



VNIVERSITAT
DE VALÈNCIA



INSTITUTE FOR
INTEGRATIVE
SYSTEMS BIOLOGY

TESIS DOCTORAL

Caracterización del viroma en plasma de donantes de sangre

María Concepción Cebriá Mendoza

Programa de Doctorado en Biodiversidad y Biología Evolutiva

Director: José Manuel Cuevas Torrijos

Abril 2022



VNIVERSITAT
E VALÈNCIA

Instituto de Biología Integrativa de Sistemas (I2SYSBIO)

Programa de Doctorado en Biodiversidad y Biología Evolutiva

Caracterización del viroma en plasma de donantes de sangre

Tesis Doctoral

María Concepción Cebriá Mendoza

Abril 2022

Director

José Manuel Cuevas Torrijos

*“Everything is theoretically impossible,
until it is done”*

Robert A. Heinlein

Índice

Índice	1
Prólogo	9
Introducción	13
1. Virus como parte de los ecosistemas	15
2. Metagenómica y descubrimiento de nuevas secuencias virales ...	19
2.1. Recogida de muestras	20
2.2. Procesado de las muestras.....	21
2.3. Extracción de los ácidos nucleicos	22
2.4. Secuenciación.....	23
2.5. Análisis bioinformático.....	23
2.6. Sesgos metodológicos en el estudio del viroma	25
2.7 Aplicaciones de la metagenómica viral	28
3. Viroma humano	30
3.1. Viroma del tracto gastrointestinal.....	33
3.2. Viroma de la piel	35
3.3. Viroma del tracto genitourinario.....	35
3.4. Viroma de la sangre.....	35
3.5. Muestras de sangre	37
3.6. Anelovirus	37
3.7. Pegivirus	41
3.8. Relación del viroma con la fisiología humana	43
Objetivos	47
Material y métodos	51
1. Procesado de las muestras de sangre.....	54

2. Prueba piloto: comparación entre la concentración por extracción directa vs centrifugación.....	55
2.1. Concentración por extracción directa del plasma.....	57
2.2. Concentración por centrifugación.....	57
2.3. Obtención de las muestras necesarias para la construcción de librerías a partir de la fracción de ADN.....	58
2.4. Obtención de las muestras necesarias para la construcción de librerías a partir de la fracción de ARN.....	58
2.5. Secuenciación masiva.....	60
2.6. Análisis bioinformático.....	60
3. Protocolo de obtención de ácidos nucleicos virales para secuenciación	61
3.1. Obtención del plasma procedente de muestras de sangre total.....	61
3.2. Extracción de ácidos nucleicos virales y amplificación	62
3.3. Secuenciación masiva.....	64
3.4. Análisis de las secuencias.....	64
3.5. Análisis filogenético.....	69
3.6. Secuenciación por el método Sanger.....	70
4. Análisis de recombinación	72
4.1. Análisis de recombinación de <i>Pegivirus</i>	72
4.2. Análisis de recombinación de anelovirus.....	72
Capítulo 1: Implementación del protocolo de enriquecimiento viral y resultados preliminares.	77
1. Resultados.....	79
1.1. Estrategia y resultado de la prueba piloto	79
1.2. Análisis de 120 muestras de plasma	82

2. Discusión.....	94
Capítulo 2: Análisis de la diversidad de los virus huérfanos presentes en plasma.....	99
1. Resultados.....	101
1.1. Resultado global de la secuenciación.....	101
1.2. Análisis de los anelovirus	106
1.3. Análisis de HPgV.....	116
2. Discusión.....	121
Capítulo 3: Análisis del aneloma individual de 50 donantes	127
1. Resultados.....	129
1.1. Análisis preliminar de las secuencias	129
1.2. Análisis de anelovirus.....	133
1.3. Prevalencia de la infección por anelovirus en las 50 muestras de plasma individual	139
1.4. Análisis del viroma de donantes individuales	141
1.5. Análisis de recombinación de los genomas de anelovirus	148
2. Discusión.....	153
Discusión general	159
Conclusiones	167
Bibliografía.....	171
Anexo	195

Abreviaturas

AU	<i>Approximately unbiased test</i>	Prueba aproximadamente imparcial
DMEM	<i>Dulbecco's Modified Eagles's Medium</i>	
DMSO	<i>Dimethyl sulfoxide</i>	Dimetilsulfóxido
dNTPs	<i>Deoxynucleotide triphosphates</i>	Deoxinucleótidos trifosfato
DTT	<i>1,4-Dithiothreitol</i>	1,4-Ditiotreitól
ELW	<i>Expected likelihood weight test</i>	Prueba de pesos de probabilidad esperada
FBS	<i>Fetal Bovine Serum</i>	Suero bovino fetal
HPgV	<i>Human Pegivirus</i>	Pegivirus humano
Hpi	<i>Hours post infection</i>	Horas post-infección
HTS	<i>High-throughput sequencing</i>	Secuenciación de alto rendimiento
ICTV	<i>International Committee on Taxonomy of Viruses</i>	Comité Internacional de Taxonomía de Virus
LB	<i>Luria-Bertani medium</i>	Medio Luria-Bertani
MDA	<i>Multiple displacement amplification</i>	Amplificación por desplazamiento múltiple
ML	<i>Maximum likelihood</i>	Máxima verosimilitud
MOI	<i>Multiplicity of infection</i>	Multiplicidad de infección
NCBI	<i>National Center for Biotechnology Information</i>	Centro Nacional de Información Biotecnológica
ORF	<i>Open Reading Frame</i>	Pauta abierta de lectura
PBS	<i>Phosphate Buffered Saline</i>	Tampón fosfato salino
PES	<i>Polyethersulfone</i>	Polietersulfona
PFU	<i>Plaque-forming unit</i>	Partículas formadoras de calva
PIF	<i>Escherichia coli C IJ 1862</i>	<i>Escherichia coli</i> C IJ 1862
rpm	<i>Rounds per minute</i>	Revoluciones por minuto
SEM	<i>Standard error of the mean</i>	Error estándar de la media
SISPA	<i>Sequence-independent single-primer amplification</i>	Amplificación independiente de secuencia con cebador único
SSIII	<i>SuperScript™ III</i>	SuperScript™ III
TTMDV	<i>Torque Teno Midi Virus</i>	Torque Teno Midi Virus
TTMV	<i>Torque Teno Mini Virus</i>	Torque Teno Mini Virus
TTV	<i>Torque Teno Virus</i>	Torque Teno Virus

VIH	<i>Human Immunodeficiency virus</i>	Virus de la inmunodeficiencia humana
VLP	<i>Viral-like particles</i>	Partículas similares a virus
VSV	<i>Vesicular stomatitis virus</i>	Virus de la estomatitis vesicular
VV	<i>Vaccinia Virus</i>	Virus Vaccinia

Prólogo

No es fácil resumir lo que pasa y se siente a lo largo de una Tesis Doctoral, pero si pudiera resumir lo sentido en una palabra sería agradecimiento. Este trabajo ha sido para mí un regalo, no solo a nivel laboral, ya que gracias a este proyecto he aprendido muchísimo y me he formado como investigadora, sino también a nivel personal, ya que también me ha hecho adquirir muchísimas aptitudes.

Por suerte, hay muchas personas implicadas en que las cosas positivas pesen mucho más que las negativas. En primer lugar, José, muchas gracias por darme la oportunidad, por la paciencia, por haber estado siempre disponible y por haber confiado en mí para llevar a cabo este proyecto. A Rafa, gracias por acoger a esa estudiante de máster en el laboratorio, ya que eso fue la semilla de este trabajo. A Pilar, gracias por los consejos, por confiar en su día en mí, por los congresos y por enseñarme tan bien a hacer diluciones seriadas (jajaja), siempre te estaré muy agradecida. A Wladimiro, por toda la paciencia que ha tenido conmigo. Durante este periodo de tiempo, he podido conocer a numerosos compañeros de laboratorio que han contribuido que el ambiente del labo haya sido genial: todos los estudiantes, Ana, Jérémy, Amanda, Mireia, Celia, Iván, JV, Salo, Ernesto, Jorge, María D, Layla, Lola y sus pasteles, Rubén, Raquel, Robby, David y Fabiana, ¡os deseo lo mejor! Espero que me perdonen los anteriormente citados, pero a estas personas les debo gran parte de mi tesis y así quiero que quede reflejado en estas líneas: a Bea Beamud, gracias por tu infinita paciencia ayudándome con la bioinformática y por ser la mejor anfitriona. A Lucas, por haber estado desde prácticamente el principio, por todas las conversaciones profundas que hemos tenido en el agujero negro, también llamado mi sitio. A Carlos y Chantal (sorry vais en pack) por ser tan divertidos y tener ese alijo de chokolatinas que nos ha salvado más de una tarde. A Inés, gracias por aportar ese toque de locura durante tu periodo en el

lab y desde entonces en mi vida. A Flo y Víctor, no existen suficientes palabras para vosotros. Una amistad que se forjó viendo la obra del muro (que aún sigue en pie, sorprendentemente), con los que he aprendido tanto infinidad de conocimiento científico, así como de la vida en general, gran parte de los mejores momentos que he vivido han sido gracias a vosotros. Y como no, si de algo estoy feliz es de haber podido compartir esta experiencia con Ale, que ha sido mi mayor apoyo dentro del lab, mi compañera de múltiples aventuras fuera, mi mano derecha, mi pañuelo de lágrimas y la mejor consejera que nadie pueda tener. Amiga, GRACIAS se queda corto.

Este trabajo no podría haberse realizado sin la colaboración con el Centro de Transfusiones de la Comunidad Valenciana. Gracias a todas las enfermeras y médicos con los que he coincidido (que no han sido pocos en casi un año de recogida de muestras) por lo bien que me acogieron y porque a pesar de darles más faena, siempre me trataron como a una más. Y como no, a todos los donantes que han participado en el estudio.

También me gustaría agradecer a mis amigas, las que sin saber que hago a ciencia cierta, me han animado siempre, a las que cuando necesitaba muestras no dudaron en apuntarse y venir a donar. A Gema, has estado formando parte de esta tesis desde el minuto 1, gracias por escucharme siempre y aconsejarme lo mejor. Y como no, gran parte del mérito de haber llegado hoy hasta aquí es de mi familia, gracias a mis padres por permitirme llegar hasta aquí y a mi hermana por apoyarme en todo. Gracias por estar cuando os he necesitado y por cuidarme tanto. Y, por último, agradecer a esa persona que me ha apoyado incondicionalmente en todo. Infinitas gracias David, por aguantar los dramitas causados por los nervios y por confiar en mí misma mucho más de lo que hago yo.

Introducción

Introducción

1. Virus como parte de los ecosistemas

Los virus son ubicuos en todos los ambientes naturales y pueden ser considerados la mayor fuente de ácidos nucleicos de la Tierra, representando el mayor origen de biomasa y de diversidad genética en el planeta (Holmes, 2011). Se estima que sólo en el océano existen más de 10^{31} partículas virales, presentando una abundancia que excede la presencia de arqueas y bacterias (Suttle, 2005). Los virus están integrados de forma natural en los ecosistemas y las interacciones entre ellos y sus hospedadores no solo repercuten en su propia abundancia, diversidad y evolución, sino también en la de sus poblaciones y comunidades hospedadoras (French & Holmes, 2020). Un ejemplo de ello son los bacteriófagos en el océano, que lisan las bacterias haciendo que el contenido celular de éstas se libere. Este material es esencial para el crecimiento del fitoplancton, que a su vez, está involucrado en el ciclo del carbono y en el cambio climático (Danovaro et al., 2011), por lo que la presencia de los virus en los ecosistemas juega un papel determinante.

Hasta el momento, se conocen alrededor de 260 virus capaces de infectar seres humanos, clasificados en 25 familias virales distintas (Carroll et al., 2018). Sin embargo, este número está en constante aumento, puesto que se estima que existen más de 1,5 millones de especies virales que pueden infectar a los principales reservorios virales, mamíferos y aves, por tanto, con potencial de causar infecciones virales emergentes en nuestra especie (Jonas & Seifman, 2019). El término de virus emergente hace referencia a un patógeno que no causaba enfermedad aparente en su hospedador natural, pero sí que es capaz de causarla en una especie nueva (Choi, 2021). Las consecuencias debidas a estas infecciones causadas por virus emergentes pueden ser el desarrollo de una nueva enfermedad o bien que una

enfermedad previamente identificada presente nuevas manifestaciones clínicas, asociadas a un aumento repentino de la incidencia, gravedad o el área geográfica en un lapso de tiempo (Frutos, Gavotte, & Devaux, 2021). El análisis de la progresión de las enfermedades infecciosas emergentes sugiere que su ritmo de aparición va en aumento y que la emergencia de nuevos virus aún no está limitada por la riqueza (en cuanto al número de virus) o la diversidad (es decir, la variabilidad genética) de los virus desconocidos, que en ambos casos se considera elevada (Anthony et al., 2013).

La gran mayoría de las enfermedades infecciosas emergentes detectadas en el último siglo tienen un carácter zoonótico (Roberts et al., 2021). Estas zoonosis representan infecciones virales que saltan las barreras de especie desde la vida salvaje, o desde animales domésticos, a humanos. Las zoonosis se pueden transmitir a humanos desde un huésped infectado mediante contacto directo o indirecto, es decir, a través de vectores mecánicos, biológicos o a través del contacto con fómites (Rupasinghe, Chomel, & Martínez-López, 2022). El papel que presentan los animales en las enfermedades infecciosas emergentes se conoce desde hace décadas, pero no ha sido hasta recientemente cuando se ha reconocido formalmente al establecerse la iniciativa One Health (<http://www.onehealthinitiative.com/>), con la que se intenta mejorar la colaboración entre profesionales de la salud humana, animal y medioambiental, junto con otras áreas, con el fin de colaborar y coordinar esfuerzos para mejorar la salud de humanos, animales y plantas en un ambiente compartido.

En los últimos diez años, ha habido brotes de infecciones de virus emergentes, como las causadas por el virus del Zika, Ébola, Dengue, Chikungunya, Hendra, Nipah y las causadas por coronavirus como el SARS-CoV, MERS-CoV y el SARS-CoV-2, causante de la pandemia actual, que implican

enormes impactos económicos, sociales y en salud global. Esta aparición de enfermedades emergentes se puede atribuir a ciertos comportamientos que dan como resultado una mayor frecuencia de exposición de los humanos a estos patógenos. Es decir, existen diversos impulsores externos que producen las condiciones adecuadas que permiten a los patógenos expandirse y adaptarse a nuevos nichos (Rupasinghe et al., 2022). Estos cambios actúan de manera sinérgica en el incremento del riesgo de emergencia de los patógenos, su transmisión y su oportunidad de infección (Kobayashi, 2018). Los distintos cambios pueden clasificarse en tres categorías diferentes. La primera de ellas engloba a los cambios que ocurren en la sociedad, debidos a causas como un incremento en la población mundial o al aumento de la globalización, entre otras (Azevedo et al., 2021; Kobayashi, 2018; Lipkin & Firth, 2013). En segundo lugar, las variaciones ocurridas en los ecosistemas y el medioambiente también influyen en la emergencia de nuevos virus. Algunos ejemplos de esto son el cambio climático (Kobayashi, 2018; Roberts et al., 2021; Rupasinghe et al., 2022) o los cambios en el uso de la tierra, como la deforestación, el aumento en la urbanización o la conversión de los bosques (Azevedo et al., 2021; Roberts et al., 2021). Por último, los cambios que ocurren en los propios virus también son un factor de riesgo para la emergencia de enfermedades infecciosas. Estas variaciones son entendidas como episodios de adaptación viral a un nuevo ambiente, que causan un incremento en la eficacia replicativa del virus y están asociadas a cambios genéticos producidos por sucesos de mutación y recombinación (Kobayashi, 2018).

La aparición de enfermedades infecciosas emergentes presenta una naturaleza accidental, es decir, se trata de una combinación única de procesos biológicos junto con factores antropogénicos que no puede predecirse (Frutos et al., 2021). Sin embargo, en un planeta cada vez más globalizado, un contagio

inicial que derive en una transmisión eficiente y sostenible puede extenderse rápidamente. Esto ha quedado bien demostrado por la actual pandemia causada por el virus SARS-CoV-2, que ha dado lugar a una crisis sanitaria, social y económica mundial sin precedentes (Magouras et al., 2020). En este contexto, pese a que no se puede evitar que se produzca la emergencia de algunos agentes biológicos infecciosos, es primordial el establecimiento de una vigilancia epidemiológica sobre los diferentes factores de riesgo asociados a la emergencia de enfermedades infecciosas (Kobayashi, 2018). En otras palabras, es urgente adoptar un enfoque más proactivo, en el que se lleve a cabo la identificación y mitigación de los distintos factores de riesgo involucrados en la emergencia de enfermedades infecciosas y su dispersión. Esto es crucial para prevenir mejor y responder más rápidamente a futuras epidemias y poder reducir así la carga global que suponen estas enfermedades (Lipkin & Anthony, 2015). En los últimos años, se han creado diferentes iniciativas como el ya concluido proyecto EPT-PREDICT (Carlson, 2020) o el *Global Virome Project* (Carroll et al., 2018), en los que se ha llevado a cabo la identificación de especies virales que presentan un potencial zoonótico. En concreto, en el caso del *Global Virome Project*, la idea principal fue la construcción de un atlas global de la mayoría de virus naturales y evaluar su riesgo de zoonosis, mejorando las herramientas de vigilancia, prevención y control proporcionadas por proyectos como EPT-PREDICT (Carroll et al., 2018; Jonas & Seifman, 2019).

Disponer de una base de datos del viroma global presenta grandes ventajas. Por un lado, la comprensión de la diversidad genética de los virus mejoraría el conocimiento sobre los factores que pueden facilitar la propagación de los virus desde sus reservorios a los humanos (Roberts et al., 2021). El coste de estos proyectos es elevado, pero es modesto si se pone en

perspectiva, ya que los brotes de enfermedades infecciosas virales pueden llevar asociadas pérdidas millonarias en cuanto a productividad, comercio, crecimiento económico y bienestar social (Jonas & Seifman, 2019) y esto es un hecho tangible actualmente debido a la pandemia causada por la COVID-19. Por otra parte, la generación de datos de secuencias de virus no evita por sí misma las potenciales pandemias, pero sí que proporciona datos que mejoran el conocimiento de la relación funcional entre los hospedadores y la diversidad viral (Anthony et al., 2013). Este mapeo de la diversidad viral, junto a la predicción de modelos ecológicos (es decir, el hecho de conocer el comportamiento del hospedador natural, del hospedador accidental y del patógeno en un ambiente), puede ayudar a prevenir e incluso predecir el contagio del agente infeccioso (Roberts et al., 2021).

2. Metagenómica y descubrimiento de nuevas secuencias virales

Los virus carecen de un marcador universal, como son los genes de ARN ribosómico 16S en las bacterias (Mirzaei et al., 2021). Por ello, existen múltiples técnicas para la detección de virus: las que están basadas en el cultivo de virus en tejidos, para su amplificación y posterior caracterización, las basadas en métodos serológicos (Lipkin & Firth, 2013) o las que emplean métodos moleculares, como la PCR o las micromatrices. Sin embargo, la técnica que ha impulsado el descubrimiento de numerosas secuencias virales y ha revolucionado el campo de la virología ha sido la metagenómica. La metagenómica se define como el análisis directo de las comunidades microbianas completas a partir de los ácidos nucleicos extraídos de una determinada muestra (Huson & Mitra, 2012). La metagenómica sorteas las limitaciones que presentan los métodos tradicionales de caracterización de virus, ya que la principal ventaja que presenta es que al dirigirse a los ácidos

nucleicos virales totales no se necesita un conocimiento *a priori* de los virus presentes en la muestra (Mirzaei et al., 2021; Rosario & Breitbart, 2011).

Los pasos llevados a cabo en el análisis metagenómico del viroma, que es el conjunto de virus presente en una muestra, presentan un impacto en la detección de las secuencias virales, por lo que el flujo de trabajo debe de ser adecuado al tipo de muestra, origen y volumen (Sathiamoorthy, Malott, Gisonni-Lex, & Ng, 2018). El protocolo de un estudio metagenómico suele incluir la recogida de muestras, el procesamiento de las mismas, la extracción de los ácidos nucleicos presentes, su secuenciación y los análisis bioinformáticos sobre los resultados de la secuenciación (**Figura 1**).

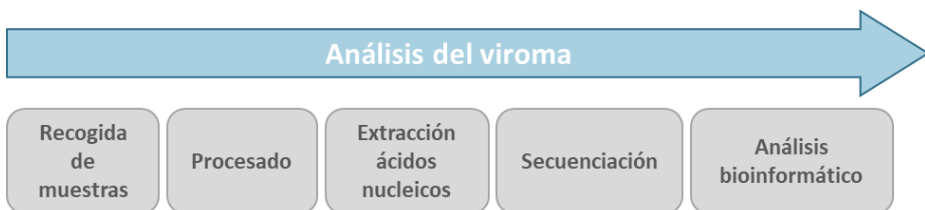


Figura 1: Esquema del flujo de trabajo seguido en un análisis de viroma. Este protocolo puede incluir los distintos pasos: recogida de muestras, procesado de las mismas, secuenciación de los ácidos nucleicos y análisis bioinformático. Figura adaptada de Santiago-Rodríguez & Hollister, 2020.

2.1. Recogida de muestras

El primer paso de un estudio metagenómico es la recogida de las muestras y su posterior almacenamiento. La recogida dependerá de la naturaleza de las muestras a emplear en el estudio metagenómico, lo cual determinará la cantidad y la forma de recogida. En la recogida, es importante seleccionar las condiciones más adecuadas de estabilización según el tipo de muestra y los objetivos del estudio (Santiago-Rodríguez & Hollister, 2020). Además, también se deben seleccionar las condiciones de almacenaje

óptimas, ya que, por ejemplo, en el caso de los virus, los ciclos de congelación-descongelación pueden afectar al título viral, por lo que es algo a tener en cuenta si se quiere medir la abundancia relativa de una especie viral. A su vez, los cambios en la temperatura de almacenaje pueden hacer que las células presentes en la muestra se lisen, lo que puede implicar un aumento de los residuos celulares, que pueden interferir en los pasos de purificación posteriores (Santiago-Rodriguez & Hollister, 2020).

2.2. Procesado de las muestras

En segundo lugar, se realiza el procesado de las muestras, que involucra diferentes procedimientos con el fin de eliminar los potenciales contaminantes presentes en las muestras. Es común que los estudios de metagenómica viral incluyan un paso de filtración para eliminar los residuos presentes en la muestra, aprovechando la ventaja de que los virus presentan un tamaño menor que las células eucariotas o procariontas (Conceição-Neto et al., 2015). Otro de los pasos que suele estar presente en un trabajo de metagenómica viral es la concentración de las partículas virales. La concentración se puede realizar con diferentes técnicas, por ejemplo, mediante ultracentrifugación (Kohl et al., 2015) o empleando filtros de purificación de proteínas con los que se consigue capturar las partículas de naturaleza viral (Chehoud et al., 2016). Además, los protocolos de análisis de viomas suelen incluir la eliminación de los ácidos nucleicos libres en la muestra mediante el uso de nucleasas (Conceição-Neto et al., 2015; Cotten et al., 2014; Zhang et al., 2016), ya que los ácidos nucleicos virales se encontrarán protegidos por las cápsides proteicas.

En este punto del protocolo, la aplicación de uno o varios de estos pasos puede ser útil, sobre todo cuando se manipulan muestras que presentan un

alto nivel de contaminación del hospedador, como saliva, sangre o biopsias (Santiago-Rodríguez & Hollister, 2020).

2.3. Extracción de los ácidos nucleicos

El siguiente paso en el protocolo consiste en la extracción de los ácidos nucleicos. Actualmente, se encuentran disponibles numerosos métodos de extracción de ácidos nucleicos, algunos de los cuales se encuentran dirigidos a determinados tipos de muestras, como sangre o saliva, y otros están dirigidos a determinadas especies virales (Bergallo et al., 2006).

En los estudios de viroma, la fracción viral suele encontrarse en una proporción baja, por lo que la extracción de los ácidos nucleicos puede tener un rendimiento residual, que en ocasiones puede comprometer los siguientes pasos del protocolo. Por ello, se suele llevar a cabo la amplificación de los ácidos nucleicos presentes en las muestras. Esta amplificación no es diferencial para los genomas virales, sino que se amplifican todos los ácidos nucleicos que se encuentran en la muestra de manera aleatoria. Con este fin existen varias técnicas, como la técnica de amplificación por desplazamiento múltiple (*multiple displacement amplification*, MDA), que emplea como cebadores hexámeros aleatorios y la polimerasa del fago $\phi 29$ (Berthet et al., 2008; Shoaib et al., 2008). Otra de las técnicas comúnmente empleadas es el método de amplificación independiente de secuencia con un cebador único (*sequence-independent single-primer amplification*, SISPA), que es una técnica que implica la unión de un cebador a las moléculas de ácidos nucleicos presentes en la muestra de manera previa a la amplificación (Reyes & Kim, 1991). En ambos casos, se permite la amplificación de los ácidos nucleicos presentes en la muestra, incluidos los genomas virales.

2.4. Secuenciación

El siguiente paso en un análisis metagenómico del viroma es la secuenciación masiva (*high-throughput sequencing*, HTS). Esta tecnología, debido a que emplea técnicas que se basan en la paralelización y miniaturización de las reacciones de secuenciación, presenta una capacidad de generar en una sola carrera (*run*) millones de secuencias, permitiendo el análisis de muestras biológicas diversas y facilitando la potencial detección de los distintos tipos de virus presentes. Actualmente, existen múltiples plataformas de secuenciación, que varían en cuanto a las reacciones químicas que se necesitan para la generación de las secuencias, pero presentan un protocolo similar a la hora de generar las librerías y los datos de secuenciación. A pesar de esto, las más empleadas son las comercializadas por Illumina®, ya que presentan un precio competitivo y generan secuencias cortas, pero de alta calidad. Aunque existe un dominio de la secuenciación por Illumina®, también se emplean otras tecnologías que se han desarrollado estos últimos años, como PacBio® de la compañía Pacific Biosciences, o MinION, desarrollado por la casa comercial Oxford Nanopore Technologies. En ambos casos, se produce secuenciación de molécula única en tiempo real, generándose secuencias de mayor tamaño, pero de peor calidad. La elección de la plataforma de secuenciación dependerá de los objetivos del proyecto, ya que las diferencias radican, principalmente, en la longitud de lectura y en la profundidad obtenida en la secuenciación.

2.5. Análisis bioinformático

El último paso en un protocolo de descripción del viroma es el análisis bioinformático, que puede incluir distintos pasos. El primero de ellos suele ser un preprocesado de las muestras, donde se lleva a cabo la comprobación de la

calidad de la secuenciación y la posterior eliminación de las lecturas de baja calidad, que son aquellas que presentan repeticiones nucleotídicas, un tamaño menor del esperado o alguna indeterminación (Nooij, Schmitz, Vennema, Kroneman, & Koopmans, 2018). El análisis metagenómico suele incluir la eliminación de aquellas secuencias que no son el objetivo del estudio. Es decir, en casos en los que se realiza un análisis del viroma, se eliminan aquellas lecturas que pueden proceder de humano o de bacterias. La eliminación de estas secuencias se realiza para garantizar que el análisis se compone principalmente de secuencias virales y, a su vez, reducir significativamente la potencia informática y el tiempo requerido para el análisis (Santiago-Rodríguez & Hollister, 2020). Tras esto, el siguiente paso suele ser el ensamblaje de las lecturas restantes en secuencias contiguas de mayor longitud (*contigs*). En este punto, se pueden emplear dos estrategias distintas para la generación de *contigs*: la primera de ellas consiste en mapear las lecturas obtenidas en la secuenciación frente a una secuencia de referencia, mientras que la segunda se basa en un ensamblaje *de novo*, mediante la unión de las lecturas en base a los solapamientos entre ellas. En el caso de las aproximaciones metagenómicas para estudiar las comunidades virales, el método elegido suele ser el ensamblaje *de novo*, debido a que se desconoce qué virus se encuentran formando parte de esa comunidad. Posteriormente, el protocolo de búsqueda de virus en el análisis del viroma podría incluir un paso de búsqueda de las secuencias (bien lecturas individuales o *contigs*) en las bases de datos de referencia, utilizando, por ejemplo, la herramienta BLAST para filtrar aquellas secuencias de interés. Este procedimiento presenta un tiempo de análisis largo, por lo que se han desarrollado algoritmos para reducir este tiempo como alternativa al clásico alineamiento. Uno de ellos es emparejar *k*-meros, es decir, secuencias de longitud *k*, frente a una base de datos de

referencia. Otro enfoque es emplear un modelo probabilístico de alineamientos de secuencia múltiple, donde se emplean dominios de proteínas, lo que permite la detección de la homología más remota entre la referencia y la secuencia a analizar (Nooij et al., 2018). El último paso que suele realizarse en el análisis del viroma es la anotación. La anotación es el proceso de identificar elementos funcionales a lo largo de la secuencia de un genoma, dándole así un significado biológico (Abril & Castellano, 2019). En primer lugar, las características de interés (genes o elementos genómicos) se identifican (predicción de la característica) y, en segundo lugar, se asignan las funciones putativas de este gen y las relaciones taxonómicas (Nooij et al., 2018).

Existen algunos estudios que, empleando una muestra con composición viral conocida, comparan la eficiencia de recuperación de distintos métodos (Li et al., 2015), pero en cualquier caso, los diferentes pasos empleados en el protocolo dependerán de la naturaleza de la muestra y del objetivo del estudio.

2.6. Sesgos metodológicos en el estudio del viroma

La elección de los pasos a seguir en el protocolo de obtención del viroma de una muestra determinará el éxito de éste. Aun así, estos procedimientos pueden alterar la composición del viroma debido a que en cada paso existe un riesgo de sesgar la información presente. Como se ha mencionado anteriormente, en el protocolo de obtención del viroma suele haber un paso inicial de filtración. Este paso puede ayudar a retener los potenciales contaminantes de las muestras, como bacterias o células eucariotas, pero también puede impedir el paso de virus de gran tamaño, como los miembros de la familia *Poxviridae* (Parras-Moltó et al., 2018). En cuanto al paso de concentración, cuando se emplea la técnica de purificación de partículas

virales por centrifugación en gradiente de cloruro de cesio, se ha visto que puede ocurrir un sesgo hacia el aislamiento de aquellos virus que presenten una flotabilidad atípica, pese a que se obtienen muestras muy puras (Mirzaei et al., 2021).

Uno de los puntos más críticos en el protocolo de obtención del viroma de una muestra es la cantidad de genomas virales obtenida tras la extracción de los ácidos nucleicos. Como se ha explicado en apartados anteriores, en algunos casos el rendimiento de la reacción puede no ser suficiente para que se obtenga la cantidad de ácidos nucleicos requerida en los pasos sucesivos, por lo que se necesita un paso previo de amplificación. En concreto, se han mencionado dos técnicas de amplificación, que suelen ser las más empleadas en los protocolos de metagenómica. El uso de la técnica MDA puede dar lugar a la sobrerrepresentación de virus de pequeño tamaño y con genoma circular monocatenario, mientras que el empleo de la técnica SISPA presenta un sesgo debido a la preferencia de su región constante por secuencias similares en la muestra, produciéndose así una amplificación desigual (Parras-Moltó et al., 2018). Otro de los sesgos que pueden llevarse a cabo es en el análisis bioinformático, ya que este punto en el flujo de trabajo engloba múltiples pasos. Por ejemplo, uno de los pasos más comunes en el análisis es la eliminación de las secuencias bacterianas y eucariotas, con objeto de acotar el conjunto de secuencias donde se encuentran aquellas de origen viral. En este paso, sin embargo, siempre existe el riesgo de perder algunas lecturas virales (Jurasz, Pawłowski, & Perlejewski, 2021). Otro de los pasos que puede presentar un potencial sesgo es en el ensamblaje de las lecturas. Esto es debido a que el ensamblaje puede presentar cierta dificultad por las características intrínsecas de las poblaciones virales, como es la alta diversidad

presente (Mirzaei et al., 2021), haciendo que se puedan generar ensamblajes fragmentados.

Por otra parte, se ha documentado que los reactivos empleados a lo largo de todo el proceso son portadores de contaminación por ácidos nucleicos virales (Asplund et al., 2019), pero la gran mayoría de estos contaminantes tienen su propio perfil específico para determinados reactivos y kits, el denominado kitoma (Jurasz et al., 2021). Algunos miembros del kitoma son los CRESS-virus, los circovirus, algunos miembros de las familias *Tombusviridae*, *Totiviridae*, *Lentiviridae*, *Iridoviridae*, *Baculoviridae* y *Genomoviridae* (Jurasz et al., 2021; Porter, Cobbin, Li, Eden, & Holmes, 2021). La detección de estas secuencias virales artefactuales ha provocado que en algunos estudios se haya vinculado la presencia de determinados virus a una enfermedad (Lombardi et al., 2009; Ngoi et al., 2016; Xu et al., 2013), si bien estos resultados se han refutado posteriormente (Erlwein et al., 2011; Naccache et al., 2013; Ngoi et al., 2017; Paprotka et al., 2011; Smuts, Kew, Khan, & Korsman, 2014).

Por tanto, es necesaria la implementación de ciertas prácticas para evitar este tipo de dificultades al analizar los datos del viroma. Un ejemplo de esto es el uso de controles adecuados (Porter et al., 2021), es decir, controles negativos (agua o medio estéril que sigue todo el flujo de trabajo) para poder detectar los contaminantes esporádicos (Asplund et al., 2019) e incluso ciertos controles positivos, como la adición de una cantidad conocida de distintos virus controles que presentan características estructurales y genómicas diferentes para determinar la fiabilidad del protocolo (Santiago-Rodríguez & Hollister, 2020). Por otra parte, para evaluar si un genoma identificado es un potencial artefacto se pueden realizar distintas pruebas. Una de ellas es llevar a cabo la evaluación de la cobertura del genoma y de la profundidad de lectura,

ya que valores elevados de estos parámetros suponen una evidencia en contra de la presencia de artefactos (Asplund et al., 2019). Por otro lado, el estándar en las prácticas de identificación de nuevas especies en un trabajo de metagenómica viral suele ser la realización de una PCR. Para ello, se diseñan unos cebadores específicos para el aislado en cuestión y, como paso adicional, se puede realizar una posterior secuenciación por Sanger (Jurasz et al., 2021).

2.7 Aplicaciones de la metagenómica viral

Desde que los virus fueron descritos, se han estudiado principalmente por ser causantes de enfermedades en humanos, plantas y animales, pese a que estos suponen probablemente una pequeña proporción dentro de la virosfera, término que engloba la diversidad total de los virus presentes en el planeta. Esta visión tan sesgada de la virosfera ha sido debida a que el muestreo para la identificación de virus se ha dirigido al subconjunto de hospedadores que manifestaban enfermedad y a que el descubrimiento de nuevas especies virales ha sido tradicionalmente un proceso laborioso hasta la aparición de la metagenómica (Zhang, Shi, & Holmes, 2018). La aparición de las herramientas metagenómicas y su desarrollo durante las últimas décadas ha permitido que este enfoque pueda aprovecharse de múltiples maneras:

- Como herramienta de vigilancia para la salud comunitaria. En este contexto, se pueden emplear las herramientas proporcionadas por la metagenómica para vigilar los potenciales reservorios zoonóticos, es decir, aquellos donde existe un riesgo de que una especie viral pueda realizar un salto de la barrera de especie y ocasionar brotes en otras especies (French & Holmes, 2020). También se puede emplear esta herramienta para realizar el muestreo de aguas residuales en búsqueda de virus detectables en heces para obtener muestras comunitarias,

sobre todo cuando existen muchos individuos infectados sin muestrear, debido a que las infecciones asintomáticas son comunes, como es el caso del SARS-CoV-2 (Hill, Ruis, Bajaj, Pybus, & Kraemer, 2021). Otra aplicación de la metagenómica es el rastreo de los brotes para descubrir el origen y las causas de la propagación de estos, como ocurrió en la epidemia de Ébola en África Occidental de 2013-2016, que fue la primera vez en la que se aplicaron las herramientas metagenómicas a gran escala (Dudas et al., 2017; Gire et al., 2014). Desde entonces, en cada una de las crisis en salud pública causadas por virus se ha empleado esta estrategia de vigilancia: en las crisis causadas por el virus del Zika en América (Grubaugh et al., 2017), por el virus del síndrome respiratorio de Oriente Medio (MERS) en Arabia Saudí (Cotten et al., 2013; Sabir et al., 2016) o la actual pandemia causada por el SARS-CoV-2 (du Plessis et al., 2021).

- Como herramienta para caracterizar la diversidad viral. En este caso, la metagenómica ha servido para mejorar la comprensión de la extensión, de la estructura que presenta y de los procesos evolutivos que se han producido en la virosfera. El impacto más notable que ha supuesto el desarrollo de la metagenómica ha sido el hecho de mostrar la enorme diversidad que presenta la virosfera, ya que gracias a las herramientas metagenómicas se han descubierto una gran cantidad de virus (Shi, Zhang, & Holmes, 2018). Cada familia viral ha experimentado una expansión notable de la diversidad descrita debido a la revolución que ha supuesto la metagenómica (Zhang et al., 2018). Esto ha servido para dilucidar alguno de los “vacíos” evolutivos entre las familias y órdenes conocidos y para demostrar que la virosfera es muy diferente a lo que se pensaba en base a los agentes cultivables o causantes de

enfermedades (Zhang et al., 2018). Paradójicamente, conforme se van identificando más virus animales, se vuelve más claro que la mayor parte de la virosfera está por descubrir (Harvey & Holmes, 2022).

