Seqüenciació massiva aplicada a l'epidemiologia d'aigües residuals i a la caracterització de viromes

Sandra Martínez Puchol

Seqüenciació massiva aplicada a l'epidemiologia d'aigües residuals i a la caracterització de viromes

Sandra Martínez Puchol

Tesi Doctoral

Desembre 2020

UNIVERSITAT DE BARCELONA

Programa de Doctorat de Biotecnologia

Departament de Genètica, Microbiologia i Estadística
Secció de Microbiologia, Virologia i Biotecnologia
Facultat de Biologia

Seqüenciació massiva aplicada a l'epidemiologia d'aigües residuals i a la caracterització de viromes

Memòria presentada per Sandra Martínez Puchol per a optar al grau de Doctora per la Universitat de Barcelona.

Directora i tutora

Dra. Sílvia Bofill Mas

Directora

Prof. Rosina Girones Llop

Doctoranda

Sandra Martínez Puchol

Barcelona, 2020

Als meus avis.

"so what of this release?
some life feels good now dont it ?
dont have to have a leaving plan
nothings gonna ease your mind
well its all fine and were all fine any way"

Bon Iver – RABi

Aquesta tesi ha estat finançada amb una beca *Ayudas para contratos predoctorales para la formación de doctores* (FPI 2015) del *Ministerio de Economía y Competitividad*.

La imatge de la portada és obra de la il·lustradora Lynn Scurfield per a la revista *The Scientist* (2016).

Totes les figures de la present tesi han estat creades amb Biorender.com.

AGRAÏMENTS

Hola, benvinguda a l'única part de la tesi que tothom llegeix. Agafa un cafè i seu, que després d'aquest temps hi ha molta gent a la que agrair...

Primer de tot, gràcies a les meves directores per donar-me la oportunitat de fer la tesi al grup i així continuar en el món de la recerca quan pensava que no podria. Gràcies **Sílvia** per la teva proximitat, confiança i implicació i perquè, encara que hem viscut alguns mals moments en aquests darrers 5 anys, hem sabut treure el millor d'ells i seguir endavant amb més fortalesa. Gràcies **Rosina** per creure en mi i ensenyar-me a veure el reptes científics des d'altres perspectives. He après molt, de virologia i de mi mateixa, al vostre costat en aquests darrers anys.

He d'agrair seguidament a la persona que em va ensenyar el que volia dir "fer recerca". Allà pel gener del 2012, jo només buscava un lloc on fer les pràctiques de màster i matar ràpid el tràmit, gràcies **Quim** per acollir-me al grup de les enteros quan ningú veia més enllà de la meva nota, per contagiar-me la teva passió per la ciència, pels bacteris i antibiòtics i per tot el que em vas ensenyar. Espero que en un futur els nostres camins professionals es tornin a creuar.

Continuant amb el que he guanyat aquests darrers anys he de parlar de l'**Eva**. De tu he après a "ser peor persona para ser mejor persona", a que la birra post-lab és obligatòria, a relativitzar preocupacions que realment no ho havien de ser... i aixina moltíssimes coses més. Gràcies per tots aquests anys de companyia i amistat de la bona, qui ho havia de dir quan estàvem destinades a ser enemigues (guiño guiño). Ah! i gràcies per guardar el secret fins avui del dia que vam tirar tots els estàndards (dos cops!) d'adenos i MS2 pel terra quan no portàvem ni dues setmanes al lab. Destinades a guanyar palitos des del primer minut, però ja saps el que diuen, sempre "go with the flow..." i arribaràs on et proposis.

Martus! Recordo quan vaig arribar al departament i tu estaves amb la teva boleta Anna a casa. Tothom parlava meravelles de tu i jo en aquell moment no em podia imaginar quanta raó tenien. Gràcies per ensenyar-me i aportar-me taaaaantes coses a nivell professional i personal, ets l'amiga i companya que tothom desitjaria tenir, sobre tot pels rampells de neteja i organització que compartim! :P Recordes aquell dia màgic a les pràctiques de viro? Espero que vinguin molts

més moment iguals o millors (se el que penses, shhhhh encara nooo) i ja saps, sempre juntes en la vida, en la ciència i en la decoració d'interiors jaja.

Bueno **Marta** ("petita"), qui ho havia de dir que la noia de "lo sud" a la que vam entrevistar per al TFG a la cuina, mentre feia una nested, m'arribaria a aportar tant en aquest temps. Gràcies per fer-me la vida tan fàcil al dirigir-te els treballs de grau i màster amb la teva implicació i rigor. Des de que et vaig conèixer sabia que estaves destinada a la recerca, me'n alegro que el camí que vas seguir et portés a on ets ara i et recuperés com a companya incansable de llibreries de seqüenciació, concentracions baixes d'ADN i de frikades de virus. Gaudeix del que queda d'aventura, tant tu com l'Eva em tindreu, com sempre, per celebrar les pujades i per a empènyer cap amunt en els moments complicats del doctorat.

Xavi, recordo com a la teva tesi em deies "no te librarás de mi tan fácilmente", i menys mal que ha estat així. Gràcies per tots els momentazos que m'has regalat durant el temps que vam compartir al departament i pels de després. Gràcies per sempre fer-li cas a tu trans consumista favorita, per les llargues converses frikejant sobre la metagenòmica i sus cosas, pels consells i per sempre estar. Espero que seguim compartint moments científics, de riures i vídeos nous de La Leti durant molts més anys.

Gràcies a tots els altres companys que formen o han format part del grup per aquests anys de complicitat. **Ayalke**, gracias chikito por siempre tener una sonrisa para todos y contagiar positividad allà donde vas. Gràcies **David** per les llargues hores de converses i berenars que van endolcir el temps que vam compartir. Gràcies **Laurus** per compartir la teva energia i motivació per a fer-nos arribar al que ens proposem. Obrigada **Marcelle** por dejarme entrar en el maravilloso mundo de los rotavirus, aderezando siempre nuestras horas de trabajo con conversaciones musicales. Gracias **Eloy** por las clases de "estadística para dummies" i también por las peleas por el threshold que le daban salsa al asunto. Gràcies **Tecles** i **Pep** per ensenyar-me que les pantalles de comandaments no fan tanta por com sembla i per ajudar-me quan no sabia ni el que era un fastaq.

Parlar de l'ànima del departament de Micro és parlar d'ells. Gracias **Hulios** por ser los mejores bailarines, jugadores de dixit, liantes profesionales y AMIGOS que pueden existir. Haberos conocido ha hecho que todos estos años valgan aún más la pena (sobre todo en época de biofesta jiji). Es imposible enumerar todos los

momentazos vividos dentro y fuera del departamento, y necesitaría demasiadas paginas para agradeceros todo lo que habéis hecho por mi, así que voy a ejemplificarlo en formato abstract. **Pedro**, gracias por el apoyo. **Manu**, gracias por las charlas. **Genoveva**, gracias por la fuerza. **Sara**, gràcies per la inocència. **Ale**, gracias por la alegría. **Gala**, gràcies per la bondat. **Eva**, tú ya tienes un párrafo entero. **Cris**, gracias por la vitalidad. **Oscar**, gracias por la implicación. **Sergio**, gracias por la calma. Siempre inclusivos, siempre intolerantes, siempre nosotros.

De tots és sabut que la Fase 2 del departament (viva la fase *3*!) és la millor de tot l'edifici, gràcies en part a tots el companys dels labs 9-10-12. Gràcies **Míriam**, **Julia, Pablo**, **Dani**, **Pedro**, **Judith**, **Laura**, **Anna**, **Lorena**, **Maite**, **Eli** (segur em deixo a molta gent) per la companyia, per les grans converses de passadís i per l'ajut durant tots aquest anys. Especialment agrair a l'**Aiora** i a la **Rachel** per sempre estar quan se les necessita (i no només parlo de ciència) i per la seva alegria que contagien arreu... sou les millors! També donar les gràcies al senyor **Robert** per la seva constant positivitat i amor al prójimo (jajaja), per les xerrades transcendentals amb birra a la mà a farmàcia i per sempre intentar lluitar pel bé comú.

Gràcies també **Alberto**, **Yexe**, **Rosa** i **Arantxa** pels grans moments passats al lab 8, i especialment a la **Karmins**, per sempre estar disposada a ajudar i a escoltar, i a l'**Edu,** per les estones de confidències a l'*speaking corner* i els "shows". Gràcies a la resta de "microlloros", amb menció honorífica a l'**Aurora** i la **Montse** (se us troba a faltar!), pels bons moments des de fent cafès fins a findes rurals o casaments <3. També a la resta de membres entèrics, **Gemma**, **Adan**, **Mario** i **Albert** pels congressos, mostrejos i fins i tot partits de pàdel compartits.

Han estat moltes estones de menjador (quan no existia la covid i erem com 20 persones apretujades en una sala, quins records eh?), de pizzes compartides pre-vacances, de sopars de nadal, de picapica de departament/secció, de Arena, de calçotades... gràcies **Quim**, **Elena**, **Pol**, **Maite**, **Sara**, **Vero** i un llarg etcètera per haver format part d'alguns d'aquests grans moments. Gràcies també a tot el equipazo de secretaria per sempre estar al peu del canó, **Rosario**, **Susana**, **Mónica**, **Manolo** i especialment **Bea**, per la paciència i l'ajut que sempre m'heu/ens heu donat.

No em puc deixar d'anomenar a algunes de les persones amb les que, abans d'entrar al departament, vaig passar moltíssimes hores de laboratori, de congressos a l'altra punta del món, de mostrejos a 4000m d'alçada i de molta felicitat i aprenentatge. Gràcies **Cláudia**, **Lidia**, **Noe** i **Mari**, érem un gran equip, chipis! També gràcies a totes les persones de l'antic CRESIB amb les que vaig gaudir d'infinitat de bons moments, especialment a la **Núria** i a l'**Ariel**. Gràcies també a tots vosaltres he arribat on estic ara.

Agradecer cómo no a mis snobs favoritas, al team miuec (¿qué queda ya de eso? jaja) que tanto me ha dado estos últimos años. Gracias **Maider**, **Bea**, **Salo**, **Estefi**, **Elena**, **Rebe**, **Rosa**, **Estela**, **Paula**, **Ale**, **Laura** y **Míriam** por tantísimas horas de risas a distancia, apoyo incondicional y conciertos variaditos que espero podamos retomar PRONTO. Un poquito de esta tesis también es vuestra. **Maider**, a ti gracias especialmente por ser una amiga como pocas, siempre dispuesta a dar todo lo que tienes sin esperar nada a cambio, ya sabes todo así que solo decirte: asko maite zaitut. També agrair a la **Mercè** els grans moments viscuts, molts d'ells bastant surrealistes i "increíbles", i la seva generositat.

Estem arribant al final i com no parlar d'un dels pilars més important de la meva vida, els meus biòlegs preferits. Gracias **Cris, Clara**, **Luís** y **Laura** por todos estos años de amistad incansable. Quién diría que esas tardes saltándonos las clases de mates (perdó) en el hall, en nuestro sitio, llegarían tan lejos. Hemos vivido grandes momentos juntos y muchos cambios, pero lo que nos unió ahí sigue. Cómo no agradecer aquí también a **Patxi** y **Carlos**, ahora compañeros de departamento, porque sus clases nos unieron y a mi me hicieron interesarme por eso de la Micro.

I arriba el moment de parlar de la família, així que primer li toca a la portuguesa. Obrigada **Inês** y **Francisco** per tudo. Vosotros, que sabéis más que nadie lo que significa llegar a este punto, habéis sido un apoyo fundamental para poder lograrlo. Gracias por acompañarme en este viaje y, aunque nos separen algunos kilómetros, en los que quedan por venir. Love u!

Gràcies als meus avis, tiets i cosins (els que estan i els que ja no) pel suport i confiança rebuda. Gracias también a mis suegros, **Mari** i **Dionisio**, por toda la ayuda que siempre me habéis brindado y sobre todo por el amor.

Gràcies als meus pares, **Pau** i **Maribel**, i a la meva germana, **Anabel**, per ajudar-me a ser el que sóc, cuidar-me i donar-m'ho tot, sense vosaltres no hagués aconseguit. I encara que sóc una rancia i no us ho dic mai, us estimo moltíssim.

I per acabar parlar de l'**Álvaro**. Gracias por nunca dejar de confiar en mi cuando ni yo lo hacía, por darme energía, risas, cariño, consejos, clases de historia, datos futbolísticos muy interesantes (ejem), música, mil conciertos, viajes inolvidables… la lista es infinita. En estos 14 años hemos vividos muchas etapas juntos, desde acompañarme en mi primer día de carrera, celebrar tu mudanza a Barna, casarnos y hasta buscar hipotecas… Heavy, eh? Gracias también por todo tu apoyo estos últimos años y por aguantar estoicamente mis ralladas científicas, que no han sido pocas. Te quiero infinito, *I wish I was the verb 'to trust' and never let you down…*

Desembre 2020

PD?: Per a fer-vos més partícips d'una part molt important de mi, he creat una breu llista de reproducció que defineix prou bé el que ha estat aquesta etapa que ara acaba. Enjoy! :)

https://open.spotify.com/playlist/12HoUiMHvFYmfelUAKUmpf?si=msFqs6CMQcew4L22BhgI9w

Sinopsi

Next generation sequencing (NGS) techniques have emerged in the last decade as keystone for the thorough study of microorganisms in a wide variety of samples and settings, replacing traditional molecular methods. In the field of virology, the constant evolution of sequencing platforms and applications enabled the improvement of virome studies. The main limitation when analysing the virome from any type of sample is the low proportion of viral sequences identified compared with the total number of sequences, especially critical for human viruses.

In this work we aimed to evaluate the use of different sequencing approaches, target enrichment (TES) and amplicon deep sequencing (ADS), for the characterization of the virome and specific viral pathogens in sewage and as tools for efficient wastewater-based epidemiology in outbreak scenario. The application of TES has proved to be a very successful strategy for the study of vertebrate viruses in sewage samples providing a higher number of detected families, a higher number of members within these families, more reads and larger genome coverage than conventional untargeted viral metagenomics. Additionally, allowed the obtention of SARS-CoV-2 sequences as part of sewage virome in a COVID-19 pandemic context, retrieving also other relevant human and animal coronavirus sequences, shedding light on the co-circulation of different strains in a determined population. In contrast, ADS proved to a very sensitive technique for the description of the diversity within a viral family, enabling the subtyping of sequences belonging to Enterovirus A71 C1 in sewage, while an encephalitis outbreak caused by this strain was happening during sampling period.

NGS, with and without TES panels, was also evaluated for the study of viral etiological agents of acute gastroenteritis in a collection of samples tested negative for the commonly associated pathogens. Its application resulted in the detection of emergent viral variants, like Norovirus GIV, and viruses not traditionally tested, like sapoviruses and astroviruses. These results highlighted the need of the incorporation of these viruses in clinical testing and the potential use of viral metagenomics as a diagnostic tool.

Lastly, to evaluate the use of enrichment panel in animal virology, TES was applied for the virome study of two economically important fish species from the Portuguese Atlantic coast. Pathogens causing viral nervous necrosis and infectious pancreatic necrosis in fishes were detected, demonstrating the utility of NGS techniques for the study of infections that may cause an economic impact in fish industry. Also, the identification of human noroviruses sequences in one of the fish samples suggested that fish virome studies can be used for evaluating potential threats regarding food safety.

ABREVIACIONS

ADN	Àcid desoxiribonucleic
ADNdc	ADN bicatenari
ADNmc	ADN monocatenari
ADS	Seqüenciació massiva d'amplicons (de l'anglès *Amplicon Deep Sequencing*)
AdV	Adenovirus
AiV	Aichivirus
ARN	Àcid ribonucleic
ARNdc	ARN bicatenari
ARNmc	ARN monocatenari (de l'anglès *single strand RNA*)
AstV	Astrovirus
BKPyV	Poliomavirus BK
bp	Parell de bases (de l'anglès *base pair*)
cDNA	ADN de cadena complementària
CRISPR	Repeticions palindròmiques curtes agrupades i regularment espaiades (de l'anglès *Clustered Regularly Interspaced Short Palindromic Repeats*)
EDAR	Estació Depuradora d'Aigües Residuals
EV	Enterovirus

Gb	Giga-bases
GC	Còpies genòmiques
HAdV	Adenovirus humà
HAV	Virus de l'hepatitis A
HCoV	Coronavirus humà
HEV	Virus de l'hepatitis E
HPV	Virus del papil·loma humà
HPyV	Poliomavirus humà
HTS	Seqüenciació d'alt rendiment (de l'anglès *High Troughtput Sequencing*)
IPNV	Virus de la necrosi hepàtica
JCPyV	Poliomavirus JC
Kb	Quilo-bases
KIPyV	Poliomavirus KI
MAstV	Mamastrovirus
MCPyV	Poliomavirus de les cèl·lules de Merkel
MDA	Amplificació de desplaçament múltiple (de l'anglès *Multiple Displacement Amplification*)
ml	Mil·lilitre
NGS	Seqüenciació massiva (de *l'anglès Next-Generation Sequencing*)

NoV	Norovirus
nPCR	PCR niuada (o *nested* PCR)
PCR	Reacció en cadena de la polimerasa (de l'anglès *Polymerase Chain Reaction*)
PeV	Parecovirus
qPCR	PCR quantitativa
RVA	Rotavirus A
SaV	Sapovirus
SISPA	Amplificació independent de seqüència (de l'anglès *Sequence-Independent single-primer amplification*)
SNV	Polimorfismes d'un únic nucleòtid (de l'anglès *Single Nucleotide Variant*)
TES	Seqüenciació amb enriquiment de dianes (de l'anglès *target enrichment sequencing*)
TTV	Torque teno virus
UNIX	Sistema operatiu portàtil, multitasca i multiusuari (de l'anglès *Uniplexed Information and Computing Service*)
VNNV	Virus de la necrosi nerviosa
WBE	Epidemiologia d'aigües residuals (de l'anglès *Wastewater Based Epidemiology*)
WUPyV	Poliomavirus WU

GLOSSARI

Basecalling	Procés d'assignació de bases posterior a la seqüenciació
Contig	Fragment d'ADN obtingut a partir de la superposició d'un grup de reads
Coverage	Cobertura de genoma, percentatge del genoma complert recuperat com a resultat de la seqüenciació massiva
Ensamblat	Alineament i unió de reads solapats per a la construcció de contigs
Índex	Seqüència d'ADN curta que s'incorpora en la producció de les llibreries de seqüenciació per etiquetar (indexar) mostres que es seqüenciaran alhora
Llibreria	Fragments d'ADN d'una mostra a seqüenciar, de una determinada mida, amb índex i adaptadors de seqüenciació en els extrems
Mapat	Comparació de reads amb un genoma de referència
Read	Conjunt de seqüències resultant després del procés de seqüenciació
Scaffold	Agrupació consecutiva de contigs en seqüències d'una llargada encara major

INTRODUCCIÓ

1. Els virus excretats pels humans

Des de inicis del segle XXI, està ben acceptat que els éssers humans som amfitrions d'un ecosistema microscòpic complex format per una àmplia comunitat de virus, bacteris, arqueobacteris, fongs i altres eucariotes, que coneixem de manera conjunta com a microbiota. El component viral que forma part d'aquest microbioma és el viroma (García-López et al., 2019; Lecuit i Eloit, 2013).

La metagenòmica vírica és un camp d'investigació que estudia la col·lecció completa de virus que formen part de la microbiota en qualsevol nínxol donat. La ràpida evolució d'aquest camp de la virologia en els darrers anys ha estat possible gràcies a l'adveniment de tècniques de seqüenciació massiva, eines principals d'estudi de la present tesi doctoral i de les que parlarem més endavant. A la figura 1 es presenten les principals famílies de virus coneguts fins el moment i que formen el viroma humà.

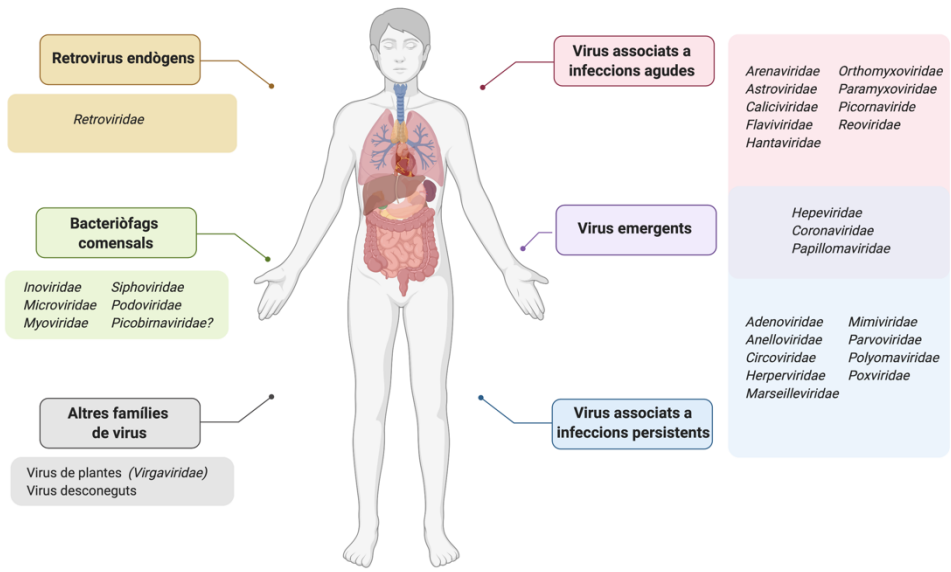


Figura 1. *Famílies de virus que formen el viroma humà. Adaptat de* Popgeorgiev et al., 2013.

Com a membres del viroma de diversos teixits o òrgans del cos humà, trobem principalment virus associats a infeccions persistents que en molts casos no estan associats o es desconeix el seu paper en el desenvolupament de malalties. Per una altra banda, es possible detectar virus que estiguin causant una infecció aguda activa i també retrovirus endògens. L'ús generalitzat de tècniques d'estudi de viromes ha permès també la detecció de virus emergents que, en molts casos, estan relacionats amb malalties d'alta rellevància sanitària. Un dels grups de virus més abundant en tots els ecosistemes son els bacteriòfags (virus que infecten bacteris), és per aquest motiu que són un grup molt nombrós de virus del microbioma humà, especialment a nivell de l'intestí (Popgeorgiev et al., 2013; Wylie et al.,2017).

A part de la detecció de virus en teixits associats a malaltia, moltes de les famílies de virus que colonitzen o infecten el tracte digestiu, urinari o epitelial poden ser excretats. Aquest fet ens permet estudiar de manera unificada, i a nivell de població, els membres d'aquestes famílies que tenen rellevància en el desenvolupament de malalties i els que s'excreten persistentment i seguir la seva presència o concentració gràcies a l'epidemiologia d'aigües residuals, un dels focus d'estudi d'aquesta tesi. L'estudi de virus excretats també genera informació important de potencials fonts de contaminació d'aigües i aliments (Rusiñol i Girones, 2017).

1.1. Epidemiologia basada en aigües residuals

Des de fa dècades, l'anàlisi de les aigües residuals s'ha utilitzat pel monitoreig de compostos químics d'origen humà com ara drogues, medicaments, contaminants exògens o nutrients. El principal focus d'anàlisis ha sigut la determinació de l'eficiència d'eliminació d'aquests compostos per les plantes de tractament d'aigües i la avaluació dels efluents d'aquestes plantes com a possibles fonts de contaminació (Choi et al., 2018). L'epidemiologia basada en aigües residuals també s'ha postulat

com a eina de control de la presència de patògens i d'indicadors de salut poblacional, entre d'altres (Fig. 2) (Lorenzo i Picó, 2019; Pina et al., 1998).

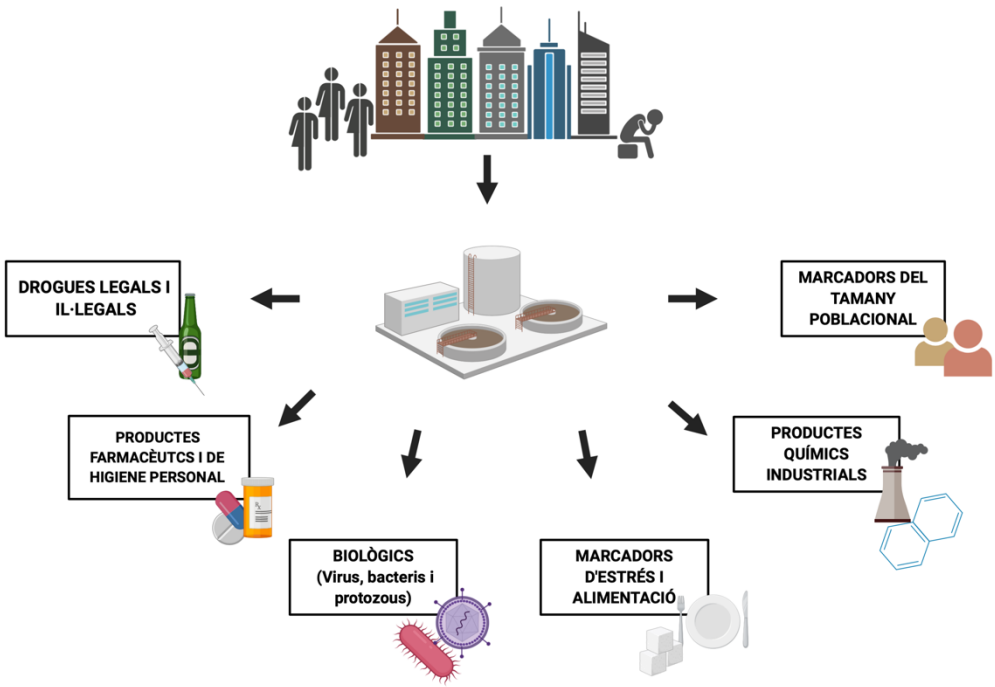


Figura 2. *Aplicacions de l'epidemiologia basada en aigües residuals. Adaptada de* (Choi et al., 2018).

L'anàlisi dels virus que circulen en una població i són excretats a la femta i l'orina es pot realitzar amb la seva detecció en aigües residuals urbanes. Gràcies al monitoreig d'aquestes, s'ha obtingut informació rellevant sobre quins virus es troben presents de manera persistent, podent actuar d'indicadors de contaminació fecal (Bofill-Mas et al., 2000; Jebri et al., 2017; Kongprajug et al., 2019; Pina et al., 1998). També ha permès estudiar quins patògens circulen a la població en moments determinats, caracteritzant els diferents genotips, la seva estacionalitat i diferències geogràfiques (Bisseux et al., 2018; Bofill-Mas et al., 2000; Thongprachum et al., 2018).

La vigilància ambiental s'ha implementat des d'organismes com la Organització Mundial de la Salut, com per exemple en la Iniciativa per l'Eradicació Mundial de la Poliomielitis, on es realitza un monitoreig de la presència del virus de la poliomielitis en regions en vies d'eradicació de la malaltia (Nakamura et al., 2015; Tiwari i Dhole, 2018). En aquest mateix context, també des del Instituto de Salud Carlos III anualment s'utilitza la detecció dels enterovirus circulants en aigües residuals com a complement del monitoreig dels casos clínics de paràlisi flàccida aguda (Instituto de Salud Carlos III, 2018).

L'epidemiologia d'aigües residuals ha generat un important impacte en la detecció de virus emergents, detectant-se per exemple la circulació del virus de l'hepatitis E en països industrialitzats (Clemente-Casares et al., 2003), o reflectint l'efecte de la vacunació en la circulació de soques de virus com els de l'hepatitis A o els rotavirus (Matthijnssens et al., 2012; Pintó et al., 2007).

Finalment, també ha demostrat ser una eina eficient per al seguiment de brots causats per un ampli ventall de virus (Guerrero-Latorre et al., 2011; Hellmér et al., 2014), aplicació que ha agafat especial rellevància aquest any 2020 arrel de la pandèmia causada pel SARS-CoV-2. Existeixen una gran quantitat de publicacions que recolzen la utilitat de l'epidemiologia basada en aigües residuals per al seguiment de l'evolució de la malaltia com a suport per a la implementació d'accions de control (Medema et al., 2020; Randazzo et al., 2020), el que ha fet que un gran nombre d'administracions inverteixin en aquests estudis i utilitzin els resultats obtinguts com una eina més per a la presa de decisions. A Catalunya, des del mes de juliol de 2020, s'està realitzant el seguiment de 56 plantes de tractament d'aigües residuals repartides territorialment, per a representar el 80% de la població, analitzant setmanalment l'evolució de la concentració del virus (https://sarsaigua.icra.cat) (Corominas et al., 2020). Aquest tipus de seguiment es pot utilitzar per a conèixer l'estat del

brot (o pandèmia), actuant com a potencial eina de detecció precoç en casos en els que la malaltia no s'estigui seguint a nivell clínic. També és útil per seguir el funcionament de les estratègies de mitigació implantades.

Per a realitzar un bon seguiment dels virus que circulen per la població és indispensable triar el mètode de detecció adequat a les necessitats de l'estudi. En l'epidemiologia d'aigües residuals per al seguiment de virus, els mètodes mol·leculars han estat aplicats amb èxit i la millora d'aquests ha permès l'elaboració d'estudis cada cop més exhaustius (Corpuz et al., 2020). A continuació es detallen les principals famílies de virus que s'excreten i com la seva detecció i caracterització s'ha vist millorada gràcies a l'aparició de noves eines basades en la seqüenciació massiva.

1.2. Principals famílies de virus excretats pels humans i caracterització de viromes

Es coneixen centenars de virus que infecten els éssers humans i són alliberats en la femta i l'orina fent el seu camí a l'ambient per l'excreció o secreció de fluids corporals o per la descamació epitelial. Els virus que infecten el tracte digestiu són en algunes ocasions excretats en elevades concentracions per persones sanes o malaltes durant llargs períodes de temps després de la resolució de la malaltia.

A la taula 1 s'indiquen les principals famílies virals que s'excreten, de quina manera ho fan i les possibles malalties relacionades amb la seva infecció.

Taula 1. *Principals famílies de virus que s'excreten a través de femta, orina o pell i les seves característiques. Adaptada de* Rusiñol i Girones, 2017.

Família (genoma, mida)	**Gènere**	**Espècies que infecten humans**	**Excreció**	**Malalties relacionades**
Adenoviridae (ADNdc, 70-90nm)	*Mastadenovirus*	Adenovirus humans (HAdV) A-G	Persistent	Gastroenteritis, infeccions respiratòries, conjuntivitis, cistitis
Anelloviridae (ADNmc, 30-32nm)	*Alphatorquevirus*	Torquetenovirus (TTV)	Persistent	Possible associació amb diverses malalties com hepatitis o malalties pulmonars
Astroviridae (ARNmc, 28-30nm)	*Mamastrovirus*	Astrovirus (HAstV) 1-9	Associada a malaltia	Gastroenteritis
Caliciviridae (ARNmc, 27-40nm)	*Norovirus*	Norovirus (NoV) GI, GII, GIV	Associada a malaltia	Gastroenteritis
	Sapovirus	Sapovirus (SaV) GI, GII, GIV, GV	Associada a malaltia	Gastroenteritis
Circoviridae (ADNmc, 15-20nm)	*Circovirus*	Circovirus humà (HuACV) 1	Associada a malaltia / Persistent	Possible associació amb infeccions sistèmiques
	Cyclovirus	Ciclovirus humà (HuACyV) 1-12	Associada a malaltia / Persistent	Possible associació amb infeccions sistèmiques
Coronaviridae (ARNmc, 80-220nm)	*Betacoronavirus*	SARS-CoV-2	Associada a malaltia	Síndrome respiratòria aguda greu
Hepeviridae (ARNmc, 27-34nm)	*Orthohepevirus*	Virus de l'hepatitis E G1,2,3,4,7 (HEV)	Associada a malaltia	Hepatitis aguda
Picornaviridae (ARNmc, 24-30nm)	*Enterovirus*	Enterovirus A-D (EV-68 i EV-71), Rhinovirus A-C,	Associada a malaltia	Paràlisis, meningitis, malaltia boca-mà-peu, anomalies cardíaques, erupció cutània, poliomielitis.
	Hepatovirus	Virus de l'hepatitis A GI-III (HAV)	Associada a malaltia	Hepatitis aguda
	Kobuvirus	Aichivirus A-C (AiV)	Associada a malaltia	Gastroenteritis
	Parechovirus	Parechovirus 1-16 (PeV)	Associada a malaltia	Gastroenteritis, infeccions respiratòries, encefalitis, meningitis, hepatitis
Reoviridae (ARNdc, 70-75nm)	Rotavirus	Rotavirus A-G (RoV)	Associada a malaltia	Gastroenteritis

Parvoviridae (ADNmc, 18-26nm)	*Erythroparvovirus*	Parvovirus B19 (PaV)	Associada a malaltia	Eritema infecciós en nens, artropatia, hepatitis
	Bocaparvovirus	Human bocavirus 1-4 (HBoV)	Associada a malaltia	Gastroenteritis, relacionats amb infeccions respiratòries
	Protoparvovirus	Bufavirus (HBuV)	Associada a malaltia	Gastroenteritis
	Dependovirus	Virus Adeno-associats (AAV)	Associada a malaltia?	Desconegut
Papillomaviridae (ADNdc, 50-60nm)	*Alphapapillomavirus*	Papil·loma virus humans (HPV) 16, 18, 6, 11, entre d'altres	Associada a malaltia / Persistent	Càncer de cèrvix, penis, anus, vulva i orofaringe. També associats a Berrugues genitals, infeccions respiratòries.
	Betapapillomavirus	HPV 20, 80, 110, 120, entre d'altres	Persistent	Poden estar associats a berrugues genitals, anals i perianals.
Polyomaviridae (ADNdc, 50 - 60nm)	*Betapolyomavirus*	Poliomavirus JC (JCPyV)	Persistent	Leucoencefalopatia multifocal progressiva (LMP)
		Poliomavirus BK (BKPyV)	Persistent	Infeccions respiratòries lleus, nefropatia, cistitis hemorràgica
		Poliomavirus WU (WUPyV)	Persistent	Possible associació amb infeccions respiratòries
		Poliomavirus KI (KIPyV)	Persistent	Possible associació amb infeccions respiratòries
	Alphapolyomavirus	Poliomavirus de les cèl·lules de Merkel (MCPyV)	Persistent	Associats al carcinoma de cèl·lules Merkel

1.2.1. Virus associats a infeccions agudes

Caliciviridae

La família *Caliciviridae* engloba un gran nombre de virus de mida petita (27–40nm) amb un genoma ARN de cadena senzilla i polaritat positiva. Dins d'aquesta família, els gèneres que estan relacionats amb el desenvolupament de la malaltia en humans son els *Norovirus* (NoV) i els *Sapovirus* (SaV) (Vinjé et al., 2019). Es coneix que s'excreten en títols de 10^7 a 10^9 còpies genòmiques per gram de femta (Bozkurt et al., 2015; Rimoldi et al., 2011).

Els NoV són els causant de la major part de brots de gastroenteritis associats a aigua i aliments, i es troben de forma abundant en matrius ambientals com son aigües residuals, de riu, de mar o aigües de rentat de mol·luscs amb contaminació fecal (Haramoto et al., 2018). Els principals genogrups de NoV que trobem en infeccions humanes son els genogrups 1, 2 i 4 (GI, GII, GIV), els quals presenten una freqüència de detecció diferenciada en l'ambient. Els NoV GI es detecten més freqüentment en aigües que no pas els NoV GII i GIV, suggerint una resistència major a condicions ambientals dels primers (Borchardt et al., 2012; Miura et al., 2013). Mitjançant la seqüenciació massiva s'han realitzat molts estudis sobre la detecció dels tipus circulants, donant informació de la distribució de genogrups i l'emergència de nous de subtipus com el NoV GII.17 (Fumian et al., 2019; Strubbia et al., 2019).

Els SaV també son els responsables de casos greus de gastroenteritis, amb un gran augment de la seva prevalença a nivell mundial, arribant a ser la tercera causa més comú de gastroenteritis (Iritani et al., 2016). La seva prevalença en aigües residuals ha demostrat ser alta en els darrers anys, amb una elevada diversitat de genogrups (Ibrahim et al., 2019).

Astroviridae

La família *Astroviridae* engloba un gran nombre de virus de mida petita (28-30 nm) amb un genoma ARN de cadena senzilla i polaritat positiva. Dins d'aquesta família, el gènere *Mamastrovirus* (MAstV) esta relacionat amb un elevat nombre de gastroenteritis en nens i en brots afectant adults. En femtes les seves concentracions es troben al voltant a les 10^8 a 10^{13} còpies genòmiques per gram de femta (Bosch et al., 2012; Payne, 2017).

Els MAstV es troben circulant activament en l'ambient, detectant-se en aigües residuals i efluents de depuradores, i aigües de pou i de riu amb contaminació fecal. Mitjançant estudis de metagenòmica, s'ha trobat una

alta diversitat de tipus, incloent emergents (com els MAstV-8), poc habituals i recombinants (Hata et al., 2018).

Reoviridae

Dins de la família *Reoviridae* (virus ARN de cadena doble fragmentat*)* i del gènere *Rotavirus*, els Rotavirus A (RVA) són el virus més habitualment associats a gastroenteritis infantil arreu del món, convertint-se en una amenaça per a la salut pública ja que s'excreten en grans quantitats en femta (10^{10} a 10^{12} còpies genòmiques per gram de femta) (Rimoldi et al., 2011).

Els RVA són molt estables i resistents als diferents mètodes de tractament d'aigües residuals, persistint en el medi ambient durant períodes prolongats i convertint-se així en potencials fonts d'infecció. Per una altra banda, l'impacte de la vacunació ha fet que variïn els genogrups circulants amb el temps i en funció de les polítiques territorials de vacunació (Matthijnssens et al., 2012; Ruggeri et al., 2015).

Picornaviridae

Un elevat nombre de gèneres, amb ampli ventall de malalties associades, s'agrupen en la família *Picornaviridae*. Els membres d'aquesta família son virus ARN de cadena senzilla d'una mida al voltant de 24-30nm.

Encara que no estan associats a episodis de diarrea, els gèneres *Enterovirus* i *Hepatovirus* (amb el virus de l'hepatitis A com a membre principal), s'excreten en femtes quan estan causant una infecció activa en ordres de 10^3 a 10^7 còpies genòmiques per gram de femta. És per aquest fet que s'ha reportat la seva presència en l'ambient (Bisseux et al., 2020; La Rosa et al., 2014; Lizasoain et al., 2018) i el seu potencial risc de transmissió a través d'aigua contaminada fecalment (Glas et al., 2001; Sinha i Dutta, 2019).

El gèneres *Kobuvirus* i *Parechovirus*, amb Aichivirus (AiV) i Parechovirus (PeV) respectivament com a espècies principals, sí que estan relacionats amb casos de gastroenteritis. Els AiV s'han detectat en coinfeccions amb RVA i també com a agents etiològics causants de brots alimentaris (Rivadulla i Romalde, 2020). Els PeV en canvi, estan relacionats amb casos lleus de gastroenteritis, però el seu rol en brots de la malaltia ha estat qüestionat recentment, suggerint la possibilitat de la simptomatologia sigui deguda a coinfeccions (Olijve et al., 2018). AiV i PeV s'han trobat en mostres ambientals en gairebé tots els mesos de l'any en que els s'han analitzat mostres (Thongprachum et al., 2018).

Hepeviridae

Els virus de la família *Hepeviridae* es caracteritzen per tenir un ARN de cadena senzilla i polaritat positiva i una mida de 27–34nm (Purdy et al., 2017). Dins del gènere *Orthohepevirus* trobem el virus de la Hepatitis E (VHE), un virus emergent responsable de brots d'hepatitis aguda relacionats amb el consum de carn poc cuinada (Bouwknegt et al., 2007; Meng i Lindsay, 2009). S'ha descrit la seva presència en el 70% de les femtes d'individus infectats, amb una excreció de 10^3 a 10^7 còpies genòmiques per gram de femta (Kim et al., 2014; Takahashi et al., 2007).

La circulació del VHE en l'ambient està àmpliament documentada, trobant-se, per exemple, en més del 30% de les aigües residuals analitzades en un estudi a Espanya (Clemente-Casares et al., 2003). Aquesta contaminació pot considerar-se com una font important de dispersió del virus (Rodriguez-Manzano et al., 2010; Rusiñol et al., 2015).

Coronaviridae

Membres de la família *Coronaviridae (ARN de cadena senzilla, 80-220nm de mida)* tradicionalment s'han relacionat amb el desenvolupament de refredats comuns (subespècies 229E, NL63, OC43 i

HKU1) i, en els darrers anys, de brots de malalties respiratòries greus (subespècies MERS, SARS) (Van der Hoek, 2015).

L'aparició del SARS-CoV-2, membre emergent del gènere *Betacoronavirus* causant de la pandèmia de la malaltia COVID-19, ha fet de la seva excreció un dels principals eixos d'estudi per al seu monitoreig ambiental. Es coneix que entre el 30% i el 55% dels malalts excreten el virus en femta, el qual pot arribar a concentracions des de 10^2 a 10^7 còpies genòmiques per gram de femta (Wölfel et al., 2020; Wu et al., 2020; Zhang et al., 2020). És així, que la seva circulació en aigües residuals està àmpliament demostrada (Farkas et al., 2020; Kitajima et al., 2020).

1.2.2. Virus persistentment excretats

Adenoviridae

En la família *Adenoviridae* (ADN de doble cadena, 70-90nm), dins del gènere *Mastadenovirus*, les espècies d'adenovirus humans (HAdV) A a G poden estar implicades en el desenvolupament de malalties de diversa severitat, des de gastroenteritis moderades a infeccions respiratòries o oculars (Lynch i Kajon, 2016).

Els HAdV es poden excretar en concentracions que es troben sobre 10^7 a 10^{11} còpies genòmiques per gram de femta i la seva circulació és global (Lion et al., 2010; Rimoldi et al., 2011). Aproximadament en més del 95% de les aigües residuals es pot detectar la presència de HAdV (Farkas et al., 2020; Rusiñol et al., 2015). Aquesta alta prevalença, juntament amb la seva fàcil detecció i capacitat de cultiu, han fet que s'utilitzin exitosament com a indicadors de contaminació fecal humana (Bofill-Mas et al., 2006; Pina et al., 1998; Rames et al., 2016).

Polyomaviridae

Els virus de la família *Polyomaviridae* presenten un ADN circular de doble cadena i una mida de 50-60nm. Les espècies de *Betapoliomavirus* JC i BK (JCPyV i BKPyV) produeixen infeccions ubiqües i persistents en individus sans que, en situacions d'immunodepressió, poden reactivar-se causant un ampli ventall de malalties (Doerries, 2006). El poliomavirus de les cèl·lules Merkel (MCPyV) ha estat relacionat amb el desenvolupament d'un carcinoma de pell agressiu (Feng et al., 2008; Varga et al., 2009).

El 85% de la població presenta anticossos contra JCPyV, amb concentracions d'excreció en orina de fins a 10^5 còpies genòmiques per mil·lilitre d'orina (Fumian et al., 2010). Tal i com passa amb els HAdV, l'anàlisi de la seva presència es pot utilitzar com a indicador de contaminació fecal humana (Bofill-Mas et al., 2000).

Papillomaviridae

Els papil·lomavirus humans (HPV, ADN circular de doble cadena, 50-60 nm) pertanyen a la família *Papillomaviridae* i es distribueixen en cinc gèneres. Els membres del gènere *Alphapapillomavirus* infecten principalment les superfícies mucoses orals i genitals, així com els genitals externs, i estan relacionats amb el desenvolupament de carcinomes. Els HPV dels gèneres *Betapapillomavirus, Gammapapillomavirus i Mupapillomavirus* infecten la mucosa no genital i la pell. Les infeccions per *Betapapillomavirus* són en molts casos persistents i poden donar lloc a una àmplia varietat de manifestacions clíniques en les superfícies epidèrmiques. La seva excreció s'ha relacionat amb la descamació d'epitelis i per orina i la presència d'aquests virus en l'ambient es podria considerar emergent i ha estat reportada en diversos ecosistemes (Di Bonito et al., 2015; Hamza i Hamza, 2018; La Rosa et al., 2013).

1.2.3. Virus excretats sense un paper etiològic clar

Existeixen famílies de virus de les que es desconeix la seva implicació en el desenvolupament de malalties però que s'excreten, a vegades de manera persistent, i es poden detectar en mostres ambientals, formant part del viroma. L'aplicació de la metagenòmica ha permès saber més d'elles i de les seves possibles implicacions en la salut pública.

En la família *Parvoviridae* (ADN circular de cadena senzilla, 18-26nm) a part de la presència de tipus que s'associen a malalties respiratòries i gastrointestinals, trobem espècies víriques, com els virus adeno-associats, de les que es desconeix la seva implicació en el desenvolupament de patologies. Virus de la família *Anelloviridae* (ADN circular de cadena senzilla, 30-32nm) com els torque-tenos (TTV) han estat postulats com possibles a agents etiològics de diverses malalties com hepatitis o infeccions pulmonars, però la seva implicació encara està en estudi (Reshetnyak et al., 2020). Amb tot això, actualment se sap que una gran quantitat de virus d'ADN de cadena senzilla circular, similars a la família *Circoviridae*, s'han detectat a través de la metagenòmica en aigües residuals, demostrant la necessitat de més estudis enfocats a aquestes famílies i el seu potencial patogènic (Blinkova et al., 2009; Rusiñol i Girones, 2017).

Els bacteriòfags, o fags, que infecten bacteris associats amb l'intestí humà també estan presents en altes concentracions en aigües residuals. Conseqüentment, els colifags somàtics i els bacteriòfags d'ARN F-específics s'utilitzen per avaluar la contaminació fecal humana (Muniesa et al., 2009), encara que, com que no totes les soques s'associen exclusivament a bacteris humans, la seva utilització com a marcadors s'ha d'aplicar amb precaució (Farkas et al., 2020). Un altre tipus de fags amb potencial utilització com a marcadors son els que infecten a *Bacteroides sp*p. Entre aquests, trobem els *crAssphage,* bacteriòfags que han co-

evolucionat amb els humans i recentment han estat descrits com a part integral del viroma del sistema digestiu humà. Els *crAssphage* es troben globalment distribuïts i la seva concentració en aigües residuals és de 10^4 a 10^7 còpies genòmiques per 100 mil·lilitres d'aigua (Edwards et al., 2019; García-Aljaro et al., 2017; Kongprajug et al., 2019).

Una família de virus que durant molt temps ha tingut un paper desconegut com a possibles patògens és la família *Picobirnaviridae*. En el passat, molts autors els hi atribuïen un rol en el desenvolupament de diarrea però sense poder demostrar una associació epidemiològica consistent (Van Leeuwen et al., 2010). Estudis recents sobre els motius d'unió a ribosomes dels picobirnavirus han postulat que aquests virus no infectarien animals i, de fet, podrien classificar-se com bacteriòfags (Krishnamurthy i Wang, 2018).

Molts virus de plantes es troben a la femta humana i conseqüentment en aigües residuals (Zhang et al., 2006). En els darrers anys un d'aquests virus, el *pepper mild mottle virus* (PMMoV; família *Virgaviridae*), s'ha trobat també en aigües superficials i subterrànies i en aigua potable. L'excreció del PMMoV es produeix per la ingesta de pebrots, o aliments que en continguin, que molts cops estan infectats pel virus (Rosario et al., 2009). Degut a la seva ubiqüitat, s'utilitzen també com a indicadors de contaminació fecal humana (Farkas et al., 2020).

Amb l'expansió de les eines de seqüenciació massiva, també ha sorgit el terme "matèria obscura viral" (*viral dark matter)*, que es refereix a la gran quantitat de seqüències, possiblement de virus encara desconeguts, que es troben formant part del viroma i que també poden ser persistentment excretats. Segons quin tipus de mostra s'estigui estudiant, aquest gran grup de seqüències que no es poden identificar poden representar del 40 al 90% del total (Krishnamurthy i Wang, 2017). Millores en les aplicacions de seqüenciació han ajudat en el descobriment de virus nous, però encara

que queda molt camí per recórrer en l'estudi de la relació del viroma i la salut (Wang, 2020).

2. Tècniques de seqüenciació massiva per a l'estudi del viroma

2.1. Eines d'estudi de virus en mostres ambientals i clíniques prèvies a la era de la metagenòmica

El descobriment de virus durant la major part del segle anterior va seguir un enfoc tradicional d'aïllament de virus i l'amplificació en cultiu cel·lular o models animals, classificant-los en funció de les seves propietats morfològiques i serològiques. Aquests mètodes es basen en l'observació de l'efecte citopàtic, l'efecte que genera la infecció viral en la cèl·lula hoste, i tenen com a principal desavantatge que existeixen espècies virals que no produeixen aquest efecte de manera evident o ho fan en un temps massa elevat. Per una altra banda, existeixen virus que no es poden cultivar o dels que no disposem de línies cel·lulars per a fer-ho (Leland i Ginocchio, 2007).

Els mètodes moleculars són els més àmpliament utilitzats ja que, en comparació a les tècniques de cultiu, són mètodes més específics, sensibles, ràpids i econòmics. Dins d'aquests, els més clàssicament utilitzats són els basats en la tècnica de la PCR, que permeten detectar la presència de genomes vírics mitjançant l'ús d'encebadors específics amb els que s'amplifiquen les regions d'interès dels virus. Existeixen variacions que ens permeten augmentar la sensibilitat per realitzar la detecció si la concentració vírica és baixa (PCR niuada/nested o nPCR) o si volem quantificar els genomes d'interès (PCR quantitativa o qPCR). Aquests mètodes no ens permeten tenir dades sobre la infectivitat dels virus d'estudi, a no ser que els complementin amb tractaments per eliminar genomes virals no encapsulats o amb l'adició de compostos que només

s'uniran a partícules víriques íntegres (potencialment infeccioses) (Bofill-Mas et al., 2006; Quijada et al., 2016; Randazzo et al., 2018).

Recentment, s'ha descrit l'ús de la tècnica d'edició genètica CRISPR/Cas per a la detecció dels àcids nucleics virals en mostres clíniques de manera ràpida, fàcilment miniaturitzable i amb baix cost, basada en l'ús de l'activitat de tall d'enzims CRISPR per al reconeixement d'ADN o ARN vírics (Gootenberg et al., 2018; Mirzaei et al., 2020)

La pròpia naturalesa dels mètodes mencionats fa que molts cops no es pugui reflectir la diversitat vírica real que hi ha en una mostra ja que són mètodes molt específics o estan enfocats a un grup determinat de virus, pel que cal conèixer prèviament el virus d'estudi (Kumar et al., 2017). El sorgiment de la seqüenciació massiva aplicada a l'estudi de poblacions víriques cobreix aquest buit de coneixement, permetent un estudi exhaustiu d'aquestes poblacions i la seva presència en mostres clíniques o ambientals. L'aplicació d'aquestes tècniques, les quals estan evolucionant de forma molt ràpida, ha esdevingut clau per a descriure de forma precisa els virus que circulen per la població i l'ambient, constituint-se com a eines de gran importància per a estudis epidemiològics i de salut pública i per al descobriment de virus nous (Cantalupo et al., 2011; Fernandez-Cassi et al., 2018; Wang et al., 2020).

2.2. Tècniques de seqüenciació massiva aplicades a la virologia

L'any 2008 la seqüenciació massiva va ser aplicada per primer cop per a la identificació d'agents virals desconeguts, els arenavirus, en pacients amb una malaltia febril que no va poder ser diagnosticada per altres mètodes (Palacios et al., 2008). Posteriorment, va començar a ser extensament utilitzada per estudis de metagenòmica, transcriptòmica i per al descobriment de virus nous, esdevenint una important eina per a

l'estudi de la diversitat vírica en tot tipus de mostres, com aigües o mostres clíniques (Barzon et al., 2013; Kapgate et al., 2015).

Una de les principals limitacions en l'aplicació de la seqüenciació massiva per a l'estudi de virus és que en molts tipus de mostres els genomes virals es troben en menor proporció respecte a genomes bacterians o del propi hoste, en especial en mostres ambientals com poden ser aigües o aliments. Previ a tot el procediment de preparació de la seqüenciació massiva vírica, necessitem assolir una concentració d'ADN mínima que no obtenim al realitzar l'extracció d'àcids nucleics. És per aquest motiu, que s'ha de realitzar un pre-tractament del concentrat viral i de l'extracció d'àcids nucleics. També, i a diferència de la seqüenciació massiva bacteriana, els virus no disposen de regions del genoma compartides que puguin ser usades per fer una amplificació selectiva o un enriquiment previ, fent tot el processament més complex.

Abans d'iniciar l'extracció de l'ADN i ARN víric, els concentrats virals s'han de tractar amb enzims que degradin els àcids nucleics lliures. Un cop feta l'extracció, es poden seguir mètodes d'amplificació de desplaçament múltiple (mètodes MDA) o independents de seqüència (mètodes SISPA) adaptats al processament de material genètic víric (Allander et al., 2001; Cantalupo et al., 2011; Kohl et al., 2015). El darrer, SISPA, és el mètode més extensament utilitzat, i es basa en realitzar una retrotranscripció del ARN i una complementació de les dobles cadenes (mitjançant un enzim sequenasa o klenow) amb un encebador amb nonàmers aleatoris que contenen una seqüència inicial compartida, anomenat *Primer A*. Aquesta seqüència inicial s'incorpora a tots els àcids nucleics de la mostra, el que permetrà després amplificar-la per PCR mitjançant un encebador específic (Fig. 3).

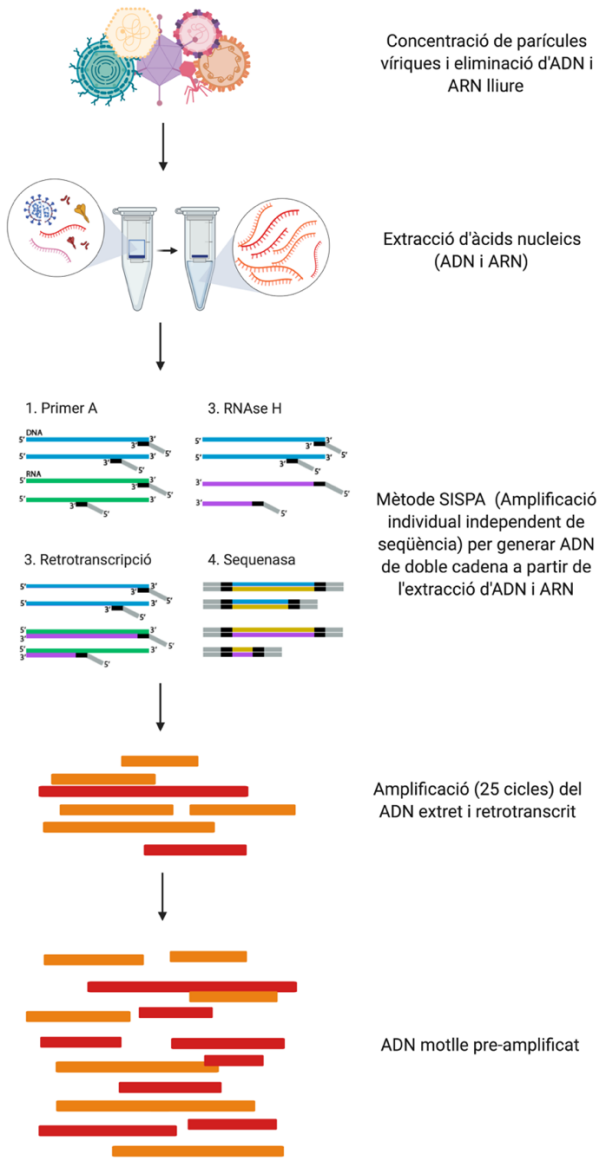


Figura 3. *Protocol de pre-amplificació d'àcids nucleics. Mètode SISPA.*

El nombre de cicles d'amplificació realitzats pot afectar la diversitat de seqüències obtingudes degut a biaixos inherent de la PCR, per això és necessari ajustar el nombre de cicles (25-30) per a obtenir la quantitat de ADN per a la realització de les llibreries de seqüenciació (Fernández-Cassi et al., 2018).

2.2.1. Seqüenciació massiva per metagenòmica

El tipus de seqüenciació massiva més utilitzat fins la data ha estat la metagenòmica no dirigida o d'alt rendiment (*High Troughtput Sequencing*, HTS). L'objectiu d'aquesta tècnica és la seqüenciació de tots els genomes que es troben en una mostra.

Tal i com s'ha comentat anteriorment, per tal de realitzar una seqüenciació massiva de tots els genomes vírics d'una mostra, s'ha de realitzar un pre-processament per incrementar la concentració d'ADN motlle. Un cop realitzat aquest pas, crític especialment en matrius ambientals, es construeixen les llibreries de seqüenciació.

Existeixen molts protocols per a fer llibreries i aquests varien en funció de la plataforma de seqüenciació escollida o de la seva automatització (Head et al., 2014). En aquest protocols, es parteix de l'ADN motlle de doble cadena que serà fragmentat físicament, per exemple amb un sonicador Covaris, o enzimàticament, que no requereix de maquinària extra encara que és menys eficient. Si s'utilitzen seqüenciadors que tenen la capacitat de seqüenciar fragments llargs, com és el cas de les plataformes per d'Oxford Nanpore, el pas de fragmentació es pot ometre. Seguidament, es realitza la reparació dels extrems de les seqüencies obtingudes i es lliguen els adaptadors de seqüenciació que permetran el reconeixement i la unió de les seqüències al suport que correspongui de cada seqüenciador. A la vegada, i si es vol seqüenciar diverses mostres alhora, es lliguen uns marcadors específics, anomenats índexs, que etiquetaran totes les seqüències obtingudes segons de la mostra que provinguin (Fig. 4). Un cop feta aquesta lligació d'adaptadors i índexs, es pot realitzar una selecció de fragments, generalment amb boles magnètiques, per a reduir fragments massa llargs o curts que empitjorarien la qualitat de la seqüenciació (Guan i Sung, 2016). Segons el protocol utilitzat, les llibreries poden ser amplificades per PCR i purificades abans de carregar-les al seqüenciador.

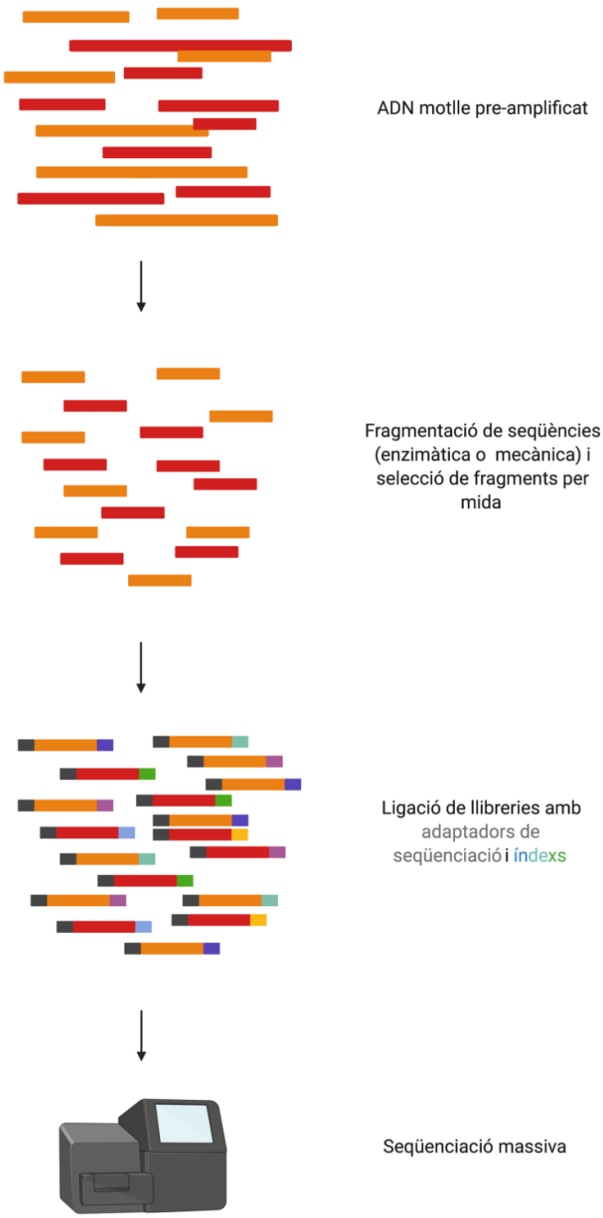


Figura 4. *Protocol de seqüenciació massiva per metagenòmica. Producció de llibreries.*

Aquest protocols han estat utilitzat per a estudis de viroma en una gran diversitat de matrius. En matrius ambientals la seva aplicació va permetre, a partir de l'any 2008, l'estudi de bacteriòfags i virus animals (Parsley et al., 2010; Rosario et al., 2009), i millores en les metodologies utilitzades van fer possible estudis més complets de viromes d'aigües residuals i

biosòlids en les que es van detectar un alt nombre de virus patògens d'interès humà (Bibby et al., 2011; Bibby i Peccia, 2013; Cantalupo et al., 2011; Ng et al., 2012). En l'actualitat, els estudis del viroma amb aquestes metodologies donen informació valuosa per a la epidemiologia basada en aigües residuals i el seu potencial paper com a sistema de detecció precoç de brots (Taula 2) (Haramoto et al., 2018).

Taula 2. *Revisió d'estudis de metagenòmica en aigües residuals (Novembre 2020).*

Títol	**Autors i any**	**Matriu**	**Principals virus detectats**
Census of the Viral Metagenome within an Activated sludge Microbial Assemblage	Parsley et al., 2010	Fangs actius	Bacteriòfags
Raw sewage harbors diverse viral populations	Cantalupo et al., 2011	Aigua residual	Bacteriòfags, *Adenoviridae, Astroviridae, Caliciviridae, Papillomaviridae, Parvoviridae, Picobirnaviridae, Picornaviridae, Polyomaviridae*
Viral Metagenome Analysis to Guide Human Pathogen Monitoring in Environmental Samples	Bibby et al., 2011	Biosòlids	Bacteriofags, *Herpesviridae, Coronaviridae, Picornaviridae, Adenoviridae, Flaviviridae, Circoviridae*
High Variety of Known and New RNA and DNA Viruses of Diverse Origins in Untreated Sewage	Ng et al., 2012	Aigua residual	Bacteriòfags, *Adenoviridae, Astroviridae, Caliciviridae, Hepeviridae, Parvoviridae, Picornaviridae, Reoviridae*
Identification of Viral Pathogen Diversity in Sewage Sludge by Metagenome Analysis	Bibby i Peccia, 2013	Biosòlids	*Papillomaviridae, Adenoviridae, Parvoviridae, Circoviridae, Coronaviridae, Parvoviridae, Reoviridae, Caliciviridae, Flaviviridae, Astroviridae, Herpesviridae*
Metagenomic approaches for direct and cell culture evaluation of the virological quality of wastewater	Aw et al., 2014	Aigua residual	Bacteriòfags, virus de plantes, *Adenoviridae, Polyomaviridae, Picornaviridae, Papillomaviridae, Ascoviridae, Hytrosaviridae, Baculoviridae, Poxviridae, Iridoviridae, Circoviridae, Herpesviridae*
Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance	Fernández-Cassi et al., 2018	Aigua residual	Bacteriòfags, virus de plantes, *Adenoviridae, Circoviridae, Parvoviridae, Caliciviridae, Astroviridae, Picornaviridae, Polyomaviridae, Hepeviridae*

Characterization of Norovirus and Other Human Enteric Viruses in Sewage and Stool Samples Through Next-Generation Sequencing	Strubbia et al., 2019	Aigua residual i femtes	Bacteriòfags, *Astroviridae, Caliciviridae, Nodaviridae, , Picornaviridae, Reoviridae*
Variations among Viruses in Influent Water and Effluent Water at a Wastewater Plant Over One Year as Assessed by Quantitative PCR and Metagenomics	Wang et al., 2020	Efluents de depuradora	Bacteriòfags, virus de plantes, *Adenoviridae, Picornaviridae, Caliciviridae, Hepeviridae, Reoviridae, Astroviridae*
Metagenomic analysis of viruses, bacteria and protozoa in irrigation water	Rusiñol et al., 2020	Aigua residual i aigües d'irrigació	Bacteriòfags, virus de plantes, *Circoviridae, Picornaviridae, Astroviridae, Caliciviridae, Adenoviridae, Hepeviridae, Anelloviridae, Papillomaviridae*

Aquests protocols han estat també satisfactòriament aplicats en la detecció de virus en mostres clíniques, demostrant el seu potencial ús com a eines de diagnòstic que permeten alhora l'estudi de diverses mostres i diversos grups vírics (Bodewes et al., 2015; Zhou et al., 2016). També s'han utilitzat exitosament per a descriure l'agent etiològic de malalties en les que es desconeixia (Moore et al., 2015; Palacios et al., 2008).

La seqüenciació massiva per metagenòmica té com a principal avantatge la possibilitat d'estudiar el viroma complet sense un biaix cap a famílies de virus concretes. En contrapartida, en mostres complexes on les seqüències virals poden suposar només un 1% del total generat pel seqüenciador, aquesta aproximació pot suposar una pèrdua d'informació de seqüències de variants o tipus vírics que es troben en una concentració menor. És per aquest motiu que es van desenvolupar aplicacions de seqüenciació dirigida, o amb enriquiment de dianes, cap a famílies o grups vírics d'interès, les quals han estat implementades en aquesta tesi doctoral.

2.2.2. Seqüenciació amb enriquiment de dianes

Incrementar la sensibilitat de la seqüenciació massiva cap a famílies o grups de virus d'interès ha estat un dels objectius principals per molts grups de recerca i per a possibles aplicacions en el diagnòstic clínic. Incorporar un biaix en la producció de llibreries de seqüenciació pot ser de gran utilitat per a no perdre rendiment de seqüenciació en reads de virus i altres organismes que no son el focus d'estudi.

Amb aquesta idea, a l'any 2015 va néixer les primera aplicació de seqüenciació amb enriquiment de dianes per a virus. Briese i col·laboradors van idear un protocol de sondes de captura que estava dirigit a la seqüenciació dirigida de virus que infecten vertebrats, el panell *VirCapSeq-VERT*. Es van dissenyar 2 milions de sondes que cobrien els genomes complerts de 207 virus que infecten vertebrats. Aquest enriquiment va permetre un augment d'entre 100 i 1000 vegades de seqüències virals en mostres d'origen humà (sang i teixit pulmonar). El mètode desenvolupat també va demostrar ser útil per a la descripció de virus nous, ja que s'identificaven genomes de virus que eren fins a un 40% diferents dels utilitzat en el disseny de les sondes (Briese et al., 2015).

Es va dissenyar un panell de captura seguint el protocol de construcció de llibreries i hibridació desenvolupat per Nimblegen-Roche, els quals van comercialitzar posteriorment el kit complet. El fonament del mètode de captura està representat a la figura 5. Un cop les llibreries de seqüenciació estan preparades, aquestes s'hibriden en líquid amb les sondes marcades amb biotina durant un màxim de 20 hores. Posteriorment, les seqüències unides a les sondes es recuperen mitjançant l'adició de boles magnètiques recobertes d'estreptavidina. Es produirà la unió del conjugat biotina-estreptavidina que a la vegada estarà unit a les seqüències de les llibreries que hibridin amb les sondes. Amb successius procediments de rentats s'eliminen aquelles seqüències que no hagin estat reconegudes per les sondes, obtenint així per a seqüenciar únicament els fragments

corresponents al virus d'interès. Aquest panell ha estat aplicat amb èxit en diversos estudis, millorant la detecció de virus d'interès que d'altra manera no es detectarien en mostres clíniques (Anderson et al., 2018; Mishra et al., 2020; Tokarz et al., 2019), ambientals (Hjelmsø et al., 2019; Strubbia et al., 2019) o associades a brots sense etiologia descrita (Cummings et al., 2019; Williams et al., 2018).

Un altre panell que segueix el mateix fonament és el de Wylie i col·laboradors, anomenat *ViroCap*. En aquest cas, la hibridació és produïda amb la mateixa tecnologia però en un microxip i les sondes dissenyades cobreixen virus de 34 famílies diferents que infecten vertebrats (Wylie et al., 2015).

Altres empreses com *Agilent Technologies*, *Illumina* o *Twist Biosciences*, disposen de tecnologies de captura similars que es poden adaptar a dissenys a mida de sondes per a la seqüenciació dirigida a famílies de virus concretes (Chalkias et al., 2018; Li et al., 2017; Miyazato et al., 2016; No et al., 2019; Oba et al., 2018). Recentment, dos kits de seqüenciació amb enriquiment de dianes, *VirCapSeq VERT* de *Roche* i el *SARS-CoV-2 Research Panel* de *Twist Biosciences*, han estat utilitzats per l'estudi genètic del SARS-CoV-2 en mostres clíniques (Carbo et al., 2020; Klempt et al., 2020), demostrant la seva gran utilitat en aplicacions de seqüenciació massiva.

