, but Not Mating System, in Bats. *Ecohealth* **2017**, *14*, 652–661. [CrossRef]
66. Rose, R.; Constantinides, B.; Tapinos, A.; Robertson, D.L.; Prosperi, M. Challenges in the analysis of viral metagenomes. *Virus Evol.* **2016**, *2*, vew022. [CrossRef] [PubMed]
67. Pedersen, A.B.; Davies, T.J. Cross-species pathogen transmission and disease emergence in primates. *Ecohealth* **2009**, *6*, 496–508. [CrossRef] [PubMed]
68. Zhao, M.; dos Santos Tavares, C.P.; Schott, E.J. Diversity and classification of reoviruses in crustaceans: A proposal. *J. Invertebr. Pathol.* **2021**, *182*, 107568. [CrossRef] [PubMed]
69. Cote, J.; Brodin, T.; Fogarty, S.; Sih, A. Non-random dispersal mediates invader impacts on the invertebrate community. *J. Anim. Ecol.* **2017**, *86*, 1298–1307. [CrossRef]
70. Conte, F.; Voslarova, E.; Vecerek, V.; Elwood, R.W.; Coluccio, P.; Pugliese, M.; Passantino, A. Humane slaughter of edible decapod crustaceans. *Animals* **2021**, *11*, 1089. [CrossRef]

2.3 Metagenomic Characterization of Parental and Production CHO Cell Lines for Detection of Adventitious Viruses

Katarina Bačnik, Denis Kutnjak, Barbara Jerič Kokelj, Nika Tuta, Tan Lončar, Matjaž Vogelsang, Maja Ravnikar

Biologicals, 2021, 69, 70-75, https://doi.org/10.1016/j.biologicals.2020.11.001.

In this publication, we used HTS to detect virus-like sequences in selected CHO cell lines. Our aim was to test various approaches of sample preparation to establish a pipeline for metagenomic analysis and to characterize standard viral metagenome of production and parental CHO cell lines. The comparison of the metagenomics composition of the differently prepared samples showed that among four tested approaches sequencing of ribosomal RNA-depleted total RNA is the most promising approach. The metagenomics investigation of one production and three parental CHO cell lines of diverse origin did not indicate the presence of adventitious viral agents in the investigated samples. The study revealed an expected background of virus-like nucleic acids in the samples, which originate from remains of expression vectors, endogenized viral elements and residuals of bacteriophages. Obtained results are showing the importance of knowing the range and source of background signal in studies of virus detection and serve as a baseline for future investigations of CHO cell lines for adventitious viruses using HTS.

The PhD candidate analyzed the HTS data for production and parental CHO cell lines, including the comparison of virus-like reads detected in the differently prepared samples and detailed analysis of virus-like sequences and their possible origin. She also wrote the first draft of the manuscript.

Biologicals 69 (2021) 70–75

Contents lists available at ScienceDirect

Biologicals

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

Short paper

Metagenomic characterization of parental and production CHO cell lines for detection of adventitious viruses

Katarina Bačnik [a,b,1], Denis Kutnjak [a,*,1], Barbara Jerič Kokelj [a], Nika Tuta [c], Tan Lončar [c], Matjaž Vogelsang [c], Maja Ravnikar [a,d]

[a] National Institute of Biology, Department of Biotechnology and Systems Biology, Večna Pot 111, 1000, Ljubljana, Slovenia
[b] Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000, Ljubljana, Slovenia
[c] Novartis Technical Research & Development, Biologics Technical Development, Lek Pharmaceuticals d.d., Kolodvorska 27, 1234, Mengeš, Slovenia
[d] University of Nova Gorica, Vipavska Cesta, 5000, Nova Gorica, Slovenia

ARTICLE INFO

Keywords:
Adventitious viruses
CHO cells
High-throughput sequencing
Metagenomics
Parental cell lines

ABSTRACT

Viral contamination is a major concern for biological products. Therefore, virus testing of raw materials and cells is essential for the safety of the final product. We used high-throughput sequencing to detect viral-like sequences in selected CHO cell lines. Our aim was to test various approaches of sample preparation, to establish a pipeline for metagenomic analysis and to characterize standard viral metagenome of production and parental CHO cell lines. The comparison of the metagenomics composition of the differently prepared samples showed that among four tested approaches sequencing of ribosomal RNA depleted total RNA is the most promising approach. The metagenomics investigation of one production and three parental CHO cell lines of diverse origin did not indicate the presence of adventitious viral agents in the investigated samples. The study revealed an expected background of virus-like nucleic acids in the samples, which originate from remains of expression vectors, endogenized viral elements and residuals of bacteriophages.

1. Introduction

Chinese hamster ovary (CHO) cells are one of the main cell lines used in the production of recombinant therapeutics [1] and have the highest occurrence of reported virus contaminations compared with other mammalian cell lines [2]. The utilization of the current combination of in vitro, in vivo, and PCR assays for the identification of adventitious viruses has a limited range of detection. High-throughput sequencing (HTS) on the other hand is able to detect all types of nucleic acid sequences in a sample [3] and has gained significant attention in the field of biologicals since a porcine circovirus 1 contamination has been detected in a licensed pediatric vaccine [4] and a novel rhabdovirus was discovered in the Sf9 insect cell line [5]. Since then, HTS has been utilized as a supplemental test and is now considered as an alternative method for adventitious virus testing [6–9]. The aim of this study was to test various approaches of sample preparation, to establish a pipeline for metagenomic analysis and to characterize standard viral metagenome of selected CHO cell lines. This study serves as a baseline for further investigations of CHO cell lines for adventitious viruses using HTS.

2. Materials and methods

2.1. CHO cell lines

Four different CHO cell lines (T-CHO, P-CHO-1, P-CHO-2, P-CHO-3) were used in the study. P-CHO-1, P-CHO-2 (engineered) and P-CHO-3 are parental cell lines originating from different lineages, including CHO-K1, CHO-DUKXB1 and CHO-DG44. As previously described [10], the CHO-K1 and CHO-DG44 (dihydrofolate reductase (DHFR) deficient) cell lineages were generated from the spontaneously immortalized cells from Chinese hamster ovaries by single-cell cloning and gamma radiation, respectively. The DHFR-deficient CHO-DUKXB1 lineage was generated from CHO-K1 cells by chemical mutagenesis. The test (T-CHO) cell line is a monoclonal antibody producing cell line derived from the engineered P-CHO-2 parental cell line.

2.2. Sample preparation and nucleic acids isolation

In the first batch of analyses, we compared different sample

* Corresponding author.
E-mail address: denis.kutnjak@nib.si (D. Kutnjak).
[1] Both authors contributed equally to this paper.

https://doi.org/10.1016/j.biologicals.2020.11.001
Received 25 August 2020; Received in revised form 26 October 2020; Accepted 9 November 2020
Available online 24 November 2020

preparation approaches using T-CHO cell line. Total (tot)RNA extracted from cells and nucleic acids isolated from either nontreated, ultracentrifuged or filtrated growth media were used as inputs for HTS (detailed in Fig. 1 and Supp. materials and methods). Additional two negative controls: processed growth medium and fresh growth medium were treated in the same way and stored for possible troubleshooting and additional confirmation analyses.

In the second batch of analyses, three parental cell lines were processed using the selected approach (totRNA sequencing of CHO cells) based on the results of the first batch of the analyses. First, cells were separated from growth medium by centrifugation (5 min, 4 °C, 400×*g*) and were further used for nucleic acids isolation (P-CHO-1, P-CHO-2, P-CHO-3).

For all analyses, RNA from cell samples was isolated using RNeasy Plus Mini kit (Qiagen) (batch 1 and batch 2). Sample of nuclease free water spiked with luciferase was used as a negative control of nucleic acid isolation from parental cell lines (batch 2). For simultaneous purification of DNA and RNA from growth media samples (batch 1) QIAamp MinElute Virus Spin Kit (Qiagen) was used.

2.3. Sequencing and bioinformatics analysis

For all samples (Table 1), depletion of ribosomal RNA using Ribo-Zero Gold rRNA Removal Kit (Illumina) and reverse transcription of RNA were performed. For first batch of production cell line samples sequencing libraries were prepared using Illumina Nextera DNA Library Prep Kit and sequenced by Illumina MiSeq (2 × 250 bp). For parental cell lines samples Illumina TruSeq Stranded total RNA library preparation kit was used and libraries were sequenced using Illumina HiSeq 2500 (2 × 150 bp). For first batch of samples less data was generated compared to sequencing of parental cell lines, where we increased the sequencing depth to get a better insight into the presence of virus-like nucleic acids.

The pipeline used for data analysis of all samples is presented in Fig. 1 and described in details in supplementary materials and methods. Additionally, reads from parental cell line samples were used for de novo assembly using software SPAdes [11] and were further classified (Supp. materials and methods). The presence of retroviral-like reads corresponding to the sequence of known transcriptionally active retrovirus sequences from CHO genomes were investigated by mapping of reads from parental cell lines to endogenous retroviral sequences (Supp.

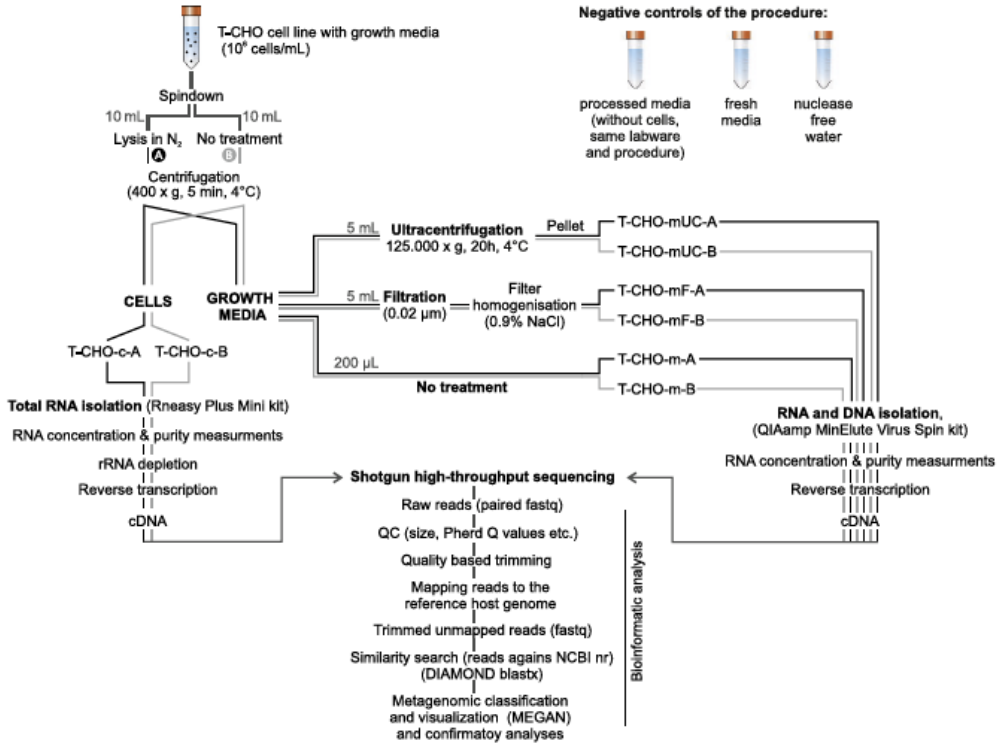


Fig. 1. Detailed schematic representation of sample preparation workflow and bioinformatics pipeline. Two aliquots of cell samples were first briefly centrifuged. In order to lyse the cells, one of the aliquots was sequentially frozen three times in liquid nitrogen (samples originating from this aliquot are marked with A). The other aliquot was left untreated (samples originating from this aliquot are marked with B). Both samples were further centrifuged to separate cells (T–CHO–c-A, T–CHO–c-B) from growth media. Total RNA was isolated from pelleted cells (T–CHO–c-A, T–CHO–c-B). Aliquots of growth media samples were stored for nucleic acids isolation (T–CHO–m-A, T–CHO–m-B). The rest of the growth media was either ultracentrifuged or filtrated through inorganic filter. After ultracentrifugation, supernatant was removed from the tubes and pellets were used for nucleic acid isolation (T–CHO–mUC-A, T–CHO–mUC-B). After sample filtration, the filter was transferred in a sterile Petri dish and fragmented into small fragments, which were then fragmented and used for the nucleic acid isolation (T–CHO–mF-A, T–CHO–mF-B).

Table 1
Summary of reads' mapping in CLC Genomics Workbench and reads' classification using DIAMOND for different samples tested in this study.

Sample designation	Description	Freezing	Processing	Isolation type	Total N of trimmed reads	Mapped to *C. griseus* genome/luciferase (CLC)	Percentage of unmapped assigned reads (DIAMOND)	Percentage of unmapped reads assigned to Eukaryota (DIAMOND)	Percentage of unmapped reads not assigned to Eukaryota (DIAMOND)	Number of unmapped reads assigned to Viruses (DIAMOND)
T–CHO–c-A	CHO cells	YES	–	RNA	2.912.756	55.05	73.72	68.96	4.76	292
T–CHO–c-B	CHO cells	NO	–	RNA	2.454.601	52.82	77.04	71.98	5.06	137
T–CHO–m-A	medium (cells)	YES	–	DNA + RNA	2.673.285	77.02	19.72	13.84	5.88	164
T–CHO–m-B	medium (cells)	NO	–	DNA + RNA	1.714.020	73.51	20.24	13.31	6.92	103
T–CHO–mF-A	medium (cells)	YES	filtration	DNA + RNA	2.507.973	84.78	17.07	11.76	5.31	46
T–CHO–mF-B	medium (cells)	NO	filtration	DNA + RNA	3.036.897	64.18	22.59	12.79	9.80	214
T–CHO–mUC-A	medium (cells)	YES	ultracent.	DNA + RNA	2.089.904	84.68	13.55	11.70	1.85	53
T–CHO–mUC-B	medium (cells)	NO	ultracent.	DNA + RNA	2.377.070	76.39	12.76	11.15	1.61	52
P-CHO-1	CHO cells	NO	–	RNA	181.580.192	67.61	70.02	65.73	4.29	150
P-CHO-2	CHO cells	NO	–	RNA	193.718.096	69.36	68.19	63.59	4.60	2629
P-CHO-3	CHO cells	NO	–	RNA	169.045.514	68.31	68.37	63.74	4.63	214
NKI-luc	NKI and luc	NO	–	RNA	179.049.102	96.23	42.59	0.19	42.40	72

materials and methods).

Reads classified as viral in Diamond/Megan analysis were extracted and further analyzed to possibly confirm/reject that they correspond to viral genomes. Those viral-like reads were then imported into CLC Genomics Workbench (Qiagen) and compared against the latest NCBI nt and nr databases using blastn and blastx, respectively. Reads showing significant similarity only to viral taxa in blastn or blastx analysis were further investigated to determine if they might originate from CHO genomic DNA. They were compared for similarity against selected known CHO short read archive (SRA) datasets (Supp. Table 1) using blastn and against reference viral database (RVDB) U-RVDBv19.0 and U-RVDBV18.0 using blastn and blastx, respectively.