- Como herramienta para detectar las interacciones entre los virus y sus hospedadores. La metagenómica ha permitido la identificación de los agentes causantes de enfermedades, incluyendo aquellas enfermedades que no presentaban una causa conocida (Santiago-Rodriguez & Hollister, 2020). Sin embargo, como se ha mencionado anteriormente, la visión generalizada que se tiene de los virus es que son agentes causantes de enfermedad, debido a que la búsqueda de virus se ha llevado a cabo en individuos que presentan un fenotipo de enfermedad. Gracias a la metagenómica, se ha visto que cada organismo está colonizado por una gran diversidad de virus, el denominado viroma (Liang & Bushman, 2021), coexistiendo frecuentemente en ausencia de sintomatología.

Actualmente, existen numerosas aplicaciones que se pueden llevar a cabo gracias al desarrollo de la metagenómica. Este campo está transformando la comprensión de la diversidad viral, demostrando que es mucho más rica y compleja de lo que se imaginaba (Zhang et al., 2018).

3. Viroma humano

El viroma humano se define como el conjunto de todos los virus que se encuentran en la superficie y en el interior del cuerpo en ausencia de síntomas de infección clínicamente significativos (Rascovan, Duraisamy, & Desnues, 2016). Esto incluye a los bacteriófagos que infectan a las bacterias del microbioma, a los virus que infectan a otros microorganismos celulares, a los virus que infectan a las células humanas y a los elementos virales endógenos,

que se encuentran integrados en los cromosomas del hospedador (Liang & Bushman, 2021; Rascovan et al., 2016).

El primer estudio de viroma humano fue llevado a cabo por Breitbart y colaboradores a partir de muestras de heces. En él se descubrió que en estas muestras había poblaciones virales que se caracterizaban por ser ricas y diversas, poniendo de manifiesto la escasez de estudios previos que revelaran esta diversidad (Breitbart et al., 2003). Desde entonces, se han llevado a cabo numerosos trabajos que describen el viroma de distintas partes del cuerpo humano. Pese al creciente número de estudios publicados, existe una gran proporción de “materia oscura”, es decir, una elevada presencia de lecturas virales que no se parecen a nada descrito en las bases de datos (Aggarwala, Liang, & Bushman, 2017; Liang & Bushman, 2021; Shkoporov & Hill, 2019). La presencia de esta “materia oscura” no es un hecho llamativo, ya que en muestras de heces existen aproximadamente 10^9 partículas similares a los virus (VLP, por sus siglas en inglés) por gramo, que es un número particularmente elevado. En los estudios de metagenómica viral, las partículas que parecen virales no suelen verificarse como competentes para la replicación, por ello se acuña el término VLP (Liang & Bushman, 2021).

El viroma humano se caracteriza por presentar una gran heterogeneidad, ya que en cada parte del cuerpo se pueden encontrar ecosistemas muy diferentes, que implican una gran variabilidad de ambientes, algo que también se refleja en las comunidades microbianas que pueblan cada uno de ellos (Human Microbiome Project Consortium, 2012). Como consecuencia de esto, las comunidades virales difieren en términos de abundancia y composición en las diferentes zonas anatómicas del cuerpo: desde 10^9 VLP/g en los sitios de barrera corporal, como intestino o faringe, hasta 10^5 o 10^7 VLP/mL de sangre u orina, respectivamente (Haynes & Rohwer,

2011). Se piensa que la gran mayoría de estas VLP detectadas son bacteriófagos, cuya distribución está determinada, en buena medida, por las comunidades bacterianas presentes en cada región anatómica (Rascovan et al., 2016). Se estima, además, que el cuerpo humano puede contener hasta 10^{15} bacteriófagos (Haynes & Rohwer, 2011), la gran mayoría de los cuales pertenece al orden *Caudovirales* y a la familia *Microviridae* (Liang & Bushman, 2021). Los virus que infectan a las células humanas también son parte fundamental del viroma humano. En este caso, algunos virus pueden causar infecciones agudas mientras que otros virus pueden establecer una latencia a largo plazo. Por otra parte, existen virus que no se pueden asociar a ninguna enfermedad en particular, por lo que se denominan virus huérfanos (Kaczorowska & Van Der Hoek, 2020).

Recientemente, se han llevado a cabo numerosos estudios que han caracterizado el viroma humano en las distintas localizaciones anatómicas, mostrando la gran diversidad viral presente (**Figura 2**). En las diferentes localizaciones anatómicas estudiadas, los virus hallados suelen ser una mezcla de virus que se replican en las células humanas locales y virus que infectan a las bacterias que colonizan una región dada.

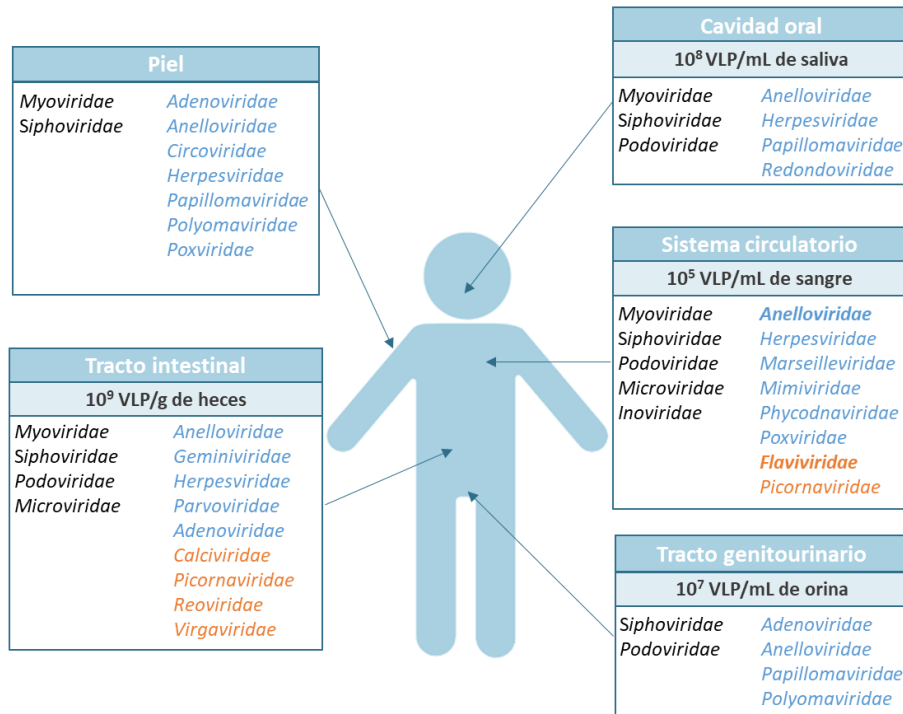


Figura 2: El viroma humano en diferentes partes del cuerpo. Resumen de las diferentes familias virales halladas en las distintas localizaciones anatómicas. En color negro, están identificadas las familias de bacteriófagos, en color azul, los virus eucariotas cuyo genoma es de ADN y en naranja, los virus eucariotas de ARN. Figura adaptada de Liang & Bushman, 2021.

3.1. Viroma del tracto gastrointestinal

En la mayoría de estudios de metagenómica viral del tracto gastrointestinal, los fagos presentan una abundancia mucho mayor que los virus eucariotas, debido a que se estima que existen más de mil especies bacterianas distintas en el tracto gastrointestinal (Shreiner, Kao, & Young, 2015). En el caso del intestino, que es la localización corporal que presenta un mayor número de estudios de viroma, las comunidades de fagos están formadas principalmente por miembros de las familias *Myoviridae*,

Siphoviridae y *Podoviridae*, del orden *Caudovirales* y por miembros de la familia *Microviridae* (Aggarwala et al., 2017; Lim et al., 2015; Reyes, Semenkovich, Whiteson, Rohwer, & Gordon, 2012). Por otro lado, la abundancia de virus eucariotas en el intestino de personas sanas es baja. Los virus de ADN que se han encontrado pertenecen a las familias *Anelloviridae*, *Geminiviridae*, *Herpesviridae*, *Parvoviridae* y *Adenoviridae* (Lim et al., 2015; Rascovan et al., 2016; Reyes et al., 2010), mientras que los virus de ARN descritos son de las familias *Calciviridae*, *Picornaviridae*, *Reoviridae* y algunos virus de plantas que parecen proceder de la alimentación, como miembros de la familia *Virgaviridae* (Liang, Conrad, et al., 2020; Lim et al., 2015; Popgeorgiev, Temmam, Raoult, & Desnues, 2013; Zhang et al., 2005). Una característica particular en los virus intestinales es que la mayoría de los virus presentes en el tracto intestinal, tanto los virus eucariotas como los bacteriófagos, carecen de envoltura lipídica, debido a que es poco probable que las membranas resistan a las condiciones que se dan a lo largo del tracto digestivo (Liang & Bushman, 2021).

Por otro lado, la cavidad bucal es un compartimento muy heterogéneo, que presenta comunidades virales significativamente diferentes en saliva o placa dental (Popgeorgiev, Temmam, et al., 2013). Pese a esto, la saliva se ha empleado como muestra para la caracterización de la mayoría de estudios de viroma (Abeles et al., 2014). Se estima que existen aproximadamente 10^8 VLP/mL de saliva en humanos sanos (Pride et al., 2012). En el caso de los fagos presentes en la saliva, se clasifican dentro de las familias *Siphoviridae*, *Myoviridae* y *Podoviridae* (Abeles et al., 2014; Pérez-Brocal & Moya, 2018; Pride et al., 2012) mientras que los virus eucariotas detectados pertenecen a las familias *Herpesviridae*, *Papillomaviridae*, *Anelloviridae* y *Redondoviridae* (Abbas et al., 2019).

3.2. Viroma de la piel

En el caso de muestras obtenidas a partir de frotis de piel se ha visto que hay una biomasa microbiana relativamente baja; pese a ello, en múltiples estudios se ha detectado la presencia de distintos virus. La mayoría de lecturas obtenidas de los estudios del viroma de la piel corresponde a bacteriófagos, pero muchas de ellas no se han podido clasificar al nivel de familia, quedando clasificadas dentro del orden *Caudovirales*. En una menor proporción, se han encontrado fagos pertenecientes a las familias *Myoviridae* y *Siphoviridae* (Hannigan et al., 2015). Por otra parte, la piel de individuos sanos puede albergar varios virus eucariotas, como miembros de las familias *Anelloviridae*, *Adenoviridae*, *Circoviridae*, *Herpesviridae*, *Poxviridae*, *Polyomaviridae* y *Papillomaviridae* (Foulongne et al., 2012; Hannigan et al., 2015), siendo esta última una de las familias más abundantes (Ma et al., 2014).

3.3. Viroma del tracto genitourinario

Los estudios metagenómicos del viroma del tracto genitourinario son escasos. Se estima que hay 10^7 VLP/mL de orina, donde la mayoría de los virus identificables son bacteriófagos (Liang & Bushman, 2021) de las familias *Siphoviridae* y *Podoviridae* (Garretto, Thomas-White, Wolfe, & Putonti, 2018). Por otra parte, en el tracto genitourinario existe una diversa comunidad de virus eucariotas compuesta por miembros de las familias virales *Adenoviridae*, *Anelloviridae*, *Papillomaviridae* y *Polyomaviridae* (Garretto et al., 2018; Liang & Bushman, 2021).

3.4. Viroma de la sangre

Por último, en estudios metagenómicos de muestras de plasma procedentes de sangre humana se ha determinado la presencia esporádica de secuencias virales de múltiples familias. En algunos de ellos, se han encontrado

tanto lecturas como VLP pertenecientes a bacteriófagos de las familias *Myoviridae*, *Siphoviridae*, *Podoviridae*, *Microviridae* e *Inoviridae* (Liang & Bushman, 2021; Rascovan et al., 2016; Zárata, Taboada, Yocupicio-Monroy, & Arias, 2017). Pese a ello, el origen de estos virus en el plasma no está del todo claro. Algunos autores sugieren que la presencia de fagos en el torrente sanguíneo puede deberse a la translocación de los fagos a través de las capas de células epiteliales del intestino, algo que parece ser frecuente y que presenta cierto grado de efectos inmunomoduladores (Górski et al., 2006; Moustafa et al., 2017).

Por otro lado, respecto a los virus eucariotas, se han identificado miembros de las familias *Anelloviridae*, *Herpesviridae*, *Picornaviridae*, *Poxviridae*, *Flaviviridae*, *Marseilleviridae*, *Mimiviridae* y *Phycodnaviridae* (Briese et al., 2015; Furuta et al., 2015; Law et al., 2013; Liang & Bushman, 2021; Popgeorgiev, et al., 2013; Stremlau et al., 2015). Sin embargo, existen dos grupos virales que causan infecciones crónicas y/o asintomáticas y que se encuentran comúnmente presentes en el plasma. En primer lugar, se estima que más del 50% de la población general está infectada por especies que pertenecen a la familia *Anelloviridae*, aunque esta prevalencia varía mucho entre las distintas poblaciones (Spandole, Cimponeriu, Berca, & Mihăescu, 2015). Por otra parte, la prevalencia del pegivirus humano (HPgV) perteneciente a la familia *Flaviviridae*, oscila entre el 1 y el 5% en los donantes de sangre sanos de los países desarrollados y aumenta hasta el 20% en los países en vías de desarrollo (Stapleton, Fong, Muerhoff, Bukh, & Simmonds, 2011). Dado que los anellovirus y HPgV se transmiten eficazmente por vía parenteral (Chivero & Stapleton, 2015; Kaczorowska & Van Der Hoek, 2020), su prevalencia es aún mayor entre los pacientes politransfundidos y los

consumidores de drogas intravenosas (Ataei, Emami Naeini, Khorvash, Yazdani, & Javadi, 2012; Bhattarai & Stapleton, 2012).

3.5. Muestras de sangre

La sangre humana y sus productos derivados representan una necesidad constante para las transfusiones de sangre y el tratamiento médico. Sin embargo, la sangre implica un importante reservorio viral y alguno de estos virus puede ser patógeno (Popgeorgiev, Temmam, et al., 2013). El uso de muestras de sangre procedentes de donantes sanos para la determinación del viroma de plasma es interesante debido a distintas razones. La primera de ellas es la facilidad para la obtención de las muestras y datos gracias a que los donantes suelen ser recurrentes en cuanto a las donaciones. En segundo lugar, el grupo de donantes de sangre representa una muestra apropiada de la población general sana, que potencialmente puede portar virus desconocidos. Por ello, la descripción de los virus presentes en la sangre tiene consecuencias directas para la salud pública (Popgeorgiev, Temmam, et al., 2013; Sauvage & Eloit, 2016), además de contribuir a la creciente evidencia que muestra que las interacciones mutualistas/comensales entre los virus y sus hospedadores parecen ser la regla y no la excepción (French & Holmes, 2020).

3.6. Anelovirus

Los anelovirus son los virus eucarióticos más abundantes del viroma de la sangre y, además, se encuentran presentes en otras localizaciones corporales, tal y como se ha descrito en apartados anteriores. Se descubrieron en 1997 mientras se buscaba un agente viral responsable de un caso de hepatitis con etiología desconocida tras una transfusión de sangre (Nishizawa et al., 1997). Desde entonces, se ha descrito una enorme diversidad dentro de la familia *Anelloviridae*. Pese a esto, no se ha encontrado ninguna relación clara

entre los anelovirus y alguna enfermedad humana, por lo que son considerados virus huérfanos (Kaczorowska & Van Der Hoek, 2020). A pesar de que se ha descubierto un gran número de anelovirus en animales salvajes y domésticos (Cibulski et al., 2014; de Souza et al., 2018; Hrazdilová et al., 2016; Li et al., 2017; Ng et al., 2011; Shi et al., 2015), cabe destacar que tan sólo un miembro del género *Gyrovirus* es agente causante de enfermedad en gallinas (Li et al., 2017).

Los anelovirus presentan un genoma de ADN monocatenario de polaridad negativa que oscila desde aproximadamente 2,0 kb hasta 3,9 kb. Según el Comité Internacional de Taxonomía de Virus (ICTV, por sus siglas en inglés) la familia *Anelloviridae* comprende 155 especies que se agrupan dentro de 31 géneros distintos (Varsani et al., 2021). Hasta ahora, se han descrito tres géneros que infectan a humanos (Spandole et al., 2015): *Alphatorquevirus* o Torque Teno Virus (TTV), *Betatorquevirus* o Torque Teno Mini Virus (TTMV) y *Gammatorquevirus* o Torque Teno Midi Virus (TTMDV). Recientemente, los géneros TTV, TTMV y TTMDV se han subdividido en 26, 38 y 15 especies, respectivamente (Varsani et al., 2021), aunque se espera que esta diversidad aumente conforme se identifiquen nuevos aislados. Esta reciente clasificación se basa en el estudio de la secuencia nucleotídica de la ORF1 completa, que se analiza empleando una nueva metodología basada en identidades de secuencias por pares (Muhire, Varsani, & Martin, 2014). De este modo, se ha establecido que un valor del 69% de identidad sirve como criterio de demarcación de las especies de esta familia. Los genomas de los miembros del género TTV presentan un tamaño de 3,6-3,9 kb, mientras que los miembros de TTMV presentan un tamaño entre 2,8-2,9 kb y los miembros de TTMDV presentan un genoma de aproximadamente 3,2 kb de longitud. En los tres casos, sus genomas tienen una organización muy similar: presentan una región

no codificante, en la que hay una sección muy conservada de aproximadamente 150 nucleótidos, y una región codificante que comprende diferentes pautas abiertas de lectura (ORF, por sus siglas en inglés) solapantes (**Figura 3**). La ORF1 es la que tiene un tamaño mayor (codifica entre 700 y 800 aminoácidos) y se piensa que presenta función estructural (Arze et al., 2021; Kaczorowska & Van Der Hoek, 2020).

Los miembros de la familia *Anelloviridae* presentan una enorme heterogeneidad genética (Focosi, Antonelli, Pistello, & Maggi, 2016), tanto a nivel global como intra-individuo (Arze et al., 2021). De hecho, la diversidad presente en esta familia es llamativa al compararse con otras familias virales (Kaczorowska & Van Der Hoek, 2020). Existen diferentes hipótesis para explicar esta enorme diversidad. La primera de ellas es que la familia *Anelloviridae* es una familia que presenta una larga historia de coevolución con los primates, por lo que millones de años de evolución han dado lugar a la diversidad actual (Kaczorowska & Van Der Hoek, 2020). Por otra parte, otra de las hipótesis que se plantea es que los anelovirus presentan una rápida evolución debida a una alta frecuencia de recombinación entre las distintas cepas del virus (Fahsbender et al., 2017), aunque la presencia de recombinación no puede explicar actualmente la enorme variabilidad que existe entre ellos (Kaczorowska & Van Der Hoek, 2020).

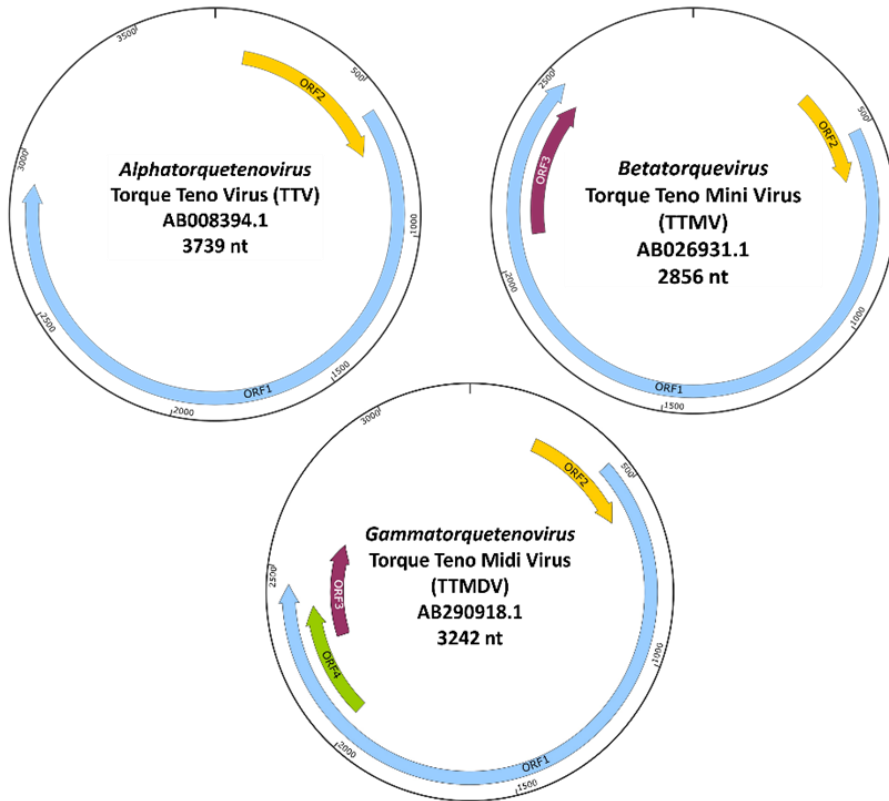


Figura 3: Representación esquemática de la organización de los genomas virales de las especies tipo de los géneros *Alphatorquevirus*, *Betatorquevirus* y *Gammatorquevirus*. En los tres géneros, el genoma consiste de varias ORF solapantes en la región codificante. Las distintas ORF se encuentran representadas por flechas coloreadas. Figura creada con SnapGene.

Como se ha descrito en apartados anteriores, los anelovirus se han identificado en una amplia variedad de muestras clínicas, por lo que se ha visto que son los virus eucariotas más abundantes del viroma humano (Kaczorowska & Van Der Hoek, 2020). Sin embargo, no existe todavía un sistema de cultivo celular para su propagación y, de hecho, ni siquiera existe un modelo animal que pueda proporcionar información sobre las interacciones que se dan entre

los anelovirus y sus huéspedes (Liang & Bushman, 2021). En el caso de los humanos, no se ha demostrado que exista una relación entre la presencia de los miembros de la familia *Anelloviridae* y algún efecto perjudicial sobre la salud (Spandole et al., 2015). Se ha visto que las primeras infecciones por anelovirus ocurren en la infancia (Lim et al., 2015; Tyschik, Rasskazova, Degtyareva, Rebrikov, & Sukhikh, 2018) y que se siguen detectando en niños y adultos de todas las edades. Por otra parte, parece que los anelovirus están reprimidos por la inmunidad del huésped (Focosi et al., 2016), ya que su carga viral aumenta a la vez que los niveles de inmunosupresión del huésped, como se ha visto en algunos estudios que comparan el viroma de personas sanas con el de pacientes que presentan alguna condición de inmunosupresión (De Vlaminc et al., 2013; L. Li et al., 2013).

3.7. Pegivirus

HPgV, también conocido como GB virus C y hepatitis G virus, es un virus linfotrópico perteneciente a la familia *Flaviviridae* y al género *Pegivirus* (Bhattarai & Stapleton, 2012) y es el virus humano que presenta una relación filogenética más estrecha con el virus de la hepatitis C (Mohr & Stapleton, 2009). Su descubrimiento ocurrió de manera casi simultánea en dos laboratorios distintos: en el primero de ellos, se describió un virus nuevo en el suero de un paciente del oeste de África con hepatitis que no estaba causada por los virus de la hepatitis A y E (Simons et al., 1995), y que fue bautizado como GB virus C. Por otro lado, Linnen y colaboradores descubrieron un virus en el plasma de un paciente con hepatitis crónica y lo nombraron hepatitis G virus (Linnen et al., 1996). El análisis de la secuencia de ambos virus permitió identificar un 96% de homología entre ambos, llegando a la conclusión de que eran dos aislados del mismo virus (Bhattarai & Stapleton, 2012).

HPgV tiene un genoma de ARN de sentido positivo de 9,3 kb, y se encuentra organizado de forma similar al virus de la Hepatitis C. El genoma de este virus se traduce para producir una única poliproteína de aproximadamente 3.000 aminoácidos de longitud (**Figura 4**). Esta poliproteína es procesada tras la traducción por proteasas celulares y virales dando lugar a proteínas estructurales y no estructurales (Bhattarai & Stapleton, 2012; Soliman et al., 2019). En la actualidad, se ha propuesto que HPgV consta de seis genotipos que muestran diferentes patrones de distribución geográfica y, a su vez, múltiples subtipos (Feng et al., 2011; Ghai et al., 2013): el genotipo 1 es predominante en el oeste de África, el genotipo 2 se ha identificado mayoritariamente en Norteamérica y Europa, el genotipo 3 en Asia, el genotipo 4 en el sudeste asiático, el genotipo 5 en Sudáfrica y el genotipo 6 en Indonesia (Soliman et al., 2019).

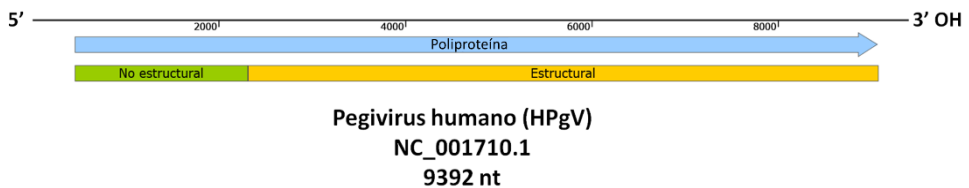


Figura 4: Representación esquemática del genoma de HPgV. El genoma de HPgV se traduce en una única poliproteína, que de manera postraducciona es procesada mediante proteasas celulares y virales. Figura creada con SnapGene.

En el caso de los pegivirus, existen algunas asociaciones débiles con algunas enfermedades, ya que HPgV parece ser protector frente a la infección por el virus de la inmunodeficiencia humana 1 (VIH-1) (Bhattarai & Stapleton, 2012; Chivero & Stapleton, 2015) debido a que disminuye la activación inmunitaria de las células T, células B, células *natural killer* y monocitos, lo que produce una reducción en la progresión de la infección por VIH-1 (Chivero & Stapleton, 2015). También se ha observado que en pacientes infectados con

Ébola, la coinfección con HPgV se ha asociado a una mayor supervivencia (Lauck et al., 2015). Estas evidencias sugieren que HPgV y el huésped presentan una relación mutuamente beneficiosa (Bhattarai & Stapleton, 2012). Al igual que ocurre con los anelovirus, la Organización Mundial de la Salud (OMS) no recomienda el cribado sanguíneo de HPgV, dado que no está asociado con ninguna enfermedad conocida (Bhattarai & Stapleton, 2012), razón por la cual es considerado también un virus huérfano (Fernández-Ruiz et al., 2021).

3.8. Relación del viroma con la fisiología humana

Además de la ubicación anatómica, como se ha explicado en apartados anteriores, existen otros factores que influyen sobre la composición del viroma. Algunos de estos factores son específicos del hospedador y otros son específicos del entorno del hospedador (Popgeorgiev, Temmam, et al., 2013).

Algunos ejemplos de factores específicos del hospedador pueden ser la edad, el sexo, la dieta o el estado inmunológico del individuo (Liang & Bushman, 2021; Popgeorgiev, Temmam, et al., 2013). En cuanto a la dieta, existen estudios que muestran diferencias en la composición del viroma intestinal de niños cuya alimentación está basada en lactancia materna frente a niños alimentados con leche de fórmula (Liang, Zhao, et al., 2020). Estas diferencias en la composición del viroma también se han identificado en individuos adultos (Minot et al., 2011). Está descrito que el estado inmunológico del individuo influye sobre la composición del viroma. Hay estudios en los que se ha demostrado que pacientes inmunodeprimidos presentan cargas virales altas circulantes en plasma de miembros de la familia *Anelloviridae*, en comparación con las que presentan individuos sanos (De Vlaminc et al., 2013; Li et al., 2013). Esta diferencia radica en que los pacientes

inmunodeprimidos presentan un control inmunológico reducido de la replicación de estos virus (Li et al., 2013). Por otra parte, factores demográficos, como la edad y el sexo del individuo también influyen sobre la composición del viroma. En distintos estudios, se han identificado diferencias estadísticamente significativas en la carga viral de anelovirus en plasma de hombres y mujeres, siendo mayor en hombres (Haloschan et al., 2014; Moustafa et al., 2017). También se ha visto que conforme aumenta la edad, aumenta la carga viral de anelovirus en plasma (Haloschan et al., 2014; Moustafa et al., 2017) así como la riqueza viral (definida como el número de especies distintas) de los anelovirus presentes en el viroma del intestino (Gregory et al., 2020).

También cabe mencionar que existen factores específicos del entorno del hospedador que pueden influir sobre el viroma. Diferentes estudios han establecido una relación entre la ubicación geográfica y la composición del viroma, ya que existen patrones de distribución geográfica para las distintas especies virales (Liang & Bushman, 2021; Popgeorgiev, Temmam, et al., 2013). Otros factores que influyen sobre la composición del viroma son la proximidad de animales o distintos factores sociales y culturales, que pueden llevar asociados variaciones en la exposición a nuevos virus (Popgeorgiev, Temmam, et al., 2013; Roberts et al., 2021).

Por otro lado, el viroma puede influir sobre la salud del individuo de diversas maneras. Algunos de estos efectos pueden ser inducidos por una interacción directa de los virus con el hospedador, mientras que otros pueden implicar la interacción del viroma con otros componentes del microbioma (Zárate et al., 2017). Las comunidades de bacteriófagos del viroma probablemente sirvan como amortiguadores, modulando la composición bacteriana, así como de almacenamiento genómico para distintas capacidades

metabólicas del microbioma (Rascovan et al., 2016). Por otra parte, otros estudios han demostrado que los fagos representan un importante reservorio de genes de virulencia bacteriana (Popgeorgiev, Temmam, et al., 2013; Virgin, 2014). Los fagos, por tanto, desempeñan un doble papel: controlan la población microbiana, pero a su vez, contribuyen a la patogenicidad y resistencia bacteriana a través de la transferencia horizontal de genes (Popgeorgiev, Temmam, et al., 2013). Los virus eucariotas también parecen jugar un papel importante en la fisiología del hospedador. Una de las hipótesis planteadas es que los miembros del viroma promueven la salud de manera indirecta, interviniendo en la maduración del sistema inmunitario de los recién nacidos (Tyschik et al., 2018; Virgin, Wherry, & Ahmed, 2009) y participando en la eliminación de patógenos virales mediante inmunidad cruzada y/o directamente mediante interferencia viral (Rascovan et al., 2016).

La relación entre el hospedador sano y el viroma puede ser considerada como una red compleja de interacciones en las que el peso de cada factor se sitúa en un punto de equilibrio metaestable. Esto permite que los virus y los humanos coexistan de forma natural, por lo que un cambio en alguno de los parámetros de esta relación podría conducir al desarrollo de alguna enfermedad, o bien a la eliminación del virus del organismo (Popgeorgiev, Temmam, et al., 2013). Además, conocer el viroma de individuos sanos puede tener distintas aplicaciones, como el uso de ciertos virus en calidad de biomarcadores del estado de salud del individuo o como predictores de diagnósticos adversos (Arze et al., 2021; Blatter et al., 2018; De Vlaminck et al., 2013).

Objetivos

Objetivos

El objetivo global de esta Tesis es la caracterización a nivel poblacional del viroma de plasma de individuos aparentemente sanos. Esto podría contribuir al establecimiento de estrategias de vigilancia epidemiológica dirigidas a la identificación de virus potencialmente implicados en la emergencia futura de enfermedades infecciosas. En este contexto, los objetivos concretos de este trabajo son:

Primer objetivo: recepción de cientos de muestras de plasma, establecimiento del flujo de trabajo para el enriquecimiento de la fracción viral, obtención y amplificación de los ácidos nucleicos presentes en las muestras de plasma y generación de librerías para la secuenciación.

Segundo objetivo: Implementación del protocolo bioinformático dirigido a la caracterización del viroma de plasma. En este protocolo se incluirán la utilización de herramientas de clasificación taxonómica a partir de las lecturas crudas obtenidas por secuenciación masiva o tras la generación de *contigs*, así como los análisis filogenéticos necesarios para caracterizar la variabilidad viral existente.

Tercer objetivo: caracterización del viroma individual de donantes de sangre. El objetivo global de la Tesis, como se ha comentado previamente, era llevar a cabo un análisis poblacional, para lo cual se requería analizar un elevado número de muestras. Dada la magnitud del estudio, el análisis por secuenciación masiva sólo sería factible si las muestras fueran procesadas como grupos y no individualmente. Sin embargo, para un número reducido de muestras, también resultaba interesante la caracterización del viroma a nivel individual, lo que permitiría establecer si existen asociaciones entre el viroma de la sangre y parámetros demográficos generales, como el sexo o la edad.

Material y métodos

Material y métodos

Esta Tesis está compuesta por tres capítulos, en los que se describe el viroma de muestras de plasma procedentes de donantes sanos. El capítulo 1, consta de dos partes y, en la primera de ellas, se ha llevado a cabo un experimento piloto en el que se ha comparado la eficiencia de recuperación de los ácidos nucleicos virales presentes en las muestras de plasma mediante dos procedimientos distintos. En concreto, se ha empleado un kit de extracción directa de ácidos nucleicos virales frente a un protocolo que involucra la centrifugación de las muestras a alta velocidad, con el fin de precipitar los virus presentes en las muestras. Los resultados obtenidos en esta primera parte del (que se encuentran en el capítulo 1 apartado 1.1) han servido para determinar los pasos a seguir en el flujo de trabajo para el procesado de las muestras de plasma empleadas a lo largo de la Tesis. En la segunda parte de este capítulo, se ha llevado a cabo el análisis del viroma presente en 120 muestras de plasma, las cuales se han agrupado en 12 mezclas compuestas por aproximadamente 10 muestras individuales. En el capítulo 2, se ha llevado a cabo el análisis del viroma en un número mayor de muestras, en concreto, en 587 muestras de plasma procedentes de sangre de donantes sanos. Estas muestras se han dividido en 60 mezclas de plasmas distintas, cada una de las cuales estaba compuesta entre ocho y trece muestras de plasma de donantes diferentes. En este análisis se ha descrito una enorme diversidad de los aislados pertenecientes a la familia *Anelloviridae*. Además, en este estudio, aunque en una prevalencia menor del 3%, también se ha determinado la presencia de virus de ARN. En concreto, se han detectado secuencias virales que pertenecen al género *Pegivirus*, de la familia *Flaviviridae*. En el análisis filogenético realizado a estos genomas, se han identificado varias secuencias que presentaban una posición intermedia entre clusters, por lo que existe la posibilidad de que pudieran ser aislados recombinantes. Para analizar con

mayor profundidad estos sucesos, se ha llevado a cabo un análisis de recombinación, que se explicará en el apartado 4a de Material y Métodos. Por último, en el capítulo 3 de la Tesis se ha analizado el aneloma, es decir, el conjunto de anelovirus presentes en el viroma de un individuo. En este capítulo, el procesado de las muestras se ha abordado de manera diferente, ya que únicamente se ha llevado a cabo la extracción de la fracción de ADN del viroma. Dado que se observó que la prevalencia de virus de ARN en el plasma de la población sana era baja, su probabilidad de detección en las 50 muestras individuales analizadas en este capítulo sería particularmente reducida. En consecuencia, y para aumentar la potencial detección de virus de ADN, se optó por obviar la fracción de virus de ARN. Esta aproximación más restrictiva ha posibilitado la descripción de una gran diversidad de miembros de la familia *Anelloviridae*, además de la identificación de genomas de anelovirus recombinantes, mediante la metodología que se explica en el apartado 4b de Material y Métodos.

1. Procesado de las muestras de sangre

Las muestras de sangre empleadas en este trabajo han sido obtenidas de donantes sanos del Centro de Transfusiones de la Comunidad Valenciana. Todos los donantes que participaron en el proyecto firmaron un consentimiento informado de acuerdo con la declaración de Helsinki. El protocolo fue aprobado por el comité de ética de la Universitat de València (IRB No. H1489496487993). La información demandada a los donantes ha consistido en sexo, edad, historia personal y familiar del donante, hábitos de salud y posibles problemas relacionados con la salud.

Las muestras de sangre han sido extraídas en tubos de 6 mL que contienen el quelante EDTA para evitar su coagulación. El procesado de las

muestras para la obtención del plasma se ha realizado el mismo día de su recogida. Para ello, la sangre de varios tubos correspondientes al mismo donante se ha juntado en un único tubo y se ha sometido a una centrifugación a $10,000 \times g$ durante 10 minutos a $4 \text{ }^\circ\text{C}$. La fracción superior, correspondiente al plasma, se ha trasvasado a un nuevo tubo, que se ha almacenado a $-80 \text{ }^\circ\text{C}$ hasta su uso.

2. Prueba piloto: comparación entre la concentración por extracción directa vs centrifugación

El objetivo de esta prueba piloto es comparar la eficiencia de concentración de los ácidos nucleicos virales mediante distintas técnicas: centrifugación vs un kit de extracción directa de ácidos nucleicos virales (**Figura 5**).