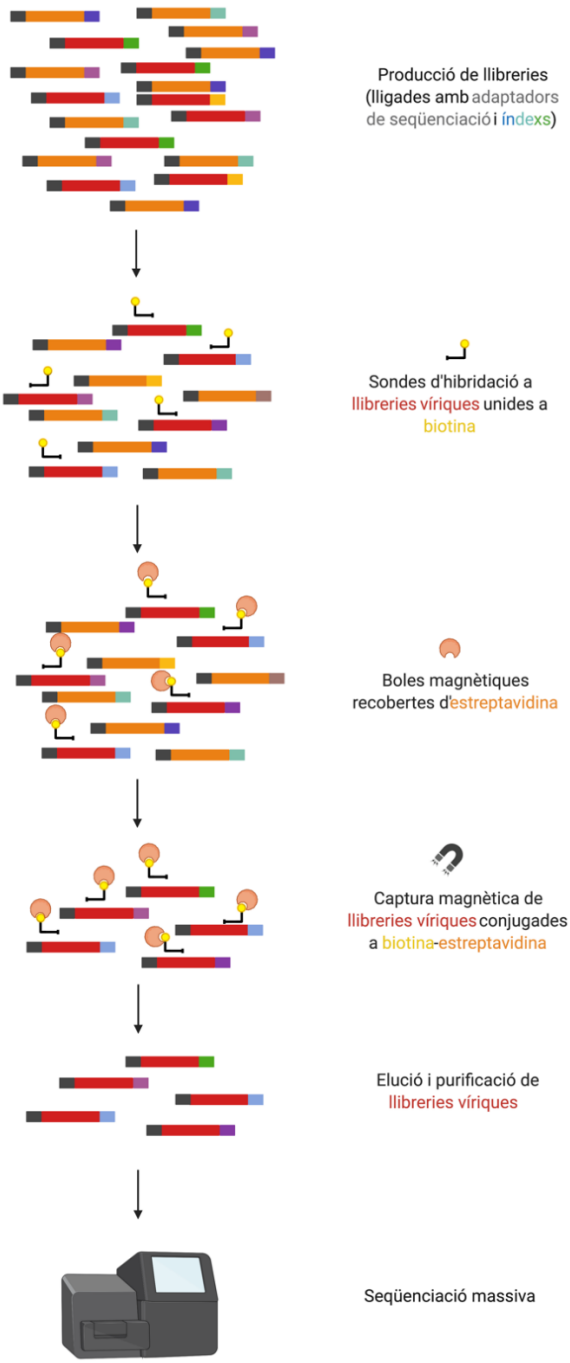


Figura 5. *Protocol de seqüenciació amb enriquiment de dianes mitjançant sondes conjugades amb biotina.*

Aquestes tècniques milloren substancialment els resultats obtinguts per a l'estudi de virus per seqüenciació massiva i la seva aplicació esta en constant evolució i millora. La principal limitació es troba en el biaix inherent de l'enriquiment per a tipus de virus coneguts, però tal i com passa amb kits ja descrits, la hibridació es pot fer menys restrictiva per a obrir la possibilitat del descobriment de nous virus. Una altra limitació que podem tenir en l'aplicació de la hibridació de sondes pot ser la falta de cobertura del genoma en estudis de variants víriques, ja que si alguna regió del virus d'estudi no hibrida amb les sondes dissenyades (noves variants, pèrdua de fragments), no tindríem la totalitat de les regions cobertes. Una metodologia que supera aquest inconvenient és la seqüenciació massiva d'amplicons.

2.2.3. Seqüenciació massiva d'amplicons

Tradicionalment, la seqüenciació d'amplicons de PCR s'ha realitzat mitjançant la tècnica de Sanger (Sanger et al., 1977). El principal inconvenient d'aquesta tècnica és que, si en la mostra d'estudi tenim diverses variants d'un mateix virus, la seqüència obtinguda serà la més abundant en la població i per seqüenciar les variants s'hauria de realitzar el clonatge de poblacions. La seqüenciació massiva d'amplicons permet realitzar estudis d'aquest tipus d'una manera molt més eficient, obtenint una gran quantitat de seqüències en un únic assaig i permetent la seqüenciació de moltes mostres alhora. Amb tot això, aporta informació rellevant sobre mutacions de baixa freqüència i l'estructura poblacional (Beerenwinkel i Zagordi, 2011).

A la figura 6 es detalla el protocol de seqüenciació massiva d'amplicons. Amb el producte que s'obté a partir d'una o vàries PCR o PCR niuades es realitza la purificació i lligació d'adaptadors de seqüenciació. La incorporació d'aquests adaptadors, que seran reconeguts per la plataforma de seqüenciació, es pot realitzar també a l'hora de fer la PCR, incorporant aquestes seqüències a les seqüències del encebadors que s'utilitzen. Per

seqüenciar varis amplicons de diverses mostres al mateix temps, es pot també lligar en els amplicons els índex que identificaran cada mostra amb una seqüència única.

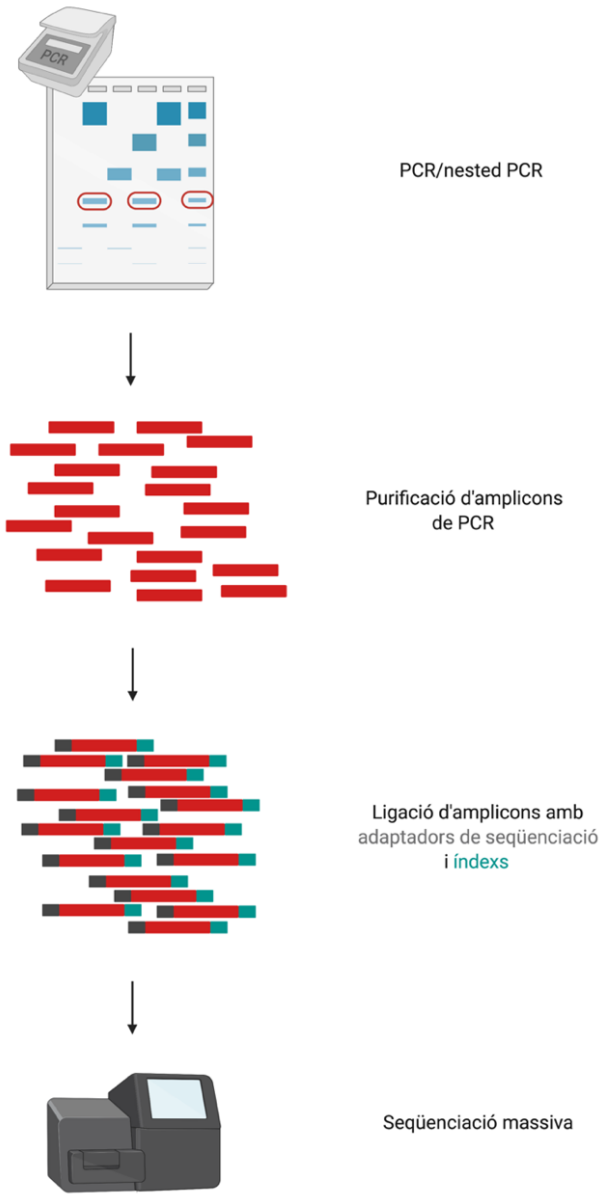


Figura 6. *Protocol de seqüenciació massiva d'amplicons.*

Aquesta aproximació ha obert la possibilitat d'estudis detallats de famílies específiques de virus, generant informació de variants genètiques dins de les pròpies espècies (estudis de quasiespècies) (Del Campo et al., 2018; Kijak et al., 2019), reconstruccions de genomes (Charre et al., 2020; Maurier et al., 2019) i estudis de diversitat viral de famílies específiques en mostres clíniques o ambientals (Fernandez-Cassi et al., 2018; Hata et al., 2018; Prevost et al., 2015).

2.3. Plataformes de seqüenciació massiva

Les tecnologies de seqüenciació d'ADN han existit des de principis de la dècada de 1970, però inicialment el seu cost, la seva complexitat i la utilització de reactius tòxics o radioactius van limitar el seu ús en recerca. Els mètodes de terminació de cadena iniciats per Sanger et al. (Sanger et al., 1977) eren més pràctics, i van formar la base per a la primera generació de seqüenciadors d'ADN automatitzats. El primer genoma complet d'un microorganisme de vida lliure, *Haemophilus influenza*, publicat el 1995 (Fleischmann et al., 1995), va ser seqüenciat utilitzant el mètode Sanger. No obstant això, la seqüenciació del genoma sencer per aquesta tecnologia va ser molt costosa i lenta.

La necessitat de disposar de tecnologies de seqüenciació d'alt rendiment es va intensificar per l'inici del Projecte Genoma Humà. Es va tardar més de 15 anys, amb una inversió de quasi 3 billons de dòlars, per realitzar la seqüenciació Sanger del genoma humà (Venter et al., 2001; International Human Genome Sequencing Consortium, 2004). Avui en dia, gràcies a l'adveniment dels mètodes de seqüenciació massiva, es pot realitzar en menys d'una setmana i el cost no supera els 700 dòlars (Wetterstrand, 2020).

La seqüenciació massiva ha esdevingut una tecnologia de reemplaçament ja que la informació generada a partir de múltiples fluxos de treball tradicionals es pot combinar en un únic flux de treball eficient i amb un

cost cada cop menor. Juntament amb la bioinformàtica i el desenvolupament informàtic està en ràpida evolució, s'espera que es produeixin encara més millores propers anys que afectaran en gran manera l'aplicació de la seqüenciació massiva en tota mena d'àmbits de recerca i clínics (Besser et al., 2018).

L'aparició de la primera plataforma de seqüenciació massiva data del 2005 de la mà de la empresa 454 Life Sciences (posteriorment comprada per Roche) i el seu piroseqüenciador. La seva aplicació en la seqüenciació de genomes es va demostrar amb la exitosa seqüenciació del bacteri *Mycoplasma genitalium* (Margulies et al., 2005). Aquesta tecnologia d'alt rendiment va permetre la generació i detecció de milers a milions de lectures de seqüenciació en una sola màquina sense necessitat de clonació. Des de llavors, han sorgit moltes altres tecnologies de seqüenciació massiva que generen tant lectures curtes (50-400 bp) com llargues (1-100 kb). A la taula 3 es resumeixen les principals característiques de les plataformes de seqüenciació que es detallaran a continuació.

Taula 3. *Principals plataformes de seqüenciació i característiques.*

Plataforma	**Companyia**	**Any sortida**	**Reacció seqüenciació**	**Amplificació**	**Mètode de detecció**	**Fragments**
454	*Roche*	*2005*	Piroseqüenciació	SI (PCR en emulsió)	Nucleòtids fluorescents	Curts
MiSeq, HiSeq, MiniSeq NextSeq, NovaSeq	*Illumina*	*2007*	Síntesi	SI (PCR en pont)	Nucleòtids fluorescents	Curts
IonTorrent	*Thermo Fisher*	*2010*	Síntesi	SI (PCR en emulsió)	Canvis de pH	Curts
PacBio	*Pacific Biosciences*	*2011*	Síntesi	NO	Nucleòtids fluorescents	Llargs
MinION, GridION, PromethION	*Oxford Nanopore Technologies*	*2015*	Pas per nanopor	NO	Canvis de corrent	Llargs

2.3.1. Seqüenciació de segona generació. Seqüenciació de fragments curts

Les tecnologies de seqüenciació de fragments curts son les que es coneixen com a tecnologies de segona generació. Els fragments curts poden ser ensamblats bioinformàticament com es fa amb les seqüències que s'obtenen de la seqüenciació de Sanger. La principal diferència és que la regió compartida entre seqüències és menor en les resultants de la seqüenciació massiva, degut a que els fragments son més curts, i l'ensamblat en alguns casos és limitat. Totes les plataformes tenen associades en el seu procediment l'amplificació de l'ADN de les llibreries de seqüenciació i actualment són les més extensament utilitzades.

A continuació s'expliquen les principals plataformes de seqüenciació de segona generació.

Roche 454

Aquesta plataforma de seqüenciació té com a fonament el mètode de piroseqüenciació. En aquest mètode, les llibreries de seqüenciació s'uneixen, mitjançant adaptadors, a una petita esfera. L'ADN és amplificat mitjançant PCR en emulsió de manera que cada esfera acaba tenint moltes còpies idèntiques de la mateixa molècula d'ADN. Aquestes esferes són immobilitzades en unes plaques de fibra òptica que contenen més d'un milió de pous i en cada pou es col·locarà una esfera per a produir la reacció de seqüenciació. Amb la piroseqüenciació, la incorporació de cada nucleòtid durant la síntesi de l'ADN resulta en l'alliberació d'un pirofosfat inorgànic que proporciona l'energia necessària per activar l'enzim luciferasa, generant una emissió de llum que serà captada. Els quatre nucleòtids flueixen seqüencialment sobre la placa, en un ordre definit, de manera que la producció d'un pols de llum identificarà quina base ha sigut inserida (Fig. 7) (Mardis, 2008).

La limitació més important d'aquesta tècnica és el baix nombre de *reads*, el seu elevat preu, així com biaixos en la seqüenciació de regions amb repeticions de més de tres nucleòtids (homopolímers). Aquesta tecnologia es considera obsoleta i està en desús.

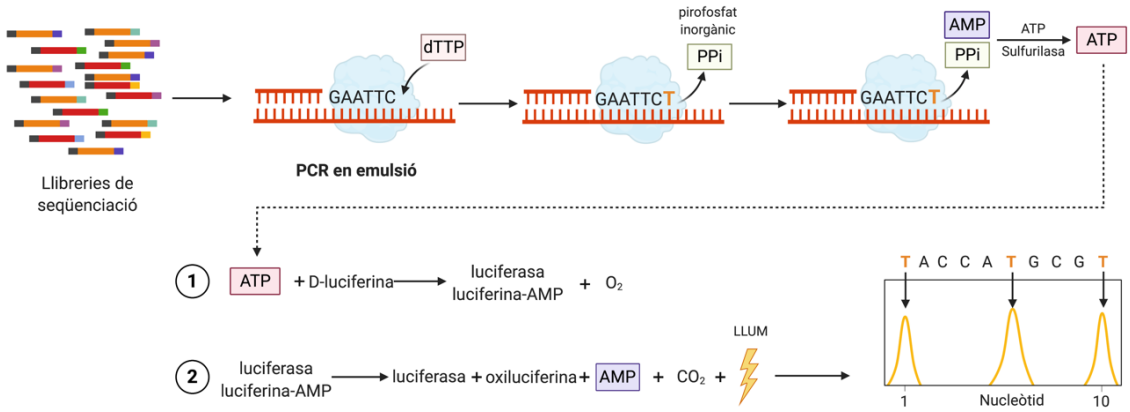


Figura 7. *Esquema del funcionament de la piroseqüenciació, base de la seqüenciació en la plataforma 454 Roche.*

Solexa-Illumina

El primer seqüenciador *Solexa*, anomenat *Genome Analyzer,* va sortir al mercat a l'any 2006. Aquesta plataforma generava 1 Gb de seqüències en un únic run. Al 2007, la companyia *Solexa* va ser comprada per *Illumina* i els seus seqüenciadors es van convertir, i ho segueixen sent avui en dia, en referent en plataformes de seqüenciació massiva arreu del món. Les seves plataformes han seguit evolucionant amb els anys per aconseguir un nombre cada cop major de seqüències en cada procés de seqüenciació independent. Les plataformes MiniSeq i MiSeq ofereixen un rendiment de mostra baix a mitjà, preus d'instruments assequibles, un flux de treball fàcil d'utilitzar sense necessitat d'automatització i un cost raonable per mostra, convertint-se en una opció atractiva per als laboratoris de diagnòstic i salut pública. Els instruments NextSeq, HiSeq i NovaSeq estan dissenyats per a un rendiment molt més alt reduint el cost per mostra, però

requereixen una automatització addicional per a la preparació de les llibreries de seqüenciació i per tant estan enfocats a grans centres de seqüenciació. De totes les tecnologies de seqüenciació massiva disponibles fins a la data, el seqüenciadors d'*Illumina* són el que obtenen una assignació de bases (*basecalling*) amb major qualitat (Uelze et al., 2020).

Aquestes plataformes de seqüenciació tenen com a fonament principal l'amplificació en pont amb marcatge fluorescent (Bentley et al., 2008). La llibreria preparada es carrega i s'immobilitza a la superfície de la cel·la de flux la qual conté seqüències unides covalentment que són complementàries a les seqüències dels adaptadors que s'han utilitzat per fer la llibreria (Mardis, 2013). Aquestes seqüències complementàries capturen l'ADN i serveixen d'encebadors per a l'amplificació en pont de la cadena complementària. Es produeixen successives rondes d'amplificació que generen clústers de molècules amplificades a partir d'una sola molècula original (Fig. 8).

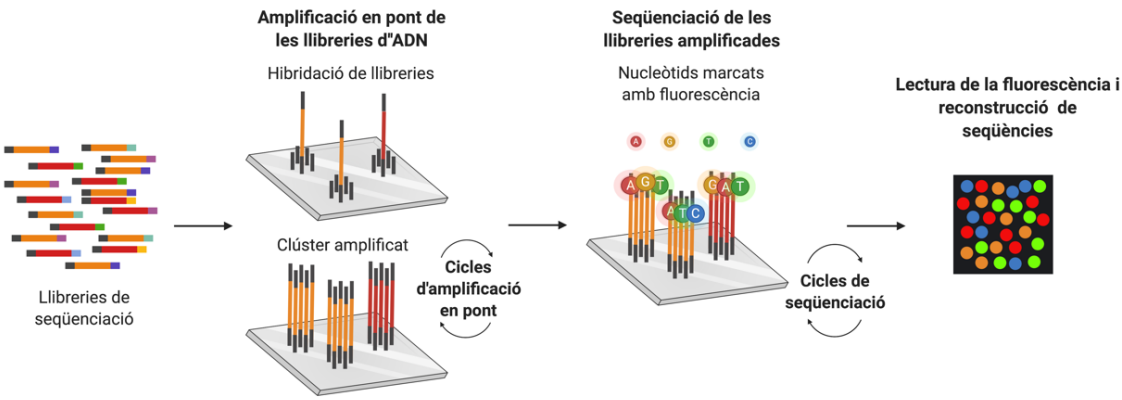


Figura 8. *Esquema de la seqüenciació mitjançant amplificació en pont en la que es basen els sistemes d'Illumina.*

Els clústers d'amplificació s'utilitzaran per a la seqüenciació que es produeix a continuació, la qual es fa mitjançant la síntesi d'ADN amb nucleòtids marcats fluorescentment i que a l'incorporar-se aturen la

síntesi. Després de cada incorporació, els clústers són monitoritzats i es fa la lectura de la base corresponent a cada posició de la seqüència. Abans de produir-se el següent cicle, es reverteix químicament el fluorocrom del nucleòtid que havia aturat la síntesi temporalment, permetent l'addició del següent nucleòtid en la següent ronda de seqüenciació (Radford et al., 2012).

La lectura seqüencial d'un sol nucleòtid per cicle assegura una baixa taxa d'error (al voltant de l'1%) inclòs en regions amb repeticions de bases. Per contra, degut a que l'addició de nucleòtids en el clúster no es fa alhora en totes les seqüències que es sintetitzen, la qualitat del senyal que s'obté va disminuint al llarg dels cicles de seqüenciació. Aquest darrer fet condiciona un dels desavantatges de la tecnologia, ja que no la fa apta per a seqüències més llargues de 300 parells de bases (Tan et al., 2019).

Ion Torrent

La plataforma Ion Torrent, de *Thermo Fisher*, va sortir al mercat a l'any 2010 per a postular-se com a competidora de les tecnologies del moment. En comptes d'utilitzar nucleòtids amb marcatge fluorescent, el fonament de la seqüenciació es basa en la tecnologia semiconductora, mesurant els ions d'hidrogen que s'alliberen en la síntesi de l'ADN (Myers i Rothberg, 2011).

La llibreria de seqüenciació s'amplifica mitjançant una PCR en emulsió. Aquesta PCR es realitza en els pous d'una cel·la de flux en la que, mitjançant un sensor de conductivitat, es detecten els canvis de pH que es produeixen al alliberar-se ions d'hidrogen. A continuació, s'assignen a cada posició els nucleòtids corresponents segons els canvis de conductivitat produïts (Fig. 9) (Mardis, 2013).

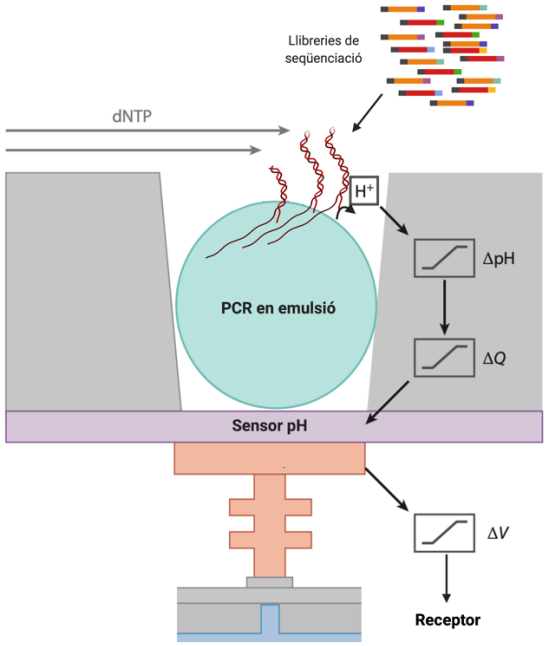


Figura 9. *Esquema de la seqüenciació mitjançant la plataforma Ion Torrent. Adaptat de Mardis et al.* (Mardis, 2013)

El principal inconvenient d'aquesta tecnologia de seqüenciació és l'alta taxa d'errors en regions homopolimèriques llargues, degut a la saturació que es produeix de la sonda de pH. En comparació a les plataformes desenvolupades per *Illumina*, la preparació de les llibreries de seqüenciació és més laboriosa (encara que es pot automatitzar) però el run de seqüenciació més ràpid (Besser et al., 2018; Mardis, 2013).

2.3.2. Seqüenciació tercera generació. Seqüenciació de fragments llargs

Les tecnologies seqüenciació de tercera generació es caracteritzen per tenir la capacitat de seqüenciar llargues molècules d'ADN sense una amplificació prèvia per PCR. Aquestes tecnologies són de gran utilitat per l'estudi de regions genòmiques grans, complexes o amb moltes

repeticions. En contrapartida, la taxa d'error que es produeix en la seqüenciació és major que en les plataformes de segona generació.

PacBio

La primera plataforma en permetre la seqüenciació d'una única molècula d'ADN va ser PacBio, de *Pacific Biosciences*, a l'any 2011.

El mètode de seqüenciació que s'utilitza és el que es coneix com a SMRT (*Single Molecule Real-time*), en el que la seqüenciació es produeix a temps real. Les reaccions es produeixen dins de nanocavitats en les que l'ADN polimerasa allarga la cadena creixent afegint nucleòtids marcats cadascun amb un fluorocrom diferent. A la vegada que es van incorporant, es registra l'emissió de fluorescència per un làser i el fluoròfor del nucleòtid és eliminat abans de que s'incorpori el següent nucleòtid. D'aquesta manera, els marcadors no s'acumulen en l'ADN i es permet la detecció de cada nucleòtid que s'incorpora a la cadena de forma individual (Logsdon et al., 2020).

S'ha incrementat la longitud mitjana de les seqüències i la precisió en la seqüenciació en els darrers anys, realitzant millores en les reaccions enzimàtiques i la detecció òptica. Actualment la taxa d'error es troba al voltant del 10% (McCombie et al., 2019).

Seqüenciació per nanopors: MinION

El concepte de seqüenciació per nanopors va aparèixer per primer cop en una publicació de Kasianowicz et al., a l'any 1996 (Kasianowicz et al., 1996), però no va ser fins l'any 2015 que la companyia *Oxford Nanopore Technologies* va comercialitzar la primera plataforma de seqüenciació utilitzant aquesta tecnologia, anomenada MinION.

Es distingeix del totes les plataformes de seqüenciació anteriors per detectar directament els nucleòtids d'una cadena d'ADN sense la seva síntesi activa, motiu per a que alguns autors la consideren com a tecnologia de "quarta generació". A mesura que una molècula senzilla d'ADN passa a través del nanopor, un sensor detecta els canvis de corrent iònic que genera cada tipus de nucleòtid, determinant així a temps real la seqüència d'ADN que passa pel porus (Feng et al., 2015; Jain et al., 2016).

Es carreguen les llibreries de seqüenciació produïdes a la cel·la de flux que conté 2000 nanopors (en grups de 4) en la seva superfície. L'adaptador que s'incorpora en la producció de les llibreries es lliga amb una proteïna motora, que permet que la doble cadena d'ADN es separi i una de les cadenes senzilles passi a través del porus. Al travessar-lo, cada nucleòtid produirà consecutivament una pertorbació en el corrent del porus que quedarà registrat (Fig. 10).

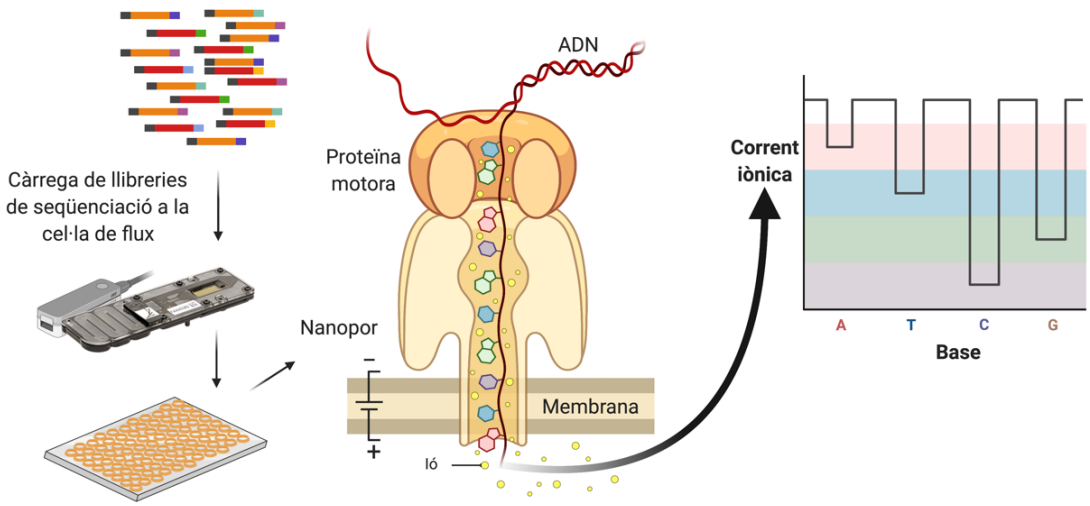


Figura 10. *Esquema de la seqüenciació mitjançant nanopors.*

El gran avantatge d'aquesta tecnologia és la seva mida (existeixen plataformes portàtils), la seva velocitat i la possibilitat de seqüenciar fragments molt llargs d'ADN. La seva taxa de error es troba al voltant del 5-10%, però la tecnologia es troba en constant actualització i recentment han sortit al mercat noves cel·les de flux i plataformes més avançades que augmenten la fidelitat de la seqüenciació i el volum de dades generades (Van Dijk et al., 2018).

2.4. Eines bioinformàtiques i possibilitats en l'anàlisi de dades

Les plataformes seqüenciació massiva poden generar de milers a milions de lectures de seqüenciació en un sol experiment. No obstant això, la qualitat de les seqüències no és uniforme entre el conjunt de dades. Per tant, el primer pas per a processar-les és l'avaluació la qualitat de les seqüències obtingudes mitjançant diferents procediments bioinformàtics. La raó per la qual les seqüències o *reads* obtinguts han ha de ser filtrats, és per eliminar aquelles que poden esbiaixar els procediments següents, com son seqüències de baixa qualitat, els adaptadors de seqüenciació o duplicats discordants. La majoria d'eines que s'utilitzen per a fer el control de qualitat només estan disponibles per a sistemes operatius basats en UNIX, que no tenen una interfície gràfica d'usuari fàcil d'utilitzar per a no informàtics, no obstant això, existeixen algunes eines web que ho fan més accessible (Goecks et al., 2010; Vincent et al., 2017).

Després de la eliminació dels *reads* de baixa qualitat, els restants s'ensamblen per a formar seqüències d'una llargada major, anomenades *contigs*. Existeixen programes específics per a realitzar ensamblats que utilitzen diversos algoritmes per a la reconstrucció de seqüències, tots es basen en l'alineament de regions solapades entre seqüències amb paràmetres més o menys restrictius segons l'algoritme que s'utilitzi. Alguns d'ells també realitzen ensamblatges utilitzant genomes de referència (Honaas et al., 2016; Nurk et al., 2017). Aquests contigs

generats es poden ordenar i enllaçar entre ells en seqüències d'una llargada encara major anomenades *scaffolds*.

El pas final és l'assignació taxonòmica de les seqüències ja processades. Aquesta assignació es pot realitzar mitjançant l'alineament (o mapat) dels contigs contra bases de dades locals o contra bases de dades disponibles com poden ser GenBank ("National Center for Biotechnology Information (NCBI). Bethesda (MD).," 1988) o UniProt (Bateman, 2019).

Tot a quest procediment bioinformàtic és laboriós i complex (Fig. 11), i la disponibilitat de bases de dades actualitzades de virus és limitada. Recentment han sorgit eines pel processament de dades de metagenòmica vírica que apropen al investigadors no bioinformàtics el processament i anàlisi de les dades generades per les plataformes de seqüenciació massiva. Entre les que realitzen un processament de dades més complert i gratuït trobem les eines web *Genome Detective* (Vilsker et al., 2018) (https://www.genomedetective.com/app/typingtool/virus/) i *Kaiju* (Menzel et al., 2016) (http://kaiju.binf.ku.dk).

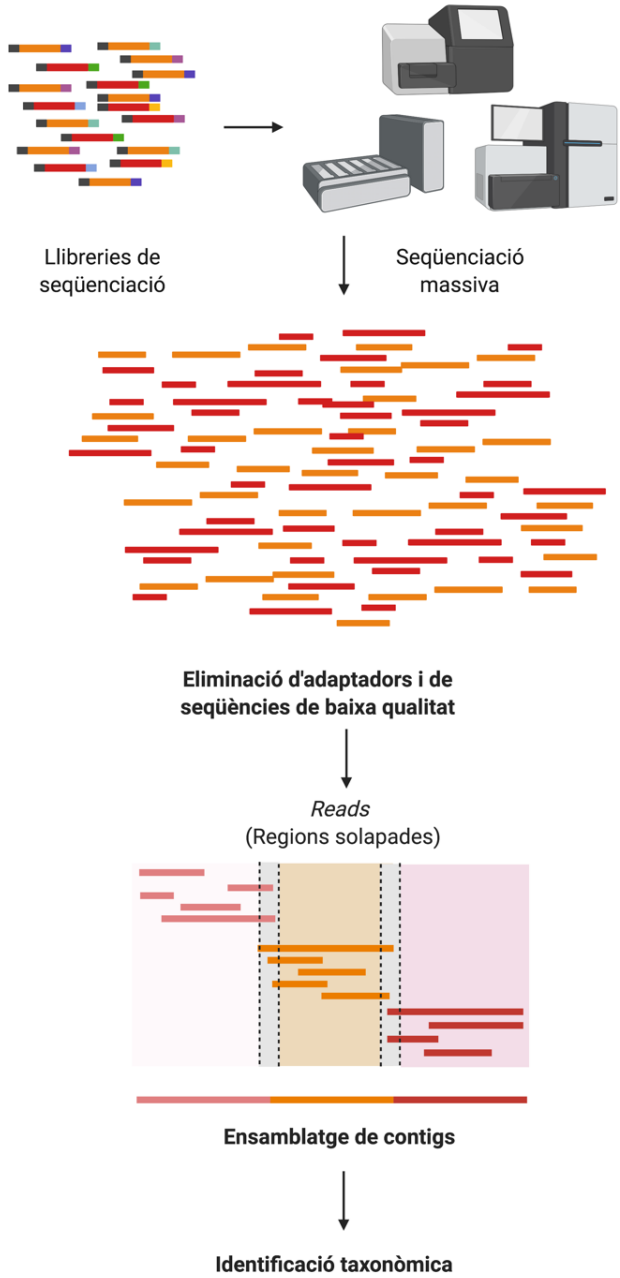


Figura 11. *Esquema del processament bioinformàtic de les dades generades en la seqüenciació massiva.*

OBJECTIUS

L'objectiu principal de la present tesi doctoral ha estat l'optimització i aplicació de procediments basats en la seqüenciació massiva per a la descripció del viroma associat a l'aigua residual i a altres mostres d'interès epidemiològic.

Els objectius específics han estat:

1. Optimitzar i aplicar diferents tècniques de seqüenciació massiva per a l'anàlisi de mostres ambientals i biològiques.

2. Caracteritzar el viroma de l'aigua residual amb focus en els patògens humans.

3. Avaluar l'ús del la seqüenciació massiva, amb i sense enriquiment de dianes, i de la seqüenciació d'amplicons per a l'estudi i monitoreig de brots infecciosos d'origen viral.

4. Aplicar la seqüenciació massiva a mostres de pacients amb gastroenteritis d'etiologia desconeguda per conèixer la presència d'espècies virals no detectades pels mètodes de diagnòstic clàssics.

5. Estudiar el viroma de diverses espècies de peixos Atlàntics amb potencials implicacions comercials i de seguretat alimentària.

INFORMES DELS ARTICLES

1. Llista d'articles inclosos en la tesi

La present tesi està basada en les següents publicacions:

Martínez-Puchol S., Rusiñol M., Fernández Cassi X., Timoneda N., Itarte M., Andrés C., Antón A., Abril J.F., Girones R., Bofill-Mas S. Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. *Science of the Total Environment* 713-136604, doi: 10.1016/j.scitotenv.2020.136604

Martínez-Puchol S., Itarte M., Rusiñol M., Forés E., Andrés C., Antón A., Quer J., Girones R., Abril J.F., Bofill-Mas S. SARS-CoV-2 is a new member of urban wastewater virome. Manuscrit en procés de revisió a la revista *Genes*.

Fernández Cassi X., **Martínez-Puchol S.**, Cornejo T., Bartolomé R., Bofill-Mas S., Girones R. Unveiling viruses associated with gastroenteritis using a metagenomics approach. *Viruses*, 12(12), 1432; https://doi.org/10.3390/v12121432.

Filipa-Silva A., Parreira R., **Martínez-Puchol S.**, Bofill-Mas S., Barreto Crespo M.T., Nunes M. The Unexplored Virome of Two Atlantic Coast Fish: Contribution of Next-Generation Sequencing to Fish Virology. *Foods*, 9, 1634, doi:10.3390/foods9111634.

2. Informe de coautoria

Martínez-Puchol S., Rusiñol M., Fernández Cassi X., Timoneda N., Itarte M., Andrés C., Antón A., Abril J.F., Girones R., Bofill-Mas S. Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. Science of the Total Environment 713-136604, doi: 10.1016/j.scitotenv.2020.136604

L'estudi ha format part del projecte nacional VIRTENGS (AGL2014-55081-R) en el que la doctoranda ha participat activament. La doctoranda ha participat en la gestió del projecte i en el disseny experimental i ha realitzat els experiments. Finalment, ha redactat el manuscrit sota la supervisió de les seves codirectores de tesi.

Martínez-Puchol S., Itarte M., Rusiñol M., Forés E., Andrés C., Antón A., Quer J., Girones R., Abril J.F., Bofill-Mas S. SARS-CoV-2 is a new member of urban wastewater virome. Manuscrit sota revisió en la revista Genes.

L'estudi ha format part del projecte nacional MetaPatFood (AGL2017-86797-C2-1-R) i del monitoreig d'aigua residual per a l'estudi de la incidència de SARS-CoV-2 en els que la doctoranda ha participat activament. La doctoranda ha participat en el disseny experimental i ha realitzat els experiments de seqüenciació massiva. Finalment, ha redactat el manuscrit sota la supervisió de les seves codirectores de tesi.

Fernández Cassi X., **Martínez-Puchol S.**, Cornejo T., Bartolomé R., Bofill-Mas S., Girones R. Unveiling viruses associated with gastroenteritis using a metagenomics approach. *Viruses*, 12(12), 1432; https://doi.org/10.3390/v12121432.

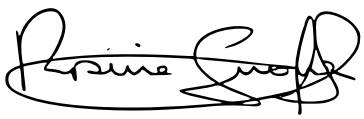
L'estudi ha format part del projecte nacional VIRTENGS (AGL2014-55081-R) en el que la doctoranda ha participat activament. La doctoranda ha participat en el disseny experimental, ha caracteritzat les mostres, ha

analitzat les seqüències virals i ha redactat el manuscrit juntament amb el primer signant de l'article, sota la supervisió de les seves codirectores de tesi.

Filipa-Silva A., Parreira R., **Martínez-Puchol S.**, Bofill-Mas S., Barreto Crespo M.T., Nunes M. The Unexplored Virome of Two Atlantic Coast Fish: Contribution of Next-Generation Sequencing to Fish Virology. Foods, 9, 1634, doi:10.3390/foods9111634.

L'estudi ha format part d'una col·laboració amb el *Instituto de Biologia Experimental e Tecnológica* de Lisboa. La doctoranda ha preparat els SOPs (*Standard Operational Procedures*) i ha format als participants en les tècniques utilitzades. També ha contribuït en el disseny experimental i, juntament amb la primera signant del treball, ha realitzat la part experimental. Finalment ha col·laborat en la redacció i correcció del manuscrit.

Signat,



Dra. Rosina Girones Llop
Barcelona, Desembre 2020

Dra. Sílvia Bofill Mas
Barcelona, Desembre 2020

3. Informe sobre el factor d'impacte

Els treballs que formen part de la present tesi doctoral s'han publicat o estan en procés de revisió en revistes científiques rellevants per a la línia de recerca en la qual la doctoranda ha participat en el darrers anys.

L'article "Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens" va ser publicat al mes de gener de 2020 a la revista *Science of the Total Environment*. L'índex d'impacte de la revista a l'any 2019 era de 6,551 (Q1).

L'article "SARS-CoV-2 is a new member of urban wastewater virome" es troba actualment en procés de revisió (novembre 2020) a la revista *Genes*. L'índex d'impacte de la revista a l'any 2019 era de 3,750 (Q1).

L'article "Unveiling viruses associated with gastroenteritis using a metagenomics approach" va ser publicat al mes de desembre de 2020 a la revista *Viruses*. L'índex d'impacte de la revista a l'any 2019 era de 3,816 (Q1).

L'article "The Unexplored Virome of Two Atlantic Coast Fish: Contribution of Next-Generation Sequencing to Fish Virology" va ser publicat al mes de novembre de 2020 a la revista *Foods*. L'índex d'impacte de la revista a l'any 2019 era de 4,092 (Q1).

Signat,

Dra. Rosina Girones Llop
Barcelona, Desembre 2020

Dra. Sílvia Bofill Mas
Barcelona, Desembre 2020

CAPÍTOL 1: SEQÜENCIACIÓ MASSIVA PER A ESTUDIS D'EPIDEMIOLOGIA BASATS EN AIGÜES RESIDUALS

Article 1: Caracterització del viroma de l'aigua residual: comparativa d'eines de seqüenciació massiva i presència de patògens importants

"Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens."

Martínez-Puchol S., Rusiñol M., Fernández Cassi X., Timoneda N., Itarte M., Andrés C., Antón A., Abril J.F., Girones R., Bofill-Mas S.

Science of the Total Environment 713-136604, doi: 10.1016/j.scitotenv.2020.136604

Les tècniques de seqüenciació massiva són eines útils pel monitoreig i identificació del patògens vírics que circulen en la població però presenten certes limitacions, sobretot pel que fa a matrius complexes. L'aigua residual conté una elevada concentració d'altres microorganismes que poden interferir a l'hora de seqüenciar virus. Per això, és necessari realitzar optimitzacions prèvies al protocol de seqüenciació.

En el present estudi, s'han comparat diferents metodologies de seqüenciació massiva per a la detecció i caracterització de virus en l'aigua residual, amb especial focus en tres patògens importants: els virus del papil·loma humà, els adenovirus i els enterovirus. Les tècniques aplicades han estat: seqüenciació massiva per metagenòmica, seqüenciació amb enriquiment de dianes i seqüenciació massiva d'amplicons.

Aplicant la seqüenciació massiva per metagenòmica s'ha obtingut el perfil complet dels virus excretats, detectant membres de quatre famílies víriques diferents a més de bacteriòfags, virus de plantes i virus que infecten altres hostes. L'ús de la seqüenciació massiva amb enriquiment de dianes, en canvi, ha permès la detecció de vuit famílies de virus humans, amb una alta varietat de tipus en cada família i una elevada cobertura dels

genomes detectats, és a dir s'han pogut ensamblar trossos grans de genomes virals. Per últim, la seqüenciació massiva d'amplicons proporciona moltes dades sobre la diversitat genètica dels tres patògens d'estudi en comparació amb les altres dues tècniques aplicades. A més, ha permès detectar i subtipar seqüències d'enterovirus A71 C1, la soca causant del brot de romboencefalitis en nens que va ocórrer a Catalunya a la primavera de 2016, època de presa de mostra de l'aigua residual analitzada.

Amb els resultats obtinguts es pot concloure que diferents estratègies de seqüenciació massiva proporcionen diferents nivells d'anàlisis. L'aplicació de l'enriquiment de dianes ha esdevingut la millor estratègia per a descriure la presència de virus d'interès humà en mostres complexes, com l'aigua residual. Les altres eines utilitzades són útils per l'estudi del viroma complet d'una mostra (metagenòmica) o quan l'interès de l'estudi és la diversitat viral dins d'una mateixa família (seqüenciació d'amplicons).

Science of the Total Environment 713 (2020) 136604

Contents lists available at ScienceDirect

Science of the Total Environment

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

Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens

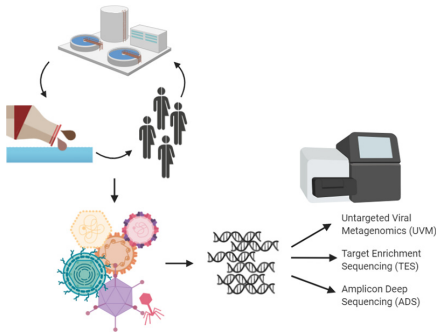
Sandra Martínez-Puchol [a,b,*], Marta Rusiñol [a,b], Xavier Fernández-Cassi [a], Natàlia Timoneda [a,c], Marta Itarte [a,b], Cristina Andrés [d], Andrés Antón [d], Josep F. Abril [c], Rosina Girones [a,b], Sílvia Bofill-Mas [a,b]

[a] Laboratory of Viruses Contaminants of Water and Food, Genetics, Microbiology &Statistics Dept., Universitat de Barcelona, Barcelona, Catalonia, Spain
[b] The Water Research Institute (IdRA); Universitat de Barcelona, Barcelona, Catalonia, Spain
[c] Computational Genomics Lab, Genetics, Microbiology & Statistics Dept., Universitat de Barcelona, Institut de Biomedicina (IBUB), Barcelona, Catalonia, Spain
[d] Respiratory Viruses Unit, Virology Section, Microbiology Department, Hospital Universitari Vall d'Hebron, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Barcelona, Spain

HIGHLIGHTS

- Different NGS methods have been applied to the study of sewage virome.
- Untargeted Viral Metagenomics provides a full picture of the complete virome.
- Target Enrichment Sequencing is the best option for detecting vertebrate viruses.
- Amplicon Deep Sequencing is useful for studying viral diversity.
- Enterovirus A71 related to an outbreak was identified with two of these methods.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 22 November 2019
Received in revised form 7 January 2020
Accepted 7 January 2020
Available online 08 January 2020

Editor: Damia Barcelo

Keywords:
Sewage virome
Target Enrichment Sequencing
Amplicon Deep Sequencing
Papillomavirus
Adenovirus
Enterovirus

ABSTRACT

NGS techniques are excellent tools to monitor and identify viral pathogens circulating among the population with some limitations that need to be overcome, especially in complex matrices. Sewage contains a high amount of other microorganisms that could interfere when trying to sequence viruses for which random PCR amplifications are needed before NGS. The selection of appropriate NGS tools is important for reliable identification of viral diversity among the population.

We have compared different NGS methodologies (Untargeted Viral Metagenomics, Target Enrichment Sequencing and Amplicon Deep Sequencing) for the detection and characterisation of viruses in urban sewage, focusing on three important human pathogens: papillomaviruses, adenoviruses and enteroviruses.

A full picture of excreted viruses was obtained by applying Untargeted Viral Metagenomics, which detected members of four different vertebrate viral families in addition to bacteriophages, plant viruses and viruses infecting other hosts. Target Enrichment Sequencing, using specific vertebrate viral probes, allowed the detection of up to eight families containing human viruses, with high variety of types within the families and with a high genome coverage.

By applying Amplicon Deep Sequencing, the diversity of enteroviruses, adenoviruses and papillomaviruses observed was higher than when applying the other two strategies and this technique allowed the subtyping of

* Corresponding author.
E-mail address: smartinezpuchol@ub.edu (S. Martínez-Puchol).

https://doi.org/10.1016/j.scitotenv.2020.136604

an enterovirus A71 C1 strain related to a brainstem encephalitis outbreak occurring at the same time in the sampling area.
From the data obtained, we concluded that the different strategies studied provided different levels of analysis: TES is the best strategy to obtain a broad picture of human viruses present in complex samples such as sewage. Other NGS strategies are useful for studying the virome of complex samples when also targeting viruses infecting plants, bacteria, invertebrates or fungi (Untargeted Viral Metagenomics) or when observing the variety within a sole viral family is the objective of the study (Amplicon Deep Sequencing).

1. Introduction

Viruses excreted by humans in faeces, urine and skin desquamation together with animal viruses and viruses infecting invertebrates, plants, bacteria and fungi constitute the sewage virome, a complex matrix that contains a large variety of pathogenic and commensal viruses and could give important information on persistent and acute infections affecting the population.

Currently, Next Generation Sequencing (NGS) methodologies are increasingly viewed as promising tools for the comprehensive study of microorganisms in a wide variety of samples and settings, replacing traditional molecular methods. The main reasons for this new trend are: the capacity to process larger number of samples simultaneously, the reduction in sequencing costs, the ease with which samples and libraries can be prepared and the development of faster and more efficient bioinformatic tools with which to process the huge amount of data. (Huang et al., 2019).

The application of NGS techniques was easily incorporated into the study of bacterial communities even for environmental samples, but the lack of shared regions in viral genomes make them more challenging for viral sequencing. To overcome this problem, the use of random-primer-based sequencing facilitates the development of different protocols adapted for viral NGS (Cantalupo et al., 2011; Kohl et al., 2015; Ng et al., 2012), making this methodology key for the study of viral communities and viral discovery while providing enough starting material for library preparation.

The excreted virome of a population provides critical information about virus circulation, the introduction of emergent viruses and how they are transmitted among the population. The accurate study of the virome with a focus on specific target viral groups urges the development of NGS protocols with higher sensitivity.

The main limitation when analysing the virome from any type of sample is the low proportion of viral sequences identified compared with the total number of sequences amplified when using random primers (Krishnamurthy and Wang, 2017; Santiago-Rodriguez and Hollister, 2019; Tamaki et al., 2012). This is especially critical for human viruses which seem to represent a low fraction of the whole virome of sewage, clearly dominated by phages and plant viruses (Cantalupo et al., 2011; Ng et al., 2012). Few published studies in viral sewage metagenomics describe the human virome together with protocols to achieve a better recovery of viral sequences by applying mainly negative viral selection methods, also called pre-extraction, which generally entail the use of, for example, filters, density gradients and nucleases (Cantalupo et al., 2011; Fernandez-Cassi et al., 2018a, 2018b; Hjelmsø et al., 2017). Viral selection approaches are effective for avoiding the high background presence of other non-desired nucleic acids from bacteria or other hosts. Apart from negative selection methods, positive selection methods, also referred to post-extraction methods or Target Enrichment Methods, are characterised by the use of probes within the PCR assays, microarrays or hybridisation (Kumar et al., 2017). These methods increase in viral sequence reads as well as in the breadth and depth of genome coverage, in some cases extending to the full genome (Paskey et al., 2019; Wylie et al., 2015). Among these, the most commonly used are those based on hybridisation probes, with different custom-made and commercial approaches available (Chalkias et al., 2018; Hjelmsø et al., 2019; Mühlemann et al., 2018), showing the potential use in viral discovery (Briese et al., 2015).

Another NGS approach useful for identification and typing of different viral species for outbreak control or environmental surveillance is Amplicon Deep Sequencing, based on the mass sequencing of traditionally sanger-sequenced PCR amplicons. This approach opened up new opportunities promoting the detailed study of specific families and their diversity within a sample, also proving useful for a different nature of samples, from studies of quasispecies of Hepatitis C Virus or Human Immunodeficiency Virus in clinical settings (Del Campo et al., 2018; Kustin et al., 2019) to environmental samples (Fernandez-Cassi et al., 2018a, 2018b; Hata et al., 2018; Prevost et al., 2015).

This study shows the wide diversity of viral pathogens identified in urban sewage over one year of study with the application of different NGS tools for the determination of the virome with different levels of analysis. To our knowledge, this is the first work aimed to compare two different NGS strategies, Untargeted Viral Metagenomics (UVM) and Target Enrichment Sequencing (TES), for the characterisation of the virome excreted in a population in terms of viral diversity, specificity and genome coverage. Also, Amplicon Deep Sequencing (ADS), was examined as an alternative strategy when a deeper study on a concrete viral family is needed. The work is focused on three specific viral families, human adenoviruses (HAdV), human papillomaviruses (HPV) and human enteroviruses (EV), since they are important pathogens and some of them are persistently excreted by the population. In addition, we analysed enterovirus nucleotide sequences obtained from clinical samples for evaluating the capacity of the studied NGS strategies to catch up viruses causing clinical disease in the population.

2. Material and methods

2.1. Sampling, viral particles concentration and nucleic acid extraction

In April 2016, a 24-hour urban raw sewage composite sample was collected from a wastewater treatment plant (WWTP), located in the north of Barcelona city (WWTP A), that treats a population equivalent of up to 2.8 million and receives domestic and industrial waste from the sewerage system. This composite sample was analysed to compare TES, UVM and ADS.

Additionally, monthly samples were collected over the year, in 2016, from a second WWTP located 30 km from Barcelona (WWTP B). The samples from this WWTP were pooled by season (spring, summer, autumn and winter) and analysed by ADS in order to evaluate specific viral groups, in depth, over one year. This plant serves up to 115,000 population equivalents.

Samples were collected in sterile containers and kept at 4 °C until concentrated within 24 h. Viral particles from 42 ml of sewage from each sample were concentrated by ultracentrifugation, as previously described (Bofill-Mas et al., 2006). DNase treatment (TurboDNAse, Ambion) and extraction of nucleic acids was performed as described previously (Fernandez-Cassi et al., 2018a, 2018b) using QIAmp RNA Viral Mini Kit (Qiagen).

2.2. Untargeted viral metagenomics and target enrichment sequencing

2.2.1. Random tagging of nucleic acids and pre-amplification

Samples were prepared prior to the library construction following the protocol described (Fernandez-Cassi et al., 2018a, 2018b). Briefly, in order to analyse both RNA and DNA viruses, NA were retrotranscribed with SuperScript III enzymes (Life Technologies) and a random nonamer primer. The second cDNA strand was obtained using Sequenase 2.0 (Thermo Fisher Scientific) and a PCR of 25 cycles was performed in order to obtain enough dsDNA for the next steps. These PCR products were purified and concentrated with the Zymo DNA Clean & Concentrate kit (Zymo Research) and quantified using Quibit 2.0 HSdsDNA kit (Life Technologies).

2.2.2. Library construction

For each sample, libraries were constructed in duplicate using KAPA HyperPlus Library Preparation Kit (KAPA Biosystems, Roche). An enzymatic fragmentation was performed in the previously obtained purification with a starting concentration ranging from 1 to 3 ng/μl following the manufacturer's instructions. After the fragmentation and the end-repair and A-tailing reaction, the adapter's ligation was performed. Using the KAPA Single-Indexed Adapter Kit (KAPA Biosystems, Roche), each sample was paired with the desired index. A post-ligation clean-up was followed with a double-sized size selection with the magnetic AMPure XP Beads (Beckman Coulter) to select those fragments of between 250 and 450 bp. Using an LM-PCR, sample libraries were amplified (seven cycles of amplification for the adapters used) and purified. The quality of the resulting libraries was assessed by using Agilent Bioanalyzer DNA 1000 assay (Agilent Technologies) and the concentration was measured using Qubit 2.0 (Life Technologies). Then, libraries were mixed together in two replicates, to obtain two pools containing 1 μg each. One pool was directly sequenced on an Illumina MiSeq 2x300bp platform and the other was captured later by using specific viral probes.

2.2.3. Capture of viral sequences by using the VirCapSeq-VERT Capture Panel

One of the pooled libraries was captured using the VirCapSeq-VERT Capture Panel (Roche). This panel consists of approximately two million probes, covering the genomes of 207 viral taxa known to infect vertebrates, thus enabling the detection of viral sequences in complex sample types (Briese et al., 2015). Using the HyperCap Target Enrichment Kit (Roche) and the HyperCap Bead Kit (Roche), the sample was prepared and then hybridised with the VirCapSeq-VERT probes at 47 °C for 20 h. Immediately after the hybridisation, the captured multiplex DNA sample was recovered with the Capture Beads (HyperCap Bead Kit, Roche), using a magnetic particle collector, and cleaned. The DNA captured, still bead-bounded, was amplified using an LM-PCR. This post-capture PCR was purified and the concentration and quality were checked, as mentioned before. Captured DNA was sequenced in the same run as non-captured DNA in an Illumina MiSeq 2x300bp platform.

2.3. Amplicon Deep Sequencing

Specific nested PCR, previously proved to be adequate for typing purposes, for HAdV, HPV and EV amplification were performed as described (De Roda Husman et al., 1995; Fernandez-Cassi et al., 2018a, 2018b; Forslund et al., 2003; WHO Regional Office for Europe, 2015) with incorporation of Illumina adapters in the nested primers. Amplicons were purified from agarose gel with the QIAquick Gel Extraction Kit (Qiagen) and then sequenced with an Illumina MiSeq 2x300bp platform.

2.4. Clinical enterovirus A71 sampling

Most of 25 EV-A71-studied cases by Vall d'Hebron Hospital Respiratory Viruses Unit corresponded to rhombencephalitis, while the remaining patients had hand-foot-mouth disease (2), gastroenteritis (2), aseptic meningitis (1), and acute bronchitis (1). EV-A71 was amplified from upper and lower respiratory tract samples from suspicious patients by using CE-marked commercial multiplex real-time RT-PCR-based assay (Anyplex II RV16 assay, Seegene, Korea). Total nucleic acids were previously extracted using NucliSens easyMAG (bioMérieux, Marcy líEtoile, France) according to the manufacturer's instructions and kept frozen (−20C) until use. An additional real-time RT-PCR (Seegene, Korea) was carried out to improve the detection of all EV strains (Gimferrer et al., 2015).

The partial coding sequences of the viral protein 1 (VP1) from EV amplicons, obtained according to the protocol recommended by the World Health Organisation (WHO Regional Office for Europe, 2015), were used to construct a phylogenetic tree.

2.5. Bioinformatics

2.5.1. UVM and TES bioinformatic processing and taxonomical assignment

Pair-end FASTAQ files generated from the sequencing were analysed using Genome Detective web-based software (https://www.genomedetective.com/)(Vilsker et al., 2018a). Briefly, low-quality reads and adapters were trimmed using Trimmomatic (Bolger et al., 2014), viral reads were selected using DIAMOND alignment method and non-viral sequences were discarded. Subsequently, viral reads were assembled with metaSPAdes (Nurk et al., 2017) and taxonomically classified with NCBI-BLASTX and NCBI-BLASTN against NCBI RefSeq viral database (Vilsker et al., 2018b; Wheeler et al., 2007), using only the contigs with 70% identity cut-off. Richness Chao1 ratio was calculated using the Catchall software, version 4.0 (Allen et al., 2013).

2.5.2. Amplicon Deep Sequencing

For the study of the amplicons, the quality of raw and clean read sequences was assessed using the FASTX-Toolkit software, version 0.0.14 (Hannon Lab, http://www.hannonlab.org). The cleaned reads were clustered using the software `CDHIT` with default parameters (Huang et al., 2010); the output was queried for sequence similarity using NCBI-BLASTN against a reference genome species-specific custom database for HADV, HPV and EV (Wheeler et al., 2007). Based on the best BLAST results (95% coverage and 95% identity cut-off), each cluster was classified into its taxonomic group.

2.5.3. Enterovirus phylogenetic analyses

A phylogenetic tree based on enterovirus VP1 nucleotide sequences of the EV-A71 C1 strains obtained from sewage and clinical samples was constructed by a using neighbour-joining method with Geneious software version 11.0 (https://www.geneious.com), where branches having a bootstrap value below 70% were discarded for the Figure. Coxackievirus A16 sequence (GenBank accession number KT327162) was used as the outgroup.

3. Results and discussion

3.1. Study of the sewage virome using Untargeted Viral Metagenomics and Target Enrichment Sequencing

An urban sewage composite sample obtained after 24 h of collection in WWTP A was analysed using UVM and TES in parallel. A total of 6.07 million viral reads were obtained when applying TES and 727,784 when applying UVM. Read counts for each of the viral groups detected are presented in Table 1. The Chao1 diversity index showed, as expected, a higher richness when using UVM, increasing from 170 obtained using the TES approach, to 311 when using UVM.

Table 1
Read counts and Genome coverage obtained from contigs of viral species of interest obtained by Target Enrichment Sequencing (TES) and Untargeted Viral Metagenomics (UVM).

		Viral sequence reads				
		Viral read count			Breadth genome coverage (%)	
Viral families	Viral species	TES	UVM	Fold increase (log)	TES	UVM
Papillomaviridae	Betapapillomavirus 2	24	0	1,38	15,47	0
Picornaviridae	Aichivirus A	23372	33	2,85	89,86	22,05
	Enterovirus A	725	0	2,82	61,12	0
	Enterovirus B	902	0	2,96	44,00	0
	Enterovirus C	2098	0	3,32	62,10	0
Adenoviridae	Human mastadenovirus A	343	0	2,54	14,22	0
	Human mastadenovirus F	1333	0	3,12	34,66	0
Polomaviridae	BK polyomavirus	829	0	2,92	79,89	0
	Human polyomavirus 6	79	0	1,90	10,88	0
	JC polyomavirus	714	237	0,48	93,82	77,37
	WU polyomavirus	17	0	1,23	10,96	0
Astroviridae	Mamastrovirus 1	83857	64	3,12	99,25	30,45
Caliciviridae	Sapporo virus	421	10	1,62	62,07	57,33
	Norwalk virus	1820	8	2,36	88,72	11,73
Hepeviridae	Orthohepevirus A	781	0	2,89	24,21	0
Reoviridae	Rotavirus A	423	0	2,63	38,64	0

Fold increase is presented in logs of increase and breadth genome coverage in percentages.

Probe enrichment increased the detection of sequences from vertebrate viruses by a factor of 81.04% (299,650 sequences) of the total sequences obtained, compared with the 2.74% (3549 sequences) obtained without applying enrichment. An exhaustive description of the total virome obtained using both methodologies is presented as Supplementary material 1.

A comparison of the viral species obtained via each methodology, based on host distribution, is shown in Fig. 1. Not only was the number of sequences from viruses that infect vertebrates higher after the TES approach, but also the number of species within each viral family increased. The *Picornaviridae* and *Parvoviridae* families showed a higher number of taxonomically assigned sequences after the application of the TES. When applying UVM, a wide variety of bacteriophage species was observed (49% of the total reads) within the families *Siphoviridae, Myoviridae, Microviridae* and *Podoviridae*, as well as species of the viral plant families (30% of the total reads) *Virgaviridae*, *Tombusviridae* and *Solemoviridae*. The predominance of these viral families in environmental settings has been reported before (Fernandez-Cassi et al., 2018a, 2018b) and is of importance because of their potential economic impact as plant pathogens. Reads from vertebrate families were identified as JC Polyomavirus, Mamastrovirus 1, Aichivirus A, non-human circoviruses and parvoviruses, representing 2.75% of the total sequences. Accordingly, UVM can be considered a reference tool for the study of a global picture of the whole virome of environmental and clinical samples, in accordance with other studies performed by our research group (Cantalupo et al., 2011; Fernandez-Cassi et al., 2018a).

The TES kit used in this study was developed in 2015 by Briese et al. (Briese et al., 2015) with the aim of capturing viral sequences from only those viruses that infect vertebrates, for use in clinical and veterinarian settings. This system uses more than two million probes of 207 different vertebrate viruses and has been reported useful for the detection of new variants. When applied to raw sewage, probes helped to capture vertebrate viruses (Briese et al., 2015), although some sequences from other viral hosts were still recovered. These kind of enrichment approaches are clearly biased towards specific viral families (Parras-Moltó et al., 2018) and while they might be useful for studying specific viral families (or groups of families), they do not provide the whole virome picture.

However, when focusing on viral families that infect humans, TES provided a more sensitive approach, allowing the detection of more viral families and catching a wide diversity of viral species within a given family while providing a higher number of reads. Members of the *Adenoviridae*, *Hepeviridae*, *Papillomaviridae*, and *Reoviridae* families as well as some species of the *Picornaviridae* (including an EV71 contig) and *Polyomaviridae* were only detected when using TES. Regarding the *Astroviridae* and *Caliciviridae* families, a big increase in sequences from Mamastrovirus 1 species (up to 3-log) and Norwalk and Sapporo viruses (up to 2-log) was observed when compared with UVM. Again, when processing these reads as contigs, a considerable increase in the genome coverage of each virus was observed after TES which is of huge interest for viral characterisation and discovery, as reported before (Briese et al., 2015). Contigs corresponding to Aichivirus A, Norwalk virus, JC Polyomavirus and Mamastrovirus 1 presented more than 90% genome coverage. Indeed, for the latter two, almost the totality of the genome was covered, at 93.82% and 99.25% respectively, whereas without enrichment the respective coverage of these viruses was 77.37% and 30.45% (Table 1). This fact could be due to the higher number of sequences obtained with TES and the wider distribution of reads from across all viral genomes.

From the data obtained, we conclude that TES is the best strategy to obtain a broad picture of human viruses present in complex samples

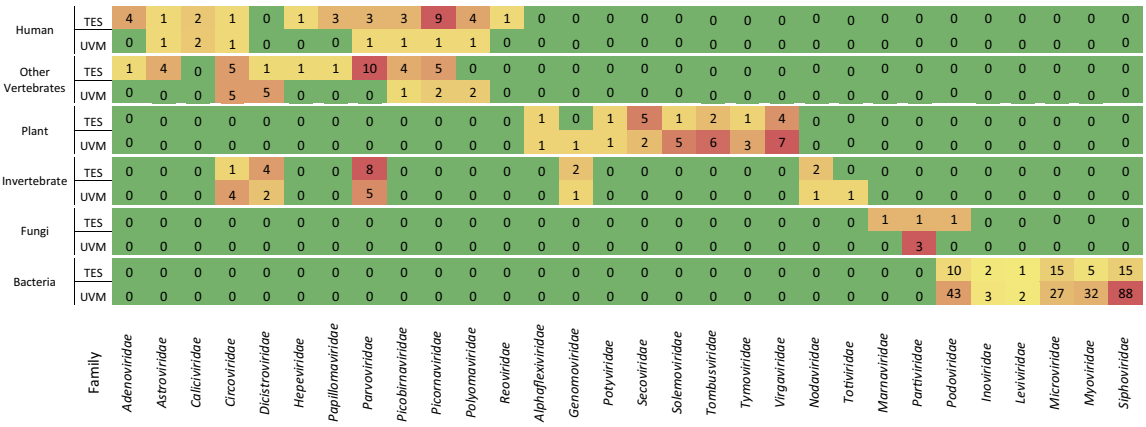


Fig. 1. Host distribution of viral species within families obtained by Target Enrichment Sequencing (TES) and Untargeted Viral Metagenomics (UVM) in WWTP A.

such as sewage, as well as being a technique that might be useful for environmental public health surveillance.

3.2. Application of Amplicon Deep Sequencing to the study of specific viral groups in sewage: human adenoviruses, papillomaviruses, and enteroviruses

3.2.1. Comparison between ADS and TES for the study of the sewage virome

Amplicon Deep Sequencing has demonstrated its usefulness in the study of specific viral groups, giving higher sensitivity and a large amount of information about the viruses present in a given sample. This methodology has previously been applied to environmental samples for studying diversity within viral groups, such as *Adenoviridae* (Fernandez-Cassi et al., 2018a, 2018b; Ogorzaly et al., 2015), *Papillomaviridae* (Hamza and Hamza, 2018; Iaconelli et al., 2015; La Rosa et al., 2015) or enterovirus (Brinkman et al., 2017).

The composite 24 h sewage sample collected from WWTP A was evaluated by ADS using specific PCR primers for HAdV, HPV and EV families. Results obtained after the application of ADS are summarised in Fig. 2.

While ADS showed similar adenovirus species to TES, detecting Human Mastadenovirus F and A species, with HAdV 41 and HAdV 18 being the most prevalent, some of the species appeared only with TES (e.g., HAdV5, HAdV 27 and HAdV 56) and others only with ADS (e.g., HAdV 18, HAdV 12 and HAdV 61).

Regarding HPV, ADS provided a higher diversity of species than TES, with HPV6 and HPV66 being the only alphapapillomavirus detected by ADS and the most abundant types, being 36.01% and 21.22% of the total sequences detected respectively. Other members of the betapapillomavirus genus were also detected, with HPV120 and HPV19 being the most prevalent types. An important difference between these two methodologies was observed: sequences of HPV 122 and HPV 49 (betapapillomavirus types) were obtained from both approaches, but HPV 17 (betapapillomavirus type) and HPV127 (gammapapillomavirus type) only by TES and a wider variety of types only by ADS, including oncogenic ones.

Regarding enterovirus, viral species Enterovirus A, B and C were detected by ADS as occurred when applying TES. *E*-E30 was the most prevalent enterovirus, followed by EV-C99 and EV-A119. Also, sequences from EV-A71 were obtained by ADS and TES, which were of interest because a clinical brainstem encephalitis outbreak was occurring in Catalonia during the sampling period. By applying TES, only one EV-A71 contig (884 bp) was obtained from WWTP A and, by ADS, 14 amplicons (301 bp) were obtained from both WWTPs.

3.2.2. Amplicon Deep Sequencing for characterisation of specific viral groups in sewage: human adenoviruses, papillomaviruses, and enteroviruses

The distribution of the three viral groups over one year in sewage was also evaluated by applying ADS in seasonally pooled samples from WWTP B and the results obtained are represented in Fig. 3. A wide variety of HAdV, HPV and EV sequences was obtained.

The in-depth analysis of HAdV showed more than 10 serotypes in the year of study (Fig. 3a). HAdV from *Human Mastadenovirus* A species, related to gastrointestinal, urinary and respiratory infections, and *Human Mastadenovirus* F, related to infantile gastroenteritis, were the groups most commonly detected over the whole year, with being HAdV 41 (Group F) and HAdDV 31 (Group A) being the most abundant types. HAdV 40 (Group F), and HAdV 12 (Group A) found throughout the year and HAdV 18 (Group A) in spring. Other *Human Mastadenoviruses* serotypes (HAdV 51, HAdV 59, HAdV 46, HAdV 19) were detected in lower proportions, as previously described in other studies (Fernandez-Cassi et al., 2018a, 2018b; Iaconelli et al., 2017).

Many HAdV produce infections that may be subclinical and could be excreted in faeces by healthy people over a long period. The prevalence of HAdV in sewage through the year is in accordance with the proposal to use adenovirus as an indicator of human faecal contamination (Pina et al., 1998).

For papillomaviruses, a large diversity of types was observed by ADS through the whole year of sampling (Fig. 3b). Sequences of the potentially oncogenic types HPV 6 and HPV 66 (genus alphapapillomavirus) were detected through the year, with higher proportions than other HPV in the spring and winter months. Members of this genus had been described previously in raw sewage (Iaconelli et al., 2015), but HPV 66, involved in vulvar cancer and classified as a group I carcinogen (Proceedings of the IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Biological Agents., 2009), has never been reported in environmental samples. More than 25 different cutaneous betapapillomaviruses were detected, with HPV 120, HPV 19, HPV 9, HPV 8, HPV 49 and HPV 80 being the most abundant during the year of observation.

