

UVAH

Aplicación de la secuenciación masiva metagenómica para el diagnóstico de infecciones neuroológicas de origen vírico no filiadas

**Application of metagenomic next-generation
sequencing for the diagnosis of neurological
infections of viral origin that remained
undiagnosed with conventional methods**

**Máster Universitario en Microbiología Aplicada a la Salud Pública e
Investigación en Enfermedades Infecciosas**

Presentado por:

D^ª ANA DEL AMO DE PALACIOS

Dirigido por:

Dra. D^ª M^ª DOLORES FERNÁNDEZ GARCÍA

Dra. D^ª MARÍA CABRERIZO SANZ

Alcalá de Henares, a 19 de junio de 2023

ÍNDICE

INTRODUCCIÓN	7
JUSTIFICACIÓN Y OBJETIVOS	11
MATERIAL Y MÉTODOS	11
1. Muestras y pacientes	11
2. Controles positivos y negativos.....	13
3. Método de extracción del material genómico.....	13
4. Metodología para la preparación de librerías	14
5. Metodología de enriquecimiento de librerías	15
6. Bioanalyzer	17
7. Secuenciación masiva por Illumina	17
8. Confirmación de los datos mediante PCR	17
9. Análisis bioinformático de los datos	17
10. Análisis filogenético	19
RESULTADOS.....	19
Resultados de metagenómica.....	19
Muestras utilizadas como controles positivos	20
Muestras idiopáticas de LCR	20
Otros virus encontrados en muestras idiopáticas	26
Controles negativos	27
Análisis filogenético.....	27
DISCUSIÓN	28
CONCLUSIONES	33
BIBLIOGRAFÍA.....	34
AGRADECIMIENTOS.....	40
ANEXO 1: definición de términos bioinformáticos.....	41
ANEXO 2: secuencias consenso obtenidas mediante la plataforma IDseq	42
ANEXO 3: ABREVIATURAS Y SIGLAS	42
INFORME TURNITIN DE ORIGINALIDAD	44

RESUMEN

Las encefalitis y meningitis son algunas de las enfermedades neurológicas más frecuentes que se diagnostican en hospitales, la mayoría causadas por agentes microbiológicos, entre los que destacan los virus. Sin embargo, un porcentaje alto de los pacientes que sufren infecciones neurológicas de sospecha clínica viral quedan sin filiación etiológica. Esto puede deberse bien a la falta de sospecha clínica de determinados virus, falta de sensibilidad de los métodos diagnósticos convencionales o a una etiología por virus de nuevas especies o nuevos genotipos. La secuenciación masiva metagenómica (mNGS) presenta numerosas ventajas respecto a las técnicas moleculares clásicas usadas en Microbiología, consideradas “gold standard”. Entre ellas su elevado rendimiento y sensibilidad. En este trabajo se ha usado la mNGS en una serie retrospectiva de muestras de líquidos cefalorraquídeos (LCRs) de pacientes con sospecha de meningitis/encefalitis (n=34). Las muestras seleccionadas fueron enviadas al Centro Nacional de Microbiología (CNM) por hospitales españoles para su diagnóstico resultando negativas a los principales virus productores de infecciones neurológicas (herpesvirus/enterovirus). Se usaron herramientas genómicas y bioinformáticas de Pikavirus, Kraken, BLAST y IDseq para identificar la etiología vírica. Entre nuestros resultados, destaca la identificación del virus Toscana en una de las muestras, que, mediante confirmación de PCR, se ha podido establecer una posible causa de meningitis. Además, se han encontrado otros virus que causan encefalitis/meningitis, como herpesvirus, norovirus, adenovirus, lisavirus europeo de murciélago y el virus de la rubéola.

Palabras clave: secuenciación masiva, virus, metagenómica, librerías, encefalitis, meningitis, líquido cefalorraquídeo, IDseq.

ABSTRACT

Encephalitis and meningitis are among the most common neurological diseases diagnosed in hospitals, most of them caused by microbiological agents, including viruses. However, a high percentage of patients suffering from clinically suspected viral neurological infections remain without aetiological affiliation. This may be due either to a lack of clinical suspicion of certain viruses, lack of sensitivity of conventional diagnostic methods, or to an aetiology due to viruses of new species or new genotypes. Massive metagenomic sequencing (mNGS) has numerous advantages over the classical molecular techniques used in microbiology, which are considered the "gold standard". These include its high yield and sensitivity. In this study, mNGS was used in a retrospective series of cerebrospinal fluid (CSF) samples from patients with suspected meningitis/encephalitis (n=34). Selected samples were sent to the Centro Nacional de Microbiología (CNM) by Spanish hospitals for diagnosis and tested negative for the main viruses causing neurological infections (herpesvirus/enterovirus). Pikavirus, Kraken, BLAST and IDseq genomic and bioinformatics tools were used to identify the viral aetiology. Our results include the identification of Toscana virus in one of the samples, which, by PCR confirmation, has been able to establish a possible cause of meningitis. In addition, other viruses causing encephalitis/meningitis have been found, such as herpesvirus, norovirus, adenovirus, European bat lyssavirus and rubella virus.

Key words: next generation sequencing, virus, metagenomics, libraries, encephalitis, meningitis, cerebrospinal fluid, IDseq.

INTRODUCCIÓN

Las encefalitis y meningitis son una de las causas más frecuentes de discapacidad y mortalidad en cuanto a enfermedades del sistema nervioso central (SNC)¹. El SNC se puede infectar por diferentes microorganismos, como virus, hongos, bacterias, helmintos y protozoos. De todos ellos, la causa más frecuente de infección del SNC son los virus^{1,2}. Las meningitis víricas se caracterizan por un cuadro con pleocitosis de predominio linfocitario y ausencia de agentes bacterianos en el líquido cefalorraquídeo (LCR). Por lo tanto, habitualmente, los términos meningitis aséptica, meningitis linfocitaria y meningitis vírica se emplean como sinónimos. Los virus que están asociados más frecuentemente a las encefalitis (inflamación del SNC que se relaciona con una disfunción neurológica) y meningitis (proceso inflamatorio de las meninges), son el virus del herpes simple (HVS), el herpesvirus varicela-zóster (VZV) o los enterovirus (EV). Las encefalitis también pueden aparecer como complicación de enfermedades víricas que usualmente cursan sin implicación neurológica. Puede ser el caso de la propia varicela, el sarampión o en menor medida la parotiditis, la gastroenteritis causada por norovirus o gripe, existiendo publicados casos de muchas otras enfermedades y agentes víricos². Además, los virus pueden producir otra serie de síndromes neurodegenerativos lentos únicamente relacionados a virus, como ocurre en el caso de la encefalopatía por el VIH, o la leucoencefalopatía multifocal progresiva causada por el poliomavirus JC³.

Las encefalitis y meningitis también pueden producirse por virus de transmisión vectorial mediante artrópodos (arbovirus) o roedores (robovirus). En nuestro país, los virus de transmisión vectorial más comunes que producen enfermedades neurológicas son: el virus West Nile (WNV) transmitido por la picadura de mosquito, el virus Toscana (TOSV) transmitido por la picadura de flebótomos, y el virus de la coriomeningitis linfocitaria (LCMV) transmitido por el contacto con roedores⁴. Normalmente, los casos de infección neurológica causada por arbovirus ocurren en los meses de primavera y verano, ya que es cuando existe un crecimiento y expansión exponencial tanto de los mosquitos como de los flebótomos, quienes tienen un periodo de reproducción y actividad a temperaturas cálidas⁵.

Existe, por tanto, un rango considerable de virus que pueden producir enfermedades neurológicas. Sin embargo, para identificar el agente etiológico de las infecciones del SNC, los laboratorios utilizan pruebas de rutina, que incluyen fundamentalmente la reacción en cadena de la polimerasa (PCR), a partir de muestras de LCR. En los últimos años, las pruebas de diagnóstico molecular se han mejorado rápidamente con la PCR en tiempo real multiplex con distintas dianas de virus o, incluso, con dianas de origen vírico, bacteriano y fúngico en el mismo kit, lo cual permite detectar muchos más patógenos en el LCR con una alta sensibilidad y

especificidad⁶. En la actualidad, se encuentran disponibles equipos comerciales para la detección de EV y herpesvirus (HV), que son los virus implicados con mayor frecuencia. Sin embargo, a pesar de los métodos de diagnóstico molecular mejorados, la etiología sigue siendo desconocida en aproximadamente entre un 40 y 60% de los casos con una sospecha de meningitis o encefalitis viral, pudiendo llegar a ser de hasta un 81%⁷⁻¹⁰.

En España, un estudio en 2012 destinado a determinar los virus causantes de infecciones del SNC mediante pruebas convencionales, concluyó que un número significativo de casos (43% meningitis, 60% de meningoencefalitis y 72% de encefalitis) permanecieron sin diagnóstico etiológico¹¹. Los datos actuales del Centro Nacional de Microbiología (CNM) muestran que aproximadamente, el 80% de los casos con sospecha clínica de infección viral del SNC permanecen sin un diagnóstico etiológico (comunicación personal al CNM). Esto provoca que a estas personas no se les dé un tratamiento adecuado, ya que, al no saber el origen, no se puede ajustar la medicación a las necesidades reales del paciente para buscar una solución y cura eficaces para cada caso^{2,12,13}.

Aunque la tasa de mortalidad asociada a la meningitis vírica es baja, la morbilidad que producen estos cuadros es elevada. La meningitis vírica se ha considerado tradicionalmente una enfermedad autolimitada benigna, pero varios informes sugieren lo contrario¹⁴⁻¹⁸. Se han encontrado secuelas neuropsiquiátricas a largo plazo, particularmente ansiedad, depresión y disfunción neurocognitiva en casos de meningitis vírica^{14,17,18}. Aunque los pacientes con meningitis bacteriana tienen inicialmente una enfermedad más grave en términos de necesidad de cuidados y mortalidad, a más largo plazo todos los pacientes con meningitis, virales y bacterianas, tuvieron secuelas que afectaron la calidad de la vida, incluyendo problemas significativos con la memoria y la salud mental¹⁴. La encefalitis se asocia con alta morbilidad y mortalidad. Si la infección no es fatal, las personas a menudo tienen graves dificultades físicas, cognitivas, emocionales, conductuales y sociales.

Por todo ello, conocer la etiología de las infecciones neurológicas causadas por virus es de especial importancia. Además, otros factores que hacen que sea importante establecer la etiología de estas infecciones son: evitar A) el uso innecesario de antibióticos, B) el uso innecesario de antivirales al orientar el tratamiento con antivíricos específicos en casos como el de los herpesvirus, C) el uso innecesario de inmunosupresores (la alternativa etiológica principal a una infección del SNC es la autoinmune) y D) el ingreso hospitalario innecesario de casos producidos por virus de buen pronóstico como, por ejemplo, los EVs causantes además de brotes epidémicos que pueden producir situaciones de sobrecarga en los hospitales. En el caso de los

arbovirus neurotropos, confirmar su rol en estas patologías es importante para las estrategias de comunicación y gestión de riesgos incluyendo en estas últimas actividades como el control vectorial a nivel local o medidas de seguridad en las transfusiones de sangre. En resumen, la identificación de la causa es clave para el correcto manejo del paciente y tomar medidas de salud pública¹⁴.

Las causas por las que podría haber un alto porcentaje de infecciones neurológicas que quedan sin filiar son múltiples. Una podría ser la amplia variedad de virus que pueden causar enfermedad neurológica y, por tanto, la falta de sospecha clínica a determinados virus. Un problema añadido es la identificación de nuevas manifestaciones neurológicas en virus que no se sabía que tenían tropismo por el SNC, como, por ejemplo, el virus del Zika, el EV-D68 o el virus de la hepatitis E¹⁹⁻²¹, lo cual dificulta la sospecha clínica. Otra causa importante sería la falta de sensibilidad de los métodos diagnósticos convencionales. La PCR es considerada la técnica molecular “gold standard” para el diagnóstico de las infecciones de sospecha vírica. Sin embargo, en esta técnica se necesita conocer previamente la secuencia del virus del que se sospecha para poder detectarlo en las muestras clínicas. Además, hay que tener en cuenta que continuamente se están descubriendo nuevas especies o nuevos genotipos o variantes de virus neurotropos². Si se trata de una nueva especie, genotipo o variante podría afectar a los resultados obtenidos por PCR, ya que los cebadores de la PCR pueden fallar en la amplificación del genoma viral, si existen mutaciones en las regiones de unión de los mimos¹⁹⁻²¹.

Otro de los inconvenientes es el tipo de muestra que se requiere para el diagnóstico de estas enfermedades, ya que, al tratarse del SNC, se necesitan muestras de LCR. Generalmente, las muestras de LCR que llegan al laboratorio tienen un volumen pequeño y limitado¹², y por ello se hace complicado repetir pruebas de laboratorio. Además, dependiendo de la infección y el agente vírico, generalmente la carga viral que se encuentra en el LCR de una persona con encefalitis o meningitis es demasiado baja, lo que dificulta su detección en los métodos tradicionales de diagnóstico. De hecho, en algunos estudios se ha visto que, tras hacer una biopsia cerebral de aquellos pacientes que han fallecido a causa de una encefalitis vírica y realizar una PCR tanto del órgano sólido como del LCR, los resultados eran negativos para el LCR y positivos para la muestra de cerebro^{2,13}. Por lo que detectar un patógeno en el LCR de un paciente infectado requiere técnicas de alta sensibilidad. Además, no siempre esos patógenos se encuentran presentes en el mismo tiempo en el que se manifiestan síntomas en el paciente. De tal manera que, a la hora de buscar un agente etiológico de esa enfermedad, ya no lo encontraremos en el LCR, a pesar de que éste se encuentre en el enfermo²².

La secuenciación metagenómica de próxima generación (mNGS) es una novedosa técnica imparcial de secuenciación que utiliza billones de fragmentos de ácidos nucleicos para recomponer el genoma de un agente etiológico. De forma general, el proceso de secuenciación puede dividirse en tres: 1) preparación de librerías, 2) inmovilización sobre una superficie bidimensional y amplificación in vitro de la librería y 3) secuenciación y captación de señal²³. Para construir la librería primero se extrae el material genético de la muestra y posteriormente se produce una fragmentación del ADN generándose pequeños fragmentos de distinto tamaño a los que a continuación se unen a sus extremos unas moléculas llamadas adaptadores y un índice que permite poder identificarlo posteriormente mediante técnicas bioinformáticas, así como la fijación a una superficie bidimensional donde tendrá lugar la reacción de secuenciación. El conjunto de todos los fragmentos de ADN unidos a sus adaptadores se conoce como librería. De manera opcional se puede hacer a continuación una etapa de enriquecimiento que consiste en seleccionar exclusivamente el patógeno o la molécula de interés antes de la secuenciación. Hay varios tipos de enriquecimiento:

1. Métodos de captura por hibridación: se utiliza una hibridación por sondas, en la que hay un enriquecimiento del material genómico mediante la adición de oligonucleótidos y una enzima de amplificación²⁴.
2. Métodos de amplificación mediante PCR: la PCR se usa cada vez más para amplificar muestras complicadas procesadas con mediante secuenciación metagenómica y librerías, ya que enriquece el producto final mediante enzimas de amplificación y nucleótidos libres. Esto aumenta la capacidad de análisis de datos metagenómicos mediante herramientas bioinformáticas²⁵.