Para realizar esta prueba, se ha empleado una mezcla de 25 mL de plasma procedente de 10 donantes distintos (es decir, 2,5 mL de cada donante). Para determinar la eficiencia de la recuperación se han añadido a esta muestra 10^3 PFU/mL de ΦX174 , MS2 y virus Vaccinia (VV) y 10^4 PFU/mL del virus de la estomatitis vesicular (VSV). Estos 25 mL de plasma se han dividido en dos fracciones distintas: la primera fracción se ha filtrado a través de un filtro de acetato de celulosa con un tamaño de poro de $0,45 \text{ }\mu\text{m}$ (Sartorius Minisart®), mientras que la segunda fracción se ha filtrado a través de un filtro de polietersulfona (PES) con un tamaño de poro de $1 \text{ }\mu\text{m}$ (Whatman® Puradisc). A su vez, cada una de estas fracciones se ha vuelto a separar: 1 mL se ha empleado para el protocolo de extracción directa del plasma y 10 mL para el protocolo de concentración por centrifugación (**Figura 5**).

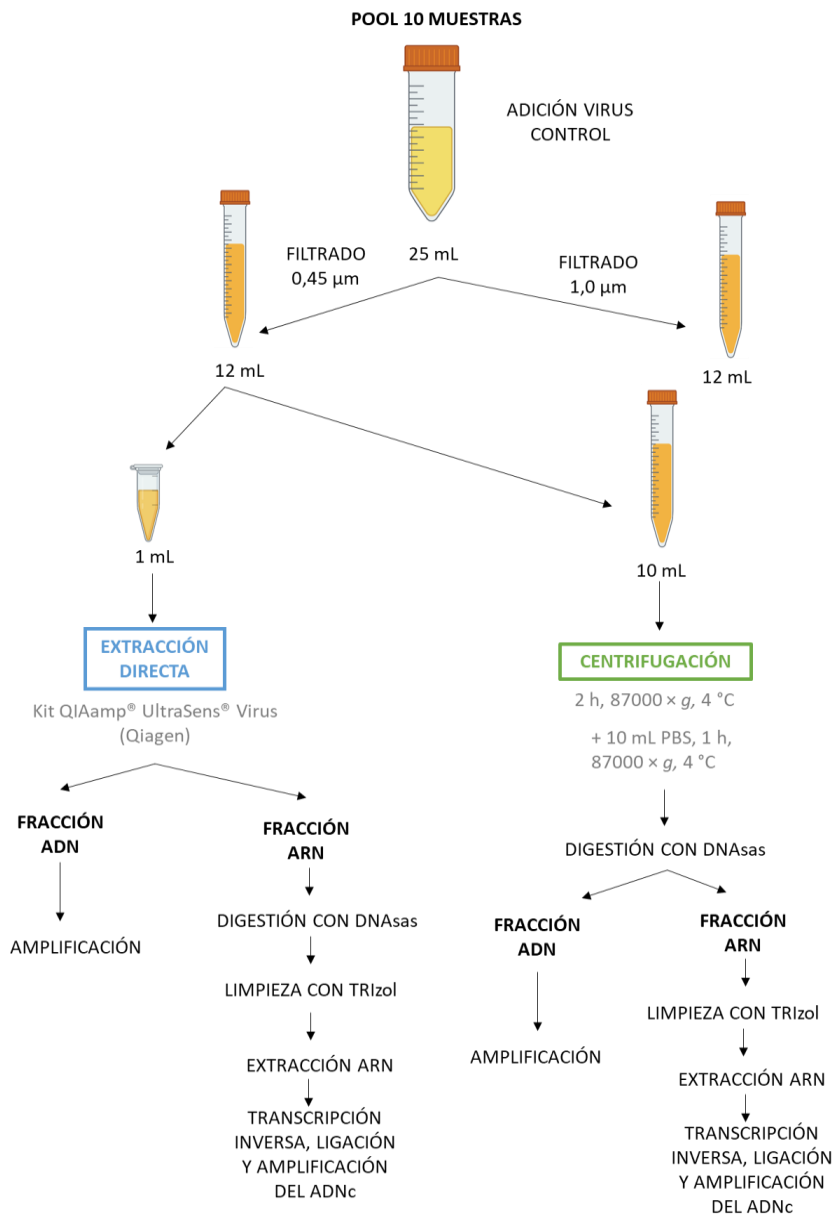


Figura 5: Esquema del flujo de trabajo seguido en el experimento piloto. En este experimento se han testado dos procedimientos distintos para la concentración de partículas virales. La muestra filtrada con el filtro de 1 µm también ha seguido el mismo procesado, pero no se refleja en el esquema. Parte de la figura ha sido creada con BioRender.com

2.1. Concentración por extracción directa del plasma

La extracción directa de los ácidos nucleicos virales se ha realizado a partir de 1 mL de plasma filtrado y para ello se ha empleado el kit QIAamp® UltraSens® Virus (Qiagen) siguiendo las instrucciones del fabricante. Este kit está optimizado para la extracción de ácidos nucleicos virales a partir de muestras biológicas líquidas libres de células, como el plasma o el suero. Para realizar la concentración de los ácidos nucleicos virales, este kit presenta una columna a la cual se unen selectivamente los ácidos nucleicos presentes en el plasma. A partir del volumen eluido de la columna, se han separado dos fracciones: una se ha empleado para la amplificación del ADN viral presente en la muestra, como se explicará posteriormente (muestras **M1** y **M2**), mientras que la otra fracción se ha empleado para realizar la extracción del ARN viral (muestras **M5** y **M6**). Para ello, en primer lugar, se ha realizado una digestión con el kit Turbo DNA free (Ambion), después se ha realizado una limpieza con TRIzol LS (Invitrogen) y el ARN se ha extraído con el kit QIAamp® Viral RNA Mini (Qiagen).

2.2. Concentración por centrifugación

Los 10 mL de plasma filtrados empleados para esta prueba se han sometido a una centrifugación a $87.000 \times g$ a 4°C durante 2 h. Una vez transcurrido este tiempo, se ha eliminado el sobrenadante, se han añadido 10 mL de PBS 1X y se ha centrifugado una hora adicional bajo las mismas condiciones de velocidad y temperatura. Tras esto, se ha eliminado de nuevo el sobrenadante y el sedimento se ha resuspendido en 245 μL del tampón de digestión del kit Turbo DNA free (Ambion). Una vez el sedimento se ha disgregado completamente, se ha añadido un cóctel de enzimas de digestión, que ha consistido en 5 μL de TURBO DNasa (Turbo DNA free kit (Ambion)), 2

Material y métodos

μL de benzonasa (Sigma) y 2 μL de nucleasa micrococcal (NEB), que se ha incubado durante 1 h a 37 °C. Transcurrido este tiempo, se ha detenido la reacción de digestión añadiendo el reactivo de inactivación de la TURBO DNasa (*stop reagent*, Turbo DNA free kit (Ambion)) a la muestra, se ha incubado a temperatura ambiente durante 3 min y se ha sometido a una centrifugación a $10.000 \times g$, 3 min. El sobrenadante obtenido se ha transferido a dos tubos: una fracción (40 μL , muestras **M3** y **M4**) se ha empleado en la extracción del ADN viral presente en la muestra y el volumen restante (aproximadamente 200 μL) se ha empleado para la extracción de ARN (muestras **M7** y **M8**). Un control negativo (**M9**) sólo con 10 mL de PBS 1X se ha sometido a todo este proceso de extracción.

En este caso, la extracción del ADN y del ARN se ha realizado con el kit QIAamp® Viral RNA Mini (Qiagen), pero para la extracción de la fracción de ARN se ha realizado un paso adicional que consiste en una limpieza previa de la muestra con TRIzol.

2.3. Obtención de las muestras necesarias para la construcción de librerías a partir de la fracción de ADN

Para la obtención de las muestras previas a la construcción de librerías procedentes de la fracción de ADN de las muestras **M1**, **M2**, **M3**, **M4** y **M9** se ha realizado una amplificación del ADN viral con el kit TruePrime WGA (Qiagen) siguiendo las instrucciones del fabricante. El producto de la amplificación se ha almacenado a -20 °C hasta su posterior uso para la construcción de librerías de secuenciación.

2.4. Obtención de las muestras necesarias para la construcción de librerías a partir de la fracción de ARN

Para obtener las muestras previas a la construcción de las librerías procedentes de la fracción de ARN, se ha realizado la síntesis del ADNc con la

enzima transcriptasa inversa SuperScript™ III (SSIII) (Invitrogen). Para llevar a cabo esta reacción, se ha realizado un mix con 8 µL del ARN, 1 µL de hexámeros (50 µM) (Invitrogen) y 1 µL de dNTPs (10 mM), que se ha incubado a 75 °C durante 5 min. Transcurrido este tiempo, la reacción se ha incubado en hielo durante 5 min, tras lo cual se ha añadido 10 µL de mix 2X de la enzima SSIII. Esta nueva mezcla ha consistido en 2 µL de tampón 10X de la enzima SSIII (Invitrogen), 4 µL de MgCl₂ 25 mM, 2 µL de 1,4-ditiotreitol (DTT), 1 µL de RNaseOUT (Invitrogen), 1 µL de SSIII (Invitrogen) y 0,5 µL de dimetilsulfóxido (DMSO) (Sigma-Aldrich). Este mix final se ha incubado a 25 °C durante 10 min, 45 °C durante 90 min y 95 °C durante 5 min. El ADNc obtenido a partir de esta reacción puede almacenarse a -20 °C o bien emplearse inmediatamente para los siguientes pasos. En este caso, se ha empleado el kit Quantitect Whole Transcriptome (Qiagen) según las instrucciones del fabricante. En este kit se realiza una reacción de ligación de los fragmentos de ADNc generados previamente a su amplificación. Brevemente, se ha realizado una mezcla de ligación, que consiste en 6 µL de tampón de ligación, 2 µL de reactivo de ligación, 1 µL de enzima de ligación 1 y 1 µL de enzima de reacción 2. Estos 10 µL de mix se han añadido a 10 µL procedentes de la reacción de retrotranscripción y se han incubado a 22 °C durante 2 h.

Una vez terminada la reacción de ligación, se ha realizado la reacción de amplificación, que se lleva a cabo mediante la técnica MDA. Para ello, se ha preparado una mezcla de amplificación que ha consistido en 29 µL de tampón de reacción REPLI-g Midi y 1 µL de ADN polimerasa REPLI-g (Quantitect Whole Transcriptome (Qiagen)). Estos 30 µL se han añadido a la reacción de ligación y se ha incubado durante 2 h a 30 °C. Transcurrido este tiempo, se ha detenido la reacción incubándola 5 min a 95 °C. El producto de amplificación se ha

almacenado a -20 °C hasta su posterior uso para la construcción de librerías de secuenciación (muestras **M5**, **M6**, **M7**, **M8** y **M9**).

2.5. Secuenciación masiva

Para este estudio piloto, se han empleado los 8 productos de amplificación obtenidos junto con el control del proceso de concentración por centrifugación (**Tabla 1**) para la preparación de librerías empleando el kit de Nextera XT DNA (Illumina) con 15 ciclos de amplificación. La secuenciación se ha realizado con un dispositivo MiSeq con el que se han generado unos 25 millones de secuencias.

Tabla 1: Resumen de las muestras que se han sometido a secuenciación, la fracción de la que se han obtenido y el tratamiento al que han sido sometidas.

Muestra	Fracción	Tratamiento
M1	ADN	Filtro de 0,45 µm y extracción directa del plasma
M2	ADN	Filtro de 1 µm y extracción directa del plasma
M3	ADN	Filtro de 0,45 µm y centrifugación
M4	ADN	Filtro de 1 µm y centrifugación
M5	ARN	Filtro de 0,45 µm, extracción directa del plasma, kit ligación
M6	ARN	Filtro de 1 µm, extracción directa del plasma, kit ligación
M7	ARN	Filtro de 0,45 µm, centrifugación, kit ligación
M8	ARN	Filtro de 1 µm, centrifugación, kit ligación
M9	ADN y ARN	Control, filtro de 1 µm y centrifugación

2.6. Análisis bioinformático

La clasificación taxonómica de las lecturas obtenidas tras la secuenciación se ha llevado a cabo con el programa Centrifuge (Kim, Song, Breitwieser, & Salzberg, 2016), bajo los parámetros de clasificación que se explicarán en la sección 3.4 de Material y Métodos.

3. Protocolo de obtención de ácidos nucleicos virales para secuenciación

3.1. Obtención del plasma procedente de muestras de sangre total

En los tres capítulos de esta Tesis se han empleado muestras de plasma procedentes de donaciones de sangre total. En los dos primeros capítulos, el volumen de sangre recogido para cada donante ha sido de aproximadamente 12 mL, mientras que en el capítulo 3 el volumen ha sido de 24 mL. En los tres casos, la sangre total se ha centrifugado a $10.000 \times g$, 4 °C, 10 min. La fracción superior, que se corresponde con el plasma, se ha transvasado a un nuevo tubo.

Las muestras pertenecientes a los capítulos 1 y 2 de la Tesis consisten en mezclas de plasma de distintos donantes, obtenidas tras juntar el plasma de aproximadamente 10 individuos. Estas mezclas son heterogéneas, es decir, no se ha discriminado por sexo o edad. Para la constitución de cada mezcla, se ha empleado 1 mL de plasma de cada donante. Este mililitro se ha filtrado utilizando un filtro de PES con un tamaño de poro de 1 μm (Whatman® Puradisc). Las diferentes mezclas generadas se han almacenado a -80 °C hasta su posterior procesado. El volumen restante de plasma de cada donante se ha empleado para realizar distintas alícuotas y se ha almacenado a -80 °C.

En el caso de las muestras del capítulo 3, como se ha mencionado anteriormente, se ha partido de un volumen de sangre superior, por lo que, para cada donante, se han obtenido unos 12 mL de plasma tras la centrifugación. En este caso, cada muestra procede de un único donante y se ha procesado de manera análoga a lo descrito para los otros dos capítulos de la Tesis: se han filtrado 10 mL de plasma utilizando un filtro de PES de 1 μm

(Whatman® Puradisc) y se han almacenado a -80 °C hasta su procesado. El volumen restante de plasma se ha alicuotado y se ha almacenado a -80°C.

3.2. Extracción de ácidos nucleicos virales y amplificación

Los resultados obtenidos en el experimento piloto (capítulo 1, apartado 1.1) han servido para el establecimiento del protocolo que se ha seguido para las muestras empleadas en este trabajo con el fin de concentrar, extraer y amplificar los ácidos nucleicos virales presentes en el plasma.

En primer lugar, se añaden los virus controles a los 10 mL de plasma filtrado. La adición de virus controles en estas muestras se ha hecho para verificar que el proceso de concentración y extracción ha funcionado de manera correcta. En cada capítulo, se han empleado cantidades distintas de virus control. En el primer capítulo, únicamente se han añadido 5×10^2 PFU/mL de Φ X174, 5×10^3 PFU/mL de MS2 y 5×10^4 PFU/mL de VSV como control a la muestra **P2**. En el segundo capítulo, se ha añadido 10^3 PFU/mL de Φ X174 y 10^4 PFU/mL de VSV a todas las mezclas. En el caso del tercer capítulo, se han añadido 50 PFU/mL de Φ X174 a todas las muestras.

Una vez añadidos los virus control, estos 10 mL se han sometido a una primera centrifugación ($87.000 \times g$, 4 °C, 2 h). Tras esto, el sobrenadante se ha eliminado, se han añadido 10 mL de PBS 1X y se ha sometido a una hora adicional de centrifugación bajo las mismas condiciones de velocidad y temperatura. Después, se ha eliminado el sobrenadante y el sedimento se ha resuspendido en 245 μ L del tampón de digestión del kit Turbo DNA free (Ambion). Cuando el sedimento se ha disgregado completamente, se ha añadido un cóctel de enzimas de digestión: 5 μ L de TURBO DNasa (Turbo DNA free kit (Ambion)), 2 μ L de benzonasa (Sigma) y 2 μ L de nucleasa micrococcal (NEB), que se ha incubado durante 1 h a 37 °C. Esta reacción de digestión se

ha detenido al añadir el reactivo de inactivación de la nucleasa, (*stop reagent*, Turbo DNA free kit (Ambion), se ha incubado la muestra a temperatura ambiente durante 3 min y, transcurrido este tiempo, se ha sometido a una centrifugación a $10.000 \times g$, 3 min. El sobrenadante se ha transferido a dos tubos nuevos: una fracción (40 μL) se utilizará para hacer la extracción de ADN y la otra (200 μL) se empleará para la extracción de ARN. En el caso de las muestras pertenecientes al capítulo 3 de la Tesis, donde el objetivo es detectar los virus de ADN, todo el volumen obtenido tras la digestión se ha empleado para realizar la extracción y amplificación del ADN.

Tras la digestión, la fracción de 40 μL se utilizará para la extracción y amplificación del ADN mediante el kit QIAamp[®] Viral RNA Mini (Qiagen) siguiendo las instrucciones del fabricante. Para la amplificación del ADN obtenido, se empleará el kit TruePrime WGA (Qiagen) siguiendo las instrucciones del fabricante.

Por otro lado, la fracción de 200 μL se utilizará para la extracción de ARN. Para ello, en primer lugar, se realiza una limpieza de la muestra con TRIzol, tras lo cual se lleva a cabo la extracción del ARN viral con el kit QIAamp[®] Viral RNA Mini (Qiagen) según las instrucciones del fabricante. El ARN así obtenido se somete a una reacción de retrotranscripción y el ADNc resultante se utilizará para las etapas siguientes de ligación y amplificación mediante el kit QuantiTect Whole Transcriptome (Qiagen), como se ha explicado en el apartado 2.4 de Material y Métodos.

A destacar que en cada capítulo se han añadido varios controles negativos, que han consistido en 10 mL de PBS 1X estéril y que se ha sometido a todo el proceso de concentración y extracción: en el caso del capítulo 1 hay tres controles, para el segundo hay ocho controles y para el último hay dos.

3.3. Secuenciación masiva

Con el fin de identificar qué especies virales están presentes en las muestras, se ha llevado a cabo la secuenciación masiva de los productos de amplificación obtenidos. Como se ha explicado en el apartado anterior, cada una de las muestras que pertenecen a los capítulos 1 y 2 están formadas por una mezcla de los productos de amplificación obtenidos a partir de las fracciones de ADN y ARN. Para cada muestra, el resultado de la amplificación obtenida en cada fracción se ha mezclado en concentraciones equimolares previamente a la construcción de las librerías para la secuenciación. La construcción de librerías se ha realizado empleando el kit Nextera XT DNA (Illumina) con 15 ciclos de amplificación. Estas librerías se han secuenciado en un equipo NextSeq, obteniéndose lecturas *single-end* (capítulo 1) y *paired-end* (capítulos 2 y 3). Los datos en crudo obtenidos tras la secuenciación se han depositado en la base de datos Sequence Read Archive de GenBank bajo los números de acceso de BioProject PRJNA691135 (capítulo 1), PRJNA731624 (capítulo 2) y PRJNA763062 (capítulo 3).

3.4. Análisis de las secuencias

El control de calidad de las lecturas obtenidas en la secuenciación es una etapa previa imprescindible en cualquier procedimiento bioinformático, ya que las lecturas constituyen la base sobre la que se va a apoyar cualquier resultado posterior. Los problemas detectados a partir de este análisis deben ser corregidos adecuadamente para tener bajo control la fiabilidad de los resultados. En general, podemos decir que son dos los parámetros que conviene tener bien controlados:

- En una librería diversa, la mayoría de las secuencias aparecen sólo una vez en el conjunto de lecturas final. Sin embargo, en ocasiones los

resultados de la secuenciación suelen mostrar secuencias que se repiten más de lo esperado. Esto se conoce como duplicación y puede ser indicativo de que ciertas secuencias se repiten de forma anómala. Un nivel muy bajo de duplicación suele indicar un nivel muy alto de cobertura. Sin embargo, un nivel alto de duplicación tiene su origen muy probablemente en algún tipo de sesgo de enriquecimiento, por ejemplo, por sobreamplificación debido a la PCR. Aunque existe un desacuerdo en el campo de la secuenciación sobre la importancia de eliminar o no las secuencias duplicadas, en este caso se ha optado por reducir el posible sesgo reduciendo el nivel de duplicación.

- En el proceso de secuenciación automática de ADN, la fiabilidad en la identificación de cada nucleótido en la cadena secuenciada varía con la longitud de la misma. La posibilidad de error, por tanto, aumenta hacia el final de la lectura. El nivel de calidad Phred (Q) es una medida ampliamente aceptada para caracterizar la calidad de la secuencia y se relaciona con la probabilidad (P) de error en la identificación de la base mediante la expresión

$$P = 10^{\frac{-Q}{10}}.$$

Por ejemplo, un valor de calidad $Q = 20$ para una base en concreto se traduce en una probabilidad de que esa base esté mal clasificada de $P = 10^{\frac{-20}{10}} = 0,01$, es decir, el error es de 1 en cada 100, o expresado de forma equivalente, la precisión en la clasificación es del 99%. Como se ha indicado, el valor de Q asignado a cada base generalmente disminuye, es decir, aumenta la tasa de error, a medida que se va leyendo la cadena de ADN. Es frecuente truncar las lecturas

cuando la tasa de error alcanza determinados niveles para evitar errores en la clasificación.

El control de calidad de las lecturas se realizó con las herramientas FastQC v0.11.9 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) y MultiQC v1.8 (Ewels, Magnusson, Lundin, & Källner, 2016). La deduplicación de las secuencias sobrerrepresentadas se realizó mediante la herramienta clumpify.sh. Para el filtrado por calidad de las lecturas, se ha empleado la herramienta bbdduk.sh, de modo que las secuencias fueron truncadas a partir de un nivel de calidad medio $\bar{Q} < 20$, calculado a partir de una ventana deslizante de tres nucleótidos. Las secuencias resultantes cuya longitud ha resultado menor de 70 nucleótidos fueron posteriormente descartadas. Ambas herramientas, clumpify.sh y bbdduk.sh, pertenecen al paquete BBTools suite v38.68 (Bushnell, Rood, & Singer, 2017).

Una vez que se ha mejorado la calidad de las lecturas, se ha procedido a su clasificación taxonómica. Es posible encontrar en la literatura una gran variedad de clasificadores taxonómicos, pero posiblemente la característica común a todos ellos es que parten de una base de datos preconstruida a partir de los genomas de referencia y que se utiliza para asignar un taxón y una puntuación (asociada a la calidad de la clasificación) a cada una de las lecturas. Las bases de datos de referencia más populares en *Shotgun Metagenomics* (SM) son la *nt* NCBI BLAST y *RefSeq* (NCBI *Reference Sequence*) para secuencias de nucleótidos de gran calidad. Estas bases de datos, de propósito general, no siempre son suficientes para proporcionar una calidad de clasificación óptima. Una estrategia habitual para mejorar el rendimiento de las bases de datos convencionales consiste en añadir genomas adicionales dirigidos al problema de interés (Martí & Garay, 2019). Estos genomas adicionales se pueden obtener a partir de MAG (*Metagenome-Assembled Genome*) caracterizados en

estudios metagenómicos relacionados. Sin embargo, hay autores que señalan que un aumento en el tamaño de la base de datos no siempre resulta beneficioso (Seppey, Manni, & Zdobnov, 2020), ya que, aunque aumenta la capacidad de clasificación de la base de datos, la adición de especies parcialmente representadas y con rangos taxonómicos incompletos (MAG) puede empeorar la representación de la muestra metagenómica en niveles taxonómicos más altos.

El análisis metagenómico de los datos obtenidos en la secuenciación se ha llevado a cabo con el programa Centrifuge (Kim et al., 2016) versión 1.0.4, empleando una puntuación mínima de 18. Centrifuge es un clasificador muy rápido que usa un esquema de indexación basado en una combinación de la transformada de Borrows-Wheeler (BW) con un índice de Ferragina-Manzini (FM) optimizados específicamente para la clasificación metagenómica. Para clasificar una secuencia, Centrifuge comienza buscando en su base de datos una pequeña concordancia exacta (mínimo 16 bp) que después extiende todo lo posible hasta encontrar un error. Este proceso se repite en el resto de la secuencia hasta que se lee en su totalidad. El proceso se realiza tanto en la dirección 5' a 3' como en la reversa complementaria, es decir, en la dirección 3' a 5'. Después y, en función del número de concordancias exactas que ha encontrado tanto en la cadena como en su reversa complementaria, Centrifuge la clasifica usando sólo las concordancias de longitud mayor de 18 pares de bases. Puesto que cada una de las concordancias puede pertenecer a una o más especies, Centrifuge utiliza la siguiente expresión para asignar una puntuación a cada especie.

$$Puntuación(Especie) = \sum_{hit \in Especie} (Longitud(hit) - 15)^2$$

Como se puede deducir de la expresión, Centrifuge da mayor peso a los segmentos más largos.

En el caso del primer capítulo, la clasificación taxonómica de las lecturas se realizó a partir de una base de datos personalizada (junio de 2019). Ésta se generó a partir de la base de datos nt de NCBI, junto con los genomas completos de arqueas, virus, bacterias y hongos presentes en la base de datos RefSeq, obtenidos mediante la herramienta de descarga que presenta Centrifuge. También se empleó draftGenomes (Martí & Garay, 2019) para añadir las secuencias SMS de la base de datos WGS del NCBI que pertenecen a taxones virales. Por otra parte, para abordar la clasificación taxonómica de los capítulos 2 y 3, se creó una nueva base de datos actualizada mediante los procedimientos descritos anteriormente (septiembre de 2020). Los análisis obtenidos con Centrifuge han sido procesados posteriormente para tratar de eliminar las lecturas pertenecientes a los grupos taxonómicos detectados en los controles y, por tanto, consideradas como contaminantes. Este análisis se ha llevado a cabo con Recentrifuge (Martí, 2019) versión 1.3.2 empleando una puntuación mínima de 22. Además de la clasificación taxonómica de las lecturas, Centrifuge también se ha utilizado para realizar el análisis de la abundancia de las distintas especies virales de interés.

El ensamblaje de las lecturas obtenidas en cada capítulo se ha realizado abordando distintas estrategias. En el caso del capítulo 1, se han generado *contigs* por ensamblaje de las lecturas de cada muestra. Este ensamblaje se ha llevado a cabo con el programa SPAdes versión 3.14.0 (Nurk et al., 2013) empleando los parámetros por defecto. En el caso de los capítulos 2 y 3, se ha empleado el programa metaSPAdes versión 3.15.0 (Nurk, Meleshko, Korobeynikov, & Pevzner, 2017) para generar los *contigs* empleando los parámetros por defecto.

En ambos casos, los *contigs* obtenidos tras el ensamblaje se han sometido a un análisis de homología empleando BLASTn v2.10.0 con un valor $E < 10^{-5}$ contra una base de datos local, que es una copia de la base de datos de nucleótidos (nt) del NCBI. Para cada aislado identificado como miembro de las familias *Anelloviridae*, *Flaviviridae* o *Microviridae*, se ha estimado su profundidad de cobertura media empleando la herramienta pileup.sh del paquete BBTools suite v38.68 (Bushnell et al., 2017).

Las secuencias descritas en cada capítulo se han depositado en GenBank bajo los números de acceso: capítulo 1, secuencias pertenecientes a la familia *Anelloviridae* MW455345-MW455458 (**Tabla Suplementaria S1**) y la secuencia descrita perteneciente a la familia *Flaviviridae* MW467971; capítulo 2, secuencias de anellovirus MZ285962-MZ286225 (**Tabla Suplementaria S2**), aislados de *Flaviviridae* MZ420565-MZ420581 (**Tabla Suplementaria S3**) y el aislado de la familia *Microviridae* MZ286294; capítulo 3, aislados de la familia *Anelloviridae* MZ824758-MZ825040 y MZ825088-MZ825153 (**Tabla Suplementaria S4**) y el *contig* de la familia *Microviridae* MZ821031.

Las ORFs putativas de las secuencias analizadas de los tres capítulos se han identificado mediante ORF Finder (<https://www.ncbi.nlm.nih.gov/orffinder/>).

3.5. Análisis filogenético

Para estudiar las relaciones filogenéticas en los miembros de la familia *Anelloviridae*, se han descargado las secuencias nucleotídicas correspondientes a la ORF1 de las especies de referencia aceptadas por ICTV de los géneros TTV, TTMV y TTMDV (capítulo 1, **Tabla Suplementaria S5**; capítulo 2, **Tabla Suplementaria S6**; capítulo 3, **Tabla Suplementaria S7**). Respecto a las secuencias obtenidas experimentalmente en los distintos

capítulos, se han seleccionado aquellos aislados en los que la longitud de la ORF ha sido mayor a 500 aminoácidos. Por otra parte, en los capítulos 1 y 2 también se ha realizado un análisis filogenético para las secuencias de HPgV obtenidas. Para ello, se descargaron las secuencias nucleotídicas para la totalidad de la longitud de la poliproteína de los aislados disponibles en GenBank hasta marzo de 2021 (**Tabla Suplementaria S3**).

Tanto para el análisis filogenético de los anelovirus, como para HPgV, se realizó un alineamiento de las secuencias (basada en las secuencias aminoacídicas) con MUSCLE (Edgar, 2004) implementado en MEGA versión X (Kumar, Stecher, Li, Knyaz, & Tamura, 2018). La posterior inferencia filogenética se ha realizado empleando las secuencias nucleotídicas y se ha llevado a cabo mediante el método de máxima verosimilitud (*maximum likelihood*, ML, por sus siglas en inglés) implementado en MEGA versión X. Los análisis se han realizado bajo el modelo de sustitución de nucleótidos de mejor ajuste, identificado como GTR + Γ + I, utilizando el criterio de información de Akaike como marco de selección del modelo en MEGA versión X. La fiabilidad de los resultados filogenéticos se evaluó utilizando un análisis de *bootstrap* con 1000 réplicas. Los árboles finales se anotaron con EvolView (Subramanian, Gao, Lercher, Hu, & Chen, 2019). La demarcación de las especies de anelovirus se realizó comprobando las matrices de identidad de nucleótidos por pares obtenidas independientemente para cada género mediante el programa STD v1.2 (Muhire et al., 2014).

3.6. Secuenciación por el método Sanger

En el capítulo 1, los *contigs* obtenidos tras el ensamblaje que se han identificado como anelovirus presentan una gran variación en el valor de la cobertura del genoma. Para comprobar que no se trata de artefactos, las

secuencias obtenidas del ensamblaje para varios genomas de anelovirus se han empleado para diseñar cebadores que amplifican la ORF1 completa. Para la amplificación, se ha llevado a cabo reacciones en un volumen final de 40 μL , que contenían 1 μL de ADN, 8 μL del tampón GC 5X, 0,4 μL de la polimerasa Phusion High-Fidelity DNA Polymerase (Thermo Scientific™), 0,8 μL de dNTPs (10 mM), 1,6 μL de los cebadores directo e inverso (10 μM) y el volumen restante de H_2O . Las condiciones de amplificación dependen de la temperatura de anillamiento de la pareja de cebadores y del tamaño del amplicón. Los cebadores de PCR, junto con algunos cebadores internos adicionales, se han utilizado para la secuenciación por Sanger (**Tabla Suplementaria S8**).

En el capítulo 2, el objetivo de la secuenciación Sanger es distinto. En este caso, se ha empleado para tratar de dilucidar aquellas regiones de la poliproteína del genoma de HPgV que no se habían obtenido tras el ensamblaje de los datos de secuenciación masiva. Para tal fin, se han diseñado cebadores específicos a partir de las regiones conocidas para posteriormente llevar a cabo las reacciones de amplificación. Para ello, a partir de la fracción de ARN almacenada de las muestras positivas para este virus, se ha realizado una reacción de retrotranscripción empleando la enzima SuperScript IV (Invitrogen) y hexámeros aleatorios, siguiendo las instrucciones del fabricante. Tras esto, para cada región no cubierta por los datos de secuenciación masiva, se han llevado a cabo reacciones de amplificación, para las cuales se han empleado 2 μL de la reacción de retrotranscripción, 8 μL del tampón GC 5X, 0,4 μL de la polimerasa Phusion High-Fidelity DNA Polymerase (Thermo Scientific™), 0,8 μL de dNTPs (10 mM), 1,6 μL de los cebadores directo e inverso (10 μM) y el volumen restante es H_2O . De nuevo, las condiciones de amplificación dependían de la temperatura de anillamiento de la pareja de cebadores y del tamaño que presenta el amplicón (**Tabla Suplementaria S9**).

En el caso de las mezclas **SP30**, **SP49** y **SP53**, los datos de secuenciación masiva hacían sospechar la presencia de dos aislados distintos de HPgV en cada una de ellas. Para tratar de confirmar o descartar esta sospecha, se hizo una extracción de ARN de cada una de las muestras individuales incluidas en las tres mezclas. Para cada muestra, se realizó la transcripción inversa y posterior amplificación, mediante el uso de cebadores específicos (Directo 5'-CAGAACCATAGCCTATTGTGA-3' e Inverso 5'-CACCTTAGATCCCCCA-3') diseñados a partir de regiones conservadas en el alineamiento global.

4. Análisis de recombinación

4.1. Análisis de recombinación de *Pegivirus*

En el capítulo 2, algunos de los genomas recuperados de HPgV presentaban evidencias de recombinación debido a que se localizaban en una posición basal en el árbol generado a partir del análisis filogenético. Para explorar esta posibilidad, se han llevado a cabo análisis de recombinación. En primer lugar, se ha descargado la secuencia nucleotídica completa de la poliproteína de los aislados de HPgV presentes en las bases de datos (abril de 2021). Sobre este conjunto de datos, se ha añadido la secuencia nucleotídica de la poliproteína de los aislados obtenidos en este trabajo. Tras esto, se ha llevado a cabo un alineamiento con MUSCLE (Edgar, 2004) implementado en MEGA versión X (Kumar et al., 2018). A partir de aquí, se ha estimado una red filogenética utilizando el programa SplitsTree4 (Huson & Bryant, 2006), basada en el alineamiento de secuencias utilizado para construir un árbol ML. La filogenia reticulada se calculó mediante el método NeighborNet.

4.2. Análisis de recombinación de anelovirus

En el capítulo 3 se ha analizado el viroma individual de distintos donantes, específicamente, la fracción de virus de ADN presentes. En aquellos

donantes en los que se ha detectado un número mayor de aislados de anelovirus se ha realizado un análisis de recombinación, ya que existe la posibilidad de que se pueda detectar algún suceso de recombinación en los diferentes genomas virales presentes en el donante. Debido a la enorme heterogeneidad genética presente en los tres géneros de anelovirus humanos, ha sido aconsejable realizar los análisis de recombinación de manera separada para cada género. Por ello, se han identificado aquellos donantes que presentaban al menos 10 aislados del mismo género, en concreto, los donantes **SPI1**, **SPI9**, **SPI17**, **SPI19**, **SPI23**, **SPI32** y **SPI35**. Para llevar a cabo el análisis de recombinación, se han seleccionado las secuencias correspondientes a la ORF1 de los aislados pertenecientes al mismo género y se han alineado con MAFFT (Katoh & Standley, 2013) con la configuración E-INS-i. Estos alineamientos se han analizado con el programa RDP4 (Martin, Murrell, Golden, Khoosal, & Muhire, 2015) con el fin de detectar los sucesos de recombinación putativos. Se han analizado los alineamientos empleando los métodos de detección RDP (Martin & Rybicki, 2000), GENECONV (Padidam, Sawyer, & Fauquet, 1999), BootScan (Martin, Posada, Crandall, & Williamson, 2005), MaxChi (Smith, 1992), Chimaera (Posada & Crandall, 2001), SiScan (Gibbs, Armstrong, & Gibbs, 2000), y 3Seq (Lam, Ratmann, & Boni, 2018) empleando los parámetros por defecto. Se han determinado como potenciales secuencias recombinantes aquellas secuencias en las que en más de tres métodos de detección se ha obtenido un P valor $<0,05$. El resultado obtenido en RDP4 ha servido para determinar los puntos de ruptura de los fragmentos recombinantes. Los fragmentos detectados como sucesos de recombinación putativos se han extraído del alineamiento con la herramienta *extract align* del programa EMBOSS (European Molecular Biology Open Software Suite).

Tras esto, se ha evaluado la señal filogenética de los fragmentos recombinantes mediante el análisis de mapeo de probabilidad. Este método se basa en el análisis de un cuarteto de secuencias, para las que se calcula la máxima verosimilitud de cada una de las tres topologías posibles de árbol filogenético sin raíz, por lo que pueden estar completamente resueltas o sin resolver (Strimmer & von Haeseler, 1997). La probabilidad de cada una de las topologías alternativas se estima con el modelo de evolución apropiado. Después, las probabilidades obtenidas, que suman un total de 1, se representan dentro de un triángulo equilátero (**Figura 6**), que está dividido en 7 regiones distintas. Cada una de las esquinas del triángulo representa una de las tres posibles topologías que pueden tener los árboles generados. Si la probabilidad está cerca de una esquina, representa que esa topología es la que presenta un apoyo mayor. Por el contrario, si la probabilidad está en la región central, la región 7 en la **Figura 6**, las probabilidades de las tres topologías presentan una magnitud similar. En esta situación, la representación más realista de los datos sería en un árbol con forma de estrella, en el que las relaciones entre los taxones no están completamente resueltas (Strimmer & von Haeseler, 1997).