Enterovirus ADS showed different distribution patterns through the studied year (Fig. 3c): CV-B5, *E*-E18 and E30 (EV species B) were the serotypes more frequently detected in spring and summer and E-E14 (EV species B) and EV-C99 (EV species C) in autumn and winter. These enterovirus species are traditionally related to aseptic meningitis cases and, with the exception of Enterovirus C99, had been also reported as etiological agents involved in Acute Flaccid Paralysis (AFP) cases by the Spanish Ministry of Health in their Annual Epidemiologic Surveillance

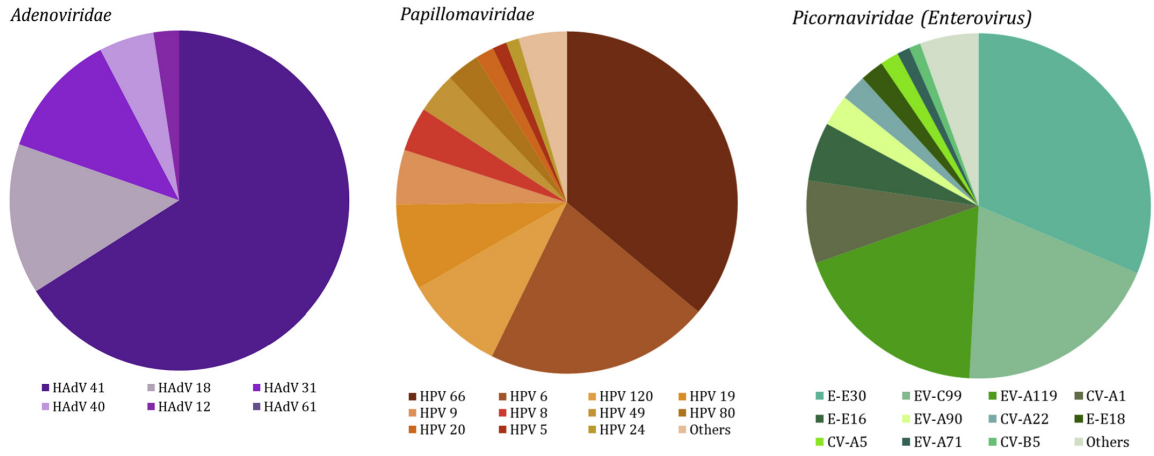
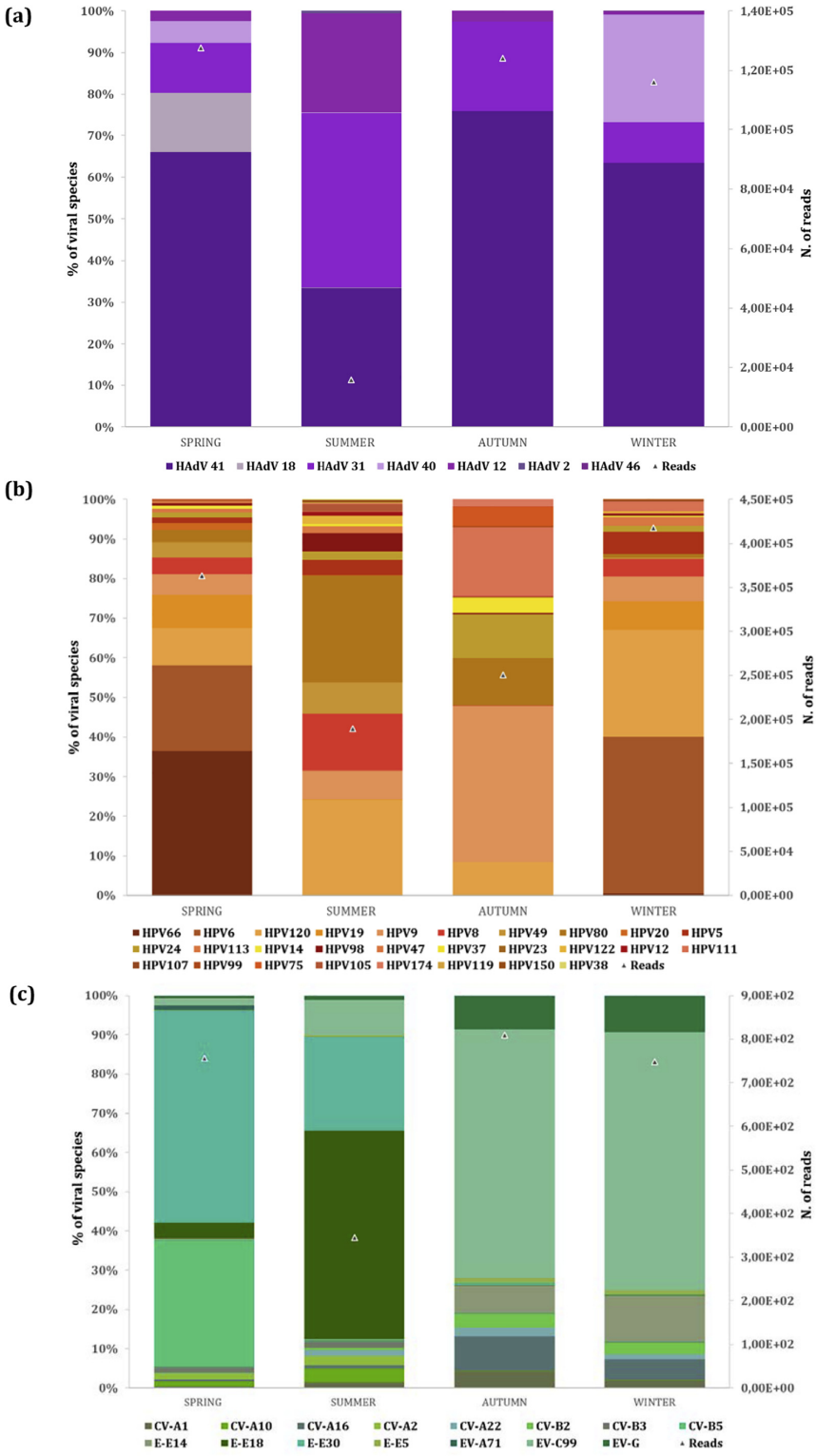


Fig. 2. Diversity of specific viral groups (human adenoviruses, papillomaviruses and enteroviruses) obtained by Amplicon Deep Sequencing in the composite raw sewage sample A.


(a)
% of viral species
N. of reads
SPRING
SUMMER
AUTUMN
WINTER
HAdV 41
HAdV 18
HAdV 31
HAdV 40
HAdV 12
HAdV 2
HAdV 46
Reads
(b)
% of viral species
N. of reads
SPRING
SUMMER
AUTUMN
WINTER
HPV66
HPV6
HPV120
HPV19
HPV9
HPV8
HPV49
HPV80
HPV20
HPV5
HPV24
HPV113
HPV14
HPV98
HPV47
HPV37
HPV23
HPV122
HPV12
HPV111
HPV107
HPV99
HPV75
HPV105
HPV174
HPV119
HPV150
HPV38
Reads
(c)
% of viral species
N. of reads
SPRING
SUMMER
AUTUMN
WINTER
CV-A1
CV-A10
CV-A16
CV-A2
CV-A22
CV-B2
CV-B3
CV-B5
E-E14
E-E18
E-E30
E-E5
EV-A71
EV-C99
EV-G
Reads

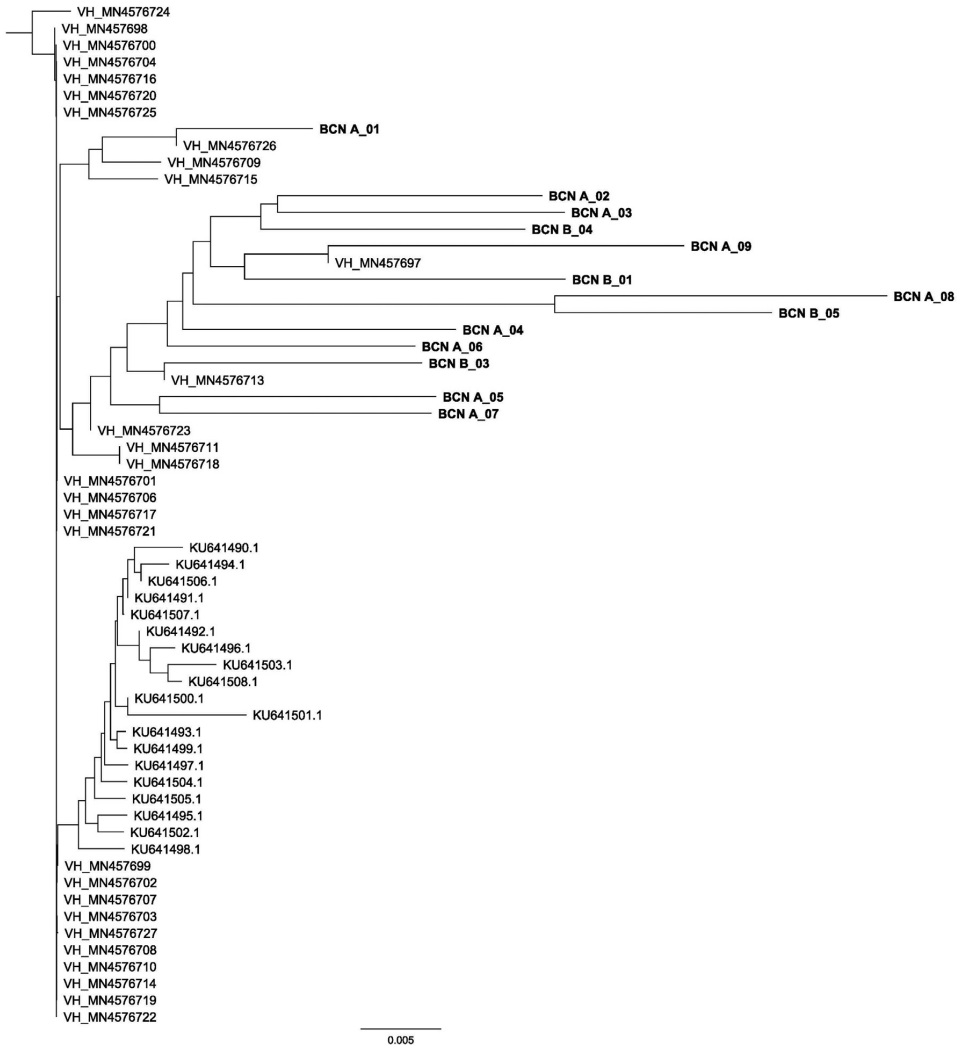


Fig. 4. Phylogenetic tree based on enterovirus VP1 nucleotide sequences of EV-A71 C1 strains identified by ADS. Represented in bold are the sequences obtained in both WWTPs (BCN_A and BCN_B). The rest of the strains were the onex obtained from clinical cases by Sanger sequencing from Hospital Vall d'Hebrón (VH) and the ones related to the Germany outbreak in 2015 (those with KU prefix; Böttcher et al., 2016). Scale bar indicate nucleotide substitution per site. CV-A16 (KT327162) has been used as outgroup, with a distance of 0.511 to BCN A_08.

(López-Perea et al., 2017). This emergent member of enterovirus species C has been reported to be involved in the development of AFP (Brown et al., 2009) and was recently described as being very prevalent in Uruguay and Brazil (Lizasoain et al., 2018; Luchs et al., 2019).

Also, a comparison of EV presence in spring in two different WWTPs (composite sample from WWTP A, which was collected in spring, and spring pooled samples from WWTP B) showed E-E30 to be the most prevalent strain detected in both plants while CV-B5 was the second most frequent strain in plant B and EV-C99 and EV-119 in plant A.

In fact, the comparison of the results obtained by analysing sewage samples collected in spring from both WWTPs by ADP resulted in a similar distribution of HAdV, HPV and EV, despite the locations being 30 km apart and treating sewage from different population equivalent amounts.

Further research should be directed towards analysing sewage during longer periods, in order to establish the most appropriate sampling design for use in the characterisation of the sewage virome from a determined population (composite, pooled or individual samples), as well as for elucidating if there exist any seasonality in the excretion of relevant pathogenic viruses.

3.3. Effect of the EV-A71 outbreak on the excreted virome in sewage

An outbreak of brainstem encephalitis occurred in the geographical area studied during the studied period affecting more than 100 children showing a peak in spring 2016 (Casas-Alba et al., 2017).

Sequences from EV-A71 were detected using ADS in both of the WWTP studied which belong to geographical areas located within 30Km

Fig. 3. Diversity of *Adenoviridae* (a), *Papillomaviridae* (b) and *Picornaviridae* (c) families in the seasonal pooled raw sewage samples from WWTP B obtained by Amplicon Deep Sequencing. Data presented as percentage of viral species and number of reads.

in the outbreak area. A total of 14 amplicons of 301 bp from VP1 region were subtyped as C1 with the Enterovirus Genotyping tool (Kroneman et al., 2011). These sequences showed a pairwise identity of 97% with sequences related to the German outbreak of 2015 (Böttcher et al., 2016) and with clinical sequences from patients involved in the outbreak, obtained at Hospital Vall d'Hebron in Barcelona (Andres et al., 2019).

A phylogenetic tree including all these sequences was constructed and represented in Fig. 4 and, even though the presence of this serotype was in a minority, these sequences proved to be phylogenetically close to those that caused the German (2015) and Catalan outbreaks (2016).

By applying TES, only one EV-A71 contig (884 bp) was obtained from WWTP A, but subtyping was not possible due to the short VP1 region present in this contig.

Thus, the application of ADS, in addition to TES, has also been useful to monitor an outbreak, detecting EV A71 sequences only when an encephalitis outbreak was occurring simultaneously with the sampling period in two different sewage samples from two different locations in Catalonia.

4. Conclusions

- Enterovirus A71 C1 was detected by Amplicon Deep Sequencing and Target Enrichment Sequencing during an encephalitis outbreak, although it was only a small percentage of the enterovirus excreted, being other serotypes much more abundant. Only Amplicon Deep Sequencing was useful for subtyping purposes.
- Sewage from two different WWTPs, collected as a unique composite and as monthly pooled samples in spring 2016, showed a similar distribution of HAdV, HPV and EV types, despite the locations being 30 km apart and treating sewage from different population equivalent amounts.
- Untargeted Viral Metagenomics is the only NGS technique that provides a complete picture of the whole virome present in sewage including vertebrate, invertebrate, bacteria, plant and fungi viruses.
- Target Enrichment NGS based on probe capture has proved a very successful strategy for the study of vertebrate viruses in sewage samples providing a higher number of detected families, a higher number of members within these families, more reads and larger genome coverage than conventional Untargeted Viral Metagenomics.
- Amplicon Deep Sequencing proved useful when observing the variety within a sole viral family is the objective of the study and, because is a very sensitive technique, it may be useful for the surveillance of specific pathogenic viruses (e.g: EV-A71).

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.136604.

Acknowledgements

We thank the Genomics Core Facility (UPF) for providing support in the sequencing experiments. We thank Eva Forés, Ayalkibet Hundesa, David Aguado and Aiora Aregita for their lab assistance and support. Graphical Abstract was created with Biorender.com.

Availability of data and material

The datasets generated during the current study are available in zenodo under the DOI number https://doi.org/10.5281/zenodo.3539112.

Funding

This work was supported through a grants funded by the Spanish Ministry of Economy and Competitiveness (MINECO) AGL2014-55081-R, AGL2017-86797-C2-1-R and JPI Water project METAWATER (4193-00001B). During the development of this study, Sandra Martínez-Puchol was a fellow of the MINECO. Sílvia Bofill-Mas is a Serra-Hunter fellow at the University of Barcelona.

Declaration of competing interest

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

References

Allen, H.K., Bunge, J., Foster, J.A., Bayles, D.O., Stanton, T.B., 2013. Estimation of viral richness from shotgun metagenomes using a frequency count approach. Microbiome 1, 5. https://doi.org/10.1186/2049-2618-1-5.

Andres, C., Vila, J., Gimferrer, L., Pinana, M., Esperalba, J., Codina, M.G., Barnes, M., Martin, M.C., Fuentes, F., Rubio, S., Alcubilla, P., Rodrigo, C., Pumarola, T., Anton, A., 2019. Surveillance of enteroviruses from paediatric patients attended at a tertiary hospital in Catalonia from 2014 to 2017. J. Clin. Virol. 110, 29–35. https://doi.org/10.1016/j.jcv.2018.11.004.

Bofill-Mas, S., Albinana-Gimenez, N., Clemente-Casares, P., Hundesa, A., Rodriguez-Manzano, J., Allard, A., Calvo, M., Girones, R., 2006. Quantification and stability of human adenoviruses and polyomavirus JCPyV in wastewater matrices. Appl. Environ. Microbiol. 72, 7894–7896. https://doi.org/10.1128/AEM.00965-06.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Böttcher, S., Obermeier, P.E., Neubauer, K., Diedrich, S., Hofmann, J., Heider, H., Reif, U., Korn, K., Rabenau, H.F., Wille, A., Baillot, A., Pettke, A., Kühn, J., Ackermann, N., Ramelow, C., Hahn, R., Enders, M., Adam, M., Michel, D., Schubert, A., Weißbrich, B., Prifert, C., 2016. Recombinant enterovirus A71 subgenogroup C1 strains, Germany, 2015. Emerg. Infect. Dis. 22, 1843–1846. https://doi.org/10.3201/eid2210.160357.

Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., OJ, J., Lipkin, W., 2015. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. MBio 6, 1–11. https://doi.org/10.1128/mBio.01491-15.Editor.

Brinkman, N.E., Fout, G.S., Keely, S.P., 2017. Retrospective surveillance of wastewater to examine seasonal dynamics of enterovirus infections. mSphere 2. https://doi.org/10.1128/msphere.00099-17.

Brown, B.A., Maher, K., Flemister, M.R., Naraghi-Arani, P., Uddin, M., Oberste, M.S., Pallansch, M.A., 2009. Resolving ambiguities in genetic typing of human enterovirus species C clinical isolates and identification of enterovirus 96, 99 and 102. J. Gen. Virol. 90, 1713–1723. https://doi.org/10.1099/vir.0.008540-0.

Cantalupo, P.G., Calgua, B., Zhao, G., Hundesa, A., Wier, A.D., Katz, J.P., Grabe, M., Hendrix, R.W., Girones, R., Wang, D., Pipas, J.M., 2011. Raw sewage harbors diverse viral populations. MBio 2, e00180-11. https://doi.org/10.1128/mBio.00180-11.

Casas-Alba, D., de Sevilla, M.F., Valero-Rello, A., Fortuny, C., García-García, J.J., Ortez, C., Muchart, J., Armangué, T., Jordan, I., Luaces, C., Barrabeig, I., González-Sanz, R., Cabrerizo, M., Muñoz-Almagro, C., Launes, C., 2017. Outbreak of brainstem encephalitis associated with enterovirus-A71 in Catalonia, Spain (2016): a clinical observational study in a children's reference Centre in Catalonia. Clin. Microbiol. Infect. https://doi.org/10.1016/j.cmi.2017.03.016.

Chalkias, S., Gorham, J.M., Mazaika, E., Parfenov, M., Dang, X., DePalma, S., McKean, D., Seidman, C.E., Seidman, J.G., Koralnik, I.J., 2018. ViroFind: a novel target-enrichment deep-sequencing platform reveals a complex JC virus population in the brain of PML patients. PLoS One 13, e0186945. https://doi.org/10.1371/journal.pone.0186945.

De Roda Husman, A.M., Walboomers, J.M.M., Van den Brule, A.J.C., Meijer, C.J.L.M., Snijders, P.J.F., 1995. The use of general primers GP5 and GP6 elongated at their 3′ ends with adjacent highly conserved sequences improves human papillomavirus detection by PCR. J. Gen. Virol. 76, 1057–1062. https://doi.org/10.1099/0022-1317-76-4-1057.

Del Campo, J.A., Parra-Sánchez, M., Figueruela, B., García-Rey, S., Quer, J., Gregori, J., Bernal, S., Grande, L., Palomares, J.C., Romero-Gómez, M., 2018. Hepatitis C virus deep sequencing for sub-genotype identification in mixed infections: a real-life experience. Int. J. Infect. Dis. 67, 114–117. https://doi.org/10.1016/j.ijid.2017.12.016.

Fernandez-Cassi, Xavier, Rusiñol, M., Martínez-Puchol, S., 2018a. Viral concentration and amplification from human serum samples prior to application of next-generation sequencing analysis. Methods in Molecular Biology, pp. 173–188. https://doi.org/10.1007/978-1-4939-8682-8_13.

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

Forslund, O., Ly, H., Higgins, G., 2003. Improved detection of cutaneous human papillomavirus DNA by single tube nested 'hanging droplet' PCR. J. Virol. Methods 110, 129–136. https://doi.org/10.1016/S0166-0934(03)00109-5.

Gimferrer, L., Campins, M., Codina, M.G., Esperalba, J., Martin, M.D.C., Fuentes, F., Pumarola, T., Anton, A., 2015. First Enterovirus D68 (EV-D68) cases detected in hospitalised patients in a tertiary care university hospital in Spain, October 2014. Enferm. Infecc. Microbiol. Clin. 33, 585–589. https://doi.org/10.1016/j.eimc.2015.01.008.

Hamza, H., Hamza, I.A., 2018. Oncogenic papillomavirus and polyomavirus in urban sewage in Egypt. Sci. Total Environ. 610–611, 1413–1420. https://doi.org/10.1016/j.scitotenv.2017.08.218.

Hata, A., Kitajima, M., Haramoto, E., Lee, S., Ihara, M., Gerba, C.P., Tanaka, H., 2018. Next-generation amplicon sequencing identifies genetically diverse human astroviruses, including recombinant strains, in environmental waters. Sci. Rep. 8, 11837. https://doi.org/10.1038/s41598-018-30217-y.

Hjelmsø, M.H., Hellmér, M., Fernandez-Cassi, X., Timoneda, N., Lukjancenko, O., Seidel, M., Elsässer, D., Aarestrup, F.M., Löfström, C., Bofill-Mas, S., Abril, J.F., Girones, R., Schultz, A.C., 2017. Evaluation of methods for the concentration and extraction of viruses from sewage in the context of metagenomic sequencing. PLoS One 12, e0170199. https://doi.org/10.1371/journal.pone.0170199.

Hjelmsø, M.H., Mollerup, S., Jensen, R.H., Pietroni, C., Lukjancenko, O., Schultz, A.C., Aarestrup, F.M., Hansen, A.J., 2019. Metagenomic analysis of viruses in toilet waste from long distance flights—A new procedure for global infectious disease surveillance. PLoS One 14, e0210368. https://doi.org/10.1371/journal.pone.0210368.

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT suite: a web server for clustering and comparing biological sequences. Bioinformatics 26, 680–682. https://doi.org/10.1093/bioinformatics/btq003.

Huang, B., Jennsion, A., Whiley, D., McMahon, J., Hewitson, G., Graham, R., De Jong, A., Warrilow, D., 2019. Illumina sequencing of clinical samples for virus detection in a public health laboratory. Sci. Rep. 9, 5409. https://doi.org/10.1038/s41598-019-41830-w.

Iaconelli, M., Petricca, S., Libera, S. Della, Di Bonito, P., La Rosa, G., 2015. First detection of human papillomaviruses and human polyomaviruses in river waters in Italy. Food Environ. Virol. https://doi.org/10.1007/s12560-015-9203-7.

Iaconelli, M., Valdazo-Gonz Alez, B., Equestre, M., Ciccaglione, A.R., Marcantonio, C., Libera, S., Della, Rosa, La, G., 2017. Molecular Characterization of Human Adenoviruses in Urban Wastewaters Using Next Generation and Sanger Sequencing. https://doi.org/10.1016/j.watres.2017.05.039.

Kohl, C., Brinkmann, A., Dabrowski, P.W., Radonic, A., Nitsche, A., Kurth, A., 2015. Protocol for metagenomic virus detection in clinical specimens. Emerg. Infect. Dis. 21, 48–57. https://doi.org/10.3201/eid2101.140766.

Krishnamurthy, S.R., Wang, D., 2017. Origins and challenges of viral dark matter. Virus Res. 239, 136–142. https://doi.org/10.1016/j.virusres.2017.02.002.

Kroneman, A., Vennema, H., Deforche, K., v d Avoort, H., Penaranda, S., Oberste, M.S., Vinje, J., Koopmans, M., 2011. An automated genotyping tool for enteroviruses and noroviruses. J. Clin. Virol. 51, 121–125. https://doi.org/10.1016/j.jcv.2011.03.006.

Kumar, A., Murthy, S., Kapoor, A., 2017. Evolution of selective-sequencing approaches for virus discovery and virome analysis. Virus Res. 239, 172–179. https://doi.org/10.1016/j.virusres.2017.06.005.

Kustin, T., Ling, G., Sharabi, S., Ram, D., Friedman, N., Zuckerman, N., Bucris, E.D., Glatman-Freedman, A., Stern, A., Mandelboim, M., 2019. A method to identify respiratory virus infections in clinical samples using next-generation sequencing. Sci. Rep. 9, 2606. https://doi.org/10.1038/s41598-018-37483-w.

La Rosa, G., Della Libera, S., Petricca, S., Iaconelli, M., Briancesco, R., Paradiso, R., Semproni, M., Di Bonito, P., Bonadonna, L., 2015. First detection of papillomaviruses and polyomaviruses in swimming pool waters: unrecognized recreational water-related pathogens? J. Appl. Microbiol., 1–9 https://doi.org/10.1111/jam.12925.

Lizasoain, A., Burlandy, F.M., Victoria, M., Tort, L.F.L., da Silva, E.E., Colina, R., 2018. An environmental surveillance in Uruguay reveals the presence of highly divergent types of human enterovirus species C and a high frequency of species a and B types. Food Environ. Virol. 10, 343–352. https://doi.org/10.1007/s12560-018-9351-7.

López-Perea, N., Masa-Calles, J., Torres de Mier, M., Fernandez-Martinez, B., Cabrerizo, M., 2017. Vigilancia de la Parálisis Flácida Aguda y Vigilancia de Enterovirus, 2016 Plan de Acción en España para la erradicación de la poliomelitis. https://doi.org/10.13140/RG.2.2.35292.90240.

Luchs, A., Leal, E., Tardy, K., Milagres, F.A. de P., Komninakis, S.V., Brustulin, R., Teles, M. da A.R., Lobato, M.C.A.B.S., das Chagas, R.T., Abrão, M. de F.N.D.S., Soares, C.V. de D.A., Deng, X., Delwart, E., Sabino, E.C., da Costa, A.C., 2019. The rare enterovirus c99 and echovirus 29 strains in Brazil: potential risks associated to silent circulation. Mem. Inst. Oswaldo Cruz 114, e190160. https://doi.org/10.1590/0074-02760190160.

Mühlemann, B., Jones, T.C., Damgaard, P. de B., Allentoft, M.E., Shevnina, I., Logvin, A., Usmanova, E., Panyushkina, I.P., Boldgiv, B., Bazartseren, T., Tashbaeva, K., Merz, V., Lau, N., Smrčka, V., Voyakin, D., Kitov, E., Epimakhov, A., Pokutta, D., Vicze, M., Price, T.D., Moiseyev, V., Hansen, A.J., Orlando, L., Rasmussen, S., Sikora, M., Vinner, L., Osterhaus, A.D.M.E., Smith, D.J., Glebe, D., Fouchier, R.A.M., Drosten, C., Sjögren, K.-G., Kristiansen, K., Willerslev, E., 2018. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature 557, 418–423. https://doi.org/10.1038/s41586-018-0097-z.

Ng, T.F.F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., Oderinde, B.S., Wommack, K.E., Delwart, E., 2012. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J. Virol. 86, 12161–12175. https://doi.org/10.1128/JVI.00869-12.

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. MetaSPAdes: a new versatile metagenomic assembler. Genome Res. 27, 824–834. https://doi.org/10.1101/gr.213959.116.

Ogorzaly, L., Walczak, C., Galloux, M., Etienne, S., Gassilloud, B., Cauchie, H.-M., 2015. Human adenovirus diversity in water samples using a next-generation amplicon sequencing approach. Food Environ. Virol. 7, 112–121. https://doi.org/10.1007/s12560-015-9194-4.

Parras-Moltó, M., Rodríguez-Galet, A., Suárez-Rodríguez, P., López-Bueno, A., 2018. Evaluation of bias induced by viral enrichment and random amplification protocols in metagenomic surveys of saliva DNA viruses. Microbiome 6, 1–18. https://doi.org/10.1186/s40168-018-0507-3.

Paskey, A.C., Frey, K.G., Schroth, G., Gross, S., Hamilton, T., Bishop-Lilly, K.A., 2019. Enrichment post-library preparation enhances the sensitivity of high-throughput sequencing-based detection and characterization of viruses from complex samples. BMC Genomics 20, 155. https://doi.org/10.1186/s12864-019-5543-2.

Pina, S., Puig, M., Lucena, F., Jofre, J., Girones, R., 1998. Viral pollution in the environment and in shellfish: human adenovirus detection by PCR as an index of human viruses. Appl. Environ. Microbiol. 64, 3376–3382.

Prevost, B., Lucas, F.S., Ambert-Balay, K., Pothier, P., Moulin, L., Wurtzer, S., 2015. Deciphering the diversities of astroviruses and noroviruses in wastewater treatment plant effluents by a high-throughput sequencing method. Appl. Environ. Microbiol. 81, 7215–7222. https://doi.org/10.1128/AEM.02076-15.

Proceedings of the IARC Working Group on the Evaluation of Carcino- genic Risks to Humans. Biological Agents, 2009. Proceedings of the IARC working group on the evaluation of carcinogenic risks to humans. Biological agents. IARC Monogr. Eval. Carcinog. Risks to Humans 100B.

Santiago-Rodriguez, T.M., Hollister, E.B., 2019. Human virome and disease: high-throughput sequencing for virus discovery, identification of phage-bacteria dysbiosis and development of therapeutic approaches with emphasis on the human gut. Viruses https://doi.org/10.3390/v11070656.

Tamaki, H., Zhang, R., Angly, F.E., Nakamura, S., Hong, P.-Y., Yasunaga, T., Kamagata, Y., Liu, W.-T., 2012. Metagenomic analysis of DNA viruses in a wastewater treatment plant in tropical climate. Environ. Microbiol. 14, 441–452. https://doi.org/10.1111/j.1462-2920.2011.02630.x.

Vilsker, M., Moosa, Y., Nooij, S., Fonseca, V., Ghysens, Y., Dumon, K., Pauwels, R., Alcantara, L.C., Vanden Eynden, E., Vandamme, A.-M., Deforche, K., de Oliveira, T., 2018a. Genome detective: an automated system for virus identification from high-throughput sequencing data. Bioinformatics https://doi.org/10.1093/bioinformatics/bty695.

Vilsker, M., Moosa, Y., Nooij, S., Fonseca, V., Ghysens, Y., Dumon, K., Pauwels, R., Alcantara, L.C., Vanden Eynden, E., Vandamme, A.-M., Deforche, K., de Oliveira, T., 2018b. Genome detective: an automated system for virus identification from high-throughput sequencing data. Bioinformatics, 1–3 https://doi.org/10.1093/bioinformatics/bty695.

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Geer, L.Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Ostell, J., Miller, V., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R.L., Tatusova, T.A., Wagner, L., Yaschenko, E., 2007. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 35, D5–D12. https://doi.org/10.1093/nar/gkl1031.

WHO Regional Office for Europe, 2015. Enterovirus surveillance guidelines. Guidel. enterovirus Surveill. Support Polio Erad. Initiat.

Wylie, T.N., Wylie, K.M., Herter, B.N., Storch, G.A., 2015. Enhanced virome sequencing using targeted sequence capture. Genome Res. 25, 1910–1920. https://doi.org/10.1101/gr.191049.115.

Article 2: SARS-CoV-2 és un nou membre del viroma de l'aigua residual urbana

"SARS-CoV-2 is a new member of urban wastewater virome."

Martínez-Puchol S., Itarte M., Rusiñol M., Forés E., Andrés C., Antón A., Quer J., Girones R., Abril J.F., Bofill-Mas S.

Manuscrit en procés de revisió a la revista *Genes*.

Arrel de la pandèmia de la COVID-19, les tècniques de seqüenciació massiva han esdevingut eines encara més importants per a la caracterització genètica del aïllats clínics del SARS-CoV-2. A part de la recerca enfocada en pacients, l'epidemiologia basada en aigües residuals ha demostrat ser una útil per a complementar els estudis clínics, servint també com un mecanisme d'alerta a l'aparició d'un brot. La utilització de la seqüenciació massiva en mostres d'aigua residual pot ser clau per un estudi a fons del viroma excretat ja que pot estar dirigit a l'estudi de la circulació i caracterització del SARS-CoV-2 i també a la detecció d'altres virus amb potencial pandèmic.

L'objectiu del present estudi ha estat l'aplicació de dues tècniques de seqüenciació massiva, amb i sense enriquiment de dianes, per a la seqüenciació de dues mostres d'aigua residual en dos moments diferents de la pandèmia, Març i Juliol 2020.

S'ha obtingut el viroma complet de les mostres analitzades, trobant membres de la família *Coronaviridae* únicament en la mostra del mes de Març i amb l'enriquiment de dianes. Un contig va ser identificat com a HCoV-OC43 i vuit contigs (corresponent a 2,03 kb del genoma) com a SARS-CoV-2. Aquest darrers van ser comparats amb els obtinguts de mostres clíniques aïllades durant el mateix període i seqüenciades amb el

mateix protocol d'enriquiment, observant 3 canvis nucleotídics en la regió de la nucleocàpsida. A més, tres coronavirus animals van ser detectats al aplicar la seqüenciació amb enriquiment de dianes.

Tot això demostra que existeix una circulació simultània de coronavirus humans i coronavirus que probablement infecten animals domèstics i rosegadors en àrees urbanes. Els protocols de seqüenciació massiva amb enriquiment de dianes són clau per a la descripció de l'ocurrència i de les característiques genètiques de membres de la família *Coronaviridae* que formen part del viroma excretat en l'aigua residual.

genes

MDPI

Article

SARS-CoV-2 is a new member of urban wastewater virome

Sandra Martínez-Puchol[1,2,*], Marta Itarte[1,2], Marta Rusiñol[3], Eva Forés[1,2], Cristina Andrés[4], Andrés Antón[4], Josep Quer[5], Rosina Girones[1,2], Josep F Abril[6] and Sílvia Bofill-Mas[1,2]

[1] Laboratory of Viruses Contaminants of Water and Food; Genetics, Microbiology & Statistics Dept.; Universitat de Barcelona. Barcelona, Catalonia, Spain.
[2] The Water Research Institute (IdRA), Universitat de Barcelona. Barcelona, Catalonia, Spain.
[3] Institute of Environmental Assessment & Water Research (IDAEA), CSIC. Barcelona, Catalonia, Spain.
[4] Respiratory Viruses Unit, Virology Section, Microbiology Department, Hospital Universitari Vall d'Hebron, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona. Barcelona, Catalonia, Spain
[5] Liver Unit, Liver Diseases - Viral Hepatitis, Vall d'Hebron Institut de Recerca (VHIR), Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain
[6] Computational Genomics Lab; Genetics, Microbiology & Statistics Dept.; Universitat de Barcelona; Institut de Biomedicina (IBUB); Barcelona, Catalonia, Spain.

* Correspondence: smartinezpuchol@ub.edu

Abstract: In the wake of COVID-19 pandemics, NGS applications have raised as important tools for the genetic characterization of SARS-CoV-2 from clinical isolates and wastewater-based epidemiology for early warning outbreak containment. The use of NGS in wastewater samples could be the key for an in-depth study of the excreted virome, not only focusing in SARS-CoV-2 circulation and typing, but also to detect other potentially pandemic viruses within the same family. With this aim, two 24-hours composite wastewater samples from March and July 2020 were sequenced by applying specific viral NGS as well as target enrichment NGS. The full virome of the analysed samples was obtained, with *Coronaviridae* members (CoV) in one of the analysed samples after applying the enrichment. One contig was identified as HCoV-OC43 and 8 contigs as SARS-CoV-2. These contigs were compared with those obtained from clinical specimens by applying the same target enrichment approach. Animal CoVs were also detected when applying this technique. The results showed that there is a co-circulation in urban areas of human and animal coronaviruses infecting domestic animals and rodents. NGS enrichment-based protocols might be crucial to describe the occurrence and genetic characteristics of *Coronaviridae* family members within the excreted virome present in wastewater.

Keywords: next generation sequencing; viral metagenomics; coronavirus; sewage; target enrichment sequencing

1. Introduction

SARS-CoV-2 was identified in China at the end of 2019 and has become the first pandemic coronavirus (CoV) [1]. Upon confirmation that COVID-19 patients shed SARS-CoV-2 in faeces, different studies provided significant correlation between the concentration of SARS-CoV-2 in wastewater and the prevalence of COVID-19 in the served population, increasing the evidence that wastewater is a good indicator of the prevalence of the excreted virus in a population (wastewater-based epidemiology, WBE). SARS-CoV-2 has been recently reported to be excreted for up to 55% of COVID-19 patients suggesting the possibility of long viral shedding in faeces, for nearly 5 weeks after the patients' respiratory samples tested negative for SARS-CoV-2 RNA [2]. The concentration of SARS-CoV-2 excreted from infected patients has been estimated to be of 1E+02 to 1E+07 copies per gram of faeces [3].

It is then presumable that, in a pandemic context, SARS-CoV-2 could become a member of sewage virome where viruses commonly present in human excreta have been described to occur [4,5]. Previous studies described useful procedures to describe sewage virome and compare high throughput viral metagenomics to target enrichment and amplicon deep sequencing approaches [6]. In this issue, we applied a probe-capture target enrichment NGS for describing sewage virome from a 2 wastewater 24-hours composite samples collected in March and July 2020 from a wastewater treatment plant located in the city of Barcelona, when COVID-19 incidence was of approximately 258 cases and 175 cases/100,000 inhabitants respectively. The results obtained were compared with those obtained from clinical samples by applying the same enrichment approach.

2. Materials and Methods

Two urban 24-hours composite sewage samples were collected in March 19th and July 14th 2020 from a wastewater treatment plant (WWTP) located in the city of Barcelona that serves up to 2.8 million population equivalents and receives domestic and industrial waste from the sewer system. Samples were collected and kept at 4ºC in a sterile container until viral particles from 70 ml of sewage were concentrated by centrifugal ultrafiltration. After a debris removal (15 min, 4500 g), the sample was ultrafiltered with a Centricon® Plus-70 device (30KDa) following the manufacturer instructions, obtaining a viral concentrate of 200 µl. Additionally, two SARS-CoV-2-positive naso/oropharyngeal swabs were obtained on March 15th from a male and on March 24th from a female patients attended at Hospital Universitari Vall d'Hebron (HUVH) (consensus sequences by an independent Illumina TruSeq high coverage sequencing run are already available at GISAID as EPI_ISL_418860 and EPI_ISL_418861, respectively).

Nucleic acid extraction (NA) was performed as described previously [7]. The concentration of SARS-CoV-2 RNA in wastewater samples was measured by (RT)-qPCR of two viral targets in nucleocapsid phosphoprotein (N1 and N2 region) and clinical samples in HUVH by commercial real-time multiplex RT-PCR (Allplex™ 2019-nCoV Assay, Seegene, South Korea) [8]. EURM-019 single stranded RNA (ssRNA) fragments of SARS-CoV-2 (Joint Research Centre, EC, https://crm.jrc.ec.europa.eu/p/EURM-019) was used to construct the standard for quantitation. JC polyomavirus (JCPyV) was quantified in the samples as an indicator of human faecal viral contamination as previously described (Bofill-Mas et al., 2006).

Previous to library preparation, NA were retrotranscribed to cDNA, tagged and complemented to obtain dsDNA. This viral randomly tagged dsDNA was then amplified (25 cycles) to obtain the sufficient amount of DNA for the preparation of libraries. Libraries were prepared in duplicate following the instructions provided by the manufacturer (Roche-Kappa Biosystems). Briefly, dsDNA was fragmented, indexed, amplified and the quality of the fragments and the concentration measured by Qubit 3.0 Fluorometer (Thermo Fisher Scientific). Then, one replicate of the libraries was

hybridized with probes designed to capture sequences from vertebrate viral pathogens (VirCapSeq Enrichment Kit, Roche). After the capture, quality and concentration were re-checked and sequencing from the captured and non-captured libraries was performed (Illumina Miseq 2x300bp). The sequencing raw data obtained was analysed with Genome Detective Virus Tool [9] and the contigs obtained with a nucleotide identity ≥70% were further processed with Geneious (v11.1.5; https://www.geneious.com). Simultaneously, raw reads from wastewater and clinical samples were cleaned of technical sequences and trimmed by quality using Trimmomatic (v0.38; [10]) in order to remove low quality segments and Illumina adapters (min Phred score=Q20 on 4 bp window, min read length=30bp, leading/trailing clip=15bp, max mismatch count=2, palindrome clip threshold=30, and simple clip threshold=10). When one of the reads, either R1 or R2, was discarded, the remaining one was collected into a single-ended (SG) reads file to use along with the resulting filtered PE reads later on samtools (v1.9; [11]) and bamtools (v2.5.1; [12]) sets of commands were used to process, sort, and index those alignments made by bowtie2 (v2.3.4.3, 64bit; [13]; with parameters k=5, L=12, and "sensitive-local" switch), mapping the PE/SG reads against the SARS-CoV-2 reference genome (GenBank entry: NC_045512.2) and ensuring that the stored alignments were position sorted on the final bam files. Trinity (r20190503git; [14,15]; min contig length=100, k-mer size 31) produced contigs that later on were mapped over the reference genome obtaining the final scaffolds, then manually curated to finish the sequences. Due to the low number of reads recovered from wastewater sample, the corresponding assembly was filled with 93% of Ns to place the contigs into the assembled scaffold (this sequence is provided as Supplementary File 2). mafft (v7.407; [16]; with localpair switch on) was chosen to calculate the multiple sequence alignment shown on Figure 3, comparing the wastewater scaffold against a randomly sampled set of 600 Spanish and 600 international sequences from GISAID database (available on August, 4th, 2020; [17]; [18]; see also acknowledgements).

3. Results and Discussion

3.1. SARS-CoV-2 quantification in clinical and wastewater samples

The viral N1 and N2 copies were quantified in a composite raw sewage sample collected in Barcelona in March 2020, one week after the state of alarm declaration in the country due to the COVID-19 pandemics, and in a composite sewage collected in July 2020 just before the beginning of the second COVID-19 peak. The samples were kept at -80°C until their analysis. The (RT)-qPCR were performed in quadruplicate with the direct NA extractions and in quadruplicate with 1/10 dilution of NA extractions, in order to avoid potential inhibitory effects. Mean concentration values obtained were 3.92E3 CG/ml for N1 and 2.71E3 CG/ml for N2 in March 2020 and 3.80E1 CG/ml for N1 and 1.25E1 CG/ml for N2 in July 2020. The concentrations for the human faecal indicator JCPyV were similar in both samples, with 1.99E2 CG/ml and 8.73E2 CG/ml in March and July respectively [19]. Enzymatic inhibition was not observed. About the two SARS-CoV-2 laboratory-confirmed clinical samples from HUVH, the *E* gene cycle-threshold values were 15.5 and 16.4 from EPI_ISL_418860 and EPI_ISL_418861, respectively.

3.2. Urban wastewater virome during COVID-19 pandemics

The virome obtained from March, by using target enrichment, resulted in almost 1 million reads belonging to 27 viral families. The distribution of the number of reads obtained for each vertebrate viral family is shown in Figure 1. As expected, a wide variety of viral families infecting vertebrates was observed when applying the probed-based capture methodology. This approach was successfully applied in the past to improve the deep sequencing in human-focused virome studies [6,20,21], a key point in the study of viral species that are present in a low concentration in environmental and clinical samples. Traditional virome studies have shown the full picture of the viral sequences present in a sample, being bacteriophages and plant viruses the most abundant ones in sewage [4,7], but depending on the sequencing platform used or the number of samples multiplexed, the total number of reads obtained could be reduced and relevant human and animal

138 viral reads could be undetected. These enrichment methods could overcome these limitations
139 enabling the possibility of viral discovery.

140 In this study we used the VirCapSeq Enrichment Kit (Roche) intended to capture vertebrate viruses
141 from complex samples prior to metagenomic sequencing. It employs approximately 2 million
142 biotinylated oligonucleotide probes designed to bind to the coding sequences of all viral taxa known
143 to infect vertebrates at intervals of 50–100 nt. Libraries prepared from random primed cDNA are
144 hybridized with the biotinylated probes and trapped with streptavidin magnetic beads. After
145 magnetic capture and washing, NA are released from the beads and subjected to post-hybridization
146 PCR prior to sequencing. VirCapSeq platform was initially described in 2015 by Briese and co-
147 workers (Briese et al., 2015) before SARS-CoV-2 emergence; it is based on 1,309,451 fragments derived
148 from 256,679 viral genomic sequences. In silico studies performed in this study showed that although
149 being designed before SARS-CoV-2 emergence, the kit contained 21,414 fragments from 1,838
150 sequences described for 346 *Coronaviridae* species. NCBI-BLASTn of the full sequences and the probes
151 was run, with default parameters but e-value=10e-25, against SARS-CoV-2 reference genome (RefSeq
152 ID: NC_045512; [1]). Among the 1,838 full targets, 277 were found on 456 alignment hits, and 29 of
153 those had hits above 90% identity; on the other hand, 104 probe fragments from 28 different
154 sequences returned 104 hits. The most abundant families obtained using target enrichment were
155 *Astroviridae*, *Picornaviridae* and *Parvoviridae*. A complete list of all the identified viral species and
156 families is presented in the Supplementary File 1.

157 *Coronaviridae* sequences were also retrieved (Table 1). One contig of a Betacoronavirus 1, typed as
158 HCoV-OC43, was mapped with a high identity to the CoV 2'-O-methyltransferase, an enzyme that
159 enables the mRNA cap formation, essential for viral RNA stability [22]. HCoV-OC43 is one of the
160 four seasonal HCoVs that are known to cause common cold in humans with winter seasonality [23],
161 and is known to have rodents as natural hosts and bovines as intermediate ones [24].

162 **Table 1.** *Coronaviridae* family contigs description after performing Probe-capture targeted NGS in
163 one urban wastewater sample (March 2020).

***Coronaviridae* species**	**Host**	**Contigs**	**Length (bp)**	**Region**	**Nt ID (%)**	**AA ID (%)**
SARS-CoV-2	Human	8	203	ORF1ab polyprotein	100	100
			217	ORF1ab polyprotein	100	100
			126	ORF1ab polyprotein	100	100
			126	ORF1ab polyprotein	100	100
			178	ORF1ab polyprotein	100	100
			266	ORF3 protein	100	100
			516	Nucleocapsid protein	99.6	99.1
			399	Nucleocapsid protein	99.8	99.3
Betacoronavirus 1 (HCoV OC43)	Human	1	235	2'-O-methyltransferase	99.5	98.7
Feline coronavirus	Other vertebrates	1	524	ORF1ab polyprotein	91.8	86.5
Lucheng Rn rat coronavirus	Other vertebrates	5	354	ORF1ab polyprotein	97.4	97.1
			416	ORF1ab polyprotein	97.0	99.3
			542	ORF1ab polyprotein	97.0	99.4
			848	ORF1ab polyprotein	96.6	96.6

			1129	ORF1ab polyprotein	97.3	99.4
Canine coronavirus	Other vertebrates	1	231	ORF1ab polyprotein	96.1	94.8

Eight SARS-CoV-2 contigs were obtained, with a total of 2.03 Kb representing 6.8% of the genome. Five contigs corresponded to the ORF1ab polyprotein region and one contig to the ORF3 protein, all of them with 100% of nucleotide and aminoacidic identity with the genome Reference Sequence NC_045512. Additionaly, two contigs, with sizes 516 bp and 316 bp, mapped against the nucleocapsid protein, separated with a gap between them of 231 bp, and presented a nucleotide identity of 99.6% and 99.8% to the reference sequence.

Regarding other members of the *Coronaviridae* family, one feline CoV contig (524 bp) matching the replicase ORF1ab polyprotein region was obtained in the analysis of sequences from the March wastewater sample. This virus causes asymptomatic persistent enteric infections in a high percentage of household and catteries' cats [25]. Even though their high prevalence in their hosts, it is known that their survival is low in sewage [26]. Additionally, one contig from the same region (231 bp) was typed as canine CoV, known for being responsible of mild or moderate enteritis on dogs of all breeds and ages and with a high establishment in the environment [27]. Finally, five contigs of Lucheng Rn rat CoV, comprising 3.28 Kb of the ORF1ab polyprotein, were obtained. This Alphacoronavirus was firstly described on 2014 in rat samples from China [28], when rodents were not considered to be a relevant CoV reservoir. It is now clear that this group of animals are important reservoirs for alphacoronaviruses and also betacoronaviruses [29].

Regarding the wastewater sample from July, it retrieved a virome composition similar to sample from March (data not shown) but no members of the *Coronaviridae* family were detected. The absence of SARS-CoV-2 could be due to the fact that COVID-19 incidence recorded when this sample was collected was lower, 175.2 cases/100,000 inhabitants, with quantifications of SARS-CoV-2 in the sample two logarithms below the values obtained in March. The lack of presence of HCoV-OC43 and animal CoV reads in July sample could be explained by the winter seasonality of the human strain and the low amount of reads of these viruses compared with other members of the virome, suggesting that the presence of these viral RNA in the samples is close to the limit of detection.

For instance, the presence of new CoV members in bats has been a focus of study since they could be the origin of important outbreaks. In this field, the use of metagenomics is important for a better understanding of evolution, epidemiology, and host-relationships of zoonotic and human viruses [30]. These approaches made possible the discovery of new alphacoronaviruses [31,32], and recently the description of a betacoronavirus closely related to SARS-CoV-2 in Rhinolophus bats from China [33].

Moreover, the low sensitivity or the lack of sequencing depth in samples with a high viral load has revealed enrichment NGS as a successful strategy for CoV surveillance and discovery in Asia [34,35]. Thus, NGS for CoV discovery should ideally be wide enough to detect unknown viruses but targeting viruses belonging to the family of interest among many other present in the analysed samples.

3.3. Comparison of wastewater strains vs. clinical strains

SARS-CoV-2 contigs retrieved after applying targeted NGS to March wastewater sample were compared with sequences obtained by applying VirCapSeq enrichment approach to two clinical samples isolated in the same period of time. A short summary of the sequencing results for the three samples is provided on Table 2; see Figure 2 for an overview of the coverage distribution of the reads mapped over reference genome.

205 **Table 2.** Summary of sequencing results for the three samples considered on this work.

Sample	Wastewater	Clinical_A	Clinical_B
Raw PE Reads (x2)	5,901,370	6,021,795	4,663,999
Raw PE Total bp	1,850,065,979	3,015,620,573	2,353,498,179
Clean PE Reads (x2)	5,006,516	375,769	316,398
Clean PE Total bp	1,399,002,141	74,651,907	66,525,983
Clean SG Reads	823,018	5,535,930	4,238,097
Clean SG Total bp	139,990,580	441,056,111	337,855,561

206 From a total of 5,006,516 paired-end reads, only 8 aligned uniquely at a single location and 20 aligned
207 more than one time over the reference genome for the wastewater sample. On the contrary,
208 VirCapSeq clinical samples starting with less clean paired-end reads, 375,769 and 316,398 for samples
209 A and B, ended up with 1,042 and 1,856 uniquely mapped reads, respectively. The final scaffolds
210 assembled from those sets of reads recovered up to 6.79% for wastewater sample, 88.38% for clinical
211 sample A, and 89.83% for clinical sample B, of the 29,903bp of reference SARS-CoV-2 genome
212 sequence (Abril et al, manuscript in preparation).

213 Other studies have reported the use of VirCapSeq and Twist Bioscience panels for sequencing of
214 clinical samples, determining the importance of having low qPCR Ct values in a given sample for
215 obtaining a high genome coverage after conducting NGS [36,37].

216 SARS-CoV-2 sequences obtained from sewage were compared with other 1,200 sequences obtained
217 from clinical isolates available at GISAID database (https://www.gisaid.org/epiflu-applications/hcov-
218 19-genomic-epidemiology/ [17]) to elucidate the presence of viral variants. Despite that high
219 similarity was found for the assembled contigs, 3 nucleotide mismatches were found within the
220 nucleocapsid protein region (Figure 3). These divergences could represent single nucleotide changes
221 (SNV), but they were only supported by 3 overlapping reads. More read coverage would be needed
222 in this region to further assess these differences, although it is worth to note that those changes were
223 not described in clinical samples at the moment.

224 **4. Conclusions**

225 The application of a targeted NGS has provided nearly the whole genome of SARS-CoV-2 from two
226 clinical samples and eight SARS-CoV-2 contigs from one wastewater sample, both type of samples
227 obtained in March 2020 from the same geographical area. Because of low genome coverage obtained
228 in the wastewater sample, the 3 nucleotide differences observed among environmental and clinical
229 samples could not be further assessed. No contigs were obtained from the wastewater sample from
230 July. The results of the excreted virome in wastewater showed that there is a co-circulation in urban
231 areas of human and animal coronaviruses infecting domestic animals and rodents.

232 Amplicon sequencing panels covering the totality of the genome are available now for studying
233 SARS-CoV-2 genetic variants. These approaches, as well as specific SARS-CoV-2 and Human CoV
234 capture panels, would have enabled the obtention of sufficient genome coverage and consequently a
235 better SNV typing [38,39] than other targeted approaches directed towards wider groups of viruses,
236 like VirCapSeq capture used in this study. However, the use of this panel designed to cover vertebrate
237 viruses, allowed the obtention of SARS-CoV-2 sequences as part of sewage virome in a COVID-19

pandemic context. Also, it retrieved other relevant human and animal CoV sequences which could shed light on the co-circulation of different CoV strains in a determined population.

Different targeted NGS approaches have been applied for providing information on SARS-CoV-2 and other CoV circulation and genetic characteristics. Specific SARS-CoV-2/human CoV panels should be used for studying SARS-CoV-2 and other Human CoV genetic diversity while panels targeting a wide variety of viral families would be a better choice for evaluating the co-circulation of known and unknown human and animal CoV, which may be of relevance regarding potentially zoonosis as well as CoV strains co-circulation immune cross-protective phenomena.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: Complete list of all the identified viral species and families in the sewage March sample by Target Enrichment NGS, File S2: Complete SARS-CoV-2 assembled scaffold from sewage contigs, File S3: Mafft whole alignment from the recovered reads with respect to the alignment with 1200 randomly chosen GISAID SARS-CoV-2 sequences. The NGS datasets generated during the current study are available in zenodo under the DOI number 10.5281/zenodo.4252765.

Author Contributions: Conceptualization, S.B-M. and S.M-P.; methodology, S.M-P, M.I, M.R, E.F; software, J.A., S.M-P; validation, S.B-M, R.G and J.A.; formal analysis, S.M-P, S.B-M and J.A; investigation, S.M-P and S.B-M.; resources, S.B-M, R.G, J.Q, A.A.; data curation, S.M-P and C.A; writing—original draft preparation, S.M-P and S.B-M.; writing—review and editing, S.M-P, S.B-M., M.I, M.R, E.F, C.A, A.A. J.Q, R.G a J.A; visualization, S.M-P and J.A.; supervision, S.B-M.; project administration, S.B-M, R.G, A.A, J.Q.; funding acquisition, S.B-M, R.G, A.A, J.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially supported by and the Ministry of Science, Innovation and Universities (AGL2017-86797-C2-1-R) through the University of Barcelona and the Direcció General de Recerca i Innovació en Salut (DGRIS) Catalan Health Ministry Generalitat de Catalunya through Vall d'Hebron Research Institute (VHIR). Sílvia Bofill-Mas is a Serra-Hunter fellow at the University of Barcelona.

Acknowledgments: We thank the members of ICRA for the coordination of the sewage sampling. We also gratefully acknowledge the authors originating and submitting laboratories of the SARS-Cov-2 sequences from GISAID's EpiCov Database (https://www.epicov.org/epi3/) used to build the Figure 3 alignment; access to the individual isolates is facilitated through GISAID web site (https://www.gisaid.org/).

Conflicts of Interest: The authors declare no conflict of interest

References

1. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.M.; Wang, W.; Song, Z.G.; Hu, Y.; Tao, Z.W.; Tian, J.H.; Pei, Y.Y.; et al. A new coronavirus associated with human respiratory disease in China. Nature 2020, 579, 265–269, doi:10.1038/s41586-020-2008-3.

2. Wu, Y.; Guo, C.; Tang, L.; Hong, Z.; Zhou, J.; Dong, X.; Yin, H.; Xiao, Q.; Tang, Y.; Qu, X.; et al. Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. Lancet Gastroenterol. Hepatol. 2020, 5, 434–435, doi:10.1016/S2468-1253(20)30083-2.

3. Wölfel, R.; Corman, V.M.; Guggemos, W.; Seilmaier, M.; Zange, S.; Müller, M.A.; Niemeyer, D.; Jones, T.C.; Vollmar, P.; Rothe, C.; et al. Virological assessment of hospitalized patients with COVID-2019. Nature 2020, 581, 465–469, doi:10.1038/s41586-020-2196-x.

4. Cantalupo, P.G.; Calgua, B.; Zhao, G.; Hundesa, A.; Wier, A.D.; Katz, J.P.; Grabe, M.; Hendrix, R.W.; Girones, R.; Wang, D.; et al. Raw sewage harbors diverse viral populations. MBio 2011, 2, e00180-11, doi:10.1128/mBio.00180-11.

5. Ng, T.F.F.; Marine, R.; Wang, C.; Simmonds, P.; Kapusinszky, B.; Bodhidatta, L.; Oderinde, B.S.; Wommack, K.E.; Delwart, E. High variety of known and new RNA and DNA viruses of diverse

origins in untreated sewage. J. Virol. 2012, 86, 12161–12175, doi:10.1128/JVI.00869-12.

6. Martínez-Puchol, S.; Rusiñol, M.; Fernández-Cassi, X.; Timoneda, N.; Itarte, M.; Andrés, C.; Antón, A.; Abril, J.F.; Girones, R.; Bofill-Mas, S. Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. Sci. Total Environ. 2020, 713, doi:10.1016/j.scitotenv.2020.136604.

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

8. CDC 2019-nCoV CDC EUA qPCR Assay. 2020.

9. Vilsker, M.; Moosa, Y.; Nooij, S.; Fonseca, V.; Ghysens, Y.; Dumon, K.; Pauwels, R.; Alcantara, L.C.; Vanden Eynden, E.; Vandamme, A.-M.; et al. Genome Detective: an automated system for virus identification from high-throughput sequencing data. Bioinformatics 2018, 1–3, doi:10.1093/bioinformatics/bty695.

10. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 2014, 30, 2114–2120, doi:10.1093/bioinformatics/btu170.

11. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009, 25, 2078–2079, doi:10.1093/bioinformatics/btp352.

12. Barnett, D.W.; Garrison, E.K.; Quinlan, A.R.; Strïmberg, M.P.; Marth, G.T. Bamtools: A C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 2011, 27, 1691–1692, doi:10.1093/bioinformatics/btr174.

13. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 2012, 9, 357–359, doi:10.1038/nmeth.1923.

14. Grabherr, M.G..; Brian J. Haas, Moran Yassour Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W., N.; Friedman, and A.R. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 2013, 29, 644–652, doi:10.1038/nbt.1883.Trinity.

15. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Philip, D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Macmanes, M.D.; et al. De novo transcript sequence recostruction from RNA-Seq: reference generation and analysis with Trinity; 2013; Vol. 8; ISBN 3942667509.

16. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 2013, 30, 772–780, doi:10.1093/molbev/mst010.

17. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob. Challenges 2017, 1, 33–46, doi:10.1002/gch2.1018.

18. Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance 2017, 22, 2–4, doi:10.2807/1560-7917.ES.2017.22.13.30494.

19. Bofill-Mas, S.; Albinana-Gimenez, N.; Clemente-Casares, P.; Hundesa, A.; Rodriguez-Manzano, J.; Allard, A.; Calvo, M.; Girones, R. Quantification and stability of human adenoviruses and polyomavirus JCPyV in wastewater matrices. Appl. Environ. Microbiol. 2006, 72, 7894–7896, doi:10.1128/AEM.00965-06.

20. Briese, T.; Kapoor, A.; Mishra, N.; Jain, K.; Kumar, A.; OJ, J.; Lipkin, W. Virome Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive Virome Analysis. MBio 2015, 6, 1–11, doi:10.1128/mBio.01491-15.Editor.

21. Hjelmsø, M.H.; Mollerup, S.; Jensen, R.H.; Pietroni, C.; Lukjancenko, O.; Schultz, A.C.; Aarestrup, F.M.; Hansen, A.J. Metagenomic analysis of viruses in toilet waste from long distance flights—A new procedure for global infectious disease surveillance. PLoS One 2019, 14, e0210368, doi:10.1371/journal.pone.0210368.

22. Krafcikova, P.; Silhan, J.; Nencka, R.; Boura, E. Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. Nat. Commun. 2020, 11, 1–7, doi:10.1038/s41467-020-17495-9.

23. Van der Hoek, L. Human coronaviruses: What do they cause? Antivir. Ther. 2015, 12.

24. Ye, Z.W.; Yuan, S.; Yuen, K.S.; Fung, S.Y.; Chan, C.P.; Jin, D.Y. Zoonotic origins of human coronaviruses. Int. J. Biol. Sci. 2020, 16, 1686–1697, doi:10.7150/ijbs.45472.

25. Vogel, L.; Van Der Lubben, M.; Te Lintelo, E.G.; Bekker, C.P.J.; Geerts, T.; Schuijff, L.S.; Grinwis, G.C.M.; Egberink, H.F.; Rottier, P.J.M. Pathogenic characteristics of persistent feline enteric coronavirus infection in cats. Vet. Res. 2010, 41, doi:10.1051/vetres/2010043.

26. Gundy, P.M.; Gerba, C.P.; Pepper, I.L. Survival of Coronaviruses in Water and Wastewater. Food Environ. Virol. 2009, 1, 10–14, doi:10.1007/s12560-008-9001-6.

27. Pratelli, A. Genetic evolution of canine coronavirus and recent advances in prophylaxis. Vet. Res. 2006, 37, 191–200, doi:10.1051/vetres:2005053.

28. Wang, W.; Lin, X.; Guo, W.; Zhou, R.; Wang, M.; Wang, C.-Q.; Ge, S.; Mei, S.-H.; Li, M.-H.; Shi, M.; et al. Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. Virology 2015, 474, 19–27, doi:10.1016/j.virol.2014.10.017.

29. Wartecki, A.; Rzymski, P. On the coronaviruses and their associations with the aquatic environment and wastewater. Water (Switzerland) 2020, 12, 1–27, doi:10.3390/W12061598.

30. Kivistö, I.; Tidenberg, E.M.; Lilley, T.; Suominen, K.; Forbes, K.M.; Vapalahti, O.; Huovilainen, A.; Sironen, T. First Report of Coronaviruses in Northern European Bats. Vector-Borne Zoonotic Dis. 2020, 20, 155–158, doi:10.1089/vbz.2018.2367.

31. Hall, R.J.; Wang, J.; Peacey, M.; Moore, N.E.; McInnes, K.; Tompkins, D.M. New alphacoronavirus in Mystacina tuberculata bats, New Zealand. Emerg. Infect. Dis. 2014, 20, 697–700, doi:10.3201/eid2004.131441.

32. De Sabato, L.; Lelli, D.; Faccin, F.; Canziani, S.; Di Bartolo, I.; Vaccari, G.; Moreno, A. Full genome characterization of two novel Alpha-coronavirus species from Italian bats. Virus Res. 2019, 260, 60–66, doi:10.1016/j.virusres.2018.11.007.

33. Zhou, H.; Chen, X.; Hu, T.; Li, J.; Song, H.; Liu, Y.; Wang, P.; Liu, D.; Yang, J.; Holmes, E.C.; et al. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. Curr. Biol. 2020, 30, 2196-2203.e3, doi:10.1016/j.cub.2020.05.023.

34. Li, B.; Si, H.-R.; Zhu, Y.; Yang, X.-L.; Anderson, D.E.; Shi, Z.-L.; Wang, L.-F.; Zhou, P. Discovery of Bat Coronaviruses through Surveillance and Probe Capture-Based Next-Generation Sequencing. mSphere 2020, 5, 1–10, doi:10.1128/msphere.00807-19.

35. Lim, X.F.; Lee, C.B.; Pascoe, S.M.; How, C.B.; Chan, S.; Tan, J.H.; Yang, X.; Zhou, P.; Shi, Z.; Sessions, O.M.; et al. Detection and characterization of a novel bat-borne coronavirus in Singapore using multiple molecular approaches. J. Gen. Virol. 2019, 100, 1363–1374, doi:10.1099/jgv.0.001307.

36. Klempt, P.; Brož, P.; Kašný, M.; Novotný, A.; Kvapilová, K.; Kvapil, P. Performance of Targeted Library Preparation Solutions for SARS-CoV-2 Whole Genome Analysis. Diagnostics 2020, 10, 769, doi:10.3390/diagnostics10100769.

37. Carbo, E.C.; Sidorov, I.A.; Zevenhoven-Dobbe, J.C.; Snijder, E.J.; Claas, E.C.; Laros, J.F.J.; Kroes, A.C.M.; de Vries, J.J.C. Coronavirus discovery by metagenomic sequencing: a tool for pandemic preparedness. J. Clin. Virol. 2020, 131, 104594, doi:10.1016/j.jcv.2020.104594.

38. Xiao, M.; Liu, X.; Ji, J.; Li, M.; Li, J.; Yang, L.; Sun, W.; Ren, P.; Yang, G.; Zhao, J.; et al. Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. Genome Med. 2020, 12, 1–15, doi:10.1186/s13073-020-00751-4.

39. Nasir, J.A.; Kozak, R.A.; Aftanas, P.; Raphenya, A.R.; Smith, K.M.; Maguire, F.; Maan, H.; Alruwaili, M.; Banerjee, A.; Mbareche, H.; et al. A comparison of whole genome sequencing of sars-cov-2 using amplicon-based sequencing, random hexamers, and bait capture. Viruses 2020, 12, doi:10.3390/v12080895.

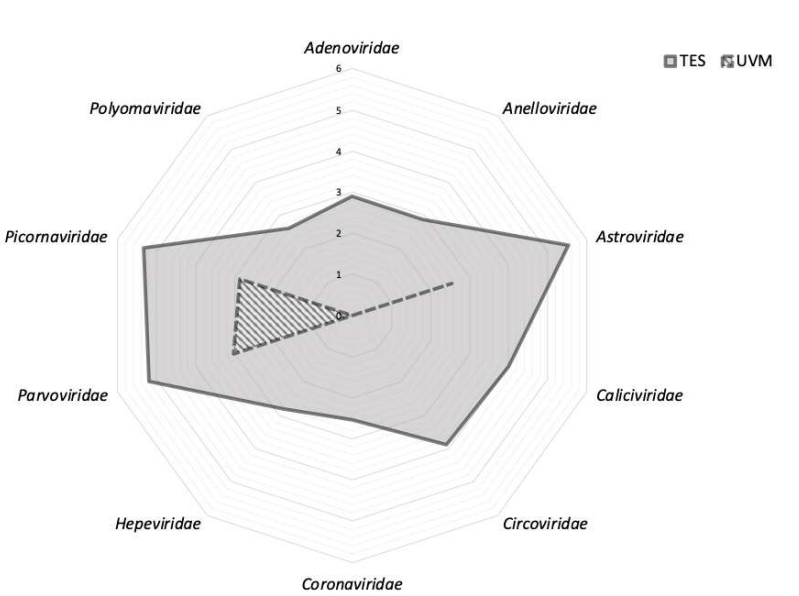


Figure 1: Diversity and relative abundance of vertebrate viral families reads. The results for non-enrichment NGS are represented in dotted lines, regular lines represent the results after Target enrichment NGS. Relative abundance is represented in log10.ç

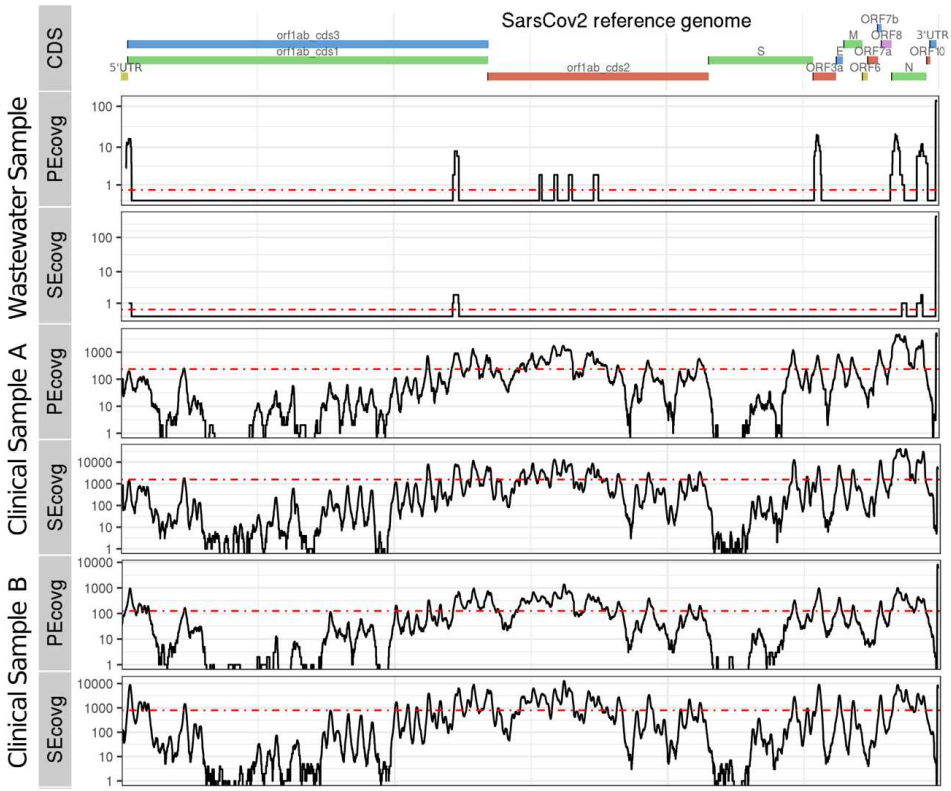


Figure 2: Summary of reads coverage along the SARS-CoV-2 reference genome sequence. The number of reads per position for the clean Pair-End reads (PEcovg) and for the Single-End reads (SEcovg) is show here for the three samples discussed on this manuscript. SE-reads were produced when cleaning the raw PE-reads, corresponding to the single member of a pair or reads that passed the cleaning and trimming criteria while the other read was discarded. Top track represents the segments defining the open reading frames encoding for the viral proteins (except the trailing 5'UTR and 3'UTR segments that are not protein-coding and were included for illustrative purposes). Dashed red line is the average coverage.

