UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escola Tècnica Superior d'Enginyeria Agronòmica i del Medi Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Agronómica y del Medio Natural

Análisis comparativo de diferentes aproximaciones experimentales para la obtención y secuenciación del viroma fecal de pacientes portadores del Síndrome de Lynch

Trabajo Fin de Grado

Grado en Biotecnología

AUTOR/A: Soria Villalba, Adriana

Tutor/a: Jantus Lewintre, Eloisa

Cotutor/a externo: MOYA SIMARRO, ANDRES

Director/a Experimental: PEREZ BROCAL, VICENTE

CURSO ACADÉMICO: 2022/2023

Comparative analysis of different experimental approaches for obtaining and sequencing the fecal virome of Lynch syndrome carrier's patients

Adriana Soria Villalba

Tutor: Vicente Pérez Brocal
Cotutor: Andrés Moya Simarro
UPV tutor: Eloisa Jantus Lewintre

ETSIAMN: Escuela Técnica Superior de Ingeniería Agronómica y del Medio Natural
UPV: Universitat Politècnica de València
Biotechnology

Final Degree Project (FDP)
Academic course: 2022/2023

Valencia, 4th of June of 2023

TITLE: Comparative analysis of different experimental approaches for obtaining and sequencing the fecal virome of Lynch syndrome carrier's patients

---

TÍTULO: Análisis comparativo de diferentes aproximaciones experimentales para la obtención y secuenciación del viroma fecal de pacientes portadores del Síndrome de Lynch

---

TÍTOL: Anàlisi comparatiu de diferents aproximacions experimentals per a l'obtenció i seqüenciació del viroma fecal de pacients portadors del síndrome de Lynch

Abstract:

Recently, several findings have highlighted the crucial role of human virome in health and host's disease, as well as its potential as future diagnosis marker or even therapy. As this field has expanded, limitations have been appearing. One of the most important is related to viral nucleic acid extraction protocol as diverse experiments have pointed out that the extraction method is highly influencing the results, determining the quantity and quality of the RNA and DNA analyzed. Then, it is demanding to find an accurate extraction method for human virome that ensures reproducible and faithful results. Following this idea, three alternative extraction protocols have been tested on fecal samples from several volunteers genetically diagnosed with Lynch syndrome, a hereditary condition that increases the risk of certain cancers, such as colorectal cancer. Two of them being highly specific for viruses, involving an enrichment of virus–like particles in the sample prior to RNA and DNA extraction. The other one being a bulk method enabling the extraction of the whole microbiome, including viruses. The aim of this work is to assess which of these extraction methods optimizes the sequencing results in terms of viral reads yield to be able to select the most efficient one for its implementation in the laboratory research line.

This project is related to the following Sustainable Development Goals (SDG): Good health and well-being (SDG 3), Quality education (SDG 4) and Reduced inequalities (SDG 10).

Key words: virome, Lynch syndrome, VLP, bulk metagenomics, NetoVIR, SISPA, extraction protocol, sequencing

---

Resumen:

Recientemente, varios hallazgos han destacado el papel crucial del viroma humano en la salud y enfermedad del huésped, así como su potencial como marcador de diagnóstico e incluso terapia. A medida que este campo se ha ido expandiendo, han surgido varias limitaciones. Una de las restricciones más importantes está relacionada con el protocolo de extracción de ácidos nucleicos virales. Diversos experimentos han señalado que el método de extracción influye en gran medida en los resultados, determinando la cantidad y calidad del ARN y ADN analizados. Por tanto, es necesario encontrar un método de extracción de viroma humano preciso, que garantice resultados reproducibles y fiables. Siguiendo esta línea de pensamiento, en este trabajo fin de grado se han propuesto y probado tres protocolos de extracción alternativos en muestras fecales de varios voluntarios diagnosticados genéticamente con el síndrome de Lynch,

una condición hereditaria que aumenta el riesgo de ciertos tipos de cáncer, como el cáncer colorrectal. Siendo dos de ellos altamente específicos para virus, implicando un enriquecimiento de '*virus-like particles*' en la muestra previo a la extracción de ARN y ADN. El otro método siendo más general, permitiendo la extracción de la población total microbiana, incluyendo virus. El objetivo de este trabajo es evaluar cuál de estos métodos de extracción optimiza los resultados de secuenciación, en términos de rendimiento de lecturas virales, para poder seleccionar la alternativa más eficiente para su posterior implementación en la línea de investigación del laboratorio.

Este proyecto se relaciona con los siguientes Objetivos de Desarrollo Sostible (ODS): Salud y bienestar (ODS 3), Educación de calidad (ODS 4) y Reducción de las desigualdades (ODS 10).

Palabras clave: viroma, síndrome de Lynch, VLP, *'bulk metagenomics'*, NetoVIR, SISPA, protocolo de extracción, secuenciación

Acknowledgments:

I would like to express my deepest gratitude to my thesis advisors Vicente Pérez, Andrés Moya and Eloisa Jantus, for their guidance, support, and invaluable feedback throughout the entire process. I would also like to thank Nicole Stephanía Pesantes, the PhD student who has been at the forefront with me during the whole project, Nuria Jimenez, who oversaw the last steps of the experiment, and Javi Pons, computer expert in charge of the sequencing part.

I would also like to thank my family, who has supported me and guided me in each step of my academic career as well as in each aspect of my life. Especially to my mum and Juanjo, thanks for encouraging me to choose this college career.

Finally, to my university friends: Any, Alex, Juanto, Maria, Carmen, Cris, Miguel and Alejandro (Capi), who make my life so much easier and happier. Without them, I would have not been as delighted with my university stay as I have. I would keep you and what we have lived in my heart no matter what.

"Y por último agradecerte a ti, abuelo. Aunque no hayas podido verme graduándome, sé que estás conmigo hoy y siempre"

INDEX

1. INTRODUCTION

1.1 HUMAN VIROME

We all have heard about the colonizing human microorganisms, the microbiome, but fewer people know that healthy humans are also colonized by a remarkable diversity of viruses collectively known as human virome. The human virome is composed mainly by phages, viruses infecting other cellular microorganisms; viruses infecting human cells and transient viruses. It is suspected that the diversity of the human virome is comparable to the diversity of viruses on Earth. Taking into account that viral agents are thought to be one of the most rich and diverse entities in our planet, the human virome diversity should not be taken lightly. Nonetheless, this heterogeneity has not been studied deep enough. Only nowadays, with new extractions protocols and detection methods this broad field of study appears to be emerging (Liang & Bushman, 2021).

Regarding the human virome diversity, it is well known the variety among viruses and their multiple classification systems, based on viral genomes (ssRNA, dsRNA, ssDNA or dsDNA viruses), genome sizes, shape of the capsid and the presence or absence of lipid membranes enclosing the viral entity. Furthermore, the VLP can be also classified based on their morphology (Liang & Bushman, 2021).

The most abundant viral entities in the human virome are thought to be resident bacteriophages, who shape the microbial communities based on predation and horizontal gene transfer. The second most extensive group corresponds to viruses that infect human cells and last but not least, the transient viruses which can be long term passengers or commensals, such as those associated with diet (e.g. plant viruses). However, when performing metagenomics studies, the majority of sequences (up to 90%) do not share homology with any references present in existing databases, so it cannot be stated whether those sequence reads come from bacteriophages and/or other uncharacterised viruses. This issue is aggravated by the fact that viruses, opposite to prokaryotes and eukaryotes, lack universal marker genes (i.e prokaryotic 16S rRNA gene) that facilitate their taxonomic classification (Shkoporov et al., 2019; Liang & Bushman, 2021).

What has been demonstrated is that, despite the fact that several studies have shown great inter-individual variation, the virome in a healthy individual remains stable over time, so the intrapersonal variation is rather low (Minot el at, 2011; Shkoporov et al, 2019). Although it is noteworthy to point out that although individual intestinal viromes are unique, some phages can be shared among individuals. This interindividual heterogeneity is due to the location in the body and other factors such as diet, genetics and geography (Carding et al., 2017; Liang & Bushman, 2021) .

For detection and estimation of viral concentration in the different ecosystems, the unit of measurement chosen are the virus-like particles (VLPs). VLPs are subviral molecules that mimic the structure of the viruses but they are unable to replicate. Talking about VLP instead of virus can be convenient because VLP can indicate the presence of viruses even when those are not actively replicating or causing a disease, meaning they show the presence of viruses even when those have entered a latency state. Moreover, as it has been previously said, the human virome is not sufficiently well characterized yet so detecting VLPs could help to identify novel viruses.

Among the different body sites rich in viral populations, we can enumerate the gastrointestinal tract, the oral cavity, the respiratory tract, the blood, the skin, the urogenital system and the nervous system, among others. We will now comment on the most important aspects of the gastrointestinal tract, as it constitutes the richest in diversity and the most abundant site for virus and bacteria colonization (Liang & Bushman, 2021).

In the final section of the gastrointestinal tract, that is, the large bowel, from 10^7 to almost 10^9 VLP per gram of intestinal content can be found . This is why it is the most chosen environment from which to extract samples, feces samples, and perform experiments about microbiome and virome diversity and characterization. It is believed that a link exists between bacterial and viral composition in the gut. The most extensively reported taxa are temperate bacteriophages, which contribute to maintain homeostasis in this environment, but are also able to provoke a microbial imbalance (dysbiosis) that could lead to the development of diseases, highlighting the importance of phages-bacteria-host interaction in host and viral evolution. It has been estimated that phage:bacteria composition ratios should be at least 1:1, the abundance of bacteriophages running in parallel with the abundance of their preyed upon bacteria (Ogilvie et al., 2015). In this context, the most abundant microbes in the gastrointestinal tract are Proteobacteria, Actinobacteria, Firmicutes and Bacteroidetes, in this order (Human Microbiome Jumpstart Reference Strains Consortium, 2010). Therefore, the most abundant bacteriophages are those infecting these classes of bacteria, such as crassPhages, *Siphoviridae*, *Myoviridae*, *Microviridae, etc* (Siranosian et al., 2020; Bhardwa et al., 2022). Phages are followed, in terms of abundance, by human-infecting viruses. Some RNA viral families that can be found on the gastrointestinal system are *Rotavirus, Astrovirus, Norovirus, Adenovirus*… while the most abundant DNA viral family found on the organisms correspond to *Anelloviridae* (Kourí, 2018). An increasing number of studies have reported a clear relationship between disease-specific alterations in the gut virome and gut-disorders or even systemic diseases (Carding et al., 2017; Shkoporov et al., 2019; Liang & Bushman, 2021).

1.2 HUMAN VIROME AND CANCER

The dysbiosis, or microbial imbalance has been proved to be involved in the development of chronic diseases, infectious diseases and carcinogenesis. Currently, as the human virome field grows and expands, its implications in carcinogenesis are being brought to light. There are multitudes of viruses, resident or transient, that either directly or indirectly can be involved in the process of tumorigenesis. Raising some interesting ideas about how virome, mainly composed of phages, can influence and modify the microbiome, by altering the relationship predator-prey that phages have with their respective bacterial hosts (Stern et al., 2019).

Integration of some viruses in the host genome is the main mechanism for cancer progression. And although integration is typical of retroviruses, it has been also observed in non-retroviruses thanks to alternative mechanisms such as homology recombination. Nonetheless, other viral influences in tumorigenesis are vital too, for instance the presence of oncogenes, genotoxins or inflammation. The viruses that contribute to carcinogenesis are collectively known as oncoviruses (Issaeva et al., 2018; Stern et al., 2019).

The oncoviruses attack key cellular targets, modifying their function and altering cellular biology as a whole. Usually, those targets are molecules with an important role in signaling pathways. Viral integrational events or specific gene expression changes can be an indicator of malignant viral action (Issaeva et al., 2018). Indeed, some studies (White, 2014; Gaglia, & Munger, 2018) have demonstrated the existing link between the presence of some viruses and specific cancer types. For instance, Epstein-Barr (EBV) virus is related to some stomach cancers and lymphomas, or Human Papillomavirus (HPV) which is involved in cervix, anus, penis, throat, vagina and vulva tumors. It is thought that those viral infections contribute to around 12% of all human cancers. Nonetheless, it should take into consideration the extreme complexity of human virome and not reduce it to a few oncovirus.

Regarding this complexity, an interesting study was carried out by a team of researchers in the university of Pittsburgh (Issaeva et al., 2018) in which a total of 22 different cancers from more than 3,000 individuals were analyzed using Next Generation Sequencing (NGS) and the conclusions they drafted were that in nearly all tumor and normal samples taken there was presence of viral sequences. At the end they could classify those viral sequences into 34 different viruses of five large virus families to finally relate each family to the most common cancer in which they could have a role.

1.3 LYNCH SYNDROME

Lynch syndrome (LS), also known as hereditary non-polyposis colorectal cancer (HNPCC), is one of the most frequent hereditary cancer syndromes. It is an autosomal dominant hereditary condition. People suffering from this condition are predisposed to develop a wide variety of cancers during their lives, with increased risk of colorectal (CRC) and endometrial cancer (EC). The prevalence of this syndrome is quite notorious as it has been found 1 case out of every 35 patients with CRC is caused by this disease, besides, 1 out of 56 cases of EC is due to LS (Moya et.al, 2017; Biller et al., 2019; Roudko et al., 2021).

The molecular pathogenesis of the LS is based on germline mutations in key proteins of the DNA mismatch repair (MMR) system leading to an unfeasible microsatellite instability. Those mutations are usually related to pathogenic variants of MLH1 and MSH2, as well as in MSH6 and PMS2. MLH1 and MSH2 are the two MMR genes more linked to LS in CRC patients, while the gene silencing of MSH2 by the loss of heterozygosity is one of the most common causes. This mutation interferes with the Epithelial Cells Adhesion Molecules also known as EPCAMs, that is the main reason why, in many articles, EPCAM are included regarding the germline inactivation of one of the alleles of the most important genes. Depending on the mutated MMR gene, significant differences in the clinical phenotype of patients have been described, evidencing huge genetic heterogenicity in Lynch syndrome (Moya et.al, 2017; Yurgelun & Hampel, 2018; Biller et al., 2019; Roudko et al., 2021).

Lynch syndrome is characterized by incomplete penetrance and variable expressivity with its appearance at early age, below 45 years, of mainly CRC and EC tumors, although other organs can be affected too. In patients with this hereditary syndrome, it is vital to initiate screening programs sooner and with more frequency than the general population. It is true that universal tumor screening is a powerful tool to detect LS, it is only available for patients who have already developed a tumor and biopsy or tumor sample is available or possible. However, if a clinical suspicion exists and/or the clinical

history points out towards this syndrome, the patients need to be incorporated into routinary screenings. At least until the ultimate diagnosis is completed, normally by germline or genetic testing. After the validation of the syndrome, the identification of the pathogenic variant or mechanism responsible for the disease is carried out with the final goal of selecting the best therapy to benefit the patient, for instance immune checkpoints, inhibitors, immunotherapy… Finally, after the treatment, a follow up is required (Moya et.al, 2017; Yurgelun & Hampel, 2018; Biller et al., 2019).

In a promising future, it is expected to diagnose LS in healthy, cancer-free individuals , rather than identify LS patients once they have developed some kind of tumor. This improvement in early diagnosis could offer the perfect possibility to reduce the morbidity and mortality of this familiar tumorigenic syndrome (Yurgelun & Hampel, 2018).

In order to expand the progress in our understanding regarding the Lynch syndrome and other hereditary cancer types, it is highly recommendable to take a look at those patients' microbiome and virome as several studies have found a correlation between the gut microbiome and CRC (Arthur & Jobin, 2011; Arthur et al., 2012; Moya et al., 2017). It has been intensively studied the role of the intestinal microbiota in CRC and several pathogenic models have been described trying to illustrate this complex interaction demonstrating, in murine models, functional interactions between microbiome, diet and mismatch repair system deficiency (Belcheva et al., 2014; Belcheva & Martin, 2014; Cheng et al., 2020). Nevertheless, there are not many investigations about gut virome and CRC, not because its role is not significant, but because viruses have not appeared in the picture until now. Nowadays, interesting questions arise: e.g. Is the intestinal microbiome and virome different in a sporadic CRC patient compared to a LS patient that develops CRC? Are there significant similarities and variations in healthy patients compared to cancer-free genetically diagnosed LS individuals? Could the virome and/or the microbiome of a person be an indicator of LS? Or instead of a proper biomarker, could the virome be a promising target for LS and tumor therapy?...

1.4 NUCLEIC ACID EXTRACTION PROTOCOLS

The arrival of metagenomic sequencing at the beginning of the 1980's opened a window for the field of microbiomics and viromics. Furthermore, with the continuous development of sequencing technologies and databases, genomics as a discipline established itself as one of the most promising lines of work in science for the next few years. The first reports that mention viral populations are dated in 2002 and were fulfilled by Mya Breitbart and her colleagues. Thanks to the community efforts of all the researchers that have preceded us, today several viral metagenomic approaches are being applied to explore and explode the hidden potential of the human virome (Breitbart et al., 2002; The New Science of Metagenomics, 2007; Bai et al., 2022). However, metagenomics and NGS analysis give an incomplete profile of the viruses that can be present in the samples or in the body. One of the key steps in which improvements can be incorporated to draw clearer and more precise conclusions is in the nucleic acid extraction part. Indeed, the reliability of the results in nucleic acid detection and analysis depends largely on the accuracy of the nucleic acid extraction, in turn based on their optimal isolation, for which removal of contaminants is crucial.

This issue is more significant regarding extraction of viral nucleic acids, as their amount in a sample tends to be extremely low compared to that of bacterial genomes, due to their difference in genome size. For this reason, additional amplification steps may be required for their final quantification. Consequently, nucleic acid isolation methods that minimize the presence of contaminants and that maximize the amount of DNA or RNA from viral particles are needed. As the presence of host nucleic acids or microbial nucleic acids can interfere with the quantification and characterization of viral sequences, making it difficult to accurately assess the viral load and diversity in a sample (Horz et al., 2010). Regarding this concern, although viromics is a field currently prospering, another limitation is present, the physical management of VLP. Obtaining VLP from a sample can be challenging considering that VLPs, and viral particles in general, can be fragile and easily disrupted during extraction, which can lead to the loss of viral nucleic acids and affect the accuracy of downstream analyses (Lewandowski et al., 2017; Roux et al., 2017). Therefore, it is important to optimize the extraction and purification methods to achieve reliable and accurate virome sequencing data.

Furthermore, many viruses are RNA viruses. RNA is a very unstable molecule and special cautions need to be taken when managing this molecule to avoid the presence of RNAses, which requires a RNAse-free procedure (Tan & Yiap, 2009). In this scenario, untrue depletion of contamination makes a huge difference and influences the results in a great manner. Because, although impurities can be removed from the final raw data and the profiling is normally not compromised, it is hypothesized that there is a great amount of undiscovered virus present in the human virome that can have vital roles in carcinogenesis and in the general health of the individual. Thus, this contamination in the extraction procedure could be prevented from discovering novel human infectious agents.

To solve this issue, several types of nucleic acid extraction methods have been developed. This broad field can be classified, firstly, on automated or manual procedures.

On one hand, current technology has reached everywhere, even at the laboratory. Automated extraction systems are used quite often as nucleic acid extraction techniques. A survey was performed in which five different automated systems for the extraction of viral RNA and DNA were compared. Although the detection rates of viral nucleic acids were similar in all of them, the researchers detected discrepant results in quantification as well as inhibitions due to interfering substances. Moreover, the automated platforms required pre-processing manual steps usually prepared by laboratory staff with a high level of technical expertise and the robots are fairly expensive (Tan & Yiap, 2009; Verheyen et al., 2012; Yang et al., 2020).

On the other hand, manual extraction methods are still very frequent in the laboratory as they are cheaper and not great technical mastery is required. Nonetheless, these procedures are more laborious and slower. An intriguing experiment was fulfilled, in which two manual extraction protocols and two automated platforms were compared. The results demonstrated that no significant differences were found regarding the sensitivity of the methods (Witlox et al., 2008; Yang et al., 2020).

Furthermore, another classification involves the specificity of the method, giving rise to the dichotomy of bulk metagenomes versus VLP approach.

Firstly, in bulk metagenomics, the DNA from the entire community of microorganisms present in a tangled sample is extracted. Whereas, the VLP approach is targeted, it focuses on specific groups present in the community, for example, the DNA of just the viral particles is extracted. It is still unknown whether for viral nucleic acid extraction it is better to choose one or other, for this reason, some advantages and drawbacks are listed below.

While in bulk extraction methods, there is an increased risk of contamination, larger quantities of nucleic acids can be achieved. As 'everything' gets to be sequenced, it is an easier and quicker method, besides, it is less biased. Nonetheless, computationally, it requires more sequence depth and the performance is reduced due to the presence of host and bacteria contaminants. On the contrary, when selecting a virus-specific extraction protocol, lower quantification values will be obtained but purity will be enlarged. VLP-enrichment strategy requires less sequencing depth, but it is a more tedious method. Additionally, it could be biased as non-encapsulated intracellular elements are lost, whereas this is not the case in bulk procedures (Gregory et al., 2020).

In conclusion, more research is required to define the gold standard viral nucleic acid extraction protocol that allows for a substantial pure extraction and reproducible results.

2. OBJECTIVES

In this context, the objectives of this work are:

The global objective of this final degree project (FDP) is to test and compare the effectiveness, advantages and limitations of three different experimental procedures for viral nucleic acid collection from fecal samples from a small set volunteers genetically diagnosed with Lynch syndrome, in order to identify the optimal method for its subsequent application as the method of choice to carry out the extractions and subsequent analyses of the virome on a larger scale, e.g. on the rest of patients with LS, as well as for other studies that make use of fecal samples for the determination of the virome. Additionally, the reproducibility of each technique will also be assessed by comparing the homogeneity of the results using two replicates from each sample and for each method.

For the achievement of the general objective, specific aims are:

1. To process the previously collected fecal samples for virome sequencing using different methods for sample preparation, nucleic acid extraction, amplification, and sequencing.
2. To evaluate computationally the quality and quantity of the sequencing output data in order to compare the yield of viral reads, obtained from those different experimental approaches in terms of absolute figures, but also as ratio compared to non-viral reads, determining the feasibility of each approach for detecting and characterizing viral sequences in fecal samples of Lynch syndrome patients.
3. To provide recommendations for the optimal experimental approach for obtaining and sequencing the fecal virome based on a comparative analysis of the different methods and their potential clinical applications.

By accomplishing these objectives, this project will contribute to the creation of a solid basis for upcoming viral extraction procedures, analysis and sequencing, to eventually help elucidate a relationship between human gut virome composition and diversity and health/ disease status like the familiar genetic condition related to carcinogenesis tackled in this project.