Tras la creación de las librerías (enriquecidas o no) se procede a utilizar las tecnologías NGS para secuenciarlas y obtener las secuencias o *reads*. De entre las diferentes tecnologías de NGS actualmente disponibles, la tecnología Illumina es la más extendida en el ámbito clínico dado que su alto rendimiento y la precisión de las lecturas generadas que hacen que sea la más apropiada para la mayoría de las aplicaciones de la secuenciación masiva en Microbiología Clínica²³.

Las ventajas de esta técnica son múltiples. Esta técnica permite la identificación de material genético de todos los microorganismos presentes en cualquier tipo de muestra biológica o ambiental con una alta sensibilidad, alto rendimiento y costo relativamente bajo. Con la mNGS es posible detectar virus en baja concentración en el LCR, lo cual, con otras técnicas como la PCR, resultarían indetectables^{1,2,12,13,22,26}. Otras ventajas que hacen que la mNGS se emplee cada vez

más en el diagnóstico clínico son A) la capacidad de detectar multitud de agentes infecciosos en un mismo ensayo, sin una sospecha previa basada en características clínicas^{22,26}, B) la capacidad de identificar cualquier agente infeccioso, tanto sospechado como insospechado, o incluso patógenos nuevos que se encuentren en la muestra estudiada. Se ha comprobado que la mNGS ha aumentado un 87% de los diagnósticos de otro tipo de enfermedades, como, por ejemplo, enfermedades respiratorias infecciosas²⁶, por lo que resulta una aplicación bastante prometedora para las enfermedades neurológicas de origen infeccioso.

JUSTIFICACIÓN Y OBJETIVOS

Las infecciones del SNC, tales como la meningitis y la encefalitis producidas por virus representan una carga importante para la salud humana. Sin embargo, en la práctica clínica, un porcentaje alto de los pacientes que sufren infecciones neurológicas con sospecha clínica viral quedan sin filiación etiológica. Los datos actuales del CNM muestran que aproximadamente el 80% de los LCRs de casos con sospecha clínica de infección neurológica viral que llegan al centro procedentes de las peticiones de diagnóstico de hospitales españoles permanecen sin filiar.

Dado que la mNGS presenta numerosas ventajas respecto a las técnicas moleculares clásicas consideradas el “gold standard”, entre ellas su elevado rendimiento y sensibilidad, el objetivo de este trabajo consistirá en su utilización para el diagnóstico virológico de infecciones neurológicas de pacientes con sospecha de meningitis y/o encefalitis en una serie retrospectiva de muestras de LCRs procesadas en el CNM por métodos convencionales principalmente para herpesvirus y enterovirus y a las que no se ha llegado un diagnóstico etiológico.

Objetivos secundarios:

1. Optimizar el protocolo de mNGS para la identificación de genomas virales en muestras de LCR.
2. Confirmación de los resultados obtenidos por PCR convencional y secuenciación Sanger.
3. Comparar las secuencias de los virus obtenidos con las secuencias halladas por Sanger.
4. Realizar un análisis filogenético con la secuencia genómica de los virus encontrados y analizar sus relaciones filogenéticas con secuencias detectadas en otros países.

MATERIAL Y MÉTODOS

1. **Muestras y pacientes (Tabla 1):** las muestras empleadas son LCRs enviados al CNM para el diagnóstico de pacientes con sospecha clínica de infección neurológica de etiología vírica (principalmente meningitis encefalitis) durante el año 2022 por hospitales de la

Comunidad de Madrid (Hospital Central de la Defensa Gómez Ulla, Hospital Puerta del Hierro, Hospital Príncipe de Asturias, Hospital Universitario Miguel Servet y Hospital Río Carrión) (Tabla 1). Al llegar al CNM, la mitad de la muestra es utilizada para el diagnóstico de rutina, y la otra mitad se guarda en la seroteca del CNM a -20°C. Las muestras empleadas para el presente proyecto proceden de la seroteca. Las muestras fueron seleccionadas según los siguientes criterios: 1) volumen suficiente (mínimo 500 µL), 2) muestras de pacientes con sospecha de infección del SNC según criterio clínico, 3) aquellos LCR a los que se había hecho test de diagnóstico de rutina en el CNM para los principales virus productores de infección neurológica y cuyo resultado fue negativo. El diagnóstico de rutina del CNM se realiza, según criterio clínico, normalmente para EV, HV (herpesvirus humano tipo 1 y tipo 2, VZV, virus de Epstein-Barr (EBV), citomegalovirus (CMV) y herpesvirus tipo 6). Para los pacientes con sospecha clínica de TOSV, virus de la rabia (lisavirus) o MuV, la muestra de LCR también se analizó para estos virus en el CNM en diferentes unidades. La descripción de los ensayos de rutina para el diagnóstico de virus para cada muestra está en el Tabla 1.

Muestra	Edad	Fecha	Hospital	Diagnóstico clínico	PCR								
					EV	VHS	VVZ	EBVL	CMVL	HHV6L	HHV7	HHV8	
LCRA119	29	feb-22	Hospital C. de la Defensa Gómez Ulla	-	-	-	-	-	-	-	-	-	-
LCRA120	48	mar-22	Hospital Puerta de Hierro	Neuropatía	-	-	-	-	-	-	-	-	-
LCRA121	39	mar-22	Hospital Puerta de Hierro	Neuropatía	-	-	-	-	-	-	-	-	-
LCRA122	51	mar-22	Hospital Puerta de Hierro	S. infección neurológica (+ cefalea)	-	-	-	-	-	-	-	-	-
LCRA123	62	mar-22	Hospital U. Príncipe de Asturias	Infección neurológica (leve pleocitosis y bajo nivel de consciencia)	-	-	-	-	-	-	-	-	-
LCRA124	56	mar-22	Hospital U. Príncipe de Asturias	Encefalitis y S. de meningoencefalitis (parkinsonismo + lesiones en SB*)	-	-	-	-	-	-	-	-	-
LCRA125	60	mar-22	Hospital Puerta de Hierro	Neuropatía	-	-	-	-	-	-	-	-	-
LCRA126	56	mar-22	Hospital Puerta de Hierro	Neuropatía	-	-	-	-	-	-	-	-	-
LCRA127	48	mar-22	Hospital Puerta de Hierro	S. de infección neurológica (+ fiebre)	-	-	-	-	-	-	-	-	-
LCRA128	86	mar-22	Hospital Puerta de Hierro	S. de infección neurológica (+ fiebre)	-	-	-	-	-	-	-	-	-
LCRA129	83	mar-22	Hospital Puerta de Hierro	S. de infección neurológica (+ fiebre)	-	-	-	-	-	-	-	-	-
LCRA130	32	mar-22	Hospital Puerta de Hierro	S. de infección neurológica (+ fiebre)	-	-	-	-	-	-	-	-	-
LCRA131	38	abr-22	Hospital Puerta de Hierro	S. infección neurológica (+ cefalea)	-	-	-	-	-	-	-	-	-
LCRA132	65	abr-22	Hospital Puerta de Hierro	S. infección neurológica (+ cefalea)	-	-	-	-	-	-	-	-	-
LCRA136	56	abr-22	Hospital Puerta de Hierro	Encefalitis	-	-	-	-	-	-	-	-	-
LCRA139	61	may-22	Hospital Puerta de Hierro	Meningitis	-	-	-	-	-	-	-	-	-
LCRA140	76	jun-22	Hospital Puerta de Hierro	S. infección neurológica (+ esclerosis múltiple)	-	-	-	-	-	-	-	-	-
LCRA142	67	jun-22	Hospital U. Príncipe de Asturias	Posible meningoencefalitis	-	-	-	-	-	-	-	-	-
LCRA145	61	jun-22	Hospital Puerta de Hierro	S. infección neurológica (+ cefalea)	-	-	-	-	-	-	-	-	-
LCRA148	40	jul-22	Hospital Puerta de Hierro	Cuadro neurológico	-	-	-	-	-	-	-	-	-
LCRA149	24	jul-22	Hospital Puerta de Hierro	S. encefalitis	-	-	-	-	-	-	-	-	-
LCRA150	38	jul-22	Hospital Puerta de Hierro	S. encefalitis	-	-	-	-	-	-	-	-	-
LCRA153	55	ago-22	Hospital Puerta de Hierro	S. encefalitis	-	-	-	-	-	-	-	-	-
LCRA154	67	ago-22	Hospital Puerta de Hierro	Baja S. encefalitis	-	-	-	-	-	-	-	-	-
LCRA155	24	ago-22	Hospital Puerta de Hierro	S. encefalitis	-	-	-	-	-	-	-	-	-
LCRA156	74	ago-22	Hospital Puerta de Hierro	S. encefalitis	-	-	-	-	-	-	-	-	-
LCRA157	76	ago-22	Hospital Puerta de Hierro	S. encefalitis	-	-	-	-	-	-	-	-	-
LCRA158	62	ago-22	Hospital Puerta de Hierro	Encefalitis	-	-	-	-	-	-	-	-	-
LCRA163	22	sep-22	Hospital Puerta de Hierro	Guillain-Barré	-	-	-	-	-	-	-	-	-
LCRA165	41	sep-22	Hospital Puerta de Hierro	Encefalitis	-	-	-	-	-	-	-	-	-
LCRA170	23	oct-22	Hospital Puerta de Hierro	S. encefalitis	-	-	-	-	-	-	-	-	-
LCRA183	74	nov-22	Hospital Puerta de Hierro	Encefalitis	-	-	-	-	-	-	-	-	-
LCRA184	18	ago-22	Hospital U. Miguel Servet	Meningitis linfocitaria (parálisis periférica facial + picadura de garrapata)	-	-	-	-	-	-	-	-	-
LCRA185	26	sep-22	Universidad General Río Carrión	Meningoencefalitis aguda (infección)	-	-	-	-	-	-	-	-	-

Tabla 1: Datos clínicos de las muestras de LCR, atendiendo a la edad del paciente, la fecha en la que se ha tomado la muestra de LCR, el hospital de donde proviene la muestra, el diagnóstico clínico por el cual se manda la muestra (S. hace referencia a *sospecha*), y las diferentes PCRs de rutina que se realizaron a cada muestra al llegar al CNM. Los guiones en la sección "PCR" significan que esa muestra resultó negativa para el virus seleccionado. Si aparece la casilla en blanco, significa que no se hizo PCR para esa muestra.

*SB: sustancia blanca.

- 2. Controles positivos y negativos (Figura 1.2):** para validar los experimentos, se añadieron controles negativos en cada uno de los pasos principales del ensayo, es decir, en la extracción, en el comienzo de las librerías y en la fase de amplificación. Estos controles negativos constaban de agua libre de RNAasas. Para los controles positivos, se usaron diferentes muestras que fueron positivas por PCR para distintos virus, tanto ARN como ADN. Estas muestras se presentan en la Tabla 2, donde se puede observar coinfección de virus tanto ARN como ADN en los controles llamados 2 y 5

Control positivo	Tipo de muestra	Virus confirmado por PCR
1 (Rotavirus)	Heces	Rotavirus A (ARN)
2 (Sapovirus)	Heces	Sapovirus (ARN), Salivirus (ARN), Adenovirus (ADN), Enterovirus A (ARN)
3 (Sarampión)	Exudado faríngeo	Sarampión (ARN)
4 (Monkeypox)	Lesión cutánea	Monkeypox virus (ADN)
5 (Adenovirus)	Heces	Adenovirus (ADN), Norovirus (ARN), Enterovirus A (ARN)

Tabla 2: tipos de muestras que se emplearon como controles positivos para la extracción, elaboración de librerías, amplificación y enriquecimiento, y posterior secuenciación.

- 3. Método de extracción del material genómico (Figura 1.3):** este procedimiento se lleva a cabo mediante el kit comercial QIAmp MinELUTE VIRUS SPIN de Qiagen, en el cual se extrae el material genético de las muestras de LCR, tanto ARN como ADN, según el protocolo propuesto por el kit: en primer lugar, se mezclan 200 μ l de muestra con proteasa y con buffer AL (que asegura la inactivación de las RNAasas) sin Carrier. Después, se incuba a 56°C durante 15 minutos. Tras este tiempo, se añade etanol y se deja incubar 5 minutos a temperatura ambiente. Todo el contenido se pasa a una columna, se centrifuga, y se retira el filtrado. A esa columna se le añade el buffer AW1 (previamente diluido con etanol absoluto, al igual que el buffer AW2), y se vuelve a centrifugar. Se añade el buffer AW2 y se centrifuga de nuevo. Después de cada una de estas centrifugaciones, se elimina el filtrado. A continuación, se añade etanol absoluto, y se centrifuga 2 veces seguidas, eliminando entre ambas el filtrado. Posteriormente, se incuba la muestra a 56°C durante 3 minutos, y se añaden 22 μ l de buffer de elución, dejándolo incubar a temperatura ambiente durante 5 minutos. Después se centrifuga y se pasa el eluido a un tubo nuevo. Tras la extracción, obtendremos 22 μ l de eluido en el cual se encuentra el material extraído que se utilizará para la preparación de librerías. Antes de realizar las librerías, se mide la concentración de ARN que hay en la muestra mediante un QuantiFluor dsDNA system de la marca Promega.