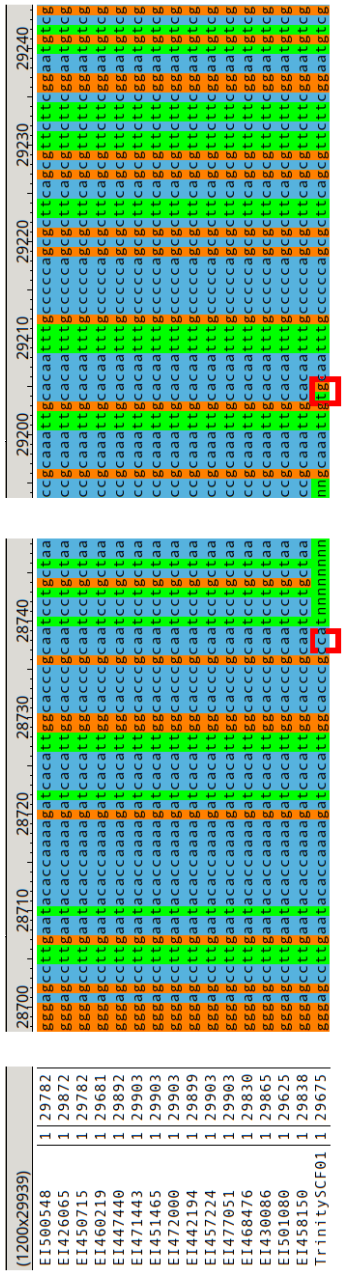


Figure 3: Nucleotide mismatches found in the nucleocapsid protein region. The red boxes highlight the 3 nucleotides that differ on the sequence reconstructed from the recovered reads with respect to the alignment with 1200 randomly chosen GISAID SARS-CoV-2 sequences (including the reference). Those nucleotide substitutions correspond to 1 synonymous and 2 non-synonymous changes: 28721.c**a**a[Q] <> 28514.c**C**a[P], 29187.tg**c**[C] == 28959.tg**T**[C], 29188.**a**ca[T] <> 28960.**G**ca[A]. Those three positions do not change on the alignment for all the other provided sequences yet are supported by only three reads which in fact does not allow defining them as variants. Full alignment is available as Supplementary File 3.

CAPÍTOL 2: IDENTIFICACIÓ DE POTENCIALS AGENTS ETIOLÒGICS CAUSANTS DE GASTROENTERITIS MITJANÇANT SEQÜENCIACIÓ MASSIVA

Article 3: Utilització de la metagenòmica per a l'estudi dels virus associats a gastroenteritis d'etiologia desconeguda

"Unveiling viruses associated with gastroenteritis using a metagenomics approach."

Fernández Cassi X.*, Martínez-Puchol S.*, Cornejo T., Bartolomé R., Bofill-Mas S., Girones R.

Viruses, 12(12), 1432; https://doi.org/10.3390/v12121432.

La gastroenteritis aguda infecciosa és una malaltia amb elevada importància a nivell mundial, especialment en nens, on aproximadament el 70% dels casos estan associats a infeccions víriques. També un elevat nombre de casos presenten etiologia desconeguda, pel que l'aparició de les tècniques de seqüenciació massiva ha obert noves oportunitats per a la detecció i descobriment de patògens vírics.

Es van analitzar per metagenòmica 124 mostres de femta de pacients amb gastroenteritis, dels que no es va trobar agent causal clàssic, per tal d'identificar possibles infeccions víriques. Les famílies de virus que infecten vertebrats més abundantment trobades han estat *Astroviridae* i *Caliciviridae*, amb la detecció de NoV GIV i sapovirus. També s'han detectat seqüències d'enterovirus, adenovirus i rotavirus en algunes de les mostres analitzades. Aquests resultats posen de manifest que caldria optimitzar les tècniques de cribratge rutinari que s'utilitzen actualment, per a fer possible la detecció d'astrovirus, sapovirus i NoV GIV.

L'aplicació de la metagenòmica per a l'estudi d'infeccions víriques té doncs un gran potencial en l'anàlisi clínic rutinari per, principalment, identificar patògens que tradicionalment no s'han testat i escapen els test rutinaris i també per detectar noves variants que poden causar gastroenteritis.

viruses

MDPI

Article

Unveiling Viruses Associated with Gastroenteritis Using a Metagenomics Approach

Xavier Fernandez-Cassi [1,*,†,‡], Sandra Martínez-Puchol [1,†], Marcelle Silva-Sales [1,2], Thais Cornejo [3], Rosa Bartolome [3], Silvia Bofill-Mas [1] and Rosina Girones [1]

[1] Laboratory of Virus Contaminants of Water and Food, Department of Genetics, Microbiology and Statistics, University of Barcelona, Barcelona 08028, Spain; smartinezpuchol@ub.edu (S.M.-P.); marcelle.figueira@ufg.br (M.S.-S.); Sbofill@ub.edu (S.B.-M.); rgirones@ub.edu (R.G.)

[2] Laboratory of Virology and Cell Culture, Tropical Pathology and Public Health Institute, Federal University of Goiás, 74605-050 Goiânia, Goiás, Brazil

[3] Hospital Universitari Vall d'Hebron, Microbiology Service, 08035 Barcelona, Spain; thais.cornejo@vhir.org (T.C.); rbartolome@vhebron.net (R.B.)

* Correspondence: xaviako@gmail.com

† These authors contributed equally to this work.

‡ Present address: Laboratory of Environmental Chemistry, School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.

Academic Editor: Lennart Svensson

Received: 4 November 2020; Accepted: 8 December 2020; Published: 13 December 2020

Abstract: Acute infectious gastroenteritis is an important illness worldwide, especially on children, with viruses accounting for approximately 70% of the acute cases. A high number of these cases have an unknown etiological agent and the rise of next generation sequencing technologies has opened new opportunities for viral pathogen detection and discovery. Viral metagenomics in routine clinical settings has the potential to identify unexpected or novel variants of viral pathogens that cause gastroenteritis. In this study, 124 samples from acute gastroenteritis patients from 2012–2014 previously tested negative for common gastroenteritis pathogens were pooled by age and analyzed by next generation sequencing (NGS) to elucidate unidentified viral infections. The most abundant sequences detected potentially associated to acute gastroenteritis were from *Astroviridae* and *Caliciviridae* families, with the detection of norovirus GIV and sapoviruses. Lower number of contigs associated to rotaviruses were detected. As expected, other viruses that may be associated to gastroenteritis but also produce persistent infections in the gut were identified including several *Picornaviridae* members (EV, parechoviruses, cardioviruses) and adenoviruses. According to the sequencing data, astroviruses, sapoviruses and NoV GIV should be added to the list of viral pathogens screened in routine clinical analysis.

Keywords: acute gastroenteritis; viral metagenomics; children; next-generation sequencing; norovirus; sapovirus; astrovirus; *Picornaviridae*

1. Introduction

Gastroenteritis is an important health problem around the globe affecting people at any age, but it is especially severe among children and the elderly. In 2016, diarrhea was the eighth leading cause of death among populations of all ages, causing an estimated 1,655,944 deaths and representing the fifth cause of death in children under 5 years with more than 446,000 estimated deaths [1]. While most of these deaths occur in developing countries, in developed countries acute gastroenteritis is still an important cause of death and an economic problem [2]. Clinical gastroenteritis can be caused

by several infectious agents including bacteria, protozoa and viruses, the transmission of which can be foodborne or waterborne, through contaminated fomites or by person-to-person contact. Among all gastroenteritis etiological agents, viruses account for approximately 70% of acute cases in children with rotavirus (RV), norovirus (NoV), and adenovirus (AdV), accounting for the majority of cases [3,4], with an increase in astrovirus (AstV) and sapovirus (SaV) in the same age groups [5,6]. Despite these important viral agents, it has also been estimated that nearly 40% of gastroenteritis cases have an unknown etiological agent, suggesting that most could be caused by undiscovered viruses [7]. In recent years, the advent of next-generation sequencing (NGS) technologies has opened new opportunities for viral pathogen detection and discovery in the clinical setting [8]. Currently, more than 10,261 validated, complete viral genomes are available in databases (e.g., the National Center for Biotechnology Information (NCBI) RefSeq in September 2020, https://www.ncbi.nlm.nih.gov/genome/viruses/).

In the present manuscript, fecal samples from acute gastroenteritis patients who had previously tested negative for common pathogenic bacteria (*Salmonella* sp., *Campylobacter* sp., *Shigella* sp., *Yersinia* sp. and Verotoxigenic *E. coli* (VTEC).), parasites (*Giardia* sp. and *Cryptosporidium* sp.) and viruses (rotavirus, adenovirus, norovirus) using conventional routine methodologies, were analyzed by NGS to elucidate unidentified viral infections.

2. Materials and Methods

2.1. Sample Collection and Characterization

A total of 124 fecal samples from patients with acute gastroenteritis of unknown origin were collected in collaboration with Hospital Universitari Vall d'Hebron, Barcelona, Catalonia, Spain during 2012 and 2014. Previously, fecal samples were screened for the most common entero-pathogens, including *Salmonella* sp., *Campylobacter* sp., *Shigella* sp., *Yersinia* sp., and Verotoxigenic *E. coli* (VTEC), for which they tested negative. Samples were also screened for *Giardia* sp. and *Cryptosporidium* sp. if the clinical symptoms were compatible with a parasitic infection. In addition, frequent viral gastroenteritis agents for the specific age groups were tested using immunochromatographic assays for human adenoviruses (ONE STEP Adenovirus CARD TEST, Certest Biotec S.L), human rotaviruses (ONE STEP Rotavirus CARD TEST, Certest Biotec S.L) and human astroviruses (ONE STEP Astrovirus CARD TEST, Certest Biotec S.L), and conventional RT-PCR (norovirus GI and GII) [9]. No etiological agent could be identified to explain the clinical acute gastroenteritis; hence, the samples were screened using metagenomic means by pooling the samples according to age criteria in four different categories: less than 1 year (3 pools: F < 1, S < 1A and S < 1B), from 1 to 3 years (4 pools: F1-3A, S1-3A, S1-3B and S1-3C), from 3 to 10 years (1 pool: S3-10), and older than 10 years (1 pool: F > 10). For pools "F", all samples were tested by RT-PCR for noroviruses [9]. Samples pooled under "S" criteria were only tested by RT-PCR under specific clinical request. A summary of the sample features composing each pool and the clinical tests performed is represented as Table 1.

Table 1. Summary of the samples composing each sequenced pool according to age criteria and the performed conventional diagnostic assays in the clinical center. The results of the bacterial and viral parameters are expressed as the "number positive samples/number samples tested". For human adenovirus, human rotavirus and human astrovirus, a commercially available immunochromatographic test was employed (ONE STEP Adenovirus CARD TEST, ONE STEP Rotavirus CARD TEST and ONE STEP Astrovirus CARD TEST, Certest Biotec S.L). For norovirus GI and GII, conventional RT-PCR was used [9].

			Bacterial Parameters							**Viral Parameters**				
Pool	**Age**	**N**	***Salmonella* sp.**	***Campylobacter* sp.**	***Shigella* sp.**	***Yersinia* sp.**	***Aeromonas* sp.**	***Verotoxigenic E. coli (VTEC)***	***Vibrio* sp.**	**Rotavirus**	**Adenovirus**	**Astrovirus**	**Norovirus GGI**	**Norovirus GGII**
F < 1	<1	10	0/9	0/9	0/9	0/9	0/0	0/0	0/0	0/10	0/10	0/10	0/10	0/10
S < 1A		9	0/9	0/9	0/9	0/9	0/7	0/0	0/8	0/9	0/9	0/9	0/0	0/0
S < 1B		9	0/9	0/9	0/9	0/9	0/8	0/0	0/8	0/9	0/9	0/9	0/1	0/1
F1-3A	1–3	23	0/23	0/23	0/23	0/23	0/0	1/1	0/1	0/23	0/23	0/23	0/23	0/23
S1-3A		14	0/13	0/13	0/13	0/13	0/11	0/1	0/11	0/13	0/13	0/13	0/1	0/1
S1-3B		14	0/14	0/14	0/14	0/14	0/7	0/0	0/8	0/14	0/14	0/14	0/1	0/1
S1-3C		13	0/14	0/14	0/14	0/14	0/8	0/0	0/11	0/14	0/14	0/14	0/0	0/0
F3-10	3–10	13	0/13	0/13	0/13	0/13	0/0	0/0	0/0	0/12	0/12	0/13	0/13	0/13
S3-10		10	0/9	0/9	0/9	0/9	0/9	0/0	0/9	0/1	0/10	0/10	0/0	0/0
F > 10	>10	9	0/9	0/9	0/9	0/9	0/0	0/1	0/1	0/8	0/8	0/9	0/9	0/9

2.2. Viral Concentration, Free DNA Removal, Nucleic Acid Extraction, and Library Preamplification and Preparation Using Untargeted Viral Metagenomics (UVM) and Target Enrichment Sequencing (TES)

Individual fecal samples of 0.5 g were suspended in 4.5 mL PBS 1x solution to obtain a 10% fecal suspension. Suspensions were homogenized and pooled by mixing 1 mL of each sample to construct the pools (Table 1). To enrich the viral fraction, fecal pools were centrifuged for 15 min at 3,000× *g* to remove cellular debris and bacteria; the supernatant was collected and filtered using a 0.22-μm low-binding protein syringe filter (Millex-GV, 0.22 μm, PVDF, 33 mm, Gamma-Sterilized, Millipore, Massachusetts, USA). Filtered pools (500 μL) were treated with 230 units of Turbo DNAse (Ambion, Lithuania) for 1 h at 37 °C to remove free DNA prior to nucleic acid extraction. The NA (nucleic acids) present in 280 μL of DNAse-treated viral concentrates was extracted using the QIAamp®Viral RNA Mini Kit from QIAGEN (Qiagen, Valencia, CA, USA). NA were eluted in 60 μL and stored at −80 °C until further processing. To enable the detection of both DNA and RNA viruses, the total NAs were reverse-transcribed as previously described by Fernandez-Cassi et al. [10]. In short, SuperScript III (Life Technologies, Carlsbad, CA, USA) was used to retrotranscribe RNA to cDNA with primerA (5′-GTTTCCCAGTCACGATCNNNNNNNNN-3′). Second strand cDNA and DNA were constructed including primer sequences using Sequenase (USB/Affymetrix, Cleveland, OH, USA). PCR amplification with AmpliTaqGold (Life Technologies, Austin, Texas, USA) was performed using primerB (5′-GTTTCCCAGTCACGATC-3′); this step was performed in duplicate. After 10 min at 95 °C to activate DNA polymerase, the following PCR program was used: 25 cycles of 30 s at 94 °C, 30 s at 40 °C, and 30 s at 50 °C, with a final step of 60 s at 72 °C. The PCR products were purified and eluted in 15 μL using a Zymo DNA clean and concentrator (cat no. D4013, Zymo Research, Irvine, CA, USA) to yield sufficient DNA for the library preparation. Amplified DNA samples were quantified using Qubit 2.0 (cat no. Q32854, Life Technologies, OR, USA).

Untargeted viral metagenomics (UVM) viral libraries were constructed using a Nextera XT DNA sample preparation kit following the manufacturer's instructions (Illumina Inc).

For the Target Enrichment Sequencing (TES), one pool (F2, 1–3 years) was reprocessed and resequenced by hybridizing the libraries with the vertebrate viral capture panel (VirCapSeq Enrichment Kit, Roche, Pleasanton, CA, USA) developed by Briese and coworkers [11] as detailed by Martínez-Puchol et al. [12]. Libraries were sequenced for three different runs using Illumina MiSeq 2 × 300 in base-pair paired-end format.

2.3. Bioinformatics Analysis

Bioinformatics analysis was performed using the Genome Detective platform (https://www.genomedetective.com/) [13]. For a more precise and accurate taxonomic classification, human viral contigs longer than 100 bp were queried for sequence similarity using BLASTN against the NCBI GenBank nucleotide collection database [14] with Geneious Software 11.1.5 (www.geneious.com) [15].

The species nomenclature and classification were performed according to the NCBI Taxonomy standards. For specific typing of human calicivirus, the assembled genomes of human noroviruses and sapoviruses were uploaded into the Noronet web-based service Typing tool (version 2.0) developed by RIVM using the updated classification of norovirus genogroups and genotypes [16,17]. Specific phylogenetic trees for Sapovirus were performed by aligning the complete genomes and the VP1 region against reference strains using MUSCLE [18]. Trees were constructed using tree view [19] software within Geneious (software version 11.0, https://www.geneious.com). For the *Enterovirus* genus within the *Picornaviridae* family, the same approach was performed, using the Enterovirus Genotyping Tool (version 0.1) developed by the RIVM [17]. For human adenovirus, retrieved contigs were mapped against their closest adenovirus type, and the hexon, penton, and fiber genes regions were annotated, extracted, and phylogenetically analyzed using the prototype strains suggested by the human Adenovirus Working Group (http://hadvwg.gmu.edu/). For human rotavirus analysis, a routinely binary phylogenetic analysis using the genes VP4 and VP7 was performed with the tool Rotavirus A Genotype Determination (ViPR, www.viprbrc.org) [20]. Contigs belonging to the

mamastrovirus and *parechovirus* genus were aligned to their complete reference genomes according to the ICTV classification. In addition, specific alignments using the ORF2 region and VP1 region for mamastrovirus and parechovirus against selected reference strains were conducted [21,22].

2.4. RT-PCR and RT-qPCR for Specific Viral Pathogens Sequencing

Specific viral RT-qPCR to detect NoVGIV was performed using primers and conditions as previously described [9,23,24] for individual fecal samples. RT-PCR for norovirus genogroup II was used in a specific subset of samples to confirm the obtained NGS results [25]. RT-PCR-positive band products were purified using Zymo DNA clean and concentrator (Zymo Research, cat no. D4013, Zymo Research, Irvine, CA, USA) and submitted to Sanger sequencing at Serveis Científico-Tècnics of the University of Barcelona.

3. Results and Discussion

3.1. Metagenomic Identification of Viral Families in Pooled Fecal Samples

A summary of the MiSeq output obtained from each sequenced pool containing the raw read number, contigs, and viral sequences detected, among other information, is presented as Supplementary material 1. Approximately 44 million raw-paired ends (0.9 to 12 M per sample) were generated. On average, 95% of the reads passed the QC cut-off and were annotated as viral using the Diamond tool [26], with 46% of these reads identified as viral (23% to 88%). When analyzing the samples according to age, a minimum and maximum percentage of viral reads ranging from to 29% to 50%, 23% to 56%, and 41% to 88% were obtained for pools with individuals aged younger than 1 year (F < 1, S < 1A and S < 1B), 1 to 3 years (F1-3A, S1-3A, S1-3B and S1-3C), and 3 to 10 years (F3-10 and S-3-10), respectively. Despite efforts to enhance viral content using filtration and nuclease treatment, 54% of the reads were classified as nonviral (12–77%). In the pool enriched by viral probe capture prior to sequencing (F1-3A), 74% of the reads were classified as viral.

A complete inventory of identified viral species classified according to the genome detective pipeline are presented in Supplementary material 2. Relevant taxonomical assignments for acute gastroenteritis will be discussed further. Additionally, these relevant pathogens which were considered important viral pathogens associated with gastroenteritis were plotted according to age and log10 abundance of reads (Figure 1).

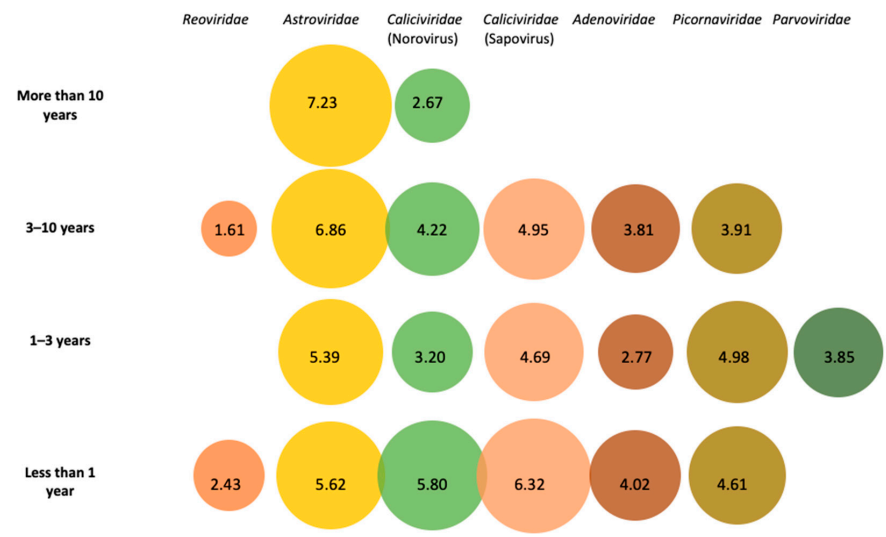


Figure 1. Age distribution of the number of reads (log10) from human viral families with etiologically related members with viral gastroenteritis.

Astrovirus, Sapovirus and NoV GIV were detected with a high number of contigs, retrieving almost their complete genomes, and are considered the potential causative agents of gastroenteritis.

3.2. Caliciviridae

The results for this family have been divided into the main two human genera: *Norovirus* (NoV) and *Sapovirus* (SaV).

A summary of the viral contigs identified in the pools belonging to *Caliciviridae* family is presented in Table 2.

Table 2. Characterization of *Caliciviridae* contigs detected in the gastroenteritis pools with the norovirus typing tool by RIVM (v 2.0) using the dual system assessing the polymerase (ORF1) and the capsid region (ORF2). NT: non-typable by the RIVM Norovirus Genotyping Tool.

Pool	Contig Length	Classification	Polymerase Genotype	Capsid Genotype	BLAST Score
F < 1	7,521	Sapovirus GV	NT	GV.1	98.03
S < 1A	7,440	Norovirus GII	GII.P21 (GII.Pb)	GII.3	80.95
F1-3A	7,187	Norovirus GIV	NT	NT	77.26
	4,827	Sapovirus GI	NT	GI.2	73.61
S1-3A	5,978	Norovirus GII	GII.P31 (GII.Pe)	GII.2	85.10
	1,107	Norovirus GII	NT	NT	88.89
	7,292	Sapovirus GII	NT	GII.1b	74.37
	550	Sapovirus GI	NT	NT	74.27
	376	Sapovirus GI	NT	NT	74.86
	698	Sapovirus GI	NT	NT	76.72
	2,109	Sapovirus GI	NT	NT	72.10
	859	Sapovirus GI	NT	GI.3	78.06
	1,321	Sapovirus GI	NT	GI.3	71.14
S1-3B	6,808	Norovirus GII	GII.P17	GII.17	75.56
F3-10	7,384	Norovirus GII	GII.P4 New_Orleans_2009	GII.4 New_Orleans_2009	84.54
	7,521	Sapovirus GV	NT	GV.1	98.08
S3-10	3,533	Norovirus GI	NT	NT	
	2,065	Norovirus GI	NT	GI.2	69.38
	798	Norovirus GI	GI.P2	GI.2	77.26
	1,551	Sapovirus GI	NT	NT	75.02
	1,059	Sapovirus GI	NT	NT	72.61
	345	Sapovirus GI	NT	GI.3	66.47
	349	Sapovirus GI	NT	GI.3	73.99
F > 10	924	Norovirus GIV	NT	NT	98.70
	576	Norovirus GIV	NT	NT	98.44
	1,249	Norovirus GIV	NT	NT	99.25
	375	Norovirus GIV	NT	NT	98.67

3.2.1. Norovirus

Norovirus is the leading viral agent related to acute gastroenteritis in all age groups worldwide, with frequent outbreaks related to schools, hospitals, or cruise ships [27]. Its classification into genogroups and P-types (polymerase types) is sustained using a binary system based on the amino acid diversity of the complete VP1 gene and nucleotide diversity of the RNA-dependent RNA polymerase (RdRp) [16,17]. However, due their continuous recombination, NoV taxonomical classification is constantly evolving [16]. Four main genogroups are known to infect humans: GI, GII, GIV, and GIX (formerly GII.15) [16]. Norovirus GII genogroups are responsible for the majority of acute gastroenteritis cases, with GII.4 being the most predominant genotype worldwide with different variants emerging frequently [28].

In pools S1-3A, F1-3A, and F1-3B, three NoV GII contigs were identified. Two contigs of 5978 bp and 6808 bp were classified as GII.P31-GII.2 and GII.P17-GII.17, respectively. Infections caused by GII.P17-GII.17 have been increasing over the last decade, being predominant in some Asian regions, where they have displaced the classical GII.P4-GII.4 and its recombinant strains [29]. Its presence has increased and expanded in subsequent years after the first reported case, presenting a worldwide distribution and producing outbreaks elsewhere [30,31]. According to Van Beek et al., NoV GIIP.17 and GII.17 strains were detected in most European Countries in 2015–2016, but not in Spain [29]. The detection in 2014 samples suggests that the virus was already circulating in the population outside China. To verify this finding, conventional RT-PCR was conducted by following the suggested typing protocol described by van Beek and collaborators [25]. One sample from the S1-3B pool tested positive for NoV GII, and the amplicon was submitted to the RIVM Norovirus Genotyping Tool to confirm the finding of GII.17 (Accession number: MW205840).

Finally, one contig from pool F3-10 was subtyped as NoV GII.P4-GII.4 (New Orleans 2009). As stated previously, this is the predominant genotype worldwide and accounts for more than 70% of norovirus outbreaks [32], with GII.4 subvariants emerging, cocirculating, and replacing each other over time [33,34]. This particular strain could have been considered as non-detected by conventional RT-PCR due to a failure of the assay.

In pool S < 1A, one contig of 7,440 bp of NoV GII.P21-GII.3 was obtained. This recombinant strain is one of the most common strains circulating during the period under study [25,35,36]. In the same pool, one NoV GI.2-P1.2 was detected. This genotype has been linked to the consumption of polluted shellfish [37–39]. The non-detection of this particular genotype has been reported previously on some commercially available kits [40,41], and in-silico analysis revealed three mismatches with two of the forward primers used in clinical routine detection.

Several contigs taxonomically classified as NoV GIV were retrieved from the F1-3A and F > 10 pools. To explore the real incidence of norovirus GIV, a specific RT-qPCR assay for NoV GIV was performed using individual patient samples to confirm the finding. Three of 26 samples (12% of incidence) within the affected pools were positive by the RT-qPCR assay, with viral loads ranging from 2.71×10^6 to 1.34×10^7 genomic copies/gr feces. The viral loads are compatible with an active replication of the virus and could explain the acute gastroenteritis of these three patients, despite its lower incidence compared with NoVGII and GI. The detected contigs could not be typed using the Noronet Typing tool. In an attempt to obtain a more precise taxonomic classification, full NoV GIV retrieved was aligned against NoV GIV full complete genomes available at genbank using MUSCLE, and a phylogenetic tree was constructed (Figure 2). Full sequenced NoV GIV shares a homology of 99.2% with the NoV GIV strain from Australia (JQ613567) and could be classified as GIV.1. Norovirus GIV was first reported in 2000 by analyzing sporadic gastroenteritis cases [42], after which two different genotypes were identified, with NoV GIV.1 affecting humans and NoV GIV.2 considered a zoonotic genotype producing gastroenteritis in animals. Norovirus infections caused by GIV.1 are rare, and its prevalence in the population remains unknown. An international molecular surveillance NoV study conducted between 2005 and 2016 showed that less than 0.1% of the sequences submitted to NoroNet belonged to the NoV GIV.1 genotype [29]. Our findings are in line with previously reported results showing an incidence of 16% of NoV GIV in clinical samples [43]. The presence of this genotype has been detected by metagenomics in samples from children affected by gastroenteritis [44], as well as in other clinical and urban wastewater samples [43,45,46], indicating that it is actively circulating in the population but probably underreported. While the introduction of routine molecular detection in clinical settings could be the key for better monitoring of GIV-related cases, the NGS techniques have been shown to be a useful tool to detect a broader spectrum of norovirus genotypes and genogroups. This is especially relevant considering that their detection by common RT-PCR can sometimes fail due to the high variability in the P1 and VP1 regions, once again lending importance to the improvement of primer-independent detection techniques [47].

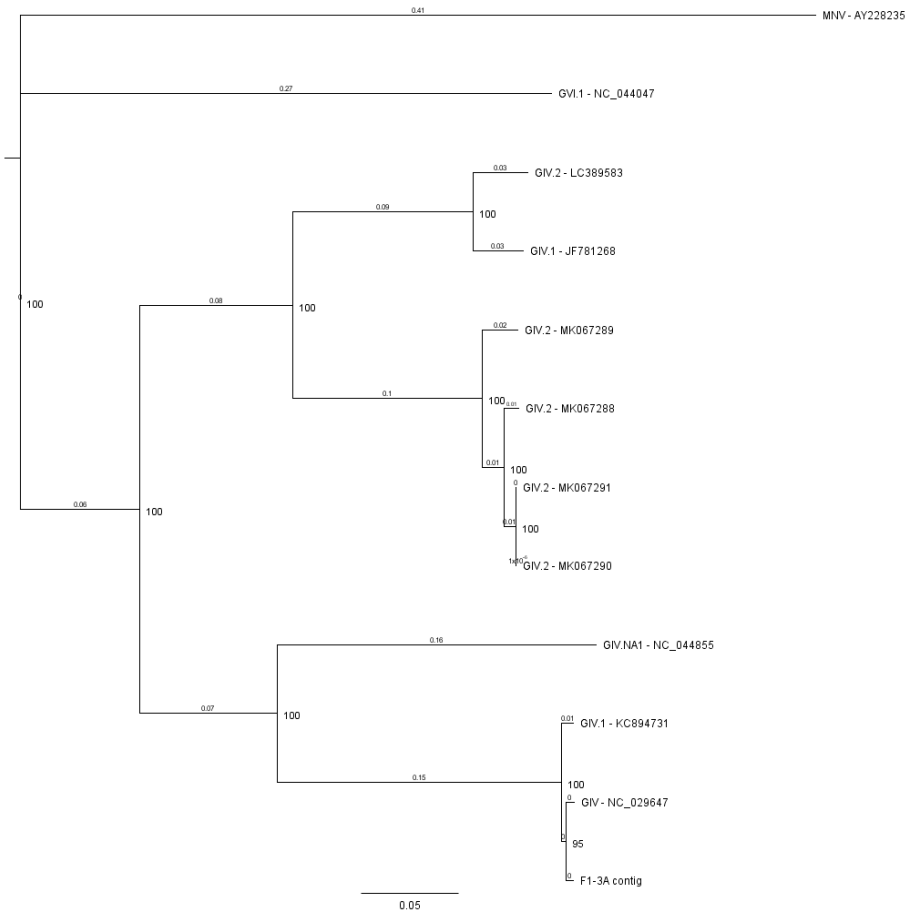


Figure 2. Phylogenetic tree of the almost-complete NoV GIV sequence from the F1-3A pool. Full genome sequences of NoV GIV are available at GenBank and were aligned using MUSCLE (Geneious 11.1). The murine norovirus sequence (accession number: AY228235) was used to root the tree. The tree was constructed with the Tamura-Nei genetic distance model using the neighbor-joining method with 1,000 bootstrap replicates. The nodes labels indicate the consensus support whereas the labels on the branch indicate the number of substitutions per site.

3.2.2. Sapovirus

Sapoviruses are one of the leading viral causes of gastroenteritis in children, with a particular impact in developing countries [48,49]. Their taxonomic classification is based on the capsid (VP1) sequence, with GI, GII, GIV and GV being the genogroups responsible for human gastroenteritis. Additionally, genogroups are further classified into genotypes, a classification system that keeps growing due to their high genetic diversity, with GI.1, GII.1, and GI.2 being the genotypes most commonly associated with clinical cases [48,50,51].

Contigs from the *sapovirus* genus were found in all sample pools except F > 10 (Table 2). Most of the detected sapoviruses were classified as GI (11/14), the most prevalent genogroup in patients within the studied age range (1 to 10 years) [49]. The typed GI contigs were GI.2 and GI.3, but no contig belonging to GI.1 was identified despite being the most prevalent in gastroenteritis samples and in this patient age group [49,52].

Contigs identified as Sapovirus GII (1/14) and GV (2/14) could not be typed using the Noronet Typing tool even though the complete genome was nearly assembled. GV is commonly linked to food-borne outbreaks [53–55], and some studies have shown a lack of detection of these types by

conventional RT-PCR, while they can be detected when applying NGS [54,55]. Alignments of the complete genomes retrieved from the sequenced pools against the complete genomes and the VP1 region from reference strains showed the clustering of GII contigs with SaV GII.1b, whereas sequenced GV contigs clustered with GV.1 (Supplementary material 3).

3.3. Astroviridae

Human astroviruses (HAstV) are important agents causing acute gastroenteritis in children and have also been involved in outbreaks affecting adults [56]. Recently, astroviruses have also been associated with a wider array of diseases, including central nervous system alterations such as meningitis or acute flaccid paralysis [57,58]. The classification is currently based on the amino acid sequence of the capsid region (ORF2), dividing the HAsV in the "classical" genotypes (Mamastrovirus 1 (MAstV-1) species) and the "novel" recombinant ones (MAstV-6, 8, and 9 species) [22,59]. A summary of the astrovirus species detected is presented in Table 3.

Table 3. Characterization of human astrovirus contigs detected in the tested fecal pools.

Pool	Contig length	Classification	ORF2 Genotype	% ORF2 Identity	% ORF2 Coverage	% Genome Identity	% Genome Coverage
F < 1	6,569	Mamastrovirus 6	MLB-3	97.8	100	98.1	99.8
S1-3A	6,642	Mamastrovirus 1	HAstV-2	92.7	100	85.7	97.6
F3-10	6,652	Mamastrovirus 1	HAstV-1	95.4	100	94.1	97.6
F > 10	6,651	Mamastrovirus 1	HAstV-1	95.1	100	94.0	97.6

Nearly the full genomes (6,469–6,652 bp) of four different astroviruses were retrieved. Human astrovirus species were found in all analyzed pools (Figure 1). HAstV-1 was detected in two pools and is the most common type associated with infantile gastroenteritis [60]. A HAstV-2 presenting a low homology over the full genome (85.7% identity) was detected in the S1-3A pool. During the development of the study, a commercially available immunochromatographic test was used to detect human astroviruses. In pooled samples "F", this test was only used under specific circumstances (clinical signs supported an astrovirus infection, while samples "S" were all tested with this immunochromatographic test). Interestingly, among samples tested by the antigenic test, one result was inconclusive, which could be partially explained by the low homology of the detected astrovirus in this pool of samples. Phylogenetic trees with HAstV ORF2 complete sequences (Supplementary material 4a) and complete HAstV genomes (Supplementary material 4b) were constructed.

Nearly the full genome (6,569 bp) of an astrovirus MLB-3 was detected in the F < 1 pool. This novel astrovirus was first described in 2014 in stools from children with acute diarrhea [61]. It has also been reported in asymptomatic patients [62] or involved in coinfections with other microorganisms that cause diarrhea, raising doubts regarding its role as a gastroenteritis causative agent [62,63].

3.4. Adenoviridae

Human adenovirus is a well-known pathogen in children as well as adults, producing a wide array of diseases, from gastroenteritis to conjunctivitis, respiratory infections, cystitis and meningitis. There are currently 104 recognized human adenovirus types according to the Human adenovirus working group (http://hadvwg.gmu.edu/) divided into seven different adenovirus species (A–G) with species D being the most abundant [64]. Sequencing has allowed the identification of several recombinant types, rapidly increasing the number of reported human adenoviruses in the literature [65,66].

Human adenovirus contigs were reported in pools F1-3A, S1-3B and F3-10. No human adenovirus species F (HAdV40 and 41), which are important etiological gastroenteritis agents in children [67], were detected in any of the pooled samples. This suggests a high specificity of immunochromatographic tests for HAdV species F. However, the full genome of a HAdV31 has been retrieved from the pool F3-10. The phylogenetic analysis of the sequenced human adenovirus clusters

with HAdV31 reference type (Supplementary material 5). HAdV31 has been identified as a causative agent of acute gastroenteritis in children despite not belonging to HAdV species F [68].

Other HAdV type partial sequences have been identified based on the hexon, penton and fiber analysis. These adenoviruses might belong to species A (HAdV18, HAdV61, HAdV61, HAdV31), B (HAdV3, HAdV66), and C (HAdV89). Human adenoviruses are known to produce permanent infections in symptomatic and asymptomatic individuals. Hence, the detection of fragments of the adenovirus genomes could be indicative of the presence of persistent HAdV infections [69,70]. However, the lower number of contigs associated to other than HAdV31 and HAdV F species make them unlikely to be the causative agents of acute gastroenteritis.

3.5. Reoviridae

Rotavirus (RV) reads were detected in pools F < 1, S < 1A, S < 1B and S3-10. All detected RV belonged to the genotype G2P[4]. This genotype is one of the six most commonly genotypes detected worldwide, together with G1P[8], G3P[8], G4P[8], G9P[8], and G12P[8] [71,72]. RV has been identified as the main leading fatal etiologic agent for diarrhea in children less than five years old, causing 128,515 deaths in 2016 [1].

In the pool of fecal samples F < 1, S < 1A, and S < 1B, reads from 9 out of the 11 RVA genes were as follows: six structural protein (VP) coding genes and three nonstructural (NSP) coding genes. In the genotyping analysis, R2, C2, M2, P[4], I2, and G2, referred to as VP1, VP2, VP3, VP4, VP6, and VP7, respectively, were detected, and N2, T2, E2, referred to as NSP2, NSP3, and NSP4, respectively, were detected for the NSP genotypes [73]. In pool S3-10, reads of the genes coding for VP2, VP3, VP4, VP6, and VP7 with the same genotype pattern observed previously were found: C2, M2, P(4), I2, and G2. RVA detected in the present study showed high similarity (99.5–100%) to the genotype G2P(4) RVA strains circulating in Europe, Asia, and South America in the last decade for both pooled samples (data not shown). In this study, no RVA sequences revealed a close nucleotide similarity with the two commercially anti-RVA vaccines (Rotarix® and Rotateq®) available worldwide. Together, these results suggest that children probably acquired RVA strains circulating in Barcelona from 2012 to 2014, having no correlation with an anti-RVA vaccination. A study conducted between 2015 and 2016 revealed the presence of RVA in high viral loads in sewage samples from Barcelona. The authors also found no evidence of RVA vaccine-related strains in the sewage samples [74]. These results suggest that RVA is widespread in the community, since the population excretes RVA and the children still require healthcare assistance due to gastroenteritis associated with RVA.

Immunochromatographic tests for RVA diagnosis have been a useful tool to provide quick and cost-effective tests for routine clinical analyses at the expense of low sensitivity when compared to molecular methods such as RT-PCR [75,76]. The low number of reads obtained within the pools might be explained by three important facts: (a) the samples could have low levels of virus, since all were negative in the commonly used test; (b) individual patient samples were pooled, diluting the number of viral reads present in positive patients; (c) dsRNA viruses have been shown to be difficult to detect using current NGS techniques [77].

3.6. Other Viral Species Identified in Fecal Samples

Detected picornaviruses have been taxonomically classified according to their genera: *Enterovirus, Kobuvirus, Parechovirus and Cardiovirus*. Enterovirus and non-enterovirus picornaviruses have been summarized in Supplementary materials 6 and 7.

Enterovirus (EV) causes a wide range of diseases in children, such as hand-foot-mouth disease (HFMD), viral meningitis and encephalitis, acute flaccid paralysis, myocarditis, common cold or neonatal sepsis. It is also reported to be present in asymptomatic infections or coinfections with another microorganism both in children and adults [78].

The EV genus is subdivided by its capsid sequence (VP1) in 13 species, with EV-A, EV-B, EV-C, EV-D, rhinovirus A (RV-A), RV-B and RV-C infecting humans [79,80]. Enterovirus from species EV-A and EV-B with RV-A and RV-B were detected by NGS in the pooled samples. An exhaustive list

containing the typed enterovirus and rhinovirus using the RIVM enterovirus typing tool is provided in Supplementary material 6.

Rhinoviruses A (HRV-A) and B (HRV-B) were present in all the pools analyzed. These species are the most important etiological agents of common cold, with a high incidence of infections in early childhood, mainly affecting the upper respiratory airways, with HRV-A being the most prevalent [81]. In the S2 pool, nearly the full genome of HRV-A78 was obtained (6,499 bp). HRV is one of the predominant circulating species of HRV-A in children [82] and has been previously associated with pneumonia [83].

Members of the genus *Parechovirus* (HPeV) are common causes of aseptic meningitis in children and mild gastrointestinal and respiratory diseases [84]. Currently, more than 18 different types have been recognized to infect humans according to the HPeV ICTV taxonomy group, all of which cluster under HPEV-A species (https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/picornaviridae/693/genus-parechovirus). Sequences from HPeV A were detected in pools associated with younger patients (<3 years). Phylogenetic analysis allowed their classification as HPeV-1 and HPeV-6 (Supplementary material 8).

The *Cardiovirus* genus was first associated with rodent infections, until their role in human pathogenicity was described in 2007 [85]. There are currently three recognized species, *Cardiovirus* A, *Cardiovirus* B and *Cardiovirus* C. More than 11 genotypes of Cardio-virus B species are associated with human infection, with Saffold viruses (SAFV) 2 and 3 being the most prevalent.

Almost the whole genome of a *Cardiovirus* B member, genotyped as SAFV-2, was sequenced in pool F > 10 from affected individuals above 10 years old (Supplementary material 7). SAFV are not routinely tested in clinical samples, but recent studies have reported a common presence in gastroenteritis patients [86,87] and in raw sewage [88,89].

The genus *Kobuvirus* is composed of six recognized *Aichi virus* (AiV) viral species (AiVA-AiVF) that are classified according to their amino acid homology in the polyprotein, P1 2C and 3CD regions (https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/picornaviridae/686/genus-kobuvirus). Members of the AiV A and B species have been recognized as human pathogens causing gastroenteritis outbreaks [90–92]. Members within this genus are considered zoonotic and have the potential to infect dogs, cats or goats [93].

Two different Aichi virus contigs were detected in pool S3-10. The alignments of the retrieved contigs were phylogenetically related to AiV A which has been associated with acute gastroenteritis.

Picobirnavirus-related contigs were detected in most of the pools analyzed except F < 1 and S3-10. Previously, picobirnavirus has been considered a putative pathogenic viral species present in the respiratory and digestive tract of humans and other animals, including pigs, cows, dogs and otters. However, recent research has suggested that picobirnavirus might be a prokaryotic virus because the ribosomal binding site (RBS) motif is conserved within its genome. This motif has been detected in all available picobirnavirus sequences and is commonly enriched in all viral families that infect prokaryotes, but not in eukaryotic viral families [94].

Contigs belonging to primate bocaparvovirus 2, formerly known as Human Bocavirus 2, were found in pool S6. These viral species have been previously found in stool samples from gastroenteritis-related studies in young children with variable prevalence depending on the area of study and year of sample collection [95].

A diverse number of ssDNA circular viruses belonging to the *Anelloviridae* and *Circoviridae* families have been detected among the different pools analyzed. The family *Anelloviridae* includes the Torque teno virus (TTV), Torque teno midi virus (TTMDV) and mini virus Torque teno virus (TTMV). The detection of anellovirus species in clinical samples does not seem unlikely, as it is estimated that 70–90% of the human population carries the virus in loads that range from 10^3 to 10^6 genomic copies/mL plasma [96]. None of the different torque teno virus species known to date have been directly associated with disease [97].

However, the leading role in acute gastroenteritis of the discussed viral agents under this subsection is at least controversial, as most of them can replicate actively in the intestinal tract [98],

have been co-detected with other well stablished gastrointestinal viral pathogens such as RV, NoV, SaV or HAdV [78,86,92,99] or are commonly detected in non-symptomatic control samples [97,100].

Further studies focusing on these specific families in non-pooled samples are needed to clarify their involvement in acute gastroenteritis either as monoinfection or coinfection pathogens.

3.7. Target Enrichment Sequencing

A target enrichment approach (TES) was performed in one pool (F1-3A) to evaluate its usefulness for the detection of gastroenteritis viral agents compared with untargeted metagenomics (UVM) applied in this study. The application of a target enrichment probe-based capture method, prior to sequencing, has been demonstrated to overcome the low sensitivity or the lack of sequencing depth in complex samples with large viral diversity [12,101].

Comparing both methodologies and focusing on viral families of interest regarding gastroenteritis etiology, TES enabled the detection of a larger number of viral reads (Figure 4a), allowing the detection of *Reoviridae* members, which were not detected with the UVM approach. Additionally, TES allowed the typing of the hexon, penton and fiber genes of a HAdV-31 that could not be typed by UVM. In contrast, when analyzing these reads as contigs, the differences were reduced to obtain the same number of contigs by applying both methodologies to the detected families (Figure 4b). This observation could be related to the sequence fragmentation and diversity of the sample, the absence of sequences from different regions of each viral assignation, or the absence of regions to which the probes are hybridizing. This phenomenon would result in a large number of reads from the same genome region and a reduction of diversity, but an increase in accuracy of the contig sequence.

Figure 4. Comparison of the number of viral reads (**a**) and contigs (**b**) from families with members related to gastroenteritis with Target Enrichment Sequencing (TES) and Untargeted Viral Metagenomics (UVM) in the F1-3A pool. The number of reads is represented as log10.

The application of TES allowed the detection of 4 RV contigs in the sample pool analyzed that were not previously detected in the UVM pool. The detected rotavirus sequences belonged to genotype G2P[4], the predominant RV circulating in Europe between 2010 and 2019 [71,72,102]. As observed for other detected RV sequences in the study, none of the contigs was related to the two commercially available rotavirus vaccines.

4. General Discussion and Conclusions

Viral metagenomics, in combination with or without a viral enrichment approach, has the potential to uncover any viral species that might be present in a given sample. These methodologies are not primer-dependent methods; hence, they detect viral species that might be undetected by using conventional molecular methodologies such as RT-PCR or RT-qPCR. The use of primer-independent methodologies represents an important advantage in detecting new viral pathogens (i.e., SARS-CoV-2) or recombinant/reassortant types. These methodologies are especially relevant for viral species, as

viruses are prone to have high mutation rates and to recombine/reassort their genomes as an evolutionary mechanism. In addition, the gastrointestinal tract is a complex ecosystem, and the evolution and severity of acute gastroenteritis might be conditioned by the presence of coinfections that could be masked by the application of single pathogen detection methods such as immunochromatography or real-time PCR. However, a large number of sequences obtained by metaviromics cannot be taxonomically classified by using reference databases and are considered viral dark matter, requiring a more exhaustive analysis [103,104]. The use of a viral-enriched approach (TES) increased the number of viral reads up to 74%, in contrast to the 23% viral reads detected without enrichment in the same pooled sample. However, despite the enrichment applied, a wider diversity of viral species could be detected without the viral enrichment with the exception of rotavirus A, which was not detected without applying TES.

According to the number of reads detected, the most abundant sequences from vertebrate viral families were from *Astroviridae*-related members, which were identified in all age groups, suggesting the need to include systematic molecular testing for this particular viral family independently of the patient age. Following AstV, NoV and SaV present an important number of reads. At the time of sample collection, clinical screening did not include any molecular assay to detect sapoviruses; thus, their detection using NGS seems reasonable. Sapoviruses have been identified as an increasing etiological agent for human gastroenteritis, and their screening in the clinical setting should be systematically performed at least for children under 10 years of age. Human noroviruses were detected in all pools analyzed despite most samples having been previously tested by conventional RT-PCR in clinical settings. Some of these NoV might have escaped detection by RT-PCR due to the presence of mismatches in primer annealing zones. Additional in-silico analysis of the NGS-retrieved sequence showed mismatches in the primer binding region, highlighting the utility of primer-independent methodologies to target viruses with a high mutation rate. Interestingly, NoV GIV, a minority NoV species related to gastroenteritis, was detected by NGS and confirmed in three patients by RT-qPCR. The inability to detect this specific NoV genogroup is expected because it is not routinely tested. Samples for which the NoVs test was ruled out according to clinical evaluation showed the presence of both NoV GI and GII. Therefore, we recommend the inclusion of NoVGI, GII and GIV in the systematic testing of all acute gastroenteritis cases.

HAdV not belonging to species F were detected in patients less than 10 years old. The use of immunochromatographic tests targeting adenovirus species F (HAdV 40 and 41) might present a lower sensitivity towards other adenoviruses (i.e., HAdV species A or C), which could be the etiological agents for some acute gastroenteritis such as the detected HAdV31. Nevertheless, HAdV are known to produce a wide array of infections including asymptomatic persistent infections of the gastrointestinal tract, raising doubts about their involvement in acute gastroenteritis. Similarly, human enteroviruses and other related picornaviruses (such as parechovirus and cardiovirus) can produce a wide array of clinical outcomes but are also persistently excreted by healthy asymptomatic individuals. Thus, without any additional clinical information, their detection in clinical samples might result from a coinfection with another gastroenteritis virus or simply as a subclinical infection. In this sense, the application of viral metagenomics has emerged as a promising tool contributing to increased knowledge of excreted viruses and their role in human pathogenesis.

Finally, a small proportion of reads could be taxonomically assigned to the family *Reoviridae*, which includes HRV. According to the phylogenetic analyses, the identified RV were not related with any of the existent vaccine strains; hence, their presence in the analyzed pools suggests their implication in the clinical gastroenteritis despite the low genome coverage and small number of reads retrieved from this particular viral family in this study. Similarly to HAdV, the small number of obtained reads could be explained by an inefficient tagging of double-stranded genomes during the retro-transcription or sequenase steps resulting in low coverage and the small number of reads assigned to this viral family. Specific protocols to increase the sensitivity towards double-stranded genomes (dsDNA and dsRNA) might need to be applied to explore the viral diversity within these genomic conformations [77].

Supplementary Materials: Supplementary Materials can be found at www.mdpi.com/1999-4915/12/12/1432/s1.

Author Contributions: Conceptualization, X.F.-C., S.M.-P., S.B.-M. and R.G.; Methodology, X.F.-C. and S.M.-P.; Software, X.F.-C., M.S.-S. and S.M.-P.; Validation, R.G., S.B.-M. and R.B.; Formal Analysis, X.F.-C., M.S.-S. and S.M.-P.; Investigation, X.F.-C. and S.M.-P.; Resources, R.G., S.B.-M. and R.B.; Data Curation, S.M.-P., X.F.-C. and T.C.; Writing–Original Draft Preparation, X.F.-C. and S.M.-P.; Writing–Review & Editing, X.F.-C., M.S.-S., S.M.-P., R.G., S.B.-M., R.B., T.C.; Visualization, X.F.-C. and S.M.-P.; Supervision, R.G.; Project Administration, R.G.; Funding Acquisition, R.G. and S.B.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially funded by the Programa RecerCaixa 2012 (ACUP-00300) the projects AGL2011-30461-C02-01/ALI from the Spanish Ministry of Science and Innovation and the project AGL2014-55081-R from the Ministry of Economy. This study was partially funded by grants from the Catalan Government to the Consolidated Research Group VirBaP (2014SRG914) with the collaboration of the Institut de Recerca de l'Aigua (IdRA).

Availability of Data and Material: The datasets generated during the current study are available in zenodo under the DOI number https://doi.org/10.5281/zenodo.4161275 (https://zenodo.org/record/4161275#.X6KAjyySlPY).

Acknowledgments: During the development of the study Xavier Fernandez Cassi was a fellow of the Catalan government "AGAUR" (fi-dgr) while Sandra Martínez-Puchol was a fellow of Economy, Industry and Competitiveness Ministry (MINECO). Marcelle Silva-Sales would like to acknowledge the Brazilian National Council for Scientific and Technological Development (CNPq, in Portuguese: Conselho Nacional de Desenvolvimento Científico e Tecnológico) for support via a postdoctoral grant. Sílvia Bofill-Mas is a Serra-Hunter fellow at the University of Barcelona.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Troeger, C.; Blacker, B.F.; Khalil, I.A.; Rao, P.C.; Cao, S.; Zimsen, S.R-M.; Albertson, S.B.; Stanaway, J.D; Deshpande, A.; Deshpande, A.; et al. stimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect. Dis.* **2018**, *18*, 1211–1228.
2. Bányai, K.; Estes, M.K.; Martella, V.; Parashar, U.D. Viral gastroenteritis. *Lancet* **2018**, *392*, 175–186.
3. Fletcher, S.; Van Hal, S.; Andresen, D.; McLaws, M.L.; Stark, D.; Harkness, J.; Ellis, J. Gastrointestinal pathogen distribution in symptomatic children in Sydney, Australia. *J. Epidemiology Glob. Heal.* **2013**, *3*, doi:10.1016/j.jegh.2012.11.004.
4. Oppong, T.B.; Yang, H.; Amponsem-Boateng, C.; Kyere, E.K.D.; Abdulai, T.; Duan, G.; Opolot, G. Enteric pathogens associated with gastroenteritis among children under 5 years in sub-saharan Africa: A systematic review and meta-Analysis. *Epidemiol. Infect.* **2020**, *148*, 1–26.
5. Tran, A.; Talmud, D.; Lejeune, B.; Jovenin, N.; Renois, F.; Payan, C.; Leveque, N.; Andreoletti, L. Prevalence of rotavirus, adenovirus, norovirus, and astrovirus infections and coinfections among hospitalized children in Northern France. *J. Clin. Microbiol.* **2010**, *48*, 1943–1946.
6. Chhabra, P.; Payne, D.C.; Szilagyi, P.G.; Edwards, K.M.; Staat, M.A.; Shirley, S.H.; Wikswo, M.; Nix, W.A.; Lu, X.; Parashar, U.D.; Vinjé, J. Etiology of viral gastroenteritis in children <5 years of age in the United States, 2008–2009. *J. Infect. Dis.* **2013**, *208*, 790–800.
7. Holtz, L.R.; Finkbeiner, S.R.; Zhao, G.; Kirkwood, C.D.; Girones, R.; Pipas, J.M.; Wang, D. Klassevirus 1, a previously undescribed member of the family Picornaviridae, is globally widespread. *Virol. J.* **2009**, *6*, 86.
8. Fischer, T.K.; Rasmussen, L.D.; Fonager, J. Taking gastro-surveillance into the 21st century. *J. Clin. Virol.* **2019**, *117*, 43–48.
9. Kageyama, T.; Kojima, S.; Shinohara, M.; Uchida, K.; Fukushi, S.; Hoshino, F.B.; Takeda, N.; Katayama, K. Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR. *J. Clin. Microbiol.* **2003**, *41*, 1548–1557.
10. Fernandez-Cassi, X.; Rusiñol, M.; Martínez-Puchol, S. *Viral Concentration and Amplification from Human Serum Samples Prior to Application of Next-Generation Sequencing Analysis*; Springer: New York, NY, USA, 2018; pp. 173–188.

11. Briese, T.; Kapoor, A.; Mishra, N.; Jain, K.; Kumar, A.; Jabado, O.J.; Lipkin, W.I. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio* **2015**, *6*, e01491-15.
12. Martínez-Puchol, S.; Rusiñol, M.; Fernández-Cassi, X.; Timoneda, N.; Itarte, M.; Andrés, C.; Antón, A.; Abril, J.F.; Girones, R.; Bofill-Mas, S. Characterisation of the sewage virome: Comparison of NGS tools and occurrence of significant pathogens. *Sci. Total. Environ.* **2020**, *713*, 136604.
13. Vilsker, M.; Moosa, Y.; Nooij, S.; Fonseca, V.; Ghysens, Y.; Dumon, K.; Pauwels, R.; Alacantara, L.C.; Vanden Eynden, E.; Vandamme, A-M.; Deforche, K.; de Oliveira, T. Genome Detective: An automated system for virus identification from high-throughput sequencing data. *Bioinformatics* **2019**, *35*, 871–873.
14. Benson, D.A.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2015**, *43*, D30–D35.
15. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Dura, C. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649.
16. Chhabra, P.; de Graaf, M.; Parra, G.I.; Chan, M.C.W.; Green, K.; Martella, V.; Wang, Q.; White, P.A.; Katayama, K.; Vennema, H.; Koopmans, M.; Vinjé, J. Updated classification of norovirus genogroups and genotypes. *J. Gen. Virol.* **2019**, *100*, 1393–1406.
17. Kroneman, A.; Vennema, H.; Deforche, K.; Avoort, H.; Peñaranda, S.; Oberste, M.S.; Vinjé, J.; Koopmans, M. An automated genotyping tool for enteroviruses and noroviruses. *J. Clin. Virol.* **2011**, *51*, 121–125.
18. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797.
19. Page, R.D.M. Treeview: An application to display phylogenetic trees on personal computers. *Bioinformatics* **1996**, *12*, 157–158.
20. Maes, P.; Matthijnssens, J.; Rahman, M.; Van Ranst, M. RotaC: A web-based tool for the complete genome classification of group A rotaviruses. *BMC Microbiol.* **2009**, *9*, 238.
21. Böttcher, S.; Obermeier, P.E.; Neubauer, K.; Diedrich, S. Recombinant enterovirus A71 subgenogroup C1 strains, Germany, 2015. *Emerg. Infect. Dis.* **2016**, *22*, 1843–1846.
22. Donato, C.; Vijaykrishna, D. The broad host range and genetic diversity of mammalian and avian astroviruses. *Viruses* **2017**, *9*, 102.
23. Loisy, F.; Atmar, R.L.; Guillon, P.; Le Cann, P.; Pommpepuy, M.; Le Guyader, F.S. Real-time RT-PCR for norovirus screening in shellfish. *J. Virol. Methods* **2005**, *123*, 1–7.
24. Miura, T.; Parnaudeau, S.; Grodzki, M.; Okabe, S.; Atmar, R.L.; Le Guyader, F.S. Environmental detection of genogroup I, II, and IV noroviruses by using a generic real-time reverse transcription-PCR assay. *Appl. Environ. Microbiol.* **2013**, *79*, 6585–6592.
25. Van Beek, J.; van der Eijk, A.A.; Fraaij, P.L.A.; Caliskan, K.; Cransberg, K.; Dalinghaus, M.; Hoek, R.A.S.; Metselaar, H.J.; Roodnat, J.; Vennema, H.; Koopmans, M. Chronic norovirus infection among solid organ recipients in a tertiary care hospital, the Netherlands, 2006–2014. *Clin. Microbiol. Infect.* **2017**, *23*, 265.e9–265.e13.
26. Buchfink, B.; Xie, C.; Huson, D.H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60.
27. Hall, A.J.; Eisenbart, V.G.; Etingüe, A.L.; Gould, L.H.; Lopman, B.A.; Parashar, U.D. Epidemiology of foodborne norovirus outbreaks, United States, 2001–2008. *Emerg. Infect. Dis.* **2012**, *18*, 1566–1573.
28. Tohma, K.; Lepore, C.J.; Gao, Y.; Ford-Siltz, L.A.; Parra, G.I. Population genomics of gii.4 noroviruses reveal complex diversification and new antigenic sites involved in the emergence of pandemic strains. *mBio* **2019**, *10*, e02202-19.
29. Van Beek, J.; de Graaf, M.; Al-Hello, H.; Allen, D.J.; Ambert-Balay, K.; Botteldoorn, N.; Brytting, M.; Buesa, J.; Cabrerizo, M.; Chan, M.; et al. Molecular surveillance of norovirus, 2005–2016: An epidemiological analysis of data collected from the NoroNet network. *Lancet Infect. Dis.* **2018**, *18*, 545–553.
30. De Graaf, M.; Van Beek, J.; Koopmans, M.P.G. Human norovirus transmission and evolution in a changing world. *Nat. Rev. Genet.* **2016**, *14*, 421–433.
31. Chen, L.; Xu, D.; Wu, X.; Liu, G.; Ji, L. An increasing prevalence of non-GII.4 norovirus genotypes in acute gastroenteritis outbreaks in Huzhou, China, 2014–2018. *Arch. Virol.* **2020**, *165*, 1121–1128.
32. Vega, E.; Barclay, L.; Gregoricus, N.; Shirley, S.H.; Lee, D.; Vinjé, J. Genotypic and epidemiologic trends of norovirus outbreaks in the united states, 2009 to 2013. *J. Clin. Microbiol.* **2013**, *52*, 147–155.

33. Lun, J.H.; Hewitt, J.; Yan, G.J.H.; Tuipulotu, D.E.; Rawlinson, W.; White, P. Recombinant GII.P16/GII.4 sydney 2012 was the dominant norovirus identified in Australia and New Zealand in 2017. *Viruses* **2018**, *10*, 548.
34. Sabrià, A.; Pintó, R.M.; Bosch, A.; Bartolomé, R.; Cornejo, T.; Torner, N.; Martínez. A.; de Simón, M.; Domínguez, A.; Guix, S. Molecular and clinical epidemiology of norovirus outbreaks in Spain during the emergence of GII.4 2012 variant. *J. Clin. Virol.* **2014**, *60*, 96–104.
35. Hoa Tran, T.N.; Trainor, E.; Nakagomi, T.; Cunliffe, N.A.; Nakagomi, O. Molecular epidemiology of noroviruses associated with acute sporadic gastroenteritis in children: Global distribution of genogroups, genotypes and GII.4 variants. *J. Clin. Virol.* **2013**, *56*, 269–277.
36. Mans, J.; Murray, T.Y.; Taylor, M.B. Novel norovirus recombinants detected in South Africa. *Virol. J.* **2014**, *11*, 168.
37. Le Guyader, F.S.; Atmar, R.L.; Le Pendu, J. Transmission of viruses through shellfish: When specific ligands come into play. *Curr. Opin. Virol.* **2012**, *2*, 103–110.
38. Yu, Y.; Cai, H.; Hu, L.; Lei, R.; Pan, Y.; Yan, S.; Wang, Y. Molecular epidemiology of oyster-related human noroviruses and their global genetic diversity and temporal-geographical distribution from 1983 to 2014. *Appl. Environ. Microbiol.* **2015**, *81*, 7615–7624.
39. Tunyakittaveeward, T.; Rupprom, K.; Pombubpa, K.; Howteerakul, N.; Kittigul, L. Norovirus Monitoring in Oysters Using Two Different Extraction Methods. *Food Environ. Virol.* **2019**, *11*, 374–382.
40. Kanwar, N.; Hassan, F.; Barclay, L.; Langley, C.; Vinjé, J.; Bryant, P.W.; George, K.; Mosher, L.; Matthews-Greer, J.M.; Rocha, M.A. Evaluation of RIDA ® GENE norovirus GI/GII real time RT-PCR using stool specimens collected from children and adults with acute gastroenteritis. *J. Clin. Virol.* **2018**, *104*, 1–4.
41. Dunbar, N.L.; Bruggink, L.D.; Marshall, J.A. Evaluation of the RIDAGENE real-time PCR assay for the detection of GI and GII norovirus. *Diagn. Microbiol. Infect. Dis.* **2014**, *79*, 317–321.
42. Vinjé, J.; Koopmans, M.P.G. Simultaneous detection and genotyping of "Norwalk-like viruses" by oligonucleotide array in a reverse line blot hybridization format. *J. Clin. Microbiol.* **2000**, *38*, 2595–2601.
43. La Rosa, G.; Pourshaban, M.; Iaconelli, M.; Muscillo, M. Detection of genogroup IV noroviruses in environmental and clinical samples and partial sequencing through rapid amplification of cDNA ends. *Arch. Virol.* **2008**, *153*, 2077–2083.
44. Ao, Y. yun, Yu, J. mei, Li, L. li, Jin, M.; Duan, Z. Detection of human norovirus GIV.1 in China: A case report. *J. Clin. Virol.* **2014**, *61*, 298–301.
45. Muscillo, M.; Fratini, M.; Graffeo, R.; Sanguinetti, M.; Martella, V.; Green, K.Y.; Della Libera, S.; La Rosa, G. GIV Noroviruses in Wastewaters and in Stool Specimens from Hospitalized Patients. *Food Environ. Virol.* **2013**, *5*, 194–202.
46. Fioretti, J.M.; Fumian, T.M.; Rocha, M.S.; dos Santos, I. de A.L.; Carvalho-Costa, F.A.; de Assis, M.R.; Rodrigues, J.; Leite, J.P.G, Miagostovich, M.P. Surveillance of Noroviruses in Rio De Janeiro, Brazil: Occurrence of New GIV Genotype in Clinical and Wastewater Samples. *Food Environ. Virol.* **2017**, *10*, 1–6.
47. Das Neves Costa, L.C.P.; Teixeira, D.M.; Portela, A.C.R.; De Lima, I.C.G.; da Silva Bandeira, R.; Sousa Júnior, E.C.; Siqueira, J.A.M.; Resque, H.R.; da Silva, L.D.; Gabbay, Y.B. Molecular and evolutionary characterization of norovirus gii.17 in the northern region of brazil. *BMC Infect. Dis.* **2019**, *19*, 1–11.
48. Xue, L.; Cai, W.; Gao, J.; Jiang, Y.; Wu, H.; Zhang, L.; Zuo, Y.; Dong, R.; Pang, R.; Zeng, H. et al. Genome characteristics and molecular evolution of the human sapovirus variant GII.8. *Infect. Genet. Evol.* **2019**, *73*, 362–367.
49. Varela, M.F.; Rivadulla, E.; Lema, A.; Romalde, J.L. Human sapovirus among outpatients with acute gastroenteritis in Spain: A one-year study. *Viruses* **2019**, *11*, 144.
50. Mann, P.; Pietsch, C.; Liebert, U.G.; Genetic diversity of sapoviruses among inpatients in germany, 2008–2018. *Viruses* **2019**, *11*, 726.
51. Oka, T.; Mori, K.; Iritani, N.; Harada, S.; Ueki, Y.; Iizuka, S.; Mise, K.; Murakami, K.; Wakita, T.; Katayama, K. Human sapovirus classification based on complete capsid nucleotide sequences. *Arch. Virol.* **2012**, *157*, 349–352.
52. Phan, T.G.; Okame, M.; Nguyen, T.A.; Nishio, O.; Okitsu, S.; Ushijima, H. Genetic diversity of sapovirus in fecal specimens from infants and children with acute gastroenteritis in Pakistan. *Arch. Virol.* **2004**, *150*, 371–377.
53. Usuku, S.; Kumazaki, M. A gastroenteritis outbreak attributed to sapovirus genogroup V in Yokohama, Japan. *Jpn. J. Infect. Dis.* **2014**, *67*, 411–412.