3. MATERIALS AND METHODS

3.1 VOLUNTEERS AND BIOLOGICAL SAMPLES

Stool samples had been previously collected from a cohort of recruited volunteers participating in a project founded by the Asociación Española contra el Cáncer (AECC) awarded to Prof Andrés Moya (call 2017). The original project, titled "*Impact of the microbiota on colorectal oncogenesis in patients with Lynch syndrome*" intended to establish and follow-up a cohort of 120 individuals diagnosed with LS and mutations in some of the genes responsible for it to carry out the functional analysis of the gut microbiota (metagenome, metatranscriptome, metabolome, and metaproteome) and its association with the eventual development of CRC in individuals with a genetic diagnosis of LS. The volunteers were consulted about their informed consent as it was the ethical committee, which approved the project.

Stool samples had been collected from each subject using a collection kit containing instructions and 10-mL collection tubes with 2 mL of stabilizer agent (RNA-later), sent to the patient's home, via courier. After collecting the biological material, the courier service delivered it to FISABIO, where the stool samples were preprocessed to isolate the bacterial fraction by adding 1 volume of sterile PBS to the RNA-later, homogenization, and soft centrifugation. The supernatant had been transferred to sterile tubes and aliquoted into 2 mL microtubes for storage at -70ºC until use.

From a total of 44 individuals that had been recruited, whose feces had been collected at several time points (every three months), ten subjects at a single time point (T09) were selected to carry out the present FDP. In total, 72 samples, including the negative controls, were analyzed. The experimental design of the project is shown in Table 1. The five initial alphanumeric characters "MLxxxx" stand for the patient's code; the next characters "T09" stand for the time point of the sample collection (ninth sampling); the next digit stands for the method (.1 Bulk metagenomics, .2 NetoVir, .3 SISPA); and the last character stands for the replicate (A or B). "CN" stands for extraction negative control, "CN-RT" stands for the retrotranscription negative control and lastly, "CN-PCR" for the negative control of the final amplification.

Table 1. Initial experimental design of the project

	LS PATIENT	PROTOCOL	REPLICATE		LS PATIENT	PROTOCOL	REPLICATE
1	**ML002-T09**	BULK METAGENOMICS	ML002-T09.1A	7	**ML009-T09**	BULK METAGENOMICS	ML009-T09.1A
			ML002-T09.1B				ML009-T09.1B

		NETOVIR	ML002-T09.2A			NETOVIR	ML009-T09.2A
			ML002-T09.2B				ML009-T09.2B
		SISPA	ML002-T09.3A			SISPA	ML009-T09.3A
			ML002-T09.3B				ML009-T09.3B
2	**ML003-T09**	BULK METAGENOMICS	ML003-T09.1A	8	**ML013-T09**	BULK METAGENOMICS	ML013-T09.1A
			ML003-T09.1B				ML013-T09.1B
		NETOVIR	ML003-T09.2A			NETOVIR	ML013-T09.2A
			ML003-T09.2B				ML013-T09.2B
		SISPA	ML003-T09.3A			SISPA	ML013-T09.3A
			ML003-T09.3B				ML013-T09.3B
3	**ML005-T09**	BULK METAGENOMICS	ML005-T09.1A	9	**ML020-T09**	BULK METAGENOMICS	ML020-T09.1A
			ML005-T09.1B				ML020-T09.1B
		NETOVIR	ML005-T09.2A			NETOVIR	ML020-T09.2A
			ML005-T09.2B				ML020-T09.2B
		SISPA	ML005-T09.3A			SISPA	ML020-T09.3A
			ML005-T09.3B				ML020-T09.3B
4	**ML006-T09**	BULK METAGENOMICS	ML006-T09.1A	10	**ML022-T09**	BULK METAGENOMICS	ML022-T09.1A
			ML006-T09.1B				ML022-T09.1B
		NETOVIR	ML006-T09.2A			NETOVIR	ML022-T09.2A
			ML006-T09.2B				ML022-T09.2B
		SISPA	ML006-T09.3A			SISPA	ML022-T09.3A
			ML006-T09.3B				ML022-T09.3B
5	**ML007-T09**	BULK METAGENOMICS	ML007-T09.1A		**CN extraction**	BULK METAGENOMICS	CN.1A
			ML007-T09.1B				CN.1B

	NETOVIR	ML007-T09.2A		NETOVIR	CN.2A
		ML007-T09.2B			CN.2B
	SISPA	ML007-T09.3A		SISPA	CN.3A
		ML007-T09.3B			CN.3B
6 **ML008-T09**	BULK METAGENOMICS	ML008-T09.1A	**CN RT**	NETOVIR	CN-RT A
		ML008-T09.1B			CN-RT B
	NETOVIR	ML008-T09.2A		SISPA	CN-RT A
		ML008-T09.2B			CN-RT B
	SISPA	ML008-T09.3A	**CN PCR**	SISPA	CN-PCR A
		ML008-T09.3B			CN-PCR B

3.2 EXTRACTION PROTOCOL 1: BULK METAGENOMICS

The first extraction protocol is modified from QIAamp ®Fast DNA Stool Mini Handbook 02/2020 (pages 24-26). It consists of a bulk DNA extraction protocol from which both gut microbiome DNA and gut virome DNA will be isolated. The lysis conditions have been optimized to enrich the non-human DNA ratio among human DNA, as with this protocol, human DNA does not get totally excluded. Furthermore, this protocol is appropriate for DNA extraction from fecal samples.

3.2.1 Reagents and materials

- QIAamp Fast DNA Stool Mini Kit (50) (#51604, QIAGEN) DNA extraction kit for 50 DNA preps.
- Glass beads, acid-washed (#G8772, Sigma).

3.2.2 DNA extraction protocol procedure

The samples were centrifuged at maximum speed (13,000 rpm), the supernatant was discarded, and it was added 1 mL of InhibitEX Buffer and 20 µL of lysozyme. The samples were then incubated for 30 min at 37°. After the incubation, 200 µL of beads previously prepared and the samples were put in contact while heating 5 min at 95°C (with gentle shaking). The mix was then centrifuged at maximum speed for 1 min to form a pellet (particle-free supernatant). 600 µL of supernatant was transferred to a 1.2 mL microcentrifuge tube containing Proteinase K. 600 µL of Buffer AL was added and the samples were incubated at 70°C for 10 min. The final step before the transference to the columns was adding 600 µL of pure ethanol and mixing by vortex. Afterwards, 600 µL of the sample was transferred to the column and centrifuged for 1 min at maximum speed. This step was repeated until all the sample passed through the column. Then 500 µL of the first washing buffer, Buffer AW1, was introduced to

the column, and 1 min of centrifugation at maximum speed was carried on. This step was repeated with Buffer AW2. Finally, the column was centrifuged at maximum speed for 3 min without adding any solution and place into the final collection tube where the elution buffer ATE was added. The sample suffer an incubation at room temperature for 1 minute and a centrifugation for 1 minute to elute the DNA. The quantification was then assessed.

3.3 EXTRACTION PROTOCOL 2: NETOVIR

The Novel Enrichment Technique Of Viromes (NetoVIR) is a fast, reproducible and high throughput technique expressly devoted to NGS gut viromics studies. This protocol was modified based on "NetoVIR: Modular Approach to Customize Sample Preparation Procedures for Viral Metagenomics" (Conceição-Neto et al, 2018). The amplification procedure was adapted using SuperScript IV Reverse Transcriptase protocol. Moreover, the purification of the PCR products was also altered, DNA Clean & Concentrator-5 kit was used instead of the original one established in NetoVIR. Every modification ever made was implemented trying to improve the quantity and quality of the RNA extracted, without losing the DNA part, as well as keeping in mind the price of the reagents to reduce the costs.

3.3.1 Reagents and materials

- **ENRICHMENT.** The NetoVIR protocol includes an enrichment of fecal samples for VLPs through different enzyme treatments. The content of this part is made of:
    - 0.8 µm filter (#VK01P042, Sartorius).
    - Sterile PBS.
    - Resolving enzyme buffer.
    - Benzonase nuclease (#70746, Sigma-Aldrich).
    - Micrococcal nuclease (#EN0181, Thermo Fisher Scientific).
    - Sterile EDTA.

- **EXTRACTION**. The extraction part of this protocol was optimized using the QIAamp Viral RNA Mini kit (250) (#52906, QIAGEN), a viral RNA purification procedure based on columns made of silica membranes. The protocol was followed as indicated although proceeding without addition of carrier RNA to the lysis buffer at the first step of extraction.

- **AMPLIFICATION RT-PCR**
    - **FIRST STRAND SYNTHESIS REACTION.** Performed with SuperScriptTM IV First-Strand cDNA Synthesis Reaction kit (50) (#18091050, Invitrogen) for 50 preps**.**
    - **SECOND STRAND SYNTHESIS REACTION.** Performed with Second Strand cDNA Synthesis Kit (#A48570, Invitrogen) for 50 preps.

- **PCR PRODUCT PURIFICATION.** Carried out using the kit of DNA Clean & Concentrator-5 (#D4003, Zymo Research) for 50 preps.

3.3.2 NetoVIR protocol

3.3.2.1 Enrichment step

Before starting the protocol, the resolving enzyme buffer was created by adding 12.11 g of 50 mM Tris, 1.47 g of 5mM $CaCl_2$ and 0.61 g of 1.5 mM $MgCl_2$ to 80 mL of ultrapure water in a glass beaker. Then it was necessary to adjust the pH to 8.0 with HCl and add up to 100 mL. Finally, sterilization of the mix in the autoclave was performed.

For the enrichment step the fecal suspension was homogenized and centrifuged at 17,000 x *g* for 3 min retrieving at least 200 µL of the supernatant which was then filtered in a 0.8 µm filter at 17,000 x *g* for 1 min. Finally, 7 µL of the premade 20x enzyme buffer, 2 µL of benzonase and 1 µL of micrococcal nuclease were added to 130 µL of sample filtrate. The mixture was incubated for 2 h at 37°C. To stop the reaction 7 µL of 0.2 M EDTA was added.

3.3.2.2 RNA and DNA extraction protocol

To extract nucleic acids 560 µL of Lysis buffer was added to the sample, mixed and incubated at room temperature for 10 min. Then it was briefly centrifuged and absolute ethanol (560 µL) was introduced. It was again briefly centrifuged before adding 630 µL of the sample mixture to the spin column for its centrifugation at 6,000 x *g* for 1 min. To wash the column, 500 µL of first buffer AW1 was added. The sample centrifuged at 6,000 x *g* for 1 min, and then 500 µL of AW2 but centrifuged at 20,000 x *g* for 3 min. Centrifuge again at 20,000 x *g* for 1 min and the elution buffer AVE was added, left to incubate for 1 minute and centrifuged for another minute at 6,000 x *g.*

3.3.2.3 Amplification by RT-PCR

First strand synthesis

The following components were combined in a reaction tube: 1 µL of 50 µM random hexamers, 1 µL of 10 mM dNTP mix, 11 µL of template RNA. The mix was heated at 65 °C for 5 min and then incubated on ice for at least 1 min. A combination the following components in a reaction tube: 4 µL of 5X SSIV buffer, 1 µL of 100 mM DTT, 1 µL RNaseOUT™ Recombinant RNase Inhibitor and 1 µL of SuperScriptTM IV Reverse Transcriptase (200 U/µL); was performed. RT reaction mix was added to the annealed RNA and then incubated at 23 °C for 10 min, 50–55 °C for 10 min and 80 °C for 10 min (to inactivate the reaction) and finished by heating at 85 °C for 5 min.

Second strand synthesis

For one 100 µL reaction, pipette the following components were pipetted directly into the first strand reaction tube on ice in the indicated order: 20 µL of First Strand cDNA Synthesis Reaction Mixture, 55µL of nuclease-free water, 20 µL of 5X Second Strand Reaction Mix and 5 µL of Second Strand Enzyme Mix. Incubation was performed at 16 °C for 60 min and to stop the reaction, 6 µL 0.5 M EDTA, pH 8.0 was added. The reaction was kept on ice until 10 µL (100 U) RNase was added. Then, incubate the sample for 5 min at room temperature.

3.3.2.4 PCR product purification

For this step it was necessary to preheat the UltraPure DEPC-Treated Water to 70ºC. Then, for each sample volume, it was needed 5 volumes of DNA Binding Buffer and for each 190 µL of sample, 1 mL of buffer. Then, the mix was transferred to a Zymo-Spin column with a collection tube and centrifuged at 14,000 x g for 30 s. Subsequently, 200 µL of DNA Washing Buffer was introduced in the column and

centrifuged at 14,000 g for 30 s. This step was repeated one more time until 15 µL of preheated UltraPure DEPC-Treated Water at 70 ºC was added and and incubation of 5 min at room temperature, followed by a centrifugation at 14000 g for 1 min were executed to end the reaction.

3.4 EXTRACTION PROTOCOL 3: SISPA

Sequence-independent Single-Primer Amplification, from now on SISPA, is a technique that amplifies heterogeneous DNA populations through target sequence modification, SISPA requires the directional ligation of an asymmetric linker/primer oligonucleotide onto blunt-ended DNA molecules. The common end sequence of the linker/primer allows for repeated rounds of annealing, extension, and denaturation in the presence of Taq DNA polymerase, and the linker/primers contain restriction endonuclease sites for molecular cloning (Reyes & Kim, 1991). This procedure is visually explained in Figure 1. This technique is characterized by its great simplicity and velocity, this protocol does not require a pre-processing step, duplication of amplification or pooling of samples. Indeed, a study of characterization of avian RNA viruses states that the extraction and amplification of the RNA was possible in only 5 h and in around 3 days the RNA viruses were identified (Chrzastek et al., 2017). Nevertheless, in our case, the Klenow fragment was replaced by Sequenase 2.0. It has been studied that the substitution of the Klenow fragment by a polymerase lacking the 3'-5' exonuclease activity results in better purification of the extended product, otherwise, a template size reduction can be observed (Rigaud et al., 1991; Sit & AbouHaidar, 1993).

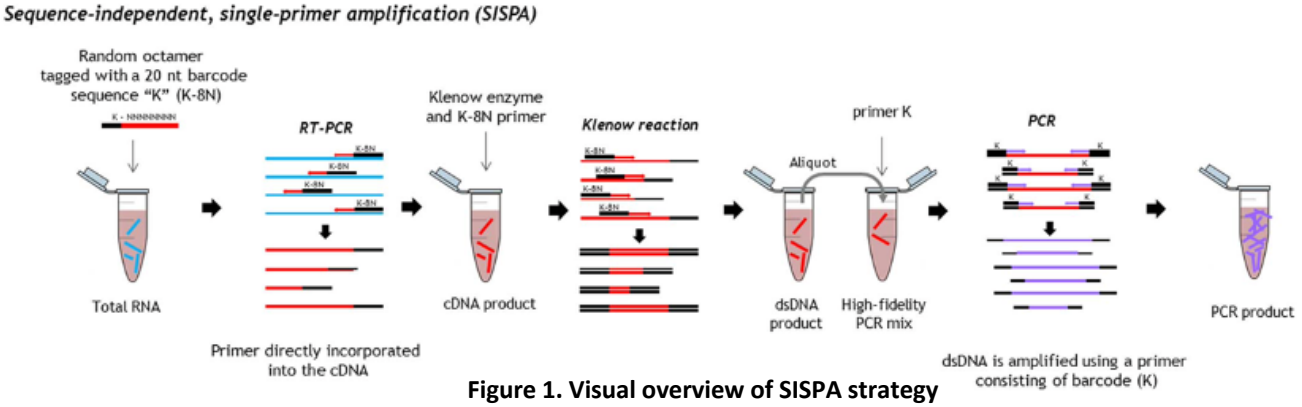

Figure 1. Visual overview of SISPA strategy

This protocol shares some parts with the NetoVIR protocol, already explained. Specifically, both protocols have in common the enrichment, the extraction and the PCR product purification procedures. Thus, those parts will not be explained again, we will focus on the amplification part, the key difference between those two virus-specific extraction protocols.

3.2.1 Reagents and materials

- **RETROTRANSCRIPTION AND VIRAL GENOME TAGGING**.
  - dNTP Mix 100mM (200µmol) (#BIO39029, Ecogen).
  - Ultrapure DEPC-treated water (#75-0024, Invitrogen).
  - Random primer A (5'-GTTTCCCAGTCACGATCNNNNNNNNN-3', Condalab).
  - Reverse Transcriptase SuperScript™ IV kit (10,000 units) (#18090050, Invitrogen).
  - RNase™ Out: Recombinant Ribonuclease Inhibitor (#10777-019, Invitrogen).

- RNaseH (#EN0202, ThermoFisher Scientific).

- **SECOND STRAND SYNTHESIS**
  - Sequenase Version 2.0 DNA Polymerase (1,000 units) (#70775Z1000UN, ThermoFisher Scientific).

- **AMPLIFICATION BY PCR**
  - AmpliTaq Gold™ DNA Polymerase kit, with the Taq polymerase, the corresponding buffer, $MgCl_2$ and the dNTPs (#4311818, ThermoFisher Scientific).
  - Primer B (5'-GTTTCCCAGTCACGATC-3', Condalab).

3.2.2 SISPA protocol

As it has been already mentioned, due to the shared steps in NetoVIR and SISPA protocols, only the amplification step (the different part) will be explained.

For the amplification, the master mix was created by combining in a 0.5 mL tube: 1 µL of random primer A (it should be added diluted to ½ from the stock), 1 µL of dNTPs (it should be added diluted to 1/10 from the stock) and 3 µL Ultrapure DEPC-treated water. After vortexing it, 5 µL were incorporated into each 0.2 mL tube. 8 µL of extracted DNA/RNA was added to their corresponding tubes and denatured by incubating for 5 min at 65°C and cooled on ice for 5 minutes. To prepare the SuperScriptIV to each tube it was added: 4 µL of 5× SSIV buffer, 1 µL of 100mM DTT, 1 µL of RNase™ Out and 1 µL of SuperScript™ IV enzyme. Then, 7 µL of this mix were added to each 0.2 mL tube. The tubes were put in the thermocycler and run the following reverse transcription program: 23°C for 10 min, 50°C for 10 min, finally, 80°C for 10 min. Subsequently 1 µL of RNase H was introduced to the tubes (kept on ice). Then, to continue with the second strand synthesis, the tubes were incubated at 95°C for 5 min in the thermocycler and quickly cooled down on ice for 5 min.

To prepare the Sequenase I mix the following reagents per tube were needed: 2 µL of Sequenase reaction buffer, 0.3 µL of Sequenase enzyme, 7.7 µL of UltraPure™ DEPC-treated water. Afterwards, 10 µL of Sequenase I mix was added to each tube, which followed the next program: from 10°C to 37°C for 8 min, ramping up 1°C every 18 s: 37°C for 8 min, then 94°C for 2 min and 10°C for 5 min. To complete the second strand, it was required to prepare the Sequenase II mix by adding per tube: 0.9 µL of enzyme dilution reagent and 0.3 µL of Sequenase enzyme and introducing 1.2 µL of the mix to each tube. The procedure continued in a thermocycler: from 10°C to 37°C for 8 min, ramping up 1°C every 18 s: 37°C for 8 min, 94°C for 8 min and 10°C for 5 min.

Finally, the amplification was carried on, samples run in duplicate. PCR-B mix for n+1 samples was prepared adding per tube: 8 µL of $MgCl_2$, 10 µL of PCR Gold Buffer 10X, 1 µL of dNTPs 100 mM, 1 µL of Taq DNA polymerase, 1 µL of Primer B and 69 µL of UltraPure™ DEPC-Treated water. Then to each tube 90 µL of the PCR-B mix and 10 µL of the corresponding sample were introduced. The tubes were Incubated in a thermocycler: 95°C for 10 min, 94°C for 30 s, 40°C for 30 s, 50°C for 30 s, 72°C for 1 min, 72°C for 10 min and maintain at 4°C. Lastly, the replicates of each sample were combined.

3.5 FLUOROMETRIC QUANTIFICATION: QUBIT

Qubit is a lab instrument used for the quantification of double stranded DNA, single stranded DNA, RNA and proteins. It is a fluorometer, meaning it uses fluorescent dyes that specifically bind to the target molecule. This technology includes several programs, suitable for a wide range of different quantifications. For this project, the program selected was high sensitivity dsDNA quantification. The output sample units were ng/µL and original sample volume was, in every case, 2µL. Finally, the standard sample employed was 10 ng/µL.

3.5.1 Reagents and materials

- Invitrogen™ Qubit™ 3 Fluorometer (Invitrogen by ThermoFisher Scientific).
- Qubit® dsDNA HS (High Sensitivity) Assay Kit (#Q33231 Invitrogen by ThermoFisher Scientific).
- Lab Consumable.

3.5.2 Qubit protocol

198 µL de buffer Qubit dsDNA HS and 2 µL of sample were added to each tube. The samples were spined and incubated 1 min at room temperature. Then they were quantified using Qubit with the selected program.

3.6 AUTOMATED ELECTROPHORESIS: TAPESTATION

TapeStation is an automated electrophoresis that allows for simpler, faster and more reliable electrophoresis for the assessment of the quality and length of DNA and RNA samples. In this project, the precise platform used was Agilent 4150 TapeStation System (#G2992AA, Agilent). Agilent owns a great variety of kits for preparing the samples before running the TapeStation. Those kits are usually classified depending on the ladder included and the program that is going to be chosen. In this case, the kit used was the High Sensitivity D5000 ScreenTape system kit, which is designed for analyzing DNA molecules from 100 – 5,000 base pairs (bp).

3.6.1 Reagents and materials

- High Sensitivity D5000 ScreenTape (#5067-5592, Agilent).
- High Sensitivity D5000 Reagents (#5067-5593, Agilent.
- Agilent softwares packages: TapeStation Controller Software and TapeStation Analysis Software.
- Lab Consumable.

3.6.2 TapeStation protocol

Allow High Sensitivity D5000 Reagents were allowed to equilibrate at room temperature for 30 min while launching the Agilent TapeStation Controller software. The manufacturer's instructions to correctly insert the ladder and the samples were followed pipetting 51 µL High Sensitivity D5000

Sample Buffer and 15 µL High Sensitivity D5000 Ladder at position A1 in a tube strip to marked it as ladder and 2 µL High Sensitivity D5000 Sample Buffer and 2 µL DNA sample in a tube strip or 96-well sample plate were then pipetted in the other positions. The tubes were closed by caps and mixed in IKA MS3 vortexer at 2,000 rpm for 1 min. Then, they were spined and loaded into the TapeStation instrument where the caps were removed. The analysis was executed.

3.7 LIBRARY PREPARATION: NEXTERA® XT DNA

The Nextera XT DNA is based on fragmentation and tagmentation of the input DNA. To tag the fragments, unique adapters are used, those adapters will be employed as primer sequence for the limited-cycle PCR reaction. This reaction will incorporate indexes on both ends enabling dual-indexed sequencing of pooled libraries (Figure 2). This protocol was modified as indicated:

- TD Buffer: 10 µL.
- DNA (0.2 ng/µL): 8 µL.
- Tagmentase: 2 µL.
- Incubation 55º: 2 min 30 s.
- Clean-up: 0.8x beads.