3. Results and discussion

3.1. Comparison of sample preparation approaches and standard viral metagenome of selected CHO cell line

Comparing the number of viral-like reads and detection of different viral taxonomic groups can give some insight about which of the tested sample preparation approaches would be more appropriate for discovery of adventitious viral agents. When comparing different sample preparation approaches, the differences are not striking, however, highest number of viral-like reads (292) and detected viral taxonomic groups (Fig. 2) was found in sample of isolated totRNA from cells. Analyzing ribosomal RNA-depleted totRNA from cells required the smallest amount of handling and therefore smaller chance of the contaminants introduction among tested approaches, additionally supporting our choice of this approach for subsequent analyses (for parental cell lines).

Majority of reads classified as viral in Diamond/Megan analysis corresponded to retroviral sequences (Fig. 2), which are known to be endogenized in CHO genomes. A substantial amount of reads were classified as human betaherpes virus 5 (family *Herpesviridae*), and some to polyomaviruses, both originating from the expression vectors used to manipulate the cell line. Some samples contained remains of bacterial and bacteriophage reads, which are expected to be present as residuals from chemicals and labware used in the procedure. Few reads were misclassified as viral, however further analysis showed they correspond to parts of bacterial or CHO genomes. Finally, very few reads were showing similarities to filoviruses or parvoviruses and were analyzed in detail to better determine their possible origin.

8 reads remained classified as protoparvovirus sequences after the additional blastx and blastn analysis in media samples (T–CHO–m-A and T–CHO–m-B). Sequences could be further assembled, resulting in two contigs that were most similar to sequences of rat parvovirus (82% identity at the nucleic acid level and 74% at the amino acid level) and canine parvovirus (74% identity on nucleic acid level and 71% on amino acid level), confirmed also by blasting against RVDB. Blastn analysis did not show significant similarity to the reads in CHO SRA archives. There have been several reports of endogenization of parvoviral sequences into animal genomes [12], including rodent genomes [13]. Moreover, there have been several reports on contamination of nucleic acid extraction columns with parvoviral nucleic acids [14]. Therefore, it seems highly unlikely that such reads correspond to an adventitious viral agent replicating in the CHO cells, due to the extremely low number of observed reads (corresponding to less than 12% of parvoviral genome length) and low similarity of these reads to known pathogenic viruses.

2 reads were classified as filoviral. Blastn analysis showed that their sequences were highly similar to *C. griseus* sequence (100% sequence identity), annotated as magnesium transporter, but was identified as filoviral using blastx (77% sequence identity). We could find a similar sequence in other existing CHO SRA datasets. We were unable to find a report about CHO genome integrated filoviral sequences, however, integrated filoviral sequences have been found in other rodent species

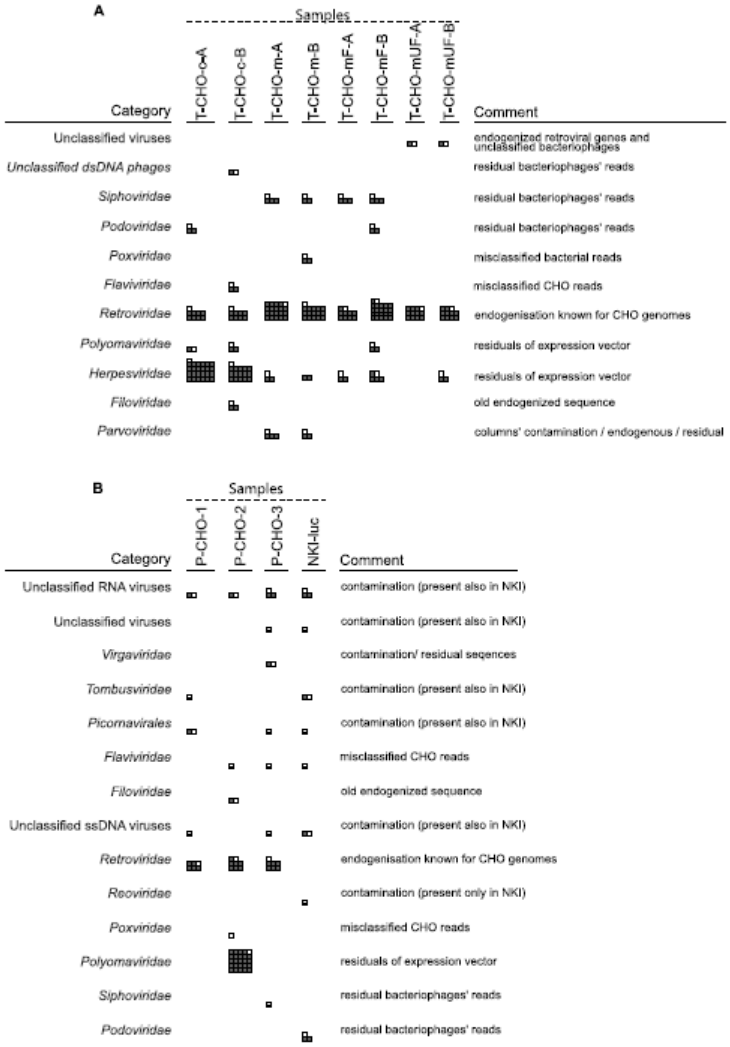


Fig. 2. Overview of the virus-like reads across the tested production cell line samples in first batch of analysis (a) and parental cell line samples with negative control (b). Reads were classified as viral by Diamond and Megan. Squares represent the abundance of the reads classified as a specific category in a specific sample (one square represents 4 reads); the read counts are square root transformed for better representation. The comment column shows final classification or interpretation of reads origin.

[15]. Thus, we suggest that such reads likely correspond to an ancient integrated filovirus sequence in the CHO genome, which likely obtained new function through the evolution of *C. griseus* genome.

3.2. *Metagenomic analysis for the detection of adeventitous viruses in selected parental CHO cell lines*

We used totRNA sequencing of cells to determine the metagenomic baseline of selected CHO parental cell lines. We obtained a high number of sequencing reads for analyzed samples (Table 1), including a negative control, which enabled us to further filter the results. Most of the reads classified as viral corresponded to retroviral sequences that did not map to the reference host genome using selected parameters. However, additional blastn and blastx similarity searches showed that all detected retroviral-like sequences had high similarities with CHO genomes and thus probably represent endogenized retroviral elements. Additional assembly of sequencing data (trimmed reads) from parental cell lines resulted in 13 contigs (78-1076 nt) classified as retroviral using Diamond/Megan analysis (Supplementary Table 2). Blastn and blastx similarity searches against nt/nr databases showed that all contigs have significant alignments to host transcripts or transcripts or genomes of rodent species (Supplementary Table 2). Reads from parental CHO cell lines mapped to the two previously reported transcriptionally active retroviral sequences [16,17]. However, for only one of the two

sequences (ETC109F) and only in P-CHO-2, complete length of the sequence was covered by the reads (Supp. Fig. 1).

In addition, 4 reads were classified as filoviral in P-CHO-2 sample. Reads were assembled into 1 contig (166 bp) showing similarity with *C. griseus* genome using blastn, and similarity to filoviruses using blastx (54% sequence identity), which we already observed in the previous analysis of T-CHO cell line. We could find similar sequences in other existing CHO SRA datasets. The P-CHO-2 parental cell line has been used to prepare the T-CHO cell line analyzed in previous experiment, which explains the detection of similar sequences, with exception of herpesvirus sequences that originate from the expression vectors used to manipulate the parental cell line. Also in P-CHO-2 sample, we detected a relatively high number of reads classifying as polyomaviruses corresponding to expression vector sequence, supporting the fact that P-CHO-2 is an engineered parental cell line. Reads classifying as polyomaviruses (expression vector sequence) were not detected in other parental cell lines, explaining the lower numbers of reads assigned to viruses in other two analyzed parental cell lines.

Few reads of unclassified ssDNA viruses and unclassified RNA viruses were detected both in samples and in the negative control. They were probably introduced during the extraction or sequencing and were thus regarded as contamination. Sequences classified as unclassified ssDNA viruses showed similarity with widely distributed [18] CRESS (circular Rep-encoding single-stranded DNA) viruses and sequences classified as unclassified RNA viruses were similar to sequences of members from *Picornavirales* order. Viral-like reads only present in negative control were probably introduced together with the synthetic luciferase RNA, which was spiked in the negative control.

6 reads in P-CHO-3 sample were classified as plant infecting tobamoviruses from *Virgaviridae* family (blastn – 94–96% sequence identity, blastx – 100% sequence identity) matching two different parts of tobamovirus genomes. Reads classifying as tobamoviruses are potentially representing a very low input contamination. The sequences belong to plant infecting viruses and are not of concern for human health.

To summarize, in all analyzed CHO cell line samples (T-CHO, P-CHO-1, P-CHO-2, P-CHO-3) we detected known endogenized retroviral sequences and residuals of bacteriophages. Residuals of expression vectors with viral sequences and filoviral-like reads were detected in engineered parental (P-CHO-2) and in corresponding derived production (T-CHO) cell line. Reads classifying as parvoviral were detected only in T-CHO cell line media samples (T–CHO–m-A, T–CHO–m-B). In parental cell line samples, which were investigated with higher sequencing depth, more contaminating sequences were detected; however, they were also detected in negative control, except from plant infecting tobamoviral reads only present in P-CHO-3.

4. Conclusions

The metagenomics investigation of one production and three parental CHO cell lines of diverse origin did not indicate the presence of adventitious viral agents in the investigated samples. The study revealed an expected background of virus-like nucleic acids in the samples, which originate from remains of expression vectors, endogenized viral elements and residuals of bacteriophages. Obtained results together with established bioinformatic pipeline focused on viral-like reads serve as a baseline for future investigations of CHO cell lines for adventitious viruses using HTS. The results including the overlapping virus-like sequences between parental and corresponding production cell line and the contaminations present in negative control are showing the importance of knowing the range and source of background signal in studies of adventitious virus detection. Regulatory authorities are increasingly recognizing the potential of HTS for the detection of a broad range of viruses. Therefore, HTS methodologies can complement, supplement or even replace some of the conventional adventitious virus detection assays (Khan et al., 2020). The method described in this study is currently a research tool, which has a potential to supplement other currently used adventitious virus detection methods. Since HTS is constantly evolving, there is a need for standardization and validation of the methodology, including the technical and bioinformatics steps.

Declaration of competing interest

Matjaž Vogelsang and Nika Tuta are employed by Lek Pharmaceuticals d.d. that prepared the analyzed production CHO cell line. The statement is made in the interest of full disclosure and not because the authors consider this a conflict of interest.

Acknowledgements

We would like to thank Tinkara Tinta from National Institute of Biology for providing Whatman® Anodisc inorganic filter membranes and Kristina Gruden and Špela Baebler for discussion of experiments plans.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biologicals.2020.11.001.

Funding source

The study has been performed in the frame of SPS BioPharm.Si project (OP20.00363), funded by Ministry of Education, Science and Sport of Republic of Slovenia and European regional fund. This work was supported by Slovenian research agency (P4-0407 and young researchers' grant awarded to K.B.).

Data statement

The raw sequencing data obtained in this study are confidential.

References

[1] Grillberger L, Kreil TR, Nasr S, Reiter M. Emerging trends in plasma-free manufacturing of recombinant protein therapeutics expressed in mammalian cells. Biotechnol J 2009;4:186–201. https://doi.org/10.1002/biot.200800241.

[2] Barone PW, Wiebe ME, Leung JC, Hussein ITM, Keumurian FJ, Bouressa J, et al. Viral contamination in biologic manufacture and implications for emerging therapies. Nat Biotechnol 2020;1–10. https://doi.org/10.1038/s41587-020-0507-2.

[3] Lambert C, Braxton C, Charlebois RL, Deyati A, Duncan P, Neve F La, et al. Considerations for Optimization of High-Throughput Sequencing Bioinformatics Pipelines for Virus Detection. 2018. https://doi.org/10.3390/v10100528.

[4] Victoria JG, Wang C, Jones MS, Jaing C, McLoughlin K, Gardner S, et al. Viral nucleic acids in live-attenuated vaccines: detection of minority variants and an adventitious virus. J Virol 2010;84:6033–40. https://doi.org/10.1128/jvi.02690-09.

[5] Ma H, Galvin TA, Glasner DR, Shaheduzzaman S, Khan AS. Identification of a novel rhabdovirus in spodoptera frugiperda cell lines. J Virol 2014;88:6576–85. https://doi.org/10.1128/jvi.00780-14.

[6] Khan AS, Benetti L, Blumel J, Deforce D, Egan WM, Knezevic I, et al. Report of the international conference on next generation sequencing for adventitious virus detection in biologicals. Biologicals 2018;55:1–16. https://doi.org/10.1016/j.biologicals.2018.08.002.

[7] Khan AS, Benetti L, Blumel J, Deforce D, Egan WM, Knezevic I, et al. Report of the second international conference on next generation sequencing for adventitious virus detection in biologics for humans and animals. Biologicals 2020;55:1–16. https://doi.org/10.1016/j.biologicals.2018.08.002.

[8] McClenahan SD, Krause PR. Towards dynamic monitoring of cell cultures using high throughput sequencing. Vaccine 2019;37:1001–5. https://doi.org/10.1016/j.vaccine.2018.12.019.

[9] Brussel A, Brack K, Muth E, Zirwes R, Cheval J, Hebert C, et al. Use of a new RNA next generation sequencing approach for the specific detection of virus infection in cells. Biologicals 2019;59:29–36. https://doi.org/10.1016/j.biologicals.2019.03.008.

[10] Wurm FM, Hacker D. First CHO genome. Nat Biotechnol 2011;29:718–20. https://doi.org/10.1038/nbt.1943.

[11] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–77. https://doi.org/10.1089/cmb.2012.0021.