54. Shibata, S.; Sekizuka, T.; Kodaira, A.; Kuroda, M.; Haga, K.; Doan, Y.H.; Takai-Todaka, R.; Katayama, K.; Wakita, T.; Oka, T. Complete genome sequence of a novel GV.2 sapovirus strain, NGY-1, detected from a suspected foodborne gastroenteritis outbreak. *Genome Announc.* **2015**, *3*, e01553-14.
55. Hergens, M.; Nederby Öhd, J.; Alm, E.; Askling, H.H.; Helgesson, S.; Insulander, M.; Lagerqvist, N.; Svenungsson, B.; Tihane, M.; Tolfvenstam, T.; Follin, P. Investigation of a food-borne outbreak of gastroenteritis in a school canteen revealed a variant of sapovirus genogroup V not detected by standard PCR, Sollentuna, Sweden, 2016. *Eurosurveillance* **2017**, *22*, doi:10.2807/1560-7917.es.2017.22.22.30543.
56. Jeong, H.S.; Jeong, A.; Cheon, D.S. Epidemiology of astrovirus infection in children. *Korean J. Pediatr.* **2012**, *55*, 77–82.
57. Cordey, S.; Vu, D.-L.; Schibler, M.; L'Huillier, A.G.; Brito. F.; Docquier, M.; Posfay-Barbe, K.M.; Petty, T.J.; Turin, L.; Zdobnoy, E.M.; Kaiser, L. Astrovirus MLB2, a New Gastroenteric Virus Associated with Meningitis and Disseminated Infection. *Emerg. Infect. Dis.* **2016**, *22*, 846–853.
58. Kapoor, A.; Li, L.; Victoria, J.; Oderinde, B.; Mason, C.; Pandey, P.; Zaidi, S. Z.; Delwart, E. Multiple novel astrovirus species in human stool *J. Gen. Virol.* **2009**, *90*, 2965–2972.
59. Bosch, A.; Pintó, R.M.; Guix, S. Human Astroviruses. *Clin. Microbiol. Rev.* **2014**, *27*, 1048–1074.
60. Gabbay, Y.B.; Leite, J.P.G.; Oliveira, D.S.; Nakamura, L.S.; Nunes, M.R.T.; Mascarenhas, J.D.A.P.; Heinemann, M.B.; Linhares, A.C. Molecular epidemiology of astrovirus type 1 in Belém, Brazil, as an agent of infantile gastroenteritis, over a period of 18 years (1982–2000): Identification of two possible new lineages. *Virus Res.* **2007**, *129*, 166–174.
61. Jiang, H.; Holtz, L.R.; Bauer, I.; Franz, C.J.; Zhao, G.; Bodhidatta, L.; Shrestha, S.K.; Kang, G.; Wang, D. Comparison of novel MLB-clade, VA-clade and classic human astroviruses highlights constrained evolution of the classic human astrovirus nonstructural genes. *Virology* **2013**, *436*, 8–14.
62. Meyer, C.T.; Bauer, I.K.; Antonio, M.; Adeyemi, M.; Saha, D.; Oundo, J.O.; Ochieng, J.B.; Omore, R.; Stine, O.C.; Wang, D.; Holtz, L.R. Prevalence of classic, MLB-clade and VA-clade Astroviruses in Kenya and the Gambia Emerging viruses. *Virol. J.* **2015**, *12*, doi:10.1186/s12985-015-0299-z.
63. Vu, D.L.; Cordey, S.; Brito, F.; Kaiser, L. Novel human astroviruses: Novel human diseases? *J. Clin. Virol.* **2016**, *82*, 56–63.
64. Robinson, C.M.; Singh, G.; Lee, J.Y.; Dehghan, S.; Rajaiya, J.; Liu, E.B.; Yousuf, M.A.; Betensky, R.A.; Jones, M.S.; Dyer, D.W. Molecular evolution of human adenoviruses. *Scientific Reports* **2013**, *3*, 1812.
65. Kaján, G.L.; Kajon, A.E.; Pinto, A.C.; Bartha, D.; Arnberg, N. The complete genome sequence of human adenovirus 84, a highly recombinant new Human mastadenovirus D type with a unique fiber gene. *Virus Res.* **2017**, *242*, 79–84.
66. Hage, E.; Dhingra, A.; Liebert, U.G.; Bergs, S.; Ganzenmueller, T.; Heim, A. Three novel, multiple recombinant types of species of human mastadenovirus D (HAdV-D 73, 74 & 75) isolated from diarrhoeal faeces of immunocompromised patients. *J. Gen. Virol.* **2017**, *98*, 3037–3045.
67. Shimizu, H.; Phan, T.G.; Nishimura, S.; Okitsu, S.; Maneekarn, N.; Ushijima, H. An outbreak of adenovirus serotype 41 infection in infants and children with acute gastroenteritis in Maizuru City, Japan. *Infect. Genet. Evol.* **2007**, *7*, 279–284.
68. Leruez-Ville, M.; Chardin-Ouachée, M.; Neven, B.; Picard, C.; Le Guinche, I.; Fisher, A.; Rouzioux, C.; Blanche, S. Description of an adenovirus A31 outbreak in a paediatric haematology unit. *Bone Marrow Transplant.* **2006**, *38*, 23–28.
69. Roy, S.; Calcedo, R.; Medina-Jaszek, A.; Keough, M.; Peng, H.; Wilson, J.M. Adenoviruses in lymphocytes of the human gastro-intestinal tract. *PLoS ONE* **2011**, *6*, e24859.
70. Kosulin, K. Intestinal HAdV Infection: Tissue Specificity, Persistence, and Implications for Antiviral Therapy. *Viruses* **2019**, *11*, 804.
71. Da Silva, M.F.M.; Fumian, T.M.; de Assis, R.M.S.; Fialho, A.M.; Carvalho-Costa, F.A.; da Silva, J.; Leite, J.P.G. VP7 and VP8* genetic characterization of group A rotavirus genotype G12P[8]: Emergence and spreading in the Eastern Brazilian coast in 2014. *J. Med Virol.* **2016**, *89*, 64–70.
72. Bányai, K.; László, B.; Duque, J.; Steele, A.D.; Nelson, E. A. S.; Gentsch, J.R.; Parashar, U.D. Systematic review of regional and temporal trends in global rotavirus strain diversity in the pre rotavirus vaccine era: Insights for understanding the impact of rotavirus vaccination programs. *Vaccine* **2012**, *30*, A122–A130.

73. Matthijnssens, J.; Ciarlet, M.; Heiman, E.; Arijs, I.; Delbeke, T.; McDonald, S.M.; Palombo, E.A.; Iturriza-Gómara, M.; Maes, P.; Patton, J.T. Full Genome-Based Classification of Rotaviruses Reveals a Common Origin between Human Wa-Like and Porcine Rotavirus Strains and Human DS-1-Like and Bovine Rotavirus Strains. *J. Virol.* **2008**, *82*, 3204–3219.
74. Silva-Sales, M.; Martínez-Puchol, S.; Gonzales-Gustavson, E.; Hundesa, A.; Gironès, R. High prevalence of rotavirus A in raw sewage samples from northeast Spain. *Viruses* **2020**, *12*, 318.
75. Moutelíková, R.; Dvořáková Heroldová, M.; Holá, V.; Sauer, P.; Prodělalová, J. Human rotavirus a detection: Comparison of enzymatic immunoassay and rapid chromatographic test with two quantitative RT-PCR assays. *Epidemiol. Mikrobiol. Imunol.* **2018**, *63*, 110–113.
76. Khamrin, P.; Tran, D.N.; Chan-it, W.; Thongprachum, A.; Okitsu, S.; Maneekarn, N.; Ushijima, H. Comparison of the rapid methods for screening of group a rotavirus in stool samples. *J. Trop. Pediatr.* **2011**, *57*, 375–377.
77. Wilcox, A.H.; Delwart, E.; Díaz-Muñoz, S.L. Next-generation sequencing of dsRNA is greatly improved by treatment with the inexpensive denaturing reagent DMSO. *Microb. Genom.* **2019**, *5*, e000315.
78. Tatte, V.S.; Gopalkrishna, V. Detection of different enteric viruses in children with diarrheal disease: Evidence of the high frequency of mixed infections. *Access Microbiol.* **2019**, *1*, e000010.
79. Adams, M.J.; Lefkowitz, E.J.; King, A.M.Q.; Harrach, B.; Harrison, R.L.; Knowles, N.J.; Kropinski, A.M.; Krupovic, M.; Kuhn, J.H.; Mushegian, A.R.; et al. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2016). *Arch. Virol.* **2016**, *161*, 2921–3949.
80. Nikonov, O.S.; Chernykh, E.S.; Garber, M.B.; Nikonova, E.Y. Enteroviruses: Classification, diseases they cause, and approaches to development of antiviral drugs. *Biochemistry* **2017**, *82*, 1615–1631.
81. Arakawa, M.; Okamoto-Nakagawa, R.; Toda, S.; Tsukagoshi, H.; Kobayashi, M.; Ryo, A.; Mizuta, K.; Hasegawa, S.; Hirano, R.; Wakiguchi, H.; et al. Molecular epidemiological study of human rhinovirus species A, B and C from patients with acute respiratory illnesses in Japan. *J. Med Microbiol.* **2012**, *61*, 410–419.
82. Martin, E.T.; Kuypers, J.; Chu, H.Y.; Foote, S.; Hashikawa, A.; Fairchok, M.P.; Englund, J.A. Heterotypic infection and spread of rhinovirus a, b, and c among childcare attendees. *J. Infect. Dis.* **2018**, *218*, 848–855.
83. Daleno, C.; Piralla, A.; Scala, A.; Senatore, L.; Principi, N.; Esposito, S. Phylogenetic analysis of human rhinovirus isolates collected from otherwise healthy children with community-acquired pneumonia during five successive years. *PLoS ONE* **2013**, *8*, e80614.
84. Olijve, L.; Jennings, L.; Walls, T. Human parechovirus: An increasingly recognized cause of sepsis-like illness in young infants. *Clin. Microbiol. Rev.* **2017**, *31*, e00047-17.
85. Jones, M.S.; Lukashov, V.V.; Ganac, R.D.; Schnurr, D.P. Discovery of a novel human picornavirus in a stool sample from a pediatric patient presenting with fever of unknown origin. *J. Clin. Microbiol.* **2007**, *45*, 2144–2150.
86. Li, L. li, Liu, N.; Yu, J. mei, Ao, Y. yun, Li, S.; Stine, O. C.; Duan, Z. Analysis of Aichi virus and Saffold virus association with pediatric acute gastroenteritis. *J. Clin. Virol.* **2017**, *87*, 37–42.
87. Vandesande, H.; Edman, K.; Rondahl, E.; Falkeborn, T.; Serrander, L.; Lindberg, A. M. Saffold virus infection in elderly people with acute gastroenteritis in Sweden. *J. Med Virol.* **2020**, doi:10.1002/jmv.26452.
88. Bonanno Ferraro, G.; Mancini, P.; Veneri, C.; Iaconelli, M.; Suffredini, E.; Brandtner, D.; La Rosa, G. Evidence of Saffold virus circulation in Italy provided through environmental surveillance. *Lett. Appl. Microbiol.* **2019**, *70*, 102–108.
89. Fernandez-Cassi, X.; Timoneda, N.; Martínez-Puchol, S.; Rusiñol, M.; Rodriguez-Manzano, J.; Figuerola, N.; Bofill-Mas, S.; Abril, J.F.; Girones, R. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. *Sci. Total. Environ.* **2018**, *618*, 870–880.
90. Reuter, G.; Boros, Á.; Pankovics, P. Kobuviruses—A comprehensive review. *Rev. Med Virol.* **2011**, *21*, 32–41.
91. Ambert-Balay, K.; Lorrot, M.; Bon, F.; Giraudon, H.; Kaplon, J.; Wolfer, M.; Lebon, P.; Gendrel, D.; Pothier, P. Prevalence and genetic diversity of Aichi virus strains in stool samples from community and hospitalized patients. *J. Clin. Microbiol.* **2008**, *46*, 1252–1258.
92. Rivadulla, E.; Romalde, J.L. A Comprehensive Review on Human Aichi Virus. *Virol. Sin.* **2020**, 1–16, doi:10.1007/s12250-020-00222-5.
93. Khamrin, P.; Maneekarn, N.; Okitsu, S.; Ushijima, H. Epidemiology of human and animal kobuviruses. *VirusDisease* **2014**, *25*, 195–200.

94. Krishnamurthy, S.R.; Wang, D. Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. *Virology* **2018**, *516*, 108–114.
95. Leitão, G.A.A.; Olivares, A.I.O.; Pimenta, Y.C.; Delgado, I.F.; Miagostovich, M. P.; Leite, J. P. G.; Moraes, M. T. B. Human Bocavirus genotypes 1 and 2 detected in younger Amazonian children with acute gastroenteritis or respiratory infections, respectively. *Int. J. Infect. Dis.* **2020**, *95*, 32–37.
96. Haloschan, M.; Bettesch, R.; Görzer, I.; Weseslindtner, L.; Kundi, M.; Puchhammer-Stöckl, E. TTV DNA plasma load and its association with age, gender, and HCMV IgG serostatus in healthy adults. *Age* **2014**, *36*, 9716.
97. Focosi, D.; Antonelli, G.; Pistello, M.; Maggi, F. Torquetenovirus: The human virome from bench to bedside. *Clin. Microbiol. Infect.* **2016**, *22*, 589–593.
98. Wildenbeest, J.G.; Benschop, K.S.M.; Bouma-De Jongh, S.; Wolthers, K.C.; Pajkrt, D. Prolonged Shedding of Human Parechovirus in Feces of Young Children after Symptomatic Infection. *Pediatr. Infect. Dis. J.* **2016**, *35*, 580–583.
99. Netshikweta, R.; Chidamba, L.; Nadan, S.; Taylor, M.B.; Page, N.A. Molecular epidemiology of human bocavirus infection in hospitalized children with acute gastroenteritis in South Africa, 2009–2015. *J. Med Virol.* **2019**, *92*, 1124–1132.
100. De, R.; Liu, L.; Qian, Y.; Zhu, R.; Deng, J.; Wang, F.; Sun, Y.; Dong, H.; Jia, L.; Zhao, L. Risk of acute gastroenteritis associated with human bocavirus infection in children: A systematic review and meta-analysis. *PLoS ONE* **2017**, *12*, e0184833.
101. Paskey, A.C.; Frey, K.G.; Schroth, G.; Gross, S.; Hamilton, T.; Bishop-Lilly, K. A. Enrichment post-library preparation enhances the sensitivity of high-throughput sequencing-based detection and characterization of viruses from complex samples. *BMC Genom.* **2019**, *20*, 155.
102. Luchs, A.; Timenetsky, M. do C.S.T. Group A rotavirus gastroenteritis: Post-vaccine era, genotypes and zoonotic transmission. *Einstein* **2016**, *14*, 278–287.
103. Roux, S.; Hallam, S.J.; Woyke, T.; Sullivan, M.B. Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *eLife* **2015**, *4*, e08490.
104. Clooney, A.G.; Sutton, T.D.S.; Shkoporov, A.N.; Holohan, R.K.; Daly, K.M.; O'Regan, O.; Ryan, F.J.; Draper, L.A.; Plevy, S.E.; Ross, R.P.; et al. Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease. *Cell Host Microbe* **2019**, *26*, 764–778.e5.

CAPÍTOL 3: ESTUDI DEL VIROMA D'ESPÈCIES DE PEIXOS ATLÀNTICS DESTINATS AL CONSUM HUMÀ

Article 4: El viroma sense explorar de dos espècies peixos de la costa Atlàntica: Contribució de la seqüenciació massiva en la virologia de peixos

"The Unexplored Virome of Two Atlantic Coast Fish: Contribution of Next-Generation Sequencing to Fish Virology."

Filipa-Silva A., Parreira R., Martínez-Puchol S., Bofill-Mas S., Barreto Crespo M.T., Nunes M.

Foods, 9, 1634, doi:10.3390/foods9111634.

Molt del coneixement que tenim dels virus està enfocat en aquells causants de malalties en humans o que generen impacte econòmic al infectar alguns animals i plantes. En canvi, aquesta classificació reflecteix només una petita part de la biosfera. En aquest context, s'ha col·laborat amb el *Instituto de Biologia Experimental e Tecnológica* (Lisboa, Portugal) en l'estudi, mitjançant eines de seqüenciació massiva, del viroma de dues espècies de peixos de pesca i aqüicultura, sorell i orada, de les regions centre i sud de Portugal.

S'han detectat seqüències víriques d'espècies amb potencial patogènic per peixos, crustacis i humans de les famílies *Astroviridae*, *Nodaviridae*, *Hepadnaviridae*, *Birnaviridae*, *Caliciviridae* i *Picornaviridae*. A més, les seqüències trobades amb més abundància en les dues espècies de peixos corresponen a bacteriòfags. Comparant entre aquestes espècies de peixos d'estudi, els sorells han presentat un major nombre de virus en els teixits analitzats.

Els resultats obtinguts mostren que els peixos disponibles per al consum poden contenir una elevada diversitat de virus, molts d'ells desconeguts. Mentre que la transmissió de virus de peixos a humans és improbable, la

troballa d'alguns virus patògens per humans suggereix un potencial risc en seguretat alimentària.

foods

MDPI

Article

The Unexplored Virome of Two Atlantic Coast Fish: Contribution of Next-Generation Sequencing to Fish Virology

Andreia Filipa-Silva [1,2], Ricardo Parreira [3], Sandra Martínez-Puchol [4,5], Sílvia Bofill-Mas [4,5], Maria Teresa Barreto Crespo [1,2] and Mónica Nunes [1,2,*]

1 ITQB NOVA, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras, Portugal; andreiasilva@ibet.pt (A.F.-S.); tcrespo@ibet.pt (M.T.B.C.)
2 iBET, Instituto de Biologia Experimental e Tecnológica, Apartado 12, 2781-901 Oeiras, Portugal
3 Global Health and Tropical Medicine (GHTM) Research Center, Unidade de Microbiologia Médica, Instituto de Higiene e Medicina Tropical (IHTM), Universidade Nova de Lisboa (NOVA), 1349-008 Lisboa, Portugal; ricardo@ihmt.unl.pt
4 Laboratory of Viruses Contaminants of Water and Food, Genetics, Microbiology & Statistics Department, Universitat de Barcelona, 08028 Barcelona, Catalonia, Spain; smpuchol@gmail.com (S.M.-P.); sbofill@ub.edu (S.B.-M.)
5 The Water Research Institute (idRA), Universitat de Barcelona, 08001 Barcelona, Catalonia, Spain
* Correspondence: mnunes@ibet.pt

Received: 20 October 2020; Accepted: 6 November 2020; Published: 9 November 2020

Abstract: Much of the knowledge on viruses is focused on those that can be propagated using cell-cultures or that can cause disease in humans or in economically important animals and plants. However, this only reflects a small portion of the virosphere. Therefore, in this study, we explore by targeted next-generation sequencing, how the virome varies between Atlantic horse mackerels and gilthead seabreams from fisheries and aquaculture from the center and south regions of Portugal. Viral genomes potentially pathogenic to fish and crustaceans, as well as to humans, were identified namelyese included *Astroviridae*, *Nodaviridae*, *Hepadnaviridae*, *Birnaviridae*, *Caliciviridae*, and *Picornaviridae* families. Also bacteriophages sequences were identified corresponding to the majority of sequencese detected, with *Myoviridae*, *Podoviridae*, and *Siphoviridae*, the most widespread families in both fish species. However, these findings can also be due to the presence of bacteria in fish tissues, or even to contamination. Overall, seabreams harbored viruses from a smaller number of families in comparison with mackerels. Therefore, the obtained data show that fish sold for consumption can harbor a high diversity of viruses, many of which are unknown, reflecting the overall uncharacterized virome of fish. While cross-species transmission of *bonafide* fish viruses to humans is unlikely, the finding of human pathogenic viruses in fish suggest that fish virome can be a potential threat regarding food safety.

Keywords: virome diversity; metagenomics; Atlantic horse mackerel; gilthead seabream; health-threat

1. Introduction

As an important source of food and revenue for millions of people globally, it was estimated that fish production in 2018 reached 179 million tons, 87 percent of which were used directly for human consumption [1]. While the demand for fish may be rising, viral diseases are a major factor constraining fish trade and aquaculture production, affecting the development of fish industry with substantial economic losses worldwide [2,3]. Furthermore, the increasing demand for seafood with the

subsequent growth of aquaculture systems, and the impact of climate change, offer new opportunities for the transmission of both novel and previously characterized pathogens, namely viruses [3].

Marine environments are known for their viral diversity and richness. Bacteriophages, commonly found in marine and freshwater systems, play an important role in the aquatic environments balance as predators of bacteria, representing one of the most abundant organisms in fish [4–6]. In addition to the presence of bacteriophages, recent studies suggest that fish harbor a greater number of viruses than any other class of vertebrates, including multiple families of RNA viruses previously thought to infect only mammals [3,7]. For instance, the discovery of hepadnaviruses and filoviruses in fish implies that these viruses have ancient vertebrate origins, and that their evolution possibly required more host jumping than previously realized, involving a possible move from aquatic to terrestrial vertebrates [3,7,8].

Valuable insights into virus ecology and evolution are provided when using metagenomics approaches [3,7], with benefits for aquaculture [3] and food security. For instance, information about outbreaks associated with human pathogenic viruses in fish is scarce. However, the Centers for Disease Control and Prevention (CDC) compiles searchable lists of viral outbreaks that can be retrieved from the National Outbreak Reporting System [9]. In this system, regarding the years between 2008 and 2018, fish alone were responsible for 14 noroviruses (NoV) outbreaks, totaling 177 human infections and 3 hospitalizations.

Much of our present knowledge of fish viruses is focused either on those that act as agents of disease for economically relevant species, or those that can be easily isolated in cell culture, representing a small proportion of the overall viral diversity [10]. To circumvent this problem, metagenomic approaches provide the possibility for an in-depth characterization of the molecular diversity of viruses present in a range of environments, potentially revealing the entire virus composition associated with an individual or *taxa*—i.e., its virome—even when they are not associated with evident disease [7,11,12]. Metagenomic-based studies and characterization of more viruses can also offer important new data about evolutionary processes, as well the factors that may mediate differences in virus composition between different species of fish, and the frequency with which viruses jump over species borders [3]. Moreover, it can provide insightful information on how viral community composition changes, which depends on regional and local processes, including interactions occurring between fish, as well as environmental factors such as temperature, salinity, or dissolved oxygen, that determine fish distribution [13,14]. Virome diversity in fish populations is also affected by how viral transmission between fish occurs: primarily through feces, contaminated water and/or broodstocks, culture of multiple fish species in close proximity, or the use of unsterilized fish products (e.g., feed) in aquaculture productions [15].

Therefore, in this study, the diversity of viruses associated with two of the most highly captured, farmed and consumed fish species in Europe was evaluated using a metagenomic approach. We aimed to determine: (i) how virus composition varies between species; (ii) how virus composition varies between fisheries from different regions and aquaculture within the same species; (iii) whether a large and dense host population is associated with a greater number of viruses when compared to a more solitary counterpart; (iv) and, if the analyzed fish virome could pose a risk regarding food safety. To address these goals, targeted next-generation sequencing (NGS) was used to evaluate the virome of wild gilthead seabreams (*Sparus aurata*) and of Atlantic horse mackerels (*Trachurus trachurus*) from fisheries of Portugal, namely from Peniche and Algarve regions, as well as farmed gilthead seabreams from Algarve.

2. Materials and Methods

2.1. Fish Sample Collection and Nucleic Acids Extraction

Fish from two distinct regions of Portugal with different population density and ocean temperature, were analyzed during this study (Figure 1). A total of 15 gilthead seabreams caught along the coast of Peniche (n = 5, west/central Atlantic coast of Portugal), Algarve (n = 5, Portuguese southern Atlantic

coast) and from an aquaculture farm in Algarve (n = 5) were purchased *post-mortem* from the respective commercial fisheries (Table 1). Additionally, 20 Atlantic horse mackerels were analyzed *post-mortem* from the same fisheries (Peniche, n = 10; Algarve, n = 10) (Table 1).

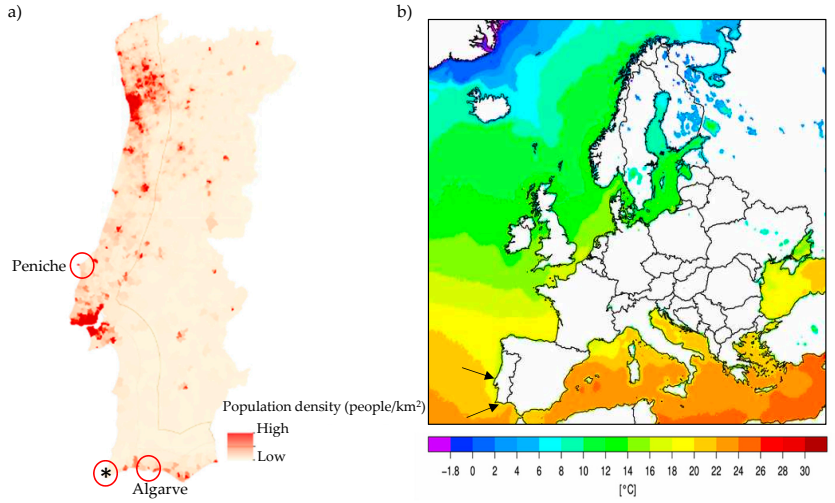


Figure 1. Localization of the fish sampling regions studied. (**a**) Population density on Portuguese coast (adapted) [16]; * indicates aquaculture sampling site. (**b**) Sea surface temperature during the summer of 2017 in Europe. Black arrows indicate the sampling sites on the Portuguese coast (adapted) [17].

Table 1. Fish samples acquired during this study, their source, and fishery type.

Species	Source	Fishery Type	N° of Specimens
Sparus aurata (gilthead seabream)	Peniche fish market	Wild fisheries	5
Sparus aurata (gilthead seabream)	Algarve fish market	Wild fisheries	5
Sparus aurata (gilthead seabream)	Algarve fish market	Aquaculture	5
Trachurus trachurus (Atlantic horse mackerel)	Peniche fish market	Wild fisheries	10
Trachurus trachurus (Atlantic horse mackerel)	Algarve fish market	Wild fisheries	10

Fish have been divided by size and species. Larger fish, namely gilthead seabreams were gathered in pools of five specimens, while the smaller fish, Atlantic horse mackerel, in pools of 10. One internal tissue (liver—detoxification tissue and local of viral replication), and two external tissues (skin—contact tissue exposed to several environmental contaminants, and gills—portal of entry of many viral pathogens) were selected considering their characteristics. Approximately 2.0 g of tissue were chopped individually by using a sterilized razor blade to minimize external contamination. The dissection material was decontaminated with a solution of bleach a 10% and sterilized water between each sample manipulation. The tissues were homogenized in 10 mL of TNE buffer (50 mM Tris-HCl, 100 mM NaCl, 0.1 mM EDTA, pH 7.6) using a Precellys Evolution Homogenizer (Bertin Instruments, France). The homogenates previously obtained were centrifuged 10 min at 2000 rpm at 4 °C to remove any potential tissue debris, and the supernatant of each homogenized tissue was divided in triplicates and used for nucleic acids extraction. From the 45 samples, nucleic acids extraction was carried out from 250 µL of clarified supernatant, using NZYol (NZYTech, Portugal), following the manufacturer's instructions. Nucleic acids were dissolved in 30 µL of DEPC-water, and the concentration and purity, of the obtained extracts determined using a NanoDrop 1000 spectrophotometer. The extracted nucleic acids were stored at −80 °C until further use.

2.2. Library Preparation and Sequencing

For all samples, libraries were prepared using the protocol described by Fernandez-Cassi et al. [18]. Briefly, in order to detect both RNA and DNA viruses, total RNA was retrotranscribed into cDNA using Invitrogen Superscript IV (Life technologies, Austin, TX, USA) and primer A, which is composed of 17-nucleotide specific sequence followed by a random nonamer for random priming A-(5′-GTTTCCCAGTCACGATANNNNNNNNN-3′) [19]. A cDNA complementary strand was synthesized using Sequenase 2.0 (USB/Affymetrix, Cleveland, OH, USA). To obtain enough DNA for sequencing library preparation, a pre-amplification PCR was performed using primer B (5′-GTTTCCCAGTCACGATA-3′) and AmpliTaqGold (Life technologies, Austin, TX, USA). The temperature profile for this reaction was: 10 min at 95 °C as hot start, followed by 30 cycles of 30 s at 94 °C for denaturation, 30 s at 50 °C for annealing, and 1 min at 72 °C for extension. Finally, an extension at 72 °C for 10 min was used. To remove excess primers and dNTPs, PCR products were cleaned and concentrated using the Zymo DNA Clean and Concentrator kit (Zymo Research, USA). The concentration of the obtained DNA samples was determined using Qubit 2.0 (Life Technologies, USA), and for each sample, libraries were constructed using the KAPA HyperPlus Library Preparation kit (Roche, Switzerland) according to the manufacturer's instructions.

Libraries were captured using VirCapSeq-VERT capture Panel (Roche, Switzerland). This panel consists of approximately two million probes, covering the genomes of 207 viral *taxa* known to infect vertebrates, thus enabling the detection of viral sequences in complex samples [20]. The samples prepared with the HyperCap Target Enrichment Kit (Roche, Switzerland) and the HyperCap Bead Kit (Roche, Switzerland), were hybridized with the VirCapSeq-VERT probes at 47 °C for 20 h. Immediately after the hybridization, the captured DNA samples were recovered with the Capture Beads (HyperCap Bead Kit, Roche) using a magnetic particle collector, and cleaned. The captured DNA, still bead-bounded, was purified using an ligation mediated PCR (LM-PCR). This post-capture PCR was purified and then sequenced in two runs using an Illumina MiSeq (2 × 300 bp) device, producing paired-end reads.

All reactions included non-template negative controls to rule out the possibility of positive amplification results due to external contamination.

2.3. Bioinformatic Pipeline

Bioinformatic processing and taxonomical assignment of the pair-end reads obtained as described above were analyzed using Genome Detective web-based software (https://www.genomedetective.com/) [21]. Briefly, low-quality reads were filtered and adapters trimmed using Trimmomatic [22]. Viral reads were selected using the DIAMOND protein-based alignment and non-viral sequences discarded. The obtained viral reads from Genome Detective were assembled with metaSPAdes [23] and taxonomically classified with NCBI-BLASTn and BLASTx tools to search for candidate reference sequences against the NCBI RefSeq virus database. In addition, assembled unused viral contigs and discovery viral contigs from Genome Detective software were analyzed using NCBI-BLASTn and tBLASTx tools against NCBI RefSeq viral databases using Python scripts. Only viral contigs with 85% identity and an e-value threshold below 10^{-3} (used as a cut-off) were used in this study.

2.4. Dataset Compilation and Phylogenetic Analysis

Nucleotide (nt) sequences used for the preparation of the different sequence-datasets were selected among those previously deposited in the GenBank database (Supplementary Materials Tables S1 and S2), on the proviso that they would be representative of (i) each of the previously described species with (ii) a significant sequence overlap with the sequences obtained in the course of this study, in order to maximize the number of unambiguously aligned nt positions in each sequence alignment. For phylogenetic analysis, multiple alignments of nt sequences were constructed with the iterative G-INS-I method as implemented in MAFFT vs. 7 [24] followed by their edition using GBlocks [25]. Phylogenetic trees were constructed using the Maximum Likelihood (ML) optimization criterium and the best fitting evolutionary model (GTR+Γ+I; GTR—general time reversal, Γ—Gamma distribution, I—proportion of invariant sites), as suggested W-IQ-TREE [26]. Phylogenetic reconstructions were carried out using either the Mega X software [27] run on a Linux server or W-IQ-TREE (version 2.1.2 MacOSX) [26] running on a personal computer, and the stability of the obtained ML tree topologies assessed by bootstrapping with 1000 re-samplings of the original sequence data. The sequence data files were deposited in GenBank within the BioProject with the SRA accession number PRJDB10531.

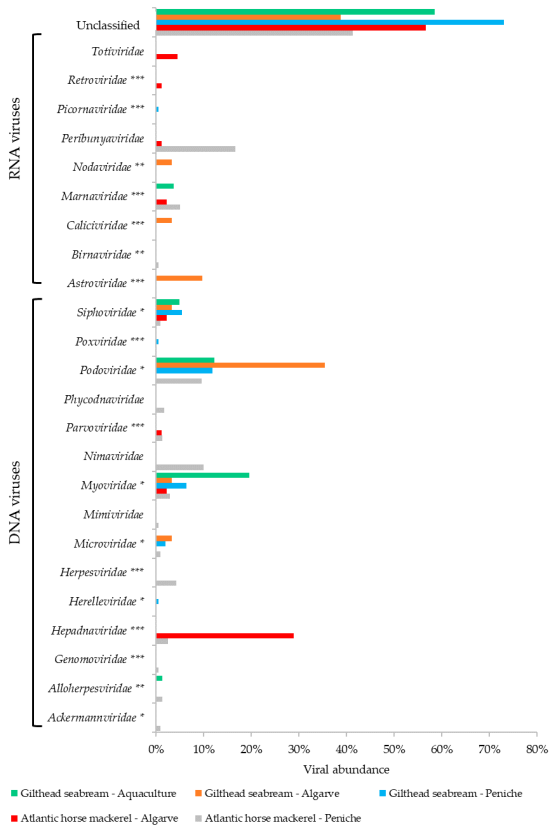
3. Results

3.1. Viral Sequences Diversity

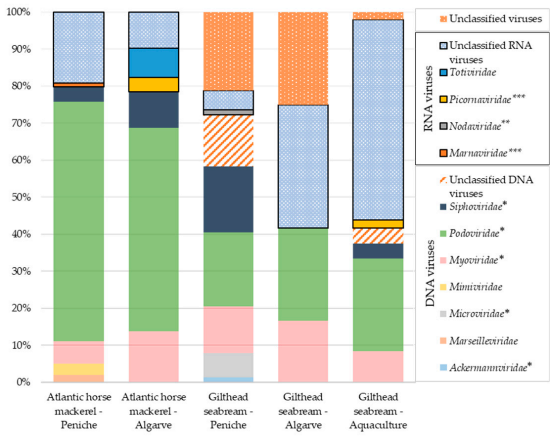
3.1.1. Viral Sequences Diversity—Global Results

We performed metagenomics analysis to characterize tissue-specific virome of two species of ray-finned (Actinopterygii) bony fish: the gilthead seabream and the Atlantic horse mackerel. This analysis was carried out using wild-caught specimens of gilthead seabreams and Atlantic horse mackerels, available from fisheries on the west (Peniche) and south (Algarve) Atlantic coast of Portugal, and farmed specimens of gilthead seabreams from the south (Algarve). To carry out this study, nucleic acids were extracted from the liver, gills, and skin of these animals, which was then pooled in libraries for next-generation sequencing. Overall, 1,216,428 viral reads were obtained that were assembled, *de novo*, into viral contigs for metagenomic analysis.

In total, the analysis revealed viral sequences that could be assigned to 25 different viral families (Figure 2). The analytical approach used in this study was able to identify both RNA and DNA viruses based on sequence similarity (Figure 2). The majority of the viral contigs were assigned to DNA viral families *Ackermannviridae*, *Alloherpesviridae*, *Genomoviridae*, *Hepadnaviridae*, *Herelleviridae*, *Herpesviridae*, *Marseilleviridae*, *Microviridae*, *Mimiviridae*, *Myoviridae*, *Nimaviridae*, *Parvoviridae*, *Phycodnaviridae*, *Podoviridae*, *Poxviridae*, and *Siphoviridae*. The remaining viral contigs were assigned to families of RNA viruses, including *Astroviridae*, *Birnaviridae*, *Caliciviridae*, *Marnaviridae*, *Nodaviridae*, *Picornaviridae*, *Peribunyaviridae*, *Retroviridae*, and *Totiviridae*. This analysis also revealed an unclassified taxonomic fraction of contigs (around 20%), with 3% of unclassified DNA viruses, 10% of unclassified RNA viruses, and 7% of viral sequences without any assignment to a known viral *taxon* after BLAST analysis [28,29].



(**a**)



(**b**)

Figure 2. Blast results concerning viral diversity within each fish species and region. (**a**) Viral diversity in each fish species per region (including those contigs assigned as unclassified even within a family). (**b**) Viral diversity of unclassified viral contigs of each species per region. Bacteriophages families are pointed out with *, whereas pathogenic families for fish/other vertebrates and humans/other vertebrates are marked with ** and ***, respectively.

3.1.2. Atlantic Horse Mackerel and Gilthead Seabream Viral Sequence Diversity

Our first analysis focused on parallel assessment of the viral sequence diversity between the two species selected, without considering their geographic or fishery/aquaculture origins. A comparison of viral abundance across the different studied organs for both species under analysis, revealed several differences. In total, viral sequences were more abundant in skin samples (44%) than in the liver (38%) or gills (18%). Bacteriophages from families *Myoviridae*, *Podoviridae*, and *Siphoviridae* were the viruses most frequently found in both fish species (Figures 2 and 3). Podoviruses, which comprise short noncontractile-tailed bacteriophages with dsDNA genomes, were the most abundant (31%), totaling 19% and 8% of the total viral sequences associated with the liver and skin, respectively. The remaining 4% of podoviruses sequences were found in the gills. In the liver tissue samples prepared from Atlantic horse mackerel, the total virome associated with the *Podoviridae* family comprised 48% of the sequences analyzed, while the equivalent value for the gilthead seabream was only 14% (Figure 3). As for the gilthead seabream, we observed that almost 26% of the total virome for *Podoviridae* was associated with the skin tissues, whereas in the case of the Atlantic horse mackerel, the corresponding value was only 5%.

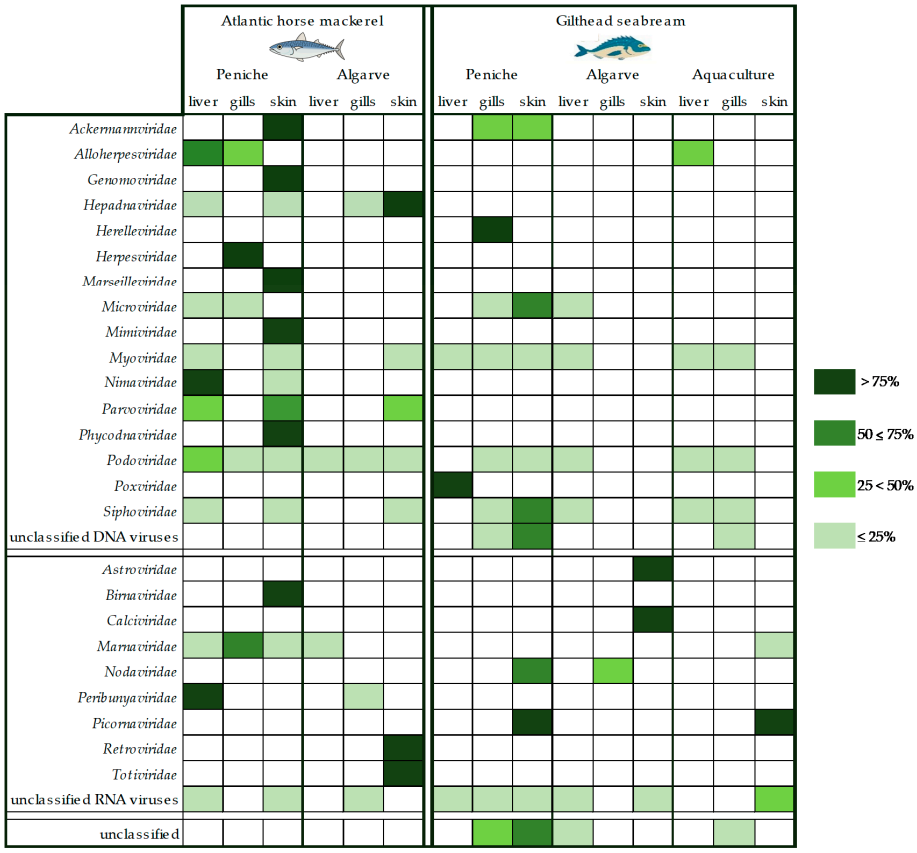


Figure 3. Viral abundance for each fish species, region, and tissue type normalized for each viral family (the taxonomic classification used is based on the Taxonomy Database from NCBI).

Additionally, the virome of mackerels from Peniche was apparently characterized by a higher, and more diverse number of viral sequences from different *taxa*, when compared with mackerels and

seabreams from Algarve region, with viral sequences distributing among 17 viral families across the different organs (Figure 3). Although gilthead seabreams, in general, seemed to harbor a less diverse group of viral families regardless of the origin of the tissues under analysis, they displayed a higher abundance of unclassified viral sequences (Figures 2b and 3).

Our data also revealed the presence of viral sequences from astroviruses, birnaviruses, caliciviruses, hepadnaviruses, nodaviruses, herpesviruses, and picornaviruses, all of which include representatives that are known to infect not only marine animals, such as fish and crustaceans, but also humans. For instance, *Birnaviridae*, *Hepadnaviridae*, and *Herpesviridae* viral sequences were identified in Atlantic horse mackerels, while others, indicating the presence of other families (*Astroviridae*, *Caliciviridae*, and *Nodaviridae*) were detected in libraries prepared from the skin and gills of gilthead seabreams.

3.1.3. Viral Sequence Diversity: Fisheries vs. Aquaculture

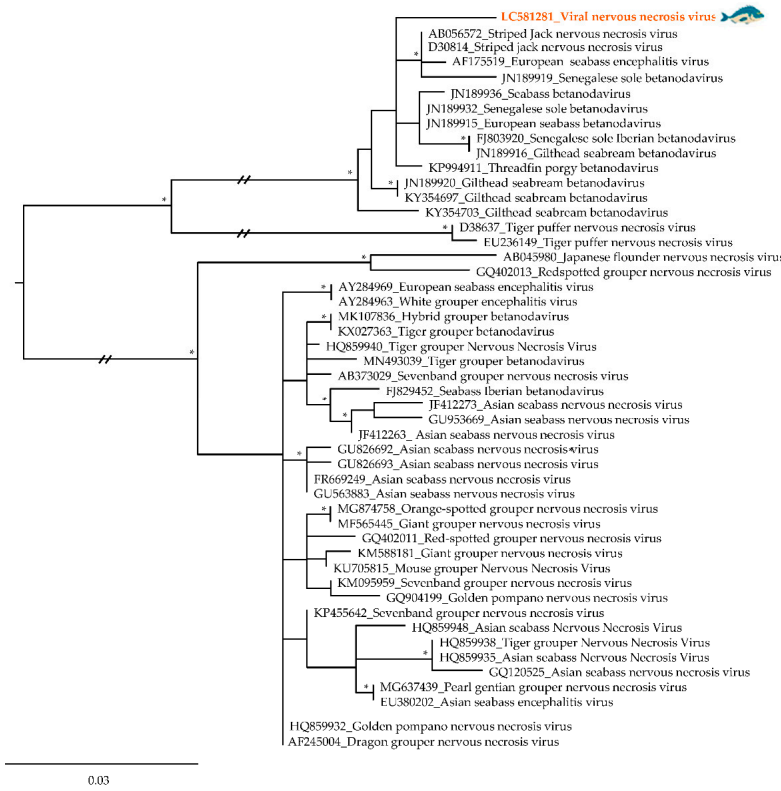
An uneven distribution of viral diversity between fisheries from Peniche and Algarve, as well as from an aquaculture environment was observed. From the total of viral sequences obtained, the most abundant ones were associated with fish specimens from Peniche, with a total of 29% associated with the skin, followed by the liver (28%) and the gills (11%). The same pattern was observed in specimens from Algarve region, with 9%, 8%, and 2% of viral sequences being associated with the skin, liver, and gills, respectively. In aquaculture, 5% of viral sequences were detected in sequencing libraries prepared from skin samples, 3% from the liver and 5% from the gills.

As expected, and regardless the geographic origin of the specimens under analysis, bacteriophages were the most commonly found viruses, with *Podoviridae* being the most widespread family representing 21% and 7% of the total viral sequences detected in samples from Peniche and Algarve, respectively. In specimens from the Peniche region, we observed that 39% of the total liver virome was composed of podoviruses, although these could also be detected in libraries prepared from the skin (22%) and gills (7%). A similar trend was observed in specimens collected from Algarve, but the values associated with the liver (16%), the skin (4%), or the gills (2%) were lower.

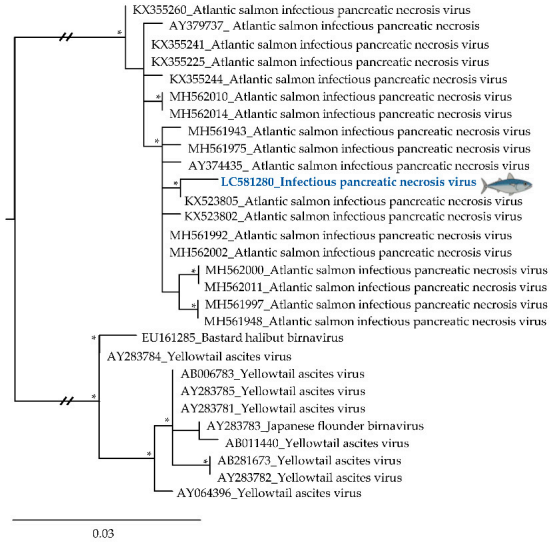
It is notable that the Atlantic horse mackerel from Peniche, a more shoaling fish, had an associated higher number of distinct viral families (17 families) when compared to the gilthead seabream, a more solitary fish, with only 9 viral families identified (Figure 3). In contrast, Atlantic horse mackerels and gilthead seabreams from Algarve region presented a smaller number of viral families (9 and 7, respectively). Additionally, gilthead seabream from an aquaculture farm displayed the less diverse virome with our analysis, with nucleotide sequences allocated to only 6 viral families, half of which included bacteriophages. However, the unclassified taxonomic fraction associated with farmed gilthead seabream is substantial (56%).

3.2. Phylogenetic Relationships of Relevant Viral Pathogenic Families

For further characterization of some relevant viral sequences from known human and fish pathogenic viruses from selected viral families, a phylogenetic analysis was carried out using the ML optimization criterium, and different sequence-datasets. The inferred phylogenetic trees (Figures 4 and 5; Supplementary Materials Figures S1 and S2) revealed the phylogenetic relationships between the obtained sequences and their closest relatives, identified by sequence similarity searches at GenBank. All trees have been rooted at mid-point.

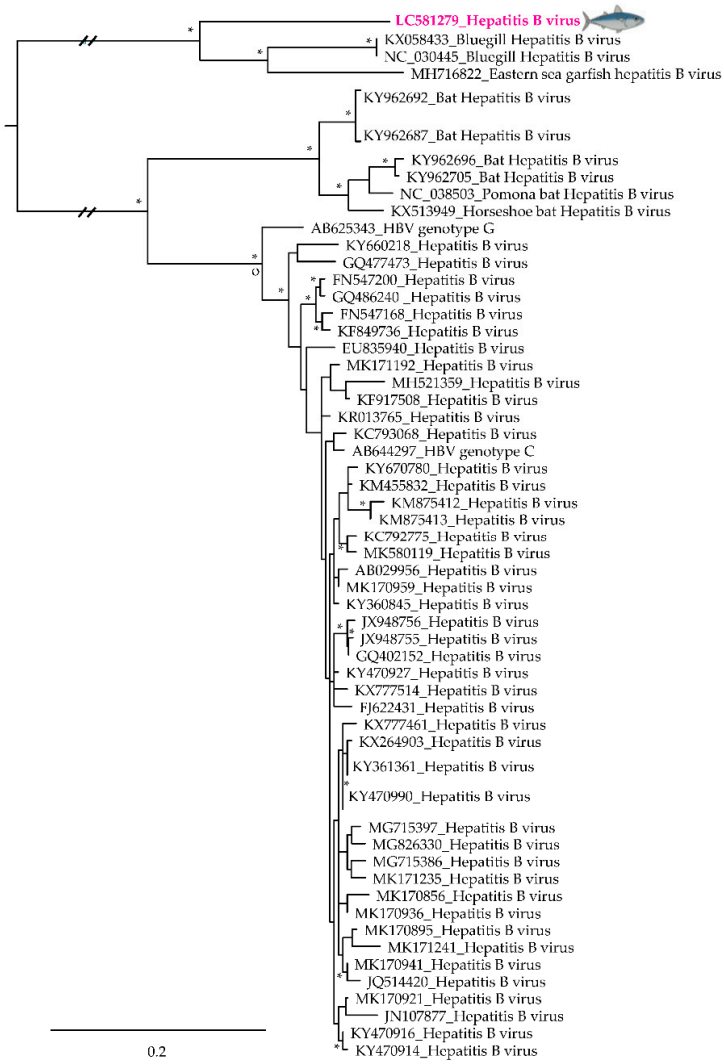


(**a**)



(**b**)

Figure 4. *Cont.*

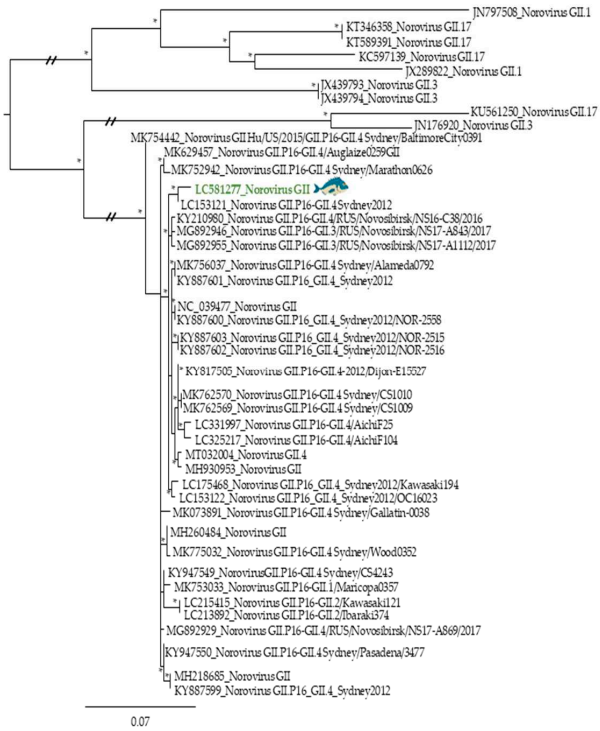


(**c**)

Figure 4. Phylogenetics trees of common fish pathogenic viruses: (**a**) *Nodaviridae*—partial sequence of Viral Nervous Necrosis Virus amplified from gilthead seabream gills from Algarve, (**b**) *Birnaviridae*—partial sequence of Infectious Pancreatic Necrotic Virus detected in Atlantic horse mackerel skin from Peniche, (**c**) *Hepadnaviridae*—partial sequence of Hepatitis B virus obtained from Atlantic horse mackerel gills from Algarve. The viral sequences analyzed in this study are colored and highlighted by a fish. All viral sequences are identified by their accession number and name. At specific branch nodes, bootstrap branches ≥0.60 are displayed by *; branch node point out with a circle referred to human viral sequences.

(**a**)

Figure 5. *Cont.*



(**b**)

Figure 5. Phylogenetics trees of human pathogenic viruses identified in this study: (**a**) *Picornaviridae*—partial sequence of Rhinovirus A obtained from gilthead seabream skin from Peniche; (**b**) *Caliciviridae*—partial sequence of Norovirus GII amplified from gilthead seabream skin from Algarve. The viral sequences analyzed in this study are colored and highlighted by a blue fish. All viral sequences are identified by their accession number and name; all viral sequences used are from human hosts. At specific branch nodes bootstrap values ≥0.60 are displayed by *.

Among the viral sequences identified, Viral Nervous Necrosis Virus (VNNV, accession number LC581281), one of upmost importance pathogenic nodaviruses to fish, which was detected in gilthead seabream from the Algarve, was placed, as expected, in a cluster with other betanodaviruses but branched separately from the other sequences previously identified (Figure 4a and Supplementary Materials Figure S1a). This group also included other viral reference sequences, previously detected in specimens of striped jack, European seabass, Senegalese sole, and gilthead seabream from Europe and Asia.

A birnavirus sequence (accession number LC581280), detected in Atlantic horse mackerel from Peniche (Figure 4b; Supplementary Materials Figure S1b), was placed, as expected from BLAST search results, with other Infectious Pancreatic Necrosis Viruses (IPNV) amplified from salmonids. Specifically, the closest homolog to the Atlantic horse mackerel IPNV sequence identified in this study corresponded to a putative virus detected in an Atlantic salmon (*Salmo salar*) caught in Chile. These sequences encoded putative proteins sharing 99.7% of amino acid identity. Interestingly, hepadnaviruses (which include Hepatitis B viruses (HBV) from various sources) were also identified from the gills of Atlantic horse mackerel but in this case from Algarve (accession number LC581279; Figure 4c and Supplementary Materials Figure S1c). The Atlantic horse mackerel HBV sequence here described branched separately from all the others in the dataset but was clearly

closely related with other hepadnaviruses from fish, specifically, blue gill HBV, forming a separated cluster from the HBV that infect mammals. Additionally, a sequence amplified from the skin of gilthead seabream (accession number LC581278), clustered within a large monophyletic cluster among Rhinovirus A viral sequences from diverse sources, including those identified in humans, and clearly segregated away from the genetic lineages that include Rhinovirus B and Rhinovirus C sequences (Figure 5a and Supplementary Materials Figure S2a). The sequence detected shared 99% similarity with Rhinovirus A sequences detected in humans from Uganda (accession numbers MH685691 and MH685686). Another viral sequence obtained from gilthead seabream from Algarve (accession number LC581277) clustered in a major cluster grouping strains of NoV GII (Figure 5b and Supplementary Materials Figure S2b), specially NoV GII.P16-GII.4, phylogenetically close to the recombinant NoV GII.P16-GII.4 Sydney 2012 variant (Figure 5b), one of the most common NoV strains associated with outbreaks of gastroenteritis worldwide.

4. Discussion

Viruses can be found in every single environment on Earth, but their importance is more evident in oceans, where they are known to be the reservoir of most of the genetic variety [5]. To explore the viral diversity of two important commercialized fish species, we analyzed viral sequences obtained from liver, gills, and skin tissue samples using a metagenomic approach. As these species included shoaling—Atlantic horse mackerel, and solitary—gilthead seabream specimens, the obtained data allowed us to tentatively address how fish population density might affect virus composition.

The presence of unknown virus genomes in fish samples was expected since previous studies on aquatic viral communities have shown that the unknown (orphan) viral fraction (i.e., without homology to any sequence available in the databases) could be substantial, ranging between 20% to 99% [30,31]. In fact, the analysis of virus open reading frames (ORFs) showed that a large portion were orphans, with these unclassified ORFs being more frequently identified in viruses than in bacteria (30% and 9%, respectively) [32].

In addition, nucleotide sequences identifying viral families associated with vertebrate infections such as *Astroviridae*, *Nodaviridae*, *Hepadnaviridae*, *Birnaviridae*, *Caliciviridae*, and *Picornaviridae* were detected during this analysis (Figures 2 and 3). Although, astroviruses have only recently been discovered in fish [7], they are known to persist in aquatic environments and have been detected in seabirds [33], amphibians [7], and mammals [34], where these viruses are a frequent cause of intestinal diseases [34]. Nodaviruses can cause neural necrosis, encephalopathy, or retinopathy in fish, being associated with behavioral abnormalities and high mortality, posing significant problems to marine aquaculture [35]. The analysis of a VNNV sequence detected in gilthead seabream gills, revealed that it was closely related to Striped jack nervous necrosis virus, the type species of the betanodaviruses. While, our sequence forms a separate branch with other betanodaviruses (Figure 4a), also detected in gilthead seabream, it segregates in a larger group of viral sequences that include others found in common fish species such as European seabass and Senegalese sole [36]. Interestingly, VNNV is endemic in the Mediterranean Sea, with several reports in fish samples [37–39], being in direct communication with the Atlantic Ocean through the Strait of Gibraltar, located at the south of Spain near the Algarve region, where part of this study was focused. In addition, other nodaviruses were also identified, namely the *Macrobrachium rosenbergii* nodavirus or *Penaeus vannamei* nodavirus, both pathogens of freshwater crustaceans that pose a threat to food security, and cause significant economic losses in the aquaculture industries [40].

Until recently, the known host range of hepadnaviruses was limited to mammals and birds [8]. However, previous studies showed that fish carry a remarkable diversity of hepadnaviruses [3,8]. Notably, the Atlantic horse mackerel Hepatitis B virus forms a sister group with other hepadnaviruses recently discovered in fish [8], segregating away from mammalian hepadnaviruses (Figure 4c and Supplementary Materials Figure S1c) which are known to cause infections that have the potential to lead to both severe chronic liver disease and hepatocellular carcinoma in humans [8].

IPNV is an important fish pathogen and the protype virus of the *Birnaviridae* family [41]. It has a worldwide distribution, and normally is associated with an acute and contagious disease that causes distended abdomen, aberrant swimming, darkened pigmentation, and necrotic lesions in internal organs. In this study, an IPNV sequence was detected in Atlantic horse mackerel skin tissue, corresponding to a distinct branch on a phylogenetic tree, separating it from other aquabirnaviruses such as the Yellowtail ascites virus (Figure 4b), the causative agent of ascites and responsible for serious losses in the fish-farming industry in Japan [42]. This sequence forms a group with IPNV sequences amplified from those clustering with others from Europe and South America specimens of salmonids [43]. This result is not surprising taking into account a report that supports the association between fish eggs from Europe and the disease transmission across Chilean farms [44].

Human NoV genomic sequence was also identified in the skin of gilthead seabream from the Algarve. Although, its presence does not mean that the virus is active or infectious, we cannot exclude the hypothesis that this virus can also be present in the fish flesh that can be consumed raw or undercooked in several meals, possessing a potential risk to human health. Noroviruses are the leading cause of viral gastroenteritis worldwide, and are estimated to cause 677 million cases and ≈ 210,000 human deaths every year [45].

Furthermore, the detection of NoV genomes in these fish samples can indicate that the water was contaminated with human feces, probably due to sewage pollution, or that fish were contaminated during handling, processing, or preparation. While the small size of the sequence under study limits the breath of observations drawn from its analysis, its place in a phylogenetic tree revealed that it branched with a group containing essentially recombinant strains of NoV GII.4 (Figure 5b) which caused six food-related outbreaks since 1995 [46]. In particular, the NoV sequence identified in this study is closely related to the recombinant GII.P16/GII.4 Sydney 2012. Interestingly, in the last decades, the GII.4 variant dominated NoV infections, being responsible for 60–80% of all NoV outbreaks. Therefore, in the last years, we observed the emergence of a novel GII.4 recombinant virus responsible for several outbreaks worldwide [45,46] that retained the Sydney 2012 capsid-coding sequence but acquired a new structural region (GII.P16/GII.4 Sydney 2012) [45,46].

Picornaviruses such as rhinoviruses, or other small non-enveloped viruses with RNA genomes such as noroviruses (*Caliciviridae*), have a compact capsid structure that has been selected to allow them to withstand a long-stand preservation in harsh environments [3]. A closer analysis of the rhinovirus sequence amplified from a sample of gilthead seabream skin from Peniche showed that this viral sequence branched separately from Rhinovirus B and C (Figure 5a). It formed a cluster mainly with Rhinovirus A serotypes 54 and 98, being phylogenetically close to rhinovirus sequences identified from stools in African children with diarrhea [47]. Although often ignored, human rhinoviruses are the most frequent causes of respiratory tract infections with severe disease manifestations in patients suffering from bronchiolitis and asthma [48].

Highly abundant viral sequences associated with the fish species studied were assigned to bacteriophage families, which are known as the most abundant viruses in the biosphere, occurring in large numbers in freshwater, soil, sewage, and marine environments [4]. Their presence in the environment suggests that they play an important role in the ecology and evolution of diverse ecosystems [5]. In particular, several bacteriophages from freshwater and marine ecosystems seem to play important roles not only in the equilibrium of the ecosystems as predators of bacteria, but also in a broad range of ecological processes such as carbon cycle [5,6].

The genomes of myoviruses, podoviruses, and shipoviruses were identified as the most widespread viral sequences found among the samples analyzed during this study. As expected, they were frequently associated with exterior organs (the skin), or tissues that were exposed to the external environment (the gills). Surprisingly, our data also suggest their presence in the liver. Considering that internal and external portions of the livers were used, it is likely that the majority of bacteriophage sequences detected in our samples would correspond to viruses found on the liver surface, in the abdominal cavity. Gut microbiota consists of a plethora of microorganisms including bacteria, fungi, and archaea, and a

myriad of viruses including bacteriophages and others infecting eukaryotic cells. Bacteriophages have been found in humans and other animal, inhabiting the oral, gastrointestinal and respiratory tracts, in the urine and blood, the latter being considered an sterile environment in heathy individuals [49,50]. Moreover, the liver is an important component of the reticuloendothelial system [51], being one of the organs that filter a variety of foreign elements that are in circulation, including phages [51], being identified as one of the sites with the highest phage accumulation [30,31,52].

It has previously been shown that species of fish forming large and dense groups are characterized by high-level contact rates and exhibit higher viral diversity compared to their more solitary counterparts [3,53–55]. This hypothesis supports classic epidemiological theories that larger populations with higher contact rates have an increased likelihood of acquiring, and transmitting, viruses [56,57]. Interestingly, in our study, this only seems valid for the specimens caught in the region of Peniche since Atlantic horse mackerels and gilthead seabreams from Algarve fisheries do not reveal this pattern. These differences in the distribution of the viral families between the two fish populations could be due to sampling bias, differences in water temperature (around to 5 °C to 10 °C lower in Peniche than in the Algarve), anthropogenic factors, and recycled water management systems [57]. Thus, this hypothesis requires further studies to be investigated. Taking this into account, our data regarding the virome associated with liver and gills suggests that the most solitary fishes studied, the gilthead seabream from Peniche, houses the smallest viral diversity. On the other hand, Atlantic horse mackerel from the same region, a densely shoaling fish, held the greatest number of viral families across the studied organs. For this analysis, the skin tissue was not considered since it is an external contact organ more propitious to environmental and handling contaminations. Additionally, as expected, farmed gilthead seabream had the lower viral diversity, indicating more uniform rearing conditions [58]. Therefore, these results support, to some extent, the concept that more frequent intra-host contacts could increase viral diversification and spread. However, a broader comparison of more fish species is essential to understand how population density may influence virus diversity, abundance, and evolution [3].

Although this study focused on only two fish species commonly consumed in Europe, a characterization of their virome in different environments was made for the first time. Since the species analyzed were bought at fisheries, it is possible that we missed viruses with low abundance and gained other viruses (including phages) due to sample processing, manipulation, external contamination, and bacterial growth during fish transportation. Our findings support that fish harbor a very large number of viruses and that viral metagenomics is a useful tool to exhaustively characterize their viral-associated diversity. Indeed, metagenomic analysis has become a precious tool for the characterization of viral composition and diversity in environmental samples. However, the procedures used for sample collection, sequencing preparation and bioinformatic analyses may induce biases in estimating viral diversity [14]. Nevertheless, our knowledge, regarding the diversity of viruses in fish, has increased greatly with the development of metagenomic approaches [14]. Moreover, these approaches would be a valuable asset if used to understand the viral diversity and putative viral pathogens in aquaculture settings, since such information may directly impact on the biosafety of aquaculture systems [59]. Finally, the data presented here show that fish commonly sold in the markets, many of which may be consumed raw or undercooked, harbor a wide range of viral genomes, much of them, unclassified. This fact does not mean that the identified viruses are infectious or pathogenic, nor does it indicate their degree of virulence. However, we cannot exclude that at least some of these viruses can pose a potential risk to human health, being some of them human viral pathogens of interest regarding food safety.

Supplementary Materials: The following are available online at http://www.mdpi.com/2304-8158/9/11/1634/s1, Figure S1: Maximum Likelihood phylogenetics trees of each fish pathogenic viral family identified in this study, Figure S2: Maximum Likelihood phylogenetics trees of each human pathogenic viral family identified in this study, Table S1: Accession numbers of sequences used for phylogenetic analysis, including the sequences obtained in this study (*) and those downloaded from GenBank, Table S2: Accession numbers of sequences used for phylogenetic analysis (family analysis), including the sequences obtained in this study (*) and those downloaded from GenBank.

Author Contributions: Conceptualization, A.F.-S., R.P., M.T.B.C., and M.N.; methodology, A.F.S., R.P., S.M.-P., S.B.-M., and M.N.; formal analysis, A.F.-S. and M.N.; investigation, A.F.-S. and M.N.; validation, A.F.-S., S.M.-P., S.B.-M., and M.N.; writing—original draft preparation, A.F.-S.; writing—review and editing, A.F.-S., R.P., S.M.-P., S.B.-M., M.T.B.C., M.N.; supervision, M.T.B.C. and M.N.; project administration, M.T.B.C. and M.N.; funding acquisition, M.T.B.C. All authors have read and agreed to the published version of the manuscript.

Funding: Funding from INTERFACE Program, through the Innovation, Technology and Circular Economy Fund (FITEC) (CIT/2018/15), and MultiBiorefinery project (POCI-01–0145-FEDER-016403) is gratefully acknowledged.

Acknowledgments: The authors would like to acknowledge the iNOVA4Health—UID/Multi/04462/2013 program, and Andreia Silva to FCT for the doctoral fellowship (PD/BD/128363/2017).