Furthermore, once this procedure is finished, a quantification using Qubit will be carried out. Then, after examining the DNA concentration, samples with undesirable values would be selected for a 10 cycles recovery PCR being purified with 0.9x Ampure Beads.

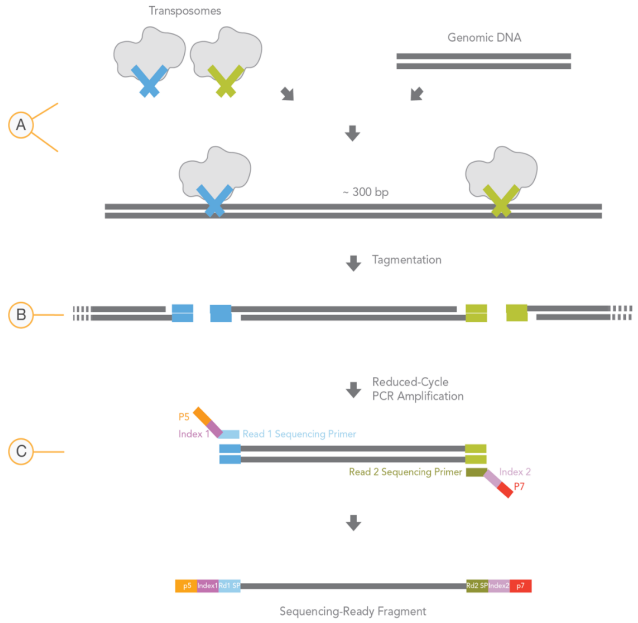

Figure 2. Nextera XT DNA Sample Preparation Visual Guide. (A) Nextera XT transposome with adapters is combined with template DNA. **(B)** Tagmentation to fragment and add adapters. **(C)** Limited cycle PCR to add sequencing primer sequences and indices

3.7.1 Reagents and materials

- Nextera XT DNA Library Preparation Kit (96 samples) (#FC-131-1096, Illumina).
- Nextera XT Index Kit (96 indexes, 192 samples) (#15052166, Illumina).
- Magnetic Stand-96 (#AM10027, ThermoFisher Scientific).
- Lab Consumable.

3.7.2 Nextera XT Protocol

3.7.2.1 Tagmentation of input DNA

To a 96-well TCY plate NTA (Nextera CT Tagment Amplicon Plate) 10 µL of TD Buffer were added to each well, as well as, 5 µL of input DNA at 0.2 ng/µL and 5 µL of ATM. Mixing the samples was carried out by gently pipette up and down followed by sealing the plate with Microseal 'B'. The plate was centrifuged at 280 x g at 20°C for 1 min and place in a thermocycler to run the following program: 55 ºC for 5 min, hold at 10 ºC. To neutralize NTA, 5 µL of NT Buffer were added to each well and mixed by up and down pipetting. The plate was then sealed again and centrifuged at 280 x g at 20°C for 1 min. At the end, the plate was incubated 5 min at room temperature.

3.7.2.2 PCR amplification

Firstly, the index primers in the TruSeq Index Plate Fixture were arranged and their position recorded on the lab tracking as follows:

a. Index 1 (i7) primer tubes (orange caps) in order horizontally, so that N701 is in column 1 and N712 is in column 12.
b. Index 2 (i5) primers (white caps) in order vertically, so that S501 is in row A and S508 is in row H.
c. Record their positions on the lab tracking form.

Afterwards, 15 µL of NPM was added to each well of the NTA plate containing index primers. By using a multichannel pipette, 5 µL of index 2 primers were introduced to each column and 5 µL of index 1 primers to each row. The plate was then covered with Microseal 'A' and centrifuged at 280 x g at 20°C for 1 min. The last step was the PCR amplification with the following program on a thermal cycler: 72 ºC 3 min, 95 ºC 30 s, 12 cycles (95 ºC 10 s, 55 ºC 30 s and 72 ºC 5 min) 72 ºC 5 min and hold at 10 ºC.

3.7.2.3 PCR clean-up

After the PCR, the plate was centrifuged at 280 x g for 1 min (20̊C) to collect condensation and 50 µL, were transferred from the NTA plate to the CAA plate (Clean Amplified Plate) already labelled. Subsequently, 30 µL of 0.6x AMPure XP beads were incorporated to each well of the CAA plate and mixed by pipetting. At this point it was crucial to place the plate on a magnetic stand for 2 min or until the supernatant has cleared and could be removed. Then, still in the magnetic stands, the beads were washed two times with 80% ethanol as follows: 200 µL of ethanol and 30 seconds incubation, the supernatant discarded using a multichannel pipette. The beads were then allowed to air-dry for 15 min before adding 52.5 µL of RSB to each well. This step was followed by a 2 min incubation and by a transference of 0 µL of the supernatant from the CAA plate to the CAN plate (Clean Amplified NTA Plate), previously labelled. Finally, library was quantified via Qubit.

3.8 SEQUENCING: NEXTSEQ ® 500

Sequencing was carried out in NextSeq ® 500 System, from Illumina. It is a Whole-Genome Sequencing (WGS) NGS technology that enables the entire genome analysis of any species. Once the library was prepared, the products were introduced into the instrument, selecting the sequencing conditions that allow for the acquisition of a single read DNA sequence with a length of 150 base pairs. Sequencing was then started and after 24-48h, the raw reads were obtained. Afterwards, via bioinformatics' tools, the raw reads were converted into individual sample files, then processed. The computer code is shown below:

CLEAN

1. The paired-end reads were filtered out and trimmed by quality with the Fastp application (version 0.23.2) (Shifu Chen et al., 2018).

    - Cut front and tail bases with quality lower than 20.
      --cut_front --cut_front_window_size 1 --cut_front_mean_quality 20
      --cut_tail --cut_tail_window_size 1 --cut_tail_mean_quality 20
    - Drop the bases on the right if the mean quality in the front to tail sliding window with size 4 is lower than 15.
      --cut_right --cut_right_window_size 4 --cut_right_mean_quality 15
    - Trim poly X tails.
      --trim_poly_x
    - Reads shorter than 50 bases are discarded.
      --length_required 50

2. The filtered reads were mapped onto the Homo sapiens (GRCh38.p14) genome by means of Bowtie2 (version 2.4.5) (Langmead et al., 2012), with the very-sensitive-local preset.
   With the aid of Samtools (version 1.16) (Danecek et al., 2021), the aligned reads were discarded (pairs corresponding to primary alignments with both mates unmapped were extracted from the SAM file).
   samtools view --require-flags 12 --exclude-flags 256

TAXONOMY

3. By means of Kaiju (version 1.9.0) (Menzel et al., 2016), each non-human paired-end read was assigned to a taxon by comparing it to the NCBI nr+euk (2022/03/10) reference database, with a maximum of 5 mismatches allowed and a minimum matching length of 20 amino acids.

   -e 5 -m 20

4. Using R (version 4.1.1) (R Core Team. R: A language and environment for statistical computing), all the reads belonging to the same taxon and sample were counted and the results were saved in a table.

4. RESULTS AND DISCUSSION

4.1 SAMPLE QUANTIFICATION

Once the DNA and/or RNA of the 72 samples was extracted with the corresponding extraction protocol, aliquots were made, which were quantified according to Qubit protocol, selecting the "High sensitivity dsDNA quantification" program. The calibration standard used was 10 ng/µL, allowing a range of values from 9 to 11 ng/µL. If the quantification of the standard was below or above those values, the calibration of the instrument from the beginning had to be performed. The quantification results are displayed in Table 2, where they are primarily sorted by protocol rather than by patient, so their analysis is easier.

Table 2. Quantification results

PROTOCOL	REPLICATE	ng DNA/ µL sample	REPLICATE	ng DNA/ µL sample
BULK METAGENOMICS	ML002-T09.1A	35.40	ML008-T09.1B	6.72
	ML002-T09.1B	14.00	ML009-T09.1A	9.95
	ML003-T09.1A	39.20	ML009-T09.1B	25.20
	ML003-T09.1B	22.90	ML013-T09.1A	5.52
	ML005-T09.1A	27.70	ML013-T09.1B	2.91
	ML005-T09.1B	18.60	ML020-T09.1A	27.20
	ML006-T09.1A	29.50	ML020-T09.1B	16.40
	ML006-T09.1B	25.00	ML022-T09.1A	25.60
	ML007-T09.1A	45.60	ML022-T09.1B	12.30
	ML007-T09.1B	24.20	CN.1A	too low
	ML008-T09.1A	6.34	CN.1B	too low

PROTOCOL	REPLICATE	ng DNA/ µL sample	REPLICATE	ng DNA/ µL sample
NETOVIR	ML002-T09.2A	0.25	ML009-T09.2A	0.61
	ML002-T09.2B	0.44	ML009-T09.2B	0.52

	ML003-T09.2A	0.74	ML013-T09.2A	0.11
	ML003-T09.2B	0.99	ML013-T09.2B	0.33
	ML005-T09.2A	4.33	ML020-T09.2A	6.27
	ML005-T09.2B	2.70	ML020-T09.2B	9.70
	ML006-T09.2A	0.68	ML022-T09.2A	1.21
	ML006-T09.2B	1.04	ML022-T09.2B	2.81
	ML007-T09.2A	1.78	CN.2A	too low
	ML007-T09.2B	0.78	CN.2B	too low
	ML008-T09.2A	0.17	CN-RT A	too low
	ML008-T09.2B	0.33	CN-RT B	too low

PROTOCOL	REPLICATE	ng DNA/ µL sample	REPLICATE	ng DNA/ µL sample
SISPA	ML002-T09.3A	6.70	ML009-T09.3B	0.21
	ML002-T09.3B	44.40	ML013-T09.3A	45.00
	ML003-T09.3A	9.34	ML013-T09.3B	too low
	ML003-T09.3B	36.20	ML020-T09.3A	46.00
	ML005-T09.3A	1.60	ML020-T09.3B	43.20
	ML005-T09.3B	40.80	ML022-T09.3A	41.70
	ML006-T09.3A	0.11	ML022-T09.3B	38.50
	ML006-T09.3B	too low	CN.3A	26.30
	ML007-T09.3A	21.07	CN.3B	1.20
	ML007-T09.3B	0.10	CN-RT A	46.60

	ML008-T09.3A	0.09	CN-RT B	3.06
	ML008-T09.3B	11.70	CN-PCR A	too low
	ML009-T09.3A	25.70	CN-PCR B	0.13

As it can be observed, the measured concentration varies greatly between extraction protocols. In general, the larger concentrations are achieved with bulk metagenomics (mean ± standard deviation: 21.01 ± 11.76 ng/µL), while NetoVIR presents the lowest quantification values (1.79 ± 2.44 ng/µL). SISPA (20.62 ± 19.26 ng/µL).

Bulk metagenomics, as expected from a total DNA extraction, shows higher concentration values for any sample, with a notable heterogeneity between duplicates (11.50 ± 8.13 ng/µL). This difference was not foreseeable, the replicas were from the same patient sample but done in duplicates, although it can be due to many reasons, such as the initial DNA input, differences in handling the reagents, etc. Although the replicates come from the same patient, the starting aliquots came from different tubes, thus seems appropriate for the wide-ranging values. When evaluating the outcome of the quantification, it should be taken into consideration that with this protocol nucleic acids from the host and all microbes present in the sample, including bacterial and viral communities, were lysed, extracted and quantified. Then, the high DNA concentration values and the heterogeneity are coherent.

In regard to NetoVIR protocol, the resulting DNA concentration between duplicates (replicate) was more homogeneous (0.89 ± 0.63), even though the achieved amount of DNA was inferior. Nevertheless, these results are also coherent with the expectancy. NetoVIR is a virus-specific extraction protocol so it is expected to have a purer, but lower, viral concentration. Even if the DNA concentration was minor than the obtained with other protocols, it should be pointed out that none was classified as 'too low'.

Finally, the most discrepant results were found with SISPA protocol. Here, the range of values was so wide that it ranged from 0.2 to 46 ng DNA/µL. Furthermore, the duplicates did not show any kind of relation between them, instead the highest standard deviation from the difference between replicas of the same sample (21.29 ± 15.06 ng/µL). For example, from patient number 13, its second replicate (B) was ranked as 'too low', meaning its concentration was below 0.05 and the Qubit is not able to quantify it faithfully; while the first DNA replicate (A) was 45 ng/µL. Despite this, the suggested workflow consisted in incorporating the 'too low' samples either way, its repetition was discarded due to the lack of input material. Moreover, it should be highlighted that the negative controls in this protocol did have a value, they are not 'too low' as it could be expected. Nonetheless, once taking into consideration the procedure of SISPA, which is virus-specific protocol with several amplification by PCR, the odd results can be due to contamination in several steps of the experimental operation. It is also hypothesized, the 'DNA' in the CN is not actual viral DNA, instead, primer-dimers could have been formed from the RT-PCR and successive additional amplification steps. Primer A is a random primer, a fact that can be causing some of the troubles. In any case, if after sequencing the negative controls

present reads corresponding to some viral species, the reads of those viruses would be extracted from the other samples.

Quantification is not something trivial, even more concerning metagenomics. It is widely recognized that this field of study suffers from insufficient starting genomic material for sequencing. Consequently, correctly quantifying the samples along the experiment may influence the final results (Duhaime et al., 2012). In the book titled "The Human Virome: Methods and Protocols" (Moya & Brocal, 2018), the above-mentioned protocols are explained based on similar experimental strategies regardless of the origin of the samples: fecal, serum, plasma… and quantification is a key point in all of them. DNA quantification is a potent tool for superficial analysis in these cases, but not for final determination. That is the main reason why, according to the book, it is advisable to include in the workflow design diverse quantification steps, still continuing with the planning until its fulfillment.

4.2 TAPESTATION

To ensure the adequate length of the extracted and quantified DNA and cDNA fragments for the library quantification, six different samples were picked and an automated electrophoresis, via TapeStation was carried out. The selection of the six samples was made based on the protocol and concentration value. Only samples from protocol 2 (NetoVIR) and protocol 3 (SISPA) were examined, three for each protocol. The first protocol had been employed in different laboratory research lines using the same library preparation protocol, without showing any problem. The fragments of bulk metagenomics are always long enough, size validation is not necessary in this case. Regarding the quantification value, the samples selected presented the widest possible concentration range, to ensure there was no correlation between concentration and length. According to these parameters, the samples chosen were: ML002-T09.2A (2T92A), ML005-T09.2B (5T92B), ML020-T09.2A (20T92A), ML003-T09.3A (3T93A), ML005-T09.3A (5T93A) and ML020-T09.3B (20T93B). For the ladder, a 5,000 bp molecular weight marker was chosen, setting as upper marker 10,000 bp and 15 bp as lower marker. The program run was High Sensitivity D5000. The general overview of the automated electrophoresis is shown in Figure 3, while Figure 4 presents a deeper analysis of each lane.

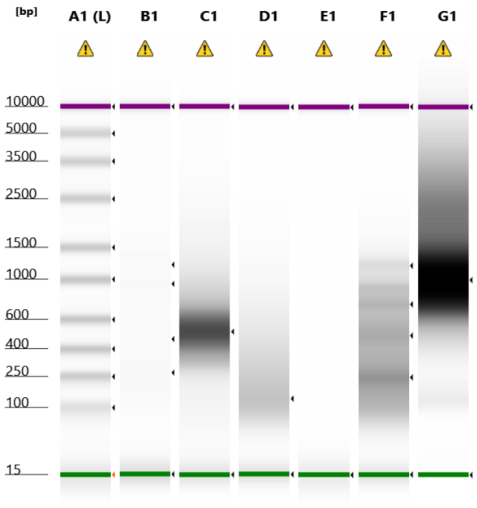

Figure 3. Automated electrophoresis results. The purple line indicates the upper marker, the green one the lower marker. A ladder of 5,000 bp is used as reference and placed in lane A1. B1 corresponds to 2T92A, C1 to 3T93A, D1 to 5T93A, E1 to 20T92A, F1 to 20T93B and G1 to 5T92B.

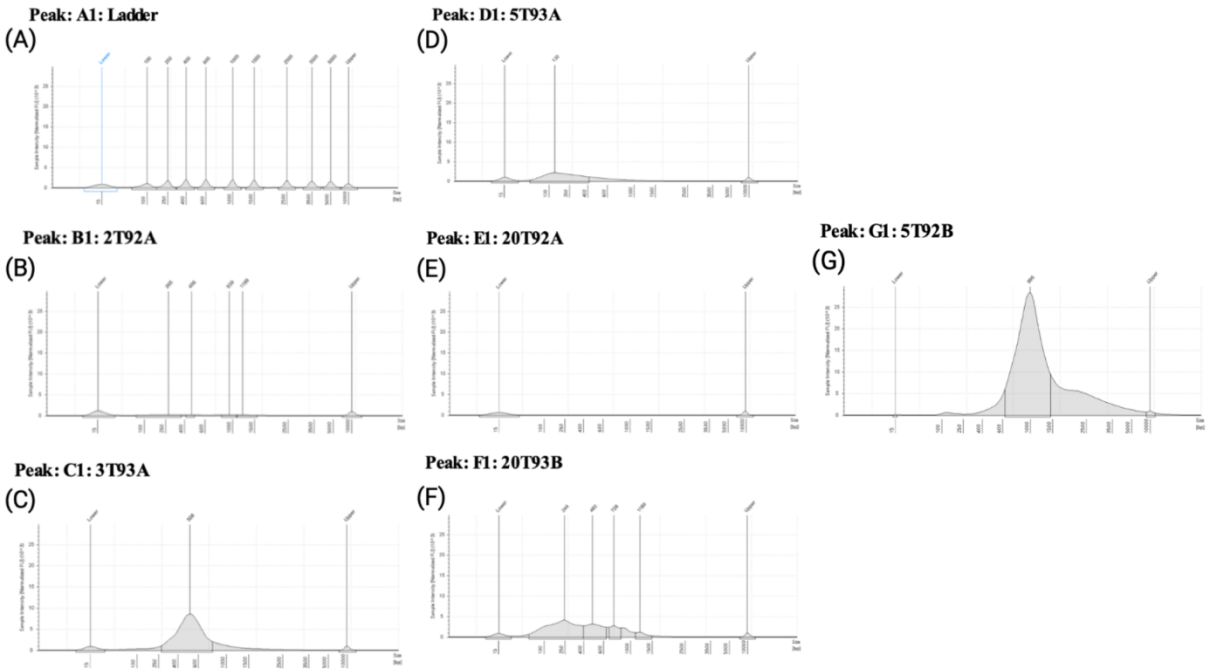

Figure 4. Lanes' analysis of automated electrophoresis plotting Sample intensity vs sample size (bp) **(A)**The ladder. **(B)** Different intensity peaks found in sample ML002-T09.2A. **(C)** Different intensity peaks found in sample ML003-T09.3A. **(D)** Different intensity peaks found in sample ML005-T09.3A. **(E)** Different intensity peaks found in sample ML020-T09.2A. **(F)** Different intensity peaks found in sample ML020-T09.3B. **(G)** Different intensity peaks found in sample ML005-T09.2B.

Focussing on Figure 4, by plotting Sample intensity vs sample size (bp), it could be noticed the different fragment length in each sample. Plot A showed a discrete number of peaks, the intensity equally distributed, indicating this scheme corresponds to the Ladder. Plot B, sample ML002-T09.2A (patient 2, protocol 2, replicate A) presented size distribution fluctuating from 250 up to 1200 bp. However, the intensity of all of those values was too low, not many DNA fragments were actually in the sample, as it was also seen in the almost empty lane B1 of Figure 3. Continuing with ML003-T09.3A, size variation was reduced and the fragments piled up with great intensity around 350-650 bp, peak of intensity found at 508bp. There were numerous DNA fragments with that size as it could be estimated looking at the intense black band in Figure 3 lane C1. D corresponds to ML005-T09.3A, this sample was extracted using SISPA, same as sample C. Still, the results were nothing alike, the fragments here were mainly, 130 bp, some fragments around 250-400 bp might be recovered, but the intensity could not be enough. In Figure 3 lane E1, an empty lane was found. It was not shocking that there was no intensity, meaning no peaks, aside to upper and lower markers, in sample E Figure 4. Nonetheless, the sample ML020-T09.2A will not be excluded from library preparation or sequencing as it is needed, at least, for the comparison with its other replicate. Sample ML020-T09.3B in scheme F showed a fragment size distribution ranging from 100 up to 1500, all of them with nice intensity, as it was observed in lane F1 Figure 3. Finally, lane G1 was the most intense one, being distributed around 600-1500 bp, finding the maximum number of fragments at 1000 bp, plot G in Figure 4 corroborated this information. This sample, ML005-T09.2B, presented the finest size-intensity distribution for the standards of library preparation. It was extracted using protocol 2, NetoVIR. One could wrongly think this method is better than protocol 3, but the results of TapeSation are not consistent. TapeStation's only role is to ensure having enough fragments with proper size to prepare the library, its aim is not to

compare the protocols. Indeed, it has been mentioned that in lane E none intensity was reflected, and the sample in lane E was also extracted with NetoVIR. Regarding protocol 3, the outcome is similar, peaks found in plots C or F bear no resemblance with DNA content of sample in D.

All things considered, no conclusion about which VLP approach was progressing better could be extracted from the automated electrophoresis. However, as it has been already mentioned, this was not the objective of the procedure. The main goal was to observe the DNA fragments size distribution and evaluate which modifications, if necessary, of the library preparation protocol will be applied. In this context, the objective was accomplished.
Agarose gels have been the gold standard method for DNA quality and quantity assessment, but they are not suitable for high-throughput workflow. On the contrary, innovative techniques such as automated electrophoresis (TapeStation systems), have several benefits to overcome the established use of traditional electrophoresis, even more when the fate of the samples analyzed is NGS. Certainly, automated electrophoresis presents better results with genomic bacterial DNA (Kong et al., 2015), but their ability to informatically determine the fragments size distribution is also very valuable for viral studies. The study of Allander et al., in 2001 was quite inspiring. In this study, SISPA was employed on serum samples to extract and analyze bovine parvovirus. The researchers compared sample dilution, sample filtering and SISPA procedure to isolate and extract viral entities. The comparison was performed in an agarose gel where it could be observed how SISPA was the best choice. Interestingly, the background noise in the electrophoresis gel was highly reduced once DNAse treatment was applied before the RT-PCR.