[12] François S, Filloux D, Roumagnac P, Bigot D, Gayral P, Martin DP, et al. Discovery of parvovirus-related sequences in an unexpected broad range of animals. Sci Rep 2016;6:1–13. https://doi.org/10.1038/srep30880.
[13] Arriagada G, Gifford RJ. Parvovirus-derived endogenous viral elements in two south American rodent genomes. J Virol 2014;88:12158–62. https://doi.org/10.1128/jvi.01173-14.
[14] Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, et al. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction Spin columns. J Virol 2013;87:11966–77. https://doi.org/10.1128/jvi.02323-13.
[15] Taylor DJ, Leach RW, Bruenn J. Filoviruses are ancient and integrated into mammalian genomes. BMC Evol Biol 2010;10. https://doi.org/10.1186/1471-2148-10-193.
[16] Duroy PO, Bosshard S, Schmid-Siegert E, Neuenschwander S, Arib G, Lemercier P, et al. Characterization and mutagenesis of Chinese hamster ovary cells endogenous retroviruses to inactivate viral particle release. Biotechnol Bioeng 2020;117:466–85. https://doi.org/10.1002/bit.27200.
[17] Lie YS, Penuel EM, Low MA, Nguyen TP, Mangahas JO, Anderson KP. Chinese hamster ovary cells contain transcriptionally active full-length type C proviruses. J Virol 1994;68:7840–9.
[18] Kazlauskas D, Koonin EV, Krupovic M. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. Nat Commun 2019:1–12. https://doi.org/10.1038/s41467-019-11433-0.

Chapter 3

Discussion

3.1 Plant Viruses in Wastewater

In our first publication, we explored the diversity of viruses in wastewater, focusing on economically relevant ssRNA plant viruses and their biological significance in wastewater.

The composition of whole metagenomes supported the previously observed biological complexity of wastewater samples (Alhamlan et al., 2013; Gulino et al., 2020). Low amount of viral nucleic acids in contrast to the background ones was observed in non-concentrated wastewater samples, where prokaryotic sequences were most abundant. However, the percentage of viral reads increased in CIM concentrated samples, using either step gradient elution concentration approach or linear gradient elution concentration approach.

By analyzing metagenomes of the 12 wastewater samples we detected members of 56 different viral families, including those with double-stranded (ds)DNA, single-stranded (ss)DNA, dsRNA and ssRNA genomes. Examination of wastewater samples by transmission electron microscopy (TEM) also revealed a diversity of virion morphologies with filamentous viral particles and phage-like virions. Detection of sequences of human pathogenic viruses (members of the *Caliciviridae, Picornaviridae, Reoviridae, Hepeviridae, Polyomaviridae* families) was consistent with known viral species typically found in wastewater (Fernandez-cassi et al., 2018).

In this publication, we describe the presence of members of 11 families of plant pathogenic ssRNA viruses. The most commonly found viruses recovered before and after wastewater treatment belonged to the *Tobamovirus* genus most abundantly represented by sequences of PMMoV, tobacco mild green mosaic virus (TMGMV), tomato mosaic virus (ToMV) and cucumber green mottle mosaic virus (CGMMV). Tobamoviruses, known to have extremely stable virions (Rosario, Nilsson, et al., 2009), are considered a major risk to a range of agriculturally important plant species belonging to the *Solanaceae*, *Cucurbitaceae* and other plant families. Their sequences were previously detected in different environmental waters including ballast water (Kim et al., 2015), irrigation systems (Boben et al., 2007), drinking water (Haramoto et al., 2013) and raw and urban sewage (Cantalupo et al., 2011; Fernandez-cassi et al., 2018). We also detected sequences of tomato brown rugose fruit virus (ToBRFV), which has not been confirmed in Slovenia at the time of analysis and is an emerging tobamovirus that has brought concern to plant health authorities worldwide. The presence of sequences of different plant pathogenic viruses, described in detail in Publication 2.1, shows that the combination of efficient virus concentration and metagenomics used in this study is applicable for searching of new emerging viruses in new areas.

We observed a higher fraction of viral reads in both step/linear gradient chromatographically concentrated influent and effluent samples compared to non-concentrated ones. The enrichment for viral sequences was most noticeable in the middle linear gradient fractions at the peak of absorbance according to RT-qPCR analysis for selected tobamoviruses (CGMMV and PMMoV) and contained the highest number of reads belonging to tobamoviruses. Both concentration approaches (step and linear) allowed us to focus on the diversity of ssRNA viruses, with the linear gradient having the potential to study specific viral groups, namely tobamoviruses, in more detail.

Infectivity tests with non-concentrated wastewater samples used for inoculation did not result in test plant infection. Therefore, the ability of the CIM chromatographic sample preparation step to increase viral concentration by several orders of magnitude (Steyer et al., 2015) and at the same time to remove impurities present in the sample (Rupar et al., 2013) likely contributed to the success of infectivity tests. In test plants inoculated with concentrated influent sample and using small RNA sequencing that enables generic detection of RNA and DNA viruses and is widely used in plant virus discovery (Pecman et al., 2017), we were able to detect two infective tobamoviruses: PMMoV and ToMV. To assess the effect of the wastewater treatment on plant virus infectivity, we also inoculated test plants with the concentrated effluent of the WWTP, which is released directly into the river. Test plants again showed disease symptoms with reduction of the size of the plant, curling and mosaic, which were linked with PMMoV and TMGMV infection using small RNA sequencing. All of the influent and effluent fractions that successfully infected test plants contained a high number of reads of the detected infective viruses. Before our study, infectivity of plant viruses was never studied in influents and effluents of WWTP, even though using reclaimed wastewater for irrigation purposes is becoming a widespread practice (Pedrero et al., 2010).

These results implicate the flux of infective pathogenic plant viruses from anthropogenic environments into environmental waters. Wastewater-derived viral genetic material is commonly deposited in the environment (Adriaenssens et al., 2021), therefore the release of infective plant viruses into environment through wastewater might have consequences in rapid transmission of pathogenic viruses to new areas, which cannot be reached by other vectors. Wastewaters metagenomics can provide a comprehensive view into the presence and abundance of plant viruses, possibly reflecting the diversity of local plants as well as the plants consumed by local residents and animals (Ng et al., 2012). Wastewater-based epidemiology has the potential to predict critical locations for viral disease onset (Xagoraraki & Brien, 2019) and monitoring viral epidemics (Kitajima et al., 2020), therefore it could be applied also to predict and prevent plant viral disease outbreaks.

3.2 Viruses of Invasive Signal Crayfish

In the first study, we used shotgun metagenomics to address the understudied presence of plant viruses in water samples, while in the second study we applied this method to search for new viruses in an aquatic host, which has not yet been studied for the presence of viruses. Our second publication reports the presence of novel and divergent RNA viruses detected by HTS of RNA isolated from the invasive signal crayfish hepatopancreas, tissue most often associated with viral infections in crayfish (Dragičević, Bielen, et al., 2020).

By metagenomics analysis of signal crayfish sampled at four locations along its invasion range, including upstream and downstream invasion cores and invasion fronts, only a small fraction of reads was classified as viral sequences, using reads-based protein similarity

searches. Analysis of reads and de novo assembled contigs showed the presence of diverse virus-like sequences most similar to different unclassified RNA viruses. Small fraction of viral reads and dominance of previously unknown RNA viruses in aquatic arthropod hosts samples is in line with previous studies (Chang et al., 2021; Shi et al., 2016b; Wolf et al., 2020). Many of the virus-like sequences identified in this study were having highest identities with unclassified RNA viruses discovered as part of a large-scale metagenomics study of invertebrate viromes (Shi et al., 2016b), where similarities with viruses described in different crustacean hosts were also observed.

Sequences belonging to potential invertebrate viruses exhibited relatively low levels of similarity to known viral sequences. Based on the presence of RdRp domain and contig length we selected four distinct viral contigs for further analyses described in detail in Publication 2.2, including their genome organization and phylogenetic relationships. Selected viral sequences represent signal crayfish associated reo-like virus 1, hepe-like virus 1, toti-like virus 1 and picorna-like virus 1. High average read coverage values were obtained for these viral sequences and they were showing similarities to viruses previously described in crustacean hosts, suggesting that it is likely they were infecting signal crayfish rather than being associated with the environment or co-infecting microorganisms. Their presence in analyzed samples was confirmed by RT-PCR.

In our study, signal crayfish-associated reo-like virus 1 had the highest average read coverage in all samples. Using blastx we observed distant similarity to RdRp sequence of Cherax quadricarinatus reovirus, which was recently associated with hepatopancreatic samples of redclaw crayfish (*Cherax quadricarinatus*) (Hayakijkosol et al., 2021) and causes necrosis of hepatopancreocytes and inflammatory cells in hepatopancreatic tubules (Hayakijkosol & Owens, 2011). Noteworthy, we have recorded a similar condition, the acute necrotizing hepatopancreatitis, along the whole invasion range of the signal crayfish in the Korana river, which is currently classified as idiopathic (Bekavac et al., 2022). Thus, further studies would be needed to associate individual crayfish samples of a known health status with the presence of signal crayfish-associated reo-like virus 1 as a potential causative agent of this disease.

Secondly, by comparing different sampling locations we have noticed that the number of distinct viral sequences and reads corresponding to these sequences varied between the sampling locations. These differences might be a consequence of different factors and processes, however based on our results we can speculate that particularly differences in signal crayfish population density and co-occurrence with the native and phylogenetically related crayfish species could influence the abundance of viral sequences. When comparing the two upstream and two downstream samples, the ones at invasion cores had higher numbers of virus-like reads detected. At the same time, higher crayfish abundances were observed at invasion cores. In terms of abundance of viral sequences and distinct viral contigs the upstream front location was most different from other locations, where again the correlation between the lowest number of viral sequences and lowest signal crayfish abundance, which probably resulted in lower encounter rates between individuals, was observed. Despite the lower total number of virus-like reads in upstream front sample, the highest number of distinct putative virus-like contigs similar to unclassified RNA viruses was found here. According to previous research (Gay et al., 2014; Lindenfors et al., 2007; Webber et al., 2017), a decrease in viral diversification is expected in less dense groups, such as invasion fronts. In contrast, we observed the highest number of distinct viral contigs in upstream-front location. However, at invasion fronts also the population of native *P. leptodactylus* is present with higher abundances than the signal crayfish (Dragičević, Faller, et al., 2020), potentially contributing to inter-specific transmission of viruses.

For the three viruses, which were present at all four locations with relatively high average read coverage, we determined the diversities of their populations by calculating

average nucleotide diversities. Since we analyzed composite hepatopancreas samples without the information about the number of infected individual, nucleotide diversity might help us understand transmission dynamics of analyzed viruses. For instance, higher nucleotide diversity might indicate higher prevalence of a particular virus. After calculating nucleotide diversities, we observed two contrasting patterns. For signal crayfish-associated reo-like virus 1, the nucleotide diversity was lower at the invasion fronts, while for signal crayfish-associated hepe-like virus 1 and toti-like virus 1, the nucleotide diversity was higher at the invasion fronts. These differences might be a consequence of different factors and processes; however, based on our results we can postulate hypotheses to explain the observed patterns in connection to ecological conditions at different sample sites. Higher nucleotide diversity of reo-like virus at core sites might suggest the transmission of virus from longer established core populations with high signal crayfish abundance to recently established front populations of low signal crayfish abundance. Signal crayfish-associated reo-like virus 1 might be an endemic virus of signal crayfish, introduced to the new habitat along with its host. On the contrary, higher nucleotide diversity of signal crayfish-associated toti-like virus 1 and hepe-like virus 1 in samples from invasion fronts could potentially reflect the introduction of these viruses from other host populations (e.g., from native *P. leptodactylus* with higher density at the fronts) to populations of low signal crayfish abundance at invasion fronts. At invasion cores, the signal crayfish have displaced the native host, which would, under the proposed scenario, explain the decrease of nucleotide diversity for signal crayfish-associated toti-like virus 1 and hepe-like virus 1. Such host associations should be further investigated and additional analysis of individual samples would be needed to advance the understanding and test the postulated hypotheses.

3.3 Virus-Like Sequences in Samples from Biopharmaceutical Industry

In previous studies we used shotgun sequencing to explore hidden viral diversity in relatively understudied environmental samples, however analogous methodology can be applied to characterize matrices that are well studied, but their purity and absence of viruses is of great importance. In the third publication, we characterized standard viral metagenome of selected production and parental CHO cell lines.

In the first experiment, we tested different CHO cell line preparation approaches coupled with HTS and bioinformatics pipeline focused on virus-like sequences. The majority of the reads in all analyzed samples of production CHO cell line mapped to the reference genome of *C. griseus*. Reads, which were not assigned to *Eukaryota*, and especially reads having similarities with known viral sequences were the ones of further interest for the discovery of adventitious agents. Majority of such virus-like reads corresponded to retroviral sequences, which are known to be endogenized in CHO genomes. A substantial amount of reads were classified as human betaherpes virus 5 (family *Herpesviridae*), and some to polyomaviruses, both originating from the expression vectors used to manipulate the cell line. Some samples contained remains of bacterial and bacteriophage reads, which are expected to be present as residuals from chemicals and labware used in the procedure. Few reads were misclassified as viral, however further analysis showed they correspond to bacterial or CHO genomes.

Finally, very few reads were showing similarities to filoviruses and parvoviruses and their possible origin is discussed in detail in Publication 2.3, including the likely connection of filoviral reads to an ancient integrated filovirus sequence in the CHO genome and

previous reports on contamination of nucleic acid extraction columns with parvoviral nucleic acids (Naccache et al., 2013).

When comparing different sample preparation approaches, where total RNA extracted from cells and nucleic acids (DNA+RNA) isolated from either non-treated, ultra-centrifuged or filtrated growth media were used as inputs for HTS, the differences were not obvious. However, the highest number of viral-like reads and detected viral taxonomic groups was found in the sample of isolated total (tot)RNA from cells. Analyzing ribosomal RNA-depleted totRNA from cells also required the smallest amount of handling and therefore smaller chance of the contaminants introduction among tested approaches, supporting our choice to use this approach for further analysis.

Secondly, we used totRNA sequencing of cells to obtain a high number of sequencing reads for three parental CHO cell lines. Most of the reads classified as viral corresponded to retroviral sequences that did not map to the reference host genome in previous steps of analysis. Additional assembly of retroviral reads followed by similarity search and mapping to previously reported transcriptionally active retroviral sequences (Duroy et al., 2020) (described in detail in publication 2.3) were showing that detected sequences probably represent endogenized retroviral elements.

Like in the previous analysis of the test CHO cell line, we detected few reads similar to filoviruses in one of the parental cell lines. This parental cell line has been used to prepare the production CHO cell line, analyzed in the previous experiment. This explains the detection of similar sequences, with the exception of herpesvirus sequences that originate from the expression vectors used to manipulate the parental cell line. Because this particular parental cell line is an engineered parental cell line, we detected a relatively high number of reads classifying as polyomaviruses corresponding to expression vector sequence. While reads classifying as polyomaviruses (expression vector sequence) were not detected in other parental cell lines, explaining the lower numbers of reads assigned to viruses.

In parental cell line samples with higher sequencing depth, more contaminating sequences of unclassified ssDNA viruses and unclassified RNA viruses were detected. However, they were also detected in negative control and therefore probably introduced as contamination during the extraction or sequencing. Some plant infecting tobamoviral reads were detected in one of the parental cell lines, potentially representing a very low input contamination, which is not of concern for human health.