4. Metodología para la preparación de librerías (Figura 1.3): se utilizará el método SISPA (Sequence Independent Single Primer Amplification) para generar ADNc a partir del ARN que hemos extraído de las muestras. En este proceso, se amplifican las muestras que tienen tanto alta como baja carga viral de fragmentos de secuencia desconocida mediante cebadores random e índices específicos que servirán para la identificación de cada muestra de manera bioinformática después de la secuenciación. Para la preparación de estas librerías, se usa el kit Librería NEBNext Ultra II Directional RNA Library Prep, junto con el set 1 de índices (NEBNext Multiplex Oligo for Illumina), y se utiliza el siguiente protocolo:

- i. En primer lugar, se calcula el volumen que se debe añadir del extracto de cada muestra, para tener un total de 5ng. Se fragmenta el ARN de cada muestra y se sintetizan la primera y la segunda hebra del ADN complementario en dos ciclos diferentes. Después de esto, se purifica la doble hebra de ADNc mediante AMPure beads (NebNext Sample Purification Beads) y etanol al 80% preparado ese mismo día. De esta purificación se obtienen 50 µl de sobrenadante, que se usarán en el día 2.
- ii. A continuación, se usa el sobrenadante de la doble hebra del ADNc, y se preparan esas hebras de forma que los extremos queden igualados. A continuación, se añaden los adaptadores, y se vuelve a purificar el producto originado tras una serie de ciclos. Las purificaciones también se realizan mediante las AMPure beads y etanol al 80% preparado ese mismo día.
- iii. Por último, se añaden los índices a cada una de las muestras, de tal manera que éstas se puedan identificar posteriormente. Los índices utilizados contienen en el mismo pocillo un índice general (i5) que sirve para identificar el pool de muestras y el kit utilizado, y cada uno tiene un índice específico (i7), que es con el que se identificará cada muestra a la hora de secuenciar. Posteriormente, se realiza una nueva purificación con las AMPure beads y etanol al 80%.

Posteriormente, se mide la concentración de ADN que existe en cada una de las librerías con QuantiFluor dsDNA system de la marca Promega. Además, se separarán 2 µl del producto para hacer el bioanalyzer (ver apartado 7 de Material y Métodos), y el resto se guarda a -20°C.

5. **Metodología de enriquecimiento de librerías (Figura 1.3):** se procede a realizar una amplificación por enriquecimiento con sondas mediante el kit Twist Pan Viral Panel v2 de Twist Bioscience. El panel panviral contiene más de 600.000 sondas para el enriquecimiento específico de más de 1.000 patógenos vitales humanos a partir de una sola muestra. Es un método simplificado, eficiente y preciso para examinar muestras de pacientes y mejorar le diagnóstico de los mismos, rastreando infecciones virales emergentes. Para ello, se ha seguido el protocolo adjunto al kit que consta de los siguientes pasos:

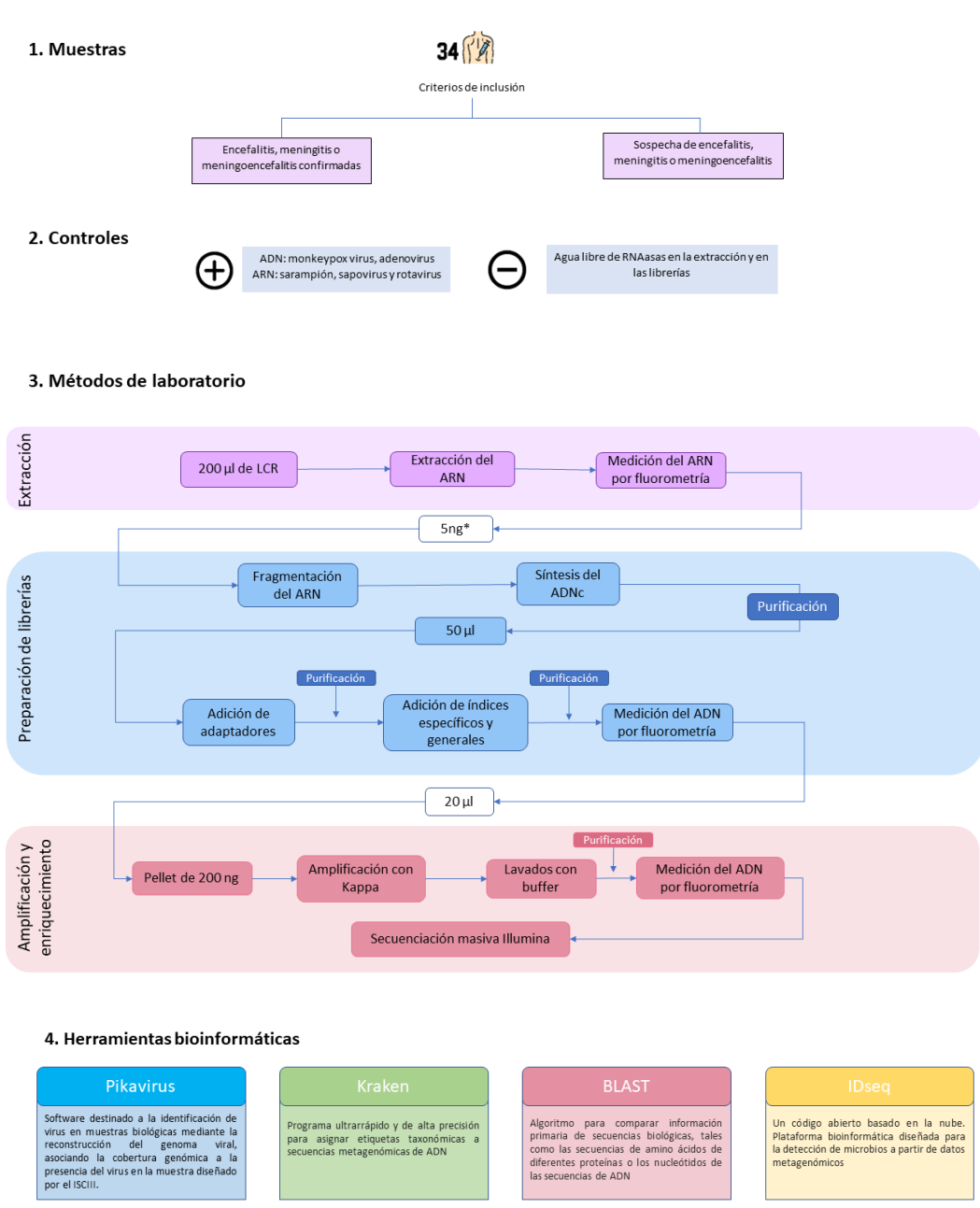


Figura 1: descripción general de los materiales y métodos utilizados, representado en un algoritmo. 1) Criterios de inclusión para la selección de muestras. 2) Descripción de los controles negativos y positivos añadidos al experimento. 3) Métodos de laboratorio. 4) Descripción de las herramientas bioinformáticas empleadas.

*En la mayoría de las muestras se pudo preparar las librerías con 5ng iniciales de ARN del extracto. Sin embargo, en algunas muestras, la concentración era tan baja que la cantidad de partida fue menor.

- a) Preparación de las librerías para la hibridación: una vez se han obtenido las concentraciones finales de ADN de cada librería, se preparan unos pools, en los cuales debe haber 200 ng de ADN. Cada pool constará de 4 a 6 librerías, dependiendo de las diferentes concentraciones finales de ADN de cada librería, y se calcula el volumen necesario de cada librería para añadir al pool. Para concentrar el ADN de cada pool, se centrifugan los tubos en un SpeedVac, el cual está a temperatura ambiente, durante 2 horas aproximadamente, hasta que el líquido haya desaparecido, de tal manera que en el fondo de cada tubo quedará un pellet con el ADN concentrado.
- b) Hibridación con sondas de captura: se prepara una solución mix, en el cual se incluye un mix de sondas, un panel y agua, y se introducen en el termociclador 2 minutos a 95°C, y rápidamente se enfrían en hielo durante 5 minutos. Cuando se ha evaporado el líquido de cada pool, se le añade a cada pellet obtenido unos bloqueadores y un buffer. Este pellet resuspendido se introduce en el termociclador a 95°C durante 5 minutos, y posteriormente se enfría a temperatura ambiente. A continuación, se añade la solución mix al pellet resuspendido, y a éste se le añade un potenciador de la hibridación. Esta mezcla estará 16 horas a 70°C en el termociclador.
- c) Unión de los pellet hibridados a las beads de Streptavidin: antes de comenzar este paso, se agitan con vortex los tubos que contienen las Streptavidin beads, y se añade el tampón de esas perlas magnéticas. Se colocarán los tubos en una gradilla magnética para realizar 3 purificaciones con el tampón de las perlas. Esto sirve para la purificar las perlas magnéticas antes de mezclarlas con el pellet. Después se añaden a cada tubo y se mezclan bien. Los tubos se colocan en una noria rotatoria durante 30 minutos a temperatura ambiente, y después, se realizan una serie de lavados en la gradilla magnética gracias a los buffer de lavado. Tras el último lavado, se añade agua y se mantiene en hielo. Se guarda la mitad de este contenido en el congelador a -20°C, y a la otra mitad se usará para continuar con el enriquecimiento.
- d) PCR posterior a la captura para amplificar, purificar y realizar un control de calidad: a la mitad del volumen que se utilizará para la amplificación, se le añaden unos primers y la enzima kapa (que se encuentra en un mix llamado KAPA HiFi HotStart ReadyMix). Se introducen los tubos con las mezclas en el termociclador. A continuación, se procede a realizar una serie de purificaciones con etanol al 80% preparado ese mismo día, junto con unas DNA Purification

beads (1X) y la gradilla magnética. Al final de estas purificaciones, se añade agua y se incuba a temperatura ambiente durante 2 minutos, y junto con la gradilla magnética, se recoge el sobrenadante, el cual se empleará para medir la concentración final de ADN de cada pool (mediante QuantiFluor dsDNA system de la marca Promega), para realizar el Bioanalyzer y para secuenciar.

6. **Bioanalyzer:** se trata de una técnica de determinación de la calidad y cantidad de fragmentos de material genético de una forma más rigurosa y fiable que los sistemas tradicionales mediante un sistema de microelectroforesis. La mayoría de fragmentos deben estar en torno a los 300 pares de bases (pb). Para ello, se ha requerido del kit Agilent High Sensitivity DNA kit de la marca Agilent Technologies. Estos análisis se llevaron a cabo en la Unidad de Genómica del Instituto de Salud Carlos III (ISCIII) mediante el programa 2100Xpert.
7. **Secuenciación masiva por Illumina:** la secuenciación masiva de las muestras se realizó por la Unidad de Genómica del ISCIII. Para la primera carrera se utilizó una plataforma de secuenciación Miseq System de Illumina (mediante un cartucho MiseqV2) y para la segunda carrera se utilizó un NextSeq System de Illumina (mediante un cartucho NextSeq 500), ya que en esta última había una mayor cantidad de muestras.
8. **Confirmación de los datos mediante PCR:** a partir de los resultados obtenidos en la secuenciación, se procede a confirmar cada muestra con los patógenos obtenidos mediante el análisis bioinformático. Los “hits” obtenidos se confirman por PCR. Se hizo una PCR convencional de citomegalovirus para aquellas muestras que presentaron un porcentaje de aparición de este virus, al igual que ocurrió con las muestras que parecían contener rotavirus²⁷, virus TOSV (general para flebovirus²⁸ y específica para TOSV²⁹), poliomavirus, betaherpesvirus humano 7 y virus de la rabia (lisavirus) (métodos in-house de las diferentes Unidades del CNM).
9. **Análisis bioinformático de los datos:** para el análisis de los datos se ha contado con la Unidad de Bioinformática del ISCIII (BU-ISCIII) (<https://github.com/BU-ISCIII>). A esta unidad se le solicita el análisis para el descubrimiento de presencia de patógenos virales. Esta solicitud consta de tres servicios: Pikavirus, BLAST y Kraken (Figura 1.4)
 - **Pikavirus** es una base de datos NextFlow disponible abiertamente en <https://github.com/BU-ISCIII/PikaVirus>, creada por el ISCIII. Este análisis consiste en los siguientes pasos: control de calidad de las muestras mediante FastQC (v 0.11.9) y MultiQC (v 1.9) para la evaluación de calidad, y FastP (v 0.20.1) para el recorte; identificación de genomas virales potenciales dentro de las lecturas recortadas

anteriormente usando MASH (v 2.3) contra la base de datos completa de ensamblajes virales de GenBank (v 248, descargada el 2 de abril de 2022); mapeo de genomas virales potencialmente presentes contra las muestras utilizando BowTie2 (v 2.4.4); y la obtención de estadísticas de coberturas usando SamTools (v 1.12) y BedTools (v 2.30.0).

- **BLAST:** es una herramienta básica de búsqueda de alineación local que encuentra regiones de similitud local entre secuencias. El programa compara secuencias de nucleótidos con bases de datos y calcula la importancia estadística de las coincidencias. En la Unidad de Bioinformática del ISCIII, se hace un BLAST de los contigs de ensamblaje de novo con la base de datos RefSeq de los genomas virales.
- **Kraken:** para ver el contenido de los distintos taxones que hay en cada muestra, se hace una clasificación de las lecturas en los distintos taxones por clusterización con el programa Kraken2, un sistema de clasificación taxonómica que utiliza coincidencias exactas de k-mers para lograr una alta precisión y velocidades rápidas de clasificación con datos probabilísticos. Posteriormente, se ha utilizado el programa KronaTools para realizar la representación de estos porcentajes de taxones en cada muestra.

Cuando se encuentra un posible “hit” en una muestra usando estas herramientas bioinformáticas, se solicita a BU-ISCIII un mapeo por referencia de esa muestra junto a un número de acceso del NCBI para reconstruir el genoma completo y formar un genoma consenso. Este número de acceso se obtiene mediante un BLAST que se realiza con las secuencias de novo analizadas por bioinformática después de la secuenciación masiva. En ese BLAST, se elige el número de acceso más próximo a la secuencia de la muestra de interés. En Resultados, cada mapeo que se ha realizado vendrá acompañado del número de acceso que se ha seleccionado para cada virus en cada muestra. Para el genotipado de la muestra de norovirus, se ha utilizado la herramienta <https://www.rivm.nl/mpf/typingtool/norovirus/>.

De manera complementaria, se ha utilizado la herramienta **Idseq** (<https://czid.org>), un servicio de código abierto basado en la nube para la detección y seguimiento de patógenos. Idseq permite un análisis robusto y reproducible de datos de mNGS para la identificación microbiana, independientemente de la muestra, especie o tipo de huésped (Figura 1.4). Además, permite eliminar el “background”, que son todos aquellos patógenos ambientales y contaminantes de reactivos o ambientales que pueden aparecer en los controles negativo. Este análisis consta de tres fases principales: 1) filtrado del huésped y control de calidad, 2) alineación basada en ensamblaje, y 3) informes y visualización mediante *heatmaps* (ver definición de términos

bioinformáticos en Anexo 1). Para nuestro estudio, se han usado unos umbrales con Z-score ≥ 50 , rPM ≥ 50 y $r \geq 50$ basándonos en los umbrales de otros estudios de mNGS^{30,31}. Otros estudios recomiendan filtrar los datos con Z-score ≥ 8 y $r \geq 10$, pero esto depende mucho del tipo de estudio que se haga, así como de los resultados que se obtengan en cuanto a virus se refiere³¹. Además, hemos definido como *hit* aquellos puntos en los que la Z-score es mayor a 80, la rPM es superior a 1.000 y la r supera las 10.000 lecturas.