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

References

1. FAO. The State of World Fisheries and Aquaculture 2020. *State World Fish. Aquac.* **2020**, 60–61. [CrossRef]
2. Kibenge, F.S.B. Emerging viruses in aquaculture. *Curr. Opin. Virol.* **2019**, *34*, 97–103. [CrossRef]
3. Geoghegan, J.L.; Di Giallonardo, F.; Cousins, K.; Shi, M.; Williamson, J.E.; Holmes, E.C. Hidden diversity and evolution of viruses in market fish. *Virus Evol.* **2018**, *4*, 1–11. [CrossRef]
4. Chow, C.-E.T.; Suttle, C.A. Biogeography of Viruses in the Sea. *Annu. Rev. Virol.* **2015**, *2*, 41–66. [CrossRef]
5. Suttle, C.A. Marine viruses—Major players in the global ecosystem. *Nat. Rev. Microbiol.* **2007**, *5*, 801–812. [CrossRef]
6. Weitz, J.S.; Wilhelm, S.W. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol. Rep.* **2012**, *4*, 2–9. [CrossRef]
7. Shi, M.; Lin, X.-D.; Chen, X.; Tian, J.-H.; Chen, L.-J.; Li, K.; Wang, W.; Eden, J.-S.; Shen, J.-J.; Liu, L.; et al. The evolutionary history of vertebrate RNA viruses. *Nature* **2018**, *556*, 197–202. [CrossRef]
8. Dill, J.A.; Camus, A.C.; Leary, J.H.; Di Giallonardo, F.; Holmes, E.C.; Ng, T.F.F. Distinct Viral Lineages from Fish and Amphibians Reveal the Complex Evolutionary History of Hepadnaviruses. *J. Virol.* **2016**, *90*, 7920–7933. [CrossRef]
9. National Outbreak Reporting System (NORS). Available online: https://wwwn.cdc.gov/norsdashboard/ (accessed on 20 September 2020).
10. Nkili-Meyong, A.A.; Bigarré, L.; Labouba, I.; Vallaeys, T.; Avarre, J.-C.; Berthet, N. Contribution of Next-Generation Sequencing to Aquatic and Fish Virology. *Intervirology* **2017**, *59*, 285–300. [CrossRef]
11. Geoghegan, J.L.; Pirotta, V.; Harvey, E.; Smith, A.; Buchmann, J.P.; Ostrowski, M.; Eden, J.S.; Harcourt, R.; Holmes, E.C. Virological sampling of inaccessible wildlife with drones. *Viruses* **2018**, *10*, 300. [CrossRef] [PubMed]
12. Shi, M.; Zhang, Y.Z.; Holmes, E.C. Meta-transcriptomics and the evolutionary biology of RNA viruses. *Virus Res.* **2018**, *243*, 83–90. [CrossRef]
13. Crane, M.; Hyatt, A. Viruses of fish: An overview of significant pathogens. *Viruses* **2011**, *3*, 2025–2046. [CrossRef] [PubMed]
14. Hwang, J.; Park, S.Y.; Park, M.; Lee, S.; Lee, T.K. Seasonal dynamics and metagenomic characterization of marine viruses in Goseong Bay, Korea. *PLoS ONE* **2017**, *12*, e0169841. [CrossRef] [PubMed]
15. Kurath, G.; Winton, J. Complex dynamics at the interface between wild and domestic viruses of finfish. *Curr. Opin. Virol.* **2011**, *1*, 73–80. [CrossRef]
16. República Portuguesa. Available online: https://www.portugal.gov.pt/pt/gc21/governo/programa/programa-nacional-para-a-coesao-territorial-/ficheiros-coesao-territorial/programa-nacional-para-a-coesao-territorial-o-interior-em-numeros-territorio-pdf.aspx (accessed on 23 September 2020).
17. Deutscher Wetterdienst (DWD), Sea Surface Temperature. Available online: https://www.dwd.de/EN/ourservices/rcccm/int/rcccm_int_sst.html (accessed on 23 September 2020).
18. Fernandez-Cassi, X.; Rusiñol, M.; Martínez-Puchol, S. Viral Concentration and Amplification from Human Serum Samples Prior to Application of Next-Generation Sequencing Analysis. In *The Human Virome: Methods and Protocols*; Moya, A., Pérez Brocal, V., Eds.; Springer: New York, NY, USA, 2018; pp. 173–188. ISBN 978-1-4939-8682-8.
19. Wang, D.; Urisman, A.; Liu, Y.T.; Springer, M.; Ksiazek, T.G.; Erdman, D.D.; Mardis, E.R.; Hickenbotham, M.; Magrini, V.; Eldred, J.; et al. Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol.* **2003**, *1*, e2. [CrossRef]

20. Briese, T.; Kapoor, A.; Mishra, N.; Jain, K.; Kumar, A.; Jabado, O.J.; Ian Lipkina, W. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio* **2015**, *6*, e01491-15. [CrossRef]
21. Vilsker, M.; Moosa, Y.; Nooij, S.; Fonseca, V.; Ghysens, Y.; Dumon, K.; Pauwels, R.; Alcantara, L.C.; Vanden Eynden, E.; Vandamme, A.M.; et al. Genome Detective: An automated system for virus identification from high-throughput sequencing data. *Bioinformatics* **2019**, *35*, 871–873. [CrossRef]
22. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
23. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P.A. MetaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **2017**, *27*, 824–834. [CrossRef] [PubMed]
24. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
25. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [CrossRef]
26. Trifinopoulos, J.; Nguyen, L.-T.; von Haeseler, A.; Quang Minh, B. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235. [CrossRef]
27. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]
28. Benson, D.A.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2009**, *37*, 26–31. [CrossRef]
29. Sayers, E.W.; Barrett, T.; Benson, D.A.; Bryant, S.H.; Canese, K.; Chetvernin, V.; Church, D.M.; Dicuccio, M.; Edgar, R.; Federhen, S.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2009**, *37*, 5–15. [CrossRef] [PubMed]
30. Uhr, J.W.; Weissman, G. Intracellular distribution and degradation of bacteriophage in mammalian tissues. *J. Immunol.* **1965**, *94*, 544–550.
31. Trigo, G.; Martins, T.G.; Fraga, A.G.; Longatto-Filho, A.; Castro, A.G.; Azeredo, J.; Pedrosa, J. Phage Therapy Is Effective against Infection by *Mycobacterium ulcerans* in a Murine Footpad Model. *PLoS Negl. Trop. Dis.* **2013**, *7*, e2183. [CrossRef] [PubMed]
32. Yin, Y.; Fischer, D. Identification and investigation of ORFans in the viral world. *BMC Genom.* **2008**, *9*, 24. [CrossRef] [PubMed]
33. Chu, D.K.W.; Leung, C.Y.H.; Perera, H.K.K.; Ng, E.M.; Gilbert, M.; Joyner, P.H.; Grioni, A.; Ades, G.; Guan, Y.; Peiris, J.S.M.; et al. A Novel Group of Avian Astroviruses in Wild Aquatic Birds. *J. Virol.* **2012**, *86*, 13772–13778. [CrossRef]
34. Rivera, R.; Nollens, H.H.; Venn-Watson, S.; Gulland, F.M.D.; Wellehan, J.F.X. Characterization of phylogenetically diverse astroviruses of marine mammals. *J. Gen. Virol.* **2010**, *91*, 166–173. [CrossRef]
35. Sahul Hameed, A.S.; Ninawe, A.S.; Nakai, T.; Chi, S.C.; Johnson, K.L. ICTV virus taxonomy profile: Nodaviridae. *J. Gen. Virol.* **2019**, *100*, 3–4. [CrossRef]
36. Toffan, A.; Pascoli, F.; Pretto, T.; Panzarin, V.; Abbadi, M.; Buratin, A.; Quartesan, R.; Gijon, D.; Padros, F. Viral nervous necrosis in gilthead sea bream (*Sparus aurata*) caused by reassortant betanodavirus RGNNV/SJNNV: An emerging threat for Mediterranean aquaculture. *Sci. Rep.* **2017**, *7*, 46755. [CrossRef]
37. Berzak, R.; Scheinin, A.; Davidovich, N.; Regev, Y.; Diga, R.; Tchernov, D.; Morick, D. Prevalence of nervous necrosis virus (NNV) and *Streptococcus* species in wild marine fish and crustaceans from the Levantine Basin, Mediterranean Sea. *Dis. Aquat. Organ.* **2019**, *133*, 7–17. [CrossRef]
38. Giacopello, C.; Foti, M.; Bottari, T.; Fisichella, V.; Barbera, G. Detection of viral encephalopathy and retinopathy virus (VERV) in wild marine fish species of the South Tyrrhenian Sea (Central Mediterranean). *J. Fish Dis.* **2013**, *36*, 819–821. [CrossRef]
39. Luna, G.M.; Corinaldesi, C.; Dell'Anno, A.; Pusceddu, A.; Danovaro, R. Impact of aquaculture on benthic virus-prokaryote interactions in the Mediterranean Sea. *Water Res.* **2013**, *47*, 1156–1168. [CrossRef] [PubMed]
40. Ho, K.L.; Gabrielsen, M.; Beh, P.L.; Kueh, C.L.; Thong, Q.X.; Streetley, J.; Tan, W.S.; Bhella, D. Structure of the *Macrobrachium rosenbergii* nodavirus: A new genus within the Nodaviridae? *PLoS Biol.* **2018**, *16*, e3000038. [CrossRef] [PubMed]
41. Isshiki, T.; Nagano, T.; Suzuki, S. Infectivity of aquabirnavirus strains to various marine fish species. *Dis. Aquat. Organ.* **2001**, *46*, 109–114. [CrossRef]

42. Hirayama, T.; Nagano, I.; Shinmoto, H.; Yagyu, K.I.; Oshima, S.I. Isolation and characterization of virulent yellowtail ascites virus. *Microbiol. Immunol.* **2007**, *51*, 397–406. [CrossRef] [PubMed]
43. Manríquez, R.A.; Vera, T.; Villalba, M.V.; Mancilla, A.; Vakharia, V.N.; Yañez, A.J.; Cárcamo, J.G. Molecular characterization of infectious pancreatic necrosis virus strains isolated from the three types of salmonids farmed in Chile. *Virol. J.* **2017**, *14*, 1–16. [CrossRef]
44. Mutoloki, S.; Evensen, Ø. Sequence similarities of the capsid gene of Chilean and European isolates of infectious pancreatic necrosis virus point towards a common origin. *J. Gen. Virol.* **2011**, *92*, 1721–1726. [CrossRef]
45. Lun, J.H.; Hewitt, J.; Yan, G.J.H.; Tuipulotu, D.E.; Rawlinson, W.D.; White, P.A. Recombinant GII.P16/GII.4 sydney 2012 was the dominant norovirus identified in Australia and New Zealand in 2017. *Viruses* **2018**, *10*, 548. [CrossRef]
46. Ruis, C.; Roy, S.; Brown, J.R.; Allen, D.J.; Goldstein, R.A.; Breuer, J. The emerging GII.P16-GII.4 Sydney 2012 norovirus lineage is circulating worldwide, arose by late-2014 and contains polymerase changes that may increase virus transmission. *PLoS ONE* **2017**, *12*, e0179572. [CrossRef]
47. Ramesh, A.; Nakielny, S.; Hsu, J.; Kyohere, M.; Byaruhanga, O.; de Bourcy, C.; Egger, R.; Dimitrov, B.; Juan, Y.-F.; Sheu, J.; et al. Etiology of fever in Ugandan children: Identification of microbial pathogens using metagenomic next-generation sequencing and IDseq, a platform for unbiased metagenomic analysis. *bioRxiv* **2018**, 385005. [CrossRef]
48. McIntyre, C.L.; Knowles, N.J.; Simmonds, P. Proposals for the classification of human rhinovirus species A, B and C into genotypically assigned types. *J. Gen. Virol.* **2013**, *94*, 1791–1806. [CrossRef]
49. Abeles, S.R.; Pride, D.T. Molecular bases and role of viruses in the human microbiome. *J. Mol. Biol.* **2014**, *426*, 3892–3906. [CrossRef]
50. Łusiak-Szelachowska, M.; Weber-Dąbrowska, B.; Jończyk-Matysiak, E.; Wojciechowska, R.; Górski, A. Bacteriophages in the gastrointestinal tract and their implications. *Gut Pathog.* **2017**, *9*, 1–5. [CrossRef]
51. Dąbrowska, K. Phage therapy: What factors shape phage pharmacokinetics and bioavailability? Systematic and critical review. *Med. Res. Rev.* **2019**, *39*, 2000–2025. [CrossRef]
52. Geier, M.R.; Trigg, M.E.; Merril, C.R. Fate of bacteriophage lambda in Non-immune germ-free mice. *Nature* **1973**, *246*, 221–223. [CrossRef] [PubMed]
53. Ezenwa, V.O.; Price, S.A.; Altizer, S.; Vitone, N.D.; Cook, K.C. Host traits and parasite species richness in even and odd-toed hoofed mammals, Artiodactyla and Perissodactyla. *Oikos* **2006**, *115*, 526–536. [CrossRef]
54. Lindenfors, P.; Nunn, C.L.; Jones, K.E.; Cunningham, A.A.; Sechrest, W.; Gittleman, J.L. Parasite species richness in carnivores: Effects of host body mass, latitude, geographical range and population density. *Glob. Ecol. Biogeogr.* **2007**, *16*, 496–509. [CrossRef]
55. Webber, Q.M.R.; Fletcher, Q.E.; Willis, C.K.R. Viral Richness is Positively Related to Group Size, but Not Mating System, in Bats. *EcoHealth* **2017**, *14*, 652–661. [CrossRef]
56. May, R.M.; Anderson, R.M. Population biology of infectious diseases: Part II. *Nature* **1979**, *280*, 455–461. [CrossRef]
57. Munang'andu, H.M. Environmental viral metagenomics analyses in aquaculture: Applications in epidemiology and disease control. *Front. Microbiol.* **2016**, *7*, 1986. [CrossRef]
58. Uren Webster, T.M.; Consuegra, S.; Hitchings, M.; Garcia de Leaniz, C. Interpopulation Variation in the Atlantic Salmon Microbiome Reflects Environmental and Genetic Diversity. *Appl. Environ. Microbiol.* **2018**, *84*, e00691-18. [CrossRef]
59. Alavandi, S.V.; Poornima, M. Viral metagenomics: A tool for virus discovery and diversity in aquaculture. *Indian J. Virol.* **2012**, *23*, 88–98. [CrossRef]

DISCUSSIÓ

Seqüenciació massiva per a estudis d'epidemiologia basats en aigües residuals.

L'epidemiologia basada en aigües residuals és una eina eficient per al monitoreig dels virus, entre d'altres, que circulen en la població. Les eines de seqüenciació massiva han experimentat avenços en els darrers anys, permetent estudiar el viroma circulant en la població i el possible seguiment de brots en un únic assaig.

El primer capítol de la tesi inclou un article on s'avalua l'aplicació de tres tècniques de seqüenciació massiva diferents per a l'estudi del viroma i de la diversitat de patògens presents a l'aigua residual. La principal limitació en els estudis de viroma en aquest tipus de matrius és la baixa proporció de seqüències víriques en comparació amb el nombre total de seqüències amplificades quan s'utilitzen encebadors aleatoris, un fet crític quan l'objectiu d'estudi són virus d'interès humà. Els membres més abundants del viroma són els bacteriòfags i els virus de plantes (Aw et al., 2014; Cantalupo et al., 2011; Fernández-Cassi et al., 2018), motiu pel qual l'aplicació d'eines de seqüenciació massiva d'amplicons i d'enriquiment de dianes poden ser de gran utilitat per a l'estudi detallat de famílies específiques i la seva diversitat en una mostra.

Per avaluar les eines més apropiades per analitzar el viroma en aigües residuals, primerament es va analitzar per seqüenciació massiva per metagenòmica i per enriquiment de dianes una mostra composta d'aigua residual d'abril de 2016 d'una EDAR de la ciutat de Barcelona, que tracta l'aigua residual produïda per 1,5 milions d'habitants. La utilització de l'enriquiment va resultar en un augment de 0,72 a 6,07 milions de seqüències víriques. El kit d'enriquiment aplicat per a la seqüenciació estava dirigit a famílies de virus que infecten vertebrats, pel que la detecció de seqüències obtingudes d'aquestes famílies es va veure augmentada quasi un 80% en comparació a la seqüenciació no dirigida. Aquesta tècnica va resultar ser més sensible per a la detecció de patògens vírics d'interès

humà, permetent la identificació de membres de les famílies *Caliciviridae* i *Astroviridae*, també detectades per metagenòmica no dirigida, a més de seqüències de *Adenoviridae*, *Hepeviridae*, *Papillomaviridae*, *Reoviridae*, *Polyomaviridae* i *Picornaviridae*. A part d'una millora substancial en la detecció d'espècies de virus d'aquestes famílies, l'enriquiment de dianes va permetre una cobertura major dels genomes seqüenciats gràcies al elevat nombre i diversitat de seqüències obtingudes amb aquesta metodologia, arribant a recuperar més del 90% de la seqüència del genoma complet de diferents virus com AiV, NoV, JCPyV o membres del gènere *Mamastrovirus*.

A la vegada, es va avaluar la utilització de la seqüenciació massiva d'amplicons per a l'estudi de la diversitat de tres grups vírics (adenovirus, papil·lomavirus i enterovirus) i la seva presència en el transcurs d'un any, mostrejant mensualment una EDAR del nord del Barcelonès (115,000 habitants). Es van poder seqüenciar més de 10 serotips diferents d'adenovirus dels grups F i A, relacionats amb gastroenteritis i altres malalties, amb HAdV 41 i HAdV 31 com a membres més abundants, tal i com ha estat descrit prèviament (Ayed i Sabbahi, 2017). Es van detectar en totes les estacions analitzades, el que concorda amb el fet de que els adenovirus són excretats de manera persistent, també per persones sanes, i del seu ús com a indicador de contaminació fecal humana (Bofill-Mas et al., 2006; Pina et al., 1998). L'estudi dels virus del papil·loma humà va constatar l'elevada diversitat de *Betapapillomavirus* en l'aigua residual, en concordància amb l'estudi realitzat a Itàlia per La Rosa i col·laboradors (La Rosa et al., 2013). Respecte als membres del gènere *Alphapapillomavirus*, HPV 6 i HPV 66 van ser detectats en major proporció a l'hivern i primavera. Encara que la presència d'aquest gènere ha estat descrita prèviament en aigua residual (Iaconelli et al., 2015), el tipus HPV 66, relacionat amb càncer vulvar, no ha estat descrit amb anterioritat a l'ambient, encara que recentment s'ha reportat la seva

elevada incidència lesions de cèrvix o en el recte (Juárez-González et al., 2020; Smelov et al., 2018).

L'estudi de la diversitat del gènere *Enterovirus* va mostrar un elevat nombre de genotips amb distribució diferent en funció de l'estació de l'any analitzada. En la primavera de l'any 2016 es va d'escriure a Catalunya l'aparició d'un brot de romboencefalitis en nens amb l'enterovirus A71 C1 com a agent causal (Casas-Alba et al., 2017). La seqüenciació massiva d'amplicons va permetre la detecció de seqüències d'aquest genotip en les mostres d'abril 2016 analitzades, les quals van ser comparades amb les obtingudes d'aïllats clínics del brot, trobant una elevada proximitat filogenètica entre elles. Amb la seqüenciació amb enriquiment de dianes també es van obtenir seqüències d'aquest genotip, però en analitzar el contig no va ser possible realitzar el subtipatge.

Els resultats d'aquest estudi demostren la utilitat de les diverses aplicacions de seqüenciació massiva en l'epidemiologia d'aigües residuals. Mentre que la seqüenciació per metagenòmica permet estudiar el viroma complet en una mostra determinada, l'enriquiment de dianes facilita la detecció i caracterització dels potencials patògens amb una major sensibilitat i en un únic estudi, el que fa que sigui una potencial eina per la vigilància ambiental en termes de salut pública. Finalment, la seqüenciació massiva d'amplicons ha permès l'anàlisi de grups vírics específics, permetent el caracterització genotípica de tots els tipus presents i la seva distribució estacional.

Arrel de la identificació del nou SARS-CoV-2 i de la declaració de la OMS de la pandèmia de COVID-19 el passat 11 de març del 2011, es van aplicar les tècniques descrites anteriorment per tal de caracteritzar el viroma i monitoritzar la presència del SARS-CoV-2 i altres coronavirus en aigües residuals. En el context d'aquesta pandèmia, s'ha demostrat la correlació entre la concentració del virus en aigües residuals i la prevalença de la malaltia, fent encara més evident la utilització de la epidemiologia basada

en aigües residuals com a eina d'estudi de patògens rellevants a nivell sanitari (Medema et al., 2020).

Es va estudiar una mostra d'aigua residual composta, proporcional al cabal i durant 24h, del mes de març de 2020, pic de casos de la primera onada de la pandèmia a Catalunya, i una altra mostra del mes de juliol, prèvia a la segona onada, que presentaven concentracions diferents de SARS-CoV-2 (10^6 i 10^4 còpies genòmiques per litre d'aigua residual, respectivament) però concentracions similars de l'indicador de contaminació fecal JCPyV. La caracterització del viroma es va realitzar amb i sense enriquiment de dianes, trobant en ambdós mesos analitzats una composició similar del viroma. Seqüències de virus de la família *Coronaviridae* només es van detectar en la mostra de març i mitjançant l'enriquiment, amb la obtenció de 8 contigs de SARS-CoV-2 que comprenen aproximadament el 7% del genoma complet del virus i un contig del coronavirus humà OC43, relacionat amb el refredat comú estacional (Van der Hoek, 2015). Altres coronavirus no humans van poder ser detectats amb la seqüenciació amb enriquiment de dianes de virus que infecten vertebrats, trobant coronavirus felins, canins i de rata, encara que amb un baix nombre de seqüències. L'ús d'aquestes aplicacions dirigides específicament a l'estudi dels coronavirus humans i animals circulants generaria informació valuosa sobre potencials riscs de zoonosi i fenòmens d'immunitat creuada entre soques (Ma et al., 2020).

Les seqüències de SARS-CoV-2 recuperades de la mostra d'aigua residual de març van ser comparades amb seqüències clíniques provinents de l'Hospital Vall d'Hebron, obtingudes en el mateix mes i amb la mateixa metodologia de seqüenciació massiva amb enriquiment de dianes. El 90% de cobertura de genoma aconseguit en les mostres clíniques va permetre estudiar que existien 3 divergències nucleotídiques respecte a les seqüències obtingudes de l'aigua residual. A la vegada aquestes últimes es van comparar amb un total de 1200 seqüències d'aïllats clínics del virus

del repositori de dades GISAID (Elbe i Buckland-Merrett, 2017). Els canvis nucleotídics detectats no es van trobar en cap de les seqüències comparades ni estan recollits en estudis preliminars existents de variants del virus trobades en aigües residuals (Izquierdo-Lara et al., 2020). Es així que, per que aquests canvis es puguin considerar polimorfismes, seria necessari disposar de més seqüències de la regió divergent per a constatar que estan presents a l'ambient i la seva presència no és fruit d'errors generats en la preparació de la mostra o la seqüenciació. Aquests resultats tornen demostrar que per a realitzar aquests estudis de tipatge i anàlisi de variants en mostres ambientals és més adequada la utilització de la seqüenciació massiva d'amplicons, els quals han estat utilitzat recentment i de manera exitosa per a la caracterització del SARS-CoV-2 en mostres clíniques (Nasir et al., 2020; Xiao et al., 2020).

La seqüenciació amb enriquiment de dianes en canvi ha permès també la detecció d'altres coronavirus humans i animals, però la seva concentració ha demostrat ser determinant per la recuperació eficient de genomes (Carbo et al., 2020; Klempt et al., 2020). Existeix un ampli espai de millora en aquestes tecnologies i, arrel de l'aparició del SARS-CoV-2, el seu ús s'ha vist potenciat i millorat com per exemple en els panells dirigits a l'amplificació massiva de virus respiratoris d'Illumina (Illumina Inc., 2020) o Twist Biosciences (Twist Bioscience, 2020). La seqüenciació amb enriquiment de dianes també ha permès recentment el descobriment de nous coronavirus en ratpenats asiàtics (Li et al., 2020; Lim et al., 2019), pel que seria interessant l'elaboració de panells optimitzats per a detectar virus desconeguts però dirigits a virus que pertanyen a famílies d'interès en les mostres a analitzar.

Identificació de potencials agents etiològics causants de gastroenteritis mitjançant seqüenciació massiva.

La gastroenteritis és una important malaltia arreu del món, afectant amb més severitat a nens i gent gran i causant un elevat nombre de morts en

països de mitja i baixa renta i un impacte econòmic important també en països amb bones condicions sanitàries. Dins dels agents etiològics causants de la malaltia, els virus són responsables de més del 70% del casos en nens, amb els rotavirus, norovirus i adenovirus com a patògens principals (Oppong et al., 2020). Encara així, en molts casos no és possible identificar l'agent causal, pel que s'ha postulat que l'estudi mitjançant seqüenciació massiva de la presència de virus nous, o no tradicionalment relacionats amb la malaltia, pot jugar un paper important per al seu diagnòstic clínic (Fischer et al., 2019).

Amb l'objectiu d'identificar patògens vírics en mostres clíniques, es va plantejar l'aplicació de la seqüenciació massiva per a l'estudi dels virus presents en mostres negatives pels patògens tradicionalment associats a gastroenteritis. Així, es van seqüenciar pools de mostres de femtes de nens i adults amb gastroenteritis aguda per metagenòmica i per seqüenciació massiva amb enriquiment de dianes. Les famílies identificades conegudes com a potencials causants de gastroenteritis van ser *Caliciviridae*, *Astroviridae*, *Adenoviridae* i *Reoviridae*.

La detecció de seqüències de genotips de NoV GII en diversos pools es podria explicar per possibles canvis en les regions d'anellament dels primers utilitzats pel diagnòstic o per estar presents en baixes concentracions, el que faria que escapessin a la detecció clínica. Dins del mateix gènere *Norovirus,* es van detectar un elevat nombre de seqüències corresponents al genogrups IV en dos pools, resultat que va ser comprovat mitjançant qPCR. La presència d'aquest genogrups va ser descrita a principis de l'any 2000 en casos esporàdics de gastroenteritis (Vinjé i Koopmans, 2000), i des de llavors la seva presència ha estat descrita en diversos estudis de metagenòmica en casos clínics i en monitoreig ambiental (Ao et al., 2014; Fioretti et al., 2018; La Rosa et al., 2008), encara que molts dels kits de detecció a nivell clínic no inclouen la seva detecció. Un cas similar succeeix amb els membres del gènere *Sapovirus*,

els quals es van trobar en gairebé tots els pools analitzats. Els SaV són una de les causes principals de gastroenteritis en nens, amb un impacte major en països de baixa renta però, degut a la seva emergència en els darrers anys, la seva detecció no es realitza encara de manera rutinària en molts dels centres de diagnòstic del país (Varela et al., 2019).

Pel que fa a l'elevat nombre de seqüències obtingudes d'AstV humans, es va poder recuperar, quasi al complet, el genoma de 3 genotips diferents presents en tots els pools analitzats. Els astrovirus estan relacionats amb brots de gastroenteritis aguda en adults i nens (Jeong et al., 2012) i la seva presència va ser avaluada mitjançant un test inmunocromatogràfic en les mostres que comprenen l'estudi, suggerint així la necessitat de millora de la sensibilitat dels kits de detecció utilitzats.

A part dels patògens mencionats anteriorment, també es van obtenir una gran quantitat de seqüències d'enterovirus i altres membres de la família *Picornaviridae*, com els PeV i els cardiovirus, dels que es coneix que poden produir una gran varietat de quadres clínics (Blinkova et al., 2009; De Crom et al., 2016) i alhora s'excreten activament per individus sans, pel que la seva relació amb el desenvolupament de gastroenteritis aguda no es pot afirmar amb les dades clíniques disponibles. També es van trobar seqüències de virus de les famílies *Parvoviridae*, *Anelloviridae* i *Circoviridae*, de les que es desconeix la seva implicació directa o indirecta en el desenvolupament de la malaltia (Reshetnyak et al., 2020; Tatte i Gopalkrishna, 2019). L'aplicació de la metagenòmica va poder permetre detectar membres de totes aquestes famílies en les femtes dels individus analitzats, encara que es desconeix si pot la seva presència estar deguda per una infecció subclínica al moment de la recollida de mostra i/o per coinfeccions amb altres patògens.

Finalment, l'aplicació de la seqüenciació massiva amb enriquiment de dianes en un dels pools analitzats va resultar en la obtenció d'un nombre major de seqüències víriques. Això es va traduir en la detecció de més

reads de membres de les famílies *Caliciviridae*, *Astroviridae* i *Adenoviridae* encara que al analitzar-los com a contigs no es van trobar diferències significatives amb la seqüenciació sense enriquiment. Aquest fet es podria explicar perquè l'enriquiment de dianes probablement generaria moltes seqüències solapades dels virus detectats, no aconseguint reconstruir en aquest cas una major part del genoma del virus. Pel que fa a la família *Reoviridae*, l'aplicació d'aquesta tècnica de seqüenciació massiva sí va permetre detectar 4 contigs de rotavirus en aquest pool que no havien estat detectats d'altra manera.

Els resultats obtinguts, a més de recolzar la incorporació i millora de tests de detecció de NoV GIV, SaV i AstV en el diagnòstic rutinari de gastroenteritis, posen en valor l'aplicació de tècniques de seqüenciació massiva en el context del diagnòstic clínic. Encara que són eines que permeten un anàlisi global dels patògens presents en una mostra en un sol assaig, la seqüenciació massiva per a l'estudi de virus presenta un gran potencial en el camp dels assajos automatitzats i amb plataformes avançades de seqüenciació per al seu ús en diagnòstic rutinari, com s'ha fet, per exemple, en el diagnòstic de malalties hereditàries, prenatals o càncer (Cabanillas et al., 2017; Chen et al., 2016; Muzzey, 2018).

Estudi del viroma d'espècies de peixos atlàntics destinats al consum humà.

La seqüenciació amb enriquiment de dianes també permet estudiar el viroma d'altres organismes i en concret d'animals destinats al consum humà. En aquest sentit, s'ha aplicat aquesta metodologia per a l'estudi del viroma i dels patògens presents en dues espècies de peixos d'una important zona de pesca i criança europea, la costa Atlàntica portuguesa. Aquest estudi es va realitzar en col·laboració amb un grup de recerca del *Instituto de Biologia Experimental e Tecnológica* de la ciutat de Lisboa.

La producció de peix és una important font d'aliments per al consum humà i d'ingressos per a milions de persones a tot el món. Tot i que la demanda de peix es troba en augment, les malalties víriques són un factor important per restringir el comerç de peix i la producció d'aqüicultura, afectant el desenvolupament de la indústria pesquera amb importants pèrdues econòmiques (Geoghegan et al., 2018). El coneixement actual en la virologia de peixos es centra en aquelles famílies estan relacionades amb el desenvolupament de malalties, i conseqüent pèrdua econòmica, d'espècies rellevants, pel que una gran proporció de la diversitat viral d'aquestes i el seu possible impacte en salut animal i humana es desconeix (Crane i Hyatt, 2011). És en aquest darrer punt que l'aplicació de la seqüenciació massiva pot generar informació rellevant per a la salut animal i a la vegada en termes de seguretat alimentària.

En aquest estudi es va analitzar, mitjançant seqüenciació massiva amb enriquiment de dianes de virus vertebrats, la diversitat vírica de diversos òrgans de dues espècies de peixos de vida lliure, sorell (*Trachurus trachurus*) i orada (*Sparus aurata*), i d'aqüicultura. Les espècies de virus trobades més abundantment van ser les corresponents a bacteriòfags, el que concorda amb la seva ubiqüitat i alta presencia també en ecosistemes marins (Chow & Suttle, 2015). Les famílies detectades que poden estar relacionades amb infeccions en vertebrats van ser *Astroviridae*, *Nodaviridae*, *Hepadnaviridae*, *Birnaviridae*, *Caliciviridae* i *Picornaviridae*. Al voltant del 13% de les seqüències víriques obtingudes no es van poder classificar en cap família coneguda.

Respecte a les seqüències víriques trobades de famílies que tenen un impacte en la salut dels propis peixos, trobem virus de la necrosi nerviosa (VNNV) un nodavirus que està relacionat amb necrosis neural, encefalopaties i rentinopaties i que s'associa a una alta mortalitat convertint-se en un problema important en aqüicultura (Hameed et al., 2019). També es van detectar seqüències del virus de la necrosi hepàtica

(IPNV), comunament associat a lesions necròtiques en espècies de salmons i grans pèrdues econòmiques, que poden estar transmesos entre els ous dels peixos en el moment del seu creixement en granges marines (Mutoloki i Evensen, 2011).

En quant a l'aplicació del mètode de seqüenciació per a la detecció de les famílies que tenen poden tenir un impacte en la salut dels consumidors, la seva utilitat queda demostrada amb la detecció de seqüències de NoV GII.4 en mostres de pell d'orada. Encara que la presència de genomes en aquest teixit no ens dona informació sobra la seva infectivitat, sí que dona informació sobre una possible contaminació d'aigua en origen o un mala manipulació del producte. Si aquesta contaminació arriba a la carn del peix, pot suposar un potencial risc per a la salut si és ingerida crua o poc cuinada.

Els resultats obtinguts en aquest darrer capítol reforcen l'ús de la seqüenciació massiva per la caracterització de la elevada diversitat vírica en peixos i el seu impacte, juntament amb la detecció de patògens, en l'estudi de la bioseguretat dels sistemes d'aqüicultura. A la vegada l'aplicació d'aquestes tècniques permet detectar en un únic assaig independent de seqüència, com s'ha comentat amb anterioritat, la potencial presència de patògens que poden ser de risc per la salut dels consumidors. Aquestes eines són doncs potencialment vàlides per a l'estudi de virus contaminants d'aigua i aliments amb repercussió en salut pública.

Així doncs, la utilització de diferents tècniques de seqüenciació massiva en aigua residual, mostres clíniques i animals ha permès la realització d'estudis exhaustius sobre la circulació i diversitat de virus en la població, analitzant la seva potencial relació amb malalties en humans i el seu ús en aplicacions com l'estudi del viroma d'animals d'interès alimentari, posant èmfasi en la necessitat de tractar tota la informació obtinguda mitjançant aquestes metodologies des d'un context *One-Health*.

CONCLUSIONS

- L'optimització i aplicació de diferents tècniques de seqüenciació massiva en aigües residuals, mostres clíniques i animals ha permès realitzar estudis exhaustius sobre la circulació i diversitat de virus en la població, la seva potencial relació amb malalties en humans i el seu ús en aplicacions com l'estudi del viroma d'animals d'interès alimentari.

- La seqüenciació massiva per metagenòmica ha permès descriure el viroma humà d'aigües residuals, incloent virus que infecten vertebrats, invertebrats, bacteris, plantes i fongs.

- La seqüenciació massiva amb enriquiment de dianes ha resultat ser l'eina més útil per a l'estudi d'un grup de virus d'interès, com els virus que infecten vertebrats, proporcionant un elevat nombre de seqüències d'aquestes famílies, amb elevada cobertura genòmica, que són d'interès en estudis d'epidemiologia basada en l'estudi d'aigües residuals.

- La seqüenciació massiva d'amplicons ha facilitat la caracterització genètica de grups de virus concrets amb una elevada sensibilitat, podent realitzar estudis de subtipatge de patògens específics associats a brots que circulen en la població, com ha estat el cas de l'enterovirus A71 C1 causant del brot d'encefalitis català de l'any 2016.

- En el context de la pandèmia de COVID-19, l'aplicació de l'enriquiment de dianes en la seqüenciació ha permès la recuperació quasi total del genoma del SARS-CoV-2 en mostres clíniques i de seqüències parcials del virus aïllades de l'aigua residual.

- El viroma de l'aigua residual estudiat durant el primer pic de la pandèmia ha evidenciat la co-circulació d'altres coronavirus que infecten humans, animals domèstics i rosegadors en àrees urbanes, demostrant la importància d'aplicar estudis de caracterització de virus circulants potencialment zoonòtics.

- Mitjançant seqüenciació massiva, amb i sense enriquiment de dianes, s'han detectat patògens causants de gastroenteritis aguda en pacients de diferents edats en els que no s'havia pogut identificar prèviament l'agent etiològic.

- Els resultats obtinguts en l'anàlisi per metagenòmica de mostres de pacients de gastroenteritis suggereix que el diagnòstic de virus causants de gastroenteritis, a més de rotavirus, adenovirus i norovirus GI i GII, hauria d'incloure la detecció de norovirus GIV, sapovirus i astrovirus.

- L'estudi del viroma de dues espècies de peixos de la costa Atlàntica destinats a consum humà ha mostrat una elevada diversitat de famílies víriques, detectant importants patògens de peixos amb un impacte econòmic directe i amb implicacions en la gestió de la producció d'aquests tipus d'aliments.

REFERÈNCIES

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. https://doi.org/10.1073/pnas.211424698

Anderson, M. E., Nagy-Szakal, D., Jain, K., Patrone, C. C., Frattini, M. G., Lipkin, W. I., & Geskin, L. J. (2018). Highly Sensitive Virome Capture Sequencing Technique VirCapSeq-VERT Identifies Partial Noncoding Sequences but no Active Viral Infection in Cutaneous T-Cell Lymphoma. *Journal of Investigative Dermatology*, *138*(7), 1671–1673. https://doi.org/10.1016/j.jid.2018.01.024

Ao, Y. yun, Yu, J. mei, Li, L. li, Jin, M., & Duan, Z. jun. (2014). Detection of human norovirus GIV.1 in China: A case report. *Journal of Clinical Virology*, *61*(2), 298–301. https://doi.org/10.1016/j.jcv.2014.08.002

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

Ayed, L., & Sabbahi, S. (2017). Part Three. Specific Excreted Pathogens: Environmental and Epidemiology Aspects: Entamoeba Histolytica. *Global Water Pathogens Project*, 36. Retrieved from http://www.waterpathogens.org

Barzon, L., Lavezzo, E., Costanzi, G., Franchin, E., Toppo, S., & Palù, G. (2013). Next-generation sequencing technologies in diagnostic virology. *Journal of Clinical Virology*, *58*(2), 346–350. https://doi.org/10.1016/j.jcv.2013.03.003

Bateman, A. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky1049

Beerenwinkel, N., & Zagordi, O. (2011). Ultra-deep sequencing for the analysis of viral populations. *Current Opinion in Virology*, *1*(5), 413–418. https://doi.org/10.1016/j.coviro.2011.07.008

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., … Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456*(7218), 53–59. https://doi.org/10.1038/nature07517

Besser, J., Carleton, H. A., Gerner-Smidt, P., Lindsey, R. L., & Trees, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical Microbiology and Infection*, *24*(4), 335–341. https://doi.org/10.1016/j.cmi.2017.10.013

Bibby, K., Viau, E., & Peccia, J. (2011). Viral metagenome analysis to guide human pathogen monitoring in environmental samples. *Letters in Applied Microbiology*. https://doi.org/10.1111/j.1472-765X.2011.03014.x

Bibby, Kyle, & Peccia, J. (2013). Identification of viral pathogen diversity in sewage sludge by metagenome analysis. *Environmental Science and Technology*. https://doi.org/10.1021/es305181x

Bisseux, M., Colombet, J., Mirand, A., Roque-Afonso, A. M., Abravanel, F., Izopet, J., … Henquell, C. (2018). Monitoring human enteric viruses in wastewater and relevance to infections encountered in the clinical setting: A one-year experiment in central France, 2014 to 2015. *Eurosurveillance*. https://doi.org/10.2807/1560-7917.ES.2018.23.7.17-00237

Bisseux, M., Didier, D., Audrey, M., Christine, A., Hélène, P.-L., Jean-Luc, B., & Cécile, H. (2020). Monitoring of enterovirus diversity in wastewater by ultra-deep sequencing: An effective complementary tool for clinical enterovirus surveillance. *Water Research*, *169*, 115246. https://doi.org/10.1016/j.watres.2019.115246

Blinkova, O., Rosario, K., Li, L., Kapoor, A., Slikas, B., Bernardin, F., … Delwart, E. (2009). Frequent detection of highly diverse variants of Cardiovirus, Cosavirus, Bocavirus, and Circovirus in sewage samples collected in the United States. *Journal of Clinical Microbiology*. https://doi.org/10.1128/JCM.01062-09

Bodewes, R., Run, P. R. W. A. Van, Schürch, A. C., Koopmans, M. P. G., Osterhaus, A. D. M. E., Baumgärtner, W., … Smits, S. L. (2015). Virus characterization and discovery in formalin-fixed paraffin-embedded tissues. *Journal of Virological Methods*, *214*, 54–59. https://doi.org/10.1016/j.jviromet.2015.02.002

Bofill-Mas, Silvia, Albinana-Gimenez, N., Clemente-Casares, P., Hundesa, A., Rodriguez-Manzano, J., Allard, A., … Girones, R. (2006). Quantification and stability of human adenoviruses and polyomavirus JCPyV in wastewater matrices. *Applied and Environmental Microbiology*, *72*(12), 7894–7896. https://doi.org/10.1128/AEM.00965-06

Bofill-Mas, Sílvia, Pina, S., & Girones, R. (2000). Documenting the epidemiologic patterns of polyomaviruses in human populations by studying their presence in urban sewage. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.66.1.238-245.2000

Borchardt, M. A., Spencer, S. K., Kieke, B. A., Lambertini, E., & Loge, F. J. (2012). Viruses in nondisinfected drinking water from municipal wells and community incidence of acute gastrointestinal illness.

Environmental Health Perspectives. https://doi.org/10.1289/ehp.1104499

Bosch, A., Guix, S., & Pintó, R. M. (2012). Epidemiology of Human Astroviruses. In *Astrovirus Research* (pp. 1–18). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-4735-1_1

Bouwknegt, M., Lodder-Verschoor, F., Van Der Poel, W. H. M., Rutjes, S. A., & Husman, A. M. D. R. (2007). Hepatitis E virus RNA in commercial porcine livers in The Netherlands. *Journal of Food Protection*. https://doi.org/10.4315/0362-028X-70.12.2889

Bozkurt, D., Selimoğlu, M. A., Otlu, B., & Sandıkkaya, A. (2015). Eight different viral agents in childhood acute gastroenteritis. *Turkish Journal of Pediatrics*.

Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O. J., & Ian Lipkina, W. (2015). Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio*, *6*(5), 1–12. https://doi.org/10.1128/mBio.01491-15

Cabanillas, R., Diñeiro, M., Castillo, D., Pruneda, P. C., Penas, C., Cifuentes, G. A., … Cadiñanos, J. (2017). A novel molecular diagnostics platform for somatic and germline precision oncology. *Molecular Genetics and Genomic Medicine*. https://doi.org/10.1002/mgg3.291

Cantalupo, P. G., Calgua, B., Zhao, G., Hundesa, A., Wier, A. D., Katz, J. P., … Pipas, J. M. (2011). Raw sewage harbors diverse viral populations. *MBio*, *2*(5), e00180-11. https://doi.org/10.1128/mBio.00180-11

Carbo, E. C., Sidorov, I. A., Zevenhoven-Dobbe, J. C., Snijder, E. J., Claas, E. C., Laros, J. F. J., … de Vries, J. J. C. (2020). Coronavirus discovery by metagenomic sequencing: a tool for pandemic preparedness. *Journal of Clinical Virology*, *131*(January), 104594. https://doi.org/10.1016/j.jcv.2020.104594

Casas-Alba, D., de Sevilla, M. F., Valero-Rello, A., Fortuny, C., García-García, J. J., Ortez, C., … Launes, C. (2017). Outbreak of brainstem encephalitis associated with enterovirus-A71 in Catalonia, Spain (2016): a clinical observational study in a children's reference centre in Catalonia. *Clinical Microbiology and Infection*. https://doi.org/10.1016/j.cmi.2017.03.016

Chalkias, S., Gorham, J. M., Mazaika, E., Parfenov, M., Dang, X., DePalma, S., … Koralnik, I. J. (2018). ViroFind: A novel target-enrichment deep-sequencing platform reveals a complex JC virus population in the brain of PML patients. *PloS One*, *13*(1), e0186945. https://doi.org/10.1371/journal.pone.0186945

Charre, C., Ginevra, C., Sabatier, M., Regue, H., Destras, G., Brun, S., … Josset, L. (2020). Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. *Virus Evolution*. https://doi.org/10.1093/ve/veaa075

Chen, S., Dong, C., Wang, Q., Zhong, Z., Qi, Y., Ke, X., & Liu, Y. (2016). Targeted Next-Generation Sequencing Successfully Detects Causative Genes in Chinese Patients with Hereditary Hearing Loss. *Genetic Testing and Molecular Biomarkers*. https://doi.org/10.1089/gtmb.2016.0051

Choi, P. M., Tscharke, B. J., Donner, E., O'Brien, J. W., Grant, S. C., Kaserzon, S. L., … Mueller, J. F. (2018). Wastewater-based epidemiology biomarkers: Past, present and future. *TrAC - Trends in Analytical Chemistry*, *105*, 453–469.

https://doi.org/10.1016/j.trac.2018.06.004

Chow, C.-E. T., & Suttle, C. A. (2015). Biogeography of Viruses in the Sea. *Annual Review of Virology*, *2*(1), 41–66. https://doi.org/10.1146/annurev-virology-031413-085540

Clemente-Casares, P., Pina, S., Buti, M., Jardi, R., Martín, M., Bofill-Mas, S., & Girones, R. (2003). Hepatitis E virus epidemiology in industrialized countries. *Emerging Infectious Diseases*. https://doi.org/10.3201/eid0904.020351

Corominas, L., Collado, N., Guerrero-Latorre, Laura Abasolo-Zabalo, Nerea Anfruns-Estrada, Eduard Anzaldi-Varas, G., Bofill-Mas, S., Bosch, A., Bosch-Lladó, L., … Teichenne-Jané, Joan Torrell-Galceran, Helena Vázquez-Portero, Mario Borrego, C. (2020). Catalan Surveillance Network of SARS-CoV-2 in Sewage.

Corpuz, M. V. A., Buonerba, A., Vigliotta, G., Zarra, T., Ballesteros, F., Campiglia, P., … Naddeo, V. (2020). Viruses in wastewater: occurrence, abundance and detection methods. *Science of the Total Environment*, *745*, 140910. https://doi.org/10.1016/j.scitotenv.2020.140910

Craig Venter, J., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., … Zhu, X. (2001). The sequence of the human genome. *Science*, *291*(5507), 1304–1351. https://doi.org/10.1126/science.1058040

Crane, M., & Hyatt, A. (2011). Viruses of fish: An overview of significant pathogens. *Viruses*. https://doi.org/10.3390/v3112025

Cummings, M. J., Tokarz, R., Bakamutumaho, B., Kayiwa, J., Byaruhanga, T., Owor, N., … O'Donnell, M. R. (2019). Precision Surveillance for Viral Respiratory Pathogens: Virome Capture Sequencing for the

Detection and Genomic Characterization of Severe Acute Respiratory Infection in Uganda. *Clinical Infectious Diseases*. https://doi.org/10.1093/cid/ciy656

de Crom, S. C. M., Rossen, J. W. A., van Furth, A. M., & Obihara, C. C. (2016). Enterovirus and parechovirus infection in children: a brief overview. *European Journal of Pediatrics*, *175*(8), 1023–1029. https://doi.org/10.1007/s00431-016-2725-7

Del Campo, J. A., Parra-Sánchez, M., Figueruela, B., García-Rey, S., Quer, J., Gregori, J., … Romero-Gómez, M. (2018). Hepatitis C virus deep sequencing for sub-genotype identification in mixed infections: A real-life experience. *International Journal of Infectious Diseases : IJID : Official Publication of the International Society for Infectious Diseases*, *67*, 114–117. https://doi.org/10.1016/j.ijid.2017.12.016

Di Bonito, P., Della Libera, S., Petricca, S., Iaconelli, M., Sanguinetti, M., Graffeo, R., … La Rosa, G. (2015). A large spectrum of alpha and beta papillomaviruses are detected in human stool samples. *Journal of General Virology*, *96*(3), 607–613. https://doi.org/10.1099/vir.0.071787-0

Doerries, K. (2006). Human polyomavirus JC and BK persistent infection. *Advances in Experimental Medicine and Biology*. https://doi.org/10.1007/0-387-32957-9_8

Edwards, R. A., Vega, A. A., Norman, H. M., Ohaeri, M., Levi, K., Dinsdale, E. A., … Dutilh, B. E. (2019). Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. *Nature Microbiology*, *4*(10), 1727–1736. https://doi.org/10.1038/s41564-019-0494-6

Elbe, S., & Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges*, *1*(1), 33–46. https://doi.org/10.1002/gch2.1018

Farkas, K., Hillary, L. S., Malham, S. K., McDonald, J. E., & Jones, D. L. (2020). Wastewater and public health: the potential of wastewater surveillance for monitoring COVID-19. *Current Opinion in Environmental Science and Health*, *17*, 14–20. https://doi.org/10.1016/j.coesh.2020.06.001

Farkas, K., Walker, D. I., Adriaenssens, E. M., McDonald, J. E., Hillary, L. S., Malham, S. K., & Jones, D. L. (2020). Viral indicators for tracking domestic wastewater contamination in the aquatic environment. *Water Research*, *181*, 115926. https://doi.org/10.1016/j.watres.2020.115926

Feng, H., Shuda, M., Chang, Y., & Moore, P. S. (2008). Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science*. https://doi.org/10.1126/science.1152586

Feng, Y., Zhang, Y., Ying, C., Wang, D., & Du, C. (2015). Nanopore-based fourth-generation DNA sequencing technology. *Genomics, Proteomics and Bioinformatics*, *13*(1), 4–16. https://doi.org/10.1016/j.gpb.2015.01.009

Fernandez-Cassi, X., Timoneda, N., Martínez-Puchol, S., Rusiñol, M., Rodriguez-Manzano, J., Figuerola, N., … Girones, R. (2018). Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. *Science of The Total Environment*, *618*(August), 870–880. https://doi.org/10.1016/j.scitotenv.2017.08.249

Fernandez-Cassi, Xavier, Rusiñol, M., & Martínez-Puchol, S. (2018). Viral concentration and amplification from human serum samples prior to application of next-generation sequencing analysis. In *Methods in Molecular Biology* (Vol. 1838, pp. 173–188). https://doi.org/10.1007/978-1-4939-8682-8_13

Fioretti, J. M., Fumian, T. M., Rocha, M. S., dos Santos, I. de A. L., Carvalho-Costa, F. A., de Assis, M. R., … Miagostovich, M. P. (2018). Surveillance of Noroviruses in Rio De Janeiro, Brazil: Occurrence of New GIV Genotype in Clinical and Wastewater Samples. *Food and Environmental Virology*, *10*(1), 1–6. https://doi.org/10.1007/s12560-017-9308-2

Fischer, T. K., Rasmussen, L. D., & Fonager, J. (2019). Taking gastro-surveillance into the 21st century. *Journal of Clinical Virology*, *117*(March), 43–48. https://doi.org/10.1016/j.jcv.2019.05.013

Fleischmann R. D., Adams M. D., W. O., Clayton R. A., Kirkness E. F., Kerlavage A. R., Bult C. J., … et al. (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science*.

Fumian, Tulio M., Fioretti, J. M., Lun, J. H., dos Santos, I. A. L., White, P. A., & Miagostovich, M. P. (2019). Detection of norovirus epidemic genotypes in raw sewage using next generation sequencing. *Environment International*, *123*(October 2018), 282–291. https://doi.org/10.1016/j.envint.2018.11.054

Fumian, Tulio Machado, Guimarães, F. R., Vaz, B. J. P., Da Silva, M. T. T., Muylaert, F. F., Bofill-Mas, S., … Miagostovich, M. P. (2010). Molecular detection, quantification and characterization of human polyomavirus JC from waste water in Rio de Janeiro, Brazil. *Journal of Water and Health*, *8*(3), 438–445. https://doi.org/10.2166/wh.2010.090

García-Aljaro, C., Ballesté, E., Muniesa, M., & Jofre, J. (2017). Determination of crAssphage in water samples and applicability for tracking human faecal pollution. *Microbial Biotechnology*. https://doi.org/10.1111/1751-7915.12841

García-López, R., Pérez-Brocal, V., & Moya, A. (2019). Beyond cells – the virome in the human holobiont. *Microbial Cell*, *6*(9), 373–396. https://doi.org/10.15698/mic2019.09.689

Geoghegan, J. L., Di Giallonardo, F., Cousins, K., Shi, M., Williamson, J. E., & Holmes, E. C. (2018). Hidden diversity and evolution of viruses in market fish. *Virus Evolution*. https://doi.org/10.1093/ve/vey031

Glas, C., Hotz, P., & Steffen, R. (2001). Hepatitis A in workers exposed to sewage: A systematic review. *Occupational and Environmental Medicine*. https://doi.org/10.1136/oem.58.12.762

Goecks, J., Nekrutenko, A., Taylor, J., Afgan, E., Ananda, G., Baker, D., … Vincent, K. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*. https://doi.org/10.1186/gb-2010-11-8-r86

Gootenberg, J. S., Abudayyeh, O. O., Kellner, M. J., Joung, J., Collins, J. J., & Zhang, F. (2018). Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Csm6. *Science*, *360*(6387), 439–444. https://doi.org/10.1126/science.aaq0179

Guan, P., & Sung, W. K. (2016). Structural variation detection using next-generation sequencing data: A comparative technical review. *Methods*, *102*, 36–49. https://doi.org/10.1016/j.ymeth.2016.01.020

Guerrero-Latorre, L., Carratala, A., Rodriguez-Manzano, J., Calgua, B., Hundesa, A., & Girones, R. (2011). Occurrence of water-borne enteric viruses in two settlements based in Eastern Chad: Analysis of hepatitis E virus, hepatitis A virus and human adenovirus in water sources. *Journal of Water and Health*. https://doi.org/10.2166/wh.2011.126

Hamza, H., & Hamza, I. A. (2018). Oncogenic papillomavirus and polyomavirus in urban sewage in Egypt. *Science of the Total Environment*, *610–611*, 1413–1420. https://doi.org/10.1016/j.scitotenv.2017.08.218

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

Hata, A., Kitajima, M., Haramoto, E., Lee, S., Ihara, M., Gerba, C. P., & Tanaka, H. (2018). Next-generation amplicon sequencing identifies genetically diverse human astroviruses, including recombinant strains, in environmental waters. *Scientific Reports*, *8*(1), 11837. https://doi.org/10.1038/s41598-018-30217-y

Head, S. R., Kiyomi Komori, H., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*, *56*(2), 61–77. https://doi.org/10.2144/000114133

Hellmér, M., Paxéus, N., Magnius, L., Enache, L., Arnholm, B., Johansson, A., … Norder, H. (2014). Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. *Applied and Environmental Microbiology*.

https://doi.org/10.1128/AEM.01981-14

Hjelmsø, M. H., Mollerup, S., Jensen, R. H., Pietroni, C., Lukjancenko, O., Schultz, A. C., … Hansen, A. J. (2019). Metagenomic analysis of viruses in toilet waste from long distance flights—A new procedure for global infectious disease surveillance. *PLOS ONE*, *14*(1), e0210368. https://doi.org/10.1371/journal.pone.0210368

Honaas, L. A., Wafula, E. K., Wickett, N. J., Der, J. P., Zhang, Y., Edger, P. P., … DePamphilis, C. W. (2016). Selecting superior de novo transcriptome assemblies: Lessons learned by leveraging the best plant genome. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0146062

Iaconelli, M., Petricca, S., Libera, S. Della, Di Bonito, P., & La Rosa, G. (2015). First Detection of Human Papillomaviruses and Human Polyomaviruses in River Waters in Italy. *Food and Environmental Virology*. https://doi.org/10.1007/s12560-015-9203-7

Ibrahim, C., Hammami, S., & Hassen, A. (2019). *Noroviruses, sapoviruses, and aichi viruses emergence in wastewater associated with viral pandemic gastroenteritis. Emerging and Reemerging Viral Pathogens: Volume 1: Fundamental and Basic Virology Aspects of Human, Animal and Plant Pathogens*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-819400-3.00019-3

Illumina Inc. (2020). *Detection and characterization of respiratory viruses, including SARS-CoV-2, using Illumina RNA Prep with Enrichment*.

Instituto de Salud Carlos III. (2018). *Annual Epidemiological Report - Acute Flaccid Paralysis Surveillance and Enterovirus Surveillance*. https://doi.org/10.13140/RG.2.2.35292.90240

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–945.

Iritani, N., Yamamoto, S. P., Abe, N., Kubo, H., Oka, T., & Kaida, A. (2016). Epidemics of GI.2 sapovirus in gastroenteritis outbreaks during 2012-2013 in Osaka City, Japan. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.24451

Izquierdo-Lara, R., Elsinga, G., Heijnen, L., Oude Munnink, B. B., E Schapendonk, C. M., Nieuwenhuijse, D., … de Graaf, M. (2020). Monitoring SARS-CoV-2 circulation and diversity through community wastewater sequencing. *MedRxiv*.

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, *17*(1), 1–11. https://doi.org/10.1186/s13059-016-1103-0

Jebri, S., Muniesa, M., & Jofre, J. (2017). General and host-associated bacteriophage indicators of faecal pollution. In *Global Water Pathogen Project. Part 2 Indicators and Microbial Source Tracking Markers*.

Jeong, H. S., Jeong, A., & Cheon, D. S. (2012). Epidemiology of astrovirus infection in children. *Korean Journal of Pediatrics*, *55*(3), 77–82. https://doi.org/10.3345/kjp.2012.55.3.77

Juárez-González, K., Paredes-Cervantes, V., Gordillo-Rodríguez, S., González-Guzmán, S., Moncayo-Valencia, X., Méndez-Martínez, R., … Vázquez-Vega, S. (2020). High prevalence of human papillomavirus type 66 in low-grade cervical lesions of Mexican women. *Archives of Virology*, *165*(11), 2633–2640. https://doi.org/10.1007/s00705-020-04771-8

Kapgate, S. S., Barbuddhe, S. B., & Kumanan, K. (2015). Next-generation sequencing technologies: Tool to study avian virus diversity. *Acta Virologica*, *59*(1), 3–13. https://doi.org/10.4149/av

Kasianowicz, J. J., Brandin, E., Branton, D., & Deamer, D. W. (1996). Characterization of individual polynucleotide molecules using a membrane channel. In *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.93.24.13770

Khan Mirzaei, M., Xue, J., Costa, R., Ru, J., Schulz, S., Taranu, Z. E., & Deng, L. (2020). Challenges of Studying the Human Virome – Relevant Emerging Technologies. *Trends in Microbiology*, 1–12. https://doi.org/10.1016/j.tim.2020.05.021

Kijak, G. H., Sanders-Buell, E., Pham, P., Harbolick, E. A., Oropeza, C., O'Sullivan, A. M., … Tovanabutra, S. (2019). Next-generation sequencing of HIV-1 single genome amplicons. *Biomolecular Detection and Quantification*. https://doi.org/10.1016/j.bdq.2019.01.002

Kim, J. H., Nelson, K. E., Panzner, U., Kasture, Y., Labrique, A. B., & Wierzba, T. F. (2014). A systematic review of the epidemiology of hepatitis E virus in Africa. *BMC Infectious Diseases*. https://doi.org/10.1186/1471-2334-14-308

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

Klempt, P., Brož, P., Kašný, M., Novotný, A., Kvapilová, K., & Kvapil, P. (2020). Performance of Targeted Library Preparation Solutions for SARS-CoV-2 Whole Genome Analysis. *Diagnostics*, *10*(10), 769.

https://doi.org/10.3390/diagnostics10100769

Kohl, C., Brinkmann, A., Dabrowski, P. W., Radonic, A., Nitsche, A., & Kurth, A. (2015). Protocol for Metagenomic Virus Detection in Clinical Specimens. *Emerging Infectious Diseases*, *21*(1), 48–57. https://doi.org/10.3201/eid2101.140766

Kongprajug, A., Mongkolsuk, S., & Sirikanchana, K. (2019). CrAssphage as a Potential Human Sewage Marker for Microbial Source Tracking in Southeast Asia. *Environmental Science and Technology Letters*. https://doi.org/10.1021/acs.estlett.9b00041

Krishnamurthy, S. R., & Wang, D. (2017). Origins and challenges of viral dark matter. *Virus Research*, *239*, 136–142. https://doi.org/10.1016/j.virusres.2017.02.002

Krishnamurthy, S. R., & Wang, D. (2018). Extensive conservation of prokaryotic ribosomal binding sites in known and novel picobirnaviruses. *Virology*. https://doi.org/10.1016/j.virol.2018.01.006

Kumar, A., Murthy, S., & Kapoor, A. (2017). Evolution of selective-sequencing approaches for virus discovery and virome analysis. *Virus Research*, *239*(June), 172–179. https://doi.org/10.1016/j.virusres.2017.06.005

Kunin, V., He, S., Warnecke, F., Peterson, S. B., Garcia Martin, H., Haynes, M., … Hugenholtz, P. (2008). A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Research*. https://doi.org/10.1101/gr.6835308

La Rosa, G., Pourshaban, M., Iaconelli, M., & Muscillo, M. (2008). Detection of genogroup IV noroviruses in environmental and clinical samples and partial sequencing through rapid amplification of cDNA

ends. *Archives of Virology*, *153*(11), 2077–2083. https://doi.org/10.1007/s00705-008-0241-4

La Rosa, Giuseppina, Della Libera, S., Iaconelli, M., Ciccaglione, A. R., Bruni, R., Taffon, S., … Muscillo, M. (2014). Surveillance of hepatitis A virus in urban sewages and comparison with cases notified in the course of an outbreak, Italy 2013. *BMC Infectious Diseases*. https://doi.org/10.1186/1471-2334-14-419

la Rosa, Giuseppina, Fratini, M., Accardi, L., D'Oro, G., Della Libera, S., Muscillo, M., & Di Bonito, P. (2013). Mucosal and Cutaneous Human Papillomaviruses Detected in Raw Sewages. *PLoS ONE*, *8*(1), 75–77. https://doi.org/10.1371/journal.pone.0052391

Lecuit, M., & Eloit, M. (2013). The human virome: New tools and concepts. *Trends in Microbiology*, *21*(10), 510–515. https://doi.org/10.1016/j.tim.2013.07.001

Leland, D. S., & Ginocchio, C. C. (2007). Role of cell culture for virus detection in the age of technology. *Clinical Microbiology Reviews*. https://doi.org/10.1128/CMR.00002-06

Li, B., Si, H.-R., Zhu, Y., Yang, X.-L., Anderson, D. E., Shi, Z.-L., … Zhou, P. (2020). Discovery of Bat Coronaviruses through Surveillance and Probe Capture-Based Next-Generation Sequencing. *MSphere*, *5*(1), 1–10. https://doi.org/10.1128/msphere.00807-19

Li, T., Unger, E. R., Batra, D., Sheth, M., Steinau, M., Jasinski, J., … Rajeevan, M. S. (2017). Universal human papillomavirus typing assay: Whole-genome sequencing following target enrichment. *Journal of Clinical Microbiology*, *55*(3), 811–823. https://doi.org/10.1128/JCM.02132-16

Lim, X. F., Lee, C. B., Pascoe, S. M., How, C. B., Chan, S., Tan, J. H., … Yap,

G. (2019). Detection and characterization of a novel bat-borne coronavirus in Singapore using multiple molecular approaches. *The Journal of General Virology*, *100*(10), 1363–1374. https://doi.org/10.1099/jgv.0.001307

Lion, T., Kosulin, K., Landlinger, C., Rauch, M., Preuner, S., Jugovic, D., … Matthes-Martin, S. (2010). Monitoring of adenovirus load in stool by real-time PCR permits early detection of impending invasive infection in patients after allogeneic stem cell transplantation. *Leukemia*. https://doi.org/10.1038/leu.2010.4

Lizasoain, A., Burlandy, F. M., Victoria, M., Tort, L. F. L., da Silva, E. E., & Colina, R. (2018). An Environmental Surveillance in Uruguay Reveals the Presence of Highly Divergent Types of Human Enterovirus Species C and a High Frequency of Species A and B Types. *Food and Environmental Virology*, *10*(4), 343–352. https://doi.org/10.1007/s12560-018-9351-7

Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nature Reviews Genetics*, *21*(10), 597–614. https://doi.org/10.1038/s41576-020-0236-x

Lorenzo, M., & Picó, Y. (2019). Wastewater-based epidemiology: current status and future prospects. *Current Opinion in Environmental Science and Health*, *9*, 77–84. https://doi.org/10.1016/j.coesh.2019.05.007

Lynch, J. P., & Kajon, A. E. (2016). Adenovirus: Epidemiology, Global Spread of Novel Serotypes, and Advances in Treatment and Prevention. *Seminars in Respiratory and Critical Care Medicine*. https://doi.org/10.1055/s-0036-1584923

Ma, Z., Li, P., Ji, Y., Ikram, A., & Pan, Q. (2020). Cross-reactivity towards SARS-CoV-2: the potential role of low-pathogenic human

coronaviruses. *The Lancet Microbe*, *1*(4), e151. https://doi.org/10.1016/s2666-5247(20)30098-7

Mardis, E. R. (2008). Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics*, *9*(1), 387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359

Mardis, E. R. (2013). Next-Generation Sequencing Platforms. *Annual Review of Analytical Chemistry*, *6*(1), 287–303. https://doi.org/10.1146/annurev-anchem-062012-092628

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., … Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376–380. https://doi.org/10.1038/nature03959

Matthijnssens, J., Nakagomi, O., Kirkwood, C. D., Ciarlet, M., Desselberger, U., & Van Ranst, M. (2012). Group A rotavirus universal mass vaccination: how and to what extent will selective pressure influence prevalence of rotavirus genotypes? *Expert Review of Vaccines*. https://doi.org/10.1586/erv.12.105

Maurier, F., Beury, D., Fléchon, L., Varré, J. S., Touzet, H., Goffard, A., … Caboche, S. (2019). A complete protocol for whole-genome sequencing of virus from clinical samples: Application to coronavirus OC43. *Virology*. https://doi.org/10.1016/j.virol.2019.03.006

McCombie, W. R., McPherson, J. D., & Mardis, E. R. (2019). Next-generation sequencing technologies. *Cold Spring Harbor Perspectives in Medicine*, *9*(11). https://doi.org/10.1101/cshperspect.a036798

Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., & Brouwer, A. (2020). Presence of SARS-Coronavirus-2 RNA in Sewage and

Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in the Netherlands. *Environmental Science and Technology Letters*. https://doi.org/10.1021/acs.estlett.0c00357

Meng, X. J., & Lindsay, D. S. (2009). Wild boars as sources for infectious diseases in livestock and humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rstb.2009.0086

Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications*. https://doi.org/10.1038/ncomms11257

Mishra, N., Reslan, L., El-Husseini, M., Raoof, H., Finianos, M., Guo, C., … Zaraket, H. (2020). Full genome characterization of human G3P[6] and G3P[9] rotavirus strains in Lebanon. *Infection, Genetics and Evolution*, *78*(September 2019), 104133. https://doi.org/10.1016/j.meegid.2019.104133

Miura, T., Parnaudeau, S., Grodzki, M., Okabe, S., Atmar, R. L., & Le Guyader, F. S. (2013). Environmental detection of genogroup I, II, and IV noroviruses by using a generic real-time reverse transcription-PCR assay. *Applied and Environmental Microbiology*, *79*(21), 6585–6592. https://doi.org/10.1128/AEM.02112-13

Miyazato, P., Katsuya, H., Fukuda, A., Uchiyama, Y., Matsuo, M., Tokunaga, M., … Satou, Y. (2016). Application of targeted enrichment to next-generation sequencing of retroviruses integrated into the host human genome. *Scientific Reports*, *6*(June), 1–10. https://doi.org/10.1038/srep28324

Moore, N. E., Wang, J., Hewitt, J., Croucher, D., Williamson, D. A., Paine, S., … Hall, R. J. (2015). Metagenomic analysis of viruses in feces from unsolved outbreaks of gastroenteritis in humans. *Journal of Clinical*

Microbiology, *53*(1), 15–21. https://doi.org/10.1128/JCM.02029-14

Muniesa, M., Payan, A., Moce-Llivina, L., Blanch, A. R., & Jofre, J. (2009). Differential persistence of F-specific RNA phage subgroups hinders their use as single tracers for faecal source tracking in surface water. *Water Research*. https://doi.org/10.1016/j.watres.2008.12.038

Mutoloki, S., & Evensen, Ø. (2011). Sequence similarities of the capsid gene of Chilean and European isolates of infectious pancreatic necrosis virus point towards a common origin. *Journal of General Virology*, *92*(7), 1721–1726. https://doi.org/10.1099/vir.0.030270-0

Muzzey, D. (2018). Understanding the Basics of NGS in the Context of NIPT. In *Noninvasive Prenatal Testing (NIPT): Applied Genomics in Prenatal Screening and Diagnosis*. https://doi.org/10.1016/B978-0-12-814189-2.00002-5

Myers, J., & Rothberg, J. (2011). Semiconductor Sequencing for Life. *Journal of Biomolecular Techniques : JBT*, *22*(Suppl), S41–S42.

Nakamura, T., Hamasaki, M., Yoshitomi, H., Ishibashi, T., Yoshiyama, C., Maeda, E., … Yoshida, H. (2015). Environmental surveillance of poliovirus in sewage water around the introduction period for inactivated polio vaccine in Japan. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.03575-14

Nasir, J. A., Kozak, R. A., Aftanas, P., Raphenya, A. R., Smith, K. M., Maguire, F., … Mubareka, S. (2020). A comparison of whole genome sequencing of sars-cov-2 using amplicon-based sequencing, random hexamers, and bait capture. *Viruses*, *12*(8). https://doi.org/10.3390/v12080895

National Center for Biotechnology Information (NCBI). Bethesda (MD). (1988).

Ng, T. F. F., Marine, R., Wang, C., Simmonds, P., Kapusinszky, B., Bodhidatta, L., … Delwart, E. (2012). High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *Journal of Virology*, *86*(22), 12161–12175. https://doi.org/10.1128/JVI.00869-12

No, J. S., Kim, W. K., Cho, S., Lee, S. H., Kim, J. A., Lee, D., … Song, J. W. (2019). Comparison of targeted next-generation sequencing for whole-genome sequencing of Hantaan orthohantavirus in Apodemus agrarius lung tissues. *Scientific Reports*. https://doi.org/10.1038/s41598-019-53043-2

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). MetaSPAdes: A new versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834. https://doi.org/10.1101/gr.213959.116

Oba, M., Tsuchiaka, S., Omatsu, T., Katayama, Y., Otomaru, K., Hirata, T., … Mizutani, T. (2018). A new comprehensive method for detection of livestock-related pathogenic viruses using a target enrichment system. *Biochemical and Biophysical Research Communications*, *495*(2), 1871–1877. https://doi.org/10.1016/j.bbrc.2017.12.017

Olijve, L., Jennings, L., & Walls, T. (2018). Human Parechovirus : an Increasingly Recognized Cause of. *Clinical Microbiology Reviews*, *31*(Nov 2017, 31 (1) e00047-17), 1–17. Retrieved from https://doi.org/10.1128/CMR.00047-17.

Oppong, T. B., Yang, H., Amponsem-Boateng, C., Kyere, E. K. D., Abdulai, T., Duan, G., & Opolot, G. (2020). Enteric pathogens associated with gastroenteritis among children under 5 years in sub-saharan Africa: A systematic review and meta-Analysis. *Epidemiology and Infection*. https://doi.org/10.1017/S0950268820000618

Palacios, G., Druce, J., Du, L., Tran, T., Birch, C., Briese, T., … Lipkin, W.