The metagenomics investigation of one production and three parental CHO cell lines of diverse origin did not indicate the presence of adventitious viral agents in the investigated samples. The contaminating virus-like sequences present in samples and in the negative control are showing the importance of knowing the range and source of background signal in studies of adventitious virus detection. Obtained results, including the overlapping virus-like sequences between parental and corresponding production cell line, serve as a baseline for future investigations of CHO cell lines for adventitious viruses using HTS-based methods, which have a potential to supplement other currently used adventitious virus detection methods.

Chapter 4

Conclusions

This thesis provides important insights into the virus discovery and diversity in complex sample matrices using different sample preparation approaches followed by non-targeted metagenomics.

Firstly, HTS analysis of wastewater metagenomes confirmed the hypothesis that a high diversity and abundance of plant viruses is present in WWTP influents and effluents, including emerging and quarantine viruses. This indicates that the combination of efficient virus concentration and metagenomics is applicable for early detection and monitoring of economically important plant viruses. Wastewater contained infective tobamoviruses, with PMMoV and TMGMV remaining infective after conventional wastewater treatment, again confirming the postulated hypothesis. Thus, the presence of infective plant viruses in wastewater can also have consequences in rapid transmission to new areas, especially when wastewater or reclaimed water is used for irrigation.

We reported, for the first time, the virome of signal crayfish hepatopancreas tissue and found a high diversity of novel divergent viral sequences, including a near complete genome sequence of signal crayfish-associated hepe-like virus 1 and toti-like virus 1, and the partial genomes of signal crayfish-associated reo-like virus 1 and picorna-like viruses. These novel viral sequences partially confirm our hypothesis, however to prove the involvement of signal crayfish-associated reo-like virus 1 in the observed necrotizing hepatopancreatitis future research is needed. Based on our results, we can confirm the hypothesis that the diversity and abundance of viral sequences is different along the invasion range and we can speculate that the differences in the signal crayfish population density along the invasion range and possibilities of inter-specific viral transmissions may have an effect on the diversity and abundance of signal crayfish-associated viral sequences. New insights into signal crayfish virome set foundations for understanding the potential risk of virus transmissions as a result of this invader's dispersal.

The metagenomics investigation of one production and three parental CHO cell lines of diverse origin did not indicate the presence of adventitious viral agents in the investigated samples, rejecting the part of hypothesis about the presence of adventitious viruses. The study revealed an expected background of virus-like nucleic acids in the samples originating from different sources, confirming the part of hypothesis about the presence of virus-like nucleic acids. The results are showing the importance of knowing the range and source of background signal in studies of adventitious virus detection.

The detection of viral sequence using shotgun metagenomics is advantageous for addressing ecological, aetiological and biopharmaceutical perspectives of virome research in complex sample matrices.

Appendix A

Supplementary Material of Included Publications

A.1 Supplementary Material for Publication 2.1

Supplementary materials and methods:

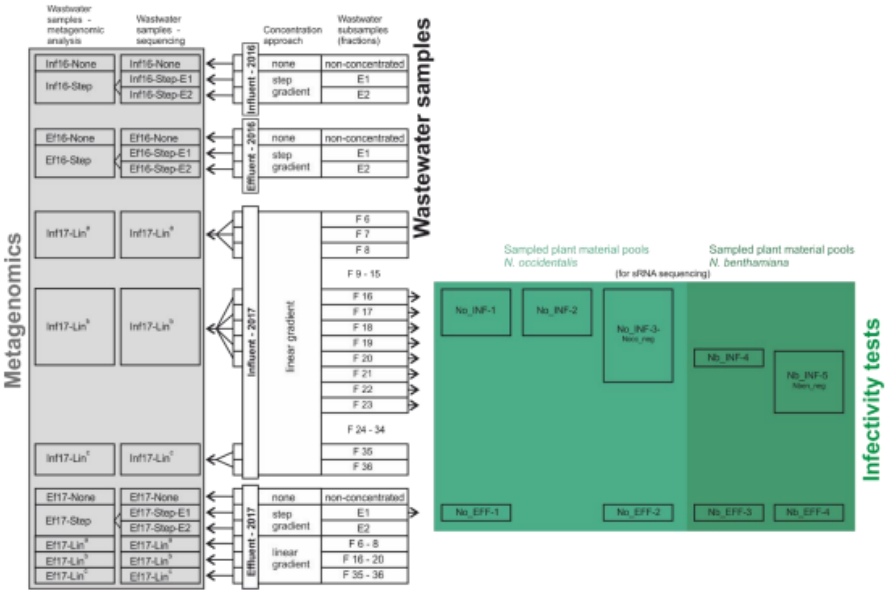
1.

Influent and effluent samples were loaded at 80 mL/min flow rate onto the CIM QA column. After column loading and washing step using 50 mM HEPES buffer, pH 7, we applied either (i) step gradient elution concentration approach (Fig. 3a), where an abrupt change in concentration of salt in the elution buffer was used or ii) linear gradient elution concentration approach (Fig. 3b), where a gradually increasing concentration of salt in the elution buffer was introduced (Supplementary Fig. 1). During step gradient elution, bound viruses were eluted at 4 mL/min flow rate using 50 mM HEPES, 1 M NaCl, pH 7 and two elution fractions (2x15 mL) were collected separately (Step-E1, Step-E2). Linear gradient elution was performed at 2 mL/min introducing increasing concentration of salt in 50 mM HEPES elution buffer (from 0.15 M to 0.75 M NaCl in 10 column volumes). Elution fractions (40 x 2 mL) were collected using an automated fraction collector Frac-920 (GE Healthcare, IL, USA). The elution was controlled by inline monitoring of the conductivity and absorbance at 280 nm. Hydroxyl methacrylate monolith pre-column (BIA separations, Slovenia) was used when concentrating samples taken in year 2017 to reduce backpressure increase on the subsequent anion-exchanger column and to increase the QA column binding capacity and lifetime as described previously (Rački et al., 2015). Between experiments, CIM columns were regenerated using 1 M NaOH loading for 2 h at low flow rate and were stored according to the manufacturer's instructions (BIA Separations, Slovenia).

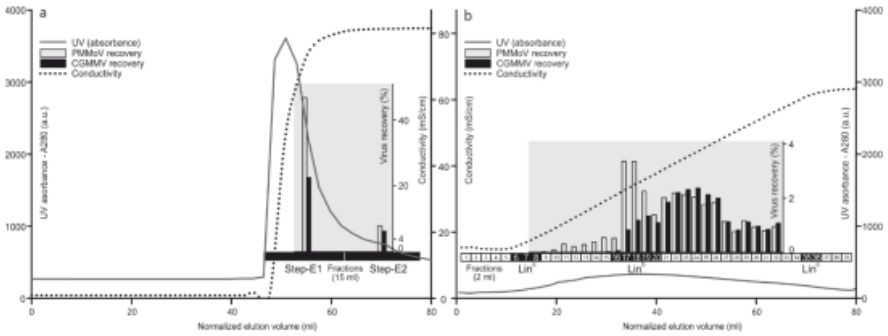
Wastewater samples from 2016 were concentrated using step gradient elution. Influent sample (2 L) was concentrated into Inf16-Step-E1 and Inf16-Step-E2 and 5 L of effluent sample into Ef16-Step-E1, Ef16-Step-E2. Samples taken in 2017 were concentrated using step gradient elution for effluent sample (Ef17-Step-E1, Ef17-Step-E2) and both, linear gradient elutions for influent (Inf17-Lin^{a}, Inf17-Lin^{b}, Inf17-Lin^{c}) and effluent sample (Ef17-Lin^{a}, Ef17-Lin^{b}, Ef17-

Lin[c]). The concentration of 5 L influent sample (year 2017) included two separate linear gradient concentration procedures for 2 x 2 L aliquots, where elution fractions of the first one were used for infectivity tests and sequencing and the other one for determining the presence and concentration of selected plant viruses by quantitative reverse transcription PCR (RT-qPCR). The concentration of 5 L effluent sample (year 2017) also included two separate concentration procedures for 2 x 2 L aliquots. The first one was concentrated using linear gradient elution and fractions were used for sequencing. The second aliquot was concentrated using step gradient elution and elution fractions were used for infectivity test and sequencing (Fig. 1).

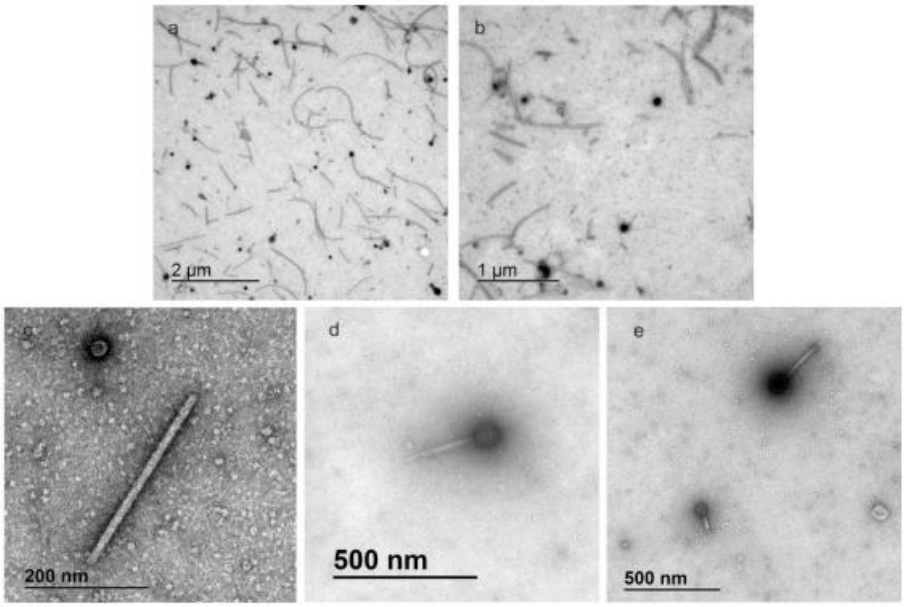
Supplementary figures:



Supplementary Figure 1: Scheme depicting processing of wastewater samples, illustrating their traceability through different stages of the study: from sampling to sequencing/infectivity tests. Wastewater samples were concentrated using different approaches (middle part). The arrows designate the processing of the wastewater samples after concentration. The same samples were used for either metagenomics with subsequent bioinformatics analyses (grey insert on the left) or infectivity tests using two different test plant species (insert on the right with two shades of green). Boxes indicate the pools of the fractions for the metagenomics analysis (left, detailed in Table 1) or pools of symptomatic or asymptomatic test plants after inoculation with corresponding concentrated wastewater samples/fractions (right, detailed in Table 2).



Supplementary Figure 2: Chromatograms representing different concentration approaches. Wastewater influent samples (2 L) were concentrated using CIM QA 8 mL monolithic column using either (a) step gradient elution, where two elution fractions (2x15 mL) were collected separately (Step-E1, Step-E2) or (b) linear gradient elution with 40 (2 mL each) fractions collected. Conductivity (reflecting ionic strength of the used elution buffer) and absorbance at 280 nm (reflecting an estimation of concentration of proteins in solution) were measured during the chromatographic runs and are depicted on the plots. Grey inserts with column charts within the chromatograms show the recoveries (%) which were calculated for each fraction for PMMoV and CGMMV using RT-qPCR results (raw data given in Supplementary Table 1, Supplementary Table 2). Collected elution fractions are designated below the column chart – the fractions highlighted in black (Step-E1 and Step-E2 for step gradient and Lin^a, Lin^b and Lin^c for linear gradient elution) were selected for shotgun HTS. The results shown on these two representative graphs correspond to the samples Inf16-Step and Inf17-Lin.



Supplementary Figure 3: Demonstrative micrography of wastewater treatment plant influent sample. (a, b) Different virion morphologies in selected elution fraction (Fraction 18) after ultracentrifugation of the sample onto transmission electron microscopy (TEM) grid. (c) Virion with morphological characteristics of tobamoviruses in influent linear gradient elution fraction (Fraction 24). Width and length of particle in nm corresponds to the dimension of tobamovirus. (d, e) Virions with morphological characteristics of bacteriophages in influent linear gradient elution fraction (Fraction 24).

Supplementary tables:

Supplementary Table 1: Recoveries of selected tobamoviruses (pepper mild mottle virus - PMMoV and cucumber green mild mottle virus - CGMMV) from the linear gradient elution chromatographic run shown in Figure 1.

	V (mL)	PMMoV average Cq (Luc corrected)	Recovery (%)	CGMMV average Cq (Luc corrected)	Recovery (%)
influent non concentrated	2000	27.47	N/A	23.94	N/A
sample after 0.8µm filter	2000	28.16	N/A	24.67	N/A
sample after CIM-OH	2000	27.86	100.000	24.64	100.0000
flow through	2000	36.78	0.380	32.76	0.5224
fraction 5	2	ND	0.000	34.10	0.0002
fraction 6	2	ND	0.000	34.04	0.0002
fraction 7	2	35.90	0.001	33.85	0.0003
fraction 8	2	32.68	0.005	33.66	0.0003
fraction 9	2	30.28	0.022	34.05	0.0002
fraction 10	2	28.09	0.087	34.01	0.0002
fraction 11	2	26.09	0.304	32.77	0.0005
fraction 12	2	26.70	0.206	33.18	0.0004
fraction 13	2	26.19	0.284	31.74	0.0010
fraction 14	2	25.71	0.385	29.16	0.0054
fraction 15	2	25.15	0.546	26.82	0.0244
fraction 16	2	25.22	0.523	25.26	0.0669
fraction 17	2	22.25	3.352	21.38	0.8274
fraction 18	2	22.25	3.346	20.83	1.1812
fraction 19	2	22.88	2.252	20.64	1.3348
fraction 20	2	23.66	1.384	20.95	1.0907
fraction 21	2	23.06	2.017	20.14	1.8428
fraction 22	2	22.94	2.172	19.87	2.1984
fraction 23	2	22.98	2.120	19.79	2.3119
fraction 24	2	23.04	2.035	19.76	2.3640
fraction 25	2	23.27	1.766	19.93	2.1209
fraction 26	2	23.21	1.832	20.02	1.9960
fraction 27	2	23.99	1.130	20.91	1.1226
fraction 28	2	24.61	0.764	21.40	0.8181
fraction 29	2	23.94	1.162	20.92	1.1167
fraction 30	2	24.29	0.933	21.16	0.9575
fraction 31	2	24.56	0.791	21.35	0.8438
fraction 32	2	24.28	0.941	20.98	1.0715

Recoveries of selected tobamoviruses (PMMoV, CGMMV) from the linear gradient elution with 40 (2 mL each) collected fractions (aliquot of influent sample sequenced as Inf17-Lin[a], Inf17-Lin[b], Inf17-Lin[c]). All Cq values represent an average of two replicates tested by RT-qPCR and are normalized using internal control RNA luciferase (Luc). Recovery of specific fraction represents the percentage of loaded virus, which was quantified in the elution fraction, considering the total amount of loaded viruses in the primary sample (sample after CIM-OH) as 100%. Negative controls (negative control of isolation, no template control) used in a given experiment were negative. N/A - not applicable; ND – not determined

Supplementary Table 2: Recoveries of selected tobamoviruses (PMMoV, CGMMV) from the step gradient elution chromatographic run shown in Figure 1.