In our case, if DNA viruses were not subject of interest, this modification could have been done with the objective of increasing the intensity of the longer fragments. However, the most abundant viral agents in the human virome are phages that are mainly composed of DNA. So this alternative was discarded. Instead the TapeStation system provided a good choice to avoid the interference signal typical of conventional electrophoresis without needing DNAse treatment, in particular in this protocol, which owns several PCR amplification stages prompted to contamination. Same ending would have occurred in NetoVIR protocol as it possesses the same characteristics of VLP extraction approaches (The Human Virome: Methods and Protocols, 2018), DNAse treatment was not possible and background noise would have been detected in the agarose gel, difficulting the conclusions. Maybe, as protocol 2 only needs one PCR, the possibility of contamination, therefore of undesirable noise, could have been diminished. Despite that, our project has embraced the best of the two options, no DNAse treatment and an automated electrophoresis analysis, resulting in clearer results.

4.3 LIBRARY QUANTIFICATION

Library quantification was necessary to check how the library preparation step had occurred, making an estimation of the quality of the library, meaning the lengths of the fragments, via the quantity of DNA of the 72 samples present in the 96-well plate. The results are shown in Table 3. The wells that showed a DNA concentration below 1 ng/µL sample (in red) would proceed with an additional PCR amplification, to avoid losing them when sequencing.

Table 3. Library quantification results

POSITION	SAMPLE	[Lib] ng DNA/µL sample	POSITION	SAMPLE	[Lib] ng DNA/µL sample
A1	ML002-T09.1A	1.56	**D1**	ML013-T09.2A	1.05
A2	ML002-T09.1B	1.27	**D2**	ML013-T09.2B	1.29
A3	ML003-T09.1A	2.41	**D3**	ML020-T09.2A	1.12
A4	ML003-T09.1B	6.54	**D4**	ML020-T09.2B	2.43
A5	ML005-T09.1A	1.35	**D5**	ML022-T09.2A	1.03
A6	ML005-T09.1B	1.18	**D6**	ML022-T09.2B	0.79
A7	ML006-T09.1A	3.01	**D7**	CN.2A	0.19
A8	ML006-T09.1B	2.82	**D8**	CN.2B	0.21
A9	ML007-T09.1A	1.27	**D9**	CN-RT 2A	0.09
A10	ML007-T09.1B	2.81	**D10**	CN-RT 2B	0.21
A11	ML008-T09.1A	2.80	**D11**	ML002-T09.3A	1.72
A12	ML008-T09.1B	2.43	**D12**	ML002-T09.3B	1.57
B1	ML009-T09.1A	11.65	**E1**	ML003-T09.3A	5.74
B2	ML009-T09.1B	6.03	**E2**	ML003-T09.3B	2.09
B3	ML013-T09.1A	10.29	**E3**	ML005-T09.3A	1.01
B4	ML013-T09.1B	5.09	**E4**	ML005-T09.3B	2.80
B5	ML020-T09.1A	0.42	**E5**	ML006-T09.3A	1.90
B6	ML020-T09.1B	1.75	**E6**	ML006-T09.3B	0.22
B7	ML022-T09.1A	4,.51	**E7**	ML007-T09.3A	2.52
B8	ML022-T09.1B	3.44	**E8**	ML007-T09.3B	0.17
B9	CN.1A	0.12	**E9**	ML008-T09.3A	0.11
B10	CN.1B	0.05	**E10**	ML008-T09.3B	3.35
B11	ML002-T09.2A	3.07	**E11**	ML009-T09.3A	3.93
B12	ML002-T09.2B	2.03	**E12**	ML009-T09.3B	1.71

C1	ML003-T09.2A	1.60	**F1**	ML013-T09.3A	0.45
C2	ML003-T09.2B	1.79	**F2**	ML013-T09.3B	0.09
C3	ML005-T09.2A	0.78	**F3**	ML020-T09.3A	0.98
C4	ML005-T09.2B	1.71	**F4**	ML020-T09.3B	1.28
C5	ML006-T09.2A	1.09	**F5**	ML022-T09.3A	2.97
C6	ML006-T09.2B	1.47	**F6**	ML022-T09.3B	0.17
C7	ML007-T09.2A	1.79	**F7**	CN.3A	2.20
C8	ML007-T09.2B	1.32	**F8**	CN.3B	2.37
C9	ML008-T09.2A	0.85	**F9**	CN-RT 3A	2.01
C10	ML008-T09.2B	1.08	**F10**	CN-RT 3B	4.94
C11	ML009-T09.2A	0.70	**F11**	CN-PCR 3A	0.14
C12	ML009-T09.2B	0.08	**F12**	CN-PCR 3B	2.31

At this project level, once the library was constructed and quantified, some conclusions can be drawn from the extraction protocols. According to Table 3, the outcome of the Bulk metagenomics protocol libraries (ending by ".1A/B"), seemed the most optimal, as those would not need any additional amplification step, with the exception of ML020-T09.1A, whose library concentration was lower than 1 ng DNA/µL sample. The negative controls presented a minimum concentration, probably due to amplicons dimers.

However, as it has been evidenced, the virus-specific methodology, protocols 2 and 3 (NetoVIR and SISPA respectively) usually are more controversial. Regardless of the laboratory good practices established along the process, the results were, apparently, unfavorable, although the ultimate outcome was more critical for the final assessment. Therefore, the most tedious method could end up being the best choice, if the finest virus proportion is collected. In both protocols several samples needed the recovery PCR as their concentration was too low to be sequenced in these conditions. It is interesting to indicate that, at least in protocol 2, the samples that manifested low or negligible DNA amount in TapeStation analysis, particularly ML020-T09.2A corresponding to the empty lane in the automated electrophoresis, had acceptable DNA concentration in their library. As well as the sample ML002-T09.2A, which had the highest DNA concentration in Table 3 while the intensity of the peaks in the electrophoresis plot, thus the amount of DNA fragments, was not appropriate. Still, it should be proven, after the sequencing process, that those values matched with real viral DNA. Moving into protocol 3, the same issues were found with the extra handicap of the negative controls. As it was stated in Table 2, the CN in protocol 3 holded concentration values. It was not unexpected that their concentration outcome, after the Nextera library preparation procedure, was similar. Nevertheless, the PCR negative control replicate B (CN-PCR 3B) had an insignificant concentration value, not classified as 'too low' but almost. It was not concerning matter, but now the 2.31 ng DNA/µL sample was fairly confusing. Again, the main guessing was primer-dimers although it seemed too high for just

that reason. An external contamination in this sample during this step should be considered. In respect of the extraction and RT-PCR controls, the presence of DNA was deduced from the beginning, after the initial quantification. In other studies which followed the same approach as we did (Gonzales-Gustavson et al., 2017; The Human Virome: Methods and Protocols, 2018), instead of using fecal samples, serum samples were collected; similar problems regarding SISPA's negative controls appeared, but the outcome was not jeopardized.

4.4 SEQUENCING

The outcome of the 1x150 bp sequencing of the 72 samples of this project (60 case samples and 12 negative controls) resulted in a total of 211,711,790 reads (average ± STD: 2,940,442 ± 1,589,687 reads per sample), which after quality filtering, as well as human host and ribosomal RNA reads removal, resulted in 77,473,577 reads (1,076,022 ± 1,139,168 reads per sample), 48,278,624 of them (670,536 ± 834,253 reads per sample) could be taxonomically assigned by Kaiju. Of those, a total of 2,235,406 reads could be assigned to 2,167 viral taxa (31,047 ± 53,351 reads per sample), corresponding the remaining reads mostly to bacteria (45,817,006), eukarya (217,547) and archaea (8,665).

Therefore, the percentage of viruses extracted could be calculated as:

$$\frac{viral\ reads}{total\ reads} \cdot 100$$

On average, the ratio of viral reads extracted in regard to the total reads sequenced was 4.63%. Nonetheless, this percentage varied when the extraction protocol was taken into account. For instance, using bulk metagenomics, viruses only represented 1.8% of all reads, while NetoVIR and SISPA increased this figure up to 9.8% and 12%, respectively. The fraction of viruses in each protocol was coherent with the expectancy. By evaluating the outcome deeper, it can be concluded that, with protocol 1 (bulk procedure), a higher proportion of non-viral sequences was found, related mainly to prokaryotes (mostly bacteria), consequently decreasing the percentage of viral reads. Meanwhile, the virus-specific protocols were a focussed approach, and therefore reads other than viral were relatively less represented. Iit is advisable to remember that both protocols contained VLP-enrichment steps followed by PCR amplifications, so it is logical to have found a larger percentage of viral reads in NetoVIR and SISPA.

The analysis, focused on protocols, continued with the study of the number of virus species that were extracted with each method to examine the most important or representative viral taxa based on previous literature. Besides, for each protocol a classification of viral reads was made according to a type of nucleic acids of its genome, summarized Baltimore classification (Baltimore, 1971): double stranded DNA (dsDNA), single stranded DNA (ssDNA), double stranded RNA (dsRNA) and single stranded RNA (ssRNA). Although current classification of viruses is mostly based on genome sequence, this system used to stratify the viruses according to their routes of genome expression, was useful to assess the performance of protocols in terms of spanning the diversity of nucleic acids of the virome (Kooning et al., 2021). Although it is a vague classification, the main objective was to introduce a parameter to compare the diversity of viruses extracted with each protocol, Baltimore classes offer this possibility simplifying the analysis. It should be taken into account that even though one method

can detect more species of viruses it is possible that those species might be concentrated around only one type of virus while another protocol can give a lower number of species but the ones detected belong to several types. For instance, one methodology aimed to detect many different species of dsDNA viruses, and the other protocols aimed to detect numerous groups such as ssDNA or even RNA viruses apart from the dsDNA viruses.

To carry out the comparisons among the three protocols, in terms of the distribution of the identified viruses, based on their genome nucleic acid, a strict approach consisting of removal of viral species that could be identified only in the negative controls, but also present in both negative controls and samples was followed, thus retaining only those viruses exclusively found in at least one sample but absent in all negative controls. This approach could probably eliminate actual viral taxa from the samples, but this way we guarantee that contaminant sequences as those identified in the negative controls do not bias the analysis. This way, the removal of the 59,819 viral reads identified in the 12 negative controls, encompassed the decrease from 2,175,587 reads of the 60 samples, to only 52,775 remaining after filtering. The difference in number of reads before and after filtering in each protocol is detailed hereinafter. Protocol 1 before filtering owned 597.769 reads belonging to 992 species and after 16,420 reads, 734 species. Protocol 2, before filtering 391.009 reads, 561 species, while 6,902 reads, 372 species after the filtering. Finally, protocol 3 had 3: before filtering 1,186,809 reads, 919 species, after filtering 29,453 reads, 466 species. In negative controls the reads that were present before filtering that were, afterwards, removed were 59,819 reads from 903 species.

The distribution of the number of virus species and according to their genome, in the comparison of the three methods, are shown in figure 5. After filtering, a total of 184 virus species were identified in at least one replicate of one sample with the three protocols used in this study. Those viruses would constitute a core of viruses that can be identified in this particular set of samples analyzed, regardless of the method used. All of them were dsDNA viruses or had uncharacterized genomes.

In addition, with bulk metagenomics a total of 734 species of virus were identified (Fig. 5) 303 of which were uniquely identified using protocol 1, while 431 species were also identified using protocols 2, 3 or both. Bulk metagenomics' species were distributed among 5 different phyla: Uroviricota, Cressdnaviricota, Hofneiviricota, Phixviricota and Preplasmiviricota. According to their genome, the vast majority of the viruses extracted were restricted to dsDNA ones (≈94%), with only a small fraction of ssDNA viruses (≈1%) and no RNA viruses (Fig. 6). The remaining 5% come from unclassified viruses, meaning that those viral agents were not present in the database for the taxonomy assignment.

Regarding protocol 2, NetoVir, it allowed us to identify fewer viral species, 372 (Fig. 5) 23 of which were exclusive, whereas the remaining 349 were also found with other protocols. The viruses were also distributed among 5 phyla: Uroviricota, Kitrinoviricota, Hofneiviricota, Phixviricota and Pisuviricota. However, the genome type revealed more diversity in terms of nucleic acids, as it can be observed in Fig. 6, mainly composed of dsDNA (57%), but greater presence of ssDNA viruses (5%), and also detected RNA viruses: dsRNA (≈35%) and ssRNA (1,3%) viruses. The residual percentage corresponded to the unclassified viruses.

Protocol 3, SISPA, was the method that retrieved the richest diversity. Despite identification of fewer viral species, 466 (Fig 5.), than method 1, those identified were distributed among 11 phyla: Uroviricota, Artverviricota, Cressdnaviricota, Cossaviricota, Kitrinoviricota, Hofneiviricota, Negarnaviricota, Nucleocytoviricota, Phixviricota, Pisuviricota and Preplasmiviricota. Furthermore, SISPA could identify 118 novel species that any of the other protocols were able to do. According to the nucleic acid of its genome (Fig. 6) it was also the most diverse because besides viruses composed of dsDNA (58%), ssDNA (2.8%), dsRNA (≈0.3%) or ssRNA (0.2%), retroviruses (dsDNA-RT and ssRNA-RT) were identified. This special virus belongs to the phylum of Artverviricota and to the family *Retroviridae*. The percentage of unclassified viruses in this protocol was relatively important; around 40% of the viral reads obtained with this protocol could not be assigned taxonomically characterized viruses included in databases.

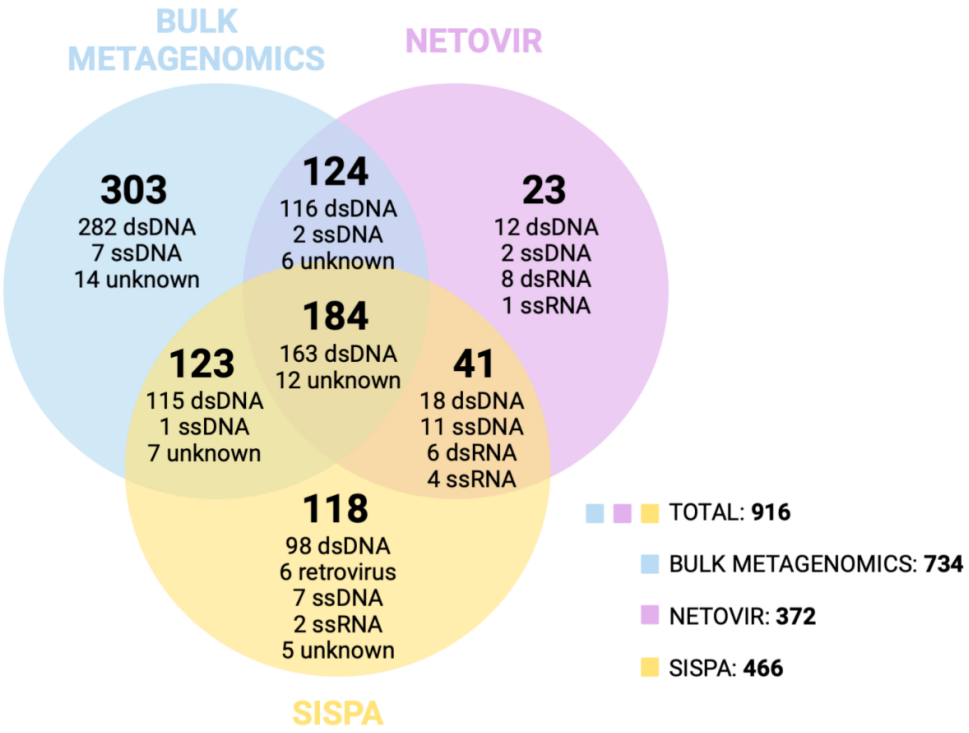

Figure 5. Venn diagram of the number of species extracted with each protocol and their combinations. Each circle corresponds to a protocol while the intersections between circles correspond to the number of species shared among those two protocols or, in the case of the central intersection, the common species detected in all the protocols. Each number is itemized into the addition of different types of viruses according to the summarized Baltimore classification.

As it can be observed in Fig. 6, the first protocol identified in a great majority dsDNA viruses. Bulk metagenomics does not incorporate retrotranscription or amplification steps and, consequently, the RNA viruses cannot be detected with this method. The ssDNA portion included with this protocol was not as large as dsDNA primary due to the approach of library construction. The library used for sequencing was formed upon double strand DNA. Therefore, there was only a remote possibility to detect single stranded viruses if they followed a rolling-circle-like replication cycle. Those species have their genetic material arranged in a circular shape, when the replication of the single DNA strand occurs, they incorporate a transient double strand phase resembling prokaryotic rolling circle plasmids

(Malathi & Renuka, 2019). If by coincidence, at the moment when the extraction was made, the dsDNA was extracted and preserved the ssDNA viruses will have been incorporated into the NGS library. To prove this assumption, the ssDNA viruses found with Bulk metagenomics were studied. Those viral agents were organized into three phyla: Hofneiviricota (Inoviridae family), Phixviricota (Microviridae family) and Cressdnaviricota (unspecified family). All of them are ssDNA viruses including a dsDNA intermediate during their replication cycle. Anyhow, ssDNA viruses were not common when using Bulk metagenomics. It is true that the method is then sort of biased, but there are other methods approved and used in viromics that have a tendency to overrepresent single stranded viruses (e.g. virome amplification through phi29, Kim and Bae, 2011; Malathi & Renuka, 2019). Furthermore, although ssDNA viruses abundance in nature could have been underestimated, they do not predominate in the viral load of feces, nor in the human virome. Nevertheless, few ssDNA viruses have been cultivated under laboratory conditions so little is known about their biochemical analysis or detailed structural information (Laanto et al., 2017).

Moving into the unclassified percentage, it is observed that this fraction in bulk is not the same as in SISPA, which shows higher percentage and diversity. Bulk metagenomics presents the lowest variability, basically focussed on dsDNA viral agents. Therefore, we suggest the unclassified fraction must belong to DNA viruses yet undiscovered or at least not taxonomically assigned. The conclusion is supported by previous bibliography (Breitbart, 2003; Rohwer, 2003; Wylie, 2015, Zhu et al., 2022), which confirms the immense amount of dsDNA phages in human virome ranking them as the most representative viral agents by far, and by the limitations of this protocol's procedure already explained.

Referring to the number of viral species and phyla identified with Bulk metagenomics (Fig. 7A), the most representative phylum is Uroviricota with its only class named as *Caudoviricetes*, being *Caudovirales* the most numerous order. These viruses are dsDNA bacteriophages, indeed consistent with the above conclusions, that are collectively named as 'tailed bacteriophages'. Their hosts are bacteria and archaea and in the context of human virome, their preferred environments are the gastrointestinal tract, skin and mouth. Under specific conditions, Caudovirales would alter the intestinal bacterial population, diminishing beneficial bacterias and initiating intestinal inflammation (Khanna et al., 2021). According to some experts, the discovery of new species belonging to this class has suffered a five fold increase since 2015. In fact, they cover the majority of phage sequencing data nowadays (Zhu et al., 2022). The small percentage of ssDNA viruses should not be left out of place as their viral families are fairly interesting. Inoviruses, viruses from the *Inoviridae* family (phylum of Hofneiviricota), are single stranded bacteriophages that infect primarily Gamma and Betaproteobacterias, although sporadically can be transferred to archaea hosts. Inoviruses are distributed along many biomes, in the human virome they are predominantly found in the gastrointestinal tract (Roux et al., 2019). The *Microviridae* family (Phixviricota phylum) are small bacteriophages that follow a lytic cycle, they usually infect *Escherichia* bacterias. Several evolving mechanisms have been proposed for these viruses, highlighting the importance of gene acquisition through horizontal gene transfer from their bacterial hosts (Rokyta et al., 2016). Lastly, about the Cressdnaviricota phylum only can be said that their name comes from CRESS-DNA viruses which are small circular ssDNA viruses that encode for the protein Rep, which is vital in the rolling-circle replication mechanism (Kazlauskas, 2019).

In conclusion, Bulk metagenomics is preferently focussed on dsDNA viruses, extracting and identifying the greatest variety of these viruses among the three protocols. Nonetheless, it is not effective in detection of other types of viruses.

Regarding protocol 2, NetoVIR, Fig. 6, shows a clear change in the pattern of viruses compared to the first protocol. Although the dsDNA portion was still the most representative, dsRNA viruses represented a significant fraction. As it has already been mentioned, bacteriophages are mostly dsDNA viruses, but dsRNA virus presence in microbial populations has been recently investigated and re-estimated (Wylie, 2015). In fact, an aquatic microbial study proposes that in microbes, dsRNA virus-like elements potentially represent an additional reservoir of genetic information (Decker & Parker, 2014). These were some reasons that could explain the great percentage of dsRNA. Moreover, it can be observed a rise in ssDNA and ssRNA viruses. These findings were again supported by the analysis of NetoVIR's methodology which includes the key step of retrotranscription. This protocol was virus-specific and included one RT-PCR, making possible the detection of RNA viruses by their retrotranscription into cDNA. Interestingly, the fraction assigned as unclassified decreased compared to the other two protocols. One reason that explains this issue could be that the amount of dsDNA viruses extracted was lower than in bulk metagenomics so not many un-assigned dsDNA viruses were detected. Besides, it could also influence the fact that the second protocol was not as diversified as the third one, hence reducing the probability of finding viruses that were undiscovered or unclassified. Those are two suggested hypotheses that could explain the percentage of unclassified viruses although a deeper analysis of the fractions should be done to support or refute this conclusion.

Regarding the phyla retrieved with NetoVIR (Fig. 7B), the type of viruses that are classified as dsDNA and ssDNA were shared among protocols 1 and 2, Uroviricota was the most representative for dsDNA while Hofneiviricota and Phixviricota were for ssDNA. The innovative discovery here is the RNA viruses, ssRNA, belonging to phylum Kitrinoviricota and dsRNA, belonging to phylum Pisuviricota, usually bacteriophages. Both groups include multiple species of viruses that are found in a broad range of environments. A recent phylogenetic study (Neri et al., 2022) was performed to deeply analyze RNA virome and they discovered new clades belonging to both phyla, they also mentioned that Kitrinoviricota and Pisuviricota showed genetic and phylogenetic distinctive features so it was possible for them to be related evolutionary groups that kept them apart from other RNA bacteriophages and other RNA viruses.

In summary, NetoVIR has gained diversity and their results are coherent with previous literature. Nevertheless, it is not as rich as SISPA, although its outcome might tally more.

Finally, SISPA's results were examined and discussed. Taking into account the outcome described in Fig. 6, protocol 3 stood as the protocol that most types of viruses detected, finding DNA and RNA viruses, both double and single stranded, but also thanks to this protocol, retroviruses were identified. Once again the dsDNA viruses were the most abundant and ssDNA viruses presented an expected percentage. Regarding RNA viruses, the value for ssRNA could be coherent, although it was a small fraction, it was logical as SISPA extracted more viral species than protocol 2, reducing the relative value of reads belonging to ssRNA compared to the total number of reads, therefore reducing the percentage. However, it was quite shocking the percentage related to dsRNA viruses, as an

considerable drop appeared compared to protocol 2. Furthermore, based on previous studies (Karlsson, 2013; Chrzastek, 2022) the SISPA method should be very efficient when detecting RNA viruses in specific, some researchers (Chrzastek, 2022) were able to complete the whole genome assembly of SARS-CoV-2 and influenza A from a sample containing a mixture of virus, the same was accomplished when applied to avian RNA viruses (Chrzastek, 2017). However, it would be advisable to deeply analyze the dsRNA outcome from this procedure. The value of the unclassified portion was quite relevant as well, around 40% of the viruses could not be assigned to any taxonomic profile present in the database. This reality could be a double edge factor, as it could mean SISPA has huge potential to identify new viruses or , on the contrary, it behaves as non-specific so the information extracted is not complete or logical then those reads do not match to any information on the database, even though this is not recorded on bibliography (Karlsson, 2013). The low percentage of retroviruses was expected, six species of these viruses were identified but their reads were not numerous.