10. Análisis filogenético: las secuencias obtenidas del TOSV se alinearon con otras secuencias representativas de diferentes genotipos del TOSV, tanto parciales como completas, disponibles en GenBank a fecha de mayo 2023 usando ClustalW del programa MEGA (versión 10). Aquellas secuencias parciales menores de 200 pb fueron excluidas. El árbol filogenético se construyó utilizando el software IQ-TREE v2.0.6, con el método *maximum-likelihood*. La reconstrucción se hizo con los parámetros establecidos por defecto y el mejor modelo fue establecido por ModelFinder seguido de un bootstrap de 1000 reensamblajes.

RESULTADOS

Resultados de metagenómica

Se analizaron muestras de LCR de 34 pacientes, con edades comprendidas entre los 18 y los 86 años. La mayoría de muestras provienen del Hospital Puerta de Hierro, Majadahonda (82,3% n=28). Las muestras del estudio se recogieron entre febrero y septiembre de 2022 (Tabla 1).

Se prepararon y secuenciaron un total de 52 librerías (34 muestras idiopáticas, 5 controles positivos y 13 controles negativos) mediante mNGS, y posteriormente se procedió a una amplificación y enriquecimiento por sondas. A continuación, se procedió a la secuenciación masiva de todas las librerías mediante Miseq (primera carrera) y NextSeq (segunda carrera) de Illumina en la Unidad de Genómica del ISCIII. Los archivos fastaq resultantes fueron procesados por la Unidad de Bioinformática mediante las herramientas bioinformáticas Pikavirus, Kraken y BLAST. Posteriormente, se hizo un análisis metagenómico mediante la plataforma Idseq, donde creó un *background* formado por las librerías pertenecientes a los controles negativos, de tal manera que se sustraen de las muestras aquellos *reads* que coinciden con los encontrados en los controles negativos.

Muestras utilizadas como controles positivos (**n=5**): estas muestras contenían patógenos previamente identificados y amplificados por PCR, como se observa en la Tabla 2. Todos estos controles positivos siguieron el mismo procedimiento de librerías y enriquecimiento que las muestras idiopáticas de LCR. El control 1 resultó positivo frente a rotavirus, con Z-Score de 99, rPM de 968.868 y un total de lecturas (*reads*, *r*) de 5.264.070 (Figura 2).

Muestra	Virus	Pikavirus	Kraken	BLAST	IDseq	PCR
LCRA120	Rubella virus	no	no	no	sí	volumen insuficiente
LCRA121	Betapapillomavirus	sí	sí (18%)	no	sí	volumen insuficiente
LCRA123	Betapapillomavirus	no	no	no	sí	volumen insuficiente
LCRA126	Gammapapillomavirus	no	no	no	sí	volumen insuficiente
LCRA127	Gammapapillomavirus	no	sí (12%)	no	sí	volumen insuficiente
LCRA128	VIH-1	no	no	no	sí	volumen insuficiente
	Betaherpesvirus humano 7	no	sí (9%)	no	sí	Negativo
	Betapapillomavirus	no	sí (3%)	no	sí	volumen insuficiente
LCRA129	Betapapillomavirus	no	sí (20%)	no	sí	volumen insuficiente
LCRA145	Toscana virus	no	sí (48%)	sí (100%)	sí	Positivo
LCRA149	Parvovirus B19	no	sí (7%)	no	sí	volumen insuficiente
LCRA155	Lisavirus de murciélago europeo	no	sí (8%)	no	sí	Negativo
LCRA156	Gammapapillomavirus	sí	no	no	sí	volumen insuficiente
LCRA157	Alphapapillomavirus	no	sí (22%)	no	sí	volumen insuficiente
LCRA158	Poliomavirus humano 6	no	sí (6%)	no	sí	volumen insuficiente
LCRA170	Lisavirus de murciélago europeo	no	sí (7%)	no	sí	Negativo
LCRA183	Adenovirus	no	sí (28%)	no	sí	volumen insuficiente
	Astrovirus	no	no	no	sí	volumen insuficiente
LCRA184	Virus Epstein-Barr	no	sí (7%)	no	sí	volumen insuficiente
	Astrovirus	no	sí (7%)	no	sí	volumen insuficiente
LCRA185	Adenovirus	no	sí (36%)	no	sí	volumen insuficiente
	Norovirus	sí	sí (32%)	sí (7%)	sí	volumen insuficiente
Control + 1	Rotavirus A	sí	sí (100%)	sí (90%)	sí	Confirmada previamente
Control + 2	Sapovirus	no	sí (5%)	sí (94%)	sí	Confirmada previamente
Control + 3	Sarampión	sí	sí (100%)	sí (99%)	sí	Confirmada previamente
Control + 4	Monkeypox virus	sí	sí (67%)	sí (5%)	sí	Confirmada previamente
Control + 5	Adenovirus	sí	sí (765)	sí (99%)	sí	Confirmada previamente

Tabla 3: comparación de las diferentes herramientas bioinformáticas utilizadas para las muestras idiopáticas en las que se ha encontrado un agente vírico, y si se ha realizado en ellas PCR confirmatoria.

En cuanto al control positivo 2, se confirmó la detección diferentes virus: adenovirus, EV A, salivirus y sapovirus. En este caso, los resultados fueron los siguientes: adenovirus (Z-Score 99, rPM 2.761 y r 26.935), EV A (Z-Score 99, rPM 108.846 y r 1.026.368), salivirus (Z-Score 100, rPM 154.247 y r 1.454.473) y para sapovirus (Z-Score 100, rPM 225.926 y r 2.130.372). En el control positivo 3, se confirmó la presencia del virus del sarampión con un Z-score 100, rPM 641.763 y r 5.980.973. En el control positivo 4, se confirmó la detección de Monkeypox virus con un Z-Score

99, rPM 623.846 y r 4.227.904 (Figura3). En el control positivo 5 también se confirmó la presencia de varios virus: adenovirus, EV A y norovirus. Para todos estos virus, sus datos analizados en Idseq fueron los siguientes: adenovirus (Z-Score 99, rPM 4.929 y r 121.552), EV A (Z-Score 99, rPM 5.491 y r 1.355.136) y norovirus (Z-Score 99, rPM 14.242 y r 351.236) (Figura 3) (ver definición de términos bioinformáticos en Anexo 2). Todos estos controles positivos se identificaron también mediante las cuatro herramientas bioinformáticas, excepto el control + 2 (sapovirus) en el Pikavirus. Además, podemos destacar que el Monkeypox virus en BLAST sólo sale en un 5%, cuando en el resto de bases de datos se confirma con un mayor porcentaje (Tabla 3).

Muestras idiopáticas de LCR (**n=34**): se identificaron 10 virus en 8 muestras mediante las diferentes herramientas bioinformáticas ya citadas (Tabla 3). Estos virus son: virus de la rubéola, VIH-1, HHV-7, EBV, astrovirus, TOSV, lisavirus europeo de murciélago, adenovirus, norovirus y parvovirus B19.

- 1) Virus de la rubéola:** se trata de la muestra idiopática LCRA120, la cual pertenece a un adulto de 48 años con neuropatía. Cuando esta muestra llegó al CNM, se le hicieron diferentes PCRs de rutina, entre las que se incluyen todos los HV y los EV, las cuales resultaron negativas (ver Tabla 1). Como se observa en la Tabla 3, este virus fue únicamente detectado mediante IDseq. Los datos analizados mediante esta plataforma fueron los siguientes: Z-Score 100, rPM 65,3 y r 102, y no se pudo obtener una secuencia consenso (Figura 2). La detección de este virus no se pudo confirmar mediante PCR ya que había un volumen insuficiente de muestra. Sin embargo, se pidió a BU-ISCI un mapeo por referencia (nº de acceso NC_001545), para el cual se obtuvieron unos valores de 1.108.778 *reads* totales, de los cuales solamente 90 pertenecían al virus de la rubeola (0,01%), con una cobertura de >10x de 1%. Para este virus sí que se pudo realizar una secuencia consenso, aunque no se pudo comprobar la presencia de este virus en esta muestra debido a que no había volumen suficiente de LCR.
- 2) Virus de la inmunodeficiencia humana tipo 1 (VIH-1):** este virus se ha encontrado en la muestra llamada LCRA128, la cual pertenece a una persona de 86 años que acudió al hospital debido a fiebres recurrentes y una sospecha de infección neurológica. Este virus solamente fue detectado mediante IDseq (Tabla 3), y no se pudo confirmar su presencia mediante PCR por volumen insuficiente de muestra. Los valores del programa IDseq para este virus en la muestra LCRA128 fueron: Z-Score 100, rPM 792,8 y r 172, sin secuencia consenso (Figura 2). Es por ello que se procedió a realizar un mapeo por referencia (nº de acceso AF033819), en el cual se obtuvieron 171.030 *reads* totales, con un porcentaje

de 0,1% que se corresponden con secuencias del VIH-1 (en total 164 reads). Además, se obtuvo una cobertura >10x del 2%. Para este virus sí se consiguió una secuencia consenso, pero no se pudo confirmar mediante PCR a falta de volumen suficiente de muestra.

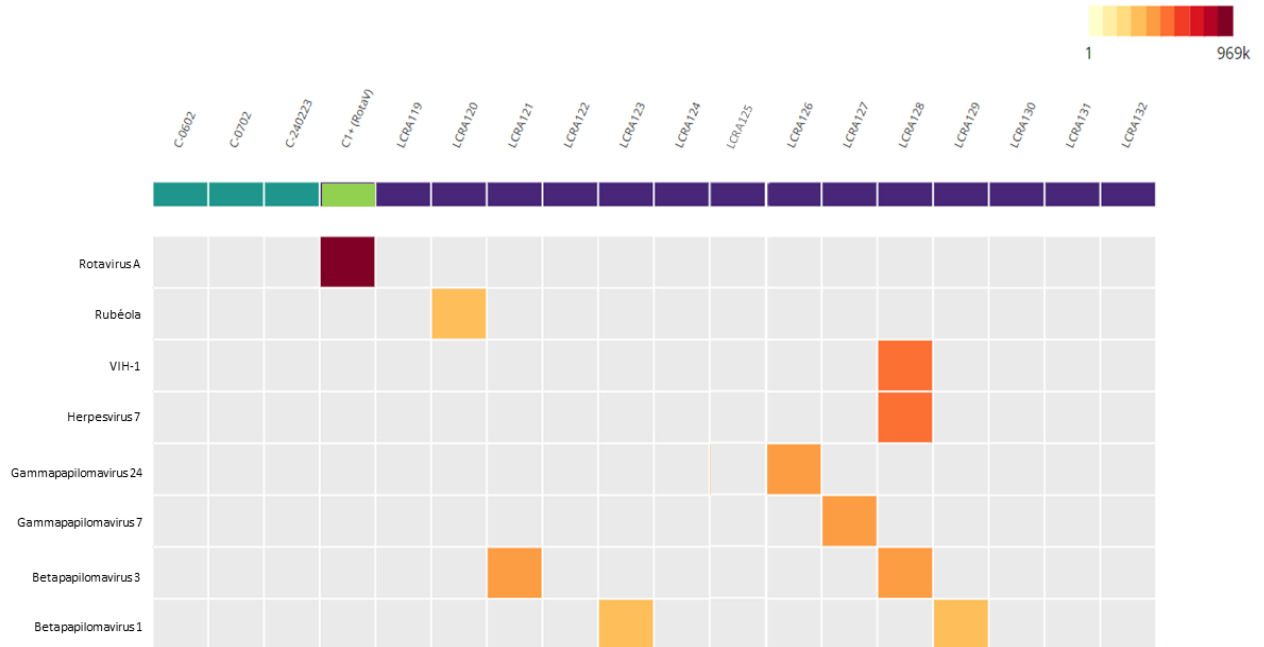


Figura 2: heatmap del rPM de los virus encontrados en las muestras idiopáticas, controles positivos y negativos de la primera carrera (secuenciada mediante Miseq) en forma de *heatmap* por IDseq. Filtros: rPM \geq 50, Z-score \geq 50 y $r \geq$ 50. En la parte superior, el color turquesa indica que es un control negativo, el color verde es el control positivo y el color azul oscuro indica que es una muestra idiopática

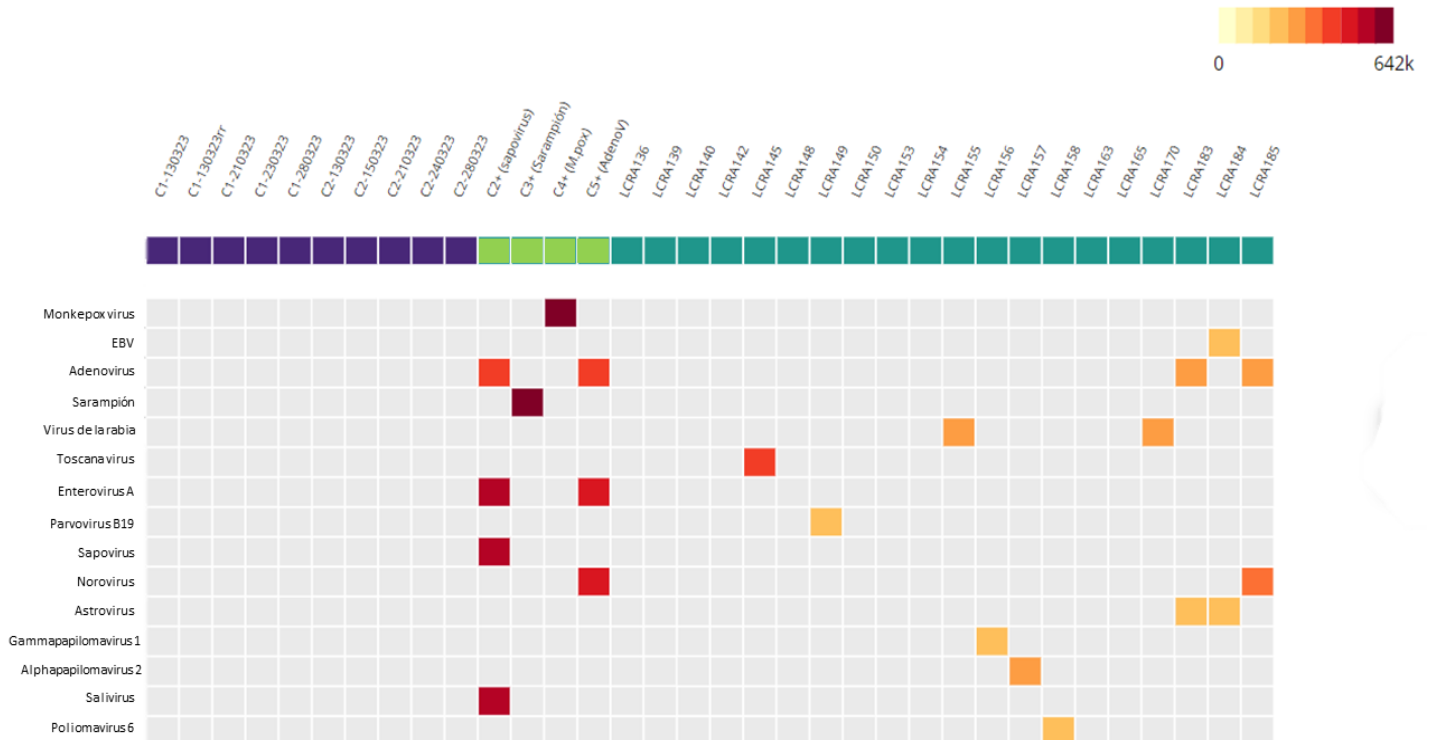


Figura 3: Heatmap del rPM de los virus encontrados en las muestras idiopáticas, controles positivos y negativos de la segunda carrera (secuenciada mediante NextSeq). En la parte superior, el color azul oscuro indica que es un control negativo, el color verde indica los controles positivos y el color turquesa indica las muestras idiopáticas.