	V (mL)	PMMoV average Cq (Luc corrected)	Recovery (%)	CGMMV average Cq (Luc corrected)	Recovery (%)
influent non concentrated	2000	28.85	N/A	22.87	N/A
sample after 0.8µm filter	2000	29.16	100.000	23.83	100.000
flow through	2000	ND	ND	ND	ND
Step-E1	15	22.50	47.974	18.55	22.952
Step-E2	15	25.41	7.783	20.55	6.289

Recoveries of selected tobamoviruses (PMMoV, CGMMV) from the step gradient elution, where two elution fractions (2x15 mL) were collected separately (Inf16-Step-E1, Inf16-Step-E2, comprised in Inf16-Step). All Cq values represent an average of two replicates tested by RT-qPCR and are normalized using internal control RNA luciferase (Luc). Recovery of specific fraction represents the percentage of loaded virus, which was quantified in the elution fraction, considering the total amount of loaded viruses in the primary sample (sample after 0.8µm filter) as 100%.Negative controls (negative control of isolation, no template control) used in a given experiment were negative. N/A - not applicable; ND – not determined

Supplementary Table 3/Source data file:Taxonomic classification of viral reads in sequenced samples. Source data file presents all viral sequences and their classification on different taxonomic levels based on Diamond blastx similarity searches followed by MEGAN taxonomic classification with following parameters for the Lowest Common Ancestor algorithm: min score = 50.0, max expected = 0.01, top percent = 10.0, min support percent = 0, min support = 1, and LCA percent = 100.0.

Supplementary References:

Rački, N., Kramberger, P., Steyer, A., Gašperšič, J., Štrancar, A., Ravnikar, M., Gutierrez-Aguirre, I., 2015. Methacrylate monolith chromatography as a tool for waterborne virus removal. J. Chromatogr. A 1381, 118–124. https://doi.org/10.1016/j.chroma.2015.01.003

A.2 Supplementary Material for Publication 2.2

viruses

MDPI

Supplementary file

Virome analysis of signal crayfish *(Pacifastacus leniusculus)* along its invasion range reveals diverse and divergent RNA viruses

Katarina Bačnik, Denis Kutnjak, Silvija Černi, Ana Bielen and Sandra Hudina

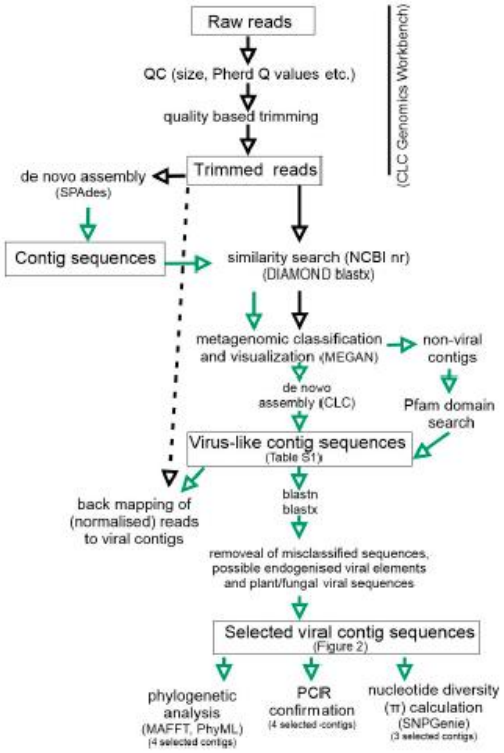


Figure S1. Schematic representation of different steps of the bioinformatic analysis of sequencing reads (black arrows) and contigs (green arrows) to identify virus-like sequences in the signal crayfish hepatopancreas samples

Table S1. Selected putative invertebrate virus-like contigs (> 300 nts) representing signal crayfish associated viruses identified in this study with their contig length (nts) and percentage of mapped reads resulting from mapping of reads from individual samples to virus-like contigs.

				Percentage of mapped reads (%)			
Contig name	**Virus name**	**GenBank accesion**	**Contig length**	**UF**	**UC**	**DC**	**DF**
Virus-like contig 4	Signal crayfish associated reo-like virus 1	OK317706	4,234	0.035179	0.793427	0.334182	0.451408
Virus-like contig 139	Signal crayfish associated hepe-like virus 1	OK317707	10,400	0.003373	0.322649	0.609349	0.300643
Virus-like contig 141	Signal crayfish associated toti-like virus 1	OK317708	8,576	0.010525	0.023344	0.025111	0.019544
Virus-like contig 1	Signal crayfish associated picorna-like virus 1	OK317711	4,587	0.000300	0.000003	0.000000	0.000000
Virus-like contig 10	Signal crayfish associated picorna-like virus 2	OK317712	531	0.000014	0.000000	0.000000	0.000000
Virus-like contig 11	Signal crayfish associated picorna-like virus 3	OK317713	320	0.000004	0.000000	0.000000	0.000000
Virus-like contig 15	Signal crayfish associated picorna-like virus 4	OK317714	898	0.000022	0.000005	0.000000	0.000001
Virus-like contig 9	Signal crayfish associated picorna-like virus 5	OK317715	771	0.000016	0.000000	0.000000	0.000000
Virus-like contig 5	Signal crayfish associated picorna-like virus 6	OK317716	714	0.000032	0.000000	0.000000	0.000000
Virus-like contig 13	Signal crayfish associated picorna-like virus 7	OK317717	1,894	0.000065	0.000000	0.000000	0.000000
Virus-like contig 66	Signal crayfish associated tombus-like virus 1	OK317718	4,504	0.000000	0.000001	0.000380	0.000000
Virus-like contig 84	Signal crayfish associated tombus-like virus 2	OK317719	2,981	0.000000	0.000000	0.000000	0.000220
Virus-like contig 55	Signal crayfish associated tombus-like virus 3	OK317720	1,425	0.000000	0.000047	0.000000	0.000000
Virus-like contig 35	Signal crayfish associated tombus-like virus 4	OK317721	665	0.000000	0.000021	0.000000	0.000000
Virus-like contig 24	Signal crayfish associated tombus-like virus 5	OK317722	538	0.000009	0.000000	0.000000	0.000000
Virus-like contig 169	Signal crayfish associated tombus-like virus 6	OK317723	301	0.000000	0.000002	0.000000	0.000000
Virus-like contig 140	Signal crayfish associated chu-like virus 1	OK317724	2,216	0.000049	0.000170	0.000099	0.000099
Virus-like contig 65	Signal crayfish associated chu-like virus 2	OK317725	746	0.000002	0.000003	0.000017	0.000003
Virus-like contig 7	Signal crayfish associated chu-like virus 3	OK317726	493	0.000006	0.000003	0.000002	0.000003
Virus-like contig 145	Signal crayfish associated chu-like virus 4	OK317727	1,009	0.000111	0.000094	0.000071	0.000147
Virus-like contig 146	Signal crayfish associated chu-like virus 5	OK317728	418	0.000002	0.000003	0.000003	0.000006
Virus-like contig 222	Signal crayfish associated chu-like virus 6	OK317729	356	0.000001	0.000000	0.000000	0.000005
Virus-like contig 27	Signal crayfish associated partiti-like virus 1	OK317730	1,188	0.000247	0.000000	0.000000	0.000000
Virus-like contig 116	Signal crayfish associated picorna-like virus 8	OK317731	734	0.000000	0.000000	0.000000	0.000012
Virus-like contig 2	Signal crayfish associated picorna-like virus 9	OK317732	412	0.000007	0.000000	0.000000	0.000000
Virus-like contig 36	Signal crayfish associated picorna-like virus 10	OK317733	457	0.000001	0.000005	0.000000	0.000002
Virus-like contig 83	Signal crayfish associated sobemo-like virus 1	OK317734	431	0.000000	0.000000	0.000000	0.000012
Virus-like contig 147	Signal crayfish associated sobemo-like virus 2	OK317709	512	0.000009	0.000000	0.000000	0.000004
Virus-like contig 30	Signal crayfish associated nama-like virus 1	OK317710	378	0.000004	0.000000	0.000000	0.000000

Table S2. (A) Virus-like contigs from different locations (UF – upstream front, UC – upstream core, DC – downstream core, DF – downstream front) identified in Diamond analysis together with average coverage values, Diamond classification, blastn similarity search results (NCBI-nt, April 2021) and sequences of individual virus-like contigs. (B) virus-like contigs from different locations (UF – upstream front, UC – upstream core, DC – downstream core, DF – downstream front) identified using pfam domain search, where the remaining contigs not classified as viral by Diamond (963721) were translated and compared with the entire Pfam database, together with average coverage values, Pfam search results and blastn (NCBI-nt, April 2021) and blastx (NCBI-nr, June 2021) similarity search results.

Table S3. Sequences of primers used for PCR amplification of selected virus-like contigs representing signal crayfish associated viruses identified in this study.

Virus-like Contig	Primer sequence (5'->3')	Orientation	Amplicon length
Virus-like contig 4 (signal crayfish associated reo-like virus 1)	TTCTGGCGCGACTTTAGCTT	forward	269
	GCCGCACGTTGCTGTAAATA	reverse	
Virus-like contig 139 (signal crayfish associated hepe-like virus 1)	GGTGACGACCTTGCGATCAT	forward	248
	GCGCCGCATAAGTGAATAGC	reverse	
Virus-like contig 141 (signal crayfish associated toti-like virus 1)	GGTTGTTGCACATGAAGCGT	forward	249
	GCCGTAATGCAGCGTGTTAG	reverse	
Virus-like contig 1 (signal crayfish associated picorna-like virus 1)	CCACGCACACGAAAAACCAT	forward	193
	CATGGTAGATGGTGGGCTCC	reverse	

Alignment S1. Sequence alignments (.fasta) with virus names and corresponding GenBank accession numbers of viruses used for phylogenetic analysis of signal crayfish associated reo-like virus 1

Alignment S2. Sequence alignments (.fasta) with virus names and corresponding GenBank accession numbers of viruses used for phylogenetic analysis of signal crayfish associated hepe-like virus 1.

Alignment S3. Sequence alignments (.fasta) with virus names and corresponding GenBank accession numbers of viruses used for phylogenetic analysis of signal crayfish associated toti-like virus 1.

Alignment S4. Sequence alignments (.fasta) with virus names and corresponding GenBank accession numbers of viruses used for phylogenetic analysis of signal crayfish associated picorna-like virus 1.

A.3 Supplementary Material for Publication 2.3

Supplementary material

Materials and methods

Sample preparation

Two aliquots of cell samples (2x 10 mL, 10^6 cells/mL) were first briefly centrifuged (spin down). In order to lyse the cells, one of the aliquots was sequentially frozen three times for 10 seconds in liquid nitrogen (samples originating from this aliquot are marked with A). The other aliquot was left untreated (samples originating from this aliquot are marked with B). Both samples were further centrifuged for 5 minutes at 4 °C and 400 x g to pellet the cells and separate cells (T-CHO-c-A, T-CHO-c-B) from growth media. Total RNA was isolated from pelleted cells (T-CHO-c-A, T-CHO-C-B). 200 µL aliquots of growth media samples were stored for nucleic acids isolation (T-CHO-m-A, T-CHO-m-B). The rest of the growth media was either ultracentrifuged or filtrated through inorganic filter with 20 nm pores. For ultracentrifugation, 5 mL of growth media was transferred into Beckmann Coulter Thickwall Polycarbonate Tube (Part Number 355630). Tubes were centrifuged for 20 hours at 4°C and 125.000 x g (Ultracentrifuge Beckman Coulter L8-80M, rotor type 70.1 TI). After ultracentrifugation, supernatant was removed from the tubes and pellets were used for nucleic acid isolation (T-CHO-mUC-A, T-CHO-mUC-B). The other part of growth media samples was filtrated through Whatman® Anodisk inorganic filter membrane (diameter 25 mm, pore size 20 nm). 5 mL of each sample was transferred in the vacuum pump filter system with embedded Whatman® Anodisk filter membrane. Whole filter system was cleaned before processing each new sample using a DNA remover and washed with a MiliQ water. Before loading sample, 1 mL of nuclease free water was let to run through the filter each time prior to applying a new sample. After sample filtration, the filter was transferred in a sterile petri dish and fragmented into small fragments. Small filter fragments were then transferred into a sterile 1,5 mL microcentrifuge tube with 200 µL of 0,9% NaCl (prepared in nuclease free water) and further fragmented with a sterile disposable plastic pestle before proceeding to the nucleic acid isolation (T-CHO-mF-A, T-CHO-mF-B).

Bioinformatic pipeline for metagenomic analysis of sequencing data

The pipeline used for data analysis started with the quality control of the sequencing reads and their trimming (Limit = 0.01, reads ≤25bp were discarded) in CLC Genomics Workbench (Qiagen BIoinformatics). Since the aim of the study was to get an insight into the baseline of viral-like reads in separate parental cell lines, no normalization of sequenced samples was performed in order not to lose any potential viral sequences. The trimmed reads were mapped to reference genome sequence of *Cricetulus griseus*

(GCF_000223135.1) in CLC Genomics Workbench (Length fraction = 0.9, Similarity fraction = 0.95.). Reads obtained in negative control sample were mapped to luciferase sequence to determine and remove luciferase reads from the data. In all samples, unmapped reads were used for further analysis using direct similarity search of sequencing reads on protein level (against complete NCBI nr database; 17/5/2019) using Diamond [1] and subsequent taxonomic classification and visualization with Megan [2].

De novo assembly of parental CHO cell lines reads and search for viral-like contigs

Trimmed reads from parental cell line samples were used for de novo assembly using software SPAdes [3]. Contig sequences generated were further analyzed using direct similarity search of sequencing reads on protein level (against complete NCBI nr database; 17/5/2019) using Diamond [1] and subsequent taxonomic classification and visualization with Megan [2]. Contigs classified as viral in Diamond analysis were extracted and further compared against latest NCBI nt (blastn), nr (blastx) and RVDB (blastn, blastx; U-RVDBv19.0, U-RVDBV18.0) databases to possibly confirm/reject that they correspond to viral genomes.

Investigating the presence of retroviral-like reads corresponding to the sequence of known transcriptionally active retrovirus sequences from CHO genomes

Mapping of sequencing reads from parental cell lines to endogenous retroviral sequences was performed in CLC Genomics Workbench (Length fraction = 0.9, Similarity fraction = 0.95.) to characterize endogenous retroviral sequences. Sequence of genomic locus (ETC109F) reported to be responsible for most type-C viral particle release from CHO cells [4] and previously reported transcriptionally active retroviral sequence (CGU09104) present in CHO cell genomes (Lie et al., 1994) were used as reference sequences.