Concerning the diversity of phyla and viral species (Fig. 7C), the most abundant phylum was Uroviricota, with the characteristic *Caudovirales*; followed by ssDNA viruses belonging to Hofneiviricota, only one specie of Inovirus was identified; Phixviricota, being *Microviridae* the main representative family; and Cressdnaviricota. Single stranded RNA viruses correlated to Kitrinoviricota phylum but also, to a lesser extent, to the relatively new Negarnaviricota phylum (Kuhn, 2022), to which important pathogens such as Ebolavirus Zaire, the virus responsible for the Ebola disease; or virus responsible for pneumonia. In this project, an *Orthopneumovirus* was identified as the only Negarnaviricota species present. It corresponded to patient 7, it was possible that the patient suffered from an upper respiratory infection at the moment of sample collection. Regarding dsRNA viruses, Pisuviricota was the only phylum observed. Lastly, the phylum of Artverviricota was identified, all the viruses were classified in the same class and order, *Revtraviricetes* and *Ortervirales*. Every one of these viruses encoded a reverse transcriptase, converting them into retroviruses. Moreover, both families of *Ortervirales* were found: *Caulimoviridae* (dsDNA-RT viruses) and *Retroviridae* (ssRNA-RT viruses). *Caulimoviridae* are mainly plant viruses, but it is usually to find them in human samples. This family is one of the most important viral families, economically and scientifically speaking, due to the harvest's diseases they can cause. Some *Caulimoviridae's* species had been proposed as model agents for the study of viral replication (Geering, 2014). Meanwhile, *Retroviridae* family enclose viral agents dangerous for human health, for example *Human Immunodeficiency Virus* (HIV), or oncoviruses. Nonetheless, they also include Spumaviruses, collectively known as foamy viruses, which are beneficial for the host's health.

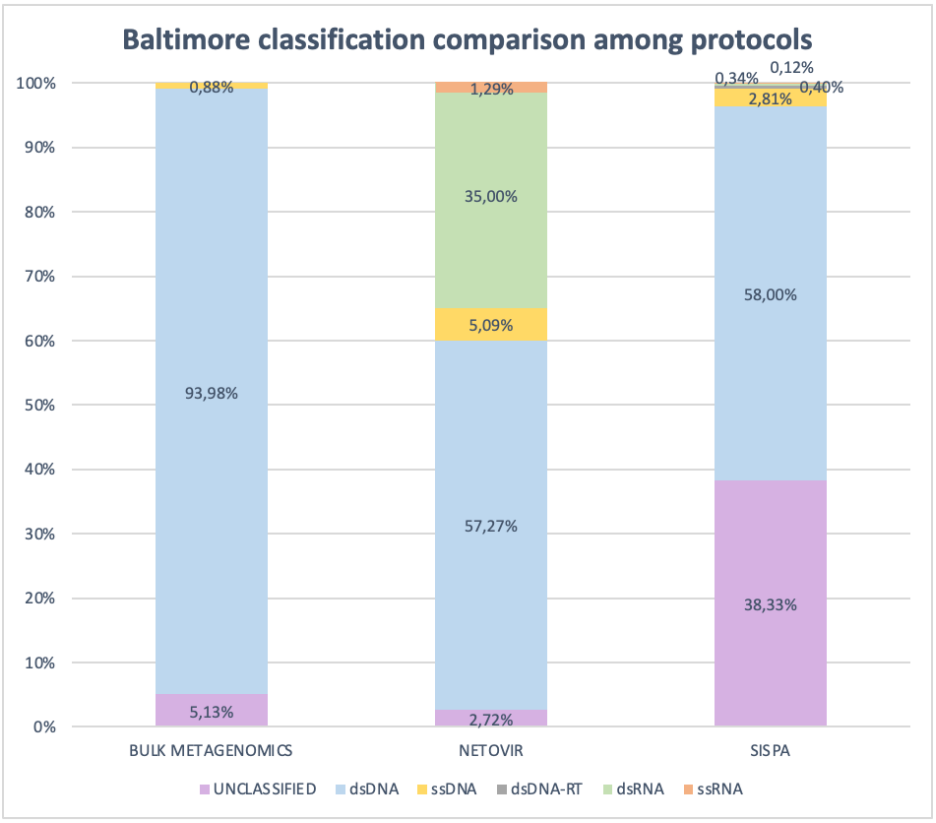

Figure 6. Comparison of the diversity in Baltimore classification found in each protocol. The percentages represent the fraction of reads from every Baltimore group in relation with the total of viral reads found in each protocol (Reads of X group in protocol X/Total reads in protocol X)

Figure 7. Comparison of the phyla found in each protocol. (A) Phyla found with Bulk metagenomics, its unclassified fraction and their addition. **(B)** Phyla found with NetoVIR, its unclassified fraction and their addition. **(C)** Phyla found with SISPA, its unclassified fraction and their addition.

5. FUTURE PERSPECTIVES

In summary, Bulk metagenomics reached the highest DNA concentration extracted, with relative homogeneity between replicas, pointing towards a satisfactory reproducibility. This method detects a wide variety of species. Indeed, it was the protocol that more viruses identified, 734 different species 303 of which were exclusive from this extraction protocol. However, it could be useful only when the objectives are dsDNA bacteriophages because RNA and ssDNA viruses are left out during its procedure. This protocol behaved as it was expected from a bulk extraction protocol. Moreover, regarding the technique, it was the simplest, quickest and most affordable of all three, features that are worth mentioning.

Meanwhile, NetoVIR extracted the lowest concentration of DNA, which was a limitation in the library preparation step. Despite this, it presented the highest homogeneity between replicates, meaning it is a highly reproducible procedure. It detected 372 viral species, shared among Bulk metagenomics and/or SISPA. The species were mainly classified into dsDNA or dsRNA bacteriophages, although ssDNA and ssRNA viruses were also detected. Its results were coherent when contrasting them against previous literature about virus-specific protocols.

Last but not least, SISPA's quantification results were not homogenous, this lack of reproducibility created some issues during the library preparation due to the difference in DNA fragments length. Nonetheless, it was the most diverse protocol regarding viral species detection. Indeed, it identified 466 species, 118 of which were exclusive. It also identified more phyla, with viruses classified in eleven phyla, discovering dsDNA and ssRNA retroviruses. However, there was a very relevant portion of unclassified viruses while the percentage of reads from dsRNA viral agents was excessively deficient. The protocol's insufficiency in reproducibility could be a compelling reason for these limitations.

Therefore, it has been demonstrated that the election of a certain viral nucleic acid extraction protocol influences vastly in the final results. Thus, more research in this field is needed to discard and select different extraction procedures to finally achieve a predetermined experimental workflow depending on the class of virus that is aimed to detect. In such a manner, outcomes from different projects could be more comparable and global conclusions could be drawn.

6. CONCLUSION

The main aim of this project was to evaluate the efficacy, reproducibility, advantages and limitations of three different experimental procedures for viral nucleic acid extraction already compared. The experiments were performed on fecal samples collected from patients genetically diagnosed with Lynch Syndrome and the three selected protocols were Bulks metagenomics, NetoVIR and SISPA. In this context, the outcome from an initial sample quantity and quality assessment, from the library quantification and from the sequencing procedure in all the approaches was discussed and checked. The main conclusions that were drawn from this FDP were:

- In terms of viral reads characterization, Bulks metagenomics successfully identified 734 different species 303 of which were exclusive from this extraction protocol. NetoVIR could only identify 372 viral species, many of them also detected by the other protocols. SISPA identified 466 species, 118 of which were exclusive from it.

- The most important and abundant viruses characterized in human fecal samples are dsDNA bacteriophages, mainly belonging to the Uroviricota phylum, followed by dsRNA viruses from Pisuviricota phylum, being, most of them, bacteriophages as well. Therefore, it can be assumed the human gastrointestinal virome is mainly composed of bacteriophages which correlates with previous findings.

- Selecting the best protocol will depend on the goal of the experiment. If virus diversity is not a relevant matter and the project is focused on dsDNA viruses, bulk metagenomics could be a safe option. On the contrary, if the research requires RNA viruses, a virus-specific protocol must be the choice.

7. REFERENCES

1. ALLANDER, T., EMERSON, S. U., ENGLE, R. E., PURCELL, R. H., & BUKH, J. (2001). A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proceedings of the National Academy of Sciences*, *98*(20), 11609-11614.

2. ARTHUR, J. C., & JOBIN, C. (2011). The struggle within: microbial influences on colorectal cancer. *Inflammatory bowel diseases*, *17*(1), 396–409.

3. ARTHUR, J. C., PEREZ-CHANONA, E., MÜHLBAUER, M., TOMKOVICH, S., URONIS, J. M., FAN, T. J., CAMPBELL, B. J., ABUJAMEL, T., DOGAN, B., ROGERS, A. B., RHODES, J. M., STINTZI, A., SIMPSON, K. W., HANSEN, J. J., KEKU, T. O., FODOR, A. A., & JOBIN, C. (2012). Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science (New York, N.Y.)*, *338*(6103), 120–123.

4. BAI, G., LIN, S., HSU, Y., & CHEN, S. (2022). The Human Virome: Viral Metagenomics, Relations with Human Diseases, and Therapeutic Applications. *Viruses*, *14*(2), 278.

5. BALTIMORE D. (1971). Expression of animal virus genomes. *Bacteriological reviews*, *35*(3), 235–241.

6. BELCHEVA, A., IRRAZABAL, T., ROBERTSON, S. J., STREUTKER, C., MAUGHAN, H., RUBINO, S., MORIYAMA, E. H., COPELAND, J. K., SURENDRA, A., KUMAR, S., GREEN, B., GEDDES, K., PEZO, R. C., NAVARRE, W. W., MILOSEVIC, M., WILSON, B. C., GIRARDIN, S. E., WOLEVER, T. M. S., EDELMANN, W., GUTTMAN, D. S., … MARTIN, A. (2014). Gut microbial metabolism drives transformation of MSH2-deficient colon epithelial cells. *Cell*, *158*(2), 288–299.

7. BELCHEVA, A., & MARTIN, A. (2014). Gut microbiota and colon cancer: the carbohydrate link. *Molecular & cellular oncology*, *2*(1), e969630.

8. BHARDWAJ, K., GARG, A., PANDEY, A. K., SHARMA, H., KUMAR, M., & VRATI, S. (2022). Insights into the human gut virome by sampling a population from the Indian subcontinent. *Journal of General Virology*, *103*(8), 10.1099/jgv.0.001774.

9. BILLER, L. H., SYNGAL, S., & YURGELUN, M. B. (2019). Recent advances in Lynch syndrome. *Familial Cancer*, *18*(2), 211-219.

10. BREITBART, M., SALAMON, P., ANDRESEN, B., MAHAFFY, J. M., SEGALL, A. M., MEAD, D. A., AZAM, F., & ROHWER, F. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences*, *99*(22), 14250-14255.

11. BREITBART, M., HEWSON, I., FELTS, B., MAHAFFY, J. M., NULTON, J., SALAMON, P., & ROHWER, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. *Journal of bacteriology*, *185*(20), 6220–6223.

12. CARDING, S. R., DAVIS, N., & HOYLES, L. (2017). Review article: the human intestinal virome in health and disease. *Alimentary Pharmacology & Therapeutics*, *46*(9), 800-815.

13. CHRZASTEK, K., LEE, D., SMITH, D. K., SHARMA, P., SUAREZ, D. L., PANTIN-JACKWOOD, M. J., & KAPCZYNSKI, D. R. (2017). Use of Sequence-Independent, Single-Primer-Amplification (SISPA) for rapid detection, identification, and characterization of avian RNA viruses. *Virology*, *509*, 159-166.

14. CHRZASTEK, K., TENNAKOON, C., BIALY, D., FREIMANIS, G., FLANNERY, J., & SHELTON, H. (2022). A random priming amplification method for whole genome sequencing of SARS-CoV-2 virus. *BMC genomics*, *23*(1), 406.

15. CHENG, Y., LING, Z., & LI, L. (2020). The Intestinal Microbiota and Colorectal Cancer. *Frontiers in Immunology*, 11, 615056.

16. CONCEIÇÃO-NETO, N., YINDA, C. K., VAN RANST, M., & MATTHIJNSSENS, J. (2018). NetoVIR: Modular Approach to Customize Sample Preparation Procedures for Viral Metagenomics. *Methods in molecular biology* (pp. 85-95). Springer Science+Business Media.

17. NATIONAL RESEARCH COUNCIL (US) COMMITTEE ON METAGENOMICS: CHALLENGES AND FUNCTIONAL APPLICATIONS. (2007). *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet*. National Academies Press (US).

18. DECKER, C. J., & PARKER, R. (2014). Analysis of double-stranded RNA from microbial communities identifies double-stranded RNA virus-like elements. *Cell reports*, *7*(3), 898–906.

19. DUHAIME, M. B., DENG, L., POULOS, B. T., & SULLIVAN, M. B. (2012). Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environmental microbiology*, *14*(9), 2526–2537.

20. GAGLIA, M. M., & MUNGER, K. (2018). More than just oncogenes: mechanisms of tumorigenesis by human viruses. *Current opinion in virology*, *32*, 48–59.

21. GEERING, A. D. (2014). Caulimoviridae (Plant Pararetroviruses). *Wiley online library*. 10.1002/9780470015902.a0000746.

22. GREGORY, A. C., ZABLOCKI, O., ZAYED, A. A., HOWELL, A., BOLDUC, B., & SULLIVAN, M. B. (2020). The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. *Cell host & microbe*, *28*(5), 724–740.e8.

23. GONZALES-GUSTAVSON, E., TIMONEDA, N., VICTORIA, M., CABALLERO, A. C., ABRIL, J. F., BUTI, M., RODRIGUEZ-FRIAS, F., & GIRONES, R. (2017). Identification of sapovirus GV.2, astrovirus VA3 and novel anelloviruses in serum from patients with acute hepatitis of unknown aetiology. *PLOS ONE*, *12*(10), e0185911.

24. HORZ, H., SCHEER, S., VIANNA, M. E., & CONRADS, G. (2010). New methods for selective isolation of bacterial DNA from human clinical specimens. *Anaerobe*, *16*(1), 47-53.

25. HUMAN MICROBIOME JUMPSTART REFERENCE STRAINS CONSORTIUM, NELSON, K. E., WEINSTOCK, G. M., HIGHLANDER, S. K., WORLEY, K. C., CREASY, H. H., WORTMAN, J. R., RUSCH,

D. B., MITREVA, M., SODERGREN, E., CHINWALLA, A. T., FELDGARDEN, M., GEVERS, D., HAAS, B. J., MADUPU, R., WARD, D. V., BIRREN, B. W., GIBBS, R. A., METHE, B., PETROSINO, J. F., … ZHU, D. (2010). A catalog of reference genomes from the human microbiome. *Science*, 328(5981), 994–999.

26. ISSAEVA, N., KATZ, J. S., & PIPAS, J. M. (2018). Viral sequences in human cancer. *Virology*, *513*, 208-216.

27. KARLSSON, O. E., BELÁK, S., & GRANBERG, F. (2013). The effect of preprocessing by sequence-independent, single-primer amplification (SISPA) on metagenomic detection of viruses. *Biosecurity and bioterrorism: biodefense strategy, practice, and science*, *11 Suppl 1*, S227–S234.

28. KAZLAUSKAS, D., VARSANI, A., ZHANG, F., & KRUPOVIC, M. (2019). Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nature Communications*, 10 (1), 3425.

29. KHANNA, K., ESLAVATH, M. R., GANJU, L., MISHRA, K., SINGH, S. B., & KUMAR, B. (2021). Metagenomic insights into taxonomic and functional shifts under chronic exposure to hypobaric hypoxia induced microbial communities. *Journal of Proteins & Proteomics*, *12*(2), 105-114.

30. KIM, K., & BAE, J. (2011). Amplification Methods Bias Metagenomic Libraries of Uncultured Single-Stranded and Double-Stranded DNA Viruses. *Applied and Environmental Microbiology*, *77*(21), 7663-7668.

31. KONG, N., WHITNEY NG, ., AZARENE FOUTOUHI, ., CAROL HUANG, ., STOREY, D., WEIMER, B., & LENORE KELLY, (2015). Quality Control of High-Throughput Library Construction Pipeline for KAPA HTP Library Using an Agilent 2200 TapeStation (Version 1).

32. KOONIN, E. V., KRUPOVIC, M., & AGOL, V. I. (2021). The Baltimore Classification of Viruses 50 Years Later: How Does It Stand in the Light of Virus Evolution?. *Microbiology and molecular biology reviews : MMBR*, *85*(3), e0005321.

33. KOURÍ, A. L. A. (2018). El viroma humano. Implicaciones en la salud y enfermedad. *Revista Habanera de Ciencias Médicas*, *17*(3), 376-385.

34. KUHN, J. H., ADKINS, S., ALKHOVSKY, S. V., AVŠIČ-ŽUPANC, T., AYLLÓN, M. A., BAHL, J., BALKEMA-BUSCHMANN, A., BALLINGER, M. J., BANDTE, M., BEER, M., BEJERMAN, N., BERGERON, E., BIEDENKOPF, N., BIGARRÉ, L., BLAIR, C. D., BLASDELL, K. R., BRADFUTE, S. B., BRIESE, T., BROWN, P. D., . . . MARSHALL, S. H. (2022). 2022 taxonomic update of phylum Negarnaviricota (Riboviria: Orthornavirae), including the large orders Bunyavirales and Mononegavirales. *Archives of Virology*, *167*(12), 2857-2906.

35. LANGMEAD B, SALZBERG S. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods.* 9:357-359.

36. LAANTO, E., MÄNTYNEN, S., DE COLIBUS, L., MARJAKANGAS, J., GILLUM, A., STUART, D. I., RAVANTTI, J. J., HUISKONEN, J. T., & SUNDBERG, L. R. (2017). Virus found in a boreal lake links ssDNA and dsDNA viruses. *Proceedings of the National Academy of Sciences*, *114*(31), 8378–8383.

37. LEWANDOWSKI, K., BELL, A. J., MILES, R. W., CARNE, S., WOOLDRIDGE, D., MANSO, C. F., HENNESSY, N., BAILEY, D. P., PULLAN, S. T., GHARBIA, S. E., & VIPOND, R. (2017). The Effect of Nucleic Acid Extraction Platforms and Sample Storage on the Integrity of Viral RNA for Use in Whole Genome Sequencing. *The Journal of Molecular Diagnostics*, *19*(2), 303-312.

38. LIANG, G., & BUSHMAN, F. D. (2021). The human virome: assembly, composition and host interactions. *Nature Reviews Microbiology*, *19*(8), 514-527.

39. MALATHI, V. G., & RENUKA DEVI, P. (2019). ssDNA viruses: key players in global virome. *Virusdisease*, *30*(1), 3–12.

40. MENZEL, P., NG, K. & KROGH, A (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* 7, 11257.

41. MINOT, S., SINHA, R., CHEN, J., LI, H., KEILBAUGH, S. A., WU, G. D., LEWIS, J. D., & BUSHMAN, F. D. (2011). The human gut virome: inter-individual variation and dynamic response to diet. *Genome research*, *21*(10), 1616–1625.

42. MOYA, A., & BROCAL, V. P. (2018). The Human Virome: Methods and Protocols. *Methods in molecular biology.* Vol. 1838.

43. MOYA, A., LATORRE, A.., PÉREZ, V., D'AURIA, G., MARTINEZ, L., ARTACHO, A., JIMÉNEZ, N., SÁNCHEZ, A. B., SÁNCHEZ, V., FERRER, M., & BARBAS, C. (2017). Impacto de la microbiota en la oncogénesis colorrectal en pacientes con síndrome de Lynch. *Proyectos AECC*.

44. NERI, U., WOLF, Y. I., ROUX, S., CAMARGO, A., LEE, B. R., KAZLAUSKAS, D., CHEN, I. A., IVANOVA, N., ALLEN, L., PAEZ-ESPINO, D., BRYANT, D. A., BHAYA, D., CONSORTIUM, R. V. D., KRUPOVIC, M., DOLJA, V. V., KYRPIDES, N. C., & ZHANG, F. (2022). A Five-Fold Expansion of the Global RNA Virome Reveals Multiple New Clades of RNA Bacteriophages. *Social Science Research Network*. P4023-4037.E18.

45. R CORE TEAM. R: A LANGUAGE AND ENVIRONMENT FOR STATISTICAL COMPUTING. R Foundation for Statistical Computing, Vienna, Austria.

46. RIGAUD, G., GRANGE, T., & PICTET, R. (1991). *Nucleic Acids Research: Vol. 19, Issue 17*.

47. OGILVIE, L. A., & JONES, B. V. (2015). The human gut virome: a multifaceted majority. *Frontiers in microbiology*, *6*, 918.

48. REYES, G. R., & KIM, J. P. (1991). Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Molecular and Cellular Probes*, *5*(6), 473-481.

49. ROHWER F. (2003). Global phage diversity. *Cell*, *113*(2), 141.

50. ROKYTA, D. R., BURCH, C. L., CAUDLE, S. B., & WICHMAN, H. A. (2006). Horizontal gene transfer and the evolution of microvirid coliphage genomes. *Journal of bacteriology*, *188*(3), 1134–1142.

51. ROUDKO, V., BOZKUS, C. C., GREENBAUM, B., LUCAS, A. L., SAMSTEIN, R. M., & BHARDWAJ, N. (2021). Lynch Syndrome and MSI-H Cancers: From Mechanisms to "Off-The-Shelf" Cancer Vaccines. *Frontiers in Immunology*, 12, 757804.

52. ROUX, S., EMERSON, J. B., ELOE-FADROSH, E. A., & SULLIVAN, M. B. (2017). Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ*, *5*, e3817.

53. ROUX, S., KRUPOVIC, M., DALY, R. A., BORGES, A. L., NAYFACH, S., SCHULZ, F., SHARRAR, A., CARNEVALI, P. B. M., CHENG, J. F., IVANOVA, N., BONDY-DENOMY, J., WRIGHTON, K. C., WOYKE, T., VISEL, A., KYRPIDES, N. C., & ELOE-FADROSH, E. A. (2019). Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. *Nature microbiology*, *4*(11), 1895-1906.