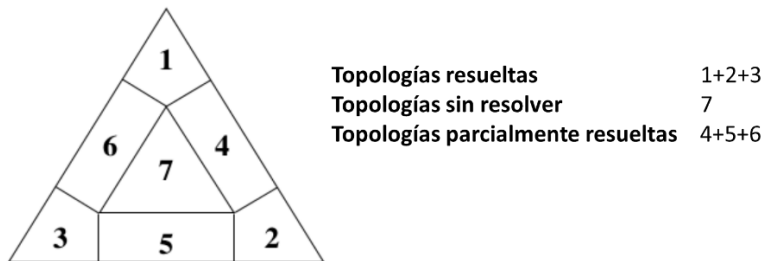


Figura 6: Esquema del triángulo equilátero en el que se representan las probabilidades obtenidas en el análisis de mapeo de probabilidad. Figura adaptada de Morrison, 2005.

Para llevar a cabo la evaluación de la señal filogenética, los fragmentos obtenidos anteriormente se han analizado con la aproximación *-lmap* implementada en el programa IQ-TREE 1.6.12 (Nguyen, Schmidt, von Haeseler, & Minh, 2015) y se han evaluado 1.000 cuartetos aleatorios. Este número es considerado suficiente para obtener una imagen comprensiva de la calidad filogenética del conjunto de datos (Strimmer & von Haeseler, 1997). En este caso, se han seleccionado aquellos sucesos que presentan menos del 50% de cuartetos que permanecen sin resolver. Esto es debido a que cuanto menor sea este valor, la señal filogenética es mejor porque la reconstrucción de los árboles es más fiable, ya que un porcentaje mayor presentará una topología resuelta. Tras esto, se ha evaluado la topología de los árboles filogenéticos de los fragmentos seleccionados en el paso anterior para determinar si son significativamente diferentes al árbol que contiene toda la ORF1. Para ello, el programa IQ-TREE 1.6.12 (Nguyen et al., 2015) presenta implementadas con los comandos *-zw* y *-au* diferentes pruebas para evaluar la topología de los árboles. En este análisis se han elegido la prueba de los pesos de probabilidad esperados (*expected likelihood weights test*, ELW) (Korbinian Strimmer & Rambaut, 2002) y la prueba aproximadamente imparcial (*approximately unbiased test*, AU) (Shimodaira, 2002). La elección de estas pruebas se ha hecho debido a la metodología, ya que es posible comparar más de dos árboles, la fiabilidad (debido a que estas pruebas son menos conservativas y sesgadas) y la implementación (más rápido y mejor implementadas que otras pruebas) (Morrison, 2005). En el caso de la prueba AU, el resultado devuelve el *P*-valor de cada topología, por lo que un árbol se rechaza para un *P*-valor < 0,05. Por otro lado, aquellos árboles donde el valor proporcionado por la prueba AU presente el símbolo +, implica que esos árboles son significativamente más verosímiles para ese alineamiento que el resto

Material y métodos

(indicados con - (P -valor $< 0,05$)). Para confirmar el suceso de recombinación, el árbol del fragmento recombinante debe presentar un valor de AU $> 0,05$ mientras que el árbol de la ORF1 completa tendría que tener un valor de AU $< 0,05$ es decir, estar rechazado significativamente. Por otro lado, el valor de la prueba ELW representa el peso de la probabilidad de cada árbol, es decir, el porcentaje de probabilidad sumada de cada uno de los árboles.

Capítulo 1: Implementación del protocolo de enriquecimiento viral y resultados preliminares.

Nota: Los resultados de este capítulo se presentan en el trabajo:

Cebriá-Mendoza, M., Arbona, C., Larrea, L. Díaz, W., Arnau, V., Peña, C., Bou, JV., Sanjuán, R., Cuevas, JM. (2021). Deep viral blood metagenomics reveals extensive anellovirus diversity in healthy humans. *Scientific Reports* **11**, 6921. <https://doi.org/10.1038/s41598-021-86427-4>

Los resultados obtenidos en este capítulo están divididos en dos partes: en la primera de ellas, se muestran los resultados obtenidos en la prueba piloto, que ha consistido en la comparación de diferentes técnicas de concentración y extracción de ácidos nucleicos virales para determinar qué procedimiento es más eficiente en cuanto a la recuperación de los ácidos nucleicos presentes en las muestras. Los resultados obtenidos en esta prueba piloto han servido para determinar el flujo de trabajo que se ha seguido en el procesamiento de las muestras empleadas en la Tesis. Por otro lado, en la segunda parte de resultados de este capítulo, a modo de se han analizado muestras de plasma procedentes de 120 donantes, las cuales se han agrupado en 12 mezclas compuestas aproximadamente por 10 muestras individuales.

1. Resultados

1.1. Estrategia y resultado de la prueba piloto

El objetivo del protocolo analizado durante la primera parte de este capítulo (ver sección 2 de Material y Métodos) es enriquecer la fracción viral de las muestras de plasma y, para ello, se han probado diferentes combinaciones experimentales que han incluido pasos de filtración y centrifugación. Posteriormente, se ha realizado una digestión con nucleasas para tratar de eliminar los ácidos nucleicos libres, mientras los genomas virales se encuentran protegidos por las cubiertas proteicas. Finalmente, se ha llevado a cabo la extracción de los ácidos nucleicos, que se ha realizado independientemente para la fracción de ADN y la de ARN, seguida de una amplificación aleatoria y la preparación de las librerías para la secuenciación. Los resultados de la secuenciación se han clasificado taxonómicamente y se han sustraído aquellos taxones que se encontraban presentes en los controles. Dado que el paso de filtración debería seleccionar las partículas virales, se

espera que una fracción significativa de los ácidos nucleicos humanos y bacterianos sea eliminada o finalmente sustraída por comparación con las muestras control durante el análisis bioinformático.

Inicialmente, el objetivo era comparar la eficacia de la recuperación de la fracción viral obtenida mediante la extracción directa de ácidos nucleicos del plasma frente a la obtenida realizando un paso previo de centrifugación a alta velocidad (ver secciones 2.1 y 2.2 de Material y Métodos). Para ello, a una mezcla de plasma, que incluía muestras de diez donantes diferentes, se le han añadido distintos virus control con títulos con significado biológico. En concreto, se han introducido 10^3 PFU/mL de los bacteriófagos ϕ X174 (virus de ADN circular monocatenario sin envuelta) y MS2 (virus de ARN lineal monocatenario sin envuelta), y del virus Vaccinia (virus de ADN lineal bicatenario con envuelta), y 10^4 PFU/mL de VSV (virus de ARN lineal monocatenario con envuelta). En este estudio piloto, también se han analizado los anelovirus (virus de ADN circular monocatenario sin envuelta), ya que se encuentran con frecuencia en el plasma sanguíneo. Para la comparación entre la extracción directa de plasma y el protocolo que implicaba alta centrifugación, se han realizado dos réplicas técnicas, que difieren en el tamaño de poro utilizado en el paso de filtración inicial (0,45 μ m frente a 1 μ m), aunque se espera que esta diferencia afecte únicamente a virus de gran tamaño, como Vaccinia (Colson et al., 2013). En efecto, las muestras filtradas inicialmente con el tamaño de poro mayor mostraban la detección de miles de lecturas de Vaccinia en los dos protocolos comparados, mientras que no se detectó ninguna lectura en una de las muestras filtradas con el tamaño de poro menor (**Figura 7 y Tabla Suplementaria S10**).

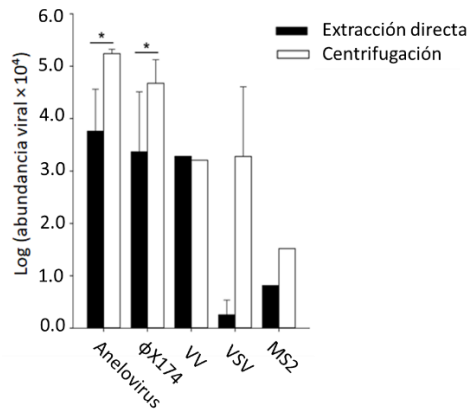


Figura 7: Comparación de la abundancia viral estimada entre la extracción directa del plasma y el protocolo que implica una centrifugación a alta velocidad. Estos datos se han obtenido transformando las lecturas totales para cada grupo taxonómico en abundancia viral, usando para ello un algoritmo de estimación-maximización implementado en Centrifuge (Kim et al., 2016). Para mayor claridad, la abundancia viral $\times 10^4$ se ha representado en escala logarítmica. Las barras de error indican el error estándar de la media (SEM, $n=2$ réplicas). El asterisco indica la significación estadística de la prueba t -test, que analiza la eficiencia de los protocolos de purificación (* $P < 0.01$). Para el virus Vaccinia, el único valor indicado para cada tratamiento fue obtenido en las muestras filtradas con tamaño de poro de $1 \mu\text{m}$.

Al comprobar la presencia de virus de ADN circular (anelovirus y ϕX174), se ha observado un claro aumento de la eficiencia de la recuperación viral en cuanto al número de lecturas y a la abundancia, que oscilaba entre uno y dos órdenes de magnitud, en el caso del protocolo que incluye un paso de centrifugación a alta velocidad (**Figura 7**; prueba t -test: $P < 0,001$ para ambos virus). En el caso de VSV, sólo se han detectado once lecturas en una réplica del tratamiento de extracción directa, mientras que se han recuperado miles de lecturas cuando las muestras se han centrifugado a alta velocidad (**Tabla Suplementaria S10**) (prueba t -test: $P = 0,066$). En el caso de MS2, se han detectado lecturas virales en una única réplica de cada tratamiento, por lo que

no se han podido extraer conclusiones claras, aunque esto podría explicarse por la baja cantidad añadida inicialmente para este virus. A pesar de que la cantidad total añadida de MS2 y otros virus era la misma, su baja detección puede explicarse por el hecho de que se trata de un virus de ARN con un genoma pequeño, y por ello, es esperable que se detecte con menor eficiencia que los grandes virus de ADN (como el virus Vaccinia) o los virus de ADN circular, como ϕ X174 y los anelovirus, que se amplifican preferentemente por el método MDA. Consecuentemente, dado que los resultados obtenidos en esta prueba piloto indican que un paso de centrifugación de alta velocidad aumenta sustancialmente la recuperación de los virus de ADN circular y VSV, se ha utilizado este enfoque en combinación con un paso de filtración inicial (tamaño de poro de 1 μ m) para evitar la posible pérdida de virus de gran tamaño.

1.2. Análisis de 120 muestras de plasma

Debido a los resultados obtenidos en la prueba piloto, se ha establecido el protocolo a seguir para concentrar y extraer los ácidos nucleicos virales presentes en las muestras de plasma. La mezcla obtenida a partir de diez plasmas agrupados en el experimento piloto, procesada con el filtro de 1 μ m y centrifugada a alta velocidad, se denominó mezcla 1 (**P1**) y se ha empleado para ser analizada en mayor detalle. Por otra parte, se han procesado bajo las condiciones determinadas en el experimento piloto otras once mezclas de plasma (**P2-P12**), cada una de las cuales estaba compuesta de alrededor de 10 muestras heterogéneas en cuanto a edad y sexo de los donantes (**Tabla Suplementaria S11**).

Una vez procesadas las muestras y recibidos los datos de secuenciación, se ha llevado a cabo una limpieza de los datos de secuenciación, como se ha

descrito en el apartado 3.4 de Material y Métodos, y se ha llevado a cabo una clasificación taxonómica mediante el software Centrifuge (Kim et al., 2016). La frecuencia de las lecturas no asignadas a ningún taxón para cada muestra osciló entre el 0,9 y el 6,8% (**Tabla Suplementaria S12**). En los estudios de metagenómica que analizan la composición del viroma suele estar presente una fracción variable de lecturas no asignadas (Fanello et al., 2014; Greninger et al., 2015; Law et al., 2013) y esto se puede explicar parcialmente por los artefactos debidos a la transcripción inversa y a la amplificación aleatoria de los ácidos nucleicos presentes (Nelson, 2014; Rosseel, Ozhelvaci, Freimanis, & Van Borm, 2015). Dado que el protocolo de purificación puede arrastrar cantidades residuales de ácidos nucleicos, ha sido esencial la introducción de controles negativos para evaluar el riesgo de contaminación. Las lecturas obtenidas en estos controles se han utilizado también para la clasificación taxonómica y para la posterior sustracción de estos potenciales taxones contaminantes en las muestras reales empleando los programas informáticos Centrifuge y Recentrifuge (Martí, 2019), respectivamente. Tras esto, el análisis se ha centrado en las lecturas asignadas a taxones exclusivamente presentes en las muestras (es decir, lecturas humanas, bacterianas y virales), aunque también se han identificado otros grupos taxonómicos (**Tabla 2** y **Tabla Suplementaria S12**). Se observa que las ambigüedades en la clasificación taxonómica de las lecturas no han sido tratadas adecuadamente por Recentrifuge, lo que ha limitado la capacidad para eliminar posibles contaminaciones correspondientes a lecturas no claramente clasificadas. Aun así, la fracción total de lecturas virales ha aumentado del 40,5 al 93,9% tras el paso de sustracción (**Figura 8A**). En cambio, la fracción total de lecturas bacterianas después de la sustracción se ha reducido del 50,4% al 6,1%. Por otra parte, las lecturas humanas se han eliminado por Recentrifuge. Las

lecturas bacterianas que no han sido eliminadas abarcan 24 filos (**Figura 8B y Tabla Suplementaria S13**), que incluyen los filos Firmicutes (55,2%), Proteobacterias (22,6%), Actinobacterias (5,0%), Cianobacterias (2,5%), Tenericutes (2,3%) y Bacteroidetes (2,0%). Las proporciones relativas de estos filos coinciden con estudios anteriores sobre el microbioma sanguíneo (Castillo, Rifkin, Cowan, & Potgieter, 2019), lo que sugiere que estas secuencias pueden corresponder a cantidades residuales de ADN que han podido “sobrevivir” al protocolo de enriquecimiento de ácidos nucleicos virales. De manera alternativa, podrían ser considerados como contaminantes que no fueron eliminados computacionalmente.

Tabla 2: Resumen de los resultados de Recentrifuge para las 12 mezclas analizadas.

Para cada mezcla, se indica el número total de lecturas clasificadas tras el análisis de Recentrifuge y el número de lecturas clasificadas como bacterias, anelovirus y otros virus. La última columna muestra el número de *contigs* obtenidos tras el paso de ensamblaje, pertenecientes a la familia *Anelloviridae*, que presentan un tamaño mayor de 1,5 kb. Para mayor claridad, se excluyen de los recuentos las lecturas vírales procedentes de los virus control.

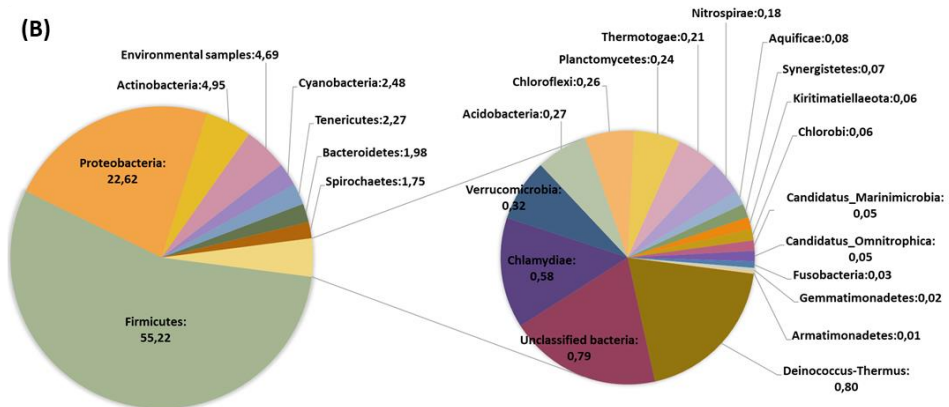
Mezcla	Nº lecturas totales	Nº lecturas bacteria	Nº lecturas anelovirus	Otros virus	Contigs de anellovirus
P1	192.322	1.704	147.618	1.943	25
P2	125.327	4.924	93.930	1.106	9
P3	127.638	3.538	111.188	1.128	9
P4	150.324	8.754	122.128	1.859	4
P5	140.469	55.550	11.187	17.302	2
P6	35.120	4.996	6.132	2.139	3
P7	368.499	6.240	339.778	1.057	22
P8	199.882	5.303	171.517	4.501	20
P9	536.796	5.096	493.302	18.383	6
P10	167.112	5.251	140.398	3.851	8
P11	74.803	8.505	25.269	943	2
P12	74.729	4.514	44.321	3.078	4

Las muestras analizadas contienen secuencias de 56 familias virales diferentes (**Figura 8C** y **Tabla Suplementaria S14**), pero existe un claro dominio de la familia *Anelloviridae*, que representa el 96,8% de la fracción total. La segunda familia más abundante es *Flaviviridae* (0,92%), aunque la mayoría de las lecturas corresponden a un pegivirus humano (HPgV) detectado en la mezcla **P9** (16.313 lecturas; cobertura del genoma mayor del 95% y profundidad de cobertura media de 191x). Este hallazgo confirma que el protocolo que se ha seguido también es eficiente para la recuperación de virus de ARN. El resto de familias virales se han detectado en menor frecuencia, con un número de lecturas que oscila entre 6 y 10.743, y no se han encontrado potenciales patógenos humanos dentro de estas familias. Esto sugiere que estas lecturas pueden ser contaminantes, algo que ocurre habitualmente en los estudios de viroma (Asplund et al., 2019). De hecho, las familias más abundantes pertenecen a virus de insectos (por ejemplo, las familias *Polydnviridae* y *Ascoviridae*, con 10.743 y 5.582 lecturas, respectivamente) y bacteriófagos (por ejemplo, los grupos taxonómicos *Caudovirales* y *Microviridae*, con 4.801 y 2.248 lecturas, respectivamente). Cabe mencionar que más del 99% de las lecturas asignadas a la familia *Circoviridae* han sido sustraídas por ReCentrifuge. La detección de miembros de esta familia se ha asociado a reactivos contaminados (Asplund et al., 2019), lo que subraya la necesidad de incluir controles adecuados. El hecho de que algunas lecturas de la familia *Circoviridae* aún permanezcan después de la sustracción bioinformática con ReCentrifuge puede ser una consecuencia de las limitaciones técnicas mencionadas anteriormente.

(A)

	% Lecturas		
	Humano	Bacteria	Virus
Centrifuge	9,1±0,8	50,4±6,0	40,5±6,5
Recentrifuge	0,0	6,1±5,7	93,9±5,7

(B)



(C)

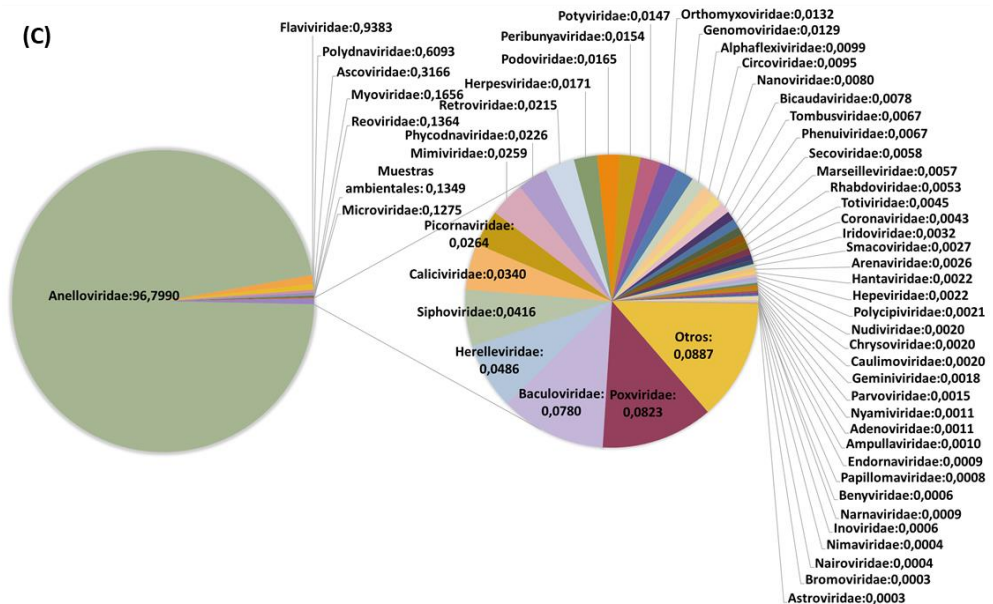


Figura 8: Resumen de la sustracción bioinformática (i.e. frecuencia media de lecturas (± SEM) antes y después de utilizar Recentrifuge para las doce mezclas analizadas) asignadas a humano, bacteria y virus (A) y descripción del microbioma (B) y del viroma (C) caracterizados en este estudio. La clasificación se muestra para las bacterias a nivel

de filo, mientras en el caso de los virus se presenta a nivel de familia. Las frecuencias se han obtenido excluyendo la contribución de los virus controles tanto en (A) como en (C).

1.2.1. Análisis de HPgV

Dado que la secuencia de HPgV detectada en nuestro estudio ha mostrado una alta cobertura del genoma, se ha realizado un análisis filogenético para asignar la secuencia a un genotipo. Para ello, se han descargado las secuencias nucleotídicas de la poliproteína completa, que abarca aproximadamente el 90% del genoma, de algunos aislados representativos de los diferentes genotipos conocidos. Mediante el análisis del árbol filogenético (**Figura 9**), se ha concluido que la secuencia descrita corresponde al subtipo a del genotipo 2. Este resultado es congruente con la distribución geográfica de los genotipos de HPgV, ya que el genotipo 2 se encuentra habitualmente en Europa y América (Reshetnyak, Karlovich, & Ilchenko, 2008). Además, la detección de un único aislado en nuestro estudio también es coherente con estudios anteriores que muestran que la prevalencia de HPgV en los países desarrollados oscila entre el 0,5 y el 5% (Mohr & Stapleton, 2009).

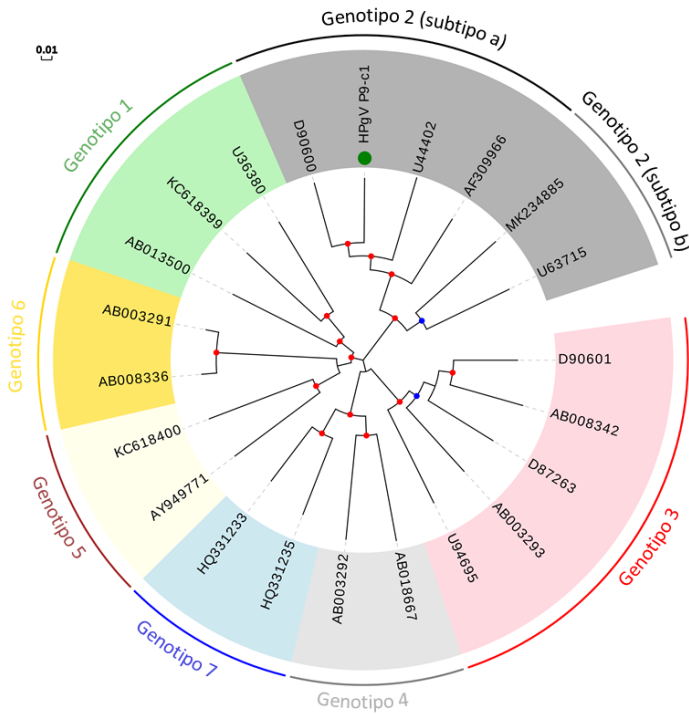


Figura 9: Árbol filogenético basado en la secuencia codificante completa (es decir, la poliproteína) que incluye aislados representativos de los siete genotipos conocidos de HPgV. La secuencia descrita en este estudio se indica con un círculo verde. Para el genotipo 2, se muestran distintos miembros de los subtipos 2a y 2b. Los valores de *bootstrap* de 0,75-0,9 y 0,9-1,0 se indican con círculos azules y rojos, respectivamente. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

1.2.2. Análisis de los anelovirus

Para cada una de las doce mezclas, se han generado *contigs* a partir de todas las lecturas, independientemente de la clasificación taxonómica preliminar, lo que ha evitado la pérdida involuntaria de lecturas virales y ha demostrado ser eficaz para la detección de nuevos anelovirus (de Souza et al., 2018). Se han seleccionado los *contigs* de un tamaño mayor de 1,5 kb y, posteriormente, se han sometido a un análisis por BLAST. Este análisis ha demostrado que sólo unos pocos *contigs* pertenecen a los virus introducidos

como control en las mezclas **P1** y **P2** o al pegivirus humano descrito anteriormente. Por otra parte, se han obtenido 114 *contigs* que corresponden a la familia *Anelloviridae*, y 23 de ellos mostraban extremos solapantes, por lo que pueden considerarse genomas completos (**Tabla 2** y **Tabla Suplementaria S1**). Además, existe una correlación positiva entre el número de *contigs* y la cantidad total de lecturas de anelovirus presentes en cada mezcla (correlación de Spearman: $\rho = 0,728$; $P = 0,004$). Para realizar el análisis filogenético, se ha empleado la secuencia nucleotídica de la ORF1. Con tal fin, se ha obtenido la ORF1 completa para todos los *contigs*, excepto en ocho (93%). Para hacer una clasificación taxonómica preliminar, se ha construido un árbol filogenético que ha incluido la ORF1 de las secuencias de anelovirus de homínidos disponibles en GenBank (**Tabla Suplementaria S5**), lo que ha permitido clasificar a los *contigs* como pertenecientes a los géneros TTV, TTMV o TTMDV (68, 29 y 17 secuencias, respectivamente; **Tabla Suplementaria S1** y **Figura Suplementaria S1**). De los 23 *contigs* considerados como genomas completos, 22 pertenecen al género TTMV y el aislado restante al género TTMDV. La eficiencia del ensamblaje se ha visto bastante afectada por las regiones ricas en GC presentes en los anelovirus, pero estas regiones son más cortas en el género TTMV (Ninomiya et al., 2007), lo que puede explicar la recuperación más eficiente de genomas completos para este género. Esto también explica por qué varios *contigs* han caído en el rango de tamaños esperado para un genoma completo, pero no presentan redundancia terminal.

Para facilitar la visualización de los resultados, los árboles filogenéticos se han construido de forma independiente para cada género, y sólo se ha utilizado un genotipo representativo de cada especie, incluyendo algunos que actualmente no son aceptados por el ICTV. Para el género TTV, que consta de siete grupos filogenéticos (Hsiao, Wang, Lin, & Liu, 2016), el árbol ha incluido

nuestras 68 nuevas secuencias, así como 36 genotipos descritos previamente, cada uno de los cuales representa una especie conocida (**Figura 10**). Este árbol, junto con los valores de divergencia, muestra que ocho de las secuencias descritas en este trabajo pueden considerarse pertenecientes a seis nuevas especies, mientras que los aislados restantes se han agrupado con 18 de las 36 especies previamente conocidas (**Tabla Suplementaria S15**). El número de secuencias asignadas a cada especie de TTV ha sido variable. Por ejemplo, cuatro de las especies de referencia se han agrupado con únicamente una de las secuencias descritas en este capítulo, mientras que las especies representadas por los genotipos TTV29-yon-KC009 y TTV3-HEL32 se han agrupado con ocho y diez de nuestras secuencias, respectivamente (**Figura Suplementaria S2**). Esto contrasta con un estudio anterior que mostraba que TTV8 era la especie cuantitativamente más prevalente en la sangre humana (De Vlaminck et al., 2013), ya que TTV8 no se ha agrupado con ninguna de las secuencias descritas en este capítulo. Tampoco se han encontrado secuencias que agrupen con especies pertenecientes a los grupos 2, 6 y 7. Aun así, existe una correlación positiva entre el número de especies incluidas en cada grupo y el número de secuencias descritas en este capítulo, incluso cuando se ignoran los datos de los grupos 6 y 7 recientemente propuestos, que constan de una sola especie (coeficiente de correlación de Spearman; $\rho = 0,821$, $P = 0,044$).

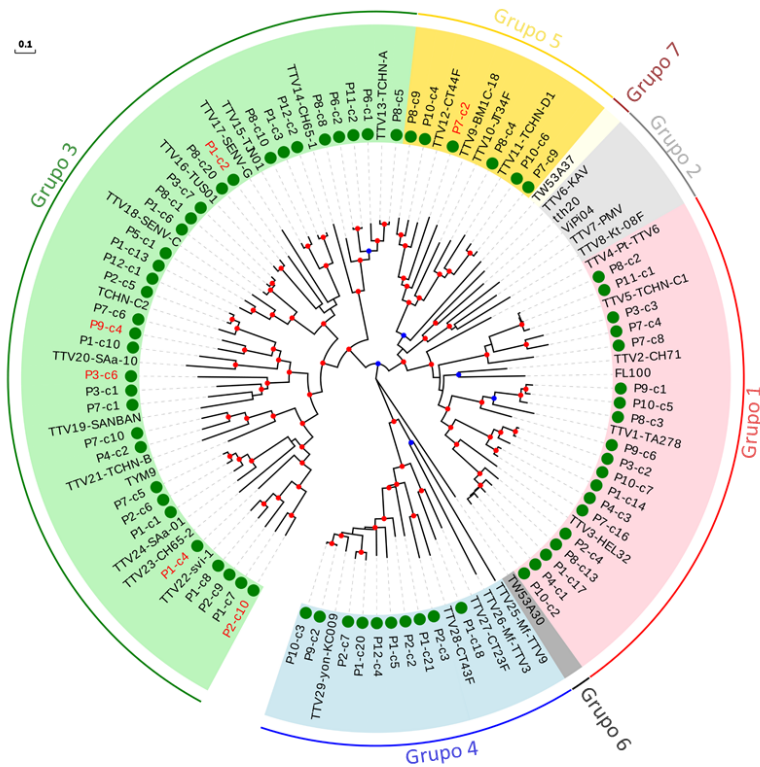


Figura 10: Árbol filogenético para la ORF1 que incluye los genotipos representativos del género TTV. Las secuencias descritas en este capítulo están marcadas con un círculo verde. Las secuencias que podrían considerarse como nuevas especies están rotuladas en rojo. Los intervalos de valores de *bootstrap* de 0,7-0,85 y 0,85-1,0 se indican con círculos azules y rojos, respectivamente. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

A continuación, se ha construido un árbol filogenético que incluye las 29 secuencias de este estudio que pertenecen al género TTMV y los 38 genotipos representativos previamente descritos que cumplen los criterios de demarcación de las especies (**Figura 11A**). Sorprendentemente, a pesar de que se han descrito un menor número de nuevas secuencias de TTMV en comparación con el género TTV (29 frente a 68), la mayoría pueden

considerarse como nuevas especies (Tabla Suplementaria S16). En total, se han definido quince especies nuevas, tres de las cuales incluían dos secuencias, mientras que los 11 aislados restantes se han agrupado con siete de las 38 especies previamente descritas.

En el caso del género TTMDV, se han identificado 17 nuevas secuencias y se han utilizado para construir un árbol que también incluye 17 genotipos representativos de especies conocidas. Al igual que para TTMV, se observa que la mayoría de las nuevas secuencias se asignan a nuevas especies (Figura 11B y Tabla Suplementaria S17). Diez de las secuencias descritas en este capítulo han definido 9 especies nuevas, mientras que las 7 secuencias restantes de TTMDV se agruparon con cinco de los 17 genotipos representativos pertenecientes a especies previamente descritas.

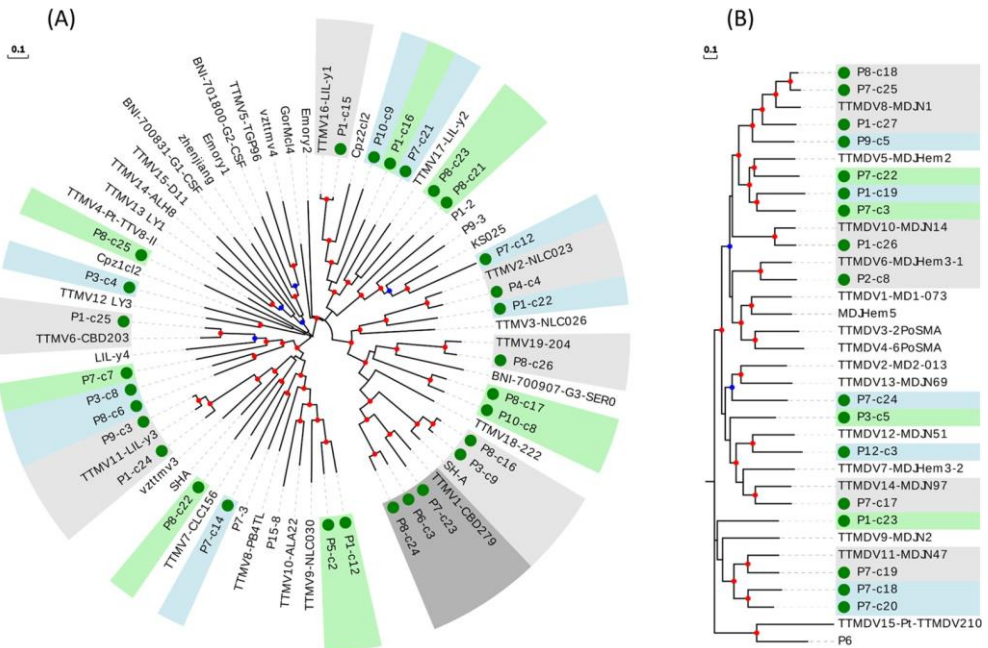


Figura 11: Árboles filogenéticos para la ORF1 en los que se incluyen los genotipos representativos de los géneros TTMV (A) y TTMDV (B). Las secuencias descritas en este estudio están marcadas con un círculo verde. Las nuevas especies (que incluyen

una o más secuencias nuevas) se indican con fondo de color verde o azul para distinguir los grupos contiguos. Los grupos de especies representativas que incluyen nuevas secuencias se indican con fondo de color gris claro u oscuro para distinguir grupos contiguos. Los intervalos de valores de *bootstrap* de 0,7-0,85 y 0,85-1,0 se indican con círculos azules y rojos, respectivamente. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

En el paso de ensamblaje de las lecturas de secuenciación masiva se pueden generar artefactos, lo cual puede afectar a la fiabilidad del análisis filogenético. Dado que esta posibilidad es especialmente relevante para la asignación de nuevas especies, se han reanalizado aquellas muestras de ADN en las que se detectaron 24 de las nuevas especies propuestas para TTV, TTMV y TTMDV (4, 11 y 9 muestras, respectivamente). Los datos de secuenciación masiva de estos genomas correspondientes a las nuevas especies muestran claras diferencias en cuanto a la profundidad de cobertura media después del ensamblaje, que oscila entre 4,9× y 1.747×. Para cada una de las especies, se ha realizado una amplificación por PCR de la ORF1 completa y una posterior secuenciación por el método Sanger (**Tabla Suplementaria S8**). En todos los casos, la secuenciación por Sanger ha confirmado las secuencias de la ORF1 deducidas con anterioridad, lo que pone de manifiesto la fiabilidad del paso de ensamblado seguido en nuestro estudio.

2. Discusión

La realización de estudios del viroma sanguíneo a gran escala es una herramienta poderosa para la detección precoz de potenciales virus emergentes que pueden causar infecciones crónicas o presentar largas fases asintomáticas. Sin embargo, hasta la fecha no se ha generalizado el establecimiento de programas de vigilancia basados en este enfoque. Por otro lado, los resultados obtenidos en este capítulo muestran que el uso de controles adecuados es esencial en estos estudios, ya que las contaminaciones pueden dar lugar a falsos positivos (Naccache et al., 2013; Rosseel et al., 2014). En este capítulo, se han utilizado tres controles negativos a lo largo del protocolo experimental, y los taxones que se han identificado en estos controles han sido sustraídos computacionalmente de las muestras. Dado que las muestras han sido filtradas inicialmente y se han utilizado nucleasas para la digestión de los ácidos nucleicos libres, se esperaría que la fracción de ácidos nucleicos no virales se viera reducida drásticamente. Sin embargo, pese a estos tratamientos, los datos contienen una fracción significativa de lecturas bacterianas y humanas. Esto es reflejo de que la secuenciación de la fracción viral se verá claramente comprometida por la incorporación de ácidos nucleicos contaminantes a lo largo de nuestro protocolo. Como se ha dicho, esta dificultad técnica inherente a los estudios de viroma se puede paliar parcialmente mediante el uso de controles, que permitirá al menos descartar la presencia de agentes virales mediante el paso de sustracción bioinformática. En el presente trabajo, se ha utilizado el ensayo MDA para la amplificación aleatoria, que puede amplificar preferentemente virus circulares de ADN monocatenario (Cheval et al., 2011), como los anelovirus. Este sesgo de amplificación puede explicar parcialmente la elevada presencia de esta familia

en los resultados; no obstante, la detección de un virus de ARN confirma la robustez del procedimiento propuesto.

En el contexto actual, la metagenómica viral debería beneficiarse en gran medida de la implementación de procedimientos que incluyan pasos de amplificación, purificación y enriquecimiento previos, como el descrito aquí, ya que esto aumenta la sensibilidad (Hall et al., 2014; Kohl et al., 2015). Como muestra de esta tendencia, un estudio reciente ha analizado la distribución de anelovirus en pequeños mamíferos, utilizando para ello un método de purificación de las muestras basado en ultracentrifugación en gradiente de sacarosa (de Souza et al., 2018). En este estudio se han detectado 11 posibles especies nuevas y se ha propuesto la inclusión de dos nuevos géneros en la familia *Anelloviridae*.