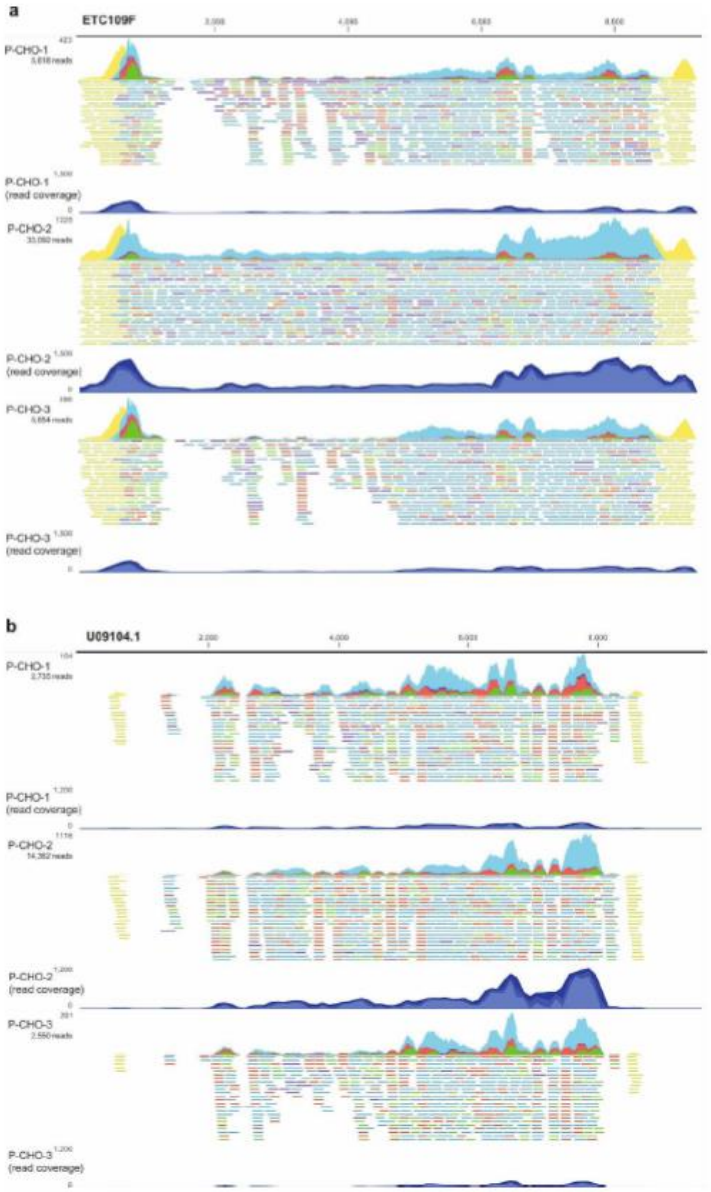
Supplementary Table 1: SRA Blast search set information. Viral-like reads (filoviral, protoparvoviral) were compared for similarity against selected known CHO short read archive (SRA) datasets to determine if they might originate from CHO genomic DNA.

Experiment accession	Run	Data type	SRA Study
SRX091169	SRR329939	WGS	SRP045758
SRX091170	SRR329940	WGS	SRP045759
SRX091171	SRR329941	WGS	SRP045760
SRX091172	SRR329942	WGS	SRP045761
SRX091173	SRR329943	WGS	SRP045762
SRX091174	SRR329944	WGS	SRP045763
SRX091175	SRR329945	WGS	SRP045764
SRX091176	SRR329946	WGS	SRP045765
SRX091177	SRR329947	WGS	SRP045766
SRX091178	SRR329948	WGS	SRP045767
SRX091179	SRR329949	WGS	SRP045768
SRX091180	SRR329950	WGS	SRP045769
SRX091181	SRR329951	WGS	SRP045770
SRX091182	SRR329952	WGS	SRP045771
SRX091183	SRR329953	WGS	SRP045772
SRX091184	SRR329954	WGS	SRP045773

*WGS – whole genome sequencing

Results

Supplementary Table 2/ viral-contigs_parental_classification.xlsx: Viral-like contig sequences assembled from sequencing data of parental cell lines with their taxonomic classification in Diamond/Megan analysis. Contigs were also compared against latest NCBI nt (blastn), nr (blastx) and RVDB (blastn, blastx; U-RVDBv19.0, U-RVDBV18.0) databases.


a
ETC109F
P-CHO-1
P-CHO-1
(read coverage)
P-CHO-2
P-CHO-2
(read coverage)
P-CHO-3
P-CHO-3
(read coverage)
b
U09104.1
P-CHO-1
P-CHO-1
(read coverage)
P-CHO-2
P-CHO-2
(read coverage)
P-CHO-3
P-CHO-3
(read coverage)

71

4

Supplementary Figure 1: Mapping of reads from parental cell line samples to previously reported transcriptionally active retroviral sequence; EUC109F – sequence obtained from a paper (a), U09104.1 – GenBank accession number (b). Mapping of forward single reads (green), reverse single reads (red), paired reads (blue) and non-specific matches (yellow) to corresponding retroviral sequence is shown for each parental cell line (P-CHO-1, P-CHO-2, P-CHO-3) together with read coverage (blue shaded) graphs, showing the number of reads contributing to the mapping for each position.

References

[1] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods 2015;12. https://doi.org/10.1038/nmeth.3176.

[2] Huson DH, Beier S, Flade I, Górska A, El-hadidi M. MEGAN Community edition - Interactive exploration and analysis of larges-scale microbiome sequencing data. Plos Comput Biol 2016;12:1–12. https://doi.org/DOI:10.1371/journal.pcbi.1004957.

[3] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–77. https://doi.org/10.1089/cmb.2012.0021.

[4] Duroy PO, Bosshard S, Schmid-Siegert E, Neuenschwander S, Arib G, Lemercier P, et al. Characterization and mutagenesis of Chinese hamster ovary cells endogenous retroviruses to inactivate viral particle release. Biotechnol Bioeng 2020;117:466–85. https://doi.org/10.1002/bit.27200.

[5] Lie, Y. S., Penuel, E. M., Low, M. A., Nguyen, T. P., Mangahas, J. O., Anderson KP. Chinese hamster ovary cells contain transcriptionally active full-length type C proviruses. J Virol 1994;68:7840–7849.

Appendix B

Permissions for Reproduction of Included Publications

A.4 Permission for Reproduction of Publication 2.1

Water Research
Volume 177, 15 June 2020, 115628

Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment

Katarina Bačnik [a, b], Denis Kutnjak [a], Anja Pecman [a, b], Nataša Mehle [a], Magda Tušek Žnidarič [a], Ion Gutiérrez Aguirre [a], Maja Ravnikar [a, c]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.watres.2020.115628

Get rights and content

Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment

Author: Katarina Bačnik,Denis Kutnjak,Anja Pecman,Nataša Mehle,Magda Tušek Žnidarič,Ion Gutiérrez Aguirre,Maja Ravnikar

Publication: Water Research

Publisher: Elsevier

Date: 15 June 2020

A.5 Permission for Reproduction of Publication 2.2

Open Access Article

Virome Analysis of Signal Crayfish (*Pacifastacus leniusculus*) along Its Invasion Range Reveals Diverse and Divergent RNA Viruses

by Katarina Bačnik [1,2], Denis Kutnjak [1,*], Silvija Černi [3], Ana Bielen [4] and Sandra Hudina [3,*]

[1] Department of Biotechnology and Systems Biology, National Institute of Biology, 1000 Ljubljana, Slovenia

[2] Jozef Stefan International Postgraduate School, 1000 Ljubljana, Slovenia

[3] Department of Biology, Faculty of Science, University of Zagreb, 10000 Zagreb, Croatia

[4] Department of Biochemical Engineering, Faculty of Food Technology and Biotechnology, University of Zagreb, 10000 Zagreb, Croatia

* Authors to whom correspondence should be addressed.

Academic Editors: Leyi Wang and Ganwu Li

Viruses **2021**, *13*(11), 2259; https://doi.org/10.3390/v13112259

Received: 1 October 2021 / Revised: 29 October 2021 / Accepted: 4 November 2021 / Published: 11 November 2021

(This article belongs to the Special Issue Applications of Next-Generation Sequencing in Virus Discovery)

View Full-Text | Download PDF | Browse Figures | Citation Export

Abstract

Crayfish are a keystone species of freshwater ecosystems and a successful invasive species. However, their pathogens, including viruses, remain understudied. The aim of this study was to analyze the virome of the invasive signal crayfish (*Pacifastacus leniusculus*) and to elucidate the potential differences in viral composition and abundance along its invasion range in the Korana River, Croatia. By the high-throughput sequencing of ribosomal RNA, depleted total RNA isolated from the crayfish hepatopancreas, and subsequent sequence data analysis, we identified novel and divergent RNA viruses, including signal crayfish-associated reo-like, hepe-like, toti-like, and picorna-like viruses, phylogenetically related to viruses previously associated with crustacean hosts. The patterns of reads abundance and calculated nucleotide diversities of the detected viral sequences varied along the invasion range. This could indicate the possible influence of different factors and processes on signal crayfish virome composition: e.g., the differences in signal crayfish population density, the non-random dispersal of host individuals from the core to the invasion fronts, and the transfer of viruses from the native co-occurring and phylogenetically related crayfish species. The study reveals a high, previously undiscovered diversity of divergent RNA viruses associated with signal crayfish, and sets foundations for understanding the potential risk of virus transmissions as a result of this invader's dispersal. View Full-Text

Keywords: signal crayfish virome; RNA viruses; invasive alien species; invasion range; high-throughput sequencing

► Show Figures

A.6 Permission for Reproduction of Publication 2.3

Biologicals
Volume 69, January 2021, Pages 70-75

Short paper

Metagenomic characterization of parental and production CHO cell lines for detection of adventitious viruses

Katarina Bačnik [a, b, 1], Denis Kutnjak [a] [1] ✉, Barbara Jerič Kokelj [a], Nika Tuta [c], Tan Lončar [c], Matjaž Vogelsang [c], Maja Ravnikar [a, d]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.biologicals.2020.11.001

Get rights and content

Metagenomic characterization of parental and production CHO cell lines for detection of adventitious viruses

Author: Katarina Bačnik,Denis Kutnjak,Barbara Jerič Kokelj,Nika Tuta,Tan Lončar,Matjaž Vogelsang,Maja Ravnikar

Publication: Biologicals

Publisher: Elsevier

Date: January 2021

References

Adriaenssens, E. M., Farkas, K., McDonald, J. E., Jones, D. L., Allison, H. E., & McCarthy, A. J. (2021). Tracing the fate of wastewater viruses reveals catchment-scale virome diversity and connectivity. *Water Research*, *203*(May), 117568. https://doi.org/10.1016/j.watres.2021.117568

Alhamlan, F. S., Ederer, M. M., Brown, C. J., Coats, E. R., & Crawford, R. L. (2013). Metagenomics-based analysis of viral communities in dairy lagoon wastewater. *Journal of Microbiological Methods*, *92*(2), 183–188. https://doi.org/10.1016/j.mimet.2012.11.016

Anderson, R. M., & May, R. M. (1982). Coevolution of Hosts and Parasites. *Parasitology*, *85*(2), 411–426. https://doi.org/10.1017/S0031182000055360

Aw, T. G., Howe, A., & Rose, J. B. (2014). Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater. *Journal of Virological Methods*, *210*, 15–21. https://doi.org/10.1016/j.jviromet.2014.09.017

Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *PNAS*, *115*(25), 6506–6511. https://doi.org/10.1073/pnas.1711842115

Barone, P. W., Wiebe, M. E., Leung, J. C., Hussein, I. T. M., Keumurian, F. J., Bouressa, J., Brussel, A., Chen, D., Chong, M., Dehghani, H., Gerentes, L., Gilbert, J., Gold, D., Kiss, R., Kreil, T. R., Labatut, R., Li, Y., Müllberg, J., Mallet, L., … Springs, S. L. (2020). Viral contamination in biologic manufacture and implications for emerging therapies. *Nature Biotechnology*, 1–10. https://doi.org/10.1038/s41587-020-0507-2

Bateman, K. S., & Stentiford, G. D. (2017). A taxonomic review of viruses infecting crustaceans with an emphasis on wild hosts. *Journal of Invertebrate Pathology*, *147*(February 2018), 86–110. https://doi.org/10.1016/j.jip.2017.01.010

Bekavac, A., Beck, A., Dragičević, P., Dragun, Z., Maguire, I., Ivanković, D., Fiket, Ž., Gračan, R., & Hudina, S. (2022). Disturbance in invasion? Idiopathic necrotizing hepatopancreatitis in the signal crayfish Pacifastacus leniusculus (Dana, 1852) in Croatia. *Journal of Fish Diseases*, *45*(2), 261–276. https://doi.org/10.1111/jfd.13552

Berting, A., Farcet, M. R., & Kreil, T. R. (2010). Virus susceptibility of Chinese hamster ovary (CHO) cells and detection of viral contaminations by adventitious agent testing. *Biotechnology and Bioengineering*, *106*(4), 598–607. https://doi.org/10.1002/bit.22723

Bettina Klug, James S. Robertson, Richard C. Condit, Stephen J. Seligman, Marian P. Laderoute, Rebecca Sheets, Anna-Lise Williamson, Louisa Chapman, Baevin Carbery, Lisa M. Mac, R. T. C. (2017). Adventitious Agents and Live Viral Vectored Vaccines: Considerations for Archiving Samples of Biological Materials for Retrospective

Analysis. *Physiology & Behavior*, *176*(3), 139–148. https://doi.org/10.1016/j.vaccine.2016.02.015.Adventitious

Boben, J., Kramberger, P., Petrovič, N., Cankar, K., Peterka, M., Štrancar, A., & Ravnikar, M. (2007). Detection and quantification of Tomato mosaic virus in irrigation waters. *European Journal of Plant Pathology*, *118*(1), 59–71. https://doi.org/10.1007/s10658-007-9112-1

Brien, E. O., Munir, M., Marsh, T., Heran, M., Lesage, G., Tarabara, V. V, & Xagoraraki, I. (2017). Diversity of DNA viruses in ef fl uents of membrane bioreactors in Traverse City, MI (USA) and La Grande Motte (France). *Water Research*, *111*, 338–345. https://doi.org/10.1016/j.watres.2017.01.014

Brussel, A., Brack, K., Muth, E., Zirwes, R., Cheval, J., Hebert, C., Charpin, J. M., Marinaci, A., Flan, B., Ruppach, H., Beurdeley, P., & Eloit, M. (2019). Use of a new RNA next generation sequencing approach for the specific detection of virus infection in cells. *Biologicals*, *59*(December 2018), 29–36. https://doi.org/10.1016/j.biologicals.2019.03.008