54. CHEN, S., ZHOU, Y., CHEN, Y., & GU, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England)*, *34*(17), i884–i890.

55. SHKOPOROV, A. N., CLOONEY, A. G., SUTTON, T. L., RYAN, F. J., DALY, K., NOLAN, J., MCDONNELL, S. A., KHOKHLOVA, E. V., DRAPER, L. A., FORDE, A., GUERIN, E., VELAYUDHAN, V., ROSS, R. P., & HILL, C. (2019). The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. *Cell Host & Microbe*, *26*(4), 527-541.e5.

56. SIRANOSIAN, B. A., TAMBURINI, F. B., SHERLOCK, G., & BHATT, A. S. (2020). Acquisition, transmission and strain diversity of human gut-colonizing crAss-like phages. *Nature communications*, *11*(1), 280.

57. SIT, T. L., & ABOUHAIDAR, M. G. (1993). Infectious RNA transcripts derived from cloned cDNA of papaya mosaic virus: effect of mutations to the capsid and polymerase proteins. *Journal of General Virology*, *74*(6), 1133-1140.

58. STERN, J., MILLER, G. A., LI, X., & SAXENA, D. (2019). Virome and bacteriome: two sides of the same coin. *Current Opinion in Virology*, *37*, 37-43.

59. TAN, S. C., & YIAP, B. C. (2009). DNA, RNA, and Protein Extraction: The Past and The Present. *Journal of Biomedicine and Biotechnology*, *2009*, 1-10.

60. DANECEK, P., BONFIELD, J. K., LIDDLE, J., MARSHALL, J., OHAN, V., POLLARD, M. O., WHITWHAM, A., KEANE, T., MCCARTHY, S. A., DAVIES, R. M., & LI, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008.

61. VERHEYEN, J., KAISER, R., BOZIC, M., TIMMEN-WEGO, M., MAIER, B., & KESSLER, H. H. (2012). Extraction of viral nucleic acids: Comparison of five automated nucleic acid extraction platforms. *Journal of Clinical Virology*, *54*(3), 255-259.

62. WHITE, M. K., PAGANO, J. S., & KHALILI, K. (2014). Viruses and human cancers: a long road of discovery of molecular paradigms. *Clinical microbiology reviews*, *27*(3), 463–481.

63. WITLOX, K. J., NGUYEN, T. D., BRUGGINK, L. D., CATTON, M. G., & MARSHALL, J. L. (2008). A comparative evaluation of the sensitivity of two automated and two manual nucleic acid extraction methods for the detection of norovirus by RT-PCR. *Journal of Virological Methods*, *150*(1-2), 70-72.

64. WYLIE, T. N., WYLIE, K. M., HERTER, B. N., & STORCH, G. A. (2015). Enhanced virome sequencing using targeted sequence capture. *Genome research*, *25*(12), 1910–1920.

65. YANG, F., SUN, J., ZHONG, H., REN, H., ZHOU, H., LIN, Y., HAN, M., CHEN, B., LIAO, H., BRIX, S., LI, J., YANG, H., KRISTIANSEN, K., & ZHONG, H. (2020). Assessment of fecal DNA extraction protocols for metagenomic studies. *GigaScience*, *9*(7), giaa071.

66. YURGELUN, M. B., & HAMPEL, H. (2018). Recent Advances in Lynch Syndrome: Diagnosis, Treatment, and Cancer Prevention. *American Society of Clinical Oncology educational book*, *38*, 101-109.

67. ZHU, Y., SHANG, J., PENG, C., & SUN, Y. (2022). Phage family classification under *Caudoviricetes*: A review of current tools using the latest ICTV classification framework. *Frontiers in microbiology*, *13*, 1032186.

APPENDIX I.

Project relation with Sustainable Development Goals (SDG)

Relationship degree of this FDP project with the Sustainable Development Goals (SDG):

- SDG 3. **Good health and well-being:** the aim of this FDP is to improve the health and well-being of patients genetically diagnosed with Lynch syndrome.
- SDG 4. **Quality education:** as scientific work, this FDP hopes to contribute to education and science, with the opportunity to be shared among researchers.
- SDG 10. **Reduced inequalities:** this FDG is not focused in only one country, on the contrary, it aims to help to every researcher around all the globe which work is related to this project's information.

SUSTAINABLE DEVELOPMENT GOALS	HIGH	MEDIUM	LOW	NOT APPLICABLE
SDG 1. **No poverty**				X
SDG 2. **Zero hunger**				X
SDG 3. **Good health and well-being**	X			
SDG 4. **Quality education**		X		
SDG 5. **Gender equality**				X
SDG 6. **Clean water and sanitation**				X
SDG 7. **Affordable and clean energy**				X
SDG 8. **Decent work and economic growth**				X
SDG 9. **Industry, innovation and infrastructure**				X
SDG 10. **Reduced inequalities**		X		
SDG 11. **Sustainable cities and communities**				X
SDG 12. **Responsible consumption and production**				X
SDG 13. **Climate action**				X
SDG 14. **Life below water**				X
SDG 15. **Life on land**				X
SDG 16. **Peace, justice and strong institutions**				X
SDG 17. **Partnerships for the goals**				X

APPENDIX II. Full list of viral species identified

The colors used in the appendix are connected to the colors used in the text's figures. They are related to the type of virus according to a summarized Baltimore classification, being the virus marked in blue dsDNA viruses, grey for any kind of retrovirus, green dsRNA, yellow ssDNA, orange ssRNA and purple are the virus that could not be taxonomically assigned to any viral agent.

Genome type	Phylum	Species
dsDNA	Cossaviricota	*Gammapapillomavirus_sp.*
dsDNA	Cossaviricota	*Papillomaviridae__uc*
dsDNA	Nucleocytoviricota	*Acanthocystis_turfacea_Chlorella_virus_MN0810.1*
dsDNA	Nucleocytoviricota	*Paramecium_bursaria_Chlorella_virus_NY2A*
dsDNA	Nucleocytoviricota	*Prasinovirus__uc*
dsDNA	Nucleocytoviricota	*Erythrocytic_necrosis_virus*
dsDNA	Nucleocytoviricota	*Marseillevirus_LCMAC202*
dsDNA	Uroviricota	*Ackermannviridae_sp._ctQad106*
dsDNA	Uroviricota	*Ackermannviridae__uc*
dsDNA	Uroviricota	*Teetrevirus__uc*
dsDNA	Uroviricota	*Bacteroides_phage_ARB14*
dsDNA	Uroviricota	*Bacteroides_phage_DAC23*
dsDNA	Uroviricota	*Bacteroides_phage_SJC10*
dsDNA	Uroviricota	*Caudovirales_sp._ct0jG3*
dsDNA	Uroviricota	*Caudovirales_sp._ct0YK8*
dsDNA	Uroviricota	*Caudovirales_sp._ct2A51*
dsDNA	Uroviricota	*Caudovirales_sp._ct7Ux15*
dsDNA	Uroviricota	*Caudovirales_sp._ctaix4*
dsDNA	Uroviricota	*Caudovirales_sp._ctcLF7*
dsDNA	Uroviricota	*Caudovirales_sp._ctCpR1*
dsDNA	Uroviricota	*Caudovirales_sp._ctCVG11*
dsDNA	Uroviricota	*Caudovirales_sp._ctFWA4*
dsDNA	Uroviricota	*Caudovirales_sp._ctIsq18*
dsDNA	Uroviricota	*Caudovirales_sp._ctIZM3*
dsDNA	Uroviricota	*Caudovirales_sp._ctlwr10*
dsDNA	Uroviricota	*Caudovirales_sp._ctNZz8*
dsDNA	Uroviricota	*Caudovirales_sp._ctqPn17*
dsDNA	Uroviricota	*Caudovirales_sp._ctt3K6*
dsDNA	Uroviricota	*Caudovirales_sp._ctu3532*
dsDNA	Uroviricota	*Caudovirales_sp._ctU7I6*
dsDNA	Uroviricota	*Caudovirales_sp._ctUJJ3*
dsDNA	Uroviricota	*Caudovirales_sp._ctUL28*
dsDNA	Uroviricota	*Caudovirales_sp._ctVfb8*
dsDNA	Uroviricota	*Caudovirales_sp._ctyaR3*
dsDNA	Uroviricota	*Suwonvirus__uc*
dsDNA	Uroviricota	*Pantoea_phage_vB_PagS_AAS21*
dsDNA	Uroviricota	*Enterobacter_phage_Phc*
dsDNA	Uroviricota	*Baoshanvirus__uc*
dsDNA	Uroviricota	*Staphylococcus_phage_vB_SsapH-Golestan101-M*

dsDNA	Uroviricota	*Lactobacillus_virus_Lpa804*
dsDNA	Uroviricota	*Herelleviridae_sp._cttEB8*
dsDNA	Uroviricota	*Kayvirus__uc*
dsDNA	Uroviricota	*Enterococcus_phage_156*
dsDNA	Uroviricota	*Sepunavirus__uc*
dsDNA	Uroviricota	*Staphylococcus_phage_Twillingate*
dsDNA	Uroviricota	*Faecalibacterium_virus_Brigit*
dsDNA	Uroviricota	*Aeromonas_phage_AP1*
dsDNA	Uroviricota	*Salmonella_phage_SSBI34*
dsDNA	Uroviricota	*Emdodecavirus__uc*
dsDNA	Uroviricota	*Klebsiella_phage_vB_KpnM_15-38_KLPPOU148*
dsDNA	Uroviricota	*Erwinia_phage_phiEa116*
dsDNA	Uroviricota	*Salmonella_phage_vB_SenM_SB18*
dsDNA	Uroviricota	*Faecalibacterium_virus_Mushu*
dsDNA	Uroviricota	*Mydovirus__uc*
dsDNA	Uroviricota	*Acinetobacter_virus_133*
dsDNA	Uroviricota	*Aeromonas_phage_CF8*
dsDNA	Uroviricota	*Myoviridae_sp._ct04y17*
dsDNA	Uroviricota	*Myoviridae_sp._ct0e511*
dsDNA	Uroviricota	*Myoviridae_sp._ct0Tg8*
dsDNA	Uroviricota	*Myoviridae_sp._ct1CM14*
dsDNA	Uroviricota	*Myoviridae_sp._ct1IL4*
dsDNA	Uroviricota	*Myoviridae_sp._ct1TR10*
dsDNA	Uroviricota	*Myoviridae_sp._ct25F5*
dsDNA	Uroviricota	*Myoviridae_sp._ct2DO6*
dsDNA	Uroviricota	*Myoviridae_sp._ct2Qy24*
dsDNA	Uroviricota	*Myoviridae_sp._ct2th6*
dsDNA	Uroviricota	*Myoviridae_sp._ct3D84*
dsDNA	Uroviricota	*Myoviridae_sp._ct3hC12*
dsDNA	Uroviricota	*Myoviridae_sp._ct3Oc10*
dsDNA	Uroviricota	*Myoviridae_sp._ct3Sw5*
dsDNA	Uroviricota	*Myoviridae_sp._ct3tv2*
dsDNA	Uroviricota	*Myoviridae_sp._ct44j18*
dsDNA	Uroviricota	*Myoviridae_sp._ct4yW2*
dsDNA	Uroviricota	*Myoviridae_sp._ct55Z2*
dsDNA	Uroviricota	*Myoviridae_sp._ct5hB2*
dsDNA	Uroviricota	*Myoviridae_sp._ct5MF11*
dsDNA	Uroviricota	*Myoviridae_sp._ct5nJ10*
dsDNA	Uroviricota	*Myoviridae_sp._ct5ra14*
dsDNA	Uroviricota	*Myoviridae_sp._ct5Tq8*
dsDNA	Uroviricota	*Myoviridae_sp._ct5xX4*
dsDNA	Uroviricota	*Myoviridae_sp._ct5xZ3*
dsDNA	Uroviricota	*Myoviridae_sp._ct6aW5*
dsDNA	Uroviricota	*Myoviridae_sp._ct6eX13*
dsDNA	Uroviricota	*Myoviridae_sp._ct6nn14*
dsDNA	Uroviricota	*Myoviridae_sp._ct7113*
dsDNA	Uroviricota	*Myoviridae_sp._ct78050*
dsDNA	Uroviricota	*Myoviridae_sp._ct7CH26*
dsDNA	Uroviricota	*Myoviridae_sp._ct7Q419*

dsDNA	Uroviricota	*Myoviridae_sp._ct8Eu10*
dsDNA	Uroviricota	*Myoviridae_sp._ct8iP21*
dsDNA	Uroviricota	*Myoviridae_sp._ct8mY9*
dsDNA	Uroviricota	*Myoviridae_sp._ct9dX1*
dsDNA	Uroviricota	*Myoviridae_sp._ct9Fw19*
dsDNA	Uroviricota	*Myoviridae_sp._ct9MV2*
dsDNA	Uroviricota	*Myoviridae_sp._ct9Ns12*
dsDNA	Uroviricota	*Myoviridae_sp._ctA1z6*
dsDNA	Uroviricota	*Myoviridae_sp._cta6i12*
dsDNA	Uroviricota	*Myoviridae_sp._ctaMv1*
dsDNA	Uroviricota	*Myoviridae_sp._ctBbR2*
dsDNA	Uroviricota	*Myoviridae_sp._ctBCv9*
dsDNA	Uroviricota	*Myoviridae_sp._ctBtV12*
dsDNA	Uroviricota	*Myoviridae_sp._ctbwh6*
dsDNA	Uroviricota	*Myoviridae_sp._ctbWL16*
dsDNA	Uroviricota	*Myoviridae_sp._ctcaJ26*
dsDNA	Uroviricota	*Myoviridae_sp._ctcFb5*
dsDNA	Uroviricota	*Myoviridae_sp._ctCjb12*
dsDNA	Uroviricota	*Myoviridae_sp._ctCL221*
dsDNA	Uroviricota	*Myoviridae_sp._ctcPl3*
dsDNA	Uroviricota	*Myoviridae_sp._ctcwu24*
dsDNA	Uroviricota	*Myoviridae_sp._ctdWz11*
dsDNA	Uroviricota	*Myoviridae_sp._ctDzM5*
dsDNA	Uroviricota	*Myoviridae_sp._ctE3x18*
dsDNA	Uroviricota	*Myoviridae_sp._cte5Z19*
dsDNA	Uroviricota	*Myoviridae_sp._cteBs22*
dsDNA	Uroviricota	*Myoviridae_sp._ctEg02*
dsDNA	Uroviricota	*Myoviridae_sp._ctEtC12*
dsDNA	Uroviricota	*Myoviridae_sp._ctEXz2*
dsDNA	Uroviricota	*Myoviridae_sp._ctfA14*
dsDNA	Uroviricota	*Myoviridae_sp._ctfJc17*
dsDNA	Uroviricota	*Myoviridae_sp._ctfuG5*
dsDNA	Uroviricota	*Myoviridae_sp._ctFYw8*
dsDNA	Uroviricota	*Myoviridae_sp._ctg8M33*
dsDNA	Uroviricota	*Myoviridae_sp._ctGk74*
dsDNA	Uroviricota	*Myoviridae_sp._ctgpD8*
dsDNA	Uroviricota	*Myoviridae_sp._ctGRa7*
dsDNA	Uroviricota	*Myoviridae_sp._ctgXa1*
dsDNA	Uroviricota	*Myoviridae_sp._ctgXL3*
dsDNA	Uroviricota	*Myoviridae_sp._ctHaT25*
dsDNA	Uroviricota	*Myoviridae_sp._ctHIt1*
dsDNA	Uroviricota	*Myoviridae_sp._ctHMa1*
dsDNA	Uroviricota	*Myoviridae_sp._cthmz15*
dsDNA	Uroviricota	*Myoviridae_sp._ctHP32*
dsDNA	Uroviricota	*Myoviridae_sp._cthRr4*
dsDNA	Uroviricota	*Myoviridae_sp._ctI7W9*
dsDNA	Uroviricota	*Myoviridae_sp._ctilS8*
dsDNA	Uroviricota	*Myoviridae_sp._ctino4*
dsDNA	Uroviricota	*Myoviridae_sp._ctISH16*

dsDNA	Uroviricota	*Myoviridae_sp._ctJfU3*
dsDNA	Uroviricota	*Myoviridae_sp._ctjH82*
dsDNA	Uroviricota	*Myoviridae_sp._ctK7P4*
dsDNA	Uroviricota	*Myoviridae_sp._ctKaW6*
dsDNA	Uroviricota	*Myoviridae_sp._ctKkB1*
dsDNA	Uroviricota	*Myoviridae_sp._ctkOm7*
dsDNA	Uroviricota	*Myoviridae_sp._ctKZW4*
dsDNA	Uroviricota	*Myoviridae_sp._ctlD98*
dsDNA	Uroviricota	*Myoviridae_sp._ctlnK45*
dsDNA	Uroviricota	*Myoviridae_sp._ctLtm40*
dsDNA	Uroviricota	*Myoviridae_sp._ctLx49*
dsDNA	Uroviricota	*Myoviridae_sp._ctm8X17*
dsDNA	Uroviricota	*Myoviridae_sp._ctMvU7*
dsDNA	Uroviricota	*Myoviridae_sp._ctNDZ29*
dsDNA	Uroviricota	*Myoviridae_sp._ctngn1*
dsDNA	Uroviricota	*Myoviridae_sp._ctNhr24*
dsDNA	Uroviricota	*Myoviridae_sp._ctNqw6*
dsDNA	Uroviricota	*Myoviridae_sp._ctnzH2*
dsDNA	Uroviricota	*Myoviridae_sp._ctO7l1*
dsDNA	Uroviricota	*Myoviridae_sp._ctoIO8*
dsDNA	Uroviricota	*Myoviridae_sp._ctOpw2*
dsDNA	Uroviricota	*Myoviridae_sp._ctOv05*
dsDNA	Uroviricota	*Myoviridae_sp._ctP4M4*
dsDNA	Uroviricota	*Myoviridae_sp._ctP6q2*
dsDNA	Uroviricota	*Myoviridae_sp._ctp7F23*
dsDNA	Uroviricota	*Myoviridae_sp._ctPGO22*
dsDNA	Uroviricota	*Myoviridae_sp._ctplG2*
dsDNA	Uroviricota	*Myoviridae_sp._ctPoO4*
dsDNA	Uroviricota	*Myoviridae_sp._ctPVE25*
dsDNA	Uroviricota	*Myoviridae_sp._ctQ6D10*
dsDNA	Uroviricota	*Myoviridae_sp._ctq8k5*
dsDNA	Uroviricota	*Myoviridae_sp._ctq9w2*
dsDNA	Uroviricota	*Myoviridae_sp._ctQf419*
dsDNA	Uroviricota	*Myoviridae_sp._ctQQg4*
dsDNA	Uroviricota	*Myoviridae_sp._ctQYc56*
dsDNA	Uroviricota	*Myoviridae_sp._ctr0w28*
dsDNA	Uroviricota	*Myoviridae_sp._ctrEx11*
dsDNA	Uroviricota	*Myoviridae_sp._ctRrG7*
dsDNA	Uroviricota	*Myoviridae_sp._ctSBU9*
dsDNA	Uroviricota	*Myoviridae_sp._ctslb3*
dsDNA	Uroviricota	*Myoviridae_sp._ctsK93*
dsDNA	Uroviricota	*Myoviridae_sp._ctt8G1*
dsDNA	Uroviricota	*Myoviridae_sp._ctTfa5*
dsDNA	Uroviricota	*Myoviridae_sp._ctTOm1*
dsDNA	Uroviricota	*Myoviridae_sp._cttp71*
dsDNA	Uroviricota	*Myoviridae_sp._ctTrm2*
dsDNA	Uroviricota	*Myoviridae_sp._cttWQ44*
dsDNA	Uroviricota	*Myoviridae_sp._ctu6J18*
dsDNA	Uroviricota	*Myoviridae_sp._ctuIn11*

dsDNA	Uroviricota	*Myoviridae_sp._ctULz28*
dsDNA	Uroviricota	*Myoviridae_sp._ctuSi15*
dsDNA	Uroviricota	*Myoviridae_sp._ctUX613*
dsDNA	Uroviricota	*Myoviridae_sp._ctv9K3*
dsDNA	Uroviricota	*Myoviridae_sp._ctVCj30*
dsDNA	Uroviricota	*Myoviridae_sp._ctVeR24*
dsDNA	Uroviricota	*Myoviridae_sp._ctW7Z6*
dsDNA	Uroviricota	*Myoviridae_sp._ctWaE18*
dsDNA	Uroviricota	*Myoviridae_sp._ctWPU11*
dsDNA	Uroviricota	*Myoviridae_sp._ctWXg38*
dsDNA	Uroviricota	*Myoviridae_sp._ctX172*
dsDNA	Uroviricota	*Myoviridae_sp._ctxi06*
dsDNA	Uroviricota	*Myoviridae_sp._ctxlX31*
dsDNA	Uroviricota	*Myoviridae_sp._ctxpQ22*
dsDNA	Uroviricota	*Myoviridae_sp._ctXRl20*
dsDNA	Uroviricota	*Myoviridae_sp._ctXVO17*
dsDNA	Uroviricota	*Myoviridae_sp._ctXwe21*
dsDNA	Uroviricota	*Myoviridae_sp._ctYGJ17*
dsDNA	Uroviricota	*Myoviridae_sp._ctyhU11*
dsDNA	Uroviricota	*Myoviridae_sp._ctyWv1*
dsDNA	Uroviricota	*Myoviridae_sp._ctZDd15*
dsDNA	Uroviricota	*Myoviridae_sp._ctzS633*
dsDNA	Uroviricota	*Myoviridae_sp._ctzyI3*
dsDNA	Uroviricota	*Serratia_phage_vB_SspM_BZS1*
dsDNA	Uroviricota	*Shewanella_sp._phage_1/41*
dsDNA	Uroviricota	*Synechococcus_phage_S-CRM01*
dsDNA	Uroviricota	*Vibrio_phage_1.005.O._10N.286.48.F2*
dsDNA	Uroviricota	*Vibrio_phage_vB_VmeM-32*
dsDNA	Uroviricota	*Pseudomonas_phage_201phi2-1*
dsDNA	Uroviricota	*Erwinia_virus_Risingsun*
dsDNA	Uroviricota	*Faecalibacterium_virus_Taranis*
dsDNA	Uroviricota	*Faecalibacterium_virus_Toutatis*
dsDNA	Uroviricota	*Kuravirus__uc*
dsDNA	Uroviricota	*Burkholderia_virus_Bcepil02*
dsDNA	Uroviricota	*Bacteroides_phage_crAss002*
dsDNA	Uroviricota	*CrAss-like_virus_sp._ctOWe7*
dsDNA	Uroviricota	*CrAss-like_virus_sp._ctRQZ5*
dsDNA	Uroviricota	*CrAss-like_virus_sp._ctt4r3*
dsDNA	Uroviricota	*CrAss-like_virus_sp._ctUXy32*
dsDNA	Uroviricota	*crAssphage_cr1_1*
dsDNA	Uroviricota	*crAssphage_cr10_1*
dsDNA	Uroviricota	*crAssphage_cr107_1*
dsDNA	Uroviricota	*crAssphage_cr109_1*
dsDNA	Uroviricota	*crAssphage_cr111_1*
dsDNA	Uroviricota	*crAssphage_cr112_1*
dsDNA	Uroviricota	*crAssphage_cr113_1*
dsDNA	Uroviricota	*crAssphage_cr115_1*
dsDNA	Uroviricota	*crAssphage_cr3_1*
dsDNA	Uroviricota	*crAssphage_cr50_1*