- 3) Herpesvirus humano 7 (HHV-7):** este virus se ha encontrado en la misma muestra que el virus anterior, es decir, la muestra LCRA128. Cuando la muestra llegó al CNM, únicamente se le hizo PCR de rutina de EV, HVS y VZV (ver Tabla 1). Sin embargo, este virus sí fue identificado mediante Kraken, resultando ser un 9% de las secuencias totales virales de la muestra (Tabla 3). También se detectó mediante IDseq, con unos valores de Z-Score 100, rPM 949,5 y r 206 (Figura 2). Sin embargo, en esta plataforma no se pudo establecer una secuencia consenso. Para este virus, sí se pudo hacer PCR, la cual resultó negativa para el HHV-7. Esta PCR era genérica para herpesvirus.
- 4) Astrovirus:** este virus se ha encontrado en dos muestras diferentes: LCRA183 (paciente de 74 años con encefalitis) y LCRA 184 (joven de 18 años con meningitis linfocitaria, parálisis periférica parcial y picadura de garrapata). El astrovirus es detectado para la muestra LCRA183 únicamente en IDseq (Tabla 3), con unos valores de Z-Score 100, rPM 56,3 y r 192, sin poder formar una secuencia consenso (Figura 3). Para la muestra LCRA184, además de detectarse mediante IDseq, pudo ser identificado el astrovirus por Kraken, con un porcentaje del 7% (Tabla 3). Los valores de IDseq para esta muestra fueron Z-score 100, rPM 99,4 y r 80 (Figura 3), aunque sin secuencia consenso. Sin embargo, para esta muestra sí se realizó un mapeo por referencia (nº de acceso AY720892): se obtuvieron 334.314 *reads*, de los cuales un 0,03% pertenecían a astrovirus (102 *reads*). Se obtuvo una cobertura >10x del 3%. De estas muestras no había un volumen suficiente como para confirmar la presencia de astrovirus por PCR.
- 5) Virus de Epstein-Barr (EBV):** este virus se ha identificado en la muestra LCRA184 (paciente descrito en apartado de Astrovirus con PCR de rutina de EV, HVS y VZV. Ver Tabla 1) mediante Kraken (7%) (Tabla 3) y IDseq (Z-score 100, 77,1 rPM y 62 r, no secuencia consenso) (Figura 3). Se solicitó un mapeo por referencia (nº de acceso LC573555), del cual se obtuvieron 334.314 *reads*, de los cuales 55 pertenecían al BV (0,02%). Sin embargo, no se obtuvo cobertura >10x, por lo que no se pudo realizar una secuencia consenso. Tampoco se pudo comprobar la presencia del EBV en esta muestra debido al volumen insuficiente.
- 6) Virus Toscana (TOSV):** la muestra en la que se detectó este virus fue en LCRA145. Esta muestra pertenece a una persona de 61 años con cefalea y sospecha de infección neurológica. Para este LCR se realizaron PCRs para todos los HVs y EVs (Tabla 1). Este virus es identificado mediante todas las herramientas bioinformáticas utilizadas excepto en Pikavirus. En Kraken se obtiene un porcentaje de 48%, en BLAST es identificado con un 100% (Tabla 3), y en IDseq se obtienen unos valores de Z-Score 100, rPM 5.564,5 y r 23.798 (Figura 3). Además, en esta última plataforma se pudo obtener una secuencia

consenso para los tres segmentos diferentes que contiene este virus en su genoma: segmento S, segmento M y segmento L. La longitud de referencia obtenida para el segmento L fue de 6.360 pb, con una profundidad de 145,5 y una cobertura del 100%. Para el segmento M se obtuvo una longitud de 4.203 pb, una profundidad de 214,2 y una cobertura de 94,7%. El segmento S es el más pequeño, con una longitud de 1.871 pb y una profundidad de 870. Además, se obtuvo una cobertura de 99,7%. Las gráficas de las secuencias consenso se encuentran en el Anexo 2, Figura S1 A-C. Para esta muestra quedaba suficiente volumen residual para poder realizar una PCR de confirmación del TOSV. En primer lugar, se realizó una PCR genérica para flebovirus (Figura 4A), que salió positiva. En esta PCR se puso como control negativo una muestra positiva para otro flebovirus. La secuenciación Sanger de la banda amplificada generó una secuencia que por análisis BLAST tenía una similitud del 100% con la obtenida por NGS. El alineamiento y comparación de ambas secuencias reveló que eran iguales. En segundo lugar, se hizo una RT-PCR específica para el TOSV, en la cual se obtuvo un resultado Ct de 35 (Figura 4B). Además, se solicitó un mapeo por referencia de esta muestra, del cual se obtuvieron 3.050.178 *reads* totales. De ellos, 21.858 pertenecían a TOSV, el cual supone un 0,72% de los *reads*. Se obtuvo una cobertura >10x del 100% para el segmento S (nº de acceso KX010932), para el segmento L (nº de acceso FJ153280) y para el segmento M (nº de acceso KU204982).

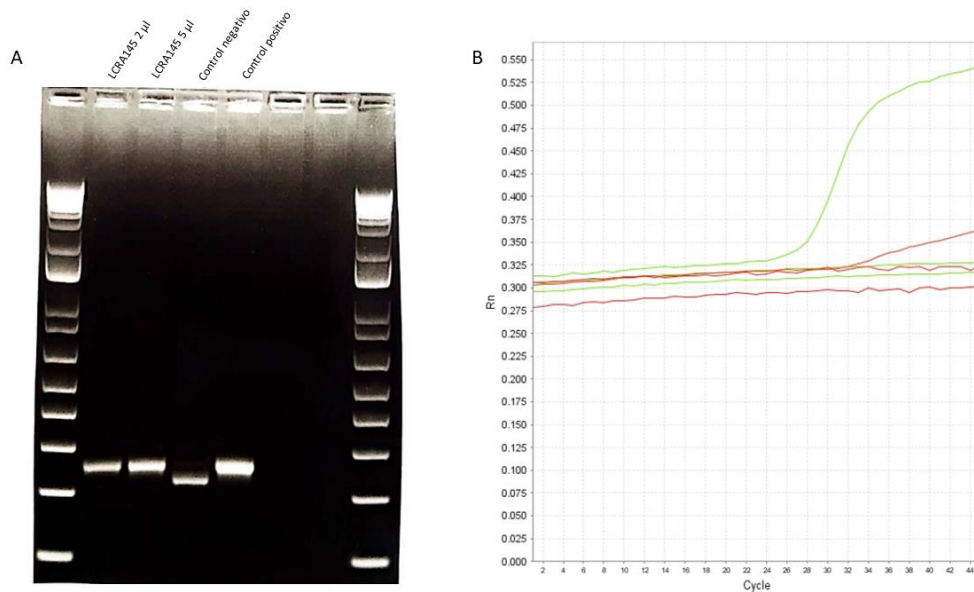


Figura 4: A) PCR genérica para flebovirus. B) RT-PCR para TOSV: muestra LCRA145 (línea roja), control positivo TOSV (línea verde), control negativo (línea rosada inferior). El resto de líneas son otras muestras de la Unidad de arbovirus del CNM que se metieron en la PCR

Se realizó el análisis filogenético de los tres segmentos (S, M y L) obtenidos de la muestra LCRA145 en combinación con secuencias disponibles en GenBank de los 3 genotipos (A, B y C). La filogenia estimada sitúa las secuencias de los tres segmentos en el genotipo B (Figura 6). Este genotipo comprende secuencias del TOSV mayoritariamente de muestras recogidas en España (1998-2008), Portugal (1983), Francia (2004-2005) y Turquía (2013-2018). En los tres segmentos, S, M y L, la mayor homología se obtiene con secuencias del TOSV (98,5%, 97,5% and 97,5% respectivamente) obtenidas de flebotomos en un estudio de un foco de leishmaniosis en la Comunidad de Madrid en el año 2012 (números de acceso LN848251.1, LN848246.1 y LN84240.1).

7) Lisavirus europeo de murciélago (EBLV): este virus (relacionado con el virus de la rabia) se identificó en dos muestras idiopáticas: LCRA155 (24 años) y LCRA170 (23 años), ambos con sospecha de encefalitis. Para ambas, se obtuvo un porcentaje en Kraken, siendo un 8% para LCRA155 y un 7% para LCRA170 (Tabla 3). Las dos muestras se analizaron mediante IDseq, las cuales tuvieron unos datos de Z-score 100. Para LCRA155, el rPM fue 140,8 y r 1.286 r, y en la muestra LCRA170 se obtiene un rPM de 152,7 y 1.020 r (Figura 3). Para ninguna de las dos se obtuvo una secuencia consenso mediante este programa. Se pidió a BU-ISCI un mapeo por referencia, y para ambos se usó el nº de acceso MF187880. La muestra LCRA155 obtuvo 6.109.040 *reads* totales, de los cuales un 0,02% pertenecían a lisavirus (1.170 *reads*), con una cobertura >10x de 2%. Para la muestra LCRA170 se obtuvieron 4.021.148 *reads*, donde solo 940 pertenecían a lisavirus (0,02%), con una cobertura >10x de 3%. Las secuencias obtenidas de aproximadamente 200pb mostraron un 100% de homología en el análisis BLAST con secuencias del EBLV de tipo 1. En ambas muestras sí había un volumen suficiente para confirmar la presencia del virus mediante PCR, pero el resultado fue negativo.

8) Adenovirus: se encontró adenovirus en dos muestras: LCRA183 (al igual que para astrovirus) y LCRA185 (paciente de 26 años con meningoencefalitis aguda). Ambos fueron detectados por Kraken (28% en LCRA183 y 36% en LCRA185) y por IDseq (Tabla 3). El Z-score de ambas muestras para adenovirus fue el mismo: 99. Sin embargo, los rPM para la primera fue 444,5 (1.516 *reads*) y para la segunda 349,4 (2.512 *reads*) (Figura 3). Para la muestra LCRA185 sí se obtuvo una secuencia consenso por IDseq (ver Anexo 2, Figura S1 D). La muestra LCRA183 se obtuvo con una longitud de 35.937, con una profundidad de 10,1 y una cobertura del 1,7%. Sin embargo, para la muestra LCRA185 se obtuvo una longitud de 33.802, con profundidad de 225,1 y una cobertura de 28,7%. Únicamente para esta última muestra se hizo el mapeo por referencia (nº de acceso

OM112281), donde se consiguieron mapear 3.318.118 *reads* totales, de los cuales 42.224 pertenecían a adenovirus (1,27%), con una cobertura >10x de 24%, junto con una secuencia consenso.

9) Norovirus: este virus se ha identificado en la muestra LCRA185 (previamente descrita) mediante todas las herramientas bioinformáticas empleadas (Tabla 3): Pikavirus, Kraken (32%), BLAST (7%) y IDseq (Z-score 81,6, rPM 1.613 y 11.596 r, Figura 3). Cuando se realizó el BLAST, se observó que esta muestra contenía un norovirus del genogrupo GII. De IDseq que obtuvo una secuencia consenso, con una longitud de 7.536 pb, profundidad de 234,7 y cobertura de 45% (ver Anexo 2, Figura S1 E). Se hizo un mapeo por referencia (nº acceso MG746026) y se obtuvieron 3.318.118 *reads* totales, donde un 0,32% pertenecían a norovirus (10.470 *reads*), con cobertura >10x de 43%, además de poder realizar una secuencia consenso. En el genotipado de las secuencias obtenidas de norovirus en esta muestra, se observó que se trataba de un norovirus GII (Figura 5). La polimerasa pertenece a GII.P34 y la cápside a GII.2. Sin embargo, no fue posible su confirmación mediante PCR a falta de volumen suficiente de LCR.


Name	Length	Family Genus Genogroup	BLAST score	polymerase	capsid	genome
AB089882 Norovirus OC97007	1045	Caliciviridae Norovirus GII	83.47826	GII.P34 (GII.Ph)	GII.2	

Figura 5: resultado de la secuencia de norovirus obtenida mediante mapeo por referencia de la muestra idiopática LCRA185 mediante www.rivm.nl/mpf/typingtool/norovirus/.

10) Parvovirus B19: únicamente se ha detectado este virus en la muestra LCRA149 (24 años con sospecha de encefalitis) mediante Kraken (7%) y IDseq (Z-score 100, rPM 71,5 y 260 r, sin secuencia consenso) (Tabla 3, Figura 3). Se hizo un mapeo por referencia (nº de acceso NC_000883) y no hubo coincidencias. No se hizo PCR a falta de volumen suficiente de muestra.