En consonancia con estudios anteriores, los resultados obtenidos en este capítulo muestran que la diversidad de los anelovirus es particularmente notable en comparación con otras familias virales (Kaczorowska & Van Der Hoek, 2020). En este sentido, el hecho de que los aislados de primates humanos y no humanos se agrupen filogenéticamente (Fahsbender et al., 2017) sugiere que los anelovirus son una familia antigua y que la gran diversidad genética presente en esta familia es la consecuencia de millones de años de evolución. Recientemente, se ha propuesto la incorporación al esquema de clasificación del ICTV de la creciente cantidad de secuencias víricas identificadas mediante análisis metagenómicos (Simmonds et al., 2017). Esta inclusión, que debería requerir un control de calidad adecuado, es importante para obtener una imagen más realista de la diversidad viral global. Aunque esta propuesta es especialmente relevante para las muestras ambientales, también lo es particularmente en el caso de la familia

Anelloviridae, donde no existe un reflejo en el ICTV del constante aumento de la diversidad descrita en esta familia.

La mayoría de las secuencias detectadas en nuestro estudio pertenecen al género TTV, que ha sido el más ampliamente estudiado respecto a los otros géneros de anelovirus. Sin embargo, las potenciales especies nuevas se han encontrado principalmente entre los géneros TTMV y TTMDV. Esto puede ser debido a que estos últimos sean más difíciles de detectar en los protocolos que carecen de pasos que implican el enriquecimiento de la fracción viral, lo que explicaría que permanecieran peor caracterizados. Por ello, nuestros resultados subrayan la importancia de utilizar métodos de enriquecimiento de la fracción viral para el estudio de la diversidad de anelovirus.

En otros trabajos, se ha propuesto que la carga viral de los anelovirus en sangre aumenta en pacientes inmunodeprimidos, como se ha descrito en pacientes trasplantados (Focosi, Antonelli, Pistello, & Maggi, 2016) y con VIH-1 (Li et al., 2013; Thom & Petrik, 2007). También se ha demostrado que la prevalencia de los anelovirus es menor en sujetos sanos que en pacientes con patologías comunes (Spandole-Dinu et al., 2018). En base a este planteamiento, se sugiere que la carga viral podría utilizarse como biomarcador del estado de salud de pacientes con afecciones crónicas, o incluso en personas sin patologías conocidas (Béland et al., 2014; Focosi, Macera, Pistello, & Maggi, 2014). Los miembros del género TTV también se han postulado como potenciales biomarcadores de contaminación antropogénica (Charest et al., 2015), de rechazo de injertos (De Vlamincq et al., 2013) y del estado inmunitario (Focosi et al., 2016). Sin embargo, es necesario aclarar mejor las relaciones causa-efecto entre la carga viral de TTV y el estado de salud.

La prevalencia de TTMV y de TTMDV es notablemente inferior a la de TTV (De Vlaminck et al., 2013; Spandole-Dinu et al., 2018). En general, aparte de algunas pruebas indirectas, los virus del género TTDMV no se han asociado a patologías (Burián et al., 2011). En cambio, muchas de las especies de anelovirus recientemente descritas pertenecientes al género TTMV se han asociado a algunas patologías específicas (Eibach et al., 2019; Ng, Dill, Camus, Delwart, & Van Meir, 2017; Pan et al., 2018; Zhang et al., 2016). No obstante, la asociación entre la presencia de un virus y una condición patológica no demuestra, necesariamente, la causalidad de ese virus. Como se ha indicado anteriormente, la carga viral de los anelovirus podría ser más bien una consecuencia del estado inmunitario del individuo. Una carga menor en individuos más sanos podría limitar la detección del virus, lo que podría conducir a una asociación estadística (pero causal) entre la presencia de un determinado virus y ciertas enfermedades. Un ejemplo ilustrativo de esta posibilidad viene dado por el genogrupo 2 de TTV, que se ha detectado con una frecuencia muy baja en la población sana (Maggi et al., 2005). Estudios basados en secuenciación y PCR cuantitativa, así como los resultados obtenidos en este capítulo, han mostrado que el genogrupo 2 está ausente o se detecta en bajas frecuencias en donantes sanos (Béland et al., 2014a; Burra et al., 2008; Gonzales-Gustavson et al., 2017), está esporádicamente ausente en pacientes trasplantados (Focosi et al., 2010), y se detecta en frecuencias más altas en pacientes inmunodeprimidos (Béland et al., 2014a; Burra et al., 2008; De Vlaminck et al., 2013; Segura-Wang, Görzer, Jaksch, & Puchhammer-Stöckl, 2018). Además, también se ha demostrado que la carga viral de TTV aumenta con el número de genogrupos de TTV que infectan simultáneamente a un paciente (Béland et al., 2014a; Maggi et al., 2005).

La era de la metagenómica ha dado lugar a una nueva perspectiva ecológica en la virología, en la que se evita considerar a los virus necesariamente como patógenos causantes de enfermedades (French & Holmes, 2020). En su lugar, los virus son considerados como componentes integrales de los ecosistemas que pueden causar esporádicamente enfermedades emergentes, pero también pueden ser beneficiosos para sus huéspedes (Kernbauer, Ding, & Cadwell, 2014; Roossinck, 2015). Los anelovirus humanos, y probablemente la mayoría de los miembros de esta familia, parecen ser esencialmente inocuos (Virgin et al., 2009). De hecho, se han sugerido efectos potencialmente beneficiosos para la salud humana (Kaczorowska & Van Der Hoek, 2020). Por ejemplo, la infección de los recién nacidos (Tyschik et al., 2018) podría promover el desarrollo y la maduración del sistema inmunitario (Virgin et al., 2009). Además, la detección del mismo tipo de TTV en muestras recogidas con 16 años de diferencia apoya la teoría de que las personas pueden permanecer infectadas crónicamente (Bédarida, Dussol, Signoli, & Biagini, 2017). Estos resultados concuerdan con una larga historia de coevolución entre el virus y el huésped, que puede desembocar en relaciones comensales o incluso mutualistas.

Capítulo 2: Análisis de la diversidad de los virus huérfanos presentes en plasma.

Nota: Los resultados de este capítulo se presentan en el trabajo:

Cebriá-Mendoza, M., Bracho, MA., Arbona, C., Larrea, L., Díaz, W., Sanjuán, R., Cuevas, JM. (2021). Exploring the Diversity of the Human Blood Virome. *Viruses*, **13**(11):2322
<https://doi.org/10.3390/v13112322>

En este capítulo, se ha llevado a cabo el análisis del viroma presente en un número mayor de muestras: se han analizado un total de 587 muestras agrupadas en 60 mezclas, que comprendían el plasma de ocho a trece individuos distintos. En estas muestras, se ha encontrado una gran diversidad de genomas virales, la gran mayoría de ellos pertenecientes a la familia *Anelloviridae*. También, se han identificado genomas pertenecientes a la familia *Flaviviridae*, en concreto, al género *Pegivirus*, aunque en una prevalencia mucho menor.

1. Resultados

1.1. Resultado global de la secuenciación

En este capítulo, se ha empleado el protocolo experimental implementado en el capítulo 1 para promover el enriquecimiento de la fracción viral presente en las muestras de plasma. Brevemente, se han analizado 60 mezclas de plasma, procedentes de 587 muestras individuales. Cada una de estas mezclas estaba compuesta por 1 mL de plasma procedente de ocho a trece muestras individuales de donantes sanos (**Tabla Suplementaria S18**). Como se ha descrito en la sección 3 de Material y Métodos, las mezclas se han filtrado con el fin de eliminar las bacterias y los restos celulares y se han sometido a una centrifugación a alta velocidad para sedimentar las partículas virales presentes. Tras la resuspensión de este sedimento, se ha llevado a cabo un tratamiento con nucleasas para digerir los ácidos nucleicos libres, y tras esto, se ha procedido a la extracción del ADN y el ARN viral de forma independiente. A continuación, los ácidos nucleicos extraídos, tras una reacción de transcripción inversa para la fracción de ARN, han sido sometidos a una amplificación aleatoria basada en $\phi 29$ (Shoib et al., 2008). Finalmente, los productos de amplificación obtenidos a partir de las

fracciones de ADN y ARN, se han mezclado equimolarmente como paso previo a la preparación de las librerías de secuenciación. Para determinar la eficacia de la recuperación de la fracción viral, se han añadido a todas las mezclas de plasma 10^3 PFU del bacteriófago ϕ X174 y 10^4 PFU de VSV. Dado que el protocolo de purificación puede arrastrar cantidades residuales de ácidos nucleicos, se han procesado ocho controles negativos en paralelo a las muestras, que han servido para evaluar el potencial riesgo de contaminación. Tras la clasificación taxonómica de las lecturas obtenidas en estos controles, se ha procedido a la sustracción de estos posibles contaminantes de las muestras reales.

Dado que el objetivo de este capítulo se centra en analizar la fracción viral, se ha realizado una clasificación taxonómica inicial con Centrifuge (Kim et al., 2016) para seleccionar aquellos taxones que se corresponden con familias virales. Posteriormente, se han eliminado las potenciales contaminaciones empleando el software Recentrifuge (Martí, 2019). Los virus ϕ X174 y VSV se han empleado como virus controles para evaluar la eficiencia de la recuperación del ADN y el ARN, respectivamente. Por otra parte, se han detectado lecturas clasificadas en la familia *Anelloviridae* en todas las mezclas, excepto en una (**Tabla 3 y Tablas Suplementarias S19 y S20**), con un número de lecturas que ha oscilado entre 10 y 1.580.534. Por otra parte, no se ha podido extraer ninguna conclusión clara al analizar la presencia del virus control de ADN (es decir, ϕ X174), ya que se ha recuperado en la mezcla que no mostraba lecturas de anelovirus, pero estaba ausente en cuatro mezclas en las que sí se han detectado lecturas de anelovirus (**Tablas Suplementarias S19 y S20**). En concordancia con su menor prevalencia, únicamente se han detectado lecturas de HPgV en diecisiete de las mezclas: una de ellas presentaba sólo 9 lecturas de HPgV, mientras que el resto mostraba un número

de lecturas que oscilaba entre 339 y 25.965 lecturas. Al analizar la presencia del virus de ARN control (es decir, VSV), no se han observado diferencias significativas en el número de lecturas recuperadas de VSV entre las mezclas donde se detectó HPgV y el resto de mezclas (prueba *t*: $P = 0,37$), lo que sugiere que la detección de HPgV no ha estado condicionada por un sesgo experimental.

Tabla 3: Resumen de los resultados del viroma para las 60 mezclas analizadas. Para cada mezcla, se indica el número de lecturas clasificadas como anelovirus, pegivirus y otros virus. Los recuentos de las lecturas virales procedentes de los virus control han sido omitidos para evitar confusión.

Mezcla	Lecturas de anelovirus	Lecturas de pegivirus	Otros virus	Mezcla	Lecturas de anelovirus	Lecturas de pegivirus	Otros virus
SP1	101.069	25.965	64	SP31	332.437	0	816
SP2	1.580.534	0	3.013	SP32	1.011	0	223
SP3	131.969	3.669	421	SP33	72.784	2033	68
SP4	9.992	4.250	61	SP34	270.083	0	57
SP5	47.927	0	225	SP35	419.986	0	317
SP6	718.633	0	330	SP36	185.281	0	144
SP7	63.139	0	80	SP37	666.063	0	1.311
SP8	76.204	0	5.089	SP38	242.853	2261	479
SP9	153.491	0	52	SP39	15.756	0	200
SP10	30.175	0	1.649	SP40	342.193	0	3.390
SP11	9.787	5.706	143	SP41	169.614	0	2.815
SP12	57.559	0	15.397	SP42	4.519	0	118
SP13	95.922	1.173	4.844	SP43	206.185	0	99
SP14	271.731	0	1.757	SP44	7.975	10.713	19
SP15	141.896	0	37	SP45	124.171	0	210
SP16	149.985	9	2.610	SP46	29.728	0	431
SP17	10	0	74	SP47	150.531	0	3.731
SP18	24.134	0	7.168	SP48	45.430	0	676
SP19	74.391	339	9.506	SP49	94.919	5.226	255
SP20	73.067	373	21	SP50	0	0	340
SP21	124.389	0	4.952	SP51	299.530	0	17
SP22	51.168	0	3.428	SP52	59.852	0	16
SP23	51.730	0	557	SP53	14.323	5.344	68
SP24	71.389	0	7.583	SP54	2.404	0	7.634
SP25	4.269	0	262	SP55	663	0	121
SP26	27.676	0	84	SP56	25.673	3.523	131
SP27	7659	0	3.030	SP57	52.296	0	1.737
SP28	96.187	0	270	SP58	1232	2.158	241
SP29	334.689	6.606	18.366	SP59	157.753	0	46
SP30	69.110	6.924	156	SP60	36.470	0	324

Se han detectado otros virus en todas las mezclas, pero en una proporción mucho menor. De hecho, al considerar globalmente los resultados de este capítulo, el 97,7% de las lecturas virales pertenecen a la familia *Anelloviridae*, casi el 1% pertenece a la familia *Flaviviridae* (en concreto a HPgV) y la fracción restante incluye a 46 familias virales (**Figura 12** y **Tabla Suplementaria S19**). La gran diversidad presente en esta fracción residual de lecturas virales sugiere que éstas pueden corresponder a identificaciones taxonómicas erróneas o bien, a la amplificación de cantidades residuales de ácidos nucleicos virales potencialmente presentes en los distintos reactivos empleados en el protocolo de enriquecimiento y que no han sido eficientemente sustraídas en el procesado bioinformático. Como se comentó en el capítulo 1, esta hipótesis se veía apoyada por el hecho de que el programa Recentrifuge presentaba dificultades frente a la existencia de ambigüedades en la clasificación taxonómica de las lecturas, lo que podría limitar la capacidad de eliminar posibles contaminaciones. En este sentido, la detección de la familia *Circoviridae*, que representa la tercera familia más abundante en este capítulo, se asocia comúnmente con reactivos contaminantes (Asplund et al., 2019). Además, la mayoría de los grupos taxonómicos identificados se asociaban con virus que infectan bacterias, algas, protozoos y hongos. Por otro lado, se han encontrado lecturas que potencialmente podían estar asociadas a patógenos humanos. En estos casos, se han mapeado estas lecturas frente a las correspondientes secuencias de referencia asignadas por Centrifuge. Este mapeo no ha sido exitoso, lo que indica que existen errores en la clasificación taxonómica.

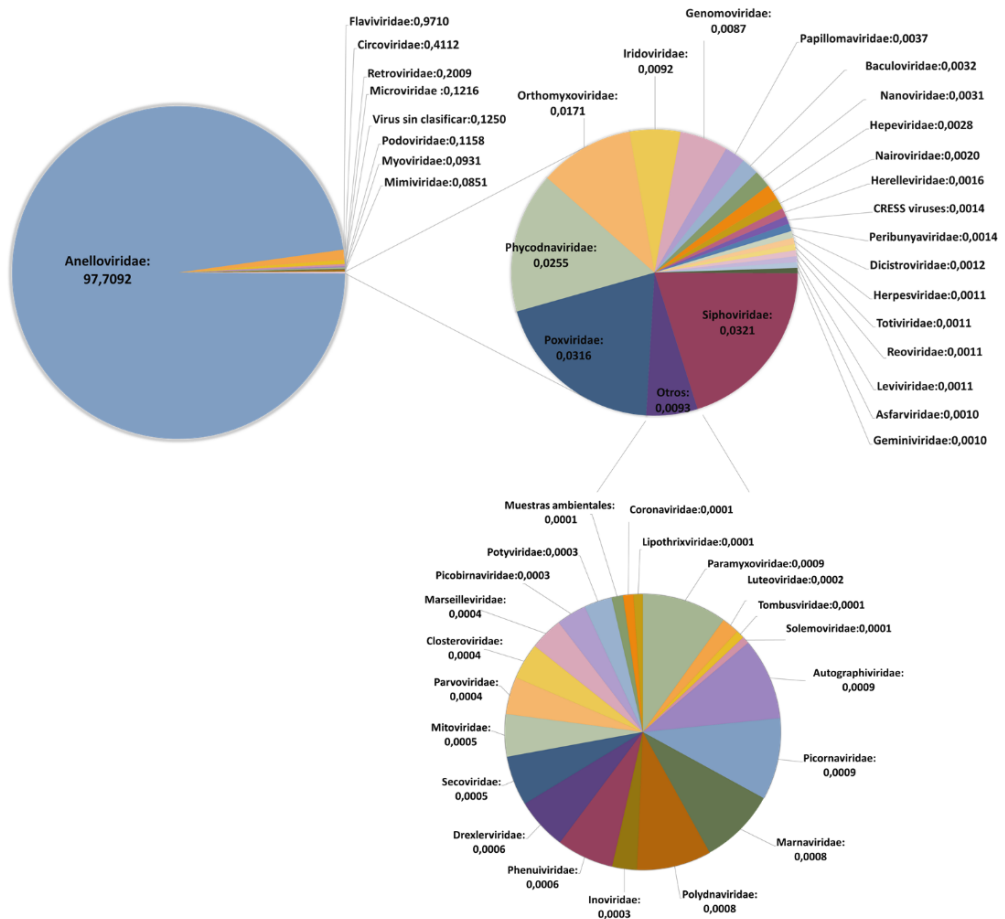


Figura 12: Descripción del viroma caracterizado en este estudio. La clasificación se muestra a nivel de familia y las frecuencias se han obtenido excluyendo la contribución de los virus control.

Curiosamente, se ha recuperado un genoma completo de un bacteriófago de ADN monocatenario y circular, que pertenece a la familia *Microviridae*, en las mezclas SP47 y SP57, donde se presentaban 3.415 y 1.716 lecturas de este virus, respectivamente. Se han encontrado 6 ORF putativas presentes en el aislado mediante ORF Finder y las ORF obtenidas se han

analizado mediante BLASTp. Estas secuencias presentan porcentajes de homología que oscilan entre el 48,3 y el 61,5% con la secuencia de referencia más cercana (**Tabla Suplementaria S21**). Este resultado pone de manifiesto la sensibilidad de este protocolo de purificación, que ha sido capaz de recuperar genomas completos de un virus que probablemente proceda de una contaminación, aunque hay que tener en cuenta que la polimerasa $\phi 29$ amplifica preferentemente los virus circulares (Shoaib et al., 2008).

Por último, aunque el enfoque experimental permitiría la detección de virus de gran tamaño, en este trabajo únicamente se han observado una evidencia residual de la presencia del género *Marseillevirus* (**Tabla Suplementaria S19**), lo que concuerda con estudios anteriores que lo asignan como un contaminante de laboratorio (Phan et al., 2015; Sauvage et al., 2014).

1.2. Análisis de los anelovirus

Para cada una de las sesenta mezclas, se han generado *contigs* a partir de todas las lecturas, independientemente de su clasificación taxonómica preliminar, lo que resulta eficaz para la detección de nuevos anelovirus (de Souza et al., 2018). Mediante los análisis realizados con BLAST, se ha podido detectar las secuencias correspondientes a los virus control y HPgV, pero la gran mayoría de las secuencias corresponden a anelovirus. En concreto, se han asignado 332 *contigs* a esta familia, de los cuales 69 mostraban extremos solapantes, por lo que pueden considerarse genomas completos (**Tablas Suplementarias S2 y S20**). Al igual que ocurría en el capítulo 1, se ha observado una correlación positiva entre el número de *contigs* obtenidos y la cantidad total de lecturas correspondientes a la familia *Anelloviridae* en cada mezcla (correlación de Spearman: $\rho = 0,414$; $P = 0,001$). De los 332 *contigs*, se ha obtenido la ORF1 completa en todos los casos, salvo en 17 (94,9%). Esta

secuencia nucleotídica de la ORF1 se ha empleado para el análisis filogenético y la identificación de nuevas especies. Para ello, en primer lugar, se ha construido un árbol filogenético por máxima verosimilitud donde se incluían nuestras secuencias y las de las especies de referencia recientemente propuestas por el ICTV (**Tabla Suplementaria S6**). Esta primera clasificación ha permitido asignar las secuencias descritas en este capítulo a los diferentes géneros de la familia *Anelloviridae* que infectan a humanos: se han asignado 160 secuencias al género TTV, 111 al TTMV y 61 al género TTMDV (**Tabla Suplementaria S2 y Figura Suplementaria S3**). Sesenta y siete de los 69 *contigs* considerados como genomas completos pertenecen al género TTMV, uno pertenece al género TTV y el restante al género TTMDV. Como se ha mencionado en el capítulo anterior, esto es coherente con la presencia de regiones ricas en GC más cortas en TTMV (Ninomiya et al., 2007), que pueden favorecer una mayor eficiencia en el ensamblaje.

La metodología establecida para la clasificación de las especies de anelovirus se ha modificado muy recientemente y el número de especies de referencia se ha actualizado en consecuencia (Varsani et al., 2021). Teniendo en cuenta esto, se ha decidido reevaluar los datos del capítulo anterior, donde se aplicaban los mismos procedimientos experimentales y bioinformáticos a un número menor de muestras. Esta reevaluación ha permitido describir 26 nuevas especies (6, 11 y 9, para TTV, TTMV y TTMDV, respectivamente; **Tabla 4 y Tablas Suplementarias S22-S24**). Estas nuevas especies se han utilizado, junto con las especies de referencia, para caracterizar las secuencias analizadas en el presente capítulo.

Tabla 4: Resumen del análisis de las secuencias de anelovirus. ¹Número de especies de referencia actualmente aceptadas por el ICTV para cada género. ²Resultados obtenidos tras reevaluar los datos del capítulo anterior utilizando las especies actualmente aceptadas y el criterio de demarcación de especies recientemente propuesto por el ICTV. ³Resultados obtenidos tras analizar las secuencias descritas en el capítulo 2. ⁴Asignación de género para las secuencias descritas. ⁵Número de especies nuevas (entre paréntesis se indica el porcentaje con respecto al número total de secuencias descritas para cada género). ⁶Número de especies que se agrupan con al menos una nueva secuencia (el porcentaje con respecto al número total de especies se indica entre paréntesis). Las nuevas especies identificadas en el capítulo 1 también se han empleado como especies de referencia en los posteriores análisis filogenéticos y de identidad por pares de las secuencias descritas en el capítulo 2.

	Capítulo 1 ²				Capítulo 2 ³		
	Especies ¹	Secuencias ⁴	Nuevas especies ⁵	Grupos coincidentes ⁶	Secuencias ⁴	Nuevas especies ⁵	Grupos coincidentes ⁶
TTV	26	68	6 (8,8)	13 (50,0)	160	6 (3,8)	20 (62,5)
TTMV	38	29	11 (37,9)	11 (28,9)	111	27 (24,3)	24 (49,0)
TTMDV	15	17	9 (52,9)	5 (33,3)	61	17 (27,9)	16 (66,6)
Total	79	114	26 (22,8)	29 (36,7)	332	50 (15,1)	60 (57,1)

Para poder observar los datos de una manera más clara, la caracterización de las secuencias descritas en el capítulo 2 se ha llevado a cabo mediante la elaboración de árboles filogenéticos y matrices de identidad de manera independiente para cada género. En estos análisis, sólo se han incluido los aislados considerados como especies de referencia y aquellas secuencias descritas como tal tras la reevaluación de los datos del capítulo 1 (**Tabla Suplementaria S6**).

Para el género TTV, que a su vez consta de siete grupos filogenéticos (Hsiao et al., 2016), el árbol y la matriz han incluido las 160 nuevas secuencias, las 26 especies de referencia y las 6 especies obtenidas tras la reevaluación de los datos del capítulo 1 (**Figura 13, Tabla 4 y Tabla Suplementaria S25**). El análisis del árbol, junto con los valores obtenidos en la matriz de identidades por pares, indica que 23 de las secuencias pueden considerarse pertenecientes

a seis nuevas especies (**Tabla 4 y Figura Suplementaria S4**), mientras que las secuencias restantes se agrupan dentro del 62,5% (20 de 32) de las especies de referencia, aunque este porcentaje aumenta hasta el 87,0% (20 de 23) cuando se excluyen los aislados que infectan a primates no homínidos, que no estaban relacionados con ninguna de nuestras secuencias descritas (**Figura 13 y Tabla Suplementaria S6**). La variabilidad del género TTV obtenida en este capítulo ha cubierto una fracción importante de la diversidad mundial para este género, pero la distribución de las secuencias dentro de cada especie es muy variable (**Figura Suplementaria S4 y Tabla Suplementaria S25**). Por ejemplo, cuatro especies se han agrupado con una única secuencia descrita, mientras que las especies representadas por los aislados TTV24-SAA-01, TTV18-SENV-C y TTV29-yon-KC009 se han agrupado con veinticinco, veinte y dieciocho de nuestras secuencias, respectivamente. En términos globales, nuestras secuencias se han agrupado dentro de especies pertenecientes a todos los grupos de TTV propuestos, excepto en el grupo 6, que sólo incluye un aislado identificado en indígenas del este de Taiwán (Hsiao et al., 2016), y que actualmente no está considerado como especie de referencia por el ICTV. Como ya se observó en el capítulo 1, existe una correlación positiva entre el número de especies incluidas en cada grupo y el número de nuevas secuencias descritas (coeficiente de correlación de Spearman; $\rho = 0,971$; $P < 0,01$).

0.1

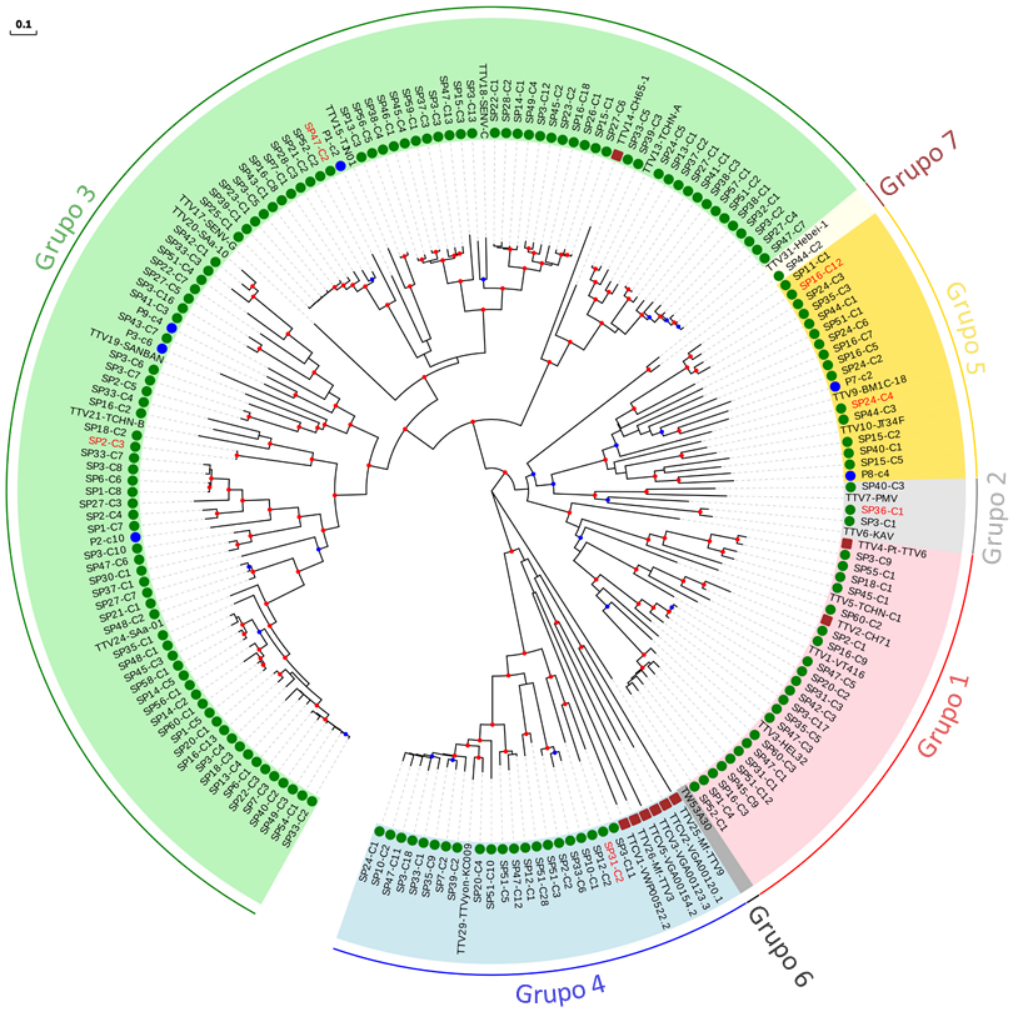


Figura 13: Árbol filogenético para la ORF1 del género TTV, donde se incluyen las especies de referencia. Las secuencias descritas en este capítulo están marcadas con un círculo verde. Las secuencias que podrían considerarse como nuevas especies están etiquetadas en rojo. Las secuencias identificadas como nuevas especies, tras reevaluar los datos del capítulo 1, están marcadas con un círculo azul. Los aislados de primates no homínidos están marcados con un cuadrado marrón. Los intervalos de valores de *bootstrap* de 0,7-0,85 y 0,85-1,0 se indican con círculos azules y rojos, respectivamente. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

A continuación, se ha construido un árbol filogenético que incluye las 111 secuencias pertenecientes al género TTMV, las 38 especies de referencia y las 11 nuevas especies descritas tras la reevaluación de los datos del capítulo 1 (**Figura 14** y **Tabla 4**). Cabe destacar que 40 de las secuencias descritas se pueden considerar como pertenecientes a 27 nuevas especies (**Tabla 4** y **Tabla Suplementaria S26**), lo que supondría un notable incremento en la diversidad del género TTMV. Las 71 secuencias restantes se han agrupado dentro del 49,0% (24 de 49) de las especies incluidas, y este porcentaje ha aumentado hasta el 53,3% (24 de 45) cuando se excluyen los aislados de primates no homínidos, que no se han agrupado con ninguna de las secuencias descritas en este capítulo (**Figura 14** y **Tabla Suplementaria S6**).

0.1

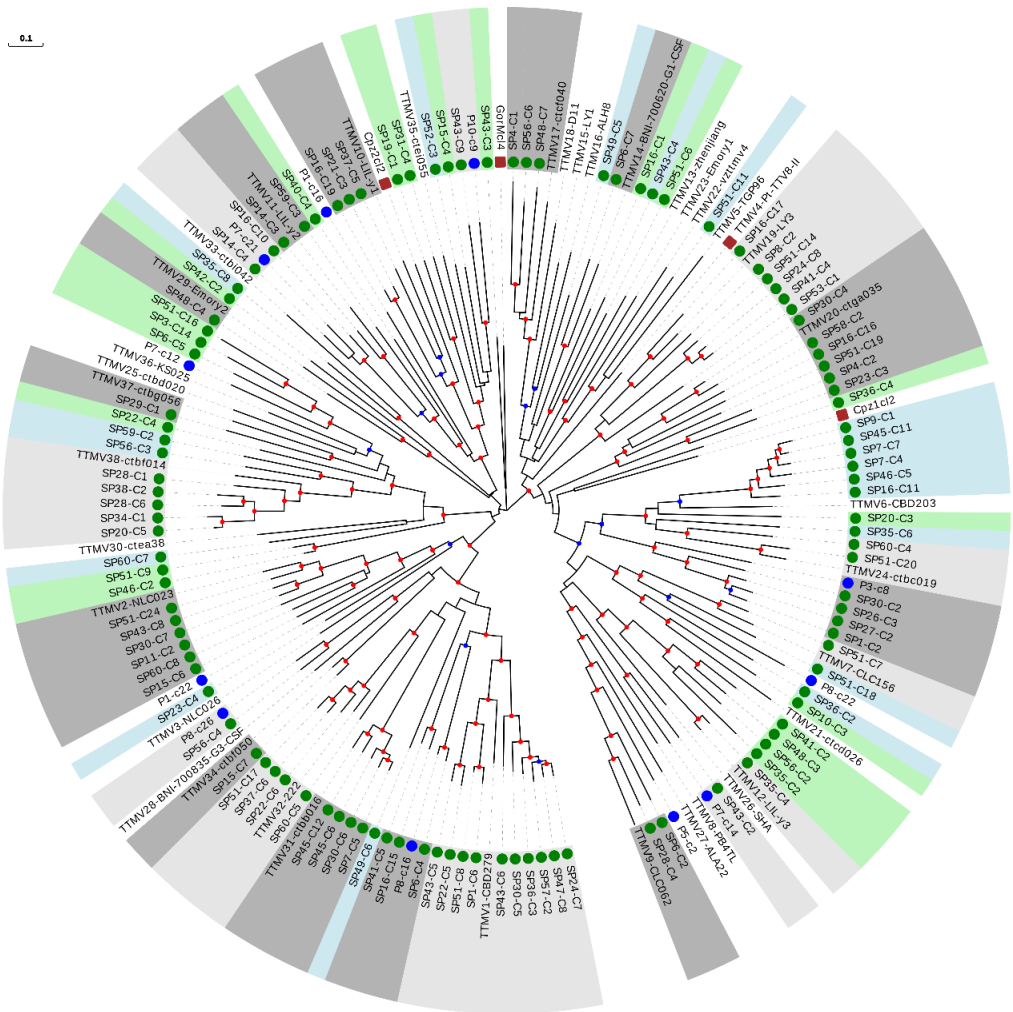


Figura 14: Árbol filogenético para la ORF1 del género TTMV, donde se incluyen las especies de referencia. Las secuencias descritas en este capítulo están marcadas con un círculo verde. Las secuencias identificadas como nuevas especies tras reevaluar los datos del capítulo 1 están marcadas con un círculo azul. Las nuevas especies (que incluyen una o más secuencias nuevas) se indican con el color de fondo verde o azul para distinguir los grupos contiguos. Los grupos de especies de referencia que incluyen nuevas secuencias se indican con colores de fondo gris claro u oscuro para distinguir los grupos contiguos. Los aislados de primates no homínidos están marcados con un cuadrado marrón. Los intervalos de valores de *bootstrap* de 0,7-0,85 y 0,85-1,0 se indican con círculos azules y rojos, respectivamente. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

Por último, para el género TTMDV se ha construido un árbol que ha incluido las 61 nuevas secuencias descritas, las 15 especies de referencia según el ICTV y las 9 especies recién descritas tras la reevaluación de los datos del capítulo 1 (**Figura 15, Tabla 4**). Tras el análisis, 24 de nuestras secuencias se han asignado a 17 nuevas especies (**Tabla 4 y Tabla Suplementaria S27**), lo que ha implicado un gran incremento en la diversidad del género TTMDV, de manera similar a lo observado para el género TTMV. Las 37 secuencias restantes se han agrupado dentro del 66,6% (16 de 24) de las especies de referencia, y de manera sorprendente, también dentro del único aislado descrito en el género TTMDV de primates no homínidos.

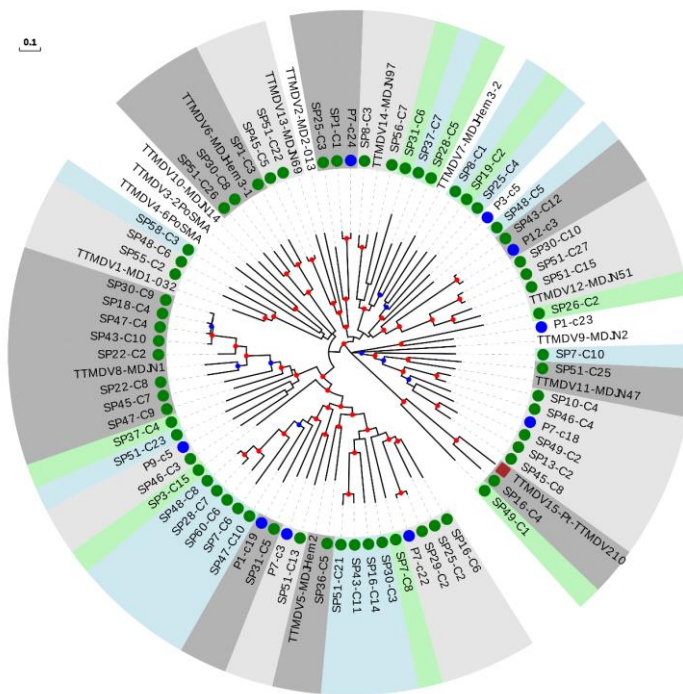


Figura 15: Árbol filogenético para la ORF1 del género TTMDV, donde se incluyen las especies de referencia. Las secuencias descritas en este capítulo están marcadas con un círculo verde. Las secuencias identificadas como nuevas especies, tras reevaluar los datos del capítulo 1, se marcaron con un círculo azul. Las nuevas especies (que

incluyen una o más secuencias nuevas) se indican con el fondo de color verde o azul para distinguir los grupos contiguos. Los grupos de especies de referencia que incluyen nuevas secuencias se indican con colores de fondo gris claro u oscuro para distinguir los grupos contiguos. El aislado de primate no homínido está marcado con un cuadrado marrón. Los intervalos de valores de *bootstrap* de 0,7-0,85 y 0,85-1,0 se indican con círculos azules y rojos, respectivamente. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

Como se ha comentado anteriormente, la reevaluación de los resultados obtenidos en el capítulo 1 ha conducido a la identificación de 26 potenciales nuevas especies de anelovirus, la mayoría de ellas pertenecientes a los géneros TTMV y TTMDV (**Tabla 4**). Pese a incorporar estas 26 especies al conjunto de especies de referencia, en este segundo capítulo se ha conseguido identificar 50 nuevas especies adicionales en el conjunto de las secuencias descritas. De nuevo, aunque casi la mitad de las secuencias se han asignado como TTV, sólo 6 de las 50 nuevas especies descritas han correspondido a este género. La comparación de los resultados obtenidos en los capítulos 1 y 2 nos puede ayudar a inferir la potencial diversidad esperada para los aislados de la familia *Anelloviridae* (**Tabla 4**). Para el género TTV, se ha observado que el porcentaje de especies nuevas descritas ha disminuido del 8,8 al 3,8%, lo que sugiere que se ha descrito una fracción significativa de la diversidad de este género en humanos. En cambio, al hacer esta comparación para los géneros TTMV y TTMDV, el porcentaje de especies nuevas descritas era moderadamente mayor en el capítulo 1 (37,9% para TTMV y 52,9% para TTMDV) que en el capítulo 2 (24,3% para TTMV y 27,9% para TTMDV). En consecuencia, estos resultados sugieren que la diversidad global de los géneros TTMV y TTMDV está aún lejos de ser descrita. Esto está de acuerdo con la menor prevalencia que caracteriza a los géneros TTMV y TTMDV respecto al género TTV (De Vlamincq et al., 2013). Esta menor prevalencia está potencialmente asociada a títulos virales

más bajos que comprometen la detección cuando no se aplican protocolos de enriquecimiento viral, como el empleado en este trabajo.