dsDNA	Uroviricota	*crAssphage_cr52_1*
dsDNA	Uroviricota	*crAssphage_cr53_1*
dsDNA	Uroviricota	*crAssphage_cr55_1*
dsDNA	Uroviricota	*crAssphage_cr56_1*
dsDNA	Uroviricota	*crAssphage_cr60_1*
dsDNA	Uroviricota	*CrAssphage_sp.*
dsDNA	Uroviricota	*CrAssphage_sp._C0521BD4*
dsDNA	Uroviricota	*CrAssphage_sp._C0521BW15*
dsDNA	Uroviricota	*Escherichia_phage_YDC107_2*
dsDNA	Uroviricota	*Podoviridae_sp._ct1h53*
dsDNA	Uroviricota	*Podoviridae_sp._ct3k57*
dsDNA	Uroviricota	*Podoviridae_sp._ct5O42*
dsDNA	Uroviricota	*Podoviridae_sp._ct7Kl21*
dsDNA	Uroviricota	*Podoviridae_sp._ct8Lf7*
dsDNA	Uroviricota	*Podoviridae_sp._ct90d35*
dsDNA	Uroviricota	*Podoviridae_sp._ct9H612*
dsDNA	Uroviricota	*Podoviridae_sp._ctAmM4*
dsDNA	Uroviricota	*Podoviridae_sp._ctAV815*
dsDNA	Uroviricota	*Podoviridae_sp._ctC2p8*
dsDNA	Uroviricota	*Podoviridae_sp._ctcKt3*
dsDNA	Uroviricota	*Podoviridae_sp._ctDd04*
dsDNA	Uroviricota	*Podoviridae_sp._ctdet19*
dsDNA	Uroviricota	*Podoviridae_sp._ctDgT26*
dsDNA	Uroviricota	*Podoviridae_sp._ctdKF3*
dsDNA	Uroviricota	*Podoviridae_sp._ctdRZ1*
dsDNA	Uroviricota	*Podoviridae_sp._ctfAL26*
dsDNA	Uroviricota	*Podoviridae_sp._ctFkM10*
dsDNA	Uroviricota	*Podoviridae_sp._ctfN46*
dsDNA	Uroviricota	*Podoviridae_sp._cthJQ11*
dsDNA	Uroviricota	*Podoviridae_sp._ctHm32*
dsDNA	Uroviricota	*Podoviridae_sp._ctHMt20*
dsDNA	Uroviricota	*Podoviridae_sp._cthVG1*
dsDNA	Uroviricota	*Podoviridae_sp._ctiHu16*
dsDNA	Uroviricota	*Podoviridae_sp._ctIi96*
dsDNA	Uroviricota	*Podoviridae_sp._ctIKM86*
dsDNA	Uroviricota	*Podoviridae_sp._ctIlO27*
dsDNA	Uroviricota	*Podoviridae_sp._ctIpM11*
dsDNA	Uroviricota	*Podoviridae_sp._ctiwu7*
dsDNA	Uroviricota	*Podoviridae_sp._ctJaJ36*
dsDNA	Uroviricota	*Podoviridae_sp._ctK5Q1*
dsDNA	Uroviricota	*Podoviridae_sp._ctKmJ5*
dsDNA	Uroviricota	*Podoviridae_sp._ctLPy3*
dsDNA	Uroviricota	*Podoviridae_sp._ctn7K25*
dsDNA	Uroviricota	*Podoviridae_sp._ctO1718*
dsDNA	Uroviricota	*Podoviridae_sp._ctQuf7*
dsDNA	Uroviricota	*Podoviridae_sp._ctrub15*
dsDNA	Uroviricota	*Podoviridae_sp._ctU7u6*
dsDNA	Uroviricota	*Podoviridae_sp._ctUT63*
dsDNA	Uroviricota	*Podoviridae_sp._ctW0z17*

dsDNA	Uroviricota	*Podoviridae_sp._ctWeH21*
dsDNA	Uroviricota	*Podoviridae_sp._ctwJH20*
dsDNA	Uroviricota	*Podoviridae_sp._ctxqo3*
dsDNA	Uroviricota	*Podoviridae_sp._ctY3D12*
dsDNA	Uroviricota	*Podoviridae_sp._ctZ5d16*
dsDNA	Uroviricota	*Podoviridae_sp._ctz6O13*
dsDNA	Uroviricota	*Podoviridae_sp._ctzXp5*
dsDNA	Uroviricota	*Podovirus_SOG*
dsDNA	Uroviricota	*Escherichia_virus_Skarpretter*
dsDNA	Uroviricota	*Traversvirus__uc*
dsDNA	Uroviricota	*Enterobacter_phage_Tyrion*
dsDNA	Uroviricota	*Uetakevirus__uc*
dsDNA	Uroviricota	*Staphylococcus_phage_SeAlphi*
dsDNA	Uroviricota	*Staphylococcus_virus_St134*
dsDNA	Uroviricota	*Beecentumtrevirus__uc*
dsDNA	Uroviricota	*Bacillus_virus_GA1*
dsDNA	Uroviricota	*Bacillus_virus_SRT01hs*
dsDNA	Uroviricota	*Salasvirus__uc*
dsDNA	Uroviricota	*Enquatrovirus__uc*
dsDNA	Uroviricota	*uncultured_N4_virus*
dsDNA	Uroviricota	*Gamaleyavirus__uc*
dsDNA	Uroviricota	*Plymouthvirus__uc*
dsDNA	Uroviricota	*Pylasvirus__uc*
dsDNA	Uroviricota	*Stenotrophomonas_phage_Philippe*
dsDNA	Uroviricota	*Audreyjarvisvirus__uc*
dsDNA	Uroviricota	*Ceduovirus__uc*
dsDNA	Uroviricota	*Lactococcus_phage_74001*
dsDNA	Uroviricota	*Lactococcus_phage_BIM_BV-114*
dsDNA	Uroviricota	*Lactococcus_phage_CB17*
dsDNA	Uroviricota	*Lactococcus_phage_CW09*
dsDNA	Uroviricota	*Lactococcus_virus_05802*
dsDNA	Uroviricota	*Lactococcus_virus_37203*
dsDNA	Uroviricota	*Lactococcus_virus_50102*
dsDNA	Uroviricota	*Lactococcus_virus_50504*
dsDNA	Uroviricota	*Lactococcus_virus_62402*
dsDNA	Uroviricota	*Lactococcus_virus_62403*
dsDNA	Uroviricota	*Lactococcus_virus_62606*
dsDNA	Uroviricota	*Lactococcus_virus_bIBB14*
dsDNA	Uroviricota	*Lactococcus_virus_bIBBA3*
dsDNA	Uroviricota	*Lactococcus_virus_bIBBAm4*
dsDNA	Uroviricota	*Lactococcus_virus_bIBBE1*
dsDNA	Uroviricota	*Lactococcus_virus_bIBBp6-4*
dsDNA	Uroviricota	*Lactococcus_virus_blBB94p4*
dsDNA	Uroviricota	*Lactococcus_virus_c2*
dsDNA	Uroviricota	*Lactococcus_virus_CHPC1020*
dsDNA	Uroviricota	*Lactococcus_virus_CHPC116*
dsDNA	Uroviricota	*Lactococcus_virus_CHPC1170*
dsDNA	Uroviricota	*Lactococcus_virus_CHPC122*
dsDNA	Uroviricota	*Lactococcus_virus_CHPC1242*

dsDNA	Uroviricota	*Lactococcus_virus_D4412*
dsDNA	Uroviricota	*Lactococcus_virus_LacS15*
dsDNA	Uroviricota	*Lactococcus_virus_M5938*
dsDNA	Uroviricota	*Lactococcus_virus_PCB1*
dsDNA	Uroviricota	*Corynebacterium_phage_Kimchi1738*
dsDNA	Uroviricota	*Cequinquevirus__uc*
dsDNA	Uroviricota	*Lactobacillus_virus_Ld25A*
dsDNA	Uroviricota	*Caulobacter_virus_CcrColossus*
dsDNA	Uroviricota	*Dismasvirus__uc*
dsDNA	Uroviricota	*Flavobacterium_virus_Hattifnatt*
dsDNA	Uroviricota	*Korravirus__uc*
dsDNA	Uroviricota	*Enterobacteria_phage_LC159*
dsDNA	Uroviricota	*Escherichia_phage_Lambda_ev207*
dsDNA	Uroviricota	*Escherichia_virus_Lambda_4A7*
dsDNA	Uroviricota	*Leuconostoc_virus_LDG*
dsDNA	Uroviricota	*Leuconostoc_virus_LN03*
dsDNA	Uroviricota	*Leuconostoc_virus_LN04*
dsDNA	Uroviricota	*Leuconostoc_virus_P965*
dsDNA	Uroviricota	*Limdunavirus__uc*
dsDNA	Uroviricota	*Moineauvirus__uc*
dsDNA	Uroviricota	*Streptococcus_phage_CHPC1041*
dsDNA	Uroviricota	*Streptococcus_phage_SFi18*
dsDNA	Uroviricota	*Faecalibacterium_virus_Oengus*
dsDNA	Uroviricota	*Propionibacterium_phage_Aquarius*
dsDNA	Uroviricota	*Propionibacterium_phage_Enoki*
dsDNA	Uroviricota	*Propionibacterium_virus_P1001*
dsDNA	Uroviricota	*Propionibacterium_virus_PAD20*
dsDNA	Uroviricota	*Propionibacterium_virus_PHL095N00*
dsDNA	Uroviricota	*Propionibacterium_virus_PHL112N00*
dsDNA	Uroviricota	*Ravarandavirus__uc*
dsDNA	Uroviricota	*Rimavirus__uc*
dsDNA	Uroviricota	*Lactococcus_virus_BK5T*
dsDNA	Uroviricota	*Sextaecvirus__uc*
dsDNA	Uroviricota	*Bacillus_phage_vB_BanS_Chewbecca*
dsDNA	Uroviricota	*Bifidobacterium_phage_Bbif-1*
dsDNA	Uroviricota	*Bifidobacterium_phage_PMBT6*
dsDNA	Uroviricota	*Corynebacterium_phage_EmiRose*
dsDNA	Uroviricota	*Cylindrospermopsis_raciborskii_virus_RM-2018a*
dsDNA	Uroviricota	*Erwinia_phage_Gungnir39*
dsDNA	Uroviricota	*Gardnerella_phage_vB_Gva_AB1*
dsDNA	Uroviricota	*Gordonia_phage_GiKK*
dsDNA	Uroviricota	*Lactobacillus_phage_ATCCB*
dsDNA	Uroviricota	*Lactococcus_phage_38502*
dsDNA	Uroviricota	*Lactococcus_phage_bIL312*
dsDNA	Uroviricota	*Lactococcus_phage_C41431*
dsDNA	Uroviricota	*Lactococcus_phage_CHPC836*
dsDNA	Uroviricota	*Lactococcus_phage_phi31*
dsDNA	Uroviricota	*Parabacteroides_phage_PDS1*
dsDNA	Uroviricota	*Psychrobacillus_phage_Perkons*

dsDNA	Uroviricota	*Rhizobium_phage_RHph_TM16*
dsDNA	Uroviricota	*Ruminococcus_phage_phiRM10*
dsDNA	Uroviricota	*Siphoviridae_sp._ct0Bp21*
dsDNA	Uroviricota	*Siphoviridae_sp._ct0Ci105*
dsDNA	Uroviricota	*Siphoviridae_sp._ct0hG5*
dsDNA	Uroviricota	*Siphoviridae_sp._ct0UO21*
dsDNA	Uroviricota	*Siphoviridae_sp._ct0Wl9*
dsDNA	Uroviricota	*Siphoviridae_sp._ct0X023*
dsDNA	Uroviricota	*Siphoviridae_sp._ct0Xn2*
dsDNA	Uroviricota	*Siphoviridae_sp._ct13O11*
dsDNA	Uroviricota	*Siphoviridae_sp._ct16C7*
dsDNA	Uroviricota	*Siphoviridae_sp._ct1E017*
dsDNA	Uroviricota	*Siphoviridae_sp._ct1Eo1*
dsDNA	Uroviricota	*Siphoviridae_sp._ct1gv6*
dsDNA	Uroviricota	*Siphoviridae_sp._ct1TR2*
dsDNA	Uroviricota	*Siphoviridae_sp._ct1yA16*
dsDNA	Uroviricota	*Siphoviridae_sp._ct1Zj2*
dsDNA	Uroviricota	*Siphoviridae_sp._ct2hZ16*
dsDNA	Uroviricota	*Siphoviridae_sp._ct2QJ10*
dsDNA	Uroviricota	*Siphoviridae_sp._ct2vX3*
dsDNA	Uroviricota	*Siphoviridae_sp._ct39g3*
dsDNA	Uroviricota	*Siphoviridae_sp._ct3an14*
dsDNA	Uroviricota	*Siphoviridae_sp._ct3es5*
dsDNA	Uroviricota	*Siphoviridae_sp._ct3lF2*
dsDNA	Uroviricota	*Siphoviridae_sp._ct3r22*
dsDNA	Uroviricota	*Siphoviridae_sp._ct3yx7*
dsDNA	Uroviricota	*Siphoviridae_sp._ct3z32*
dsDNA	Uroviricota	*Siphoviridae_sp._ct47J5*
dsDNA	Uroviricota	*Siphoviridae_sp._ct47y1*
dsDNA	Uroviricota	*Siphoviridae_sp._ct4aE30*
dsDNA	Uroviricota	*Siphoviridae_sp._ct4Am4*
dsDNA	Uroviricota	*Siphoviridae_sp._ct4be24*
dsDNA	Uroviricota	*Siphoviridae_sp._ct4me22*
dsDNA	Uroviricota	*Siphoviridae_sp._ct4T77*
dsDNA	Uroviricota	*Siphoviridae_sp._ct4Z13*
dsDNA	Uroviricota	*Siphoviridae_sp._ct57T16*
dsDNA	Uroviricota	*Siphoviridae_sp._ct58H1*
dsDNA	Uroviricota	*Siphoviridae_sp._ct5co22*
dsDNA	Uroviricota	*Siphoviridae_sp._ct5FX1*
dsDNA	Uroviricota	*Siphoviridae_sp._ct5jB2*
dsDNA	Uroviricota	*Siphoviridae_sp._ct5Px37*
dsDNA	Uroviricota	*Siphoviridae_sp._ct5qs5*
dsDNA	Uroviricota	*Siphoviridae_sp._ct5YG1*
dsDNA	Uroviricota	*Siphoviridae_sp._ct5zp6*
dsDNA	Uroviricota	*Siphoviridae_sp._ct6d71*
dsDNA	Uroviricota	*Siphoviridae_sp._ct6JW23*
dsDNA	Uroviricota	*Siphoviridae_sp._ct6Ob18*
dsDNA	Uroviricota	*Siphoviridae_sp._ct6oU4*
dsDNA	Uroviricota	*Siphoviridae_sp._ct6ro14*

dsDNA	Uroviricota	*Siphoviridae_sp._ct6tD39*
dsDNA	Uroviricota	*Siphoviridae_sp._ct73D3*
dsDNA	Uroviricota	*Siphoviridae_sp._ct73V17*
dsDNA	Uroviricota	*Siphoviridae_sp._ct7BG1*
dsDNA	Uroviricota	*Siphoviridae_sp._ct7es18*
dsDNA	Uroviricota	*Siphoviridae_sp._ct7EW56*
dsDNA	Uroviricota	*Siphoviridae_sp._ct7JV2*
dsDNA	Uroviricota	*Siphoviridae_sp._ct7yc1*
dsDNA	Uroviricota	*Siphoviridae_sp._ct8hB11*
dsDNA	Uroviricota	*Siphoviridae_sp._ct8hR1*
dsDNA	Uroviricota	*Siphoviridae_sp._ct8LQ5*
dsDNA	Uroviricota	*Siphoviridae_sp._ct8LX107*
dsDNA	Uroviricota	*Siphoviridae_sp._ct8NQ14*
dsDNA	Uroviricota	*Siphoviridae_sp._ct9JD14*
dsDNA	Uroviricota	*Siphoviridae_sp._ct9mC1*
dsDNA	Uroviricota	*Siphoviridae_sp._ct9Y44*
dsDNA	Uroviricota	*Siphoviridae_sp._cta6m1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctaA31*
dsDNA	Uroviricota	*Siphoviridae_sp._ctacm4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctAnS47*
dsDNA	Uroviricota	*Siphoviridae_sp._ctaqI5*
dsDNA	Uroviricota	*Siphoviridae_sp._ctAUQ2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctAvy12*
dsDNA	Uroviricota	*Siphoviridae_sp._ctb1k4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctB9E3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctB9N2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctBAZ2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctBCI9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctbgC51*
dsDNA	Uroviricota	*Siphoviridae_sp._ctBLh2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctbrg2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctBtS10*
dsDNA	Uroviricota	*Siphoviridae_sp._ctCb814*
dsDNA	Uroviricota	*Siphoviridae_sp._ctCIv11*
dsDNA	Uroviricota	*Siphoviridae_sp._ctcj91*
dsDNA	Uroviricota	*Siphoviridae_sp._ctCjJ10*
dsDNA	Uroviricota	*Siphoviridae_sp._ctcK97*
dsDNA	Uroviricota	*Siphoviridae_sp._ctCsv15*
dsDNA	Uroviricota	*Siphoviridae_sp._ctcUB23*
dsDNA	Uroviricota	*Siphoviridae_sp._ctcuE16*
dsDNA	Uroviricota	*Siphoviridae_sp._ctcx61*
dsDNA	Uroviricota	*Siphoviridae_sp._ctD2Q91*
dsDNA	Uroviricota	*Siphoviridae_sp._ctD6g5*
dsDNA	Uroviricota	*Siphoviridae_sp._ctd9R8*
dsDNA	Uroviricota	*Siphoviridae_sp._ctdau33*
dsDNA	Uroviricota	*Siphoviridae_sp._ctDCt3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctdHi7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctDhw1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctDlU28*

dsDNA	Uroviricota	*Siphoviridae_sp._ctdm01*
dsDNA	Uroviricota	*Siphoviridae_sp._ctDmQ3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctdmY20*
dsDNA	Uroviricota	*Siphoviridae_sp._ctDo63*
dsDNA	Uroviricota	*Siphoviridae_sp._ctDsE1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctdvJ3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctDXu9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctdYc1*
dsDNA	Uroviricota	*Siphoviridae_sp._cte421*
dsDNA	Uroviricota	*Siphoviridae_sp._ctE6L85*
dsDNA	Uroviricota	*Siphoviridae_sp._ctedi74*
dsDNA	Uroviricota	*Siphoviridae_sp._ctEEM24*
dsDNA	Uroviricota	*Siphoviridae_sp._ctEfY6*
dsDNA	Uroviricota	*Siphoviridae_sp._cteHV32*
dsDNA	Uroviricota	*Siphoviridae_sp._ctEIp38*
dsDNA	Uroviricota	*Siphoviridae_sp._ctEJG5*
dsDNA	Uroviricota	*Siphoviridae_sp._ctEJj1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctEkS11*
dsDNA	Uroviricota	*Siphoviridae_sp._ctEqU3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctES717*
dsDNA	Uroviricota	*Siphoviridae_sp._ctETQ12*
dsDNA	Uroviricota	*Siphoviridae_sp._ctevH2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctf8W5*
dsDNA	Uroviricota	*Siphoviridae_sp._ctfbh2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctfDm7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctfeV1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctFPV4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctFSL3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctfza2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctg0K17*
dsDNA	Uroviricota	*Siphoviridae_sp._ctg8V11*
dsDNA	Uroviricota	*Siphoviridae_sp._ctGa111*
dsDNA	Uroviricota	*Siphoviridae_sp._ctGas1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctgaY24*
dsDNA	Uroviricota	*Siphoviridae_sp._ctgbm9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctGDt6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctGFb30*
dsDNA	Uroviricota	*Siphoviridae_sp._ctGMq5*
dsDNA	Uroviricota	*Siphoviridae_sp._ctgN495*
dsDNA	Uroviricota	*Siphoviridae_sp._ctGoR6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctGQT3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctgyq20*
dsDNA	Uroviricota	*Siphoviridae_sp._ctGyV19*
dsDNA	Uroviricota	*Siphoviridae_sp._ctH8e11*
dsDNA	Uroviricota	*Siphoviridae_sp._cthae16*
dsDNA	Uroviricota	*Siphoviridae_sp._ctHAs12*
dsDNA	Uroviricota	*Siphoviridae_sp._cthEJ2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctHiz26*
dsDNA	Uroviricota	*Siphoviridae_sp._cthjx9*

dsDNA	Uroviricota	*Siphoviridae_sp._ctHn727*
dsDNA	Uroviricota	*Siphoviridae_sp._ctHNH2*
dsDNA	Uroviricota	*Siphoviridae_sp._cthrG7*
dsDNA	Uroviricota	*Siphoviridae_sp._cthRu26*
dsDNA	Uroviricota	*Siphoviridae_sp._cthu813*
dsDNA	Uroviricota	*Siphoviridae_sp._cti0B23*
dsDNA	Uroviricota	*Siphoviridae_sp._cti6K1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctIbi23*
dsDNA	Uroviricota	*Siphoviridae_sp._ctigT3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctiMP24*
dsDNA	Uroviricota	*Siphoviridae_sp._ctiMX17*
dsDNA	Uroviricota	*Siphoviridae_sp._ctitf6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctiuu37*
dsDNA	Uroviricota	*Siphoviridae_sp._ctIwT7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctIyp7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctJ0s2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctj6w2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctj912*
dsDNA	Uroviricota	*Siphoviridae_sp._ctJbC4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctjbm8*
dsDNA	Uroviricota	*Siphoviridae_sp._ctjfQ5*
dsDNA	Uroviricota	*Siphoviridae_sp._ctJj91*
dsDNA	Uroviricota	*Siphoviridae_sp._ctJLl6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctJLm32*
dsDNA	Uroviricota	*Siphoviridae_sp._ctjsp22*
dsDNA	Uroviricota	*Siphoviridae_sp._ctk4d14*
dsDNA	Uroviricota	*Siphoviridae_sp._ctkcl3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctKeG8*
dsDNA	Uroviricota	*Siphoviridae_sp._ctkfT29*
dsDNA	Uroviricota	*Siphoviridae_sp._ctkI07*
dsDNA	Uroviricota	*Siphoviridae_sp._ctkJH11*
dsDNA	Uroviricota	*Siphoviridae_sp._ctkkB9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctkKt3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctKvA22*
dsDNA	Uroviricota	*Siphoviridae_sp._ctKXi8*
dsDNA	Uroviricota	*Siphoviridae_sp._ctKy93*
dsDNA	Uroviricota	*Siphoviridae_sp._ctL0q1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctL4w2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctL5G6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctL7J9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctLAw30*
dsDNA	Uroviricota	*Siphoviridae_sp._ctLgc23*
dsDNA	Uroviricota	*Siphoviridae_sp._ctlHU7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctLKT1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctLmu1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctLR131*
dsDNA	Uroviricota	*Siphoviridae_sp._ctLsx2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctLUY1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctM3g2*