Cantalupo, P. G., Calgua, B., Zhao, G., Hundesa, A., Wier, A. D., Katz, J. P., Grabe, M., Hendrix, R. W., Girones, R., Wang, D., & Pipas, J. M. (2011). Raw sewage harbors diverse viral populations. *MBio*, *2*(5), 1–11. https://doi.org/10.1128/mBio.00180-11

Chang, T., Hirai, J., Hunt, B. P. V, & Suttle, C. A. (2021). Arthropods and the evolution of RNA viruses. *BioRxiv*, 2021.05.30.446314. https://doi.org/10.1101/2021.05.30.446314

Charlebois, R. L., Sathiamoorthy, S., Logvinoff, C., Gisonni-Lex, L., Mallet, L., & Ng, S. H. S. (2020). Sensitivity and breadth of detection of high-throughput sequencing for adventitious virus detection. *Npj Vaccines*, *5*(1), 1–8. https://doi.org/10.1038/s41541-020-0207-4

Cheval, J., Muth, E., Gonzalez, G., Coulpier, M., Beurdeley, P., Cruveiller, S., & Eloit, M. (2019). Adventitious Virus Detection in Cells by High-Throughput Sequencing of Newly Synthesized RNAs: Unambiguous Differentiation of Cell Infection from Carryover of Viral Nucleic Acids. *MSphere*, *4*(3), 1–11. https://doi.org/10.1128/msphere.00298-19

Cleveland, M. H., Anekella, B., Brewer, M., Chin, P. J., Couch, H., Delwart, E., Huggett, J., Jackson, S., Martin, J., Monpoeho, S., Morrison, T., Ng, S. H. S., Ussery, D., & Khan, A. S. (2020). Report of the 2019 NIST-FDA workshop on standards for next generation sequencing detection of viral adventitious agents in biologics and biomanufacturing. *Biologicals*, *64*(February), 76–82. https://doi.org/10.1016/j.biologicals.2020.02.003

Costa, V. A., Mifsud, J. C. O., Gilligan, D., Williamson, J. E., Holmes, E. C., & Geoghegan, J. L. (2021). Metagenomic sequencing reveals a lack of virus exchange between native and invasive freshwater fish across the Murray-Darling Basin, Australia. *Virus Evolution*, *7*(1), 1–15. https://doi.org/10.1093/ve/veab034

Coutinho, F. H., Silveira, C. B., Gregoracci, G. B., Thompson, C. C., Edwards, R. A., Brussaard, C. P. D., Dutilh, B. E., & Thompson, F. L. (2017). Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nature Communications*, *8*(May), 1–12. https://doi.org/10.1038/ncomms15955

Crowl, T. A., Crist, T. O., Parmenter, R. R., Belovsky, G., & Lugo, A. E. (2008). The spread of invasive species and infectious disease as drivers of ecosystem change. *Frontiers in Ecology and the Environment*, *6*(5), 238–246. https://doi.org/10.1890/070151

Daughton, C. G. (2020). Wastewater surveillance for population-wide Covid-19: The present and future. *Science of the Total Environment*, *736*, 139631. https://doi.org/10.1016/j.scitotenv.2020.139631

De Cárcer, D. A., López-Bueno, A., Pearce, D. A., & Alcamí, A. (2015). Biodiversity and distribution of polar freshwater DNA viruses. *Science Advances*, *1*(5). https://doi.org/10.1126/sciadv.1400127

Dolja, V. V, & Koonin, E. V. (2018). Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Research*, *244*(January), 36–52. https://doi.org/https://doi.org/10.1016/j.virusres.2017.10.020

Dragičević, P., Bielen, A., Petrić, I., & Hudina, S. (2020). Microbial pathogens of freshwater crayfish: A critical review and systematization of the existing data with directions for future research. *Journal of Fish Diseases*, *44*(3), 221–247. https://doi.org/10.1111/jfd.13314

Dragičević, P., Faller, M., Kutleša, P., & Hudina, S. (2020). Update on the signal crayfish, pacifastacus leniusculus (Dana, 1852) range expansion in Croatia: A 10-year report. *BioInvasions Records*, *9*(4), 793–807. https://doi.org/10.3391/bir.2020.9.4.13

Duroy, P. O., Bosshard, S., Schmid-Siegert, E., Neuenschwander, S., Arib, G., Lemercier, P., Masternak, J., Roesch, L., Buron, F., Girod, P. A., Xenarios, I., & Mermod, N. (2020). Characterization and mutagenesis of Chinese hamster ovary cells endogenous retroviruses to inactivate viral particle release. *Biotechnology and Bioengineering*, *117*(2), 466–485. https://doi.org/10.1002/bit.27200

Fernandez-cassi, X., Timoneda, N., Martinez-Puchol, S., Rusinol, M., Rodriguez-Manzano, J., Figuerola, N., Bofill-Mas, S., Abril, J. F., & Girones, R. (2018). Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. *Science of the Total Environment*, *618*(8), 870–880. https://doi.org/10.1016/j.scitotenv.2017.08.249

Fong, T.-T., & Lipp, E. K. (2005). Enteric viruses of humans and animals in aquatic environments: health risks, detection, and potential water quality assessment tools. *Microbiology and Molecular Biology Reviews*, *69*(2), 357–371.

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature*, *399*(6736), 541–548. https://doi.org/10.1038/21119

Gay, N., Olival, K. J., Bumrungsri, S., Siriaroonrat, B., Bourgarel, M., & Morand, S. (2014). Parasite and viral species richness of Southeast Asian bats: Fragmentation of area distribution matters. *International Journal for Parasitology: Parasites and Wildlife*, *3*(2), 161–170. https://doi.org/10.1016/j.ijppaw.2014.06.003

Gazitúa, M. C., Vik, D. R., Roux, S., Gregory, A. C., Bolduc, B., Widner, B., Mulholland, M. R., Hallam, S. J., Ulloa, O., & Sullivan, M. B. (2021). Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. *ISME Journal*, *15*(4), 981–998. https://doi.org/10.1038/s41396-020-00825-6

Geoghegan, J. L., & Holmes, E. C. (2017). Predicting virus emergence amid evolutionary

noise. *Open Biology*, *7*(10). https://doi.org/10.1098/rsob.170189

Grandjean, F., Gilbert, C., Razafimafondy, F., Vucić, M., Delaunay, C., Gindre, P., Bouchard, J., Raimond, M., & Moumen, B. (2019). A new bunya-like virus associated with mass mortality of white-clawed crayfish in the wild. *Virology*, *533*(May), 115–124. https://doi.org/10.1016/j.virol.2019.05.014

Grubaugh, N. D., Ladner, J. T., Lemey, P., Pybus, O. G., Rambaut, A., Holmes, E. C., & Andersen, K. G. (2019). Tracking virus outbreaks in the twenty-first century. *Nature Microbiology*, *4*(1), 10–19. https://doi.org/10.1038/s41564-018-0296-2

Gulino, K., Rahman, J., Badri, M., Morton, J., Bonneau, R., & Ghedin, E. (2020). Initial Mapping of the New York City Wastewater Virome. *MSystems*, *5*(3), 1–18. https://doi.org/10.1128/msystems.00876-19

Gutiérrez Aguirre, I., Kutnjak, D., Rački, N., Rupar, M., & Ravnikar, M. (2018). Monolith Chromatography as Sample Preparation Step. In *Viral Metagenomics: Methods and Protocols* (Vol. 1746, pp. 63–75). Springer Science Business Media. https://doi.org/doi: 10.1007/978-1-4939-7683-6_6.

Haramoto, E., Kitajima, M., Kishida, N., Konno, Y., Katayama, H., Asami, M., & Akiba, M. (2013). Occurrence of pepper mild mottle virus in drinking water sources in Japan. *Applied and Environmental Microbiology*, *79*(23), 7413–7418. https://doi.org/10.1128/AEM.02354-13

Hayakijkosol, O., Jaroenram, W., Owens, L., & Elliman, J. (2021). Reverse transcription polymerase chain reaction (RT-PCR) detection for Australian Cherax reovirus from redclaw crayfish (Cherax quadricarinatus). *Aquaculture*, *530*(March 2020), 735881. https://doi.org/10.1016/j.aquaculture.2020.735881

Hayakijkosol, O., & Owens, L. (2011). Investigation into the pathogenicity of reovirus to juvenile Cherax quadricarinatus. *Aquaculture*, *316*(1–4), 1–5. https://doi.org/10.1016/j.aquaculture.2011.03.031

Heijnen, L., & Medema, G. (2011). Surveillance of influenza A and the pandemic influenza A (H1N1) 2009 in sewage and surface water in the Netherlands. *Journal of Water and Health*, *9*(3), 434–442. https://doi.org/10.2166/wh.2011.019

Hjelmsø, M. H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M., Elsässer, D., Aarestrup, F. M., Löfström, C., Bofill-Mas, S., Abril, J. F., Girones, R., & Schultz, A. C. (2017). Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. *PLoS ONE*, *12*(1), 1–17. https://doi.org/10.1371/journal.pone.0170199

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. *Scientific Reports*, *9*(1), 1–8. https://doi.org/10.1038/s41598-019-41830-w

Hudina, S., Kutleša, P., Trgovčić, K., & Duplić, A. (2017). Dynamics of range expansion of the signal crayfish (Pacifastacus leniusculus) in a recently invaded region in Croatia. *Aquatic Invasions*, *12*(1), 67–75. https://doi.org/10.3391/ai.2017.12.1.07

Hudina, S., Žganec, K., Lucić, A., Trgovčić, K., & Maguire, I. (2013). Recent invasion of the karstic river systems in croatia through illegal introductions of the signal crayfish. *Freshwater Crayfish*, *19*(1), 21–27. https://doi.org/10.5869/fc.2013.v19.021

Jeżewska, M., Trzmiel, K., & Zarzyńska-nowak, A. (2018). Detection of infectious tobamoviruses in irrigation and drainage canals in Greater Poland. *Journal of Plant Protection*, *58*, 202–205. https://doi.org/10.24425/119126

Jones, D. L., Baluja, M. Q., Graham, D. W., Corbishley, A., McDonald, J. E., Malham, S. K., Hillary, L. S., Connor, T. R., Gaze, W. H., Moura, I. B., Wilcox, M. H., & Farkas, K. (2020). Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19. *Science of the Total Environment*, *749*(July), 141364. https://doi.org/10.1016/j.scitotenv.2020.141364

Junglen, S., & Drosten, C. (2013). Virus discovery and recent insights into virus diversity in arthropods. *Current Opinion in Microbiology*, *16*(4), 507–513. https://doi.org/10.1016/j.mib.2013.06.005

Khan, A. S., Benetti, L., Blumel, J., Deforce, D., Egan, W. M., Knezevic, I., Krause, P. R., Mallet, L., Mayer, D., Minor, P. D., Neels, P., & Wang, G. (2018). Report of the international conference on next generation sequencing for adventitious virus detection in biologicals. *Biologicals*, *55*(July), 1–16. https://doi.org/10.1016/j.biologicals.2018.08.002

Khan, A. S., Benetti, L., Blumel, J., Deforce, D., Egan, W. M., Knezevic, I., Krause, P. R., Mallet, L., Mayer, D., Minor, P. D., Neels, P., & Wang, G. (2020). Report of the second international conference on next generation sequencing for adventitious virus detection in biologics for humans and animals. *Biologicals*, *55*(May), 1–16. https://doi.org/10.1016/j.biologicals.2018.08.002

Kim, Y., Aw, T. G., Teal, T. K., & Rose, J. B. (2015). Metagenomic investigation of viral communities in ballast water. *Environmental Science and Technology*, *49*(14), 8396–8407. https://doi.org/10.1021/acs.est.5b01633

Kitajima, M., Ahmed, W., Bibby, K., Carducci, A., Gerba, C. P., Hamilton, K. A., Haramoto, E., & Rose, J. B. (2020). SARS-CoV-2 in wastewater: State of the knowledge and research needs. *Science of The Total Environment*, 139076. https://doi.org/10.1016/j.scitotenv.2020.139076

Koenig, R. (1986). Plant viruses in rivers and lakes. *Advances in Virus Research*, *31*, 321–333.

Li, C. X., Shi, M., Tian, J. H., Lin, X. D., Kang, Y. J., Chen, L. J., Qin, X. C., Xu, J., Holmes, E. C., & Zhang, Y. Z. (2015). Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *ELife*, *2015*(4), 1–26. https://doi.org/10.7554/eLife.05378

Lindenfors, P., Nunn, C. L., Jones, K. E., Cunningham, A. A., Sechrest, W., & Gittleman, J. L. (2007). Parasite species richness in carnivores: Effects of host body mass, latitude, geographical range and population density. *Global Ecology and Biogeography*, *16*(4), 496–509. https://doi.org/10.1111/j.1466-8238.2006.00301.x

Lodge, D. M., Deines, A., Gherardi, F., Yeo, D. C. J., Arcella, T., Baldridge, A. K., Barnes, M. A., Lindsay Chadderton, W., Feder, J. L., Gantz, C. A., Howard, G. W., Jerde, C. L., Peters, B. W., Peters, J. A., Sargent, L. W., Turner, C. R., Wittmann, M. E., & Zeng, Y. (2012). Global introductions of crayfishes: Evaluating the impact of species invasions on ecosystem services. *Annual Review of Ecology, Evolution, and*

Systematics, *43*(November), 449–472. https://doi.org/10.1146/annurev-ecolsys-111511-103919

Longshaw, M. (2016). Parasites, Commensals, Pathogens and Diseases of Crayfish. In Stebbing & M. Longshaw P. (Eds.), *Biology and Ecology of Crayfish* (pp. 171–250). Boca Raton: Taylor & Francis Group, LLC.

Ma, H., Galvin, T. A., Glasner, D. R., Shaheduzzaman, S., & Khan, A. S. (2014). Identification of a Novel Rhabdovirus in Spodoptera frugiperda Cell Lines. *Journal of Virology*, *88*(12), 6576–6585. https://doi.org/10.1128/jvi.00780-14

Martínez-Puchol, S., Rusiñol, M., Fernández-Cassi, X., Timoneda, N., Itarte, M., Andrés, C., Antón, A., Abril, J. F., Girones, R., & Bofill-Mas, S. (2020). Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. *Science of the Total Environment*, *713*. https://doi.org/10.1016/j.scitotenv.2020.136604

McClenahan, S. D., & Krause, P. R. (2019). Towards dynamic monitoring of cell cultures using high throughput sequencing. *Vaccine*, *37*(7), 1001–1005. https://doi.org/10.1016/j.vaccine.2018.12.019

Mehle, N., Gutiérrez-Aguirre, I., Prezelj, N., Delić, D., Vidic, U., & Ravnikar, M. (2014). Survival and transmission of potato virus Y, pepino mosaic virus, and potato spindle tuber viroid in water. *Applied and Environmental Microbiology*, *80*(4), 1455–1462. https://doi.org/10.1128/AEM.03349-13

Mehle, Nataša;, & Ravnikar, M. (2012). Plant viruses in aqueous environment - Survival, water mediated transmission and detection. *Water Research*, *46*, 4902–4917. https://doi.org/doi: 10.1016/j.watres.2012.07.027.