Por otra parte, no se ha observado ninguna evidencia de compartimentación geográfica de las secuencias descritas (**Tabla Suplementaria S6**). Para testar esto, se construyeron tablas de contingencia dos-a-dos en las que las especies de referencia eran clasificadas en función de si agrupaban con alguna de nuestras secuencias y de si tenían o no origen español/europeo. En ningún caso se observaron asociaciones significativas (pruebas exactas de Fisher, $P > 0,05$ para todos los análisis realizados de forma global e independiente para cada género de anelovirus). Esta falta de asociación puede ser debida a que las especies agrupadas con un mayor número de secuencias para cada género (TTV24-SAA-01, TTMV1-CBD279 y TTMDV8-MDJN1, con 25, 10 y 8 secuencias, respectivamente) son de origen asiático (**Figuras 13-15** y **Tabla Suplementaria S6**). Curiosamente, también se ha encontrado una secuencia de TTV que se agrupaba con el grupo 7 recientemente propuesto y detectado en indígenas del este de Taiwán (Hsiao et al., 2016) (**Figura 13**).

Mediante ensayos de PCR para la detección diferencial de anelovirus humanos, en chimpancés se ha observado la presencia habitual de ADN de TTV y TTMV (Ninomiya et al., 2009), lo que sugiere la existencia de transmisión cruzada entre ambas especies. En concordancia con esto, el análisis filogenético corrobora que, tanto los aislados de TTV como los de TTMV que infectan a las especies de primates no homínidos, se encuentran intercalados con las especies de TTV y TTMV de humanos, respectivamente, aunque ninguna de las secuencias descritas en nuestro estudio se ha agrupado a nivel de especie con aislados de primates no homínidos (**Figuras 13** y **14**). Por el contrario, se ha propuesto que el género TTMDV de chimpancés y humanos

está separado filogenéticamente (Ninomiya et al., 2009), aunque esto podría ser debido a un sesgo causado por un muestreo reducido con respecto a los géneros TTV y TTMV. De acuerdo con esto, una secuencia descrita en este capítulo se agrupaba con la única especie de referencia que infecta a un primate no homínido (**Figura 15** y **Tabla Suplementaria S26**). Esto sugiere que las relaciones filogenéticas entre los aislados humanos y de primates no homínidos son similares para los tres géneros y que las diferencias aparentes se deben, probablemente, a variaciones en la profundidad del muestreo.

1.3. Análisis de HPgV

Respecto a HPgV, se han detectado 17 mezclas positivas (**Tabla 5**), si bien la mezcla **SP16** ha sido excluida de posteriores análisis porque presentaba únicamente 9 lecturas. El resto de mezclas positivas han sido analizadas y comparadas frente al genoma de referencia (número de acceso en GenBank U44402). Se ha visto que, tras el ensamblaje de las lecturas en cada mezcla, la cobertura de los aislados obtenidos oscilaba entre el 70,2 y el 99,6% respecto al genoma de referencia completo, presentando valores de profundidad media de cobertura que oscilaban entre 12,4× y 1010,7× (**Tabla 5**). En el caso de la mezcla **SP16**, donde sólo se detectaron 9 lecturas de HPgV, se ha obtenido un único *contig* de 518 bases, que posteriormente se ha identificado como perteneciente al genotipo 2 tras el análisis por BLAST. Para la mezcla **SP53**, la secuencia consenso que se ha obtenido ha revelado la presencia de 219 ambigüedades, que podían estar causadas por la detección simultánea de dos aislados diferentes de HPgV. Para corroborarlo, se ha extraído individualmente el ARN de las diez muestras de plasma incluidas en esta mezcla, se ha obtenido el ADNc y se ha realizado una PCR específica para amplificar HPgV empleando cebadores diseñados en regiones conservadas. De esta forma, se han

identificado dos muestras positivas para HPgV en esta mezcla, confirmando nuestra sospecha inicial. Con esta información, se ha realizado un análisis de los *contigs* encontrados en esta mezcla, lo que ha permitido detectar la presencia de dos haplotipos diferentes que cubrían parcialmente la longitud del genoma de HPgV. Tras esto, se han diseñado cebadores específicos de ambos haplotipos, se han llevado las reacciones de amplificación a partir del ADNc y el producto de amplificación se ha secuenciado por Sanger con el fin de recuperar aquellas regiones que estaban ausentes, así como asignar los *contigs* pertenecientes a cada haplotipo sin ambigüedades.

En resumen, se ha detectado HPgV en diecisiete mezclas, una de las cuales mostraba una infección mixta. Esto supone una prevalencia del 3,1%, valor que concuerda con la prevalencia descrita en un estudio anterior donde se analizaban muestras individuales de sangre en población española (Forns et al., 1997). Excepto para la mezcla **SP16**, con objeto de recuperar la longitud completa de la secuencia codificante, se han diseñado cebadores específicos empleando las secuencias parciales de HPgV obtenidas de cada mezcla, lo que ha permitido su amplificación y posterior secuenciación por Sanger. De esta forma, se han caracterizado 17 aislados diferentes, dos de ellos pertenecientes a la mezcla **SP53**. A partir de la secuencia nucleotídica que codifica para la poliproteína (que incluye más del 90% del genoma de HPgV), se ha realizado un análisis filogenético por máxima verosimilitud. Para ello, se han utilizado las 17 secuencias descritas, junto con todas las secuencias nucleotídicas de la poliproteína de los aislados disponibles en las bases de datos (**Tabla Suplementaria S3**). En este análisis se puede observar que 15 de las secuencias descritas pertenecen al genotipo 2 (**Figura 16**) y, en concreto, diez secuencias se clasifican dentro del subtipo a y las cinco secuencias restantes, en el subtipo b. Por otro lado, la secuencia HPgV-SP30 se ha clasificado dentro del genotipo

1 y la secuencia HPgV-SP49 ha caído en una posición basal que se encuentra relativamente cercana a los aislados del genotipo 3.

Dada la posición intermedia de los aislados HPgV-SP30 y HPgV-SP49 entre los genotipos definidos en la filogenia, se ha sugerido que estos aislados se podrían tratar de genomas recombinantes. Para comprobar esto, se ha realizado un análisis en el que se ha evaluado la filogenia (**Figura 16**). La red filogenética construida (**Figura Suplementaria S5**) ha mostrado que tanto el aislado HPgV-SP30 como HPgV-SP49, de forma similar a lo que ocurría con secuencias recombinantes previamente descritas, parecen estar involucrados en una historia evolutiva reticulada en la que subyacen sucesos de recombinación. Otros análisis de recombinación (datos no mostrados) realizados con el software RDP4 (Martin et al., 2015) sugerían que HPgV-SP49 era un recombinante intergenotipo (genotipo 1/genotipo 3), mientras que HPgV-SP30 era un recombinante intragenotipo del genotipo 1. Una explicación alternativa a la existencia de recombinación podía ser la presencia de dos aislados diferentes de HPgV en cada mezcla. Para descartar esta posibilidad, se ha extraído individualmente el ARN de las diez muestras de plasma incluidas en cada mezcla, se ha obtenido el ADNc y se ha realizado una PCR específica para HPgV utilizando cebadores conservados. En ambos casos, sólo se ha identificado una muestra positiva para HPgV en cada mezcla, lo que apoya nuestra interpretación de los resultados.

Tabla 5: Resumen del análisis de HPgV. Para cada muestra, se indica el número total de lecturas clasificadas como HPgV, la profundidad media de cobertura, la cobertura del genoma completo y la cobertura de la región correspondiente a la poliproteína. Las estimaciones se han obtenido utilizando el número de acceso U44402 como secuencia de referencia. *Esta mezcla está formada por dos aislados diferentes de HPgV y no se pueden proporcionar estimaciones individuales de forma independiente.

Muestra/mezcla	Núm. de lecturas	Profundidad media de cobertura	Cobertura del genoma (%)	Cobertura de la poliproteína (%)
SP1	25.965	1010,7	98,4	98,6
SP3	3.669	130,1	94,2	95,6
SP4	4.250	157,7	92,5	93,8
SP11	5.706	204,0	99,2	100,0
SP13	1.173	40,7	96,4	97,4
SP16	9	2,4	5,5	6,0
SP19	339	12,4	82,7	89,9
SP20	373	13,5	70,2	76,4
SP29	6.606	228,7	98,2	99,2
SP30	6.924	230,1	89,7	91,1
SP33	2.033	71,1	92,6	94,8
SP38	2.261	82,7	99,6	100,0
SP44	10.713	392,0	99,1	100,0
SP49	5.226	165,7	83,4	84,0
SP53*	5.344	181,2	97,5	99,8
SP56	3.523	121,1	91,7	93,3
SP58	2.158	73,2	93,6	94,7

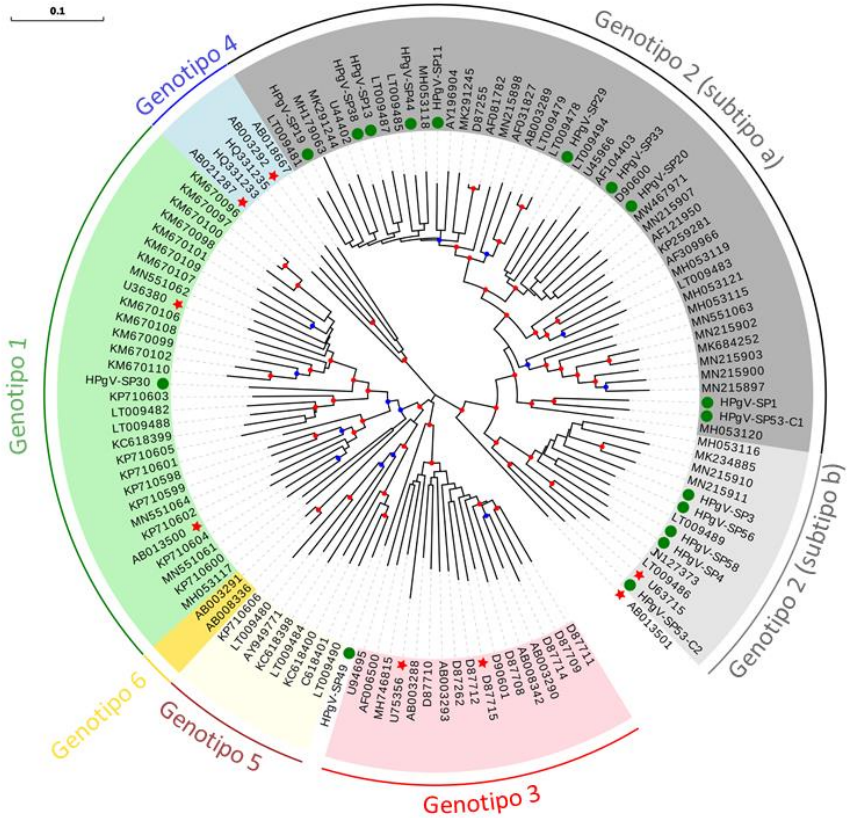


Figura 16: Árbol filogenético basado en la secuencia codificante completa (es decir, la poliproteína) que incluye los aislados de HPgV actualmente disponibles. Las secuencias descritas en este estudio se indican con un círculo verde. Los aislados previamente identificados como recombinantes se indican con una estrella roja. Para el genotipo 2, los subtipos 2a y 2b se muestran explícitamente. Los intervalos de valores de *bootstrap* de 0,75-0,9 y 0,9-1,0 se indican con círculos azules y rojos, respectivamente. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

2. Discusión

La diversidad viral está claramente infravalorada (Shi, Zhang, et al., 2018). Hasta hace poco, el uso de la PCR utilizando cebadores degenerados dirigidos a regiones conservadas ha demostrado ser eficaz para el descubrimiento de nuevos virus (Drexler et al., 2013). Sin embargo, este enfoque puede promover un fuerte sesgo en grupos que son altamente heterogéneos, como los anelovirus (Spandole et al., 2015). Una alternativa eficaz para el descubrimiento de nuevos virus es la metagenómica viral, enfoque que puede proporcionar un muestreo más realista, donde las diferencias en las tasas de detección serían consecuencia de parámetros naturales, como el título viral o la estabilidad de las partículas virales.

Los anelovirus son una familia antigua y se cree que su evolución ha seguido a la de los animales que infectan (Okamoto et al., 2002; Thom, Morrison, Lewis, & Simmonds, 2003). En los primates, esta hipótesis de coevolución ha sido cuestionada debido a que los aislados que infectan a humanos y chimpancés muestran una menor diversidad genética en el género TTMDV y a que cada especie presenta un estado monofilético (Ninomiya et al., 2009). Esto contrasta con los géneros TTV y TTMV, lo que sugiere que el género TTMDV se ha originado después de la especiación de chimpancés y humanos. Sin embargo, los resultados obtenidos en este capítulo sugieren que existe un origen común, ya que las aparentes discrepancias entre los diferentes géneros de anelovirus que infectan a primates se deben probablemente a un sesgo de muestreo. De acuerdo con esto, nuestros resultados sugieren que los géneros TTMV y TTMDV, descubiertos, respectivamente, tres y diez años después del género TTV (Spandole et al., 2015), se caracterizan por una notable diversidad, no apreciada con anterioridad debido al sesgo de amplificación. Por otra parte,

en el plasma humano se han observado cargas virales menores del género TTV que en otros compartimentos corporales (Reshetnyak, Maev, Burmistrov, Chekmazov, & Karlovich, 2020), lo que podría haber dado lugar a una subestimación de la diversidad real de anelovirus en nuestro estudio.

Inicialmente, se proponía que los genotipos del género TTV presentaban diferencias en cuanto a su distribución geográfica, pero estos estudios se basaban principalmente en la amplificación de los genomas con cebadores específicos y, por tanto, estaban sujetos a un sesgo de amplificación (Spandole et al., 2015). En este sentido, los resultados obtenidos en este capítulo muestran claramente que la diversidad de anelovirus carece de compartimentación geográfica, al menos en términos generales. Esto es particularmente notable para el caso del género TTV, donde nuestros resultados han permitido identificar el 87% de las especies descritas a nivel mundial. Además, la prevalencia de los anelovirus es muy variable y se requieren métodos de detección mediante amplificación inespecífica para proporcionar estimaciones más fiables, si bien parece indudable que la mayoría de los individuos sanos presentan infecciones crónicas (Spandole et al., 2015). Esta elevada prevalencia es consecuencia de las múltiples vías de transmisión que presenta el virus, incluyendo la parenteral, la sexual y la vertical, en combinación con un amplio politropismo (Reshetnyak et al., 2020). En el caso del género TTV, los datos disponibles sobre su prevalencia, tropismo y patogenicidad por la persistencia de su infección son muy contradictorios, lo que dificulta una evaluación del impacto sobre la salud (Reshetnyak et al., 2020).

Los títulos virales del género TTV son mayores en pacientes inmunodeprimidos respecto a los que presenta una persona sana, por lo que se ha sugerido que la patogénesis puede ser condicional (Reshetnyak et al.,

2020), es decir, el virus puede actuar como un factor agravante en la progresión de la patogénesis o bien como un agente oportunista (Spandole-Dinu et al., 2018). En este sentido, el aumento en la diversidad de los anelovirus obtenida en estudios en los que se promueve el enriquecimiento de la fracción viral, puede proporcionar más pistas sobre las potenciales asociaciones patológicas de determinadas variantes virales. En cualquier caso, los anelovirus son considerados habitualmente parte del viroma humano debido a su alta prevalencia y persistencia asintomática. De hecho, como ya se comentó en el capítulo 1, se ha propuesto utilizar la carga viral del género TTV como un marcador endógeno del estado inmunitario humano, algo que puede ser útil en salud pública. Un ejemplo de esta aplicación podría ser la cuantificación del nivel de ADN del género TTV presente en sangre de pacientes sometidos a un trasplante de órganos para controlar el posible desarrollo de una respuesta de rechazo al trasplante (Frye et al., 2019; Strassl et al., 2018).

Por otro lado, la falta de patogenicidad es uno de los criterios de clasificación de los virus como pegivirus (Stapleton et al., 2011), a pesar de que el descubrimiento de un pegivirus de caballo asociado a brotes de hepatitis aguda sugiere que, al menos un miembro del género *Pegivirus*, puede causar hepatitis (Chandriani et al., 2013). Recientemente, se ha descrito un segundo pegivirus humano, HPgV-2, que se encuentra estrechamente asociado a la infección por el virus de la hepatitis C (Kapoor et al., 2015). Como era de esperar, no se ha detectado este nuevo virus en las muestras analizadas en este capítulo, ya que presenta una prevalencia muy baja en la población general (Bonsall et al., 2016). En cualquier caso, HPgV-2 parece ser un virus linfotrópico (Wan et al., 2020) y, al igual que HPgV, sigue considerándose un virus huérfano, ya que no tiene una patología asociada. HPgV parece ser un

virus humano antiguo y la distribución de los genotipos a nivel global coincide con las antiguas migraciones humanas (Pavesi, 2001; Sharp & Simmonds, 2011). Como ejemplo de ello, la distribución del genotipo 3 podría explicarse en base a las migraciones ancestrales entre las zonas de África y el sudeste asiático (Pavesi, 2001). La infección por HPgV puede persistir durante décadas, pero la mayoría de los individuos sanos eliminan la viremia en un período de dos años tras la infección (Chivero & Stapleton, 2015). La evaluación de la prevalencia de HPgV mediante técnicas moleculares y/o serológicas ha mostrado una gran variabilidad en la población general (Mohr & Stapleton, 2009). La prevalencia mostrada en este capítulo está de acuerdo con estudios que mostraban que el ARN viral se detecta con muy baja frecuencia en donantes de sangre (Marano et al., 2017). Además, como ya se ha mencionado previamente, la prevalencia obtenida para HPgV es concordante con la descrita previamente para otra población española (Forns et al., 1997).

En la actualidad, existe un número relativamente bajo de genomas de HPgV en las bases de datos públicas que presenten la secuencia codificante completa. A pesar de esto, se puede observar que existe un claro predominio de los genotipos 2 y 3, lo que puede ser debido a que el muestreo se ha centrado en las zonas geográficas donde estos genotipos presentan mayor prevalencia. En este sentido, los resultados obtenidos en este capítulo muestran un predominio de aislados pertenecientes al genotipo 2. Este resultado está de acuerdo con estudios realizados en otros países europeos (Bonsall et al., 2016; Jordier et al., 2019). Este sesgo puede confundir los resultados de la inferencia filogenética, como demuestra la mayor diversidad genética sugerida para el genotipo 1 (Parreira, Branco, Piedade, & Esteves, 2012). Además de esto, detectar recombinación puede ser difícil entre variantes virales que son muy similares entre sí, como es el caso de HPgV (Ghai

et al., 2013). A pesar de estas dificultades, se ha visto que la recombinación probablemente provoca la observación de incongruencias filogenéticas entre las regiones subgenómicas de HPgV, tanto a escala intra como intergenotípica (Blackard et al., 2016; Neibecker, Schwarze-Zander, Rockstroh, Spengler, & Blackard, 2011; Parreira et al., 2012; Smith et al., 2000; Worobey & Holmes, 2001; Wu, Padhi, Xu, Gong, & Tien, 2016). Aunque parece claro que la recombinación no ha sido lo suficientemente generalizada como para desdibujar la estructura poblacional de HPgV (Worobey & Holmes, 2001), es un factor importante que debe considerarse a la hora de definir nuevos aislados. En este sentido, varios estudios han sugerido que el genotipo de HPgV que infecta a un individuo puede influir en el curso de la enfermedad del SIDA (Alcalde et al., 2010; Blackard et al., 2017; Muerhoff, Tillmann, Manns, Dawson, & Desai, 2003; Schwarze-Zander et al., 2006), pero en otros estudios no se ha encontrado dicha asociación potencial (Berzsenyi, Bowden, Roberts, & Revill, 2009; Miao et al., 2017). Además, existen algunos aislados no oficiales, es decir, que no están definidos por el ICTV (los aislados con números de acceso U63715, AB021287 y AB003292) que corresponden en realidad a secuencias recombinantes (Blackard et al., 2016; Parreira et al., 2012). Por esta razón, en el caso de aquellos estudios que buscan asociaciones entre el genotipo de HPgV y la progresión de alguna enfermedad, es necesario realizar una clasificación taxonómica precisa que se base en genomas completos o casi completos para comprobar los posibles efectos de la recombinación. Esto también es importante cuando se considera el potencial uso de HPgV en estrategias de vacunación que complementen la terapia contra el VIH (Greenhalgh, Schmidt, & Day, 2019).

El posible papel simbiótico o comensal que presenta HPgV podría aportar algunos beneficios a la salud humana. Apoyando este rol potencial, la

infección por HPgV está asociada a una menor activación inmunitaria (Bhattarai, Rydze, Chivero, & Stapleton, 2012; Maidana-Giret et al., 2009), pero esta reducción en los mecanismos de respuesta inmunitaria también podría explicar la asociación observada entre la infección por HPgV y el linfoma no Hodgkin (Chang et al., 2014; Krajden et al., 2010). Los recientes descubrimientos de aislados estrechamente relacionados con los pegivirus en diferentes especies (Kapoor et al., 2013; Sibley et al., 2014) plantean la posibilidad de implementar un modelo de infección animal que podría ayudar a dilucidar los posibles beneficios de la infección crónica por HPgV.

En resumen, en este capítulo se ha visto que las muestras de sangre de la población general albergan una notable diversidad viral. Hasta hace poco, la patogénesis ha sido el principal objetivo de los estudios virales, pero esta visión tradicional está cambiando debido al aumento de la diversidad viral identificada en individuos sanos obtenida a partir de los análisis metagenómicos. En consecuencia, merece la pena seguir investigando un marco diferente que considere a los virus como inocuos o, lo que es más interesante, como potencialmente beneficiosos, lo que podría conducir al uso de dichos virus en futuras aplicaciones biomédicas.

Capítulo 3: Análisis del aneloma individual de 50 donantes

En este tercer capítulo, el objetivo ha consistido en el análisis del aneloma, es decir, el conjunto de anelovirus presentes en cada individuo. En este caso, se ha obviado la fracción viral de ARN debido a la baja prevalencia obtenida en los capítulos anteriores. Por ello, únicamente se ha llevado a cabo la extracción del ADN viral presente en las muestras de plasma individual. Esta aproximación ha permitido identificar un considerable número de genomas de anelovirus y estudiar si existe recombinación entre los aislados descritos.

1. Resultados

1.1. Análisis preliminar de las secuencias

En el estudio presentado en este capítulo han participado 50 donantes de sangre (25 hombres y 25 mujeres) donde para cada sexo se trató de obtener una representación amplia en cuanto a franja de edad, que osciló entre los 20 y los 61 años (**Tabla 6**) y, para cada uno de ellos, se han recogido 24 mL de sangre completa. En primer lugar, y de manera análoga a los capítulos anteriores, los 24 mL de sangre completa se han sometido a una centrifugación para eliminar la fracción celular de la sangre. Al eliminar esta fracción, se han obtenido aproximadamente 10 mL de plasma por donante, los cuales se han procesado, en términos generales, como se describió en el capítulo 1, salvo por alguna modificación que se comentará cuando corresponda. Brevemente, estos 10 mL se han filtrado con filtros de 1 μm y se han almacenado a $-80\text{ }^{\circ}\text{C}$ hasta el procesado. Los plasmas se han centrifugado a alta velocidad para sedimentar los potenciales virus presentes en la muestra. Los sedimentos se han tratado con nucleasas para la digestión de los ácidos nucleicos libres y, tras esto, se ha realizado la extracción del ADN viral presente en las muestras. En este capítulo, puesto que sólo nos interesaba el análisis de los virus de ADN, se ha omitido la extracción de ARN y las amplificaciones asociadas. Se ha

llevado a cabo una amplificación aleatoria con el procedimiento ya descrito basado en la polimerasa del fago $\phi 29$ (Shoaib et al., 2008) y, a partir de las muestras amplificadas, se ha realizado la librería para la secuenciación. Como control de la eficiencia de recuperación del ADN viral, se ha añadido a todas las muestras 50 PFUs del bacteriófago $\phi X174$ y, además, se han procesado dos controles negativos en paralelo, que consistían en 10 mL de PBS 1X estéril.

Una vez obtenidos los datos de secuenciación, se ha realizado una clasificación taxonómica de las lecturas mediante Centrifuge (Kim et al., 2016) y a continuación se ha utilizado Recentrifuge (Martí, 2019) para eliminar las lecturas consideradas como potenciales contaminantes (es decir, aquellas presentes en los controles negativos). Respecto a $\phi X174$, el virus empleado como control, se han recuperado lecturas de este virus en únicamente 30 de las 50 muestras (**Tabla 6**). Esto no es sorprendente, ya que se añadió una cantidad muy pequeña del virus (es decir, 50 PFUs/muestra), dado que en los anteriores capítulos se había observado que, con cantidades superiores, una fracción importante de las lecturas de secuenciación correspondían a $\phi X174$, circunstancia que pretendíamos evitar en el presente capítulo. Por otra parte, en 43 muestras se han encontrado lecturas clasificadas dentro de la familia *Anelloviridae*, presentando de 11.563 a 6.255.446 lecturas (**Tabla 6**). El hecho de que cinco de las muestras no mostraran lecturas de anelovirus ni de $\phi X174$, podría hacer sospechar algún fallo en el proceso de enriquecimiento, si bien esta asociación no era estadísticamente significativa (Test exacto de Fisher, $P=0,221$). También se observa que en algunas muestras se han recuperado lecturas de $\phi X174$, pero no de anelovirus, y viceversa. En estos casos, no se puede llegar a una conclusión clara, ya que el procedimiento parece haber funcionado. Tampoco se podría descartar que las muestras donde no se han

detectado lecturas de anelovirus fueran, en realidad, libres de infección, si bien esto es algo extraño, debido a su alta prevalencia (Spandole et al., 2015).

A modo de curiosidad, en la clasificación de las lecturas de la muestra **SPI12**, se han recuperado 4.213 lecturas, posteriormente ensambladas en un *contig* con extremos solapantes. Este *contig* se ha considerado como un genoma completo de 5.930 nucleótidos y se ha identificado mediante BLASTn dentro de la familia *Microviridae*. Además, un análisis BLASTp de las cinco ORFs putativas de mayor tamaño, ha mostrado homologías que oscilan entre el 87,9 y el 99,7% con la secuencia descrita más cercana (**Tabla Suplementaria S28**).

Tabla 6: Resumen de los datos sobre las muestras analizadas. Para cada una de las muestras, se indica el sexo, la edad, el número de lecturas clasificadas en Recentrifuge (Martí, 2019) para el virus empleado como control (i.e. ϕ X174) y dentro de la familia *Anelloviridae*, así como el número de contigs con un tamaño mayor de 1,5 kb obtenido tras el ensamblaje con MetaSPAdes (Nurk et al., 2017).

Muestra	Sexo	Edad	# lecturas ϕ X174	# lecturas <i>Anelloviridae</i>	<i>Contigs Anelloviridae</i>
SPI3	M	20	3.543	40.062	1
SPI37	M	24	0	3.522.499	1
SPI49	M	24	774.866	240.001	2
SPI13	M	25	0	0	0
SPI15	M	25	0	2.249.419	2
SPI4	M	26	23.125	2.531.650	5
SPI32	M	27	151.496	5.608.907	16
SPI27	M	28	20.727	3.397.050	4
SPI39	M	29	306.156	3.812.065	2
SPI14	M	30	277.908	2.133.866	2
SPI10	M	31	0	4.044.161	2
SPI44	M	34	85.389	3.368.783	2
SPI42	M	35	0	0	0
SPI26	M	36	79.342	2.717.066	2
SPI29	M	38	73.664	6.152.845	6
SPI5	M	39	265.736	396.282	1

Capítulo 3

SPI6	M	40	0	550.509	2
SPI43	M	41	0	6.255.446	1
SPI36	M	44	130.535	1.080.289	4
SPI40	M	48	0	0	0
SPI41	M	49	22.832	5.800.847	9
SPI24	M	51	2.457	0	0
SPI48	M	57	0	409.988	1
SPI17	M	60	0	3.485.404	54
SPI18	M	60	29.559	2.389.963	4
SPI22	H	22	0	11.563	2
SPI35	H	22	0	6.085.309	14
SPI38	H	24	202.114	4.495.535	5
SPI16	H	25	0	2.140.416	5
SPI23	H	25	7.873	2.330.546	34
SPI50	H	25	32.004	4.748.677	9
SPI30	H	26	108.082	2.292.775	2
SPI12	H	27	127.496	1.821.828	15
SPI8	H	28	39.849	1.855.734	5
SPI28	H	31	10.999	128.699	2
SPI31	H	31	1.664.679	0	0
SPI45	H	31	33.577	5.259.934	7
SPI7	H	37	36.394	5.701.149	10
SPI9	H	37	55.970	2.706.686	29
SPI47	H	37	0	354.970	1
SPI33	H	39	0	0	0
SPI2	H	40	0	3.252.950	4
SPI46	H	41	0	0	0
SPI21	H	44	56.681	345.736	3
SPI34	H	48	0	1.188.890	6
SPI11	H	50	3.206	203.385	5
SPI25	H	52	0	689.753	10
SPI1	H	57	0	2.466.654	28
SPI19	H	60	4.126	3.112.408	28
SPI20	H	61	56.585	1.129.257	2

H: Hombre; M: Mujer

1.2. Análisis de anelovirus

Sin tener en cuenta la clasificación taxonómica preliminar, se han generado *contigs* a partir de todas las lecturas para cada una de las 50 muestras, así como para las dos muestras controles. Tras la obtención de los datos del ensamblaje, se han analizado los *contigs* obtenidos mediante BLASTn. Mediante esta clasificación, se han asignado 349 *contigs* dentro de la familia *Anelloviridae* (**Tabla 6, Tabla Suplementaria S4**), 65 de los cuales presentaban extremos solapantes, por lo que podrían ser considerados genomas completos. También se ha observado una correlación positiva entre el número de lecturas de la familia *Anelloviridae* y el número de *contigs* de esta familia obtenidos en cada muestra (Coeficiente de correlación de Spearman $\rho = 0,605$, $P = 0,000$).

Al igual que se hizo en los capítulos anteriores, para realizar el análisis filogenético de las muestras clasificadas como anelovirus, se ha empleado la secuencia nucleotídica completa de la ORF1. Algunos de los *contigs*, en concreto un 6% de los obtenidos, presentaban ORF1 parciales. Sin embargo, estas ORF1 parciales daban lugar, al menos, a un tamaño superior a 500 aminoácidos, por lo que se ha decidido su inclusión en el análisis filogenético. En primer lugar, se ha construido un árbol filogenético con el método de máxima verosimilitud, donde se incluían las secuencias descritas en este capítulo, junto con las especies de referencia según ICTV y las especies descritas en el capítulo 2 de este trabajo (**Tabla Suplementaria S7, Figura Suplementaria S6**). Esta clasificación preliminar ha permitido asignar los *contigs* descritos en este capítulo a los tres géneros que infectan a humanos de la familia *Anelloviridae*: 197 secuencias pertenecen al género TTV, 88 al género TTMV y 64 al género TTMDV. En cuanto a las secuencias que se podrían considerar genomas completos, tres pertenecen al género TTV, cinco al género

TTMDV y las restantes al género TTMV. Estos resultados concuerdan con los obtenidos en los capítulos 1 y 2. En segundo lugar, una vez clasificadas las secuencias dentro de cada género, para facilitar su caracterización, se ha procedido a la construcción de árboles filogenéticos por género, basados en el método de máxima verosimilitud y, paralelamente, se han construido matrices de identidades. En ambos casos, se han empleado los *contigs* identificados en este capítulo, así como las secuencias de referencia para cada género (es decir, las descritas como especies de referencia por ICTV junto a las potenciales nuevas especies descritas en el capítulo anterior).

Para el género TTV, tanto el árbol como la matriz se han construido incluyendo las 197 secuencias recuperadas en el análisis preliminar de los datos, junto con las 38 especies de referencia. La matriz de distancias obtenida, junto con el árbol filogenético, ha permitido identificar una potencial nueva especie, donde se incluirían tres de las secuencias descritas en este capítulo (**Figura 17, Tabla 7, Tabla Suplementaria S29**), mientras que el resto de las secuencias se han agrupado con el 63,2% (en 24 de 38) de las especies de referencia empleadas. La distribución de las secuencias dentro de cada especie de referencia es altamente variable: las especies de referencia representadas por los aislados TTV3-HEL32, TTV13-TCHN-A y TTVyon-KC009, se han agrupado con 28, 25 y 25 de las secuencias descritas en este capítulo, respectivamente, mientras que otras cuatro especies de referencia representadas por los aislados TTV9-BM1C-18, SP2_C3, SP16_C12 y SP36_C1 (los tres últimos descritos originalmente en el capítulo anterior), se han agrupado con una única secuencia descrita en este capítulo. El género TTV comprende siete grupos filogenéticos distintos (Hsiao et al., 2016) y las secuencias descritas en este capítulo se han distribuido en todos ellos, excepto en los grupos 6 y 7, los cuales presentan una única especie de referencia. De hecho, se puede

observar una correlación positiva entre el número de especies de referencia incluidas en cada grupo y el número de secuencias descritas en este capítulo, resultado que ya se obtuvo en los capítulos anteriores (Coeficiente de correlación de Spearman $\rho = 0,891$, $P = 0,007$).

Tabla 7: Resumen de la clasificación taxonómica para las secuencias recuperadas pertenecientes a la familia *Anelloviridae*.

	Especies de referencia ¹		Secuencias descritas ²	Nuevas especies ³
	ICTV	Capítulos previos		
TTV	26	12	197	1 (0,5%)
TTMV	38	37	88	13 (14,8%)
TTMDV	15	26	64	8 (12,5%)
Total	79	75	349	22 (6,3%)

¹Número de especies de referencia empleadas en este capítulo: aceptadas por ICTV y potenciales nuevas especies descritas en los capítulos previos.

²Número de secuencias descritas para cada género en el presente capítulo.

³Número de potenciales especies nuevas identificadas en este capítulo. Entre paréntesis, porcentaje de especies nuevas respecto al total de secuencias descritas para cada género.

En el caso del género TTMV, el árbol y la matriz se han construido incluyendo las 88 secuencias recuperadas en el análisis preliminar de los datos, junto con las 75 especies de referencia (**Figura 19, Tabla 7**). Para este género, se han identificado 13 potenciales especies nuevas que incluyen 15 de las secuencias descritas (**Figura 18, Tabla 7, Tabla Suplementaria S30**). El resto de los aislados descritos pertenecientes a este género se han agrupado con el 34,7% (26 de 75) de las especies de referencia consideradas. Algo a destacar es que una de las secuencias identificadas en el capítulo anterior como potencial especie nueva, P8-c16, incluye 11 de las secuencias descritas en este capítulo, lo que implica un 12,5% del total.