I. (2008). A New Arenavirus in a Cluster of Fatal Transplant-Associated Diseases. *New England Journal of Medicine*. https://doi.org/10.1056/nejmoa073785

Parsley, L. C., Consuegra, E. J., Thomas, S. J., Bhavsar, J., Land, A. M., Bhuiyan, N. N., … Liles, M. R. (2010). Census of the viral metagenome within an activated sludge microbial assemblage. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.02520-09

Payne, S. (2017). Family Astroviridae. In *Viruses*. https://doi.org/10.1016/b978-0-12-803109-4.00014-3

Pina, S., Jofre, J., Emerson, S. U., Purcell, R. H., & Girones, R. (1998). Characterization of a strain of infectious hepatitis E virus isolated from sewage in an area where hepatitis E is not endemic. *Applied and Environmental Microbiology*. https://doi.org/10.1128/aem.64.11.4485-4488.1998

Pina, S., Puig, M., Lucena, F., Jofre, J., & Girones, R. (1998). Viral pollution in the environment and in shellfish: Human adenovirus detection by PCR as an index of human viruses. *Applied and Environmental Microbiology*. https://doi.org/10.1128/aem.64.9.3376-3382.1998

Pintó, R. M., Alegre, D., Domíngueza, A., El-Senousy, W. M., Sánchez, G., Villena, C., … Bosch, A. (2007). Hepatitis A virus in urban sewage from two Mediterranean countries. *Epidemiology and Infection*. https://doi.org/10.1017/S0950268806006753

Popgeorgiev, N., Temmam, S., Raoult, D., & Desnues, C. (2013). Describing the silent human virome with an emphasis on giant viruses. *Intervirology*, *56*(6), 395–412. https://doi.org/10.1159/000354561

Prevost, B., Lucas, F. S., Ambert-Balay, K., Pothier, P., Moulin, L., & Wurtzer, S. (2015). Deciphering the Diversities of Astroviruses and Noroviruses in Wastewater Treatment Plant Effluents by a High-Throughput Sequencing Method. *Applied and Environmental Microbiology*, *81*(20), 7215–7222. https://doi.org/10.1128/AEM.02076-15

Purdy, M. A., Harrison, T. J., Jameel, S., Meng, X. J., Okamoto, H., Van Der Poel, W. H. M., & Smith, D. B. (2017). ICTV virus taxonomy profile: Hepeviridae. *Journal of General Virology*. https://doi.org/10.1099/jgv.0.000940

Quijada, N. M., Fongaro, G., Barardi, C. R. M., Hernández, M., & Rodríguez-Lázaro, D. (2016). Propidium monoazide integrated with qPCR enables the detection and enumeration of infectious enteric RNA and DNA viruses in clam and fermented sausages. *Frontiers in Microbiology*. https://doi.org/10.3389/fmicb.2016.02008

Radford D., A., Chapman, D., Dixon, L., & Chantrey, J. (2012). Application of Next-Generation Sequencing Technologies in Virology. *Journal of General Virology*, (May), 401–412. https://doi.org/10.1016/B978-0-12-801498-1.00038-3

Rames, E., Roiko, A., Stratton, H., & Macdonald, J. (2016). Technical aspects of using human adenovirus as a viral water quality indicator. *Water Research*. https://doi.org/10.1016/j.watres.2016.03.042

Randazzo, W., Khezri, M., Ollivier, J., Le Guyader, F. S., Rodríguez-Díaz, J., Aznar, R., & Sánchez, G. (2018). Optimization of PMAxx pretreatment to distinguish between human norovirus with intact and altered capsids in shellfish and sewage samples. *International Journal of Food Microbiology*. https://doi.org/10.1016/j.ijfoodmicro.2017.11.011

Randazzo, W., Truchado, P., Cuevas-Ferrando, E., Simón, P., Allende, A., & Sánchez, G. (2020). SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water Research*. https://doi.org/10.1016/j.watres.2020.115942

Reshetnyak, V. I., Maev, I. V., Burmistrov, A. I., Chekmazov, I. A., & Karlovich, T. I. (2020). Torque teno virus in liver diseases: On the way towards unity of view. *World Journal of Gastroenterology*. https://doi.org/10.3748/WJG.V26.I15.1691

Rimoldi, S. G., Stefani, F., Pagani, C., Chenal, L. L., Zanchetta, N., Bartolo, I. Di, … Gismondo, M. R. (2011). Epidemiological and clinical characteristics of pediatric gastroenteritis associated with new viral agents. *Archives of Virology*. https://doi.org/10.1007/s00705-011-1037-5

Rivadulla, E., & Romalde, J. L. (2020). A Comprehensive Review on Human Aichi Virus. *Virologica Sinica*, *12250*. https://doi.org/10.1007/s12250-020-00222-5

Rodriguez-Manzano, J., Miagostovich, M., Hundesa, A., Clemente-Casares, P., Carratala, A., Buti, M., … Girones, R. (2010). Analysis of the evolution in the circulation of HAV and HEV in eastern Spain by testing urban sewage samples. *Journal of Water and Health*, *8*(2), 346–354. https://doi.org/10.2166/wh.2009.042

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

Ruggeri, F. M., Bonomo, P., Ianiro, G., Battistone, A., Delogu, R., Germinario, C., … Fiore, L. (2015). Rotavirus genotypes in sewage treatment plants and in children hospitalized with acute diarrhea in Italy in 2010 and 2011. *Applied and Environmental Microbiology*.

https://doi.org/10.1128/AEM.02695-14

Rusiñol, M., Fernandez-Cassi, X., Timoneda, N., Carratalà, A., Abril, J. F., Silvera, C., … Girones, R. (2015). Evidence of viral dissemination and seasonality in a Mediterranean river catchment: Implications for water pollution management. *Journal of Environmental Management*, *159*, 58–67. https://doi.org/10.1016/j.jenvman.2015.05.019

Rusiñol, M., & Girones, R. (2017). Summary of Excreted and Waterborne Viruses. In J. S. Meschke & R. Girones (Eds.), *Global Water Pathogen Project. Part 3 Viruses*. Michigan State University. https://doi.org/10.14321/waterpathogens.19

Rusiñol, M., Martínez-Puchol, S., Timoneda, N., Fernández-Cassi, X., Pérez-Cataluña, A., Fernández-Bravo, A., … Girones, R. (2020). Metagenomic analysis of viruses, bacteria and protozoa in irrigation water. *International Journal of Hygiene and Environmental Health*, *224*(December 2019), 113440. https://doi.org/10.1016/j.ijheh.2019.113440

Sahul Hameed, A. S., Ninawe, A. S., Nakai, T., Chi, S. C., & Johnson, K. L. (2019). ICTV virus taxonomy profile: Nodaviridae. *Journal of General Virology*. https://doi.org/10.1099/jgv.0.001170

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Sinha, A., & Dutta, S. (2019). Waterborne & foodborne viral hepatitis: A public health perspective. *Indian Journal of Medical Research*. https://doi.org/10.4103/ijmr.IJMR_1430_18

Smelov, V., Elfström, K. M., Eklund, C., Sokolova, O., & Dillner, J. (2018). Determinants of the presence of human papillomaviruses in the anal canal of Russian men. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.25234

Strubbia, S., Phan, M. V. T., Schaeffer, J., Koopmans, M., Cotten, M., & Le Guyader, F. S. (2019). Characterization of Norovirus and Other Human Enteric Viruses in Sewage and Stool Samples Through Next-Generation Sequencing. *Food and Environmental Virology*, *11*(4), 400–409. https://doi.org/10.1007/s12560-019-09402-3

Strubbia, S., Schaeffer, J., Oude Munnink, B. B., Besnard, A., Phan, M. V. T., Nieuwenhuijse, D. F., … Le Guyader, F. S. (2019). Metavirome sequencing to evaluate norovirus diversity in sewage and related bioaccumulated oysters. *Frontiers in Microbiology*, *10*(OCT). https://doi.org/10.3389/fmicb.2019.02394

Takahashi, M., Tanaka, T., Azuma, M., Kusano, E., Aikawa, T., Shibayama, T., … Okamoto, H. (2007). Prolonged fecal shedding of hepatitis E virus (HEV) during sporadic acute hepatitis E: Evaluation of infectivity of HEV in fecal specimens in a cell culture system. *Journal of Clinical Microbiology*. https://doi.org/10.1128/JCM.01086-07

Tan, G., Opitz, L., Schlapbach, R., & Rehrauer, H. (2019). Long fragments achieve lower base quality in Illumina paired-end sequencing. *Scientific Reports*. https://doi.org/10.1038/s41598-019-39076-7

Tatte, V. S., & Gopalkrishna, V. (2019). Detection of different enteric viruses in children with diarrheal disease: evidence of the high frequency of mixed infections. *Access Microbiology*, *1*(2), 1–7. https://doi.org/10.1099/acmi.0.000010

Thongprachum, A., Fujimoto, T., Takanashi, S., Saito, H., Okitsu, S., Shimizu, H., … Ushijima, H. (2018). Detection of nineteen enteric

viruses in raw sewage in Japan. *Infection, Genetics and Evolution*. https://doi.org/10.1016/j.meegid.2018.05.006

Tiwari, S., & Dhole, T. N. (2018). Assessment of enteroviruses from sewage water and clinical samples during eradication phase of polio in North India. *Virology Journal*. https://doi.org/10.1186/s12985-018-1075-7

Tokarz, R., Hyams, J. S., Mack, D. R., Boyle, B., Griffiths, A. M., Leleiko, N. S., … Lipkin, W. I. (2019). Characterization of Stool Virome in Children Newly Diagnosed with Moderate to Severe Ulcerative Colitis. *Inflammatory Bowel Diseases*. https://doi.org/10.1093/ibd/izz099

Twist Bioscience. (2020). *NGS-Based Target Capture for SARS-CoV-2 Detection and Characterization*.

Uelze, L., Borowiak, M., Bönn, M., Brinks, E., Deneke, C., Hankeln, T., … Malorny, B. (2020). German-Wide Interlaboratory Study Compares Consistency, Accuracy and Reproducibility of Whole-Genome Short Read Sequencing. *Frontiers in Microbiology*, *11*(September), 1–14. https://doi.org/10.3389/fmicb.2020.573972

Van der Hoek, L. (2015). Human coronaviruses: What do they cause? *Antiviral Therapy*, *12*(March).

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends in Genetics*. https://doi.org/10.1016/j.tig.2018.05.008

Van Leeuwen, M., Williams, M. M. W., Koraka, P., Simon, J. H., Smits, S. L., & Osterhaus, A. D. M. E. (2010). Human picobirnaviruses identified by molecular screening of diarrhea samples. *Journal of Clinical Microbiology*. https://doi.org/10.1128/JCM.02452-09

Varela, M. F., Rivadulla, E., Lema, A., & Romalde, J. L. (2019). Human sapovirus among outpatients with acute gastroenteritis in Spain: A one-year study. *Viruses*, *11*(2). https://doi.org/10.3390/v11020144

Varga, E., Kiss, M., Szabó, K., & Kemény, L. (2009). Detection of merkel cell polyomavirus DNA in merkel cell carcinomas. *British Journal of Dermatology*, *161*(4), 930–932. https://doi.org/10.1111/j.1365-2133.2009.09221.x

Vilsker, M., Moosa, Y., Nooij, S., Fonseca, V., Ghysens, Y., Dumon, K., … de Oliveira, T. (2018). Genome Detective: an automated system for virus identification from high-throughput sequencing data. *Bioinformatics*, 1–3. https://doi.org/10.1093/bioinformatics/bty695

Vincent, A. T., Derome, N., Boyle, B., Culley, A. I., & Charette, S. J. (2017). Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. *Journal of Microbiological Methods*, *138*, 60–71. https://doi.org/10.1016/j.mimet.2016.02.016

Vinjé, J., Estes, M. K., Esteves, P., Green, K. Y., Katayama, K., Knowles, N. J., … Ictv Report Consortium. (2019). ICTV Virus Taxonomy Profile: Caliciviridae. *The Journal of General Virology*, *100*(11), 1469–1470. https://doi.org/10.1099/jgv.0.001332

Vinjé, J., & Koopmans, M. P. G. (2000). Simultaneous detection and genotyping of "Norwalk-like viruses" by oligonucleotide array in a reverse line blot hybridization format. *Journal of Clinical Microbiology*. https://doi.org/10.1128/jcm.38.7.2595-2601.2000

Wang, D. (2020). 5 Challenges in Understanding the Role of the Virome in Health and Disease. *PLoS Pathogens*, *16*(3), 1–6. https://doi.org/10.1371/journal.ppat.1008318

Wang, H., Neyvaldt, J., Enache, L., Sikora, P., Mattsson, A., Johansson, A., … Norder, H. (2020). Variations among Viruses in Influent Water and Effluent Water at a Wastewater Plant over One Year as Assessed by Quantitative PCR and Metagenomics. *Applied and Environmental Microbiology*, *86*(24), 1–14. https://doi.org/10.1128/AEM.02073-20

Wetterstrand, K. (2020). DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Retrieved from www.genome.gov/sequencingcostsdata

Williams, S. H., Cordey, S., Bhuva, N., Laubscher, F., Hartley, M.-A., Boillat-Blanco, N., … Lipkin, W. I. (2018). Investigation of the Plasma Virome from Cases of Unexplained Febrile Illness in Tanzania from 2013 to 2014: a Comparative Analysis between Unbiased and VirCapSeq-VERT High-Throughput Sequencing Approaches. *MSphere*. https://doi.org/10.1128/msphere.00311-18

Wölfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M. A., … Wendtner, C. (2020). Virological assessment of hospitalized patients with COVID-2019. *Nature*, *581*(7809), 465–469. https://doi.org/10.1038/s41586-020-2196-x

Wu, Y., Guo, C., Tang, L., Hong, Z., Zhou, J., Dong, X., … Huang, X. (2020). Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *The Lancet Gastroenterology and Hepatology*, *5*(5), 434–435. https://doi.org/10.1016/S2468-1253(20)30083-2

Wylie, K. M., Weinstock, G. M., & Storch, G. A. (2012). Emerging view of the human virome. *Translational Research*, *160*(4), 283–290. https://doi.org/10.1016/j.trsl.2012.03.006

Wylie, T. N., Wylie, K. M., Herter, B. N., Storch, G. A., & Wylie, T. N. (2015). Enhanced Virome Sequencing Using Targeted Sequence

Capture. *Genome*, (October).

Xiao, M., Liu, X., Ji, J., Li, M., Li, J., Yang, L., … Li, J. (2020). Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Medicine*, *12*(1), 1–15. https://doi.org/10.1186/s13073-020-00751-4

Zárate, S., Taboada, B., Yocupicio-Monroy, M., & Arias, C. F. (2017). Human Virome. *Archives of Medical Research*, *48*(8), 701–716. https://doi.org/10.1016/j.arcmed.2018.01.005

Zhang, T., Breitbart, M., Lee, W. H., Run, J. Q., Wei, C. L., Soh, S. W. L., … Ruan, Y. (2006). RNA viral community in human feces: Prevalence of plant pathogenic viruses. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.0040003

Zhang, Y., Chen, C., Song, Y., Zhu, S., Wang, D., Zhang, H., … Xu, W. (2020). Excretion of SARS-CoV-2 through faecal specimens. *Emerging Microbes & Infections*. https://doi.org/10.1080/22221751.2020.1844551

Zhou, Y., Wylie, K. M., Feghaly, R. E. E., Mihindukulasuriya, K. A., Elward, A., Haslam, D. B., … Weinstock, G. M. (2016). Metagenomic approach for identification of the pathogens associated with diarrhea in stool specimens. *Journal of Clinical Microbiology*, *54*(2), 368–375. https://doi.org/10.1128/JCM.01965-15

ANNEXOS

Material suplementari de l'Article 1

Viral species obtained by Target Enrichment Sequencing and Untargeted Viral Metagenomics in raw sewage from WWTP A.

Host	Family	Genus/Specie	
		Target Enrichment Sequencing	Untargeted Viral Metagenomics
Human	Adenoviridae	Human mastadenovirus A (HAdV 31) Human mastadenovirus C (HAdV 5) Human mastadenovirus D (HAdV 27, HAdV 51, HAdV 56) Human mastadenovirus F (HAdV 40, HAdV 41)	
	Astroviridae	Mamastrovirus 1	Mamastrovirus 1
	Caliciviridae	Sapporo virus (SV GI.2, SV GII.1, SV GII.2, SV GII.3) Norwalk virus (NV G II.17)	Sapporo virus (SV GI.2) Norwalk virus (NV G II.17)
	Circoviridae	Human fecal virus Jorvi2	Human fecal virus Jorvi2
	Hepeviridae	Hepatitis E virus	
	Papillomaviridae	Betapapillomavirus 2 (HPV 17, HPV 122) Betapapillomavirus 3 (HPV 49) Gammapapillomavirus 12 (HPV 127)	
	Parvoviridae	Adeno-associated virus (AAV2, AAV3) Human bocavirus 3 Parvovirus NIH-CQV	 Parvovirus NIH-CQV
	Picobirnaviridae	Human picobirnavirus Human circovirus VS6600022 Human associated cyclovirus 10	 Human circovirus VS6600022
	Picornaviridae	Aichivirus A Cardiovirus (Human TMEV-like cardiovirus, Saffold virus) Cosavirus A Enterovirus A (CV-A16, EV-A71) Enterovirus B (E-E30, CV-B5, CV-A9) Enterovirus C (CV-A22) Hepatovirus A(HAV) Rhinovirus A (RV-A61) Salivirus A	Aichivirus A
	Polyomaviridae	BK polyomavirus Human polyomavirus 6 JC polyomavirus WU polyomavirus	 JC polyomavirus
	Reoviridae	Rotavirus A	
	Others	Hudisavirus, Gemykibivirus	Hudisavirus, Gemykibivirus
Other Vertebrates	Adenoviridae	Simian mastadenovirus F (Simian adenovirus 18)	
	Astroviridae	Canine astrovirus Porcine astrovirus 4 Qinghai Himalayan marmot astrovirus 2 Rodent astrovirus	
	Circoviridae	Bat associated circovirus 2 Canary circovirus Canine circovirus Pigeon circovirus Porcine stool-associated virus	 Porcine stool-associated virus Duck circovirus Finch circovirus Gull circovirus Fur seal faeces associated virus
	Dicistroviridae	Goose dicistrovirus	
	Hepeviridae	Orthohepevirus C	
	Papillomaviridae	Iotapapillomavirus 1	
	Parvoviridae	Bovine adeno-associated virus Canine bocavirus 1 Feline bocaparvovirus 2 Feline panleukopenia virus Parus major densovirus Porcine bocavirus H18 Primate bocaparvovirus 2 Rat parvovirus 1 Rodent protoparvovirus 1 Ungulate protoparvovirus 1	 Porcine bocavirus H18
	Picobirnaviridae	Picobirnavirus dog/KNA/2015 Picobirnavirus green monkey/KNA/2015 Porcine picobirnavirus	 Porcine picobirnavirus
	Picornaviridae	Cardiovirus B Cardiovirus C Rabovirus A Rosavirus B	 Rosavirus B

		Hunnivirus A	Hunnivirus A
	Others	Rasavirus	Rasavirus Bovine faeces associated virus Chimpanzee faeces associated virus Duck faeces associated virus Gemycircularvirus Gemykibivirus
Plant	Potyviridae	Bell pepper mottle virus	Bell pepper mottle virus
	Secoviridae	Broad bean wilt virus 2 Cowpea mosaic virus Red clover mottle virus Squash mosaic virus Tomato mosaic virus	 Squash mosaic virus Tomato mosaic virus
	Virgaviridae	Cucumber green mottle mosaic virus Pepper mild mottle virus Tobacco mild green mosaic virus Tobacco mosaic virus	Cucumber green mottle mosaic virus Pepper mild mottle virus Tobacco mild green mosaic virus Tobacco mosaic virus Paprika mild mottle virus Tomato brown rugose fruit virus Turnip vein-clearing virus
	Tombusviridae	Honeysuckle ringspot virus Tomato bushy stunt virus	Honeysuckle ringspot virus Tomato bushy stunt virus Carnation mottle virus Japanese iris necrotic ring virus Olive mild mosaic virus Saguaro cactus virus
	Alphaflexiviridae	Pepino mosaic virus	Pepino mosaic virus
	Tymoviridae	Turnip yellow mosaic virus	Turnip yellow mosaic virus Clover yellow mosaic virus Scrophularia mottle virus
	Solemoviridae	Velvet tobacco mottle virus	 Cymbidium chlorotic mosaic virus Lucerne transient streak virus Rottboellia yellow mottle virus Southern bean mosaic virus Soybean yellow common mosaic virus
	Genomoviridae		Soybean leaf-associated gemycircularvirus 1
Invertebrates	Parvoviridae	Acheta domestica densovirus Blattella germanica densovirus 1 Decapod penstyldensovirus 1 Diaphorina citri densovirus Dipteran brevidensovirus 1 Junonia coenia densovirus Penaeus stylirostris penstyldensovirus 1 Periplaneta fuliginosa densovirus	 Cherax quadricarinatus densovirus Decapod penstyldensovirus 1 Mythimna loreyi densovirus Periplaneta fuliginosa densovirus Solenopsis invicta densovirus
	Dicistroviridae	Aphid lethal paralysis virus Drosophila C virus Kashmir bee virus Taura syndrome virus	Aphid lethal paralysis virus Drosophila C virus
	Circoviridae	Avon-Heathcote Estuary associated circular virus	Avon-Heathcote Estuary associated circular virus Cockroach associated cyclovirus 1 Fiddler Crab associated circular virus Hermit crab associated circular virus
	Genomoviridae	Dragonfly-associated circular virus Thrips-associated genomovirus 3	Dragonfly-associated circular virus
	Nodaviridae	Le Blanc nodavirus Orsay virus	 Orsay virus
	Totiviridae		Armigeres subalbatus virus
	Other	Beihai narna-like virus Beihai picorna-like virus Changjiang crawfish virus	Beihai narna-like virus Beihai picorna-like virus Beihai tombus-like virus Beihai mantis shrimp virus Beihai zhaovirus-like virus Changjiang crawfish virus Changjiang narna-like virus Changjiang picorna-like virus Changjiang tombus-like virus Ctenophore-associated virus Dragonfly circularisvirus

			Dragonfly cyclicusvirus
			Dragonfly larvae associated virus
		Hubei arthropod virus 1	Hubei arthropod virus 1
		Hubei picorna-like virus	Hubei picorna-like virus
			Hubei hepe-like virus 2
			Hubei leech virus 4
			Hubei narna-like virus 11
			Hubei odonate virus 1
			Hubei tetragnatha maxillosa virus 1
			Hubei tombus-like virus 25
		Lake Sarah-associated virus	Lake Sarah-associated virus
			Sanxia picorna-like virus 9
			Sanxia water strider virus 17
			Shahe picorna-like virus
			Wenling crustacean virus 3
		Wenzhou bivalvia virus 3	Wenzhou bivalvia virus 3
		Wenzhou picorna-like virus	Wenzhou picorna-like virus
			Wenzhou channeled applesnail virus 3
			Wenzhou narna-like virus 5
			Wenzhou tombus-like virus
Fungi	Marnaviridae	Aurantiochytrium virus	
	Podoviridae	Actinomyces virus Av1	
	Partiviridae		Heterobasidion partitivirus 8
			Pleurotus ostreatus virus 1
		Rosellinia necatrix partitivirus 6	Rosellinia necatrix partitivirus 6
	Others		Fusarium poae fusarivirus 1
		Phytophthora parasitica virus	Phytophthora parasitica virus
		Sclerophthora macrospora virus A	Sclerophthora macrospora virus A
		Acinetobacter phage AB3	
			Achromobacter phage JWAlpha
			Achromobacter phage phiAxp-3
			Acinetobacter phage Petty
			Acinetobacter phage Presley
		Acinetobacter phage vB_AbaP_Acibel007	
			Actinomyces virus Av1
		Alteromonas phage vB_AmaP_AD45-P1	Alteromonas phage vB_AmaP_AD45-P1
			Bacillus phage MG-B1
			Bacillus phage Stitch
			Bacillus phage VMY22
		Bacillus virus B103	Bacillus virus B103
			Bordetella virus BPP1
			Burkholderia virus DC1
			Celeribacter phage P12053L
			Citrobacter phage CVT22
			Erwinia phage Ea9-2
		Erwinia phage PEp14	Erwinia phage PEp14
		Erwinia phage vB_EamP-S6	Erwinia phage vB_EamP-S6
			Erwinia virus Frozen
			Escherichia phage Bp4
		Escherichia phage CICC 80001	
			Escherichia phage ECBP2
			Escherichia phage IME11
	Podoviridae		Escherichia phage KBNP1711
			Escherichia phage Pollock
			Escherichia phage vB_EcoP_PhAPEC5
			Escherichia phage vB_EcoP_SU10
		Escherichia virus K1-5	Escherichia virus K1-5
			Escherichia virus K1E
			Escherichia virus N4
			Escherichia virus phiEco32
		Hollowayvirus	Hollowayvirus
			Klebsiella phage F19
			Klebsiella phage K11
			Lactococcus virus KSY1
			Puniceispirillum phage HMO-2011
			Rhodoferax phage P26218
			Salinivibrio phage CW02
			Salmonella phage FSL SP-058
			Silicibacter phage DSS3phi2
		Sinorhizobium phage PBC5	
			Streptococcus phage Cp-1
			Vibrio phage phi 1
			Vibrio phage VBP32

			Xylella phage Prado
			Xylella phage Xfas53
		Acinetobacter phage Acj61	Acinetobacter phage Acj61
			Acinetobacter phage LZ35
			Acinetobacter phage phiAC-1
			Acinetobacter phage vB_AbaM_Acibel004
			Acinetobacter phage vB_AbaM_phiAbaA1
			Acinetobacter phage YMC13/03/R2096
			Acinetobacter phage YMC-13-01-C62
			Acinetobacter virus 133
			Aeromonas phage Aes508
			Aeromonas phage CC2
			Aeromonas phage pAh6-C
			Aeromonas phage phiAS4
		Aeromonas phage vB_AsaM-56	Aeromonas phage vB_AsaM-56
			Bdellovibrio phage phi1402
			Citrobacter phage Moon
			Clostridium phage c-st
	Myoviridae	Cronobacter phage CR8	
		Enterobacteria phage phi92	
			Enterobacteria phage phiEcoM-GJ1
			Erwinia phage vB_EamM_ChrisDB
			Escherichia phage phAPEC8
			Escherichia virus P1
			Flavobacterium phage FCL-2
			Klebsiella phage JD001
			Klebsiella phage KP15
			Klebsiella phage Matisse
			Klebsiella phage vB_KpnM_KB57
			Pseudomonas phage PA11
			Pseudomonas phage PaBG
			Pseudomonas phage VCM
			Psychrobacter phage pOW20-A
			Shewanella phage Spp001
			Shewanella sp. phage 1/41
		Shewanella sp. phage 3/49	Shewanella sp. phage 3/49
	Leviviridae	Acinetobacter phage Acj61	Acinetobacter phage AP205
			Allolevivirus
			Acinetobacter phage Bphi-B1251
		Acinetobacter phage vB_AbaS_TRS1	Acinetobacter phage vB_AbaS_TRS1
			Arthrobacter phage Decurro
			Arthrobacter phage Mudcat
			Bacillus phage BCJA1c
			Bacillus phage vB_BanS-Tsamsa
			Bacteroides phage B124-14
			Bacteroides phage B40-8
			Brochothrix phage NF5
			Burkholderia phage Bcep176
			Burkholderia phage KL1
			Cellulophaga phage phi10:1
		Cellulophaga phage phi39:1	Cellulophaga phage phi39:1
			Citrobacter virus Stevie
			Clostridium phage phiCT453B
			Croceibacter phage P2559S
		Croceibacter phage P2559Y	Croceibacter phage P2559Y
			Enterobacter phage phiEap-2
			Enterococcus phage EfaCPT1
			Enterococcus phage EFAP-1
		Enterococcus phage EFC-1	Enterococcus phage EFC-1
			Enterococcus phage IME-EFm5
			Enterococcus phage phiFL4A
			Escherichia phage Envy
Bacteria			Escherichia phage K1-dep
			Escherichia phage vB_Eco_ACG-M12
			Escherichia virus Rtp
			Flavobacterium phage 11b
			Geobacillus virus E3
			Gordonia phage BritBrat
			Gordonia phage GMA6
			Gordonia phage GRU3
			Gordonia phage Terapin

			Idiomarinaceae phage 1N2-2
			Idiomarinaceae phage Phi1M2-2
			Klebsiella phage 1513
		Klebsiella phage KP36	
			Klebsiella phage phiKO2
			Lactobacillus phage Ld3
		Lactobacillus virus phiJL1	
			Lactococcus phage 1706
			Lactococcus phage 949
			Lactococcus phage D4412
		Lactococcus phage M5938	
			Lactococcus phage phi7
	Siphoviridae		Lactococcus phage Q54
			Lactococcus phage WRP3
			Lactococcus virus bIL170
			Lactococcus virus c2
			Lactococcus virus Sl4
			Leuconostoc phage phiLN12
			Leuconostoc phage phiLN25
			Listeria phage vB_LmoS_188
			Methanobacterium phage psiM2
			Mycobacterium phage Lolly9
			Polaribacter phage P12002L
			Pseudoalteromonas phage H103
			Pseudomonas phage NP1
			Pseudomonas phage PaMx11
			Pseudomonas phage phiPSA1
		Pseudomonas phage phiPto-bp6g	Pseudomonas phage phiPto-bp6g
			Pseudomonas phage PS-1
			Pseudomonas virus Ab18
			Pseudomonas virus Yua
		Psychrobacter phage Psymv2	Psychrobacter phage Psymv2
			Rhizobium phage vB_RglS_P106B
			Rhodobacter phage RcTitan
			Rhodococcus phage ReqiPepy6
			Rhodococcus phage ReqiPoco6
			Salmonella phage 64795_sal3
			Salmonella phage 7-11
			Salmonella phage IME207
			Salmonella phage Jersey
		Salmonella phage vB_SosS_Oslo	
			Salmonella virus SPN3US
		Serratia phage Eta	Serratia phage Eta
			Shigella phage pSf-1
			Stenotrophomonas phage S1
		Streptococcus phage 315.2	Streptococcus phage 315.2
		Streptococcus phage Str-PAP-1	Streptococcus phage Str-PAP-1
			Streptococcus phage YMC-2011
			Streptococcus virus 858
			Streptococcus virus ALQ132
			Streptococcus virus DT1
			Streptococcus virus O1205
			Streptococcus virus phiAbc2
			Streptomyces phage Jay2Jay
		Synechococcus phage S-CBS3	Synechococcus phage S-CBS3
			Vibrio phage pYD21-A
			Vibrio phage pYD38-A
			Xanthomonas phage Xp15
		Yersinia phage phiR201	Yersinia phage phiR201
		Bdellovibrio phage phiMH2K	Bdellovibrio phage phiMH2K
			Chlamydia phage 4
		Chlamydia virus Chp1	Chlamydia virus Chp1
		Chlamydia virus Chp2	Chlamydia virus Chp2
			Chlamydia virus CPAR39
		Enterobacteria phage alpha3	
		Enterobacteria phage ID18 sensu lato	Enterobacteria phage ID18 sensu lato
			Enterobacteria phage ID2 Moscow/ID/2001
			Enterobacteria phage St-1
		Enterobacteria phage WA13 sensu lato	Enterobacteria phage WA13 sensu lato
		Escherichia virus phiX174	Escherichia virus phiX174
		Gokushovirinae Bog1183_53	Gokushovirinae Bog1183_53

	Microviridae	Gokushovirinae Bog5712_52	Gokushovirinae Bog5712_52
		Gokushovirinae Bog8989_22	Gokushovirinae Bog8989_22
		Gokushovirinae Fen672_31	Gokushovirinae Fen672_31
		Gokushovirinae Fen7875_21	Gokushovirinae Fen7875_21
			Gokushovirinae GAIR4
		Gokushovirinae GNX3R	Gokushovirinae GNX3R
			Microviridae Bog1249_12
			Microviridae Bog5275_51
			Microviridae Bog9017_22
		Microviridae Fen2266_11	Microviridae Fen2266_11
			Microviridae Fen418_41
			Microviridae Fen7918_21
			Microviridae IME-16
			Microviridae phi-CA82
		Parabacteroides phage YZ-2015a	Parabacteroides phage YZ-2015a
			Parabacteroides phage YZ-2015b
	Ackermannviridae		Salmonella phage Det7
	Inoviridae		Pseudomonas phage Pf3
			Ralstonia phage 1 NP-2014
		Ralstonia phage RSS0	
		uncultured phage WW-nAnB	uncultured phage WW-nAnB
	Other		Cellulophaga phage phi14:2
			Cellulophaga phage phi48:2
			Pseudomonas phage O4
			Salicola phage CGphi29
		uncultured crAssphage	uncultured crAssphage
			Yellowstone Lake virophage 5

Material suplementari de l'Article 2

A complete list of all the identified viral species and families in March sewage samblpe after Target Enrichment Sequencing.

Family	Assignment	# Contigs
Adenoviridae	Human mastadenovirus A	1
	Human mastadenovirus F	13
	Simian mastadenovirus F	5
Alphaflexiviridae	Pepino mosaic virus	1
Anelloviridae	Avian gyrovirus 2	2
	Chicken anemia virus	1
Astroviridae	Astrovirus MLB1	2
	Astrovirus MLB2	2
	Astrovirus VA3	3
	Canine astrovirus	1
	Canine astrovirus	2
	Canine astrovirus	1
	Feline astrovirus 2	1
	Feline astrovirus 2	1
	Feline astrovirus 2	1
	Feline astrovirus 2	1
	Feline astrovirus 2	1
	Feline astrovirus 2	1
	HMO Astrovirus A	3
	Mamastrovirus 1	1
	Mamastrovirus 2	1
	Mamastrovirus 9	2
	Porcine astrovirus 4	2
Betaflexiviridae	Garlic common latent virus	6
Caliciviridae	Norovirus GI	3
	Norovirus GII	2
	Norway rat hunnivirus	4
	Sapporo virus	5
	Sapporo virus	1
Circoviridae	Artemia melana sponge associated circular genome	1
	Cyclovirus TN25	2
	Human associated cyclovirus 10	1
	Human circovirus VS6600022	2
	Human circovirus VS6600022	2
	Human fecal virus Jorvi2	2
	Porcine circovirus 2	1
	Porcine stool-associated circular virus 4	2
Coronaviridae	Betacoronavirus 1 (HCoV OC43)	1
	Feline coronavirus	1
	Lucheng Rn rat coronavirus	5
	SARS-CoV-2	8
	Canine coronavirus	1
Dicistroviridae	Aphid lethal paralysis virus	4
	Drosophila C virus	5
	Rhopalosiphum padi virus	1
Genomoviridae	Gopherus associated genomovirus 1	2
	Human associated gemykibivirus 2	1
	Rhinolophus associated gemykibivirus 1	1

Genomoviridae	Sewage derived gemycircularvirus 2	3
	Sewage derived gemycircularvirus 3	2
	Sewage-associated gemycircularvirus-10a	1
Hepeviridae	Hepatitis E virus	4
	Hepatitis E virus rat/R63/DEU/2009	4
Iflaviviridae	Culex Iflavi-like virus 1	2
	Sacbrood virus	3
	Varroa destructor virus 1	4
	Yongsan iflavirus 1	1
Inoviridae	uncultured phage WW-nAnB	3
	Uncultured phage WW-nAnB strain 2	4
	Uncultured phage WW-nAnB strain 3	5
Iridoviridae	Invertebrate iridescent virus 31	2
Leviviridae	Acinetobacter phage AP205	1
Marnaviridae	Sanfarnavirus 2	4
Microviridae	Chimpanzee faeces associated microphage 1	1
	Escherichia virus G4	2
	Escherichia virus ID52	2
	Escherichia virus Talmos	1
	Gokushovirinae Bog1183_53	3
Mimiviridae	Acanthamoeba polyphaga mimivirus	1
Myoviridae	Acinetobacter phage Acj61	4
Nodaviridae	Le Blanc nodavirus (segment RA11)	1
	Orsay virus (segment RNA 1)	1
	Orsay virus (segment RNA 2)	1
	Santeuil nodavirus (segment RNA 2)	1
Paramyxoviridae	Human respirovirus 3	1
Partitiviridae;	Rosellinia necatrix partitivirus 8 (segment RNA 1)	1
Partiviridae	Fusarium poae partitivirus 2 (segment 2)	1
Parvoviridae	Adeno-associated virus - 1	2
	Adeno-associated virus - 2	1
	Adeno-associated virus - 5	1
	Aedes albopictus densovirus	1
	Avian adeno-associated virus strain DA-1	2
	Blattodean ambidensovirus 1	2
	Blattodean ambidensovirus 2	3
	Bovine adeno-associated virus	1
	Carnivore bocaparvovirus 1	4
	Carnivore bocaparvovirus 2	3
	Carnivore protoparvovirus 1	4
	Decapod penstyldensovirus 1	2
	Dipteran ambidensovirus 1	1
	Human bocavirus 3	2
	Mythimna loreyi densovirus	1
	Mythimna loreyi densovirus	2
	Orthopteran ambidensovirus 1	3
	Parus major densovirus	2
	Porcine bocavirus 3	1
	Porcine parvovirus	2
	Primate bocaparvovirus 1	1
	Primate bocaparvovirus 2	2

	Primate protoparvovirus 1	3
	Rodent protoparvovirus 1	2
	Rodent protoparvovirus 2	4
	Ungulate bocaparvovirus 3	2
	Ungulate bocaparvovirus 4	3
Picobirnaviridae	Chicken picobirnavirus (segment RNA 2)	1
	Human picobirnavirus (segment 1)	1
	Human picobirnavirus (segment 2)	1
	Human picobirnavirus (segment 2)	2
	Otarine picobirnavirus (segment 2)	1
	Picobirnavirus dog/KNA/2015	2
	Picobirnavirus green monkey/KNA/2015	3
	Porcine picobirnavirus (segment S)	1
	Porcine picobirnavirus (segment S)	2
Picornaviridae	Aichivirus A	2
	Aichivirus F	2
	Cardiovirus C	4
	Chicken megrivirus	1
	Chicken picornavirus 1	2
	Chicken picornavirus 5	2
	Cosavirus A	4
	Cosavirus D	1
	Enterovirus A	3
	Enterovirus B	5
	Enterovirus C	5
	Parechovirus A	2
	Rabovirus	4
	Rhinovirus A	2
	Rhinovirus B	2
	Salivirus A	1
	Salivirus A	2
	Sicinivirus A	4
	Theilovirus	1
	Theilovirus	3
Polyomaviridae	Human polyomavirus 1	2
	Human polyomavirus 2	1
Reoviridae	Mammalian orthoreovirus (segment L3)	2
Satellites	Sewage-associated circular DNA molecule-1	1
	Sewage-associated circular DNA molecule-3	1
Secoviridae	Broad bean wilt virus 2 (segment RNA 1)	1
	Broad bean wilt virus 2 (segment RNA 1)	2
	Broad bean wilt virus 2 (segment RNA 1)	1
	Broad bean wilt virus 2 (segment RNA 2)	1
	Carrot torradovirus 1 (segment RNA 2)	2
	Squash mosaic virus (segment RNA 1)	4
	Squash mosaic virus (segment RNA 2)	2
Siphoviridae	Streptococcus phage Str-PAP-1	1
Smacoviridae	Human associated porprismacovirus 2	1
Solemoviridae	Papaya lethal yellowing virus	2
Tombusviridae	Grapevine Algerian latent virus	2
	Moroccan pepper virus	1

	Pelargonium leaf curl virus	2
Tymoviridae	Poinsettia mosaic virus	1
	Turnip yellow mosaic virus	3
Unclassified	Gyrovirus 4	2
	Hubei picorna-like virus 15	1
	Hubei picorna-like virus 61	5
	Hubei picorna-like virus 62	4
	Hubei picorna-like virus 63	4
	Hudisavirus sp.	3
	Hudisavirus sp.	1
	Human feces pecovirus	1
	Odonata-associated circular virus-18	1
	Picalivirus A	3
	Picornavirales Tottori-HG1	4
	Wenzhou picorna-like virus 47	3
	Wuhan insect virus 23 (segment Seg 2)	1
Virgaviridae	Bell pepper mottle virus	4
	Cucumber green mottle mosaic virus	3
	Paprika mild mottle virus	1
	Pepper mild mottle virus	1
	Tobacco mild green mosaic virus	5
	Tobacco mosaic virus	2
	Tomato brown rugose fruit virus	1
	Tomato mosaic virus	1
	Turnip vein-clearing virus	1

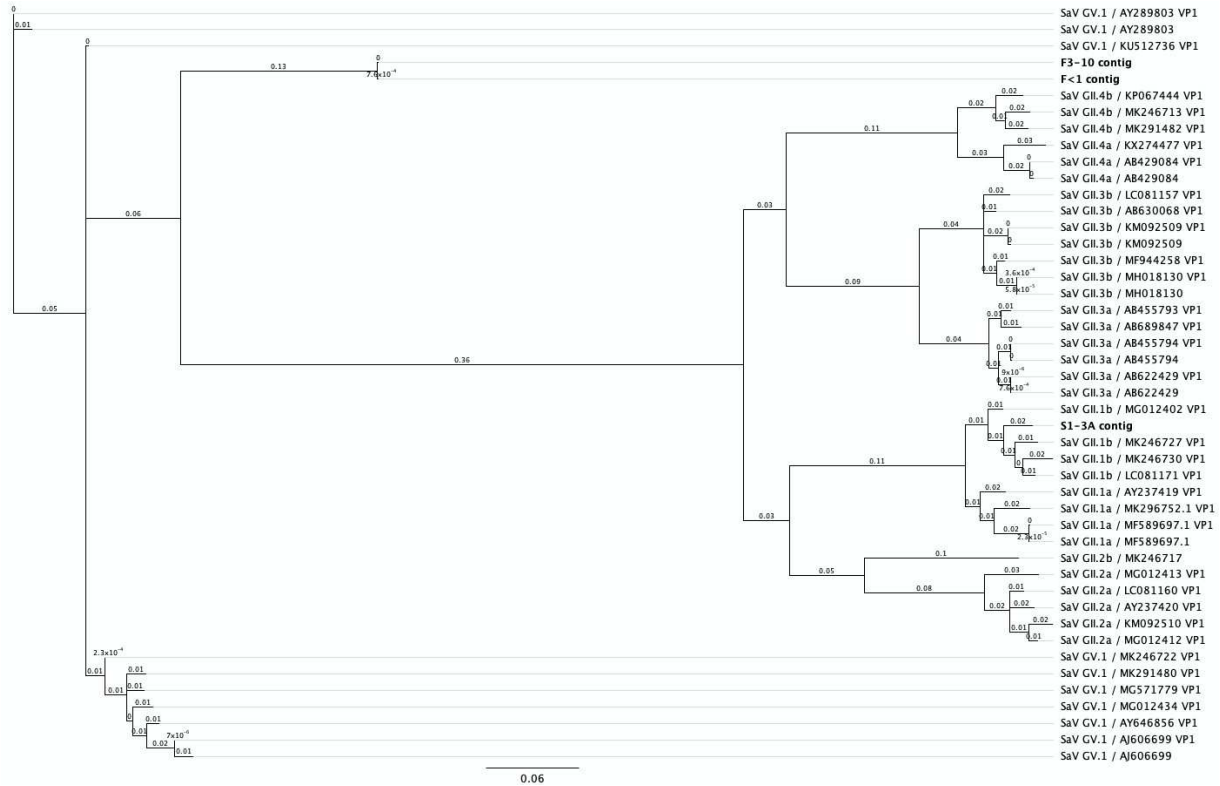
Material suplementari de l'Article 3

Supplementary material 1. Metagenomic sequencing summary statistics obtained with genome detective.

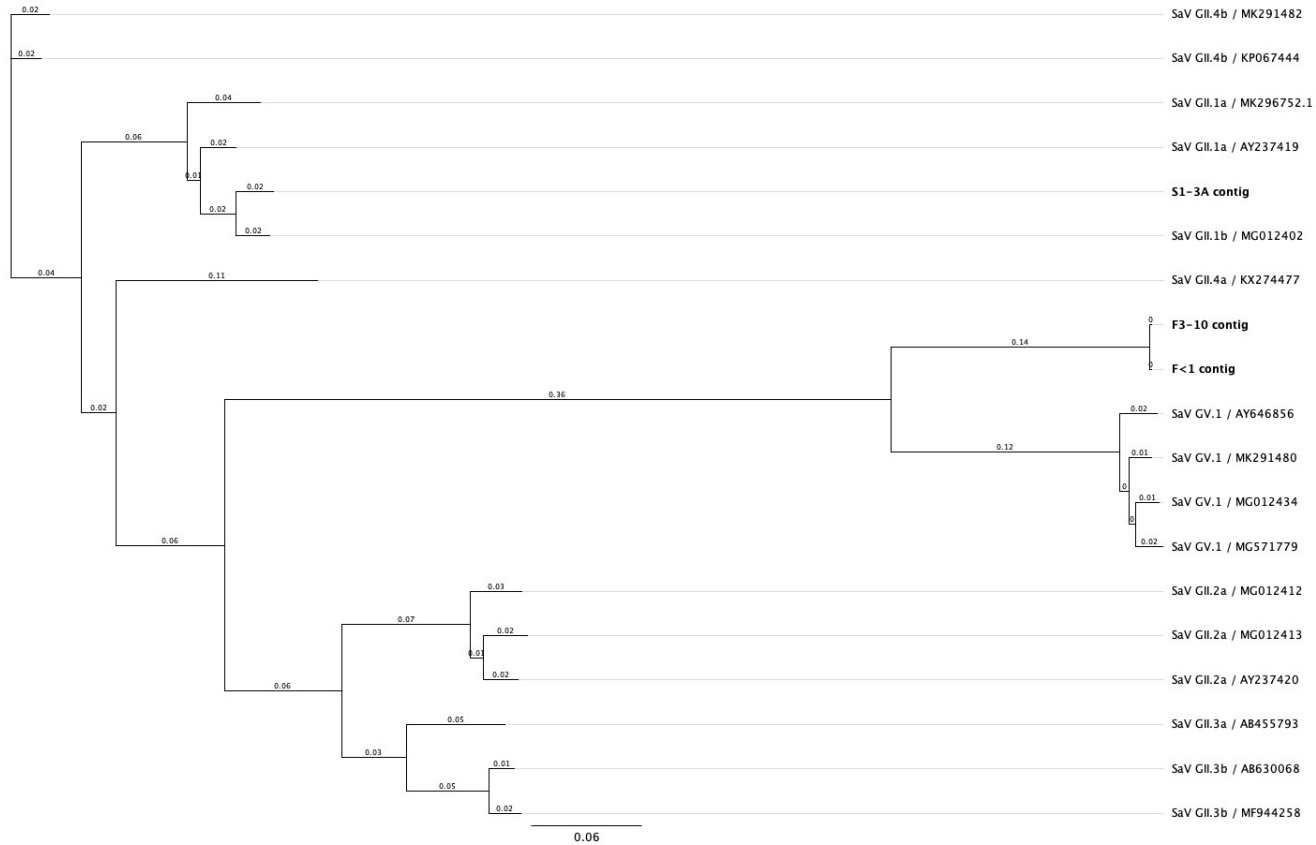
Sample	Initial paired end reads	Removed reads after quality (percentage)	Used reads	Non-viral reads (percentage)	reads used for assembly	Mapped reads into contigs
F<1	6095990	468766 (7%)	5627224	2867732 (50%)	2759492	2364300
F1-3A	5613532	220524 (3%)	5393008	4161306 (77%)	1231702	487708
F3-10	9237546	513822 (5%)	8723724	1049524 (12%)	7674200	7463778
F>10	12265116	655736 (5%)	11609380	4036566 (34%)	7572814	7414041
S<1A	1982130	119564 (6%)	1862566	1049222 (56%)	813344	873829
S<1B	1258278	69804(5%)	1188474	850264 (71%)	338210	199647
S1-3A	1549150	60662 (3%)	1488488	819480 (55%)	669008	514749
S1-3B	1037992	41840 (4%)	996152	725326 (72%)	270826	85046
S1-3C	881740	37730 (4%)	844010	379784 (44%)	464226	17425
S3-10	2419026	70822 (2%)	2348204	1387786 (59%)	960418	578919

Supplementary material 3. Phylogenetic trees of sapovirus contigs. Trees were constructed by using tree view software within Geneious (software version 11.0 (https://www.geneious.com) maximum-likelihood (Tamura-Nei model, Bootstrap of 1000 replicates) analysis. Phylogenetic trees with the full sapovirus types (a), full length VP1 (b), complete genomes are presented. The labels on the branch indicate the number of substitutions per site.

a.

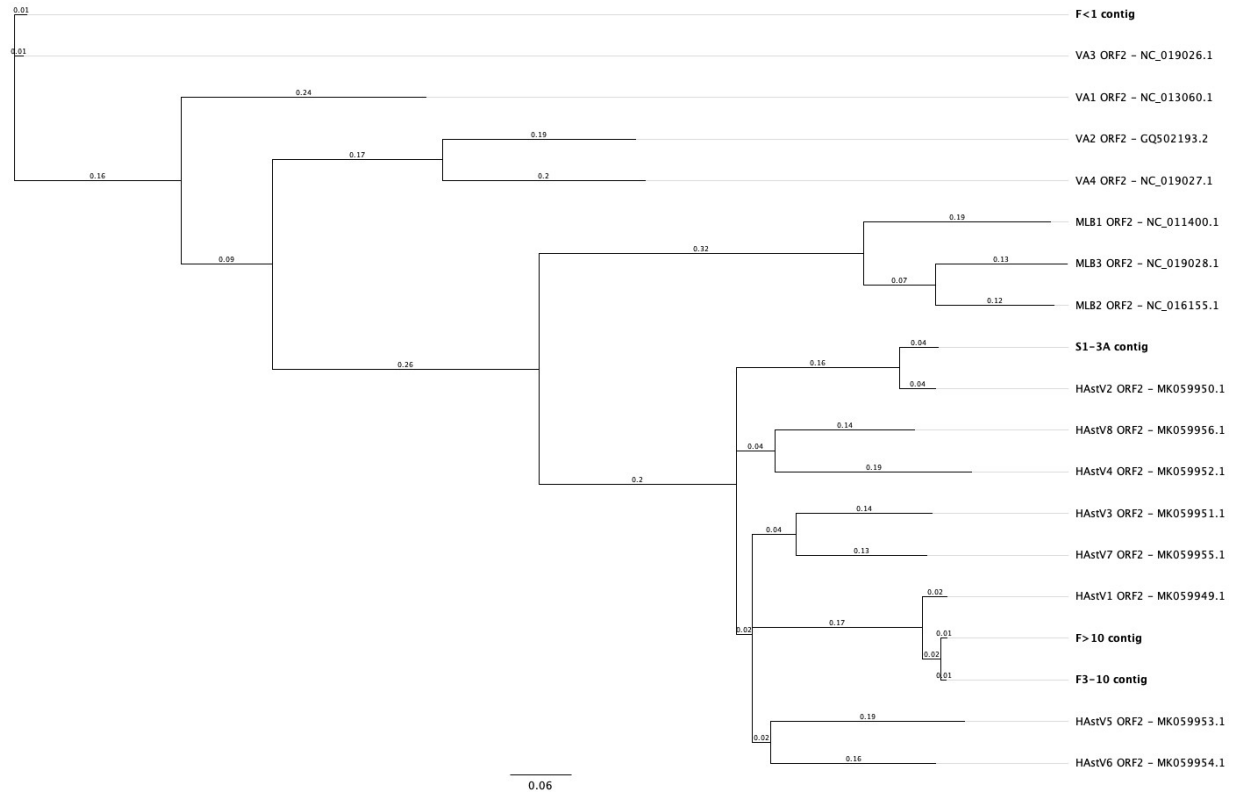


b.

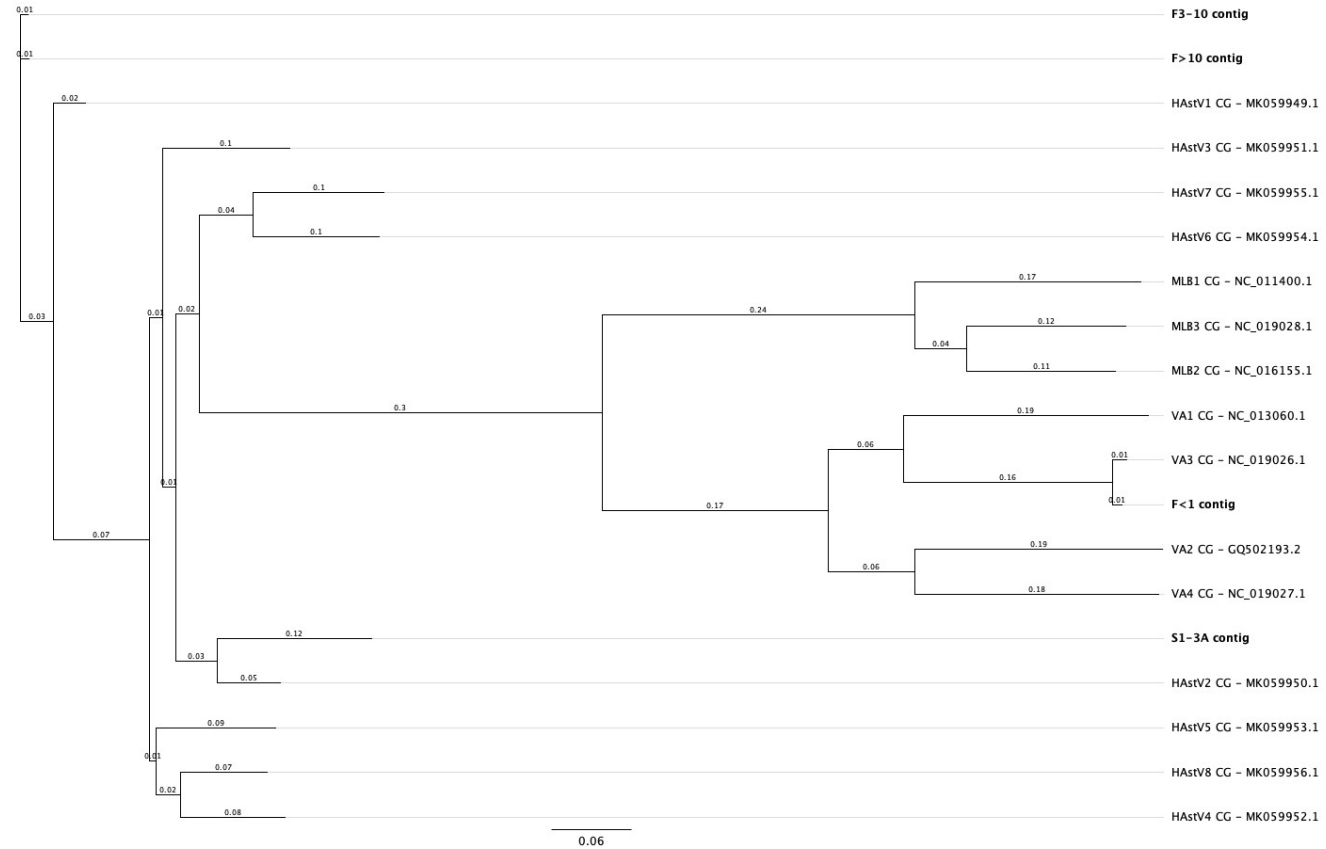

SaV GII.4b / MK291482
SaV GII.4b / KP067444
SaV GII.1a / MK296752.1
SaV GII.1a / AY237419
S1-3A contig
SaV GII.1b / MG012402
SaV GII.4a / KX274477
F3-10 contig
F<1 contig
SaV GV.1 / AY646856
SaV GV.1 / MK291480
SaV GV.1 / MG012434
SaV GV.1 / MG571779
SaV GII.2a / MG012412
SaV GII.2a / MG012413
SaV GII.2a / AY237420
SaV GII.3a / AB455793
SaV GII.3b / AB630068
SaV GII.3b / MF944258
0.06

Supplementary material 4. Phylogenetic trees of mamastrovirus contigs. Trees were constructed by using tree view software within Geneious (software version 11.0 (https://www.geneious.com) maximum-likelihood (Tamura-Nei model, Bootstrap of 1000 replicates) analysis. Phylogenetic trees with the full mamastrovirus types (a), full length ORF2 (b), complete genomes are presented. The labels on the branch indicate the number of substitutions per site.

a.



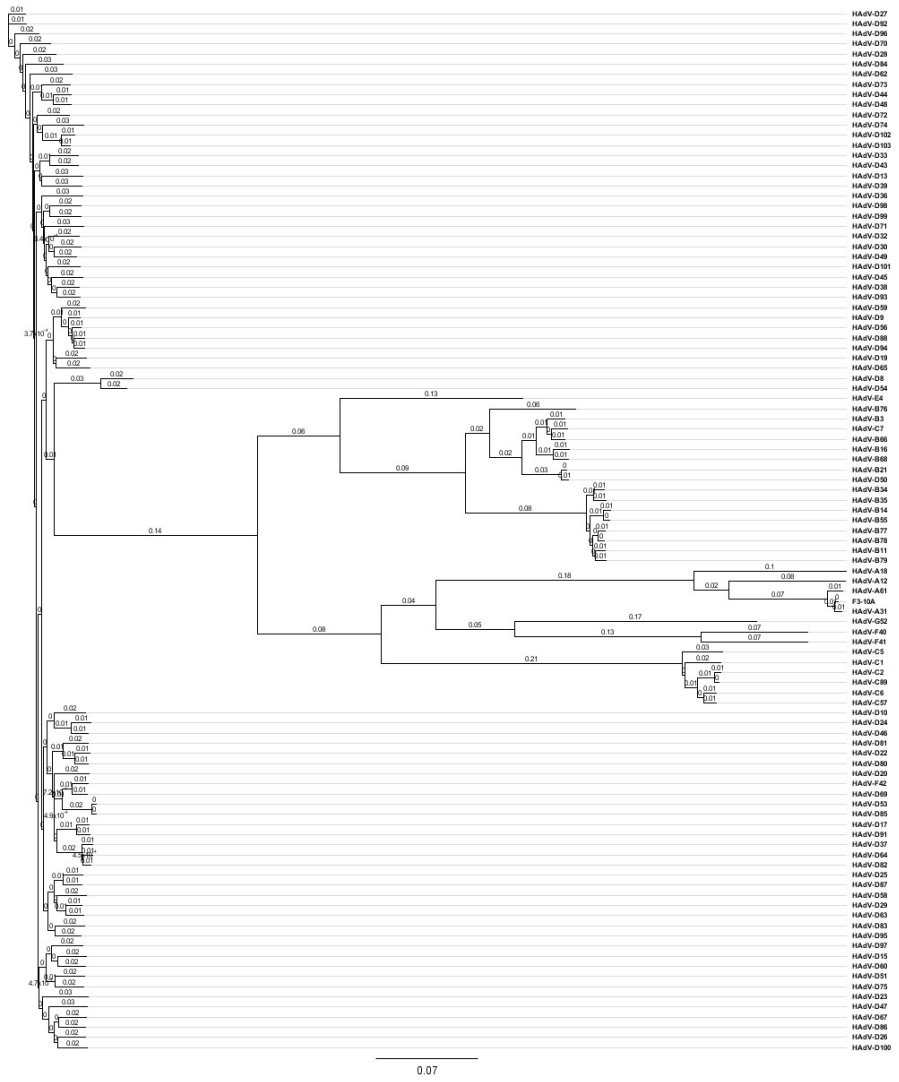
b.


F3-10 contig
F>10 contig
HAstV1 CG - MK059949.1
HAstV3 CG - MK059951.1
HAstV7 CG - MK059955.1
HAstV6 CG - MK059954.1
MLB1 CG - NC_011400.1
MLB3 CG - NC_019028.1
MLB2 CG - NC_016155.1
VA1 CG - NC_013060.1
VA3 CG - NC_019026.1
F<1 contig
VA2 CG - GQ502193.2
VA4 CG - NC_019027.1
S1-3A contig
HAstV2 CG - MK059950.1
HAstV5 CG - MK059953.1
HAstV8 CG - MK059956.1
HAstV4 CG - MK059952.1
0.06

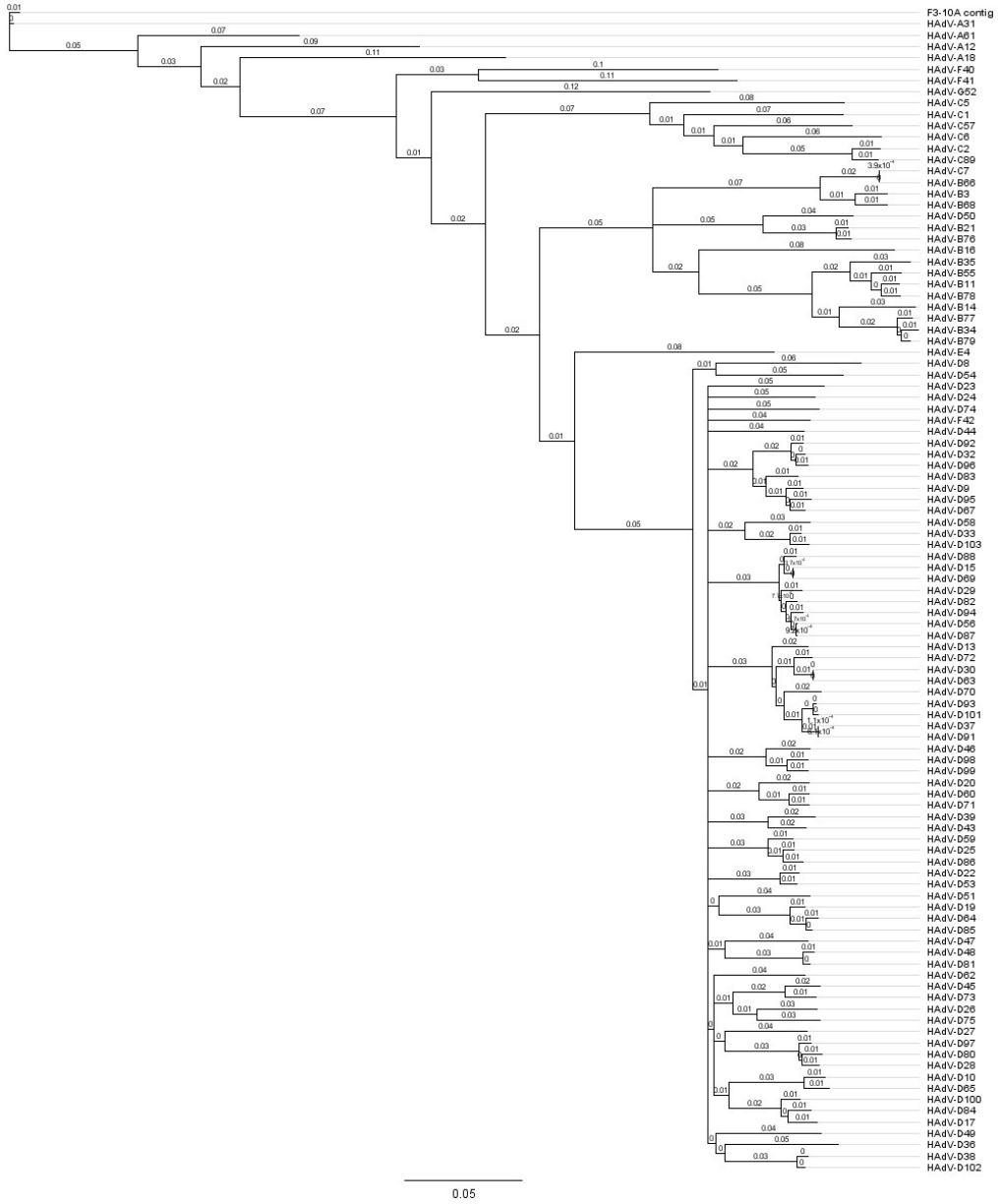
Supplementary material 5

Accession numbers of reference adenovirus types were retrieved from the Human Adenovirus Working Group (http://hadvwg.gmu.edu/). Full adenovirus genomes or the annotated regions corresponding to the hexon, penton and fiber were extracted and aligned with MUSCLE. Phylogenetic tree were constructed by using tree view software within Geneious (software version 11.0 (https://www.geneious.com) maximum-likelihood (Tamura-Nei model, Bootstrap of 1000 replicates) analysis. Phylogenetic trees with the full adenovirus reference types (a), using the hexon (b), penton (c) and fiber (d) are presented. The labels on the branch indicate the number of substitutions per site.

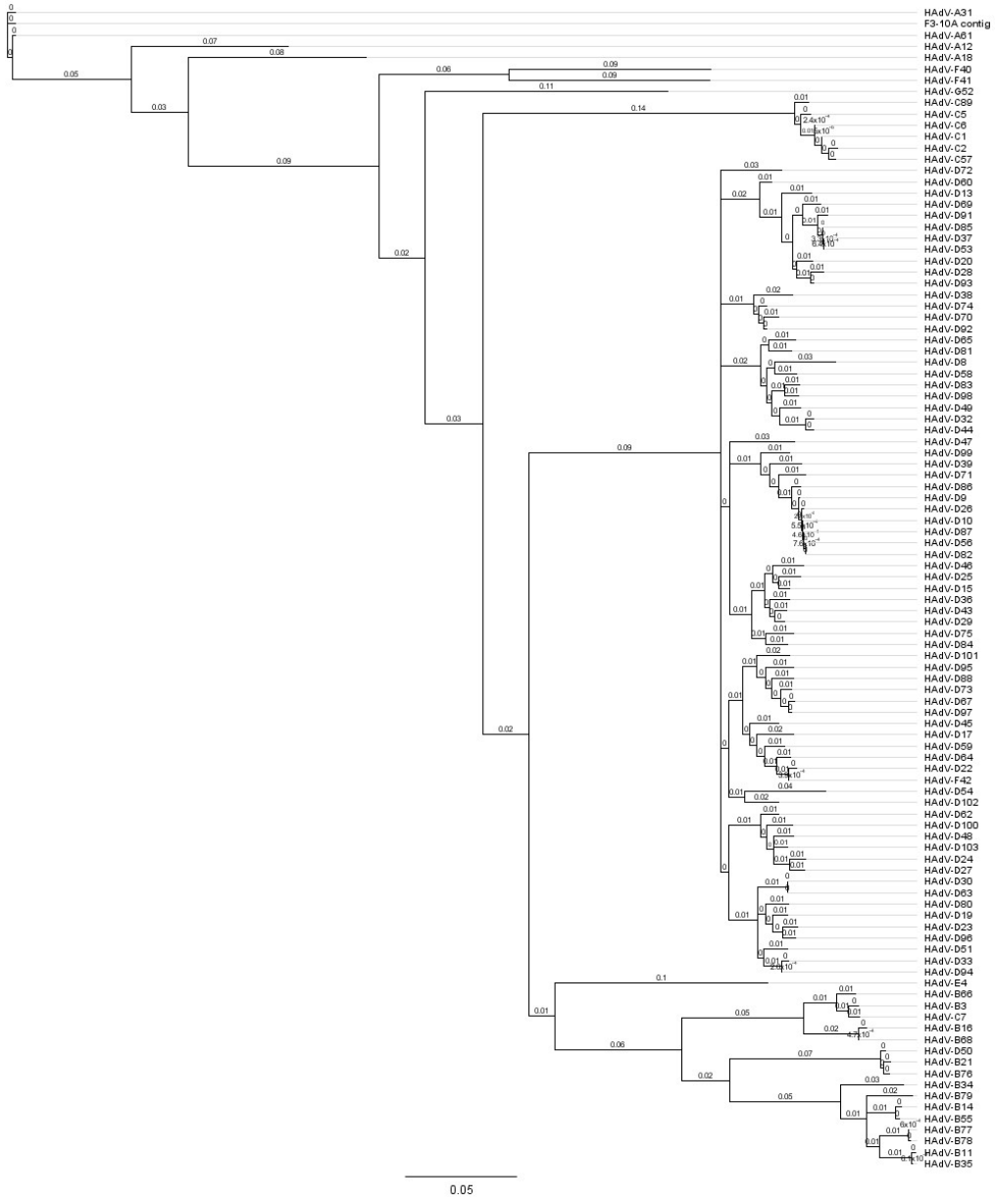
a)



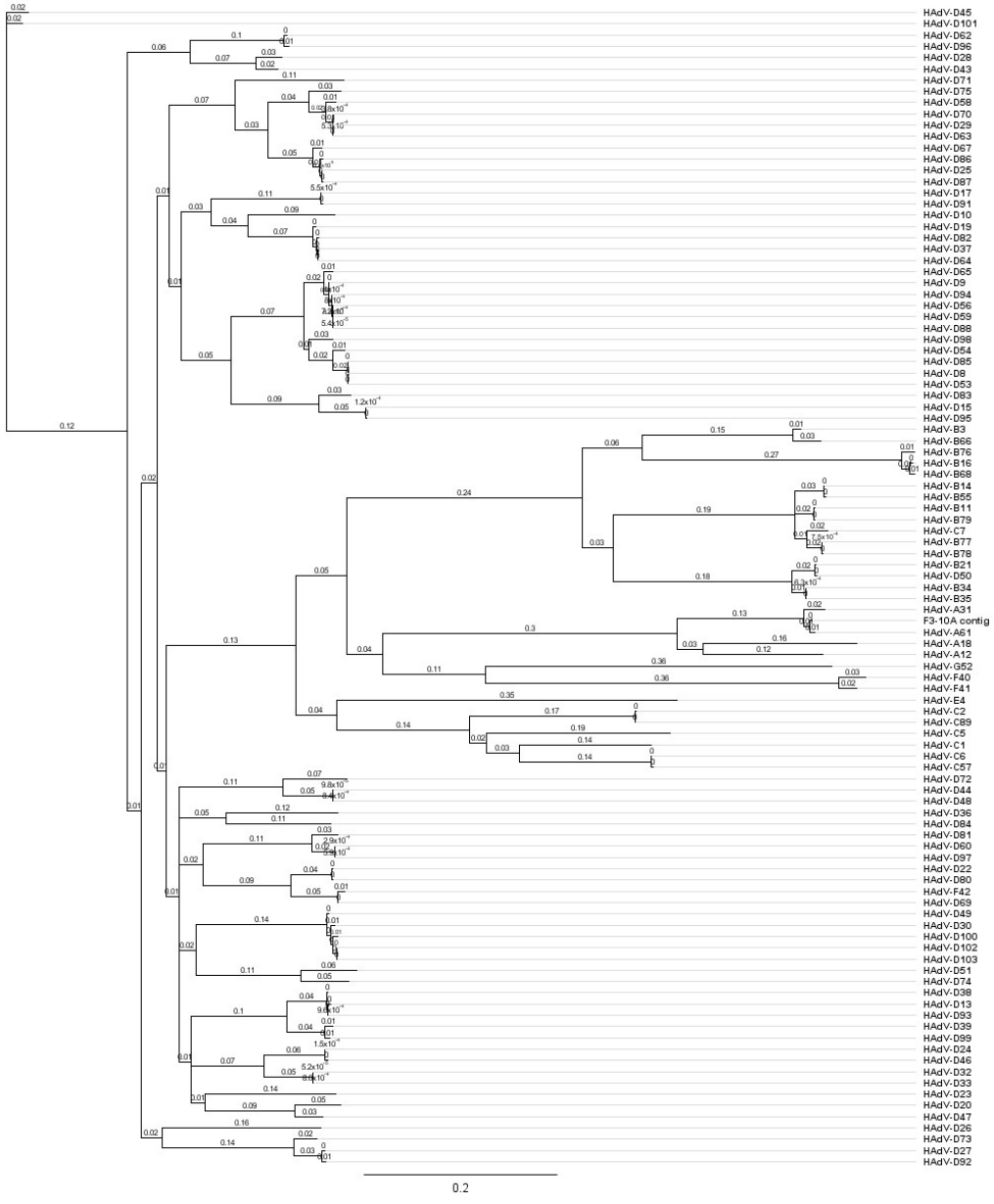
b)


F3-10A contig
HAdV-A31
HAdV-A61
HAdV-A12
HAdV-A18
HAdV-F40
HAdV-F41
HAdV-G52
HAdV-C5
HAdV-C1
HAdV-C57
HAdV-C6
HAdV-C2
HAdV-C89
HAdV-C7
HAdV-B66
HAdV-B3
HAdV-B68
HAdV-B50
HAdV-B21
HAdV-B76
HAdV-B16
HAdV-B35
HAdV-B55
HAdV-B11
HAdV-B78
HAdV-B14
HAdV-B77
HAdV-B34
HAdV-B79
HAdV-E4
HAdV-D8
HAdV-D54
HAdV-D23
HAdV-D24
HAdV-D74
HAdV-F42
HAdV-D44
HAdV-D92
HAdV-D32
HAdV-D96
HAdV-D83
HAdV-D9
HAdV-D95
HAdV-D67
HAdV-D58
HAdV-D33
HAdV-D103
HAdV-D88
HAdV-D15
HAdV-D69
HAdV-D29
HAdV-D82
HAdV-D94
HAdV-D56
HAdV-D87
HAdV-D13
HAdV-D72
HAdV-D30
HAdV-D63
HAdV-D70
HAdV-D93
HAdV-D101
HAdV-D37
HAdV-D91
HAdV-D46
HAdV-D98
HAdV-D99
HAdV-D20
HAdV-D60
HAdV-D71
HAdV-D39
HAdV-D43
HAdV-D59
HAdV-D25
HAdV-D86
HAdV-D22
HAdV-D53
HAdV-D51
HAdV-D19
HAdV-D64
HAdV-D85
HAdV-D47
HAdV-D48
HAdV-D81
HAdV-D62
HAdV-D45
HAdV-D73
HAdV-D26
HAdV-D75
HAdV-D27
HAdV-D97
HAdV-D80
HAdV-D28
HAdV-D10
HAdV-D65
HAdV-D100
HAdV-D84
HAdV-D17
HAdV-D49
HAdV-D36
HAdV-D38
HAdV-D102
0.05

c)


HAdV-A31
F3-10A contig
HAdV-A61
HAdV-A12
HAdV-A18
HAdV-F40
HAdV-F41
HAdV-G52
HAdV-C89
HAdV-C5
HAdV-C6
HAdV-C1
HAdV-C2
HAdV-C57
HAdV-D72
HAdV-D60
HAdV-D13
HAdV-D69
HAdV-D91
HAdV-D85
HAdV-D37
HAdV-D53
HAdV-D20
HAdV-D28
HAdV-D93
HAdV-D38
HAdV-D74
HAdV-D70
HAdV-D92
HAdV-D65
HAdV-D81
HAdV-D8
HAdV-D58
HAdV-D83
HAdV-D98
HAdV-D49
HAdV-D32
HAdV-D44
HAdV-D47
HAdV-D99
HAdV-D39
HAdV-D71
HAdV-D86
HAdV-D9
HAdV-D26
HAdV-D10
HAdV-D87
HAdV-D56
HAdV-D82
HAdV-D46
HAdV-D25
HAdV-D15
HAdV-D36
HAdV-D43
HAdV-D29
HAdV-D75
HAdV-D84
HAdV-D101
HAdV-D95
HAdV-D88
HAdV-D73
HAdV-D67
HAdV-D97
HAdV-D45
HAdV-D17
HAdV-D59
HAdV-D64
HAdV-D22
HAdV-F42
HAdV-D54
HAdV-D102
HAdV-D62
HAdV-D100
HAdV-D48
HAdV-D103
HAdV-D24
HAdV-D27
HAdV-D30
HAdV-D83
HAdV-D80
HAdV-D19
HAdV-D23
HAdV-D66
HAdV-D51
HAdV-D33
HAdV-D94
HAdV-E4
HAdV-B66
HAdV-B3
HAdV-C7
HAdV-B16
HAdV-B68
HAdV-D50
HAdV-B21
HAdV-B76
HAdV-B34
HAdV-B79
HAdV-B14
HAdV-B55
HAdV-B77
HAdV-B78
HAdV-B11
HAdV-B35
0.05

d)


HAdV-D45
HAdV-D101
HAdV-D62
HAdV-D96
HAdV-D28
HAdV-D43
HAdV-D71
HAdV-D75
HAdV-D58
HAdV-D70
HAdV-D29
HAdV-D63
HAdV-D67
HAdV-D86
HAdV-D25
HAdV-D87
HAdV-D17
HAdV-D91
HAdV-D10
HAdV-D19
HAdV-D82
HAdV-D37
HAdV-D64
HAdV-D65
HAdV-D9
HAdV-D94
HAdV-D56
HAdV-D59
HAdV-D88
HAdV-D98
HAdV-D54
HAdV-D85
HAdV-D8
HAdV-D53
HAdV-D83
HAdV-D15
HAdV-D95
HAdV-B3
HAdV-B66
HAdV-B76
HAdV-B16
HAdV-B68
HAdV-B14
HAdV-B55
HAdV-B11
HAdV-B79
HAdV-C7
HAdV-B77
HAdV-B78
HAdV-B21
HAdV-D50
HAdV-B34
HAdV-B35
HAdV-A31
F3-10A contig
HAdV-A61
HAdV-A18
HAdV-A12
HAdV-G52
HAdV-F40
HAdV-F41
HAdV-E4
HAdV-C2
HAdV-C89
HAdV-C5
HAdV-C1
HAdV-C6
HAdV-C57
HAdV-D72
HAdV-D44
HAdV-D48
HAdV-D36
HAdV-D84
HAdV-D81
HAdV-D60
HAdV-D97
HAdV-D22
HAdV-D80
HAdV-F42
HAdV-D69
HAdV-D49
HAdV-D30
HAdV-D100
HAdV-D102
HAdV-D103
HAdV-D51
HAdV-D74
HAdV-D38
HAdV-D13
HAdV-D93
HAdV-D39
HAdV-D99
HAdV-D24
HAdV-D46
HAdV-D32
HAdV-D33
HAdV-D23
HAdV-D20
HAdV-D47
HAdV-D26
HAdV-D73
HAdV-D27
HAdV-D92
0.2

Supplementary material 6. Characterization of enterovirus (a) and rhinovirus (b) contigs detected in the gastroenteritis pools using the enterovirus typing tool by RIVM (v.0.1).