Otros virus encontrados en muestras idiopáticas: se han descrito principalmente mediante Kraken y IDseq una serie de virus que aparecen en las muestras idiopáticas en menor cantidad (Tabla 3). Estos virus pertenecen a la familia de los papilomavirus y poliomavirus. Destacan los betapapilomavirus y los gammapapilomavirus, con un Z-score 80, rPM de 177 y 80 r de media, y

en ninguno se obtuvo una secuencia consenso (Figuras 2 y 3). En cuanto al poliomavirus 6, se encuentra en la muestra LCRA158 (paciente de 62 años con encefalitis) identificado por Kraken (6%) y IDseq (Z-score 100, 60 rPM y 980 r, sin secuencia consenso) (Tabla 3, Figura 3). Para esta muestra se hizo un mapeo por referencia para los virus BK y JC ya que estos son los poliomavirus para los que se han descrito más casos de encefalitis mientras que el poliomavirus 6 generalmente se presenta como afección cutánea (nº de acceso MT448660 para poliomavirus JC y OM938940 para poliomavirus BK), donde se obtuvieron 11.991.552 *reads*, y no hubo ninguna coincidencia con los poliomavirus. Tampoco se realizó PCR de confirmación por volumen insuficiente.

Controles negativos: en 9 de 13 controles negativos que se utilizaron en las dos carreras se secuenciaron de virus. Algunas familias de virus que se han identificado con mayor número de *reads* son diferentes papilomavirus y poliomavirus, citomegalovirus, alfaherpesvirus 2, retrovirus endógeno porcino, Kadipiro virus, Shamonda virus, circovirus, retrovirus, alfacoronavirus 229E y múltiples fagos, entre los que destaca Vibrio virus CTXphi.

Análisis filogenético

Las secuencias obtenidas del TOSV en la muestra LCRA145 se analizaron junto con secuencias representativas de TOSV totales o parciales de GenBank. Como observamos en la Figura 6, la muestra más cercana al TOSV detectado en nuestro estudio es la muestra Fue-Sp45 (obtenida en un foco de leishmaniosis que se estudió en la Comunidad de Madrid en los años 2012-2013³²). Ambas pertenecen al genotipo B en los tres segmentos (Figura 6).



Figura 6: árbol filogenético de los diferentes segmentos del virus Toscana en comparación con otras secuencias representativas de diferentes genotipos de dicho virus, tanto parciales como completas, disponibles en GenBank a fecha de mayo 2023, y con una secuencia del brote de la Comunidad de Madrid³³. A) Árbol filogenético con respecto al segmento S (pequeño). B) Árbol filogenético con respecto al segmento M (mediano). C) Árbol filogenético con respecto al segmento L (grande).

DISCUSIÓN

Las técnicas metagenómicas permiten de identificar y estudiar de forma eficaz el origen microbiológico de encefalitis y meningitis, facilitando así el diagnóstico de aquellos pacientes que sufren este tipo de enfermedades neurológicas. En este trabajo, se ha empleado la mNGS como método de detección de virus en muestras de LCR de pacientes con encefalitis o meningitis sospechosas, enviadas por diferentes hospitales de la Comunidad de Madrid al CNM durante el año 2022, y que resultaron negativas en las PCRs de rutina para los principales virus causantes de infección neurológica (herpesvirus y enterovirus).

En primer lugar, hay que resaltar la aparición de TOSV del genotipo B en la muestra LCRA145, ya que es un virus poco común en la zona de estudio. Se trata de un virus ARN con genoma segmentado en tres fragmentos: L (grande), M (mediano) y S (pequeño), que codifican para una polimerasa dependiente de ARN, glicoproteínas de la envoltura G1 y G2, y para la nucleocápside (N), respectivamente, junto con proteínas no estructurales (NS)^{32,33}. Existen 3 genotipos, los cuales se distribuyen por Europa de una forma heterogénea. El genotipo A se asocia a regiones de Italia (de donde se cree que proviene el virus), el genotipo B se encuentra fundamentalmente en España, Portugal y Francia, y el genotipo C se ha descrito en otros países europeos, fundamentalmente Croacia, Chipre o Grecia³³⁻³⁵. El TOSV es responsable de más de un 50% de los casos de encefalitis y meningitis³, sobre todo en épocas estivales (esta muestra fue tomada en junio de 2022), ya que es el periodo en el cual hay un aumento en la actividad del vector, el flebótomo *Phlebotomus perniciosus* y otras especies del género, y con una seroprevalencia en una población de hasta un 25% en el sur de España³⁵. Entre un 40-60% de estos casos de neuropatías no se diagnostican, ya que este virus no se encuentra entre los métodos de detección de virus de rutina en los centros sanitarios o en los paneles neurológicos usados por los hospitales, lo que dificulta su control y vigilancia³⁶. En España, lo más común es encontrar este virus en las regiones del sur del país, especialmente en Andalucía^{32,35}, donde se notifican casos desde 1988. Sin embargo, rara vez se ha descrito la detección de este virus en el interior de la Península Ibérica en casos de meningitis a pesar de los datos de seroprevalencia. Concretamente, en la Comunidad de Madrid existe una seroprevalencia de un 5,7% de acuerdo con los artículos publicados^{37,38}. La mayoría de casos de TOSV en esta región sucedieron en 2012-2013, cuando sucedió el brote de leishmaniosis en la ciudad de Fuenlabrada³². Esto sucedió ya que ambos patógenos comparten vector: los flebótomos. Además, existía un reservorio ideal para los dos microorganismos, que eran las liebres y conejos, por lo que se dio un entorno idóneo para la expansión de TOSV y *Leishmania infantum*³⁸. Como observamos en la Figura 6, la secuencia de la muestra LCRA145 es muy similar (superior al 97% en los tres segmentos) a las

secuencias de TOSV obtenidas en dicho foco de leishmaniosis entre 2012-2013, siendo ambas del genotipo B, característico de todas las muestras de TOSV encontradas en España hasta la fecha. La alta homología del TOSV de nuestro estudio (C. de Madrid, 2022) con las del TOSV detectado en la misma Comunidad una década antes (2012) sugiere una alta estabilidad genética del TOSV en nuestro país. Hasta la fecha, solamente se han descrito cuatro casos de meningitis causadas por TOSV en Madrid en 2003³⁷, que puede deberse a los síntomas que este virus produce, ya que, al tratarse de cefaleas, fiebre y otros signos comunes a diferentes patologías, los clínicos no sospechan de la infección por TOSV. Sin embargo, existe un importante problema en cuanto al diagnóstico de TOSV en encefalitis y meningitis, y es que el tipo de muestra escogida para estas enfermedades es el LCR, donde este virus se encuentra en muy baja concentración, lo que dificulta su detección. Es por ello que la metagenómica es una buena opción de diagnóstico de este virus, ya que se puede detectar el genoma independientemente de la cantidad que haya de éste en la muestra de LCR.

En segundo lugar, se ha observado la presencia de norovirus en la muestra LCRA185 en coinfección con adenovirus. Los norovirus son virus pequeños con ARN pertenecientes a la familia *Caliciviridae*³⁹. Normalmente causan infecciones gastrointestinales que pueden complicarse hacia el SNC, causando encefalitis o encefalopatías, siendo el genogrupo GII el que más frecuentemente se detecta en estos casos. Se han visto casos de niños con convulsiones y encefalitis derivadas de infecciones gastrointestinales^{40,41}. Sin embargo, las características clínicas de estos desordenes neurológicos siguen siendo desconocidas a día de hoy, aunque se suele tener mal pronóstico⁴⁰. Además, en estos casos no se ha podido aislar ni hay PCR positivas para norovirus en LCR, tanto en los casos de los niños como en el de un adulto⁴². Debido a las características biológicas de este virus, clínicos e investigadores no descartan la posibilidad de desarrollar encefalitis o meningitis a partir de una infección extracerebral por norovirus⁴². En España se han reportado dos casos de encefalitis originada por norovirus detectado en LCR y heces. En este caso, las muestras se genotiparon y caracterizaron, y todas pertenecían al GII.4⁴¹ (Figura 5). Es por ello que, debido a las evidencias científicas y nuestros resultados, que concuerdan con el genogrupo más detectado GII, consideramos este virus como un hit importante en la muestra LCRA185.

En cuanto al EBLV es un virus ARN relacionado con el virus de la rabia, endémico en murciélagos europeos y capaz de causar rabia en humanos. Este virus según nuestros resultados, aparece en dos muestras diferentes, LCRA155 y LCRA170, ambas de pacientes con encefalitis. La presencia del EBLV en LCR de pacientes con sospecha de encefalitis ya se ha visto en otros casos humanos^{43,44}. Causa una encefalomielitis o una encefalitis fatal, donde, en el mejor de los casos,

la supervivencia es del 41%⁴⁵. Este tipo de rabia encefalítica se presenta en el 80% de los pacientes, mientras que el 20% restante presenta una rabia parálítica. Normalmente, los virus se expanden por el tejido cerebral, incluyendo el sistema límbico y la médula espinal, provocando múltiples y variables síntomas que dependen de muchos factores^{44,45}. A pesar de que el LCR es la muestra usada en la rutina clínica para el diagnóstico de las encefalitis, esta muestra se asocia a un bajo valor diagnóstico siendo el tejido cerebral es la muestra más informativa en las encefalitis a pesar del acceso difícil a esta muestra. Es por ello que, en estas muestras del estudio, a pesar de una posible presencia del EBLV en el LCR, no se habrían confirmado por PCR⁴⁵.

Otros virus que podrían causar encefalitis, meningitis o cualquier otro desorden neurológico son los adenovirus, el virus de la rubéola y el EBV. Los adenovirus se han reconocido como un potencial patógeno viral, quien provoca gran morbilidad y mortalidad. Normalmente, los pacientes más afectados son aquellos que están inmunocomprometidos. Sin embargo, se ha observado manifestaciones neurológicas, que incluyen meningoencefalitis o encefalitis, en personas inmunocompetentes, sobre todo en niños. Además, se ha detectado en muestras de LCR mediante PCR convencional^{46,47}. Por lo tanto, su coexistencia con otros virus en LCR, como en el caso de nuestras muestras LCRA183 y LCRA185 no ha de extrañar, ya que es un virus que se puede encontrar en este tipo de muestras, y que, además, causen encefalitis. En el caso del virus de la rubeola, los casos de enfermedades neurológicas son menos frecuentes⁴⁸. Se trata de un virus que normalmente afecta a niños no vacunados entre los 5 y 14 años. Sin embargo, puede haber una infección en los adultos con síntomas más severos. Las complicaciones neurológicas son raras, ya que se reporta 1 caso entre 24.000. Normalmente la encefalitis se ocasiona entre 1 y 6 días después de aparecer el exantema, pero pocos estudios reportan la vía por la cual este virus infecta el SNC^{48,49}. Es por ello que no podemos descartar la presencia del virus de la rubeola en la muestra LCRA120, ya que, a pesar de no poder confirmar este virus por PCR, se pudo hacer un mapeo por referencia y una secuencia consenso desde BU-ISCI. En cuanto al EBV, pertenece a la familia de los HV. Este, en concreto, se trata del HV 4, un virus ubicuo que se transmite sobre todo en la infancia y adolescencia en forma de mononucleosis (el 50% de los casos)⁵⁰. La mayoría de los casos son asintomáticos, pero este virus se acantona para toda la vida en el interior de los linfocitos. La prevalencia del EBV en adultos está por encima del 90% de la población mundial. Una reactivación de este virus es bastante inusual; sin embargo, se han dado casos de movilización de estos virus latentes, ocasionando encefalitis, con mayor riesgo en inmunosuprimidos. De hecho, se ha encontrado este virus en muestras de LCR, confirmado por PCR^{50,51}. Por lo tanto, nuestra muestra LCRA184 es susceptible de presentar EBV. Por otro lado, los astrovirus también se han reportado como causantes de meningitis, y su presencia en LCR de

adultos inmunocompetentes no es inusual⁵², lo mismo que ocurre con los parvovirus, especialmente con el parvovirus B19, el cual provoca de manera ocasional encefalitis agudas⁵³. También se ha detectado mediante mNGS la presencia de ARN de VIH y EV-71 en LCR de personas con problemas neurológicos^{54,55}. Con esta misma técnica también se ha visto la existencia de HHV-7 en LCR de adultos inmunocompetentes que padecen encefalitis⁵⁶.

A pesar de las ventajas que supone el empleo de la metagenómica en el diagnóstico de enfermedades como encefalitis o meningitis, hay una serie de inconvenientes y limitaciones a tener en cuenta. En cuanto a los inconvenientes, sigue siendo un método costoso y lento, que requiere un procedimiento complejo de preparación de muestras, equipos caros y menor capacitación en bioinformática con personal de laboratorio, por lo que no es adecuado para grandes estudios de detección de virus²⁶. Además, la media de duración del ensayo completo en el laboratorio consta de 3-4 días, sin incluir los días en los que se realiza el análisis bioinformático (son aproximadamente 10 días en total)²². Otra dificultad es que, debido a que se secuencian todo lo que hay en la muestra, la mayor parte de los *reads* pertenecen al hospedador, lo que dificulta que la sonda se una a las secuencias de los patógenos problema²².

En cuanto a las limitaciones del estudio caben señalar varias. Una limitación sería la asociada al uso de muestras residuales que haría que la calidad de los ácidos nucleicos pudiera deteriorarse con el paso del tiempo (sobre todo para los virus ARN) en caso de no estar las muestras almacenadas a -80°C. Para este estudio se han usado muestras residuales almacenadas en la seroteca del CNM a -20°C lo cual podría haber hecho que la calidad de los ácidos nucleicos no fuera la óptima. Para vencer esta limitación se han usado para el estudio solo muestras recientes del año 2022. Otra limitación asociada al uso de muestras residuales para el estudio habría sido el bias en la selección de las muestras. Para el diagnóstico de las infecciones neurológicas agudas normalmente se usa el LCR. Sin embargo, dependiendo del virus, su presencia también podría detectarse en otras muestras como suero, heces o exudado faríngeo. En algunos casos, es poco común encontrar virus en el LCR de casos de encefalitis. Varios estudios han encontrado que los EV como EV-71 y EV-D68 que causan enfermedades neurológicas se detectaron con mayor frecuencia en las heces, hisopos rectales o exudados faríngeos que en el LCR^{19,54}. Por lo tanto, un LCR negativo o muestra de sangre no descarta una infección viral del SNC. Para vencer esta limitación se propone hacer en el futuro un estudio prospectivo con un amplio número de sujetos de estudio y en el que se recogieran diferentes tipos de muestras.