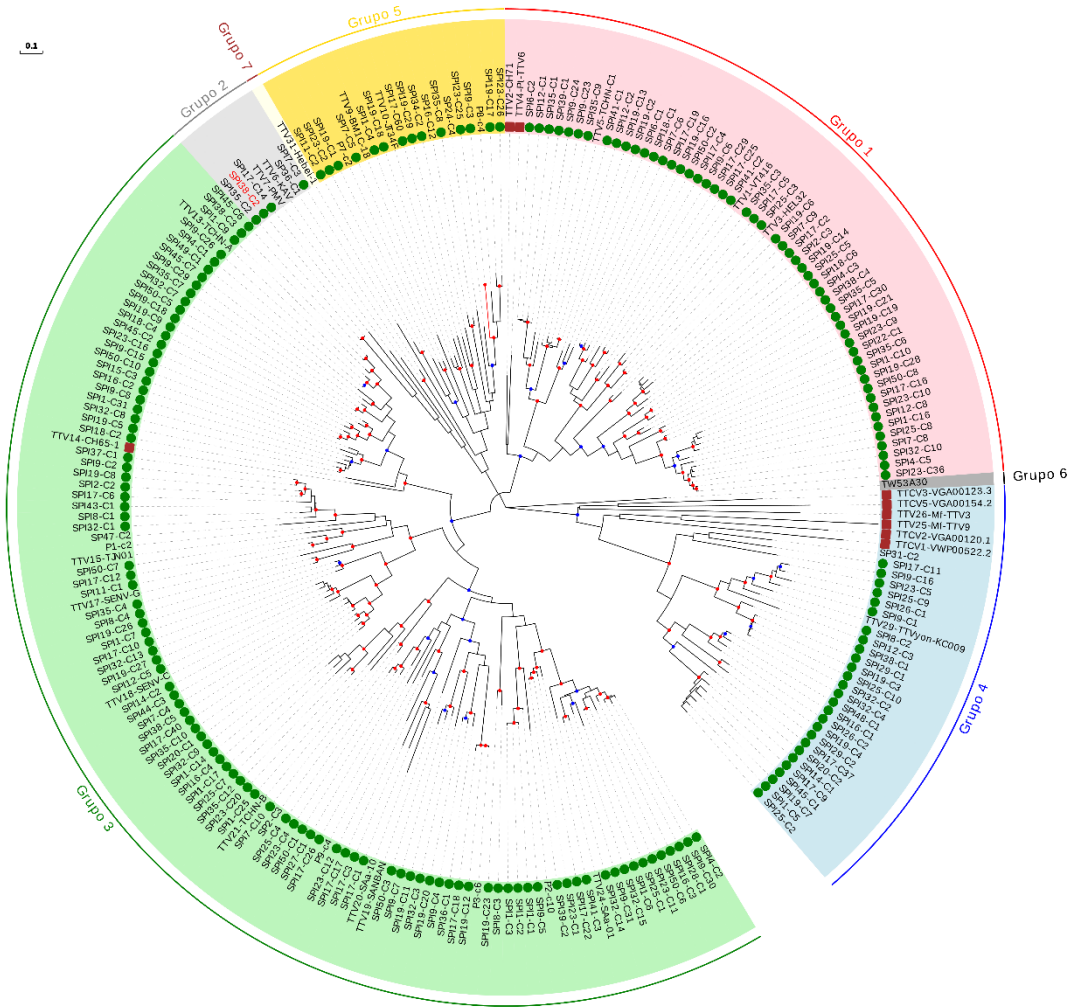


Figura 17: Árbol filogenético de la ORF1 de las especies de referencia y las secuencias descritas pertenecientes al género TTV. Las secuencias descritas en este capítulo se encuentran indicadas con un círculo verde y la secuencia que podría considerarse como potencial nueva especie está marcada en rojo. Los aislados considerados especies de referencia, cuyo hospedador es un primate no homínido, están marcados con un cuadrado marrón. Los valores de *bootstrap* entre 0,7-0,85 están indicados con un círculo azul mientras que aquellos que se encuentran en el rango 0,85-1,0 están indicados con un círculo rojo. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

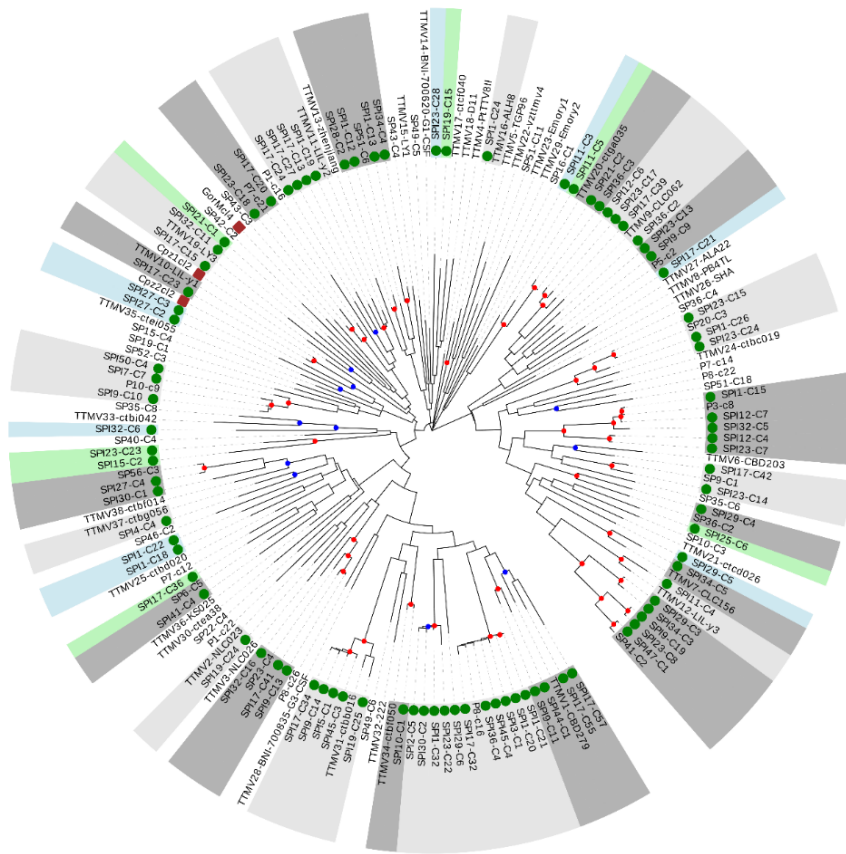


Figura 18: Árbol filogenético para la ORF1 de las especies de referencia y las secuencias descritas pertenecientes al género TTMV. Las secuencias descritas en este capítulo se encuentran indicadas con un círculo verde y aquellas secuencias que podrían considerarse como aislados de potenciales nuevas especies se encuentran sombreadas en verde o en azul, para facilitar la distinción entre grupos contiguos. Las secuencias que se agrupan con alguna especie de referencia están sombreadas con gris claro u oscuro, para distinguir entre grupos contiguos. Los aislados considerados especies de referencia, cuyo hospedador es un primate no homínido, están marcados con un cuadrado marrón. Los valores de *bootstrap* entre 0,7-0,85 están indicados con un círculo azul, mientras que aquellos que se encuentran en el rango 0,85-1,0 están indicados con un círculo rojo. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

Por último, para el género TTMDV se ha construido un árbol filogenético y una matriz de identidad en la que se han añadido las 64 secuencias descritas en este capítulo pertenecientes a este género y las 41 especies de referencia (**Figura 19, Tabla 7, Tabla Suplementaria S31**). Tras el análisis de los datos obtenidos, se han podido identificar ocho potenciales especies nuevas que incluyen 10 secuencias (**Tabla 7**). Por otra parte, el resto de los aislados se han agrupado con un 58,5% (24 de 41) de las especies de referencia.

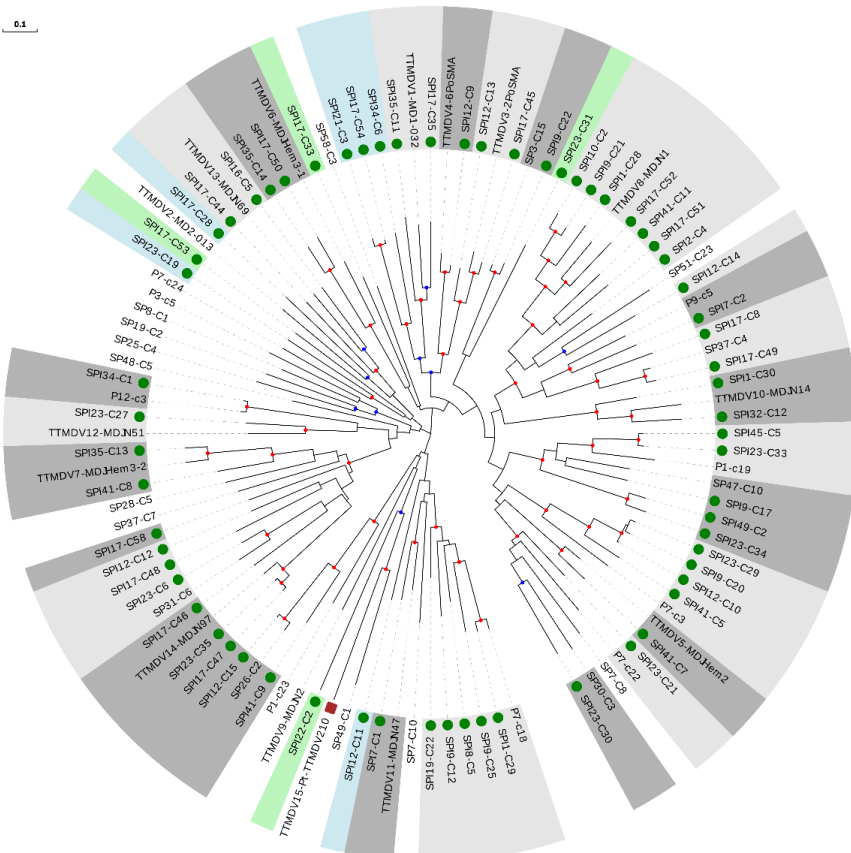


Figura 19: Árbol filogenético para la ORF1 de las especies de referencia y las secuencias descritas pertenecientes al género TTMDV. Las secuencias descritas en este capítulo se encuentran indicadas con un círculo verde y aquellas secuencias que podrían considerarse aislados de potenciales nuevas especies se encuentran

sombreadas en verde o en azul, para facilitar la distinción entre grupos contiguos. Las secuencias que se agrupan con alguna especie de referencia están sombreadas con gris claro u oscuro, para distinguir entre grupos contiguos. Los aislados considerados especies de referencia, cuyo hospedador es un primate no homínido, están marcados con un cuadrado marrón. Los valores de *bootstrap* entre 0,7-0,85 están indicados con un círculo azul, mientras que aquellos que se encuentran en el rango 0,85-1, 0 están indicados con un círculo rojo. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

1.3. Prevalencia de la infección por anelovirus en las 50 muestras de plasma individual

Una vez realizado el análisis filogenético de los aislados descritos, se ha determinado la prevalencia que presenta cada género en las muestras de plasma analizadas en este estudio. De las 50 muestras estudiadas, un 14% (es decir, 7 muestras) no presentan lecturas, y por tanto ningún *contig* perteneciente a la familia *Anelloviridae* (**Tabla 6**). Por el contrario, en las 43 muestras restantes se han recuperado lecturas de anelovirus que se han ensamblado en distintos *contigs* (**Tabla 6**). Tras el análisis filogenético, se ha podido determinar la presencia de cada uno de los géneros en las distintas muestras. De este modo, se puede observar que un 88,4% (38 de 43) de las muestras positivas presenta al menos un aislado perteneciente al género TTV. Además, el 65,1% (28 de 43) y el 44,2% (19 de 43) de las muestras presentan al menos un aislado que pertenece a los géneros TTMV y TTMDV, respectivamente (**Figura 20**). Por otro lado, se puede observar que un 27,9% de las muestras (12 de 43) presentan secuencias de los tres géneros de anelovirus, un 41,9% de las muestras (18 de 43) incluyen dos de los géneros de anelovirus (TTV+TTMV, TTV+TTMDV o TTMV+TTMDV) y un 30,2% (13 de 43) presenta secuencias de un único género. En este último caso, 10 de las muestras presentan sólo secuencias del género TTV, mientras que las tres

restantes sólo presentan secuencias del género TTMV; es decir, no hay ninguna muestra donde se presenten exclusivamente secuencias del género TTMDV.

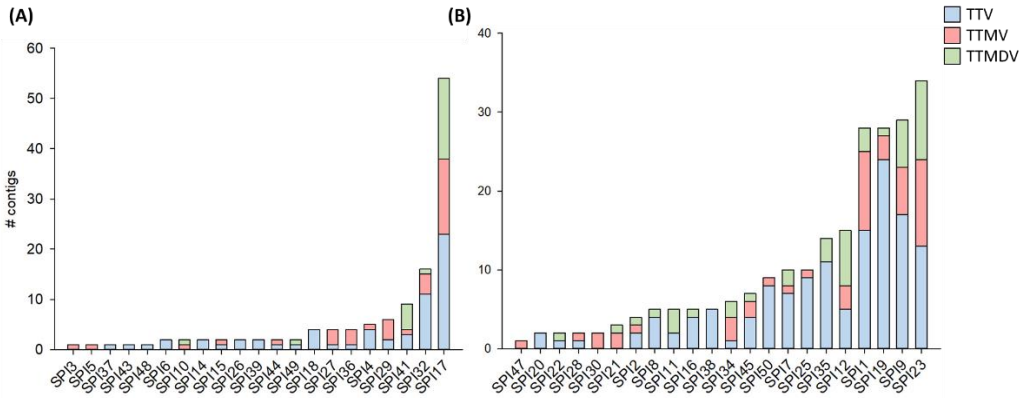


Figura 20: Distribución de los contigs pertenecientes a la familia *Anelloviridae* generados para mujeres (A) y hombres (B). Para cada muestra, se puede observar la proporción de secuencias pertenecientes a TTV, TTMV y TTMDV, mostrados en azul, rojo y verde, respectivamente.

En este capítulo, a diferencia de los dos anteriores, se ha trabajado con muestras individuales de donantes de plasma, lo cual ha permitido analizar si diferentes aspectos demográficos, como el sexo o la edad, influyen sobre la viremia de la familia *Anelloviridae*. En cuanto al factor sexo, se ha comparado el número de *contigs* obtenidos en las 25 muestras de mujeres frente a las 25 de hombres. Se observa que hay diferencias estadísticamente significativas (Test U de Mann-Whitney, $P = 0,016$) entre el número de *contigs* recuperados en mujeres respecto al de hombres (123 vs. 226). En segundo lugar, se ha estudiado si la edad influye sobre el número de *contigs* recuperado, ya que el rango de edad de los donantes que han participado en este muestreo ha sido de 20 a 61 años. Pese a que se ha trabajado con un amplio rango de edades,

no se ha podido identificar una correlación entre el número de *contigs* recuperado y la edad de los donantes (Coeficiente de correlación de Spearman $\rho = 0,045$, $P = 0,756$). Por otra parte, si comparamos ambos factores mediante un modelo de análisis de regresión lineal, empleando el número de *contigs* recuperados como variable dependiente y el sexo y la edad como variables independientes, se puede observar que no existe un efecto de la edad y el sexo sobre el número de *contigs* recuperados (**Tabla 8**).

Tabla 8: Análisis de regresión entre el número de *contigs* recuperados para la familia *Anelloviridae* (variable dependiente) y el sexo y la edad (variables independientes) de los donantes que han participado en el estudio.

Variable independiente	Coeficientes no estandarizados		Coeficientes estandarizados Beta	t	Sig.
	B	Desv. Error			
(Constante)	-7.805	6.477		-1.205	0.234
Edad	0.236	0.125	0.264	1.894	0.064
Sexo	4.015	2.940	0.190	1.366	0.179

Variable dependiente: Número de *contigs* de anelovirus

1.4. Análisis del viroma de donantes individuales

Como se puede observar en la **Tabla 6**, en las muestras se han podido recuperar un número bastante variable de *contigs* y, en cinco de los casos, el número de *contigs* recuperado ha sido mayor de 25. Para estas cinco muestras, **SPI1**, **SPI9**, **SPI17**, **SPI19**, y **SPI23**, se han realizado árboles filogenéticos individuales con el fin de realizar un análisis en mayor profundidad.

En el caso de la muestra **SPI17**, que procede de una mujer donante de 60 años, se han recuperado 54 *contigs* de anelovirus: 24 aislados pertenecen al género TTV, 15 al género TTMV y los 15 restantes al género TTMDV. Como cabría esperar, en esta muestra se recupera el mayor número de potenciales

nuevas especies, con 7 especies: una potencial nueva especie para TTV, 2 para TTMV y 4 para TTMDV (Figura 21).

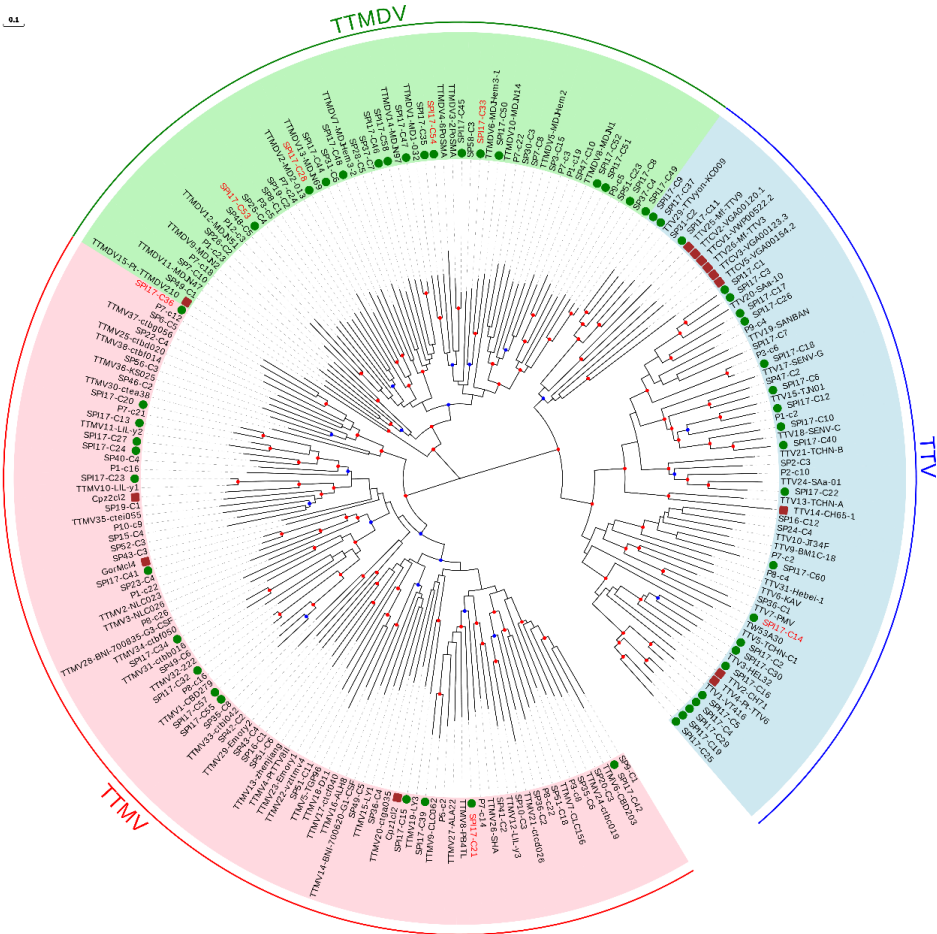


Figura 21: Árbol filogenético de la muestra SPI17. Las secuencias descritas para esta muestra se encuentran marcadas con un círculo verde y aquellas secuencias que podrían considerarse como aislados de potenciales nuevas especies se encuentran en rojo. Los aislados considerados especies de referencia, cuyo hospedador es un primate no homínido, están marcados con un cuadrado marrón. Los valores de *bootstrap* entre 0,7-0,85 están indicados con un círculo azul, mientras que aquellos que se encuentran en el rango 0,85-1,0 están indicados con un círculo rojo. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

En segunda posición en cuanto al número de genomas recuperados, encontramos la muestra **SPI23**. Esta muestra procede de un donante varón de 25 años y se han recuperado 34 *contigs* identificados como miembros de la familia *Anelloviridae*. En este caso, 13 aislados pertenecen al género *TTV*, 11 al género *TTMV* y 10 al género *TTMDV*. En esta muestra, se han podido identificar cuatro potenciales nuevas especies, dos pertenecientes al género *TTMV* y dos al *TTMDV* (**Figura 22**).

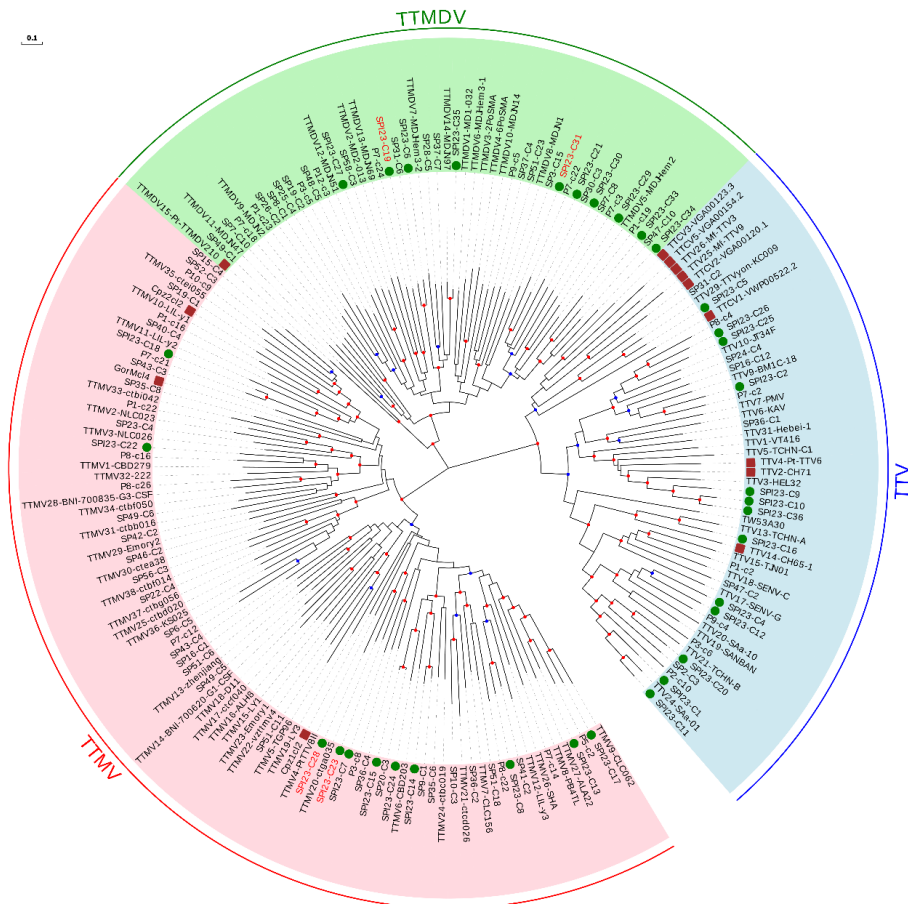


Figura 22: Árbol filogenético de la muestra SPI23. Las secuencias descritas para esta muestra se encuentran marcadas con un círculo verde y aquellas secuencias que podrían considerarse como aislados de potenciales nuevas especies se encuentran en rojo. Los aislados considerados especies de referencia, cuyo hospedador es un primate

no homínido, están marcados con un cuadrado marrón. Los valores de *bootstrap* entre 0,7-0,85 están indicados con un círculo azul, mientras que aquellos que se encuentran en el rango 0,85-1,0 están indicados con un círculo rojo. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

En tercer lugar, se encuentra la muestra **SPI9**. Esta muestra procede de un donante varón de 37 años, en la que se han recuperado 29 *contigs* pertenecientes a la familia *Anelloviridae*. Para esta muestra, la distribución de las secuencias ha sido la siguiente: 17 aislados pertenecen al género TTV, 6 al género TTMV y otras 6 al género TTMDV (**Figura 23**). En esta muestra, pese a que presenta un número elevado de *contigs*, no se ha encontrado ninguna potencial nueva especie.

En el caso de las muestras **SPI1** y **SPI19** se ha recuperado el mismo número de aislados, 28. Por una parte, la muestra SPI1 procede de un donante varón de 57 años de edad. En esta muestra, la distribución de los aislados en los tres géneros estudiados de la familia *Anelloviridae* es la siguiente: 14 genomas pertenecen al género TTV, 11 al género TTMV y tres al género TTMDV. En los análisis realizados, se ha identificado una única secuencia que podría considerarse como potencial nueva especie y pertenece al género TTMV (**Figura 24**).

0.1

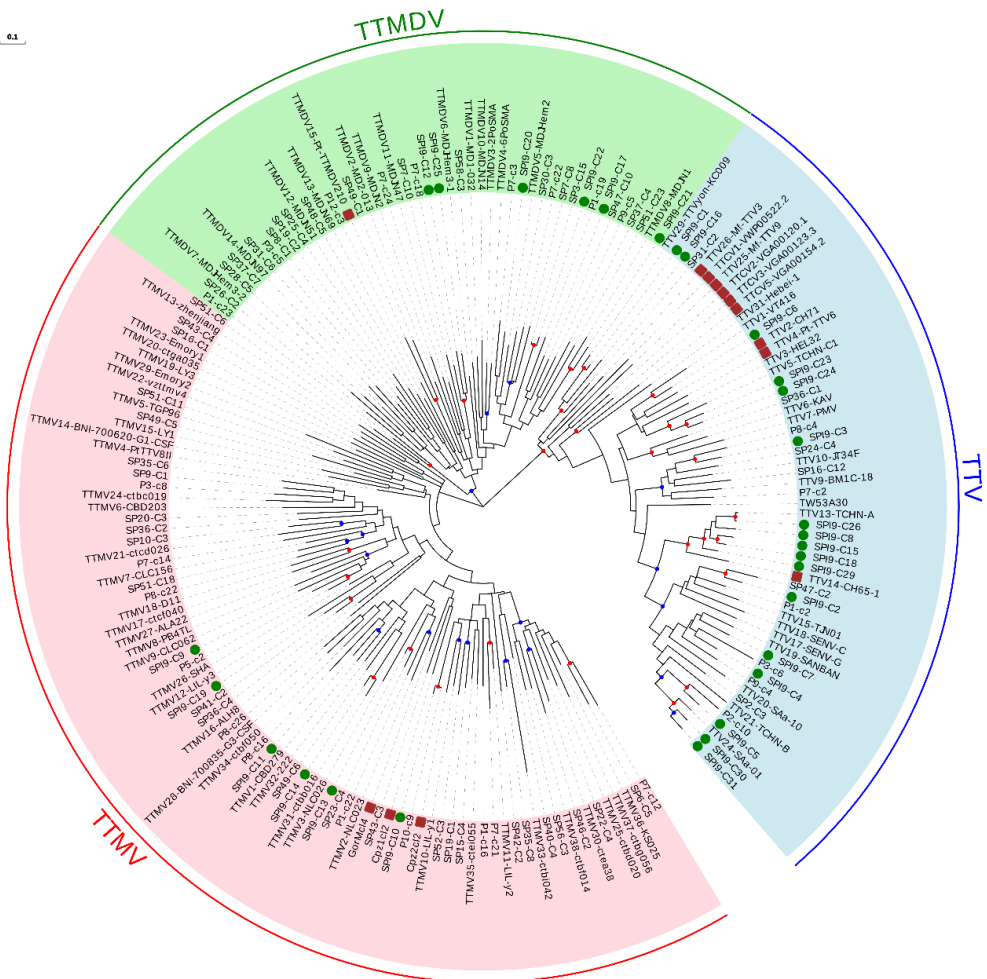


Figura 23: Árbol filogenético de la muestra SPI9. Las secuencias descritas para esta muestra se encuentran marcadas con un círculo. Los aislados considerados especies de referencia, cuyo hospedador es un primate no homínido, están marcados con un cuadrado marrón. Los valores de bootstrap entre 0,7-0,85 están indicados con un círculo azul, mientras que aquellos que se encuentran en el rango 0,85-1,0 están indicados con un círculo rojo. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

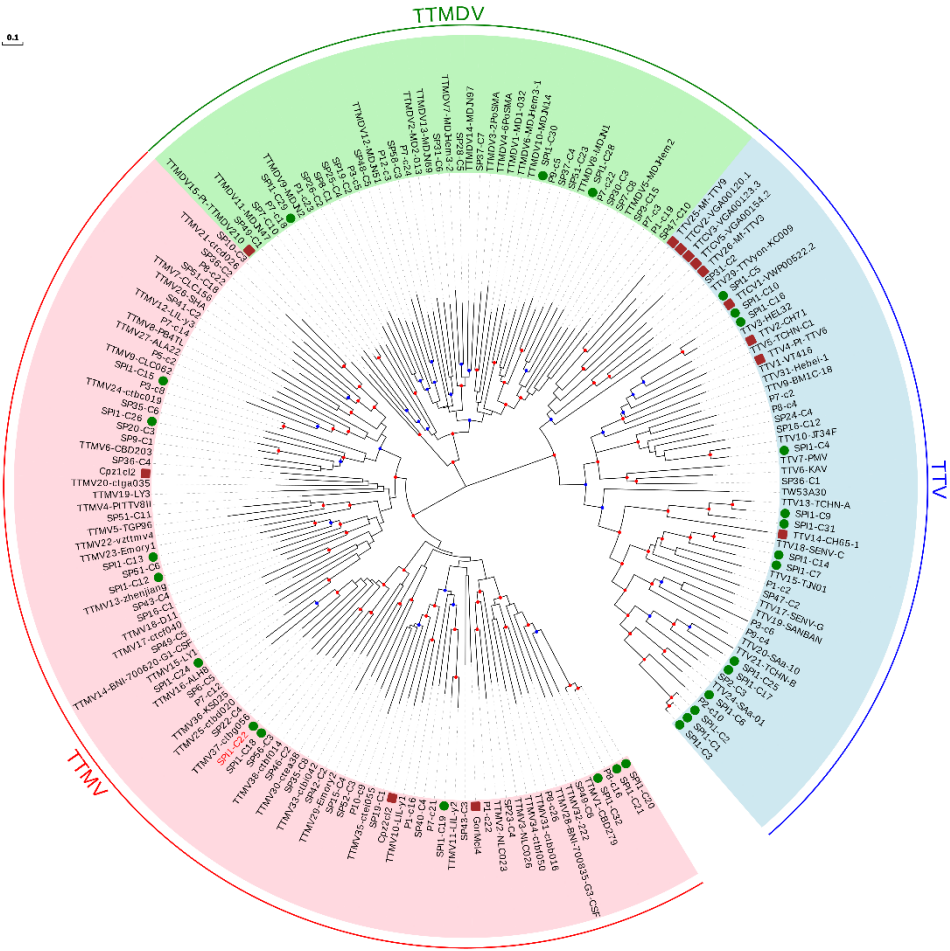


Figura 24: Árbol filogenético de la muestra SPI1. Las secuencias descritas para esta muestra se encuentran marcadas con un círculo verde y la única secuencia que podría considerarse como potencial nueva especie se encuentran en rojo. Los aislados considerados especies de referencia, cuyo hospedador es un primate no homínido, están marcados con un cuadrado marrón. Los valores de *bootstrap* entre 0,7-0,85 están indicados con un círculo azul mientras que aquellos que se encuentran en el rango 0,85-1,0 están indicados con un círculo rojo. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

Por último, en la muestra **SPI19**, procedente de un donante varón de 60 años, se han recuperado 28 *contigs* de anelovirus. En este caso, 24 de ellos pertenecen al género **TTV**, tres de ellos al género **TTMV** y un único aislado pertenece al género **TTMDV**. En esta muestra, se ha identificado dentro del género **TTMV** un aislado que podría considerarse una potencial nueva especie (**Figura 25**).

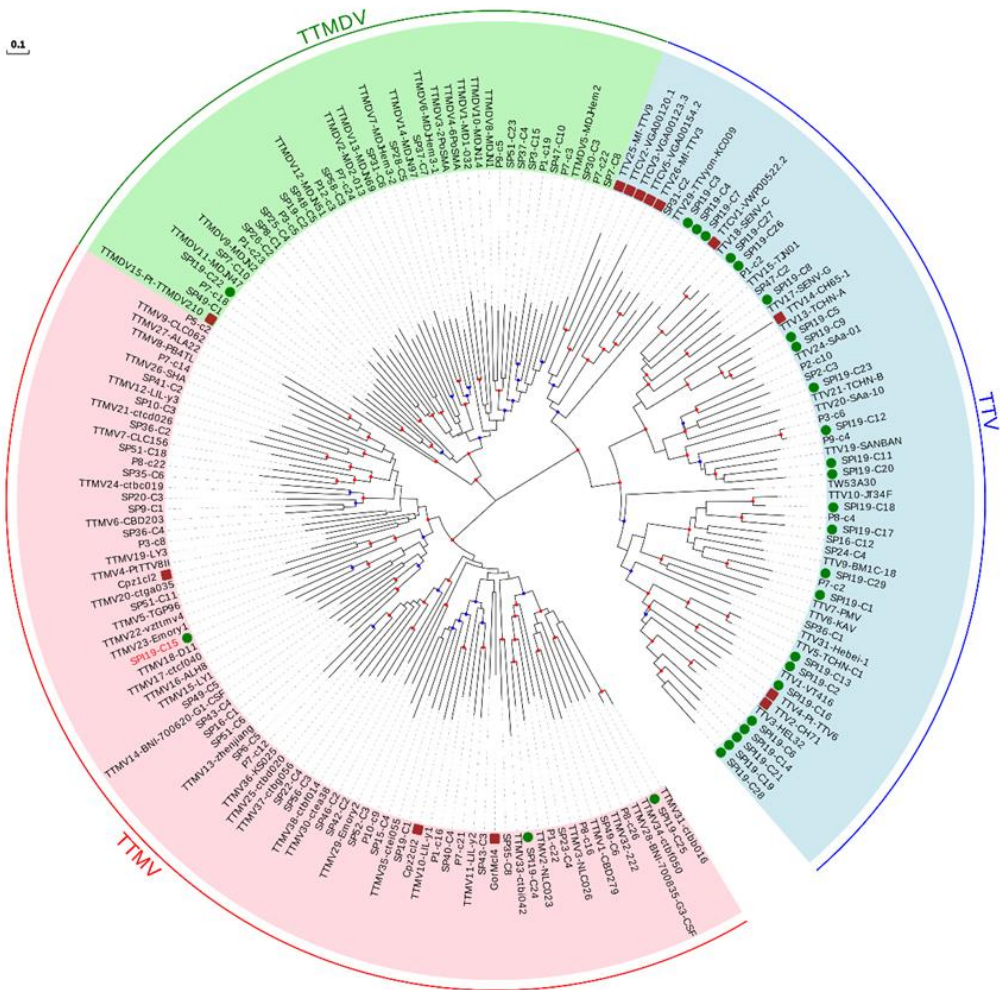


Figura 25: Árbol filogenético de la muestra **SPI19**. Las secuencias descritas para esta muestra se encuentran indicadas con un círculo verde y la única secuencia que podría

considerarse potencial nueva especie se encuentra marcada en rojo. Los aislados considerados especies de referencia, cuyo hospedador es un primate no homínido, están marcados con un cuadrado marrón. Los valores de *bootstrap* entre 0,7-0,85 están indicados con un círculo azul, mientras que aquellos que se encuentran en el rango 0,85-1,0 están indicados con un círculo rojo. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

1.5. Análisis de recombinación de los genomas de anelovirus

Este análisis se ha realizado individualmente para cada donante. Además, dado que se presenta una enorme diversidad entre los tres géneros de anelovirus, se ha realizado de manera independiente para cada género. El requisito para incluir una muestra en el análisis es que se presentaran, al menos, diez secuencias de anelovirus para un género dado dentro de un donante. Este requisito lo cumplen las muestras **SPI1, SPI9, SPI17, SPI19, SPI23, SPI32 y SPI35 (Tabla 9)**, por lo que se han identificado las secuencias de cada género, se han alineado y analizado con el programa RDP4 (Martin et al., 2015).

Tabla 9: Resumen de las muestras seleccionadas para el análisis de recombinación. Los conjuntos de secuencias utilizadas en el análisis de recombinación para cada género, dentro de la muestra, están marcados en negrita.

	Número total secuencias	TTV	TTMV	TTMDV
SPI1	28	14	11	3
SPI9	29	17	6	6
SPI17	54	24	15	15
SPI19	28	24	3	1
SPI23	34	13	11	10
SPI32	16	11	4	1
SPI35	14	11	0	3

En el análisis, se han considerado como sucesos de recombinación putativos aquellos en los que se ha obtenido evidencia estadística en al menos

tres de los métodos implementados en RDP4. Esto ha permitido la detección global de 68 sucesos de recombinación (**Tabla 10** y **Tabla Suplementaria S32**). Además, el análisis con RDP4 también ha permitido identificar los puntos de ruptura del fragmento recombinante (**Tabla Suplementaria S32**).

Tabla 10: Resumen de los sucesos de recombinación obtenidos por género y muestra.