dsDNA	Uroviricota	*Siphoviridae_sp._ctM6i4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctm7X10*
dsDNA	Uroviricota	*Siphoviridae_sp._ctm8l1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctMAv2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctMCY8*
dsDNA	Uroviricota	*Siphoviridae_sp._ctMgQ24*
dsDNA	Uroviricota	*Siphoviridae_sp._ctmHK1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctmIh35*
dsDNA	Uroviricota	*Siphoviridae_sp._ctMM521*
dsDNA	Uroviricota	*Siphoviridae_sp._ctmpG14*
dsDNA	Uroviricota	*Siphoviridae_sp._ctmqi22*
dsDNA	Uroviricota	*Siphoviridae_sp._ctMRT7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctMsr1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctmTU3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctmYS12*
dsDNA	Uroviricota	*Siphoviridae_sp._ctneY2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctngK14*
dsDNA	Uroviricota	*Siphoviridae_sp._ctNHj22*
dsDNA	Uroviricota	*Siphoviridae_sp._ctNLX12*
dsDNA	Uroviricota	*Siphoviridae_sp._ctnN38*
dsDNA	Uroviricota	*Siphoviridae_sp._ctnNB1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctNnX9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctnPP24*
dsDNA	Uroviricota	*Siphoviridae_sp._ctNwR4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctNyC15*
dsDNA	Uroviricota	*Siphoviridae_sp._ctNZc11*
dsDNA	Uroviricota	*Siphoviridae_sp._cto3L1*
dsDNA	Uroviricota	*Siphoviridae_sp._cto6l14*
dsDNA	Uroviricota	*Siphoviridae_sp._ctOba29*
dsDNA	Uroviricota	*Siphoviridae_sp._ctoD011*
dsDNA	Uroviricota	*Siphoviridae_sp._ctoiA13*
dsDNA	Uroviricota	*Siphoviridae_sp._ctoiW10*
dsDNA	Uroviricota	*Siphoviridae_sp._ctOkv13*
dsDNA	Uroviricota	*Siphoviridae_sp._ctomJ2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctoMP27*
dsDNA	Uroviricota	*Siphoviridae_sp._ctoNj20*
dsDNA	Uroviricota	*Siphoviridae_sp._ctoof1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctoRD1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctoWO12*
dsDNA	Uroviricota	*Siphoviridae_sp._ctP0x5*
dsDNA	Uroviricota	*Siphoviridae_sp._ctP6113*
dsDNA	Uroviricota	*Siphoviridae_sp._ctP6p7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctPAi1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctPB44*
dsDNA	Uroviricota	*Siphoviridae_sp._ctpbb7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctpbe1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctpCx1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctPEx3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctpnN3*

dsDNA	Uroviricota	*Siphoviridae_sp._ctPrm3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctPui28*
dsDNA	Uroviricota	*Siphoviridae_sp._ctPyh10*
dsDNA	Uroviricota	*Siphoviridae_sp._ctqA315*
dsDNA	Uroviricota	*Siphoviridae_sp._ctqBH20*
dsDNA	Uroviricota	*Siphoviridae_sp._ctqcj14*
dsDNA	Uroviricota	*Siphoviridae_sp._ctQHO9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctqK313*
dsDNA	Uroviricota	*Siphoviridae_sp._ctQLz13*
dsDNA	Uroviricota	*Siphoviridae_sp._ctqOv4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctqPo10*
dsDNA	Uroviricota	*Siphoviridae_sp._ctqpo8*
dsDNA	Uroviricota	*Siphoviridae_sp._ctQqU1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctqtA1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctr2f5*
dsDNA	Uroviricota	*Siphoviridae_sp._ctr4Z12*
dsDNA	Uroviricota	*Siphoviridae_sp._ctr8v12*
dsDNA	Uroviricota	*Siphoviridae_sp._ctRcp9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctrCv3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctREU2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctrG012*
dsDNA	Uroviricota	*Siphoviridae_sp._ctRg61*
dsDNA	Uroviricota	*Siphoviridae_sp._ctRg81*
dsDNA	Uroviricota	*Siphoviridae_sp._ctrgQ8*
dsDNA	Uroviricota	*Siphoviridae_sp._ctrgt10*
dsDNA	Uroviricota	*Siphoviridae_sp._ctRGt12*
dsDNA	Uroviricota	*Siphoviridae_sp._ctrKX6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctRlj31*
dsDNA	Uroviricota	*Siphoviridae_sp._ctrok7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctrpM6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctRuT6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctS0613*
dsDNA	Uroviricota	*Siphoviridae_sp._ctS1E53*
dsDNA	Uroviricota	*Siphoviridae_sp._ctS2049*
dsDNA	Uroviricota	*Siphoviridae_sp._ctSA812*
dsDNA	Uroviricota	*Siphoviridae_sp._ctsAY3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctsMn4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctsoB6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctSOv1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctSuy3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctSx228*
dsDNA	Uroviricota	*Siphoviridae_sp._ctsYb1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctTC45*
dsDNA	Uroviricota	*Siphoviridae_sp._cttDR14*
dsDNA	Uroviricota	*Siphoviridae_sp._cttG313*
dsDNA	Uroviricota	*Siphoviridae_sp._ctTgb17*
dsDNA	Uroviricota	*Siphoviridae_sp._ctTkm23*
dsDNA	Uroviricota	*Siphoviridae_sp._cttpk5*
dsDNA	Uroviricota	*Siphoviridae_sp._ctTqe2*

dsDNA	Uroviricota	*Siphoviridae_sp._ctTrb4*
dsDNA	Uroviricota	*Siphoviridae_sp._cttuu15*
dsDNA	Uroviricota	*Siphoviridae_sp._cttWj13*
dsDNA	Uroviricota	*Siphoviridae_sp._ctTXt1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctu3K14*
dsDNA	Uroviricota	*Siphoviridae_sp._ctu9a31*
dsDNA	Uroviricota	*Siphoviridae_sp._ctUir1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctUoe7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctuOq1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctur44*
dsDNA	Uroviricota	*Siphoviridae_sp._ctv2R2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctvbt38*
dsDNA	Uroviricota	*Siphoviridae_sp._ctvBz3*
dsDNA	Uroviricota	*Siphoviridae_sp._ctvf68*
dsDNA	Uroviricota	*Siphoviridae_sp._ctvhu9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctVif31*
dsDNA	Uroviricota	*Siphoviridae_sp._ctVJE9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctvNP11*
dsDNA	Uroviricota	*Siphoviridae_sp._ctvok7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctVOP12*
dsDNA	Uroviricota	*Siphoviridae_sp._ctVqj4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctvuW5*
dsDNA	Uroviricota	*Siphoviridae_sp._ctvxh7*
dsDNA	Uroviricota	*Siphoviridae_sp._ctW4q29*
dsDNA	Uroviricota	*Siphoviridae_sp._ctw757*
dsDNA	Uroviricota	*Siphoviridae_sp._ctWBz6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctWdm1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctWDo30*
dsDNA	Uroviricota	*Siphoviridae_sp._ctwfx1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctwHj1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctWKa2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctWlk2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctwQg18*
dsDNA	Uroviricota	*Siphoviridae_sp._ctwQT14*
dsDNA	Uroviricota	*Siphoviridae_sp._ctwWa4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctWWc42*
dsDNA	Uroviricota	*Siphoviridae_sp._ctwYi19*
dsDNA	Uroviricota	*Siphoviridae_sp._ctwzt2*
dsDNA	Uroviricota	*Siphoviridae_sp._ctx254*
dsDNA	Uroviricota	*Siphoviridae_sp._ctx7r16*
dsDNA	Uroviricota	*Siphoviridae_sp._ctX8T1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctXbO14*
dsDNA	Uroviricota	*Siphoviridae_sp._ctxfQ4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctxjx4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctXPh6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctXQ92*
dsDNA	Uroviricota	*Siphoviridae_sp._ctXQq5*
dsDNA	Uroviricota	*Siphoviridae_sp._ctxS04*
dsDNA	Uroviricota	*Siphoviridae_sp._ctxvK3*

dsDNA	Uroviricota	*Siphoviridae_sp._ctxyw6*
dsDNA	Uroviricota	*Siphoviridae_sp._ctXZQ9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctXZx16*
dsDNA	Uroviricota	*Siphoviridae_sp._ctYcY12*
dsDNA	Uroviricota	*Siphoviridae_sp._ctyHC15*
dsDNA	Uroviricota	*Siphoviridae_sp._ctYKh4*
dsDNA	Uroviricota	*Siphoviridae_sp._ctylc9*
dsDNA	Uroviricota	*Siphoviridae_sp._ctZ0X1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctZCK1*
dsDNA	Uroviricota	*Siphoviridae_sp._ctZD11*
dsDNA	Uroviricota	*Siphoviridae_sp._ctzEj35*
dsDNA	Uroviricota	*Siphoviridae_sp._ctzm5103*
dsDNA	Uroviricota	*Siphoviridae_sp._ctzMZ8*
dsDNA	Uroviricota	*Siphoviridae_sp._ctzrC10*
dsDNA	Uroviricota	*Siphoviridae_sp._ctzVd36*
dsDNA	Uroviricota	*Siphoviridae_sp._ctZZK17*
dsDNA	Uroviricota	*Siphovirus_Jomon_CT89*
dsDNA	Uroviricota	*Streptococcus_phage_Javan191*
dsDNA	Uroviricota	*Streptococcus_phage_SpSL1*
dsDNA	Uroviricota	*Streptomyces_phage_LibertyBell*
dsDNA	Uroviricota	*Lactococcus_lactis_phage_645*
dsDNA	Uroviricota	*Lactococcus_phage_6W06*
dsDNA	Uroviricota	*Lactococcus_phage_936_group_phage_Phi129*
dsDNA	Uroviricota	*Lactococcus_phage_936_group_phage_Phi4*
dsDNA	Uroviricota	*Lactococcus_phage_936_group_phage_Phi43*
dsDNA	Uroviricota	*Lactococcus_phage_936_group_phage_PhiA1127*
dsDNA	Uroviricota	*Lactococcus_phage_AV09*
dsDNA	Uroviricota	*Lactococcus_phage_GL7*
dsDNA	Uroviricota	*Lactococcus_phage_GP13*
dsDNA	Uroviricota	*Lactococcus_virus_05601*
dsDNA	Uroviricota	*Lactococcus_virus_16802*
dsDNA	Uroviricota	*Lactococcus_virus_30804*
dsDNA	Uroviricota	*Lactococcus_virus_340*
dsDNA	Uroviricota	*Lactococcus_virus_37201*
dsDNA	Uroviricota	*Lactococcus_virus_38507*
dsDNA	Uroviricota	*Lactococcus_virus_51701*
dsDNA	Uroviricota	*Lactococcus_virus_57001*
dsDNA	Uroviricota	*Lactococcus_virus_62605*
dsDNA	Uroviricota	*Lactococcus_virus_63302*
dsDNA	Uroviricota	*Lactococcus_virus_66901*
dsDNA	Uroviricota	*Lactococcus_virus_712*
dsDNA	Uroviricota	*Lactococcus_virus_79201*
dsDNA	Uroviricota	*Lactococcus_virus_88605*
dsDNA	Uroviricota	*Lactococcus_virus_96401*
dsDNA	Uroviricota	*Lactococcus_virus_96403*
dsDNA	Uroviricota	*Lactococcus_virus_ASCC191*
dsDNA	Uroviricota	*Lactococcus_virus_CaseusJM1*
dsDNA	Uroviricota	*Lactococcus_virus_CB13*
dsDNA	Uroviricota	*Lactococcus_virus_CHPC129*

dsDNA	Uroviricota	*Lactococcus_virus_CHPC362*
dsDNA	Uroviricota	*Lactococcus_virus_CHPC52*
dsDNA	Uroviricota	*Lactococcus_virus_CHPC781*
dsDNA	Uroviricota	*Lactococcus_virus_CHPC958*
dsDNA	Uroviricota	*Lactococcus_virus_CHPC959*
dsDNA	Uroviricota	*Lactococcus_virus_CHPC965*
dsDNA	Uroviricota	*Lactococcus_virus_fd13*
dsDNA	Uroviricota	*Lactococcus_virus_i0139*
dsDNA	Uroviricota	*Lactococcus_virus_jm2*
dsDNA	Uroviricota	*Lactococcus_virus_jm3*
dsDNA	Uroviricota	*Lactococcus_virus_LP8511*
dsDNA	Uroviricota	*Lactococcus_virus_MV10L*
dsDNA	Uroviricota	*Lactococcus_virus_P008*
dsDNA	Uroviricota	*Lactococcus_virus_P1532*
dsDNA	Uroviricota	*Lactococcus_virus_P656*
dsDNA	Uroviricota	*Lactococcus_virus_Phi155*
dsDNA	Uroviricota	*Lactococcus_virus_phi7*
dsDNA	Uroviricota	*Lactococcus_virus_PhiA16*
dsDNA	Uroviricota	*Lactococcus_virus_PhiF0139*
dsDNA	Uroviricota	*Lactococcus_virus_PhiJF1*
dsDNA	Uroviricota	*Lactococcus_virus_PhiLj*
dsDNA	Uroviricota	*Lactococcus_virus_PhiS0139*
dsDNA	Uroviricota	*Lactococcus_virus_R31*
dsDNA	Uroviricota	*Lactococcus_virus_sk1*
dsDNA	Uroviricota	*Lactococcus_virus_Sl4*
dsDNA	Uroviricota	*Skunavirus__uc*
dsDNA	Uroviricota	*Leuconostoc_virus_1A4*
dsDNA	Uroviricota	*Leuconostoc_virus_Ln9*
dsDNA	Uroviricota	*Unaquatrovirus__uc*
dsDNA	Uroviricota	*Rhodococcus_virus_Weasel*
DsDNA-RT	Artverviricota	*Aglaonema_bacilliform_virus*
DsDNA-RT	Artverviricota	*Badnavirus__uc*
DsDNA-RT	Artverviricota	*Cacao_swollen_shoot_CE_virus*
DsDNA-RT	Artverviricota	*Rice_tungro_bacilliform_virus*
SsRNA-RT	Artverviricota	*Avian_endogenous_retrovirus_EAV-HP*
SsRNA-RT	Artverviricota	*Retroviridae__uc*
dsRNA	Pisuviricota	*Picobirnaviridae_sp.*
dsRNA	Pisuviricota	*Picobirnaviridae_sp._ct6rZ1*
dsRNA	Pisuviricota	*Picobirnavirus_PREDICT_PbV-104*
dsRNA	Pisuviricota	*Picobirnavirus_PREDICT_PbV-109*
dsRNA	Pisuviricota	*Simian_picobirnavirus*
dsRNA	Pisuviricota	*Picobirnaviridae__uc*
dsRNA	Pisuviricota	*Human_picobirnavirus*
dsRNA	Pisuviricota	*Otarine_picobirnavirus*
dsRNA	Pisuviricota	*Picobirnavirus__uc*
dsRNA	Pisuviricota	*Porcine_picobirnavirus*
dsRNA	Pisuviricota	*Sapelovirus_sp.*
dsRNA	Pisuviricota	*Teschovirus_sp.*
dsRNA	Pisuviricota	*Artemisia_virus_A*

dsRNA	Pisuviricota	*Soybean_yellow_common_mosaic_virus*
ssDNA	Cressdnaviricota	*CRESS_virus_sp._ctBnw2*
ssDNA	Cressdnaviricota	*Cressdnaviricota_sp.*
ssDNA	Hofneiviricota	*Inoviridae_sp._ct1ro12*
ssDNA	Hofneiviricota	*Inoviridae_sp._ctaKY11*
ssDNA	Hofneiviricota	*Inoviridae_sp._ctDEu7*
ssDNA	Hofneiviricota	*Inoviridae__uc*
ssDNA	Hofneiviricota	*Inovirus_sp.*
ssDNA	Phixviricota	*Apis_mellifera_associated_microvirus_8*
ssDNA	Phixviricota	*Capybara_microvirus_Cap1_SP_162*
ssDNA	Phixviricota	*Human_gut_microviridae_SH-CHD12*
ssDNA	Phixviricota	*Microviridae_sp._ct13s5*
ssDNA	Phixviricota	*Microviridae_sp._ct2x44*
ssDNA	Phixviricota	*Microviridae_sp._ct4S516*
ssDNA	Phixviricota	*Microviridae_sp._ct9ZF1*
ssDNA	Phixviricota	*Microviridae_sp._ctbyf5*
ssDNA	Phixviricota	*Microviridae_sp._ctCVC7*
ssDNA	Phixviricota	*Microviridae_sp._ctemt10*
ssDNA	Phixviricota	*Microviridae_sp._ctFGM2*
ssDNA	Phixviricota	*Microviridae_sp._ctHN216*
ssDNA	Phixviricota	*Microviridae_sp._ctJby12*
ssDNA	Phixviricota	*Microviridae_sp._ctJKB8*
ssDNA	Phixviricota	*Microviridae_sp._ctKRd3*
ssDNA	Phixviricota	*Microviridae_sp._ctNWS1*
ssDNA	Phixviricota	*Microviridae_sp._ctUND6*
ssDNA	Phixviricota	*Microviridae_sp._ctuZ46*
ssDNA	Phixviricota	*Microviridae_sp._ctwzP10*
ssDNA	Phixviricota	*Microviridae_sp._ctX0F7*
ssDNA	Phixviricota	*Microviridae_sp._ctX401*
ssDNA	Phixviricota	*Microvirus_mar40*
ssDNA	Phixviricota	*Microvirus_mar45*
ssDNA	Phixviricota	*Microvirus_mar49*
ssDNA	Phixviricota	*Microvirus_mar52*
ssDNA	Phixviricota	*Microvirus_mar54*
ssDNA	Phixviricota	*Microvirus_mar55*
ssDNA	Phixviricota	*Microvirus_mar56*
ssDNA	Phixviricota	*Microvirus_mar59*
ssDNA	Phixviricota	*Poophage_MBI-2016a*
ssDNA	Phixviricota	*Poophage_SC_4_H6H8_2017*
ssDNA	Phixviricota	*Tortoise_microvirus_13*
ssDNA	Phixviricota	*Tortoise_microvirus_88*
ssDNA	Preplasmiviricota	*Corticoviridae_sp.*
ssRNA	Kitrinoviricota	*Cucumber_green_mottle_mosaic_virus*
ssRNA	Kitrinoviricota	*Tobacco_mild_green_mosaic_virus*
ssRNA	Kitrinoviricota	*Tropical_soda_apple_mosaic_virus*
ssRNA	Kitrinoviricota	*Lettuce_virus_X*
ssRNA	Kitrinoviricota	*Carlavirus__uc*
ssRNA	Kitrinoviricota	*Pegivirus__uc*
ssRNA	Negarnaviricota	*Human_orthopneumovirus*

Unknown	Viruses__na	*Human_anellovirus_zj-ad_1*
Unknown	Viruses__na	*DP_phage*
Unknown	Viruses__na	*Escherichia_phage_3W*
Unknown	Viruses__na	*Hudisavirus_sp.*
Unknown	Viruses__na	*McMurdo_Ice_Shelf_pond-associated_circular_DNA_virus-3*
Unknown	Viruses__na	*Phage_sp._ctcqm2*
Unknown	Viruses__na	*Phage_sp._ctez94*
Unknown	Viruses__na	*Phage_sp._ctfRs3*
Unknown	Viruses__na	*Phage_sp._ctIHi3*
Unknown	Viruses__na	*Phage_sp._ctqZP6*
Unknown	Viruses__na	*Phage_sp._ctrsQ3*
Unknown	Viruses__na	*Phage_sp._ctXnn1*
Unknown	Viruses__na	*Phage_sp._cty4N14*
Unknown	Viruses__na	*Reptile-associated_circular_DNA_molecule*
Unknown	Viruses__na	*virus_sp._ct3kA5*
Unknown	Viruses__na	*virus_sp._ct5rm7*
Unknown	Viruses__na	*virus_sp._ct6Ax4*
Unknown	Viruses__na	*virus_sp._ct6zc1*
Unknown	Viruses__na	*virus_sp._ct9pU4*
Unknown	Viruses__na	*virus_sp._ctCsQ3*
Unknown	Viruses__na	*virus_sp._ctd0M1*
Unknown	Viruses__na	*virus_sp._ctE0n6*
Unknown	Viruses__na	*virus_sp._ctEfN2*
Unknown	Viruses__na	*virus_sp._ctFlR8*
Unknown	Viruses__na	*virus_sp._ctHG14*
Unknown	Viruses__na	*virus_sp._ctJLD79*
Unknown	Viruses__na	*virus_sp._ctLpa4*
Unknown	Viruses__na	*virus_sp._ctLTC15*
Unknown	Viruses__na	*virus_sp._ctmTa7*
Unknown	Viruses__na	*virus_sp._ctpeS3*
Unknown	Viruses__na	*virus_sp._ctPLL24*
Unknown	Viruses__na	*virus_sp._ctqq75*
Unknown	Viruses__na	*virus_sp._ctr1v16*
Unknown	Viruses__na	*virus_sp._ctReX5*
Unknown	Viruses__na	*virus_sp._ctRTq15*
Unknown	Viruses__na	*virus_sp._ctuWX8*
Unknown	Viruses__na	*virus_sp._ctv2g1*
Unknown	Viruses__na	*virus_sp._ctvdG25*
Unknown	Viruses__na	*virus_sp._ctviY17*
Unknown	Viruses__na	*virus_sp._ctWpE22*
Unknown	Viruses__na	*Virus_sp._ctWxR2*
Unknown	Viruses__na	*virus_sp._ctX0L1*
Unknown	Viruses__na	*virus_sp._ctyg714*
Unknown	Viruses__na	*virus_sp._ctzmw3*