Otra limitación que podría afectar a la especificidad sería la contaminación con agentes microbianos de flora normal y reactivos lo cual es algo común en la mNGS. Según otros estudios,

el LCR no tiene un viroma constante en pacientes sanos, por lo que es un dato a tener en cuenta a la hora de hacer metagenómica, además de tratarse de un fluido corporal teóricamente estéril, y del cual se cuenta en general con muy poco volumen para hacer PCRs confirmatoria. Sin embargo, la mayoría de microorganismos encontrados de forma común en LCR de población sana resultan ser bacteriófagos⁵⁷. Siguiendo en esta línea, como se secuencian todo lo que hay en la muestra, las contaminaciones son visibles. Esto se ha podido ver en nuestros controles negativos, con la aparición de múltiples secuencias víricas en el agua libre de RNAasas. Algunos de estos virus aparecen contaminando el ambiente, otros materiales (como tubos de laboratorio, pipetas, columnas, etc.) y los reactivos de los diferentes kits pertenecen a las familias *Circoviridae*, *Tombusviridae*, *Totiviridae* y *Retroviridae*. Los bacteriófagos se encuentran también en un porcentaje de 60% en cuanto a reactivos generales de laboratorio, los cuales deben de tenerse en cuenta⁵⁷⁻⁶⁰. De igual forma, se han visto secuencias virales en reactivos de laboratorio pertenecientes a adenovirus C, herpesvirus humano 1 y 5, VIH-1, parvovirus humano B19 y torqueteno virus⁵⁹. Es por ello que siempre se debe añadir un control negativo en cada procedimiento, para controlar las contaminaciones ambientales que pudiera haber en cada laboratorio, tal y como se ha hecho en el presente estudio. Y, aun así, es importante confirmar los resultados mediante PCR⁶⁰.

Algunas limitaciones más a considerar son las dificultades para identificar secuencias no incluidas en las bases de datos de referencia o comprender el significado de secuencias raras encontradas en las muestras. Como hemos visto en este estudio, en función de la herramienta bioinformática empleada para analizar los datos, sólo encontramos los microorganismos que aparezcan en esa base de datos. Es decir, si en la muestra se encuentra un patógeno nuevo, no podemos verlo en los resultados, ya que no se podrá comparar con las bases de datos utilizadas por bioinformática y por lo tanto, este resultado será un falso negativo². Esto lo podemos comprobar en la Tabla 3, donde en la mayoría de los casos, Pikavirus y BLAST no son capaces de detectar secuencias víricas que sí son identificadas mediante Kraken y IDseq, tal como ha ocurrido por ejemplo en las muestras LCRA128 para HHV-7, LCRA149 para parvovirus B19, LCRA183 para adenovirus, LCRA184 para EBV y astrovirus, y LCRA185 para adenovirus. Esto puede deberse a que en estas herramientas donde no se ha encontrado el virus problema no existe una base de datos de las secuencias de dichos virus, por lo que son incompletas para algunos de ellos. Tampoco se detecta sapovirus en el control positivo 2 mediante Pikavirus. Para vencer esta limitación, hemos usado la plataforma de IDseq donde es posible eliminar el *background* constituido por los controles negativos y las posibles contaminaciones que pueda haber, indicando de una forma más fiable la presencia de los virus en las muestras.

En conclusión, podemos confirmar que la metagenómica puede ser un método útil de diagnóstico de encefalitis y meningitis de origen vírico a partir de LCR, ya que es capaz de detectar virus que no son capaces de identificarse mediante otras técnicas de diagnóstico, como PCR o ELISA, de una forma precisa, a pesar de la duración del ensayo.

CONCLUSIONES

En este estudio, mediante mNGS y el uso de herramientas bioinformáticas, se ha podido comprobar la presencia de virus en algunas muestras de LCR empleadas en este trabajo, entre los que destacan TOSV, norovirus y adenovirus, que, según la literatura, se asocian (frecuentemente en el caso de TOSV y esporádicamente en el caso de los norovirus y adenovirus) a casos de enfermedades neurológicas.

La herramienta bioinformática IDseq permite la identificación de virus mediante secuencias obtenidas por metagenómica de una manera más sensible y eficaz en comparación con otras bases de datos (Pikavirus, BLAST y Kraken). En IDseq es posible eliminar el *background* de contaminantes que aparecen en los controles negativos, lo que permite la detección más fiable de virus en las muestras problema.

Como conclusión, la metagenómica es una técnica de identificación de virus eficaz y sensible que, usando las herramientas bioinformáticas adecuadas, permite detectar bajas concentraciones de virus en LCR.

BIBLIOGRAFÍA

1. Piantadosi A, Mukerji SS, Ye S, Leone MJ, Freimark LM, Park D, et al. Enhanced virus detection and metagenomic sequencing in patients with meningitis and encephalitis. *MBio*. 2021; 12(4):e0114321. <http://dx.doi.org/10.1128/mBio.01143-21>
2. Brown JR, Bharucha T, Breuer J. Encephalitis diagnosis using metagenomics: application of next generation sequencing for undiagnosed cases. *J Infect*. 2018; 76(3):225–40. <http://dx.doi.org/10.1016/j.jinf.2017.12.014>
3. Swanson PA II, McGavern DB. Viral diseases of the central nervous system. *Curr Opin Virol*. 2015; 11:44–54. <http://dx.doi.org/10.1016/j.coviro.2014.12.009>
4. Negrodo Antón AI, de Ory Manchón F, Sánchez-Seco Fariñas MP, Franco Narváez L, Gegúndez Cámara MI, Navarro Mari JM, et al. Diagnóstico microbiológico de arbovirosis y robovirosis emergentes. *Enferm Infecc Microbiol Clin*. 2015; 33(3):197–205. <http://dx.doi.org/10.1016/j.eimc.2013.07.011>
5. Rodríguez-Alarcón LGS, Moros MJS, Rodríguez BS, Corella SM, Balbuena SF, Suárez RP, et al. INFORME DE SITUACIÓN Y EVALUACIÓN DEL RIESGO DE ENFERMEDAD POR FLEBOVIRUS TRANSMITIDOS POR FLEBOTOMOS EN ESPAÑA. 2019 may.
6. Espy MJ, Uhl JR, Sloan LM, Buckwalter SP, Jones MF, Vetter EA, et al. Real-time PCR in clinical microbiology: Applications for routine laboratory testing. *Clin Microbiol Rev*. 2006; 19(3):595–595. <http://dx.doi.org/10.1128/cmr.00022-06>
7. Monticelli J, Geremia N, D'Agaro P, Petix V, Luzzati R. Aseptic central nervous system infections in adults: ¿what predictor for unknown etiological diagnosis? *Neurol Sci*. 2018; 39(5):863–70. <http://dx.doi.org/10.1007/s10072-018-3274-9>
8. Codina MG, de Cueto M, Vicente D, Echevarría JE, Prats G. Diagnóstico microbiológico de las infecciones del sistema nervioso central. *Enferm Infecc Microbiol Clin*. 2011; 29(2):127–34. <http://dx.doi.org/10.1016/j.eimc.2010.10.003>
9. O'Sullivan SS, O'Connell B, Redmond J. Aseptic meningitis: a 2-year review of diagnoses reached in a tertiary neurological and infectious disease centre. *Ir J Med Sci*. 2007; 176(3):215–9. <http://dx.doi.org/10.1007/s11845-007-0063-5>
10. Shukla B, Aguilera EA, Salazar L, Wootton SH, Kaewpoowat Q, Hasbun R. Aseptic meningitis in adults and children: Diagnostic and management challenges. *J Clin Virol*. 2017; 94:110–4. <http://dx.doi.org/10.1016/j.jcv.2017.07.016>
11. de Ory F, Avellón A, Echevarría JE, Sánchez-Seco MP, Trallero G, Cabrerizo M, et al. Viral infections of the central nervous system in Spain: a prospective study. *J Med Virol*. 2013; 85(3):554–62. <http://dx.doi.org/10.1002/jmv.23470>

12. . Wilson MR, Sample HA, Zorn KC, Arevalo S, Yu G, Neuhaus J, et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *N Engl J Med.* 2019; 380(24):2327–40. <http://dx.doi.org/10.1056/NEJMoa1803396>
13. Carbo EC, Buddingh EP, Karelioti E, Sidorov IA, Feltkamp MCW, Borne PA von D, et al. Improved diagnosis of viral encephalitis in adult and pediatric hematological patients using viral metagenomics. *J Clin Virol.* 2020; 130 (104566):104566. <http://dx.doi.org/10.1016/j.jcv.2020.104566>
14. McGill F. Incidence, aetiology, and sequelae of viral meningitis in UK adults: a multicentre prospective observational cohort study. *Lancet Infect Dis.* 2005.
15. Aurelius E, Forsgren M, Gille E, Sköldenberg B. Neurologic morbidity after herpes simplex virus type 2 meningitis: a retrospective study of 40 patients. *Scand J Infect Dis.* 2002; 34(4):278–83. <http://dx.doi.org/10.1080/00365540110080485>
16. Schmidt H, Cohrs S, Heinemann T, Goerdts C, Djukic M, Heimann B, et al. Sleep disorders are long-term sequelae of both bacterial and viral meningitis. *J Neurol Neurosurg Psychiatry.* 2006; 77(4):554–8. <http://dx.doi.org/10.1136/jnnp.2005.071142>
17. Schmidt H, Heimann B, Djukic M, Mazurek C, Fels C, Wallesch C-W, et al. Neuropsychological sequelae of bacterial and viral meningitis. *Brain.* 2006; 129(Pt 2):333–45. <http://dx.doi.org/10.1093/brain/awh711>
18. Sittinger H, Müller M, Schweizer I, Merkelbach S. Mild cognitive impairment after viral meningitis in adults. *J Neurol.* 2002; 249(5):554–60. <http://dx.doi.org/10.1007/s004150200064>
19. Cabrerizo M, García-Iñiguez JP, Munell F, Amado A, Madurga-Revilla P, Rodrigo C, et al. First cases of severe flaccid paralysis associated with Enterovirus D68 infection in Spain, 2015–2016. *Pediatr Infect Dis J.* 2017; 36(12):1214–6. <http://dx.doi.org/10.1097/inf.0000000000001668>
20. Pardigon N. Pathophysiological mechanisms of Flavivirus infection of the central nervous system. *Transfus Clin Biol.* 2017; 24(3):96–100. <http://dx.doi.org/10.1016/j.tracli.2017.05.005>
21. Kamar N, Bendall RP, Peron JM, Cintas P, Prudhomme L, Mansuy JM, et al. Hepatitis E virus and neurologic disorders. *Emerg Infect Dis.* 2011; 17(2):173–9. <http://dx.doi.org/10.3201/eid1702.100856>
22. Graff K, Dominguez SR, Messacar K. Metagenomic next-generation sequencing for diagnosis of pediatric meningitis and encephalitis: A review. *J Pediatric Infect Dis Soc.* 2021; 10 (Supplement_4): S78–87. <http://dx.doi.org/10.1093/jpids/piab067>

23. Cercenado E, Rafael M, Moreno C, Palomo AO, Causapé CL, González F, et al. Seimc.org. <https://seimc.org/contenidos/documentoscientificos/procedimientosmicrobiologia/seimc-procedimiento71.pdf>
24. Machyna M, Simon MD. Catching RNAs on chromatin using hybridization capture methods. *Brief Funct Genomics*. 2018; 17(2):96–103. <http://dx.doi.org/10.1093/bfgp/elx038>
25. O’Connell GC, Smothers CG. Optimized methodology for product recovery following emulsion PCR: applications for amplification of aptamer libraries and other complex templates. *J Biol Methods*. 2020; 7(1):e128. <http://dx.doi.org/10.14440/jbm.2020.316>
26. Kiselev D, Matsvay A, Abramov I, Dedkov V, Shipulin G, Khafizov K. Current trends in diagnostics of viral infections of unknown etiology. *Viruses*. 2020; 12(2):211. <http://dx.doi.org/10.3390/v12020211>
27. Mijatovic-Rustempasic S, Esona MD, Williams AL, Bowen MD. Sensitive and specific nested PCR assay for detection of rotavirus A in samples with a low viral load. *J Virol Methods*. 2016; 236:41–6. <http://dx.doi.org/10.1016/j.jviromet.2016.07.007>
28. Sánchez-Seco M-P, Echevarría J-M, Hernández L, Estévez D, Navarro-Marí J-M, Tenorio A. Detection and identification of Toscana and other phleboviruses by RT-nested-PCR assays with degenerated primers: Nested-RT-PCRs for Phlebovirus Detection. *J Med Virol*. 2003; 71(1):140–9. <http://dx.doi.org/10.1002/jmv.10465>
29. Pérez-Ruiz M, Collao X, Navarro-Marí J-M, Tenorio A. Reverse transcription, real-time PCR assay for detection of Toscana virus. *J Clin Virol*. 2007; 39(4):276–81. <http://dx.doi.org/10.1016/j.jcv.2007.05.003>
30. Hasan MR, Sundararaju S, Tang P, Tsui K-M, Lopez AP, Janahi M, et al. A metagenomics-based diagnostic approach for central nervous system infections in hospital acute care setting. *Sci Rep*. 2020; 10(1):11194. <http://dx.doi.org/10.1038/s41598-020-68159-z>
31. Grundy BS, Parikh H, Jacob S, Banura P, Moore CC, Liu J, et al. Pathogen detection using metagenomic next-generation sequencing of plasma samples from patients with sepsis in Uganda. *Microbiol Spectr*. 2023; 11(1):e0431222. <http://dx.doi.org/10.1128/spectrum.04312-22>
32. Remoli ME, Jiménez M, Fortuna C, Benedetti E, Marchi A, Genovese D, et al. Phleboviruses detection in *Phlebotomus perniciosus* from a human leishmaniasis focus in South-West Madrid region, Spain. *Parasit Vectors*. 2016; 9(1):205. <http://dx.doi.org/10.1186/s13071-016-1488-3>