Mehle, Nataša, Gutiérrez-Aguirre, I., Kutnjak, D., & Ravnikar, M. (2018). Water-Mediated Transmission of Plant, Animal, and Human Viruses. In *Advances in Virus Research* (pp. 85–128).

Merten, O. W. (2002). Virus contaminations of cell cultures - A biotechnological view. *Cytotechnology*, *39*(2), 91–116. https://doi.org/10.1023/A:1022969101804

Naccache, S. N., Greninger, A. L., Lee, D., Coffey, L. L., Phan, T., Rein-Weston, A., Aronsohn, A., Hackett, J., Delwart, E. L., & Chiu, C. Y. (2013). The Perils of Pathogen Discovery: Origin of a Novel Parvovirus-Like Hybrid Genome Traced to Nucleic Acid Extraction Spin Columns. *Journal of Virology*, *87*(22), 11966–11977. https://doi.org/10.1128/jvi.02323-13

Ng, T. F. F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B. S., Wommack, K. E., & Delwart, E. (2012). High Variety of Known and New RNA and DNA Viruses of Diverse Origins in Untreated Sewage. *Journal of Virology*, *86*(22), 12161–12175. https://doi.org/10.1128/JVI.00869-12

Nicaise, V. (2014). Crop immunity against viruses : outcomes and future challenges. *Frontiers in Plant Science*, *5*(11), 1–18. https://doi.org/10.3389/fpls.2014.00660

Parmar, K., Dafale, N., Pal, R., Tikariha, H., & Purohit, H. (2018). An Insight into Phage Diversity at Environmental Habitats using Comparative Metagenomics Approach. *Current Microbiology*, *75*(2), 132–141. https://doi.org/10.1007/s00284-017-1357-0

Pecman, A., Kutnjak, D., Gutiérrez-Aguirre, I., Adams, I., Fox, A., Boonham, N., & Ravnikar, M. (2017). Next generation sequencing for detection and discovery of plant

viruses and viroids: Comparison of two approaches. *Frontiers in Microbiology*, *8*(October), 1–10. https://doi.org/10.3389/fmicb.2017.01998

Pedrero, F., Kalavrouziotis, I., Alarcón, J. J., Koukoulakis, P., & Asano, T. (2010). Use of treated municipal wastewater in irrigated agriculture-Review of some practices in Spain and Greece. *Agricultural Water Management*, *97*(9), 1233–1241. https://doi.org/10.1016/j.agwat.2010.03.003

Porter, A. F., Shi, M., Eden, J. S., Zhang, Y. Z., & Holmes, E. C. (2019). Diversity and evolution of novel invertebrate DNA viruses revealed by meta-transcriptomics. *Viruses*, *11*(12), 1–21. https://doi.org/10.3390/v11121092

Rastrojo, A., & Alcamí, A. (2017). Aquatic viral metagenomics: Lights and shadows. *Virus Research*, *239*, 87–96. https://doi.org/10.1016/j.virusres.2016.11.021

Rios, G., Lacoux, C., Leclercq, V., Diamant, A., Lebrigand, K., Lazuka, A., Soyeux, E., Lacroix, S., Fassy, J., Couesnon, A., Thiery, R., Mari, B., Pradier, C., Waldmann, R., & Barbry, P. (2021). Monitoring SARS-CoV-2 variants alterations in Nice neighborhoods by wastewater nanopore sequencing. *The Lancet Regional Health - Europe*, *10*, 1–9. https://doi.org/10.1016/j.lanepe.2021.100202

Rosario, K., Nilsson, C., Lim, Y. W., Ruan, Y., & Breitbart, M. (2009). Metagenomic analysis of viruses in reclaimed water. *Environmental Microbiology*, *11*(11), 2806–2820. https://doi.org/10.1111/j.1462-2920.2009.01964.x

Rosario, K., Symonds, E. M., Sinigalliano, C., Stewart, J., & Breitbart, M. (2009). Pepper mild mottle virus as an indicator of fecal pollution. *Applied and Environmental Microbiology*, *75*(22), 7261–7267. https://doi.org/10.1128/AEM.00410-09

Roux, S., Matthijnssens, J., & Dutilh, B. E. (2021). Metagenomics in Virology. *Encyclopedia of Virology*, *January*, 133–140. https://doi.org/10.1016/b978-0-12-809633-8.20957-6

Rupar, M., Ravnikar, M., Tušek-Žnidarič, M., Kramberger, P., Glais, L., & Gutiérrez-Aguirre, I. (2013). Fast purification of the filamentous Potato virus Y using monolithic chromatographic supports. *Journal of Chromatography A*, *1272*, 33–40. https://doi.org/10.1016/j.chroma.2012.11.058

Sadeghi, M., Altan, E., Deng, X., Barker, C. M., Fang, Y., Coffey, L. L., & Delwart, E. (2018). Virome of >12 thousand Culex mosquitoes from throughout California. *Virology*, *523*(August), 74–88. https://doi.org/10.1016/j.virol.2018.07.029

Sánchez-Paz, A. (2010). White spot syndrome virus: An overview on an emergent concern. *Veterinary Research*, *41*(6). https://doi.org/10.1051/vetres/2010015

Santos, P. D., Ziegler, U., Szillat, K. P., Szentiks, C. A., Strobel, B., Skuballa, J., Merbach, S., Grothmann, P., Tews, B. A., Beer, M., & Höper, D. (2021). In action - An early warning system for the detection of unexpected or novel pathogens. *Virus Evolution*, *7*(2), 1–15. https://doi.org/10.1093/ve/veab085

Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., Qin, X. C., Li, J., Cao, J. P., Eden, J. S., Buchmann, J., Wang, W., Xu, J., Holmes, E. C., & Zhang, Y. Z. (2016a). Redefining the invertebrate RNA virosphere. *Nature*, *540*(7634), 539–543. https://doi.org/10.1038/nature20167

Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., Qin, X. C., Li, J., Cao, J. P., Eden, J. S., Buchmann, J., Wang, W., Xu, J., Holmes, E. C., & Zhang, Y. Z.

(2016b). Redefining the invertebrate RNA virosphere. *Nature*, *540*(7634), 539–543. https://doi.org/10.1038/nature20167

Shkoporov, A. N., Ryan, F. J., Draper, L. A., Forde, A., Stockdale, S. R., Daly, K. M., Mcdonnell, S. A., Nolan, J. A., Sutton, T. D. S., Dalmasso, M., Mccann, A., Ross, R. P., & Hill, C. (2018). Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome*, *6*(68), 1–17. https://doi.org/https://doi.org/10.1186/s40168-018-0446-z

Steyer, A., Gutiérrez-Aguirre, I., Rački, N., Beigot Glaser, S., Brajer Humar, B., Stražar, M., Škrjanc, I., Poljšak-Prijatelj, M., Ravnikar, M., & Rupnik, M. (2015). The detection rate of enteric viruses and Clostridium difficile in a waste water treatment plant effluent. *Food and Environmental Virology*, *7*(2), 164–172. https://doi.org/10.1007/s12560-015-9183-7

Suttle, C. A. (2016). Environmental microbiology: Viral diversity on the global stage. *Nature Microbiology*, *1*(11). https://doi.org/10.1038/nmicrobiol.2016.205

Symonds, E. M., Rosario, K., & Breitbart, M. (2019). Pepper mild mottle virus : Agricultural menace turned effective tool for microbial water quality monitoring and assessing (waste) water treatment technologies. *PLoS Pathogens*, *15*(4), 1–7. https://doi.org/https://doi.org/10.1371/journal. ppat.1007639

Twardochleb, L. A., Olden, J. D., & Larson, E. R. (2013). A global meta-analysis of the ecological impacts of nonnative crayfish. *Freshwater Science*, *32*(4), 1367–1382. https://doi.org/10.1899/12-203.1

Van der Veer, G., & Nentwig, W. (2015). Environmental and economic impact assessment of alien and invasive fish species in Europe using the generic impact scoring system. *Ecology of Freshwater Fish*, *24*(4), 646–656. https://doi.org/10.1111/eff.12181

Victoria, J. G., Wang, C., Jones, M. S., Jaing, C., McLoughlin, K., Gardner, S., & Delwart, E. L. (2010). Viral Nucleic Acids in Live-Attenuated Vaccines: Detection of Minority Variants and an Adventitious Virus. *Journal of Virology*, *84*(12), 6033–6040. https://doi.org/10.1128/jvi.02690-09

Webber, Q. M. R., Fletcher, Q. E., & Willis, C. K. R. (2017). Viral Richness is Positively Related to Group Size, but Not Mating System, in Bats. *EcoHealth*, *14*(4), 652–661. https://doi.org/10.1007/s10393-017-1276-3

WHO. (2013). Recommendations for the evaluation of animal cell cultures as substrates for the manufacture of biological medicinal products and for the characterization of cell banks Replacement of Annex 1 of WHO Technical Report Series, No. 878. In *WHO technical report series; no.978* (Vol. 978, Issue Annex 3, pp. 79–187).

Wolf, Y. I., Silas, S., Wang, Y., Wu, S., Bocek, M., Kazlauskas, D., Krupovic, M., Fire, A., Dolja, V. V., & Koonin, E. V. (2020). Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. *Nature Microbiology*, *5*(10), 1262–1270. https://doi.org/10.1038/s41564-020-0755-4

Wong, K., Fong, T. T., Bibby, K., & Molina, M. (2012). Application of enteric viruses for fecal pollution source tracking in environmental waters. *Environment International*, *45*(1), 151–164. https://doi.org/10.1016/j.envint.2012.02.009

Wu, H., Pang, R., Cheng, T., Xue, L., Zeng, H., Lei, T., Chen, M., Wu, S., Ding, Y., Zhang, J., Shi, M., & Wu, Q. (2020). Abundant and Diverse RNA Viruses in Insects

Revealed by RNA-Seq Analysis: Ecological and Evolutionary Implications. *MSystems*, *5*(4). https://doi.org/10.1128/msystems.00039-20

Xagoraraki, I., & Brien, E. O. (2019). Wastewater-based epidemiology for early detection of viral outbreaks. In *Women in Water Quality* (pp. 75–97). https://doi.org/https://doi.org/10.1007/978-3-030-17819-2_5

Zhang, T., Breitbart, M., Lee, W. H., Run, J. Q., Wei, C. L., Soh, S. W. L., Hibberd, M. L., Liu, E. T., Rohwer, F., & Ruan, Y. (2006). RNA viral community in human feces: Prevalence of plant pathogenic viruses. *PLoS Biology*, *4*(1), 108–118. https://doi.org/10.1371/journal.pbio.0040003

Zhang, Y. Z., Shi, M., & Holmes, E. C. (2018). Using Metagenomics to Characterize an Expanding Virosphere. *Cell*, *172*(6), 1168–1172. https://doi.org/10.1016/j.cell.2018.02.043

Zhang, Z.-Q. (2013). Phylum Arthropoda. In *Animal Biodiversity: An Outline of Higher-level Classification and Survey of Taxonomic Richness* (Vol. 3703, Issue 1, p. 17). Zootaxa. https://doi.org/http://dx.doi.org/10.11646/zootaxa.3703.1.6

Zinger, L., Gobet, A., & Pommier, T. (2012). Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular Ecology*, *21*(8), 1878–1896. https://doi.org/10.1111/j.1365-294X.2011.05362.x

Bibliography

Publications Related to the Thesis

Journal Articles

Bacnik, K., Kutnjak, D., Pecman, A., Mehle, N., Tusek Znidaric, M., Gutierrez Aguirre, I., & Ravnikar, M. (2020). Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment. *Water Research*, *177*, 115628. https://doi.org/10.1016/j.watres.2020.115628

Bačnik, K., Kutnjak, D., Jerič Kokelj, B., Tuta, N., Lončar, T., Vogelsang, M., & Ravnikar, M. (2021). Metagenomic characterization of parental and production CHO cell lines for detection of adventitious viruses. *Biologicals*, *69*(November 2020), 70–75. https://doi.org/10.1016/j.biologicals.2020.11.001

Bačnik, K., Kutnjak, D., Černi, S., Bielen, A., & Hudina, S. (2021). Virome analysis of signal crayfish (Pacifastacus leniusculus) along its invasion range reveals diverse and divergent rna viruses. *Viruses*, *13*(11). https://doi.org/10.3390/v13112259

Biography

Katarina Bačnik was born on December 1990 in Ljubljana, Slovenia. She graduated in 2017 from Microbiology at the Biotechnical Faculty, University of Ljubljana, Slovenia. During her studies in 2015, she did an internship at the Institut für Biologie, Albert-Ludwigs-Universität Freiburg, Germany, where she was studying enzymes involved in CO_2 fixation in autotrophic bacteria and the results were published in the Science journal. In 2016, she did another internship at the Institut für Microbiologie, Universität Innsbruck, Austria, working with black soldier fly (*Hermetia illucens*) larvae and their gut microbiota. Katarina did her master thesis at the National Institute of Biology (NIB), Department for Biotechnology and Systems Biology, where she tested the infectivity of tobamoviruses in influent and effluent samples from wastewater treatment plant.

In 2017, she enrolled in the PhD study programme of Sensor Technologies at the Jožef Stefan International Postgraduate School (IPS) and started working as a young researcher at NIB, Department of Biotechnology and Systems Biology under the supervision of Prof. Dr. Maja Ravnikar and Dr. Denis Kutnjak. She was a member of the Student Council (2017-2020) and took active part in the organization of yearly IPS student conferences.

Her research work focuses on the detection of viruses and virus diversity in different complex samples from the environment and industry using high-throughput sequencing (HTS) technologies. During the PhD studies, she did a short-term scientific mission at the University of Murcia, CEBAS - CSIC, Spain, comparing procedures used for infectivity tests of plant viruses. Katarina presented her work on national and international conferences, where she also won several awards for best presentations, e.g. for the best oral presentation at the International Society for Food & Environmental Virology Conference 2018 at Arizona State University, Tempe, Arizona. In 2020, she also received an award for young microbiologists for excellent research achievements from the Slovenian Microbiological Society.

Since the COVID-19 epidemic started, Katarina has been involved in an ongoing wastewater monitoring for SARS-CoV-2 and its variants in wastewater from Slovenian wastewater treatment plants. She is currently also involved in the Slovenian Research Agency project (L4-3179).