(a)

Sample	Contig length	BLAST score	Genotype	VP1 Serotype, Sub-Genogroup	Classification
F1-3A	7400	80.481865	CV-A10	CV-A10	Picornaviridae Enterovirus A
	7193	83.12013	E-6	E-6	Picornaviridae Enterovirus B
	7011	95.69187	CV-A6	CV-A6	Picornaviridae Enterovirus A
	2696	83.11688	CV-B5	CV-B5	Picornaviridae Enterovirus B
	471	83.6518	-	-	Picornaviridae Enterovirus B
	1248	86.3124	-	-	Picornaviridae Enterovirus B
F3-10	680	91.87592	-	-	Picornaviridae Enterovirus B
	817	83.190186	-	-	Picornaviridae Enterovirus B
	722	84.02778	E-6	E-6	Picornaviridae Enterovirus B
	965	81.12033	E-6	E-6	Picornaviridae Enterovirus B
	2835	85.472496	-	-	Picornaviridae Enterovirus B
S<1A	7213	83.90564	CV-B5	CV-B5	Picornaviridae Enterovirus B
S<1B	2737	93.78882	E-13	E-13	Picornaviridae Enterovirus B
	2068	92.30769			
	1367	91.93899			
S1-3A	6865	80.99679	E-9	E-9	Picornaviridae Enterovirus B
	7085	85.63747	E-3	E-3	Picornaviridae Enterovirus B
	6650	89.35338	CV-A6	CV-A6	Picornaviridae Enterovirus A
	6902	80.17743	CV-A10	CV-A10	Picornaviridae Enterovirus A
	2608	93.66603	E-13	E-13	Picornaviridae Enterovirus B
	365	90.95891			Picornaviridae Enterovirus B

	1363				Picornaviridae Enterovirus B
S1-3B	7158	95.96199	CV-A16	CV-A16 , B1a	Picornaviridae Enterovirus A
	6284	78.27817	CV-B1	CV-B1	Picornaviridae Enterovirus B
	436	87.006966	CV-A4	CV-A4	Picornaviridae Enterovirus A
	458	86.18421	CV-A4		Picornaviridae Enterovirus A
	1105	85.791855	CV-A4	CV-A4	Picornaviridae Enterovirus A
	2122	83.577156	CV-A8		Picornaviridae Enterovirus A
	564	83.27402	EV-A71		
S1-3C	1224	84.72222	E-9	E-9	Picornaviridae Enterovirus B
	635	86.456696			Picornaviridae Enterovirus B
	614	83.68679			Picornaviridae Enterovirus B
	794	86.901764			Picornaviridae Enterovirus B
S1-3B (unassigned)	1454	82.09366	E-16	E-16	Picornaviridae Enterovirus B
	741	80.13514	E-16	E-16	Picornaviridae Enterovirus B
	358	92.20056	NTP	NTP	Picornaviridae Enterovirus B

(b)

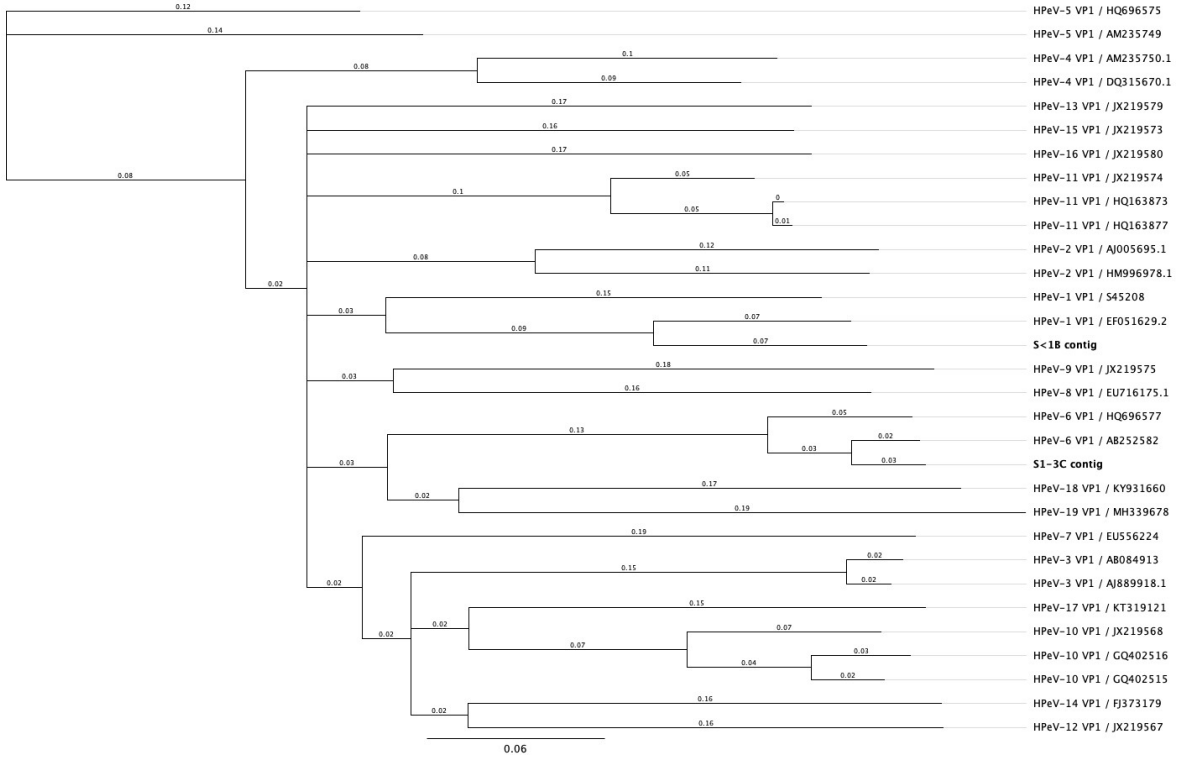
Sample	Contig length	BLAST score	Genotype	VP1 Serotype, Sub-Genogroup	Classification
F3	597	90.60403	HRV-A60		Picornaviridae Rhinovirus A
	483	90.47619	HRV-A60		Picornaviridae Rhinovirus A
	592	88.983055	NTP		Picornaviridae Rhinovirus A
S1	459	94.77124	HRV-B52		Picornaviridae Rhinovirus B
	574	91.78322	HRV-B52		Picornaviridae Rhinovirus B
	1155	92.03463	HRV-B52		Picornaviridae Rhinovirus B
	766	92.81985	HRV-B52		Picornaviridae Rhinovirus B
	626	93.29073	NTP		Picornaviridae Rhinovirus B
	413	91.26214	NTP		Picornaviridae Rhinovirus B
	451	93.11111	NTP		Picornaviridae Rhinovirus B
	409	93.887535	NTP		Picornaviridae Rhinovirus B
S2	6499	86.74773	HRV-A78		Picornaviridae Rhinovirus A

Supplementary material 7. Characterization of non-enterovirus P*icornaviridae* contigs detected on the fecal pools analyzed, including parechovirus and cardiovirus.

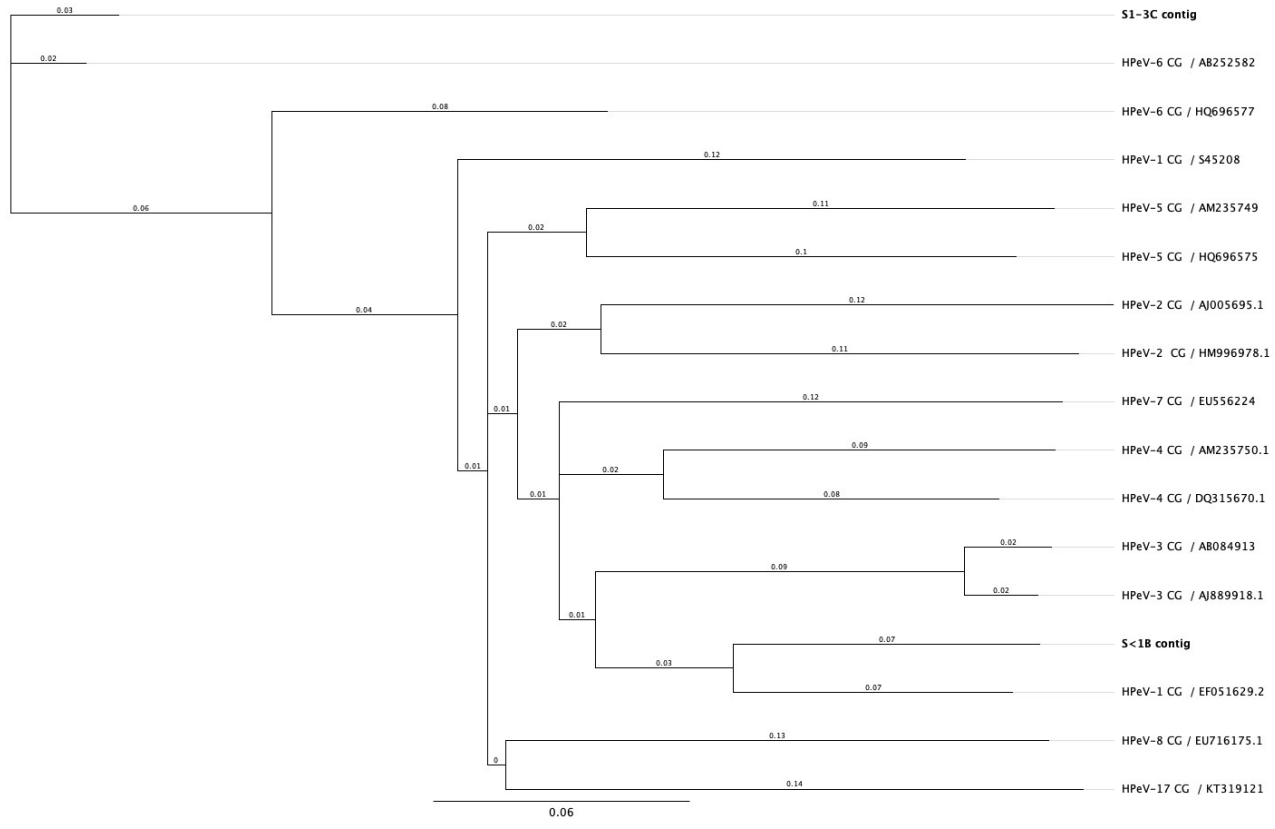
Pool	Contig length	Classification	VP1 Genotype	% VP1 Identity	% VP1 Coverage	% Genome Identity	% Genome Coverage
S<1B	7248	Parechovirus A	HPeV-1	87.8	100	87.8	98.5
S1-3B	796	Parechovirus A	-	-	-	95.6	10.8
	471	Parechovirus A	-	92.0	7.1	95.8	6.4
	1695	Parechovirus A	-	-	-	95.4	23.1
	343	Parechovirus A	-	-	-	95.0	4.7
	6949	Parechovirus A	HPeV-6	95.4	100	95.9	94.6
S3-10	7942	Cardiovirus B	SAFV-2	94.7	100	90.7	98.7

Supplementary material 8. Phylogenetic trees of retrieved parechovirus contigs. Trees were constructed by using tree view software within Geneious (software version 11.0 (https://www.geneious.com) maximum-likelihood (Tamura-Nei model, Bootstrap of 1000 replicates) analysis. Phylogenetic trees were constructed using the (a), full length of the VP1 genomic sequence (b), full human parechovirus complete genomes. The labels on the branch indicate the number of substitutions per site.

a.



b.


S1-3C contig
HPeV-6 CG / AB252582
HPeV-6 CG / HQ696577
HPeV-1 CG / S45208
HPeV-5 CG / AM235749
HPeV-5 CG / HQ696575
HPeV-2 CG / AJ005695.1
HPeV-2 CG / HM996978.1
HPeV-7 CG / EU556224
HPeV-4 CG / AM235750.1
HPeV-4 CG / DQ315670.1
HPeV-3 CG / AB084913
HPeV-3 CG / AJ889918.1
S<1B contig
HPeV-1 CG / EF051629.2
HPeV-8 CG / EU716175.1
HPeV-17 CG / KT319121
0.06

Material suplementari de l'Article 4

Table S1. Accession numbers of sequences used for phylogenetic analysis, including the sequences obtained in this study (*) and those downloaded from GenBank: a) sequences used for phylogenetic analysis of Viral Nervous Necrosis Virus obtained from a gilthead seabream gills from Algarve, b) sequences used for phylogenetic analysis of Infectious Pancreatic Necrotic Virus amplified from an Atlantic horse mackerel skin sample from Peniche, c) sequences used for phylogenetic analysis of Hepatitis B virus obtained from Atlantic horse mackerel gills from Algarve, d) sequences used for phylogenetic analysis of Rhinovirus A amplified from gilthead seabream skin from Peniche; e) sequences used for phylogenetic analysis of Norovirus GII obtained from gilthead seabream skin from Algarve.

a)

Country	Origin	Access number
Japan	Striped jack	AB056572
Japan	Striped jack	D30814
Japan	European seabass	AF175519
Italy	European Seabass	JN189936
Spain	Senegalese sole	JN189919
Italy	Senegalese sole	JN189932
Croatia	European seabass	JN189915
Portugal	Senegalese sole	FJ803920
Portugal	Seabream	JN189916
China	Threadfin porgy	KP994911
Cyprus	Seabream	JN189920
Cyprus	Seabream	KY354697
Italy	Seabream	KY354703
Japan	Tiger puffer	D38637
Japan	Tiger puffer	EU236149
Japan	Japanese flounder	AB045980
Australia	Redspotted grouper	GQ402013
Israel	European Seabass	AY284969
Israel	White Grouper	AY284963
China	Hybrid grouper	MK107836
China	Tiger Grouper	KX027363
China	Tiger grouper	HQ859940
Malaysia	Tiger grouper	MN493039
Japan	Sevenband grouper	AB373029
Spain	Seabass	FJ829452
India	Asian seabass	JF412273
India	Asian seabass	GU953669
India	Asian seabass	JF412263
India	Asian seabass	GU826692
India	Asian seabass	GU826693
India	Asian seabass	FR669249
India	Asian seabass	GU563883
China	Orange-spotted grouper	MG874758
Taiwan	Giant grouper	MF565445
Australia	Redspotted grouper	GQ402011
Taiwan	Giant grouper	KM588181
Vietnam	Mouse grouper	KU705815
South Korea	Sevenband grouper	KM095959
Malaysia	Golden pompano	GQ904199
China	Sevenband grouper	KP455642
Malaysia	Asian seabass	HQ859948
Malaysia	Tiger grouper	HQ859938
Malaysia	Asian seabass	HQ859935
Malaysia	Asian seabass	GQ120525

China	Pearl gentian	MG637439
Malaysia	Asian seabass	EU380202
Malaysia	Golden pompano	HQ859932
Taiwan	Dragon grouper	AF245004
Portugal	Seabream	LC581281*

b)

Country	Origin	Access number
Norway	Atlantic salmon	KX355260
Norway	Atlantic salmon	AY379737
Norway	Atlantic salmon	KX355241
Norway	Atlantic salmon	KX355225
Norway	Atlantic salmon	KX355244
Norway	Atlantic salmon	MH562010
Norway	Atlantic salmon	MH562014
Norway	Atlantic salmon	MH561943
Norway	Atlantic salmon	MH561975
Norway	Atlantic salmon	AY374435
Chile	Atlantic salmon	KX523805
Chile	Atlantic salmon	KX523802
Norway	Atlantic salmon	MH561992
Norway	Atlantic salmon	MH562002
Norway	Atlantic salmon	MH562000
Norway	Atlantic salmon	MH562011
Norway	Atlantic salmon	MH561997
Norway	Atlantic salmon	MH561948
China	Japanese flounder/ Bastard halibut	EU161285
South Korea	Japanese flounder	AY283784
Japan	Yellowtail	AB006783
Japan	Japanese amberjack/ yellowtail	AY283785
Japan	Yellowtail	AY283781
Japan	Japanese flounder	AY283783
Japan	Yellowtail	AB011440
Japan	Yellowtail	AB281673
Japan	Yellowtail	AY283782
Korea	Rockfish	AY064396
Portugal	Atlantic horse mackerel	LC581280*

c)

Country	Origin	Access number
USA	Bluegill	KX058433
USA	Bluegill	NC_030445
Australia	Eastern sea garfish	MH716822
China	Greater horseshoe bat	KY962692
China	Greater horseshoe bat	KY962687
China	Woolly horseshoe bat	KY962696
China	Least horseshoe bat	KY962705
China	Pomona roundleaf bat	NC_038503
China	Horseshoe bat	KX513949
Mexico	Human	AB625343
Russia	Human	KY660218
Poland	Human	GQ477473
Nigeria	Human	FN547200

France	Human	GQ486240
Nigeria	Human	FN547168
Angola	Human	KF849736
Taiwan	Human	EU835940
China	Human	MK171192
Netherlands	Human	MH521359
China	Human	KF917508
China	Human	KR013765
Taiwan	Human	KC793068
Indonesia	Human	AB644297
China	Human	KY670780
China	Human	KM455832
China	Human	KM875412
China	Human	KM875413
Taiwan	Human	KC792775
China	Human	MK580119
Japan	Human	AB029956
China	Human	MK170959
China	Human	KY360845
China	Human	JX948756
China	Human	JX948755
China	Human	GQ402152
China	Human	KY470927
China	Human	KX777514
China	Human	FJ622431
China	Human	KX777461
South Korea	Human	KX264903
China	Human	KY361361
China	Human	KY470990
China	Human	MG715397
China	Human	MG826330
China	Human	MG715386
China	Human	MK171235
China	Human	MK170856
China	Human	MK170936
China	Human	MK170895
China	Human	MK171241
China	Human	MK170941
Bangladesh	Human	JQ514420
China	Human	MK170921
India	Human	JN107877
China	Human	KY470916
China	Human	KY470914
Portugal	Atlantic horse mackerel	LC581279*

d)

Country	**Origin**	**Access number**
USA	Human	FJ445114
El Salvador	Human	JX129407
Peru	Human	JX129435
USA	Human	FJ445115
Switzerland	Human	JF285315
Switzerland	Human	JF285314
South Korea	Human	KC414928
Singapore	Human	MH648009

Singapore	Human	MH648008
USA	Human	MN749149
USA	Human	MN749151
Austria	Human	X02316
USA	Human	JN541268
Australia	Human	KF543937
USA	Human	MN306051
Australia	Human	KF543911
Australia	Human	KF543912
USA	Human	DQ473494
Uganda	Human	MH685691
Uganda	Human	MH685686
USA	Human	FJ445139
USA	Human	FJ445138
USA	Human	FJ445173
Australia	Human	KF543933
Singapore	Human	MH648059
Singapore	Human	MH648037
El Salvador	Human	JX129408
USA	Human	KY369875
USA	Human	FJ445163
South Korea	Human	JX177625
Singapore	Human	MH648048
USA	Human	MK167031
USA	Human	JN798579
USA	Human	JX074051
USA	Human	FJ445129
USA	Human	JN815252
USA	Human	FJ445125
USA	Human	DQ473499
USA	Human	FJ445145
Singapore	Human	MH648023
USA	Human	MN306057
USA	Human	MN306047
USA	Human	FJ445123
USA	Human	JQ747747
Australia	Human	KF543922
Australia	Human	KF543929
Peru	Human	JX129450
USA	Human	JN837693
USA	Human	FJ445148
USA	Human	JN112340
USA	Human	JN621246
USA	Human	KY369890
Australia	Human	KF543930
Singapore	Human	MH648032
Singapore	Human	MH648046
USA	Human	JN837692
USA	Human	FJ445175
USA	Human	JN798582
USA	Human	FJ445180
USA	Human	DQ473495
USA	Human	JQ994496
USA	Human	JN541272
USA	Human	FJ445141

Singapore	Human	MH648086
Australia	Human	KF543896
Australia	Human	KF543895
India	Human	KM109982
USA	Human	JN614995
USA	Human	FJ445135
Australia	Human	KF543909
Columbia	Human	JX129423
South Korea	Human	JX177621
Switzerland	Human	EU840727
USA	Human	FJ445140
USA	Human	KY369886
Australia	Human	GQ323774
Australia	Human	EF186077
USA	Human	JQ994495
USA	Human	JN798570
Kenya	Human	MK989744
USA	Human	MN306034
USA	Human	K02121
USA	Human	L05355
Portugal	Seabream	LC581278*

e)

Country	Origin	Access number
USA	Human	JN797508
Hong Kong	Human	KT346358
Hong Kong	Human	KT589391
French Guiana	Human	KC597139
South Korea	Human	JX439793
South Korea	Human	JX439794
South Korea	Human	KU561250
Netherlands	Human	JN176920
USA	Human	MK754442
USA	Human	MK629457
USA	Human	MK752942
Japan	Human	LC153121
Russia	Human	MG892946
Russia	Human	MG892955
USA	Human	MK756037
United Kingdom	Human	KY887601
United Kingdom	Human	NC_039477
United Kingdom	Human	KY887600
United Kingdom	Human	KY887603
United Kingdom	Human	KY887602
United Kingdom	Human	KY817505
USA	Human	MK762570
USA	Human	MK762569
Japan	Human	LC331997
Japan	Human	LC325217
USA	Human	MT032004
Portugal	Human	MH930953
Japan	Human	LC175468
Japan	Human	LC153122
USA	Human	MK073891
USA	Human	MH260484

USA	Human	MK775032
USA	Human	KY947549
USA	Human	MK753033
Japan	Human	LC215415
Japan	Human	LC213892
Russia	Human	MG892929
USA	Human	KY947550
United Kingdom	Human	MH218685
United Kingdom	Human	KY887599
Portugal	Seabream	LC581277*

Table S2. Accession numbers of sequences used for phylogenetic analysis (family analysis), including the sequences obtained in this study (*) and those downloaded from GenBank: a) sequences used for phylogenetic analysis of Viral Nervous Necrosis Virus detected in gilthead seabream gills from Algarve, b) sequences used for phylogenetic analysis of Infectious Pancreatic Necrotic Virus amplified from Atlantic horse mackerel skin from Peniche, c) sequences used for phylogenetic analysis of Hepatitis B virus obtained from Atlantic horse mackerel gills from Algarve, d) sequences used for phylogenetic analysis of Rhinovirus A detected in a gilthead seabream skin sample from Peniche; e) sequences used for phylogenetic analysis of Norovirus GII obtained from gilthead seabream skin from Algarve.

a)

Country	Origin	Access number
Japan	Barfin flounder	NC_013459
Japan	Barfin flounder	EU826138
Japan	Redspotted grouper	NC_008041
China	Redspotted grouper	AY690596
Japan	Tiger puffer	NC_013461
Japan	Tiger puffer	EU236149
Portugal	Gilthead seabream	LC581281*
Japan	Striped jack	NC_003449
Japan	Striped jack	D30814
USA	Mosquitoes	NC_004145
USA	Mosquitoes	X15960
USA	Mosquitoes	NC_004144
USA	Black beetle	X15959
USA	Black beetle	X00956
USA	Mosquitoes	NC_002691
USA	Mosquitoes	AF174534
Peru	Southern armyworm	NC_003692
Peru	Southern armyworm	AF171943

b)

Country	Origin	Access number
Germany	Chicken	X84034
France	Chicken	AJ310185
Canada	Insects	NC_004177
China	Mosquitoes	JX403941
Germany	Mosquitoes	JQ659254
France	Rotifer	FM995220
USA	Mosquitoes	GQ342962
Singapore	Asian seabass	MK103419
Vietnam	Blotched snakehead	NC_005982
Vietnam	Blotched snakehead	AJ459382
Japan	Yellowtail	NC_004168

Japan	Yellowtail	AY283781
Canada	Fish	NC_001915
Canada	Fish	M18049
Portugal	Atlantic horse mackerel	LC581280*
USA	Bivalve	AF342731
United Kingdom	Bivalve	AJ920336
United Kingdom	Bivalve	NC_038870

c)

Country	Origin	Access number
Brazil	Capuchin monkey	KY703886
Brazil	Capuchin monkey	NC_043528
USA	Woolly monkey	AF046996
USA	Woolly monkey	NC_028129
USA	Human	NC_003977
USA	Human	X02763
China	White-toothed shrew	MH484442
China	Asian gray shrew	MH484438
Myanmar	Bat	NC_020881
China	Pomona roundleaf bat	KF939649
China	Pomona roundleaf bat	NC_038503
Gabon	Noack's roundleaf bat	NC_024443
Gabon	Noack's roundleaf bat	KC790376
Ivory Coast	Antelope	MK620908
USA	Ground squirrel	NC_001484
USA	Ground squirrel	K02715
USA	Woodchuck	M19183
USA	Woodchuck	NC_004107
Australia	Domestic cat	MH307930
Australia	Domestic cat	NC_040719
Panama	Tent making bat	KC790381
Panama	Tent making bat	NC_024445
Portugal	Atlantic horse mackerel	LC581279 *
USA	Bluegill	KX058433
USA	Bluegill	NC_030445
USA	White sucker	NC_027922
USA	White sucker	KR229754
Germany	Heron	NC_001486
Germany	Heron	M22056
Canada	Duck	AF047045
China	Duck	NC_001344
Poland	Parakeet	JX274035
Poland	Parakeet	NC_016561
China	Tibetan frog	MH700450
China	Tibetan frog	NC_030446

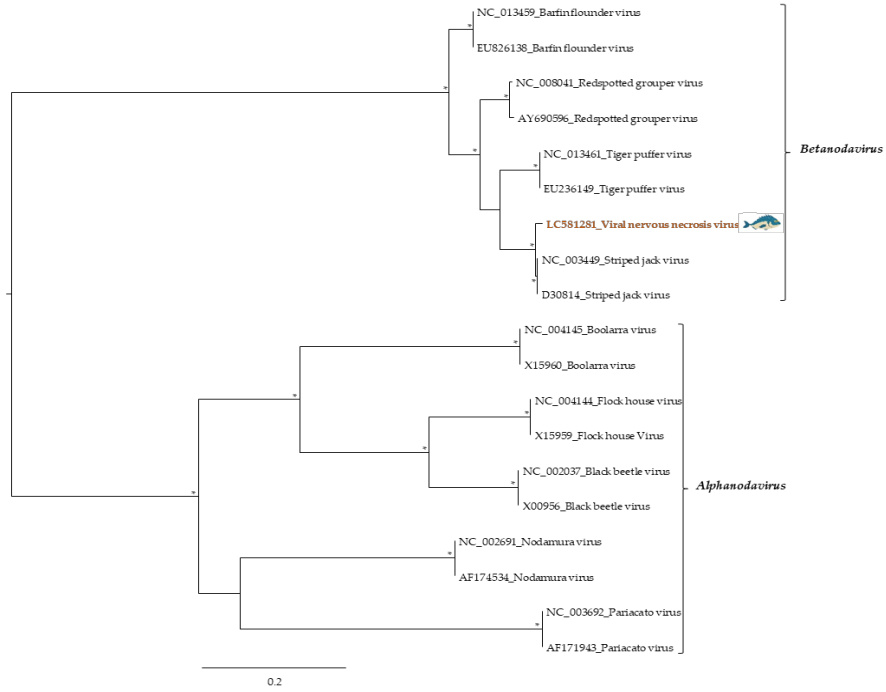
d)

Country	Origin	Access number
Finland	Human	NC_001612
Japan	Human	NC_001472
Russia	Human	AY896767
China	African green monkey	MN427525
China	African green monkey	MN427527
China	African green monkey	MN427528
China	Rhesus monkey	KU587555

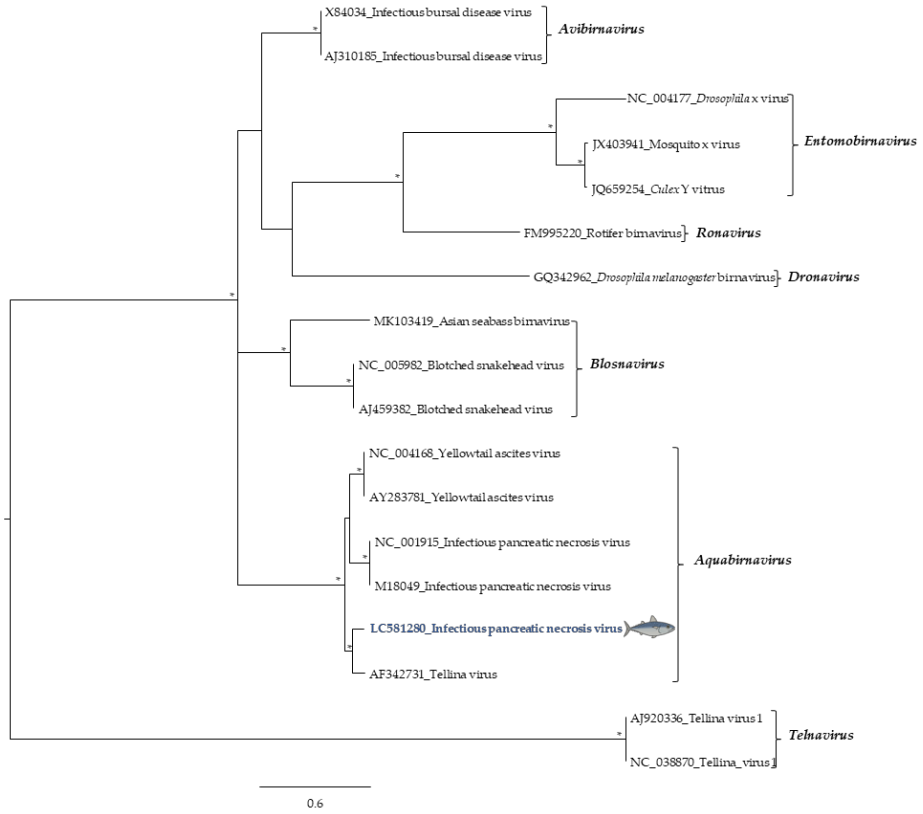
China	Rhesus monkey	NC_029905
Nigeria	Wastewater	MT542698
Germany	Bovine	MT019688
Germany	Bovine	NC_021220
USA	Human	DQ092795
Dubai	Dromedary	NC_038310
Dubai	Dromedary	KP345888
Japan	Wild boar	LC316832
Japan	Wild boar	LC316831
Japan	Wild boar	LC316819
United Kingdom	Human	NC_001430
Cameroon	Human	MK032898
Malawi	Human	MN914205
China	Rodent	KX156158
China	Rodent	NC_038989
USA	Simian	NC_038309
USA	Simian	AF326759
USA	Human	JX193795
USA	Human	JX074053
Malaysia	Human	KJ675506
Hong Kong	Human	NC_009996
USA	Human	JX074055
Portugal	Gilthead seabream	LC581278*
USA	Environment	MF043119

e)

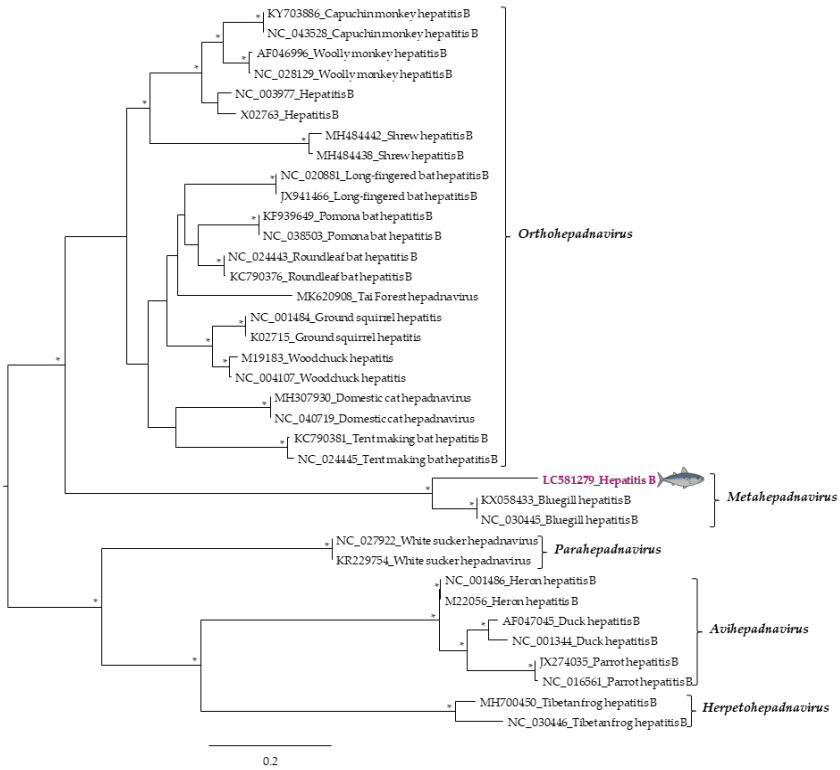
Country	**Origin**	**Access number**
Japan	Wild boar	AB863586
Canada	Wild boar	NC_012699
USA	Rhesus monkey	NC_043512
USA	Rhesus monkey	EU391643
France	Hare	NC_002615
Italy	Wolf	MF356366
Portugal	Rabbit	KF442964
China	Alpine Musk Deer	MN478485
United Kingdom	Cattle	DQ013304
United Kingdom	Cattle	NC_007916
United Kingdom	Feline	M86379
USA	Feline	NC_001481
USA	Pig	NC_002551
USA	Pig	U76874
Sweden	Human	KX431437
Sweden	Human	KY040366
Brazil	Chicken	NC_033081
Germany	Chicken	HQ010042
Germany	Turkey	NC_043516
Germany	Turkey	JQ347522
Germany	Human	AF093797
Portugal	Gilthead seabream	LC581277*
Vietnam	Human	AF504671
USA	Fathead minnow	KX371097
USA	Fathead minnow	NC_035675
Canada	Atlantic salmon	MN995807
Norway	Atlantic salmon	NC_024031


NC_013459_Barfinflounder virus
EU826138_Barfin flounder virus
NC_008041_Redspotted grouper virus
AY690596_Redspotted grouper virus
NC_013461_Tiger puffer virus
EU236149_Tiger puffer virus
LC581281_Viral nervous necrosis virus
NC_003449_Striped jack virus
D30814_Striped jack virus
Betanodavirus
NC_004145_Boolarra virus
X15960_Boolarra virus
NC_004144_Flock house virus
X15959_Flock house Virus
NC_002037_Black beetle virus
X00956_Black beetle virus
Alphanodavirus
NC_002691_Nodamura virus
AF174534_Nodamura virus
NC_003692_Pariacato virus
AF171943_Pariacato virus
0.2

(a)

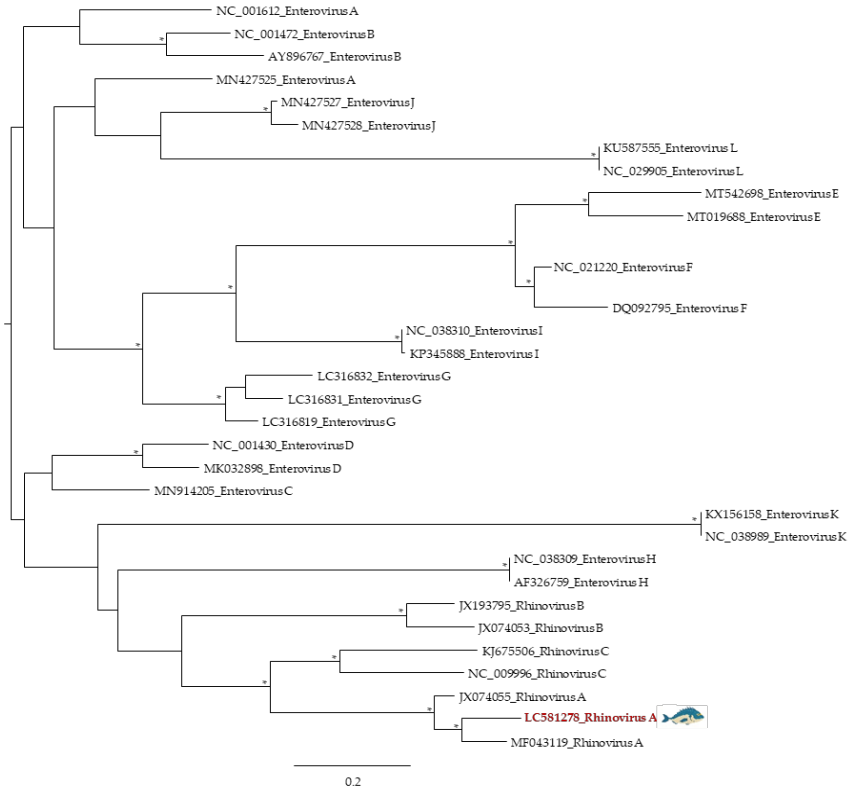

X84034_Infectious bursal disease virus
AJ310185_Infectious bursal disease virus
Avibirnavirus
NC_004177_Drosophila x virus
JX403941_Mosquito x virus
JQ659254_Culex Y vitrus
Entomobirnavirus
FM995220_Rotifer birnavirus
Ronavirus
GQ342962_Drosophila melanogaster birnavirus
Dronavirus
MK103419_Asian seabass birnavirus
NC_005982_Blotched snakehead virus
AJ459382_Blotched snakehead virus
Blosnavirus
NC_004168_Yellowtail ascites virus
AY283781_Yellowtail ascites virus
NC_001915_Infectious pancreatic necrosis virus
M18049_Infectious pancreatic necrosis virus
LC581280_Infectious pancreatic necrosis virus
AF342731_Tellina virus
Aquabirnavirus
AJ920336_Tellina virus 1
NC_038870_Tellina_virus 1
Telnavirus
0.6

(b)

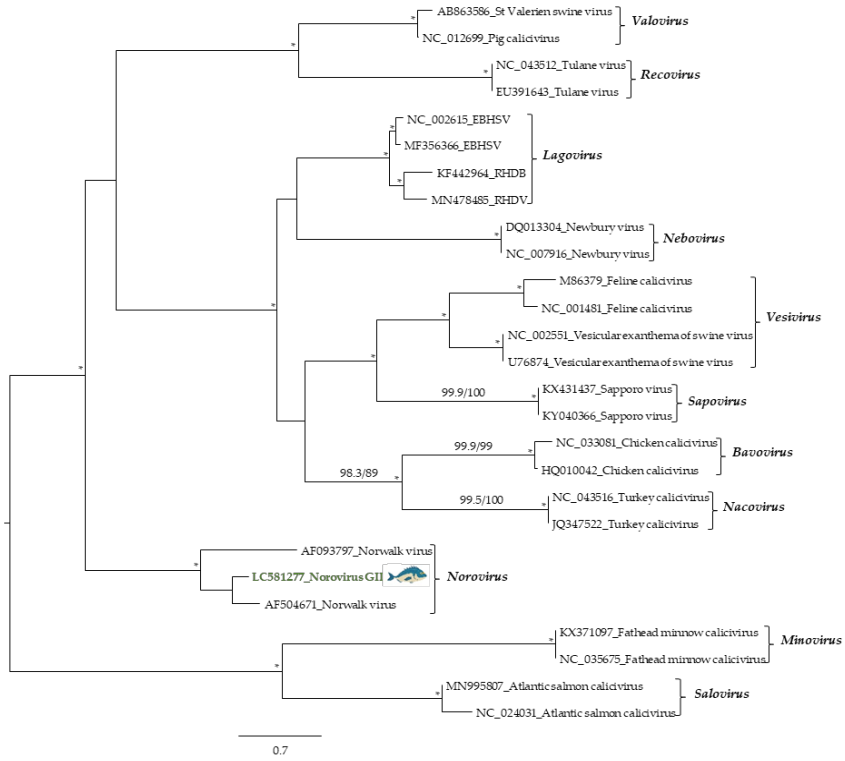


(c)

Figure S1. Maximum Likelihood phylogenetics trees of each fish pathogenic viral family identified in this study: a) *Nodaviridae* (including a partial sequence of Viral Nervous Necrosis Virus detected in gilthead seabream gills from Algarve), b) *Birnaviridae* (including a partial sequence of Infectious Pancreatic Necrotic Virus obtained from Atlantic horse mackerel skin from Peniche), and c) *Hepadnaviridae* (including a partial sequence of Hepatitis B virus obtained from Atlantic horse mackerel gills from Algarve). The viral sequences analysed in this study are coloured and highlighted by fish. All viral sequences are identified by their accession number and name. At specific branch nodes, bootstrap values ≥0.60 are displayed by *.


NC_001612_Enterovirus A
NC_001472_Enterovirus B
AY896767_Enterovirus B
MN427525_Enterovirus A
MN427527_Enterovirus J
MN427528_Enterovirus J
KU587555_Enterovirus L
NC_029905_Enterovirus L
MT542698_Enterovirus E
MT019688_Enterovirus E
NC_021220_Enterovirus F
DQ092795_Enterovirus F
NC_038310_Enterovirus I
KP345888_Enterovirus I
LC316832_Enterovirus G
LC316831_Enterovirus G
LC316819_Enterovirus G
NC_001430_Enterovirus D
MK032898_Enterovirus D
MN914205_Enterovirus C
KX156158_Enterovirus K
NC_038989_Enterovirus K
NC_038309_Enterovirus H
AF326759_Enterovirus H
JX193795_Rhinovirus B
JX074053_Rhinovirus B
KJ675506_Rhinovirus C
NC_009996_Rhinovirus C
JX074055_Rhinovirus A
LC581278_Rhinovirus A
MF043119_Rhinovirus A
0.2

(a)



(b)

Figure S2. Maximum Likelihood phylogenetics trees of each human pathogenic viral family identified in this study: a) *Picornaviridae* (partial sequence of Rhinovirus A obtained in gilthead seabream skin from Peniche); b) *Caliciviridae* (partial sequence of Norovirus GII obtained in gilthead seabream skin from Algarve). Due to the hight genetic diversity observed in *Picornaviridae*, the phylogenetic tree for this taxon was constructed based on the alignment of the closest sequences from a phylogenetic point of view. The viral sequences analysed in this study are coloured and highlighted by a blue fish. All viral sequences are identified by their accession number and name. At specific branch nodes, bootstrap values ≥0.60 are displayed by *.

Altres publicacions

A continuació es detallen, per ordre cronològic, altres publicacions no incloses en aquesta tesi però realitzades durant el doctorat:

- Fernandez-Cassi X., Timoneda N., **Martínez-Puchol S.**, Rusiñol M., Rodríguez-Manzano J., Figuerola N., Purcell N.H., Bofill-Mas S., Abril J.F., Girones R.. *Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance*. Science of the Total Environment Science of the Total Environment 618 (2018) 870–880. Factor d'impacte: 6,551; Q1

- Fernandez-Cassi X., Rusiñol M., **Martínez-Puchol S**. Viral Concentration and Amplification from Human Serum Samples Prior to Application of Next-Generation Sequencing Analysis. Chapter in "The Human Virome: Methods and Protocols, Methods in Molecular Biology", 1838:173-188, Springer Nature (2018).

- Aguado D., Fores E., Guerrero-Latorre L., Rusiñol M., **Martínez-Puchol S.**, Codony F., Girones R., Bofill-Mas S. *VirWaTest, A Point-of-Use Method for the Detection of Viruses in Water Samples*. JoVE - Journal of Visualized Experiments 2019 May 11;(147). Factor d'impacte: 1.163; Q2.

- Rusiñol M., **Martínez-Puchol S.**, Timoneda N., Fernandez-Cassi X, Perez-Cataluña A., Fernandez-Bravo A., Moreno-Mesonero L., Moreno Y., Alonso JL., Figueras M.J., Abril J.F., Bofill-Mas S., Girones R. *Metagenomic analysis of viruses, bacteria and protozoa in irrigation water*. International Journal of Hygiene and Environmental Health 224 (2020) 113440. Factor d'impacte: 4.379; Q1

- Silva-Sales M., **Martínez-Puchol S.**, Gonzales-Gustavson E., Hundesa A., Gironès R. *High Prevalence of Rotavirus A in Raw Sewage Samples from Northeast Spain*. Viruses 2020 Mar; 12(3): 318. Factor d'impacte: 3.826; Q1.

- Rusiñol M., **Martínez-Puchol S.**, Forés E., Itarte M., Girones R., Bofill-Mas S. Concentration methods for the quantification of coronavirus and other potentially pandemic enveloped virus from wastewater. Current Opinion in Environmental Science & Health 2020 Oct;17:21-28.

Science of the Total Environment 618 (2018) 870–880

Contents lists available at ScienceDirect

Science of the Total Environment

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

Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance

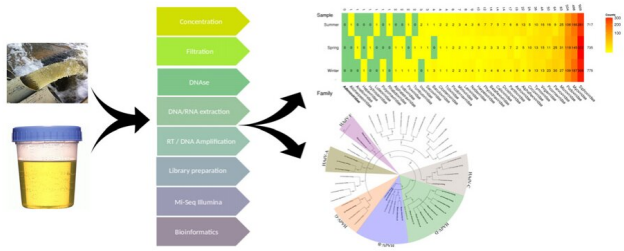
X. Fernandez-Cassi [a,*,1], N. Timoneda [a,b,1], S. Martínez-Puchol [a], M. Rusiñol [a], J. Rodriguez-Manzano [c], N. Figuerola [a], S. Bofill-Mas [a], J.F. Abril [a,b], R. Girones [a]

[a] Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Barcelona, Catalonia, Spain
[b] Computational Genomics Lab, University of Barcelona and Institute of Biomedicine (IBUB), University of Barcelona, Barcelona, Catalonia, Spain
[c] Centre for Bio-Inspired Technology, Department of Electrical and Electronic Engineering, Imperial College London, London, United Kingdom

HIGHLIGHTS

- Sensitive methods have been developed to study the metavirome of sewage
- The metavirome can pinpoint important pathogens circulating in the human population
- Mainly DNA viruses are excreted in urine, being JCPyV the most abundant
- The sewage metavirome is a useful tool for metagenomics in public health surveillance

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 2 June 2017
Received in revised form 23 August 2017
Accepted 24 August 2017
Available online 4 November 2017

Editor: D. Barcelo

Keywords:
Viral metagenomics
Human adenovirus
Viral pathogens
Sewage
Next-generation sequencing

ABSTRACT

The application of next-generation sequencing (NGS) techniques for the identification of viruses present in urban sewage has not been fully explored. This is partially due to a lack of reliable and sensitive protocols for studying viral diversity and to the highly complex analysis required for NGS data processing. One important step towards this goal is finding methods that can efficiently concentrate viruses from sewage samples. Here the application of a virus concentration method based on skimmed milk organic flocculation (SMF) using 10 L of sewage collected in different seasons enabled the detection of many viruses. However, some viruses, such as human adenoviruses, could not always be detected using metagenomics, even when quantitative PCR (qPCR) assessments were positive. A targeted metagenomic assay for adenoviruses was conducted and 59.41% of the obtained reads were assigned to murine adenoviruses. However, up to 20 different human adenoviruses (HAdV) were detected by this targeted assay being the most abundant HAdV-41 (29.24%) and HAdV-51 (1.63%). To improve metagenomics' sensitivity, two different protocols for virus concentration were comparatively analysed: an ultracentrifugation protocol and a lower-volume SMF protocol. The sewage virome contained 41 viral families, including pathogenic viral species from families *Caliciviridae*, *Adenoviridae*, *Astroviridae*, *Picornaviridae*, *Polyomaviridae*, *Papillomaviridae* and *Hepeviridae*. The contribution of urine to sewage metavirome seems to be restricted to a few specific DNA viral families, including the polyomavirus and papillomavirus species. In experimental infections with sewage in a rhesus macaque model, infective human hepatitis E and JC polyomavirus were identified. Urban raw sewage consists of the excreta of thousands of inhabitants; therefore, it is a representative sample

* Corresponding author.
E-mail address: x.fernandez-cassi@ub.edu (X. Fernandez-Cassi).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.scitotenv.2017.08.249

Chapter 13

Viral Concentration and Amplification from Human Serum Samples Prior to Application of Next-Generation Sequencing Analysis

Xavier Fernandez-Cassi, Marta Rusiñol, and Sandra Martínez-Puchol

Abstract

The protocol presented here allows the isolation, purification, nucleic acid extraction, and amplification of DNA/RNA from viruses present in human sera samples. The method allows the random amplification of the viral genomes present by using a Sequence-Independent, Single-Primer Amplification (SISPA) approach enabling the study of both DNA/RNA viruses. An amplification step is needed, as the concentration of viral DNA/RNA in serum samples is low for direct library preparation. The application of the described protocol guarantees enough randomly amplified double-strand DNA for further library preparation using Nextera XT kit from Illumina.

Key words DNA viruses, RNA viruses, Metagenomics, Blood virome, Sequence-Independent, Single-Primer Amplification (SISPA), Metagenomics

1 Introduction

The application of high-throughput sequencing (HTS) techniques in the virology field has provided us with the detection of known and unknown viruses in a wide range of samples, allowing the study of viral populations circulating through the bloodstream [1, 2]. A general overview of the described methodology is summarized in Fig. 1. The protocol presented in this chapter is based on an ultracentrifugation procedure that allows the concentration of viral particles from previously filtered serum samples. Free viral DNA and non-viral DNA is removed using a DNase treatment. Nucleic acids (NA) of viral origin are extracted from the viral capsids, tagged and amplified by using the Sequence-Independent, Single-Primer Amplification procedure, hereafter known as SISPA. Unlike other molecular techniques, the SISPA method is not hampered by the previous knowledge of the desired targeted virus sequence, allowing the amplification of all viral DNA/RNA

Andrés Moya and Vicente Pérez Brocal (eds.), *The Human Virome: Methods and Protocols*, Methods in Molecular Biology, vol. 1838, https://doi.org/10.1007/978-1-4939-8682-8_13,

jove Journal of Visualized Experiments www.jove.com

Video Article

VirWaTest, A Point-of-Use Method for the Detection of Viruses in Water Samples

David Aguado[1], Eva Fores[1], Laura Guerrero-Latorre[1,2], Marta Rusiñol[1], Sandra Martínez-Puchol[1], Francesc Codony[3,4], Rosina Girones[1], Sílvia Bofill-Mas[1]

[1]Laboratory of Viruses Contaminants of Water and Food (VIRCONT), Department of Genetics, Microbiology and Statistics, Section of Microbiology, Virology and Biotechnology, University of Barcelona

[2]Grupo de Investigación Biodiversidad, Medio Ambiente y Salud (BIOMAS), Facultad de Ingenierías y Ciencias Agropecuarias (FICA), Ingeniería en Biotecnología, Universidad de las Américas

[3]Municipal Laboratory - Waters of Mataró

[4]GenIUL

Correspondence to: Sílvia Bofill-Mas at sbofill@ub.edu

URL: https://www.jove.com/video/59463
DOI: doi:10.3791/59463

Keywords: Immunology and Infection, Issue 147, Human viruses, adenoviruses, hepatitis E virus, fecal pollution, viral concentration, viral detection, PCR, point-of-use, humanitarian, sanitation, outbreak.

Date Published: 5/11/2019

Citation: Aguado, D., Fores, E., Guerrero-Latorre, L., Rusiñol, M., Martínez-Puchol, S., Codony, F., Girones, R., Bofill-Mas, S. VirWaTest, A Point-of-Use Method for the Detection of Viruses in Water Samples. *J. Vis. Exp.* (147), e59463, doi:10.3791/59463 (2019).

Abstract

Viruses excreted by humans and animals may contaminate water sources and pose a risk to human health when this water is used for drinking, food irrigation, washing, etc. The classical fecal bacteria indicator does not always check for the presence of viral pathogens so the detection of viral pathogens and viral indicators is relevant in order to adopt measures of risk mitigation, especially in humanitarian scenarios and in areas where water-borne viral outbreaks are frequent.

At present, several commercial tests allowing the quantification of fecal indicator bacteria (FIB) are available for testing at the point of use. However, such commercial tests are not available for the detection of viruses. The detection of viruses in environmental water samples requires concentrating several liters into smaller volumes. Moreover, once concentrated, the detection of viruses relies on methods such as nucleic acid extraction and molecular detection (e.g., polymerase chain reaction [PCR]-based assays) of the viral genomes.

The method described here allows the concentration of viruses from 10 L water samples, as well as the extraction of viral nucleic acids at the point of use, with simple and portable equipment. This allows the testing of water samples at the point of use for several viruses and is useful in humanitarian scenarios, as well as at any context where an equipped laboratory is not available. Alternatively, the method allows concentrating viruses present in water samples and the shipping of the concentrate to a laboratory at room temperature for further analysis.

Video Link

The video component of this article can be found at https://www.jove.com/video/59463/

Introduction

During the first phases of any humanitarian emergency, access to clean water supplies, sanitation, and hygiene are critical for the survival of those affected. Therefore, monitoring water quality is a priority to prevent waterborne outbreaks. It is well-known that contaminated water is frequently the origin of diseases, but it is often difficult to determine the sources of viral outbreaks such as Hepatitis E virus (HEV), even with the availability of conventional laboratory methods. The control of water quality is based on the quantification of FIB[1,2,3,4]. However, it has been extensively documented that there is no correlation between the absence of FIB and the presence of viral waterborne pathogens such as rotavirus (RoV), norovirus (NoV), or HEV[5,6]. Thus, using the water quality criteria based on FIB might result in an underestimation of risks associated with the presence of waterborne viral pathogens. The surveillance of indicator viruses, such as human adenoviruses (HAdV), or specific pathogens would be helpful in defining the exposure to viral pathogens and identifying the potential source of human infection[7,8,9,10] and in validating the efficacy of sanitation measures[11].

Until now, the detection of viruses in these scenarios relied on skilled staff and complex logistics. VirWaTest (virwatest.org) is aimed at the development of a simple, affordable, and portable method for the concentration and subsequent detection of viruses from water samples at the point of use.

The virus concentration is based on the principle of organic flocculation of 10 L water samples, by which viruses are recovered in smaller volumes[12,13]. The flocs are collected and added to a buffer that lyses the viruses and prevents the nucleic acids from degradation when they stored at room temperature for not more than 2 weeks.

International Journal of Hygiene and Environmental Health 224 (2020) 113440

Contents lists available at ScienceDirect

International Journal of Hygiene and Environmental Health

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

Metagenomic analysis of viruses, bacteria and protozoa in irrigation water

Marta Rusiñol[a,*], Sandra Martínez-Puchol[a], Natalia Timoneda[a,b], Xavier Fernández-Cassi[a], Alba Pérez-Cataluña[c], Ana Fernández-Bravo[c], Laura Moreno-Mesonero[d], Yolanda Moreno[d], Jose Luís Alonso[d], Maria José Figueras[c], Josep Francesc Abril[b], Sílvia Bofill-Mas[a], Rosina Girones[a]

[a] Laboratory of Viruses Contaminants of Water and Food. Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Catalonia, Spain
[b] Computational Genomics Lab, Department of Genetics, Microbiology and Statistics, Faculty of Biology and Institute of Biomedicine (IBUB), University of Barcelona, Catalonia, Spain
[c] Microbiology Unit, Faculty of Medicine and Health Sciences, University Rovira and Virgili, Reus, Catalonia, Spain
[d] Instituto de Ingeniería Del Agua y Medio Ambiente, Universitat Politècnica de València, Valencia, Spain

ARTICLE INFO

Keywords:
Irrigation water
Metagenomics
Virus
Bacteria
Protozoa

ABSTRACT

Viruses (e.g., noroviruses and hepatitis A and E virus), bacteria (e.g., *Salmonella spp*. and pathogenic *Escherichia coli*) and protozoa (e.g., *Cryptosporidium parvum* and *Giardia intestinalis*) are well-known contributors to food-borne illnesses linked to contaminated fresh produce. As agricultural irrigation increases the total amount of water used annually, reclaimed water is a good alternative to reduce dependency on conventional irrigation water sources. European guidelines have established acceptable concentrations of certain pathogens and/or indicators in irrigation water, depending on the irrigation system used and the irrigated crop. However, the incidences of food-borne infections are known to be underestimated and all the different pathogens contributing to these infections are not known. Next-generation sequencing (NGS) enables the determination of the viral, bacterial and protozoan populations present in a water sample, providing an opportunity to detect emerging pathogens and develop improved tools for monitoring the quality of irrigation water. This is a descriptive study of the virome, bacteriome and parasitome present in different irrigation water sources. We applied the same concentration method for all the studied samples and specific metagenomic approaches to characterize both DNA and RNA viruses, bacteria and protozoa.

In general, most of the known viral species corresponded to plant viruses and bacteriophages. Viral diversity in river water varied over the year, with higher bacteriophage prevalences during the autumn and winter. Reservoir water contained *Enterobacter cloacae*, an opportunistic human pathogen and an indicator of fecal contamination, as well as *Naegleria australiensis* and *Naegleria clarki*. Hepatitis E virus and *Naegleria fowleri*, emerging human pathogens, were detected in groundwater. Reclaimed water produced in a constructed wetland system presented a virome and bacteriome that resembled those of freshwater samples (river and reservoir water). Viral, bacterial and protozoan pathogens were occasionally detected in the different irrigation water sources included in this study, justifying the use of improved NGS techniques to get a comprehensive evaluation of microbial species and potential environmental health hazards associated to irrigation water.

1. Introduction

Agricultural irrigation accounts for 36% of the total annual water usage in Europe, reaching up to 80% in some parts of the Mediterranean region (European Environment Agency, 2012). In Spain 79% of the total irrigated area (3 Mha) is irrigated with surface water and only 21% with groundwater (Drewes et al., 2017). River flow rate fluctuation and the overexploitation of groundwater resources are existing problems in Europe, where integrated water resource plans are currently being implemented. Moreover, climate change is expected to intensify problems of water scarcity and affect irrigation requirements in the Mediterranean region (Collins et al., 2009). To reduce dependency on freshwater, reclaimed water can be used for irrigation, provided that the risk associated with pathogen contamination can be minimised. Reclaimed water can be relatively nutrient rich and may reduce the need for additional applications of inorganic fertilizers (Parsons et al., 2001). In Catalonia, around 300 wastewater treatment plants (WWTP) treat 700 hm^3 of water every year to produce 204 hm^3

* Corresponding author.
E-mail address: mrusinol@ub.edu (M. Rusiñol).

https://doi.org/10.1016/j.ijheh.2019.113440
Received 26 July 2019; Received in revised form 17 December 2019; Accepted 20 December 2019

viruses

MDPI

Article

High Prevalence of Rotavirus A in Raw Sewage Samples from Northeast Spain

Marcelle Silva-Sales *,†, Sandra Martínez-Puchol, Eloy Gonzales-Gustavson, Ayalkibet Hundesa and Rosina Gironès

Microbiology, Virology and Biotechnology Section, Department of Genetics, Microbiology and Statistics, Faculty of Biology, University of Barcelona, Av. Diagonal 643, Barcelona 08028, Spain; smpuchol@gmail.com (S.M.-P.); gonzaleseloy@gmail.com (E.G.-G.); ahundesa@ub.edu (A.H.); rgirones@ub.edu (R.G.)

* Correspondence: marcelle.figueira@ufg.br

† Present address: Laboratory of Virology and Cell Culture, Department of Bioscience and Technology, Institute of Tropical Pathology and Public Health, Federal University of Goiás; Rua 235 s/n, Setor Leste Universitário, 74605-050 Goiânia, Goiás, Brazil.

Received: 28 January 2020; Accepted: 14 March 2020; Published: 16 March 2020

Abstract: Rotavirus A (RVA) is the most common virus associated with infantile gastroenteritis worldwide, being a public health threat, as it is excreted in large amounts in stool and can persist in the environment for extended periods. In this study, we performed the detection of RVA and human adenovirus (HAdV) by TaqMan qPCR and assessed the circulation of RVA genotypes in three wastewater treatment plants (WWTPs) between 2015 and 2016 in Catalonia, Spain. RVA was detected in 90% and HAdV in 100% of the WWTP samples, with viral loads ranging between 3.96×10^4 and 3.30×10^8 RT-PCR Units/L and 9.51×10^4 and 1.16×10^6 genomic copies/L, respectively. RVA VP7 and VP4 gene analysis revealed the circulation of G2, G3, G9, G12, P[4], P[8], P[9] and P[10]. Nucleotide sequencing (VP6 fragment) showed the circulation of I1 and I2 genotypes, commonly associated with human, bovine and porcine strains. It is important to mention that the RVA strains isolated from the WWTPs were different from those recovered from piglets and calves living in the same area of single sampling in 2016. These data highlight the importance of monitoring water matrices for RVA epidemiology and may be a useful tool to evaluate and predict possible emergence/reemergence of uncommon strains in a region.

Keywords: rotavirus A; wastewater; viral quantification; molecular characterization

1. Introduction

Rotavirus A (RVA) is the most common pathogen associated with acute gastroenteritis in children worldwide. It accounts for approximately 215,000 deaths annually of children up to 5 years old [1]. In Europe, RVA is responsible for 75,000–150,000 infantile hospitalization due to acute gastroenteritis associated with RVA infections. In Spain, the annual incidence of acute gastroenteritis associated with RVA in primary care ranges between 15.4 and 19.5 cases per 1000 children up to 5 years and 20 cases per 1000 children up to 3 years [2].

RVA is one of the nine groups (A–I) of rotavirus established according to the International Committee on Taxonomy of Viruses (ICTV) (ICTV, updated in July 2018 and available at https://talk.ictvonline.org/taxonomy/). The virion is non-enveloped, icosahedral, with a triple-layered protein capsid and consists of 11 segments of double-stranded genomic RNA (dsRNA), which encode six structural (VP1–4, 6 and 7) and six non-structural (NSP1–6) proteins. The binary classification system for RVA is based on molecular analysis of the two outer capsid proteins, VP7 and VP4, which define the G (for glycoprotein) and P (for protease sensitive) genotypes, respectively. To date, there are 36 G genotypes

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Environmental Science & Health

Concentration methods for the quantification of coronavirus and other potentially pandemic enveloped virus from wastewater

Marta Rusiñol[1,2], Sandra Martínez-Puchol[1,3], Eva Forés[1,3], Marta Itarte[1,3], Rosina Girones[1,3] and Sílvia Bofill-Mas[1,3]

Abstract

As the novel SARS-CoV-2 was detected in faeces, environmental researchers have been using centrifugal ultrafiltration, polyethylene glycol precipitation and aluminium hydroxide flocculation to describe its presence in wastewater samples. High recoveries (up to 65%) are described with electronegative filtration when using surrogate viruses, but few literature reports recovery efficiencies using accurate quantification of enveloped viruses. Considering that every single virus will have a different behaviour during viral concentration, it is recommended to use an enveloped virus, and if possible, a betacoronaviruses as murine hepatitis virus, as a surrogate. In this review, we show new data from a newly available technology that provides a quick ultrafiltration protocol for SARS-CoV-2. Wastewater surveillance is an efficient system for the evaluation of the relative prevalence of SARS-CoV-2 infections in a community, and there is the need of using reliable concentration methods for an accurate and sensitive quantification of the virus in water.

Addresses

[1] Section of Microbiology, Virology and Biotechnology, Department of Genetics, Microbiology and Statistics. Faculty of Biology, University of Barcelona, Catalonia, Spain

[2] GHS, Institute of Environmental Assessment & Water Research (IDAEA), CSIC, Barcelona, Spain

[3] The Water Institute of the University of Barcelona, Spain

Corresponding author: Rusiñol, Marta (martarusinol@gmail.com)

Current Opinion in Environmental Science & Health 2020, **17**:21–28

This review comes from a themed issue on **Environmental Health: COVID-19**

Edited by **Avelino Núñez-Delgado**

For a complete overview see the Issue and the Editorial

https://doi.org/10.1016/j.coesh.2020.08.002

Keywords

Concentration methods, Enveloped virus, SARS-CoV-2, Recovery efficiency, Surrogate virus.

Introduction

Many viruses that infect humans are excreted in large amounts through faeces, urine or skin desquamation, contributing to wastewater virome. Wastewater is a complex matrix that comprises a large variety of pathogenic and commensal viruses and provides important information about virus circulation, the introduction of emergent viruses and how they are transmitted among the population [1]. Waterborne viruses are generally nonenveloped and excreted in high numbers by infected individuals with or without disease, and in some cases long after the resolution of symptoms [2]. The study of excreted viruses is a very useful tool known as wastewater-based epidemiology, which has the potential to act as a complementary approach for current infectious disease surveillance systems and an early warning system for disease outbreaks [3].

The incidence of emerging microbes is a serious health concern worldwide. The increase of human–livestock contacts [4], population mobility and trade networks [5,6], climate change [7] or the wild meat trade and loss of animal habitats [8] has raised the risk of a global pandemics. Since 1980, nearly 90 novel human pathogen species have been discovered, more than 70 of those corresponded to novel human viruses, that compared to other pathogens have the potential to evolve more rapidly, being 80 of these associated with nonhuman reservoirs [9,10]. Influenza viruses (H1N1, H7N1 and H7N9), human immunodeficiency virus, Ebola virus, coronaviruses as SARS-CoV, MERS and the SARS-CoV-2 causing the COVID-19 pandemic have been the most significant.

SARS-CoV-2 was identified in China at the end of 2019 [11] and has become the first pandemic coronavirus (CoV). After the first case report of the presence of SARS-CoV-2 RNA in faeces [12], and because of the presence in the past of SARS-CoV-1 in faeces and sewage [13–15], the scientific community started to investigate if this virus could spread into the environment. Specific stability of SARS-CoV-2 has only been tested in aerosols and surfaces [16], but it is known that enveloped viruses are capable of retaining infectivity for days to months in aqueous environments [17–19]. On