33. Collao X, Palacios G, Sanbonmatsu-Gámez S, Pérez-Ruiz M, Negredo AI, Navarro-Marí JM, et al. Genetic diversity of Toscana virus. *Emerg Infect Dis.* 2009; 15(4):574–7. <http://dx.doi.org/10.3201/eid1504.081111>
34. Laroche L, Jourdain F, Ayhan N, Bañuls A-L, Charrel R, Prudhomme J. Incubation period for neuroinvasive Toscana virus infections. *Emerg Infect Dis.* 2021; 27(12):3147–50. <http://dx.doi.org/10.3201/eid2712.203172>
35. Sanbonmatsu-Gámez S, Pedrosa-Corral I, Navarro-Marí JM, Pérez-Ruiz M. Update in diagnostics of Toscana virus infection in a hyperendemic region (southern Spain). *Viruses.* 2021; 13(8):1438. <http://dx.doi.org/10.3390/v13081438>
36. García San Miguel L, Sierra MJ, Vazquez A, Fernandez-Martínez B, Molina R, Sanchez-Seco MP, et al. Phlebovirus-associated diseases transmitted by phlebotominae in Spain: Are we at risk? *Enferm Infecc Microbiol Clin (Engl).* 2021; 39(7):345–51. <http://dx.doi.org/10.1016/j.eimce.2021.05.001>
37. Echevarría J-M, de Ory F, Guisasaola M-E, Sánchez-Seco M-P, Tenorio A, Lozano A, et al. Acute meningitis due to Toscana virus infection among patients from both the Spanish Mediterranean region and the region of Madrid. *J Clin Virol.* 2003; 26(1):79–84. [http://dx.doi.org/10.1016/s1386-6532\(02\)00041-0](http://dx.doi.org/10.1016/s1386-6532(02)00041-0)
38. Martínez JG, García SG, Walter S, Gil-Prieto R, Lacomba DL, Marín-García P, et al. Seroprevalence against Toscana virus in Spain: The case of the autonomous community of Madrid. *J Vector Borne Dis.* 2022; 59(2):172–7. <http://dx.doi.org/10.4103/0972-9062.335771>
39. Ito S, Takeshita S, Nezu A, Aihara Y, Usuku S, Noguchi Y, et al. Norovirus-associated encephalopathy. *Pediatr Infect Dis J.* 2006; 25(7):651–2. <http://dx.doi.org/10.1097/01.inf.0000225789.92512.6d>
40. Shima T, Okumura A, Kurahashi H, Numoto S, Abe S, Ikeno M, et al. A nationwide survey of norovirus-associated encephalitis/encephalopathy in Japan. *Brain Dev.* 2019; 41(3):263–70. <http://dx.doi.org/10.1016/j.braindev.2018.11.001>
41. Sánchez-Fauquier A, González-Galán V, Arroyo S, Rodà D, Pons M, García J-J. Norovirus-associated encephalitis in a previously healthy 2-year-old girl. *Pediatr Infect Dis J.* 2015; 34(2):222–3. <http://dx.doi.org/10.1097/INF.0000000000000547>
42. Kimura E, Goto H, Migita A, Harada S, Yamashita S, Hirano T, et al. An adult norovirus-related encephalitis/encephalopathy with mild clinical manifestation. *BMJ Case Rep.* 2010; 2010: bcr0320102784–bcr0320102784. <http://dx.doi.org/10.1136/bcr.03.2010.2784>

43. Chaudhary SC, Khandelwal A, Tandon R, Sawlani KK. Rabies encephalitis. *BMJ Case Rep.* 2021; 14(4). <http://dx.doi.org/10.1136/bcr-2020-239249>
44. Regnault B, Evrard B, Plu I, Dacheux L, Troadec E, Cozette P, et al. First case of lethal encephalitis in Western Europe due to European bat lyssavirus type 1. *Clin Infect Dis.* 2022; 74(3):461–6. <http://dx.doi.org/10.1093/cid/ciab443>
45. Scott TP, Nel LH. Lyssaviruses and the fatal encephalitic disease rabies. *Front Immunol.* 2021; 12:786953. <http://dx.doi.org/10.3389/fimmu.2021.786953>
46. Vidal LR, de Almeida SM, Cavalli BM, Dieckmann TG, Raboni SM, Salvador GLO, et al. Human adenovirus meningoencephalitis: a 3-years' overview. *J Neurovirol.* 2019; 25(4):589–96. <http://dx.doi.org/10.1007/s13365-019-00758-7>
47. Schwartz KL, Richardson SE, MacGregor D, Mahant S, Raghuram K, Bitnun A. Adenovirus-associated central nervous system disease in children. *J Pediatr.* 2019; 205:130–7. <http://dx.doi.org/10.1016/j.jpeds.2018.09.036>
48. Chaari A, Berrajah L, Bahloul M, Bouaziz M. Rubella encephalitis. *Neurol India.* 2011; 59(5):766–7. <http://dx.doi.org/10.4103/0028-3886.86559>
49. Winter AK, Moss WJ. Rubella. *Lancet.* 2022; 399(10332):1336–46. [http://dx.doi.org/10.1016/s0140-6736\(21\)02691-x](http://dx.doi.org/10.1016/s0140-6736(21)02691-x)
50. Matthews E, Beckham JD, Piquet AL, Tyler KL, Chauhan L, Pastula DM. Herpesvirus-associated encephalitis: An update. *Curr Trop Med Rep.* 2022; 9(3):92–100. <http://dx.doi.org/10.1007/s40475-022-00255-8>
51. Wang Y, Yang J, Wen Y. Lessons from Epstein-Barr virus DNA detection in cerebrospinal fluid as a diagnostic tool for EBV-induced central nervous system dysfunction among HIV-positive patients. *Biomed Pharmacother.* 2022; 145(112392):112392. <http://dx.doi.org/10.1016/j.biopha.2021.112392>
52. Cordey S, Vu D-L, Schibler M, L'Huillier AG, Brito F, Docquier M, et al. Astrovirus MLB2, a new gastroenteric virus associated with meningitis and disseminated infection. *Emerg Infect Dis.* 2016; 22(5):846–53. <http://dx.doi.org/10.3201/eid2205.150807>
53. Cao J, Zhu X-Q. Acute viral encephalitis associated with human parvovirus B19 infection: unexpectedly diagnosed by metagenomic next-generation sequencing. *J Neurovirol.* 2020; 26(6):980–3. <http://dx.doi.org/10.1007/s13365-020-00885-6>
54. Teoh H-L, Mohammad SS, Britton PN, Kandula T, Lorentzos MS, Booy R, et al. Clinical characteristics and functional motor outcomes of Enterovirus 71 neurological disease in children. *JAMA Neurol.* 2016; 73(3):300–7. <http://dx.doi.org/10.1001/jamaneurol.2015.4388>

55. Henderson M, Pepper N, Bawa M, Muir D, Everitt A, Mackie NE, et al. Cerebrospinal fluid virology in people with HIV. *HIV Med.* 2023; <http://dx.doi.org/10.1111/hiv.13471>
56. Li Y, Qu T, Li D, Jing J, Deng Q, Wan X. Human herpesvirus 7 encephalitis in an immunocompetent adult and a literature review. *Virol J.* 2022;19(1):200. <http://dx.doi.org/10.1186/s12985-022-01925-9>
57. Ghose C, Ly M, Schwanemann LK, Shin JH, Atab K, Barr JJ, et al. The virome of cerebrospinal fluid: Viruses where we once thought there were none. *Front Microbiol.* 2019; 10:2061. <http://dx.doi.org/10.3389/fmicb.2019.02061>
58. Porter AF, Cobbin J, Li C-X, Eden J-S, Holmes EC. Metagenomic identification of viral sequences in laboratory reagents. *Viruses.* 2021;13(11). <http://dx.doi.org/10.3390/v13112122>
59. Asplund M, Kjartansdóttir KR, Mollerup S, Vinner L, Fridholm H, Herrera JAR, et al. Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. *Clin Microbiol Infect.* 2019; 25(10):1277–85. <http://dx.doi.org/10.1016/j.cmi.2019.04.028>
60. Holmes EC. Reagent contamination in viromics: all that glitters is not gold. *Clin Microbiol Infect.* 2019; 25(10):1167–8. <http://dx.doi.org/10.1016/j.cmi.2019.06.019>

AGRADECIMIENTOS

Quería agradecer la ayuda en este proyecto a mi tutora, M^a Dolores Fernández García, por su apoyo todos estos meses, y por su paciencia y dedicación durante todo el trabajo. Me he sentido muy acogida y cómoda en el laboratorio.

También quiero mencionar de una manera especial a Juan Camacho Padilla, quien ha sido mi supervisor en las tareas del laboratorio, y quien me ha enseñado con paciencia cada una de las técnicas que he trabajado.

Además, me gustaría agradecer su ayuda a todas aquellas personas que, en mayor o menor medida, me han ayudado a continuar y terminar este trabajo, ya sea en el laboratorio, haciéndome favores, o a través de consejos en cuanto a la memoria de este TFM, como Vanesa Recio, Juan Emilio Echevarría, Estéfani García Ríos, José Miguel Berciano, María Cabrerizo, y todos aquellos compañeros del laboratorio.

ANEXO 1: definición de términos bioinformáticos

- Background: secuencias de fondo que existen en las muestras y en los controles negativos debido a contaminaciones ambientales o de reactivos y materiales del laboratorio.
- Clusterización: empaquetamiento de muestras siguiendo una serie de requisitos.
- Cobertura > 10x: porcentaje de bases del total del genoma que fueron leídas con una profundidad superior a 10.
- Contig: conjunto de secuencias o fragmentos de ADN que se superponen de forma parcial, de tal manera que en conjunto muestran una representación continua de una región genómica.
- Hit: aquel punto de color más intenso en el *heatmap* que tiene un valor significativo en comparación con los demás teniendo en cuenta los filtros empleados (ver métodos).
- Heatmap: método de visualización de datos que, según la variable que se escoja, mide la magnitud de dicha variable en intensidad de color en dos dimensiones.
- K-mers: subcadenas de longitud k que se encuentran dentro de una secuencia del genoma. Se usan para ensamblar secuencias de ADN y mejorar la expresión génica heteróloga, identificar especies en muestras por secuenciación masiva, crear vacunas, etc.
- NT: en IDseq, base de datos de secuencias de nucleótidos del NCBI, conformando una colección de secuencias de varias fuentes, incluidas GeneBank, RefSeq, TPA y PDB.
- NR: en IDseq, base de datos del NCBI de secuencias de proteínas no redundantes.
- Profundidad: es el número de veces en las que se lee un nucleótido concreto durante el proceso de secuenciación.
- r (reads): lecturas. Es el número total de lecturas que se alinean con el taxón en la base de datos NCBI NT/NR.
- rPM (reads per million): lecturas por millón, es decir, el número de lecturas que se alinean con el taxón en la base de datos NCBI NT/NR por millón de lecturas secuenciadas. Se usa como métrica estándar en los análisis bioinformáticos.
- Z-Score: es un sistema de puntuación de IDseq que se usa para evaluar la prevalencia de microbios en las muestras, en comparación con los controles introducidos en el *background*.

ANEXO 2: secuencias consenso obtenidas mediante la plataforma IDseq

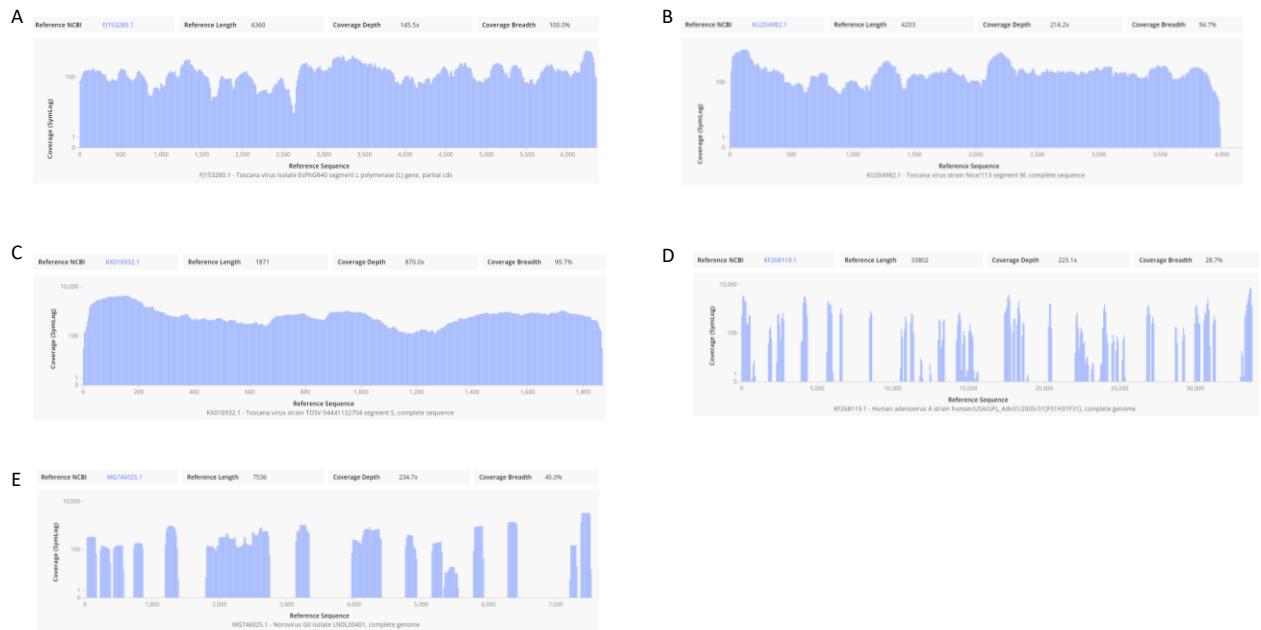


Figura S1: secuencias consenso obtenidas a partir de IDseq. A) Segmento L del virus Toscana de la muestra LCRA145. B) Segmento M del virus Toscana de la muestra LCRA145. C) Segmento S del virus Toscana de la muestra LCRA145. D) Secuencia de adenovirus de la muestra LCRA185. E) Secuencia de norovirus de la muestra LCRA185.

ANEXO 3: ABREVIATURAS Y SIGLAS

- ADNc: ADN complementario.
- CHIKV: virus de Chikungunya.
- CMV: citomegalovirus.
- CNM: Centro Nacional de Microbiología.
- DENV: virus del dengue.
- EBV: virus de Epstein-Barr.
- ELISA: enzimoimmunoanálisis de adsorción.
- EV: enterovirus.
- FTF: flebovirus transmitidos por flebotomos.
- HER K113: retrovirus humano K113.
- HV: herpesvirus.
- HVS: herpesvirus simple.
- ISCIII: Instituto de Salud Carlos III.
- LCMV: virus de coriomeningitis linfocitaria.
- LCR: líquido cefalorraquídeo.
- mNGS: secuenciación metagenómica de próxima generación.

- MuV: virus de las paperas (parotiditis).
- pb: pares de bases.
- PCR: reacción en cadena de la polimerasa.
- RT-PCR: PCR en tiempo real.
- SNC: sistema nervioso central.
- TOSV: virus Toscana.
- VIH-1: virus de la inmunodeficiencia humana tipo 1.
- VZV: virus de varicela zóster.
- WNV: virus West Nile.
- YFV: virus de la fiebre amarilla.