Género	Muestra	Nº sucesos	Género	Muestra	Nº sucesos
TTV	SPI1	4	TTMV	SPI1	3
	SPI9	8		SPI17	4
	SPI17	5		SPI23	4
	SPI19	9	TTMDV	SPI17	7
	SPI23	7		SPI23	8
	SPI32	4			
	SPI35	5			

Tras detectar los sucesos de recombinación putativos, se ha realizado un nuevo análisis donde se ha contrastado la confianza de cada suceso de recombinación de manera independiente. Para ello, se han extraído los alineamientos de los fragmentos recombinantes y, para cada uno de ellos, se han realizado pruebas de congruencia ELW y análisis de mapas de probabilidad. El análisis de los mapas de probabilidad ha proporcionado el valor de topologías de árbol sin resolver (**Tabla Suplementaria S32**). Se han seleccionado aquellos sucesos que presentaban menos del 50% de árboles sin resolver para hacer la prueba de congruencia ELW, es decir, un total de 41 de los 68 sucesos identificados por RDP4. Este valor se ha escogido debido a que, cuanto más bajo sea el valor del número de árboles sin resolver, mayor señal filogenética presenta el suceso de recombinación y, por tanto, la reconstrucción de los árboles es más fiable. Este subconjunto de datos se ha analizado mediante la prueba de congruencia ELW con el fin de determinar si

existen diferencias estadísticamente significativas entre los árboles obtenidos tras el análisis filogenético de los fragmentos resultantes del suceso de recombinación y el árbol obtenido tras el análisis filogenético de la ORF1, que se ha empleado como referencia (**Tabla 11**). Este análisis, que se ha llevado a cabo con IQ-TREE (Nguyen et al., 2015), ha permitido obtener los valores de ELW y de AU. El valor de la prueba ELW representa el peso de la probabilidad de cada árbol, por lo que cuanto más parecido a 1 sea este valor, mayor fiabilidad presenta. Por ello, se consideran como potenciales recombinantes aquellos sucesos en los que el valor de ELW sea mayor a 0,5 (Beamud, Bracho, & González-Candelas, 2019). Por otro lado, la prueba AU proporciona el *P* valor de cada árbol, por lo que, si éste es menor a 0,05, el árbol no presenta suficiente evidencia estadística y, por tanto, se rechaza. Respecto a los árboles obtenidos de los fragmentos recombinantes, 17 de ellos presentan un valor ELW mayor a 0,5 y 21 un valor de AU mayor de 0,05. Por otro lado, en el caso de los árboles de la ORF1, 19 de ellos presentan un valor ELW mayor a 0,5 y 30 un valor de AU mayor a 0,05. Esto hace que, en muchos de los potenciales sucesos de recombinación analizados, no pueda rechazarse el árbol de la ORF1 y, por tanto, afirmarse que la topología del árbol de la ORF1 es incongruente con la topología de la zona recombinante. Siguiendo este criterio, únicamente seis sucesos (i.e. SPI1-TTMV-1, SPI9-TTV-2, SPI17-TTV-1, SPI17-TTV-4, SPI17-TTMDV-1, SPI19-TTV-1) (**Figura 26**) de los 41 analizados cumplían simultáneamente los criterios descritos para los árboles obtenidos del fragmento recombinante (i.e. $ELW > 0,5$ y $AU > 0,05$) y la ORF1 (i.e. $ELW < 0,5$ y $AU < 0,05$). En estos seis casos, al rechazarse estadísticamente el árbol de la ORF1, se puede afirmar que la topología del árbol de la ORF1 es incongruente con la topología del árbol de la zona recombinante. De estos seis sucesos, cuatro se han detectado en TTV y uno en TTMV y TTMDV. Además, si nos

fijamos a nivel de muestra, cabe destacar que tres de los sucesos se han detectado en la muestra **SPI17**, dos de ellos en TTV y uno en TTMDV.

Tabla 11: Resumen de los valores ELW y AU obtenidos al realizar la evaluación de la topología de los árboles filogenéticos obtenidos para los fragmentos considerados como potenciales recombinantes.

Suceso de recombinación	Árbol del fragmento recombinante		Árbol de la ORF1	
	ELW	AU	ELW	AU
SPI1-TTV-3	0,77	0,759	0,3	0,241
SPI1-TTMV-1	0,993	0,995	0,00746	0,00484
SPI1-TTMV-2	0,995	0,994	0,00492	0,0559
SPI1-TTMV-3	NA	NA	NA	NA
SPI9-TTV-2	0,896	0,952	0,104	0,0485
SPI9-TTV-3	0,541	0,51	0,459	0,49
SPI9-TTV-6	0,0981	0,129	0,902	0,871
SPI9-TTV-7	3,21E-16	0,00194	1	0,998
SPI9-TTV-8	0,536	0,558	0,464	0,442
SPI17-TTV-1	0,999	0,998	0,000608	0,00234
SPI17-TTV-4	0,992	0,992	0,00815	0,00832
SPI17-TTV-5	NA	NA	NA	NA
SPI17-TTMV-2	0,932	0,94	0,0685	0,0596
SPI17-TTMDV-1	0,972	0,974	0,0281	0,0256
SPI17-TTMDV-2	0,0162	0,0181	0,984	0,982
SPI17-TTMDV-3	NA	NA	NA	NA
SPI17-TTMDV-4	5,72E-06	0,000611	1	0,999
SPI17-TTMDV-5	0,0301	0,0269	0,97	0,973
SPI17-TTMDV-6	0,000962	0,0046	0,999	0,995
SPI17-TTMDV-7	5,62E-16	0,00303	1	0,997
SPI19-TTV-1	0,996	0,999	0,00449	0,000798
SPI19-TTV-2	2,16E-05	6,91E-12	1	1
SPI19-TTV-4	0,0595	0,06	0,94	0,939
SPI19-TTV-6	0,00114	0,0118	0,999	0,988
SPI23-TTV-1	0,818	0,814	0,182	0,186
SPI23-TTV-2	0,329	0,367	0,671	0,633
SPI23-TTV-3	0,000478	0,000283	1	1

SPI23-TTV-4	0,0047	0,00341	0,995	0,997
SPI23-TTV-6	0,0324	0,0267	0,968	0,973
SPI23-TTMV-2	NA	NA	NA	NA
SPI23-TTMV-3	NA	NA	NA	NA
SPI23-TTMDV-1	0,94	0,937	0,0602	0,0629
SPI23-TTMDV-3	0,0166	0,022	0,983	0,978
SPI23-TTMDV-5	1,01E-06	0,000133	1	1
SPI23-TTMDV-6	0,124	0,112	0,876	0,888
SPI32-TTV-1	0,72	0,777	0,28	0,223
SPI32-TTV-2	0,0716	0,0307	0,928	0,969
SPI32-TTV-4	0,00878	0,00352	0,991	0,996
SPI35-TTV-2	0,728	0,785	0,272	0,215
SPI35-TTV-3	0,711	0,735	0,289	0,265
SPI35-TTV-4	0,735	0,761	0,265	0,239

NA: No se ha podido llevar a cabo la evaluación del árbol debido a que uno de los aislados presentaba un hueco en el alineamiento del fragmento.

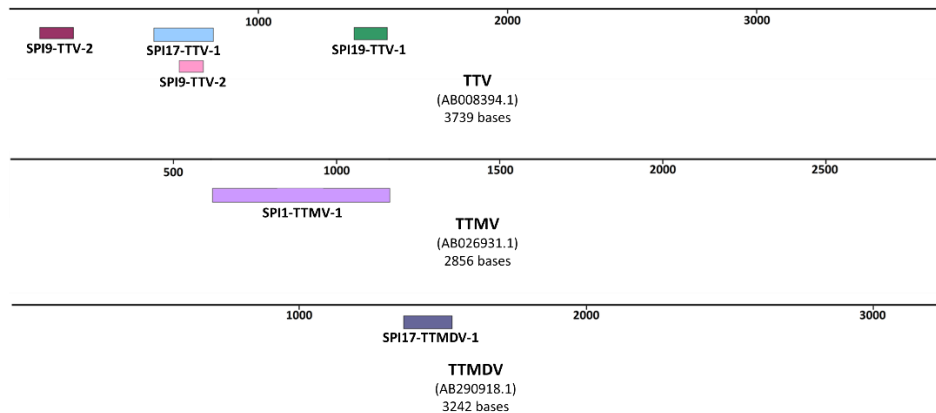


Figura 26: Representación esquemática de los sucesos de recombinación detectados.

En esta figura se muestra la localización de los seis sucesos de recombinación sobre el genoma de referencia de cada género, indicado entre paréntesis.

2. Discusión

Los estudios de metagenómica viral en muestras procedentes de humanos han permitido identificar diversas familias virales. En concreto, una de las familias que presenta una mayor representación en el viroma humano es la familia *Anelloviridae*. Los miembros de esta familia presentan una enorme diversidad, que se ha caracterizado en otros estudios (Moustafa et al., 2017) y que es llamativa en comparación con otras familias virales que presentan genomas de simple cadena de ADN (Kaczorowska & Van Der Hoek, 2020). A partir de las muestras analizadas en este capítulo, se han descrito 349 secuencias distintas de anellovirus, lo que aumenta la diversidad descrita hasta la fecha. Para realizar los análisis filogenéticos, se han empleado las secuencias de referencia descritas por la ICTV, junto con aquellas secuencias consideradas como potenciales especies nuevas en los capítulos anteriores, lo que supone un total de 157 secuencias de referencia (**Tabla 7**). En este capítulo, se han descrito 22 potenciales especies nuevas que incluyen a 28 secuencias distintas, un dato que resulta llamativo, teniendo en cuenta el número creciente de secuencias de referencia empleadas. En concreto, se ha descrito una potencial nueva especie para TTV, 13 para TTMV y 8 para TTMDV. Esto supone un aumento del 2,6%, 16,7% y 19,5% en el número de especies descritas para TTV, TTMV y TTMDV, respectivamente. En el caso de TTV no ha habido un gran aumento, pero para TTMV y TTMDV sí, lo que sugiere que la diversidad de TTV se encuentra explorada en su mayoría mientras que la diversidad de los géneros TTMV y TTMDV se encuentra en expansión, resultado que concuerda con los obtenidos en otros estudios (Arze et al., 2021).

En este capítulo, se ha realizado el análisis del viroma a partir de muestras individuales, lo que ha permitido estudiar si la composición del

viroma se ve afectada por distintos parámetros demográficos, como la edad o el sexo. En cuanto al sexo, en este estudio se han incluido hombres y mujeres en proporciones iguales. En el caso de los hombres, se ha recuperado un número significativamente mayor de secuencias de anelovirus que en las mujeres, lo cual es concordante con los resultados obtenidos en otros estudios (Focosi et al., 2020; Haloschan et al., 2014; Moustafa et al., 2017). Existen diferentes explicaciones a esta observación, como que las mujeres presentan una tendencia a tener respuestas inmunes más altas que los hombres (Moustafa et al., 2017) o bien, que existen diferencias en cuanto a los distintos efectos que producen las hormonas sexuales esteroideas sobre el sistema inmunitario del huésped (Haloschan et al., 2014).

Por otro lado, las muestras de plasma analizadas proceden de individuos con una amplia franja de edad, de 20 a 61 años, por lo que se ha evaluado si la edad implica algún efecto sobre la prevalencia de anelovirus. En las muestras empleadas en este capítulo, no se han encontrado diferencias estadísticamente significativas que determinen que la edad sea un factor influyente sobre la viremia por anelovirus. Esto puede ser debido a que la distribución de muestras por edad en el estudio no es homogénea, ya que el 64% de los sujetos que han participado en el estudio tiene menos de 40 años, lo que podría suponer un sesgo. Además, este resultado no coincide con los obtenidos en otros estudios en los que se observa un aumento lineal de la prevalencia de anelovirus conforme aumenta la edad (Focosi et al., 2020). Este aumento de la prevalencia puede deberse a una disminución en la respuesta inmunitaria, debido a un fenómeno de inmunosenescencia, como se ha descrito en estudios anteriores (Haloschan et al., 2014).

En diferentes trabajos, en los que se ha detectado la presencia de anelovirus mediante ensayos de PCR, se ha visto que la prevalencia de estos

virus varía ampliamente dependiendo de la región geográfica estudiada. Por ejemplo, en población rumana, la prevalencia es del 58,4% (Spandole-Dinu et al., 2018), mientras que en población qatarí aumenta hasta el 85% (Al-Qahtani et al., 2016). Una de las causas que pueden explicar esta variación es la inexistencia de un método estándar para la cuantificación de anelovirus, lo que supone que las regiones diana sean diferentes dependiendo del estudio (Al-Qahtani et al., 2016). En este capítulo, se han analizado 50 muestras de plasma individuales, 43 de las cuales han sido positivas para anelovirus, lo que se traduce en una prevalencia del 86%. Si se analiza la prevalencia por género, ésta ha sido del 88,4%, 65,1% y 44,2% para los géneros TTV, TTMV y TTMDV, respectivamente. Los datos obtenidos para TTV y TTMV concuerdan con los datos procedentes del análisis del viroma de muestras de plasma procedentes de donantes sanos qataríes, pero las prevalencias obtenidas para TTMDV en los donantes qataríes son superiores a las obtenidas en este capítulo (Al-Qahtani et al., 2016). Esta diferencia puede deberse a que el tamaño de la muestra utilizado en nuestro trabajo no es particularmente elevado, lo que podría producir un cierto sesgo en los resultados. Por otro lado, se ha visto que en la mayoría de las muestras positivas para anelovirus se han recuperado secuencias pertenecientes a dos e incluso tres de los géneros. La combinación de secuencias de la familia *Anelloviridae* encontradas en cada individuo, su “aneloma”, es personal y distintivo para cada sujeto, algo que concuerda con los resultados obtenidos en otros estudios (Arze et al., 2021).

Como se ha mencionado anteriormente, la familia *Anelloviridae* se caracteriza por tener una enorme diversidad. Esto es probablemente consecuencia de millones de años de evolución en sus hospedadores, lo que ha propiciado la diversidad actual observada (Kaczorowska & Van Der Hoek, 2020). Esta característica, junto con la posibilidad de que estos virus presenten

una rápida evolución, puede ser la combinación que dé lugar a la gran diversidad presente en esta familia. Esta rápida evolución puede producirse mediante dos mecanismos: una alta tasa de mutación o una alta tasa de recombinación. El genoma de los miembros de la familia *Anelloviridae* no codifica para su propia polimerasa, por lo que la replicación del genoma depende de la maquinaria de la célula huésped (Kakkola et al., 2008). Sin embargo, en la ORF1 de los miembros de esta familia se pueden encontrar motivos conservados que son parecidos a los motivos de las proteínas asociadas a la replicación mediante el mecanismo del círculo rodante, las proteínas Rep, pertenecientes a la familia *Circoviridae* o *Geminiviridae* (Kaczorowska & Van Der Hoek, 2020). En estas familias, la proteína Rep interactúa con las ADN polimerasas celulares afectando a su fidelidad de copia (Maio et al., 2020). Esto podría ser una de las posibles explicaciones de la enorme variabilidad presente en la familia *Anelloviridae*, pero la falta de estudios *in vitro* debido a que los anelovirus no presentan un sistema cultivo robusto causa que haya una falta de estudios del perfil transcripcional de los anelovirus (Kaczorowska & Van Der Hoek, 2020; Kyathanahalli, Snedden, & Hirsch, 2021). Otra posibilidad es que esta variabilidad pueda estar influenciada por el efecto de la recombinación, como se ha descrito anteriormente (Arze et al., 2021; Fahsbender et al., 2017). La principal condición para que ocurra un suceso de recombinación es que se encuentren presentes al menos dos genomas virales coinfectando la misma célula huésped (Pérez-Losada, Arenas, Galán, Palero, & González-Candelas, 2015). En este sentido, se observa que un 86% de las muestras analizadas presentaba más de una secuencia de anelovirus. En concreto, en siete de las muestras (i.e. **SPI1**, **SPI9**, **SPI17**, **SPI19**, **SPI23**, **SPI32** y **SPI35**) se han recuperado más de 10 secuencias distintas en alguno de los géneros. Como se ha descrito

anteriormente, estas muestras han sido escogidas para realizar un análisis de recombinación. La enorme variabilidad genética entre los tres géneros de anelovirus humanos, podría producir artefactos en el alineamiento necesario para llevar a cabo el análisis de recombinación. Por este motivo, el análisis se ha realizado entre las secuencias pertenecientes al mismo género viral dentro de cada donante. Los análisis de recombinación han permitido identificar seis sucesos, de entre los 68 inicialmente detectados mediante RDP4. En estos seis casos, se ha observado evidencia estadística para determinar que el árbol de la ORF1 y el árbol del fragmento recombinante presentan incongruencias, lo cual corrobora la presencia de un potencial suceso de recombinación. El tamaño de estos seis fragmentos es variable, desde 96 a 542 nt, por lo que no parece que éste sea un factor que influya sobre la construcción de árboles y su posterior evaluación. Por otra parte, estos seis sucesos se encuentran en distintas regiones de la secuencia de la ORF1. Esto sugiere que los sucesos de recombinación podrían ocurrir a lo largo de todo el genoma. Además, sería interesante conocer el mecanismo a través del cual se produce la recombinación. Esto se podría llevar a cabo mediante evolución experimental, donde más de un linaje de anelovirus coinfectara una célula hospedadora, pero actualmente no se dispone de un sistema adecuado para la propagación experimental de anelovirus humanos.

El objetivo de este capítulo era caracterizar el viroma de ADN de manera individual a partir de muestras de donantes de sangre sanos, por lo que un incremento en el número de muestras estudiadas proporcionaría una descripción más precisa de los virus presentes en el plasma. Una forma alternativa de enfocar y ampliar estudios similares sería la realización de un análisis longitudinal en el que se realizara un muestreo periódico de los sujetos durante un periodo de tiempo determinado, con el fin de analizar la dinámica

Capítulo 3

que presenta la población de anelovirus. Esto permitiría observar las variaciones que pudieran ocurrir en la composición del aneloma e, incluso, la propia evolución de los anelovirus. Por otro lado, otro análisis que podría haberse llevado a cabo es la cuantificación de las cargas virales mediante qPCR, lo cual serviría para precisar la abundancia de los diferentes aislados de anelovirus y determinar el perfil del viroma que presentan los individuos sanos. Finalmente, cabe comentar que en este capítulo se ha realizado exclusivamente el análisis del viroma de ADN. Dado que la prevalencia para virus de ARN es muy baja en donantes sanos, a diferencia de lo que ocurre con anelovirus, abordar también este otro estudio habría probablemente proporcionado muy poca información.

Discusión general

En esta Tesis se ha llevado a cabo la puesta a punto y la implementación de un protocolo de obtención del viroma presente en muestras de plasma procedentes de donaciones de sangre de donantes sanos. El objetivo de este protocolo ha consistido en recuperar la fracción viral presente en las muestras de plasma, sin conocimiento previo de los virus presentes. Por ello, en el protocolo se combina la extracción de ADN y de ARN de las muestras, lo que ha permitido que se hayan podido recuperar virus que presentan características diferentes. En concreto, en esta Tesis se ha analizado el plasma procedente de 757 donantes, donde se han podido recuperar un total de 815 *contigs* correspondientes a genomas virales completos o parciales. De este total, un 97,79% de los *contigs* corresponden a virus de ADN mientras que la fracción restante, el 2,21%, se trata de virus de ARN. Esta diferencia tan grande respecto al número de genomas recuperados puede ser debida a las distintas prevalencias que presentan los virus de ADN y de ARN en la población sana. Los virus de ADN se han encontrado en todas las localizaciones anatómicas del cuerpo humano, estableciendo infecciones persistentes, como en el caso de los miembros de la familia *Herpesviridae*, o bien, sin estar asociados a alguna enfermedad, como los miembros de la familia *Anelloviridae*. Por el contrario, los virus de ARN tienden a causar infecciones agudas, por lo que muchos de los estudios metagenómicos se suelen centrar únicamente en la fracción de virus de ADN (Zárate et al., 2017) y, por tanto, los virus de ARN del viroma se encuentran peor caracterizados. En esta Tesis, se puede apreciar que el protocolo que se ha puesto a punto es capaz de recuperar tanto virus de ADN y de ARN, lo que permite describir el viroma de una manera más completa.

De la fracción mayoritaria de *contigs*, correspondiente a virus de ADN, en los capítulos 2 y 3 se han recuperado dos genomas distintos, ambos pertenecientes a la familia *Microviridae*. En otros trabajos también se ha

descrito la presencia de virus de esta familia en el torrente sanguíneo (Liang & Bushman, 2021; Rascovan et al., 2016; Zárate et al., 2017), algo que es llamativo por tratarse de bacteriófagos. La detección de estos virus puede ser debida a distintas causas: la primera hipótesis que surge es que se ha amplificado un virus contaminante presente en alguno de los reactivos empleados y que, al no detectarse en ninguno de los controles negativos, no se ha podido sustraer con el programa Recentrifuge (Martí, 2019). En segundo lugar, cabría la posibilidad de que se estuviera detectando el profago, es decir, el genoma del bacteriófago integrado en el cromosoma bacteriano, ya que el ADN correspondiente al hospedador bacteriano también se ha detectado en algunos estudios metagenómicos en muestras de sangre (Rascovan et al., 2016). Por último, debido a que los fagos de la familia *Microviridae* presentan hospedadores que se suelen detectar en el intestino o en la piel, podría haber ocurrido una translocación de las partículas virales desde estas localizaciones anatómicas al torrente sanguíneo, un fenómeno que puede ser frecuente y presentar cierto grado de efectos inmunomoduladores (Górski et al., 2006; Rascovan et al., 2016). En cualquier caso, independientemente de su origen, la detección de estas secuencias indica que el protocolo empleado es sensible, ya que se ha recuperado el genoma completo en ambos casos.

Por otro lado, en trabajos como el de Arze y colaboradores, la búsqueda de virus en el plasma ha ido dirigida específicamente a la amplificación de genomas de anelovirus mediante el uso de cebadores degenerados diseñados en regiones conservadas (Arze et al., 2021). Por el contrario, en el protocolo seguido en esta Tesis para recuperar los virus presentes en el plasma no se ha realizado una búsqueda específica de ninguna familia viral. Aun así, la gran mayoría de los genomas virales identificados en este estudio pertenecen a la familia *Anelloviridae*. En concreto, se han descrito un total de 795 aislados que

corresponden a genomas parciales o completos (es decir, con extremos solapantes) de miembros de la familia *Anelloviridae*. Tras los análisis filogenéticos realizados en cada capítulo de la Tesis con el criterio recientemente implementado por la ICTV (Varsani et al., 2021), se han identificado un total de 98 potenciales nuevas especies. Esto supone un aumento del 122% en cuanto al número de especies aceptadas en la familia *Anelloviridae*. Esta enorme diversidad parece ser una característica intrínseca de esta familia viral (Arze et al., 2021; Focosi et al., 2016; Kaczorowska & Van Der Hoek, 2020). Una de las explicaciones más plausibles de esta elevada diversidad es el efecto de la recombinación entre los aislados de anelovirus, que parece ser una importante fuerza impulsora en la evolución de esta familia viral (Arze et al., 2021). De acuerdo con esto, los análisis realizados en el capítulo 3 nos han permitido detectar evidencias de recombinación a nivel de individuo, a pesar de que se analizaba un número relativamente pequeño de secuencias. Por desgracia, nuestros resultados no nos permitían discernir si alguno de esos potenciales sucesos de recombinación habría ocurrido en el propio donante.

Muchos aspectos de la relación entre el hospedador y los anelovirus permanecen desconocidos. Un ejemplo de ello es que no se sabe dónde se replican los anelovirus, aunque se piensa que puede ser en la médula ósea o en el hígado, ya que en estas localizaciones anatómicas se han encontrado anelovirus de doble cadena, la forma intermedia de replicación de los genomas de simple cadena de ADN (Kaczorowska & Van Der Hoek, 2020). Otro ejemplo es que estos virus no presentan un fuerte tropismo hacia ningún tipo celular o tejido, ya que se encuentran en numerosos tejidos y fluidos corporales (Kaczorowska & Van Der Hoek, 2020). Por otro lado, los anelovirus se encuentran presentes en un amplio porcentaje de la población sana, en

ausencia de enfermedad o cualquier sintomatología (Bédarida et al., 2017), por lo que se desconoce qué efectos puede tener esta infección sobre el organismo. A su vez, no existen cultivos celulares en los que se pueda amplificar el virus y, por tanto, investigar la interacción entre los anelovirus y su hospedador (Kaczorowska & Van Der Hoek, 2020). Pese a ello, la infección por anelovirus parece que no está asociada a ninguna condición de enfermedad, por lo que los anelovirus se consideran virus huérfanos. Esta idea se ve reforzada debido a la gran prevalencia que presenta este virus, lo que hace muy probable que todos los humanos estén infectados por algún miembro de los géneros *Alphatorquevirus*, *Betatorquevirus* o *Gammatorquevirus* desde la infancia. Además, cada individuo presenta su propio aneloma, es decir, un conjunto de anelovirus personal y distintivo (Arze et al., 2021).

La segunda familia viral más predominante en el viroma de las muestras de plasma analizadas en esta Tesis ha sido la familia *Flaviviridae*, y en concreto, un miembro del género *Pegivirus*. Los virus pertenecientes a este género también son considerados virus huérfanos, ya que HPgV no está asociado a ninguna enfermedad (Fernández-Ruiz et al., 2021). Esta falta de patogenicidad puede ser debida a que HPgV es un virus humano antiguo que está bien adaptado al ser humano (Marano et al., 2017). Esta adaptación se puede observar en la marcada distribución geográfica de los genotipos de HPgV (Marano et al., 2017; Yu, Wan, Wang, Yang, & Zhang, 2022), que es el reflejo de las antiguas migraciones humanas (Pavesi, 2001; Sharp & Simmonds, 2011). En este trabajo, sólo se ha analizado la fracción de viroma de ARN en los capítulos 1 y 2, en los que, de manera global, se ha obtenido una prevalencia de HPgV del 2,7%. Este valor se encuentra en consonancia con otros trabajos, donde se estima que la prevalencia de HPgV en donantes de sangre es del 3,1%

(Yu et al., 2022). A pesar de que es un virus que se detecta con frecuencia en los donantes de sangre, no está incluido en las pruebas de cribado de la sangre de las donaciones, ya que, como se ha mencionado, no está asociado con ninguna enfermedad. Además, en las personas sanas, la viremia por HPgV es autolimitante y se aclara en un periodo de dos años tras la infección, mientras que, en personas inmunocomprometidas, este aclaramiento puede durar décadas (Chivero & Stapleton, 2015). En algunos casos, se ha descrito una asociación entre la infección persistente por HPgV y el desarrollo de linfoma en individuos inmunocomprometidos. Si bien sólo se trata de una asociación, el desarrollo de la enfermedad podría explicarse como consecuencia de la infección persistente de los linfocitos, que podría causar mutaciones en el ADN de estas células, provocando en último término su transformación y el desarrollo del linfoma (Yu et al., 2022). Por otro lado, HPgV se suele detectar junto con virus que también se transmiten por la sangre, como el VIH o el virus de la Hepatitis C. También se ha observado que la infección persistente por HPgV inhibe la activación inmunitaria excesiva y anormal en pacientes coinfectados por los virus anteriormente citados, lo que, en algunos casos, muestra efectos clínicos beneficiosos como una deceleración en la progresión de la enfermedad causada por el VIH (Alcalde et al., 2010; Blackard et al., 2017; Schwarze-Zander, Blackard, & Rockstroh, 2012; Yu et al., 2022). Sin embargo, esta asociación beneficiosa no se ha podido encontrar en otros estudios (Berzsenyi et al., 2009; Miao et al., 2017). En términos globales, HPgV parece presentar un potencial papel simbiótico sobre la salud humana y esto es debido a que la infección por el virus parece estar asociada a una menor activación del sistema inmunitario (Bhattarai et al., 2012; Maidana-Giret et al., 2009). Actualmente, no se pueden obtener evidencias experimentales en este sentido, dado que no existe ningún modelo de infección animal ni un cultivo

de células *in vitro* en los que se puedan investigar los mecanismos moleculares implicados en este potencial papel simbiótico.

Los estudios sobre el viroma de individuos sanos han identificado la presencia de una amplia variedad de aislados pertenecientes a la familia *Anelloviridae* y, en menor medida, al género *Pegivirus* de la familia *Flaviviridae*, lo cual indica que estos virus parecen ser inocuos para su hospedador. Como se ha podido observar a lo largo de esta Tesis, y como se ha visto en otros trabajos (Arze et al., 2021), el viroma presenta una marcada variabilidad intra e interindividual. Ésta, además, se ve afectada por factores como el sexo o la edad. Esta relación de comensalismo que se da entre el hospedador y los virus que forman parte del viroma puede aprovecharse para el desarrollo de aplicaciones biomédicas, por ejemplo, como biomarcadores del estado inmunitario del hospedador en el caso de los anellovirus (Burra et al., 2008; Focosi et al., 2010) o como vectores virales en el caso de HPgV (Bailey et al., 2015; Yu et al., 2022). En conclusión, se necesita un número mayor de estudios sobre el viroma que ayuden a mejorar la comprensión de su papel en la salud y en la enfermedad.

Conclusiones

Conclusiones

- Los estudios de virómica van a constituir una herramienta fundamental en la caracterización de la biodiversidad viral; de ahí la importancia de implementar protocolos adaptados al análisis de distintas muestras biológicas o ambientales.
- Relacionado con lo anterior, el establecimiento de estrategias de vigilancia, que traten de anticipar la emergencia de nuevos virus, también supone un desafío en Salud Pública, como lo acredita la actual situación pandémica.
- La enorme diversidad viral descrita en esta Tesis, en particular en la familia *Anelloviridae*, pero también a menor escala en los pegivirus humanos, compromete la interpretación de estudios de asociación de genotipos virales específicos con potenciales enfermedades.
- En este sentido, nuestros resultados han evidenciado a la recombinación como fuerza propulsora relevante en la evolución de diversos virus. Por tanto, los estudios de asociación de genotipos virales con patologías concretas deben tener en cuenta este fenómeno, para lo cual se requiere analizar genomas virales completos o casi completos.
- Los dos grupos de virus fundamentalmente detectados en nuestro trabajo entran en la categoría de virus huérfanos, es decir, virus no asociados a enfermedad alguna. Sin embargo, para profundizar en esta consideración, resulta relevante analizar distintos parámetros demográficos de las poblaciones en estudio, como pueden ser la edad y el sexo.
- Como conclusión global:
 - Por un lado, los datos que ofrece la virómica permitirán establecer potenciales asociaciones entre las variantes virales presentes y sintomatologías de etiología desconocida.

Conclusiones

- Por otro lado, esta enorme diversidad viral también ha producido un cambio a nivel conceptual, que lleva a considerar a los virus, no como meros patógenos, sino fundamentalmente como componentes inocuos e incluso beneficiosos en sus hospedadores, en particular, y en los ecosistemas, en general.

Bibliografía

Bibliografía

- Abbas, A. A., Taylor, L. J., Dothard, M. I., Leiby, J. S., Fitzgerald, A. S., Khatib, L. A., Collman, R. G., & Bushman, F. D. (2019). Redondoviridae, a Family of Small, Circular DNA Viruses of the Human Oro-Respiratory Tract Associated with Periodontitis and Critical Illness. *Cell Host and Microbe*, 25(5), 719–729.e4. <https://doi.org/10.1016/j.chom.2019.04.001>
- Abeles, S. R., Robles-Sikisaka, R., Ly, M., Lum, A. G., Salzman, J., Boehm, T. K., & Pride, D. T. (2014). Human oral viruses are personal, persistent and gender-consistent. *The ISME Journal*, 8(9), 1753–1767. <https://doi.org/10.1038/ismej.2014.31>
- Abril, J. F., & Castellano, S. (2019). Genome Annotation. In *Encyclopedia of Bioinformatics and Computational Biology* (pp. 195–209). Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20226-4>
- Aggarwala, V., Liang, G., & Bushman, F. D. (2017). Viral communities of the human gut: metagenomic analysis of composition and dynamics. *Mobile DNA*, 8(1), 12. <https://doi.org/10.1186/s13100-017-0095-y>
- Al-Qahtani, A. A., Alabsi, E. S., AbuOdeh, R., Thalib, L., El Zowalaty, M. E., & Nasrallah, G. K. (2016). Prevalence of anelloviruses (TTV, TTMDV, and TTMV) in healthy blood donors and in patients infected with HBV or HCV in Qatar. *Virology Journal*, 13(1), 1–6. <https://doi.org/10.1186/s12985-016-0664-6>
- Alcalde, R., Nishiya, A., Casseb, J., Inocêncio, L., Fonseca, L. A. M., & Duarte, A. J. S. (2010). Prevalence and distribution of the GBV-C/HGV among HIV-1-infected patients under anti-retroviral therapy. *Virus Research*, 151(2), 148–152. <https://doi.org/10.1016/j.virusres.2010.04.008>
- Anthony, S. J., Epstein, J. H., Murray, K. A., Navarrete-Macias, I., Zambrana-Torrelío, C. M., Solovyov, A., ... Lipkin, W. I. (2013). A strategy to estimate unknown viral diversity in mammals. *MBio*, 4(5). <https://doi.org/10.1128/mBio.00598-13>
- Arze, C. A., Springer, S., Dudas, G., Patel, S., Bhattacharyya, A., Swaminathan, H., ... Yozwiak, N. L. (2021). Global genome analysis reveals a vast and dynamic anellovirus landscape within the human virome. *Cell Host & Microbe*, 29(8), 1305–1315.e6. <https://doi.org/10.1016/j.chom.2021.07.001>
- Asplund, M., Kjartansdóttir, K. R., Mollerup, S., Vinner, L., Fridholm, H., Herrera, J. A. R., ... Hansen, A. J. (2019). Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. *Clinical Microbiology and Infection*, 25(10), 1277–1285. <https://doi.org/10.1016/j.cmi.2019.04.028>
- Ataei, B., Emami Naeini, A., Khorvash, F., Yazdani, M. R., & Javadi, A. A. (2012). Prevalence of transfusion transmitted virus infection in hemodialysis

- patients and injection drug users compared to healthy blood donors in Isfahan, Iran. *Gastroenterology Research and Practice*, 2012, 2–5. <https://doi.org/10.1155/2012/671927>
- Azevedo, D. S., Duarte, J. L. C., Freitas, C. F. G., Soares, K. L., Sousa, M. S., Sousa, E. S. S., & Lucena, R. B. (2021). One Health Perspectives on New Emerging Viral Diseases in African Wild Great Apes. *Pathogens*, 10(10), 1283. <https://doi.org/10.3390/pathogens10101283>
- Bailey, A. L., Lauck, M., Mohns, M., Peterson, E. J., Beheler, K., Brunner, K. G., ... O'Connor, D. H. (2015). Durable sequence stability and bone marrow tropism in a macaque model of human pegivirus infection. *Science Translational Medicine*, 7(305). <https://doi.org/10.1126/scitranslmed.aab3467>
- Beamud, B., Bracho, M. A., & González-Candelas, F. (2019). Characterization of New Recombinant Forms of HIV-1 From the Comunitat Valenciana (Spain) by Phylogenetic Incongruence. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.01006>
- Bédarida, S., Dussol, B., Signoli, M., & Biagini, P. (2017). Analysis of Anelloviridae sequences characterized from serial human and animal biological samples. *Infection, Genetics and Evolution*, 53, 89–93. <https://doi.org/10.1016/j.meegid.2017.05.017>
- Béland, K., Dore-Nguyen, M., Gagné, M. J., Patey, N., Brassard, J., Alvarez, F., & Halac, U. (2014a). Torque teno virus in children who underwent orthotopic liver transplantation: New insights about a common pathogen. *Journal of Infectious Diseases*, 209(2), 247–254. <https://doi.org/10.1093/infdis/jit423>
- Béland, K., Dore-Nguyen, M., Gagné, M. J., Patey, N., Brassard, J., Alvarez, F., & Halac, U. (2014b). Torque teno virus load as a biomarker of immunosuppression? new hopes and insights. *Journal of Infectious Diseases*, 210(4), 668–670. <https://doi.org/10.1093/infdis/jiu210>
- Bergallo, M., Costa, C., Gribaudo, G., Tarallo, S., Baro, S., Ponzi, A. N., & Cavallo, R. (2006). Evaluation of six methods for extraction and purification of viral DNA from urine and serum samples. *NEW MICROBIOLOGICA* (Vol. 29).
- Berthet, N., Reinhardt, A. K., Leclercq, I., Van Ooyen, S., Batéjat, C., Dickinson, P., ... Manuguerra, J. C. (2008). Phi29 polymerase based random amplification of viral RNA as an alternative to random RT-PCR. *BMC Molecular Biology*, 9, 1–7. <https://doi.org/10.1186/1471-2199-9-77>
- Berzsenyi, M. D., Bowden, D. S., Roberts, S. K., & Revill, P. A. (2009). GB virus C genotype 2 predominance in a hepatitis C virus/HIV infected population associated with reduced liver disease. *Journal of Gastroenterology and Hepatology (Australia)*, 24(8), 1407–1410.

- <https://doi.org/10.1111/j.1440-1746.2009.05920.x>
- Bhattarai, N., Rydze, R. T., Chivero, E. T., & Stapleton, J. T. (2012). GB virus C viremia is associated with higher levels of double-negative T cells and lower T-cell activation in HIV-infected individuals receiving antiretroviral therapy. *Journal of Infectious Diseases*, 206(9), 1469–1472. <https://doi.org/10.1093/infdis/jis515>
- Bhattarai, N., & Stapleton, J. T. (2012). GB virus C: The good boy virus? *Trends in Microbiology*, 20(3), 124–130. <https://doi.org/10.1016/j.tim.2012.01.004>
- Blackard, J. T., Ma, G., Polen, C., DuBois, J. C., Gast, J., Radens, C. M., ... Sherman, K. E. (2016). Recombination among GB virus C (GBV-C) isolates in the United States. *Journal of General Virology*, 97(7), 1537–1544. <https://doi.org/10.1099/jgv.0.000477>
- Blackard, J. T., Ma, G., Welge, J. A., Taylor, L. E., Mayer, K. H., Klein, R. S., ... King, C. C. (2017). Cytokine/chemokine expression associated with Human Pegivirus (HPgV) infection in women with HIV. *Journal of Medical Virology*, 89(11), 1904–1911. <https://doi.org/10.1002/jmv.24836>
- Blatter, J. A., Sweet, S. C., Conrad, C., Danziger-Isakov, L. A., Faro, A., Goldfarb, S. B., ... Wang, D. (2018). Anellovirus loads are associated with outcomes in pediatric lung transplantation. *Pediatric Transplantation*, 22(1), e13069. <https://doi.org/10.1111/petr.13069>
- Bonsall, D., Gregory, W. F., Ip, C. L. C., Donfield, S., Iles, J., Ansari, M. A., ... Simmonds, P. (2016). Evaluation of viremia frequencies of a novel human pegivirus by using bioinformatic screening and PCR. *Emerging Infectious Diseases*, 22(4), 671–678. <https://doi.org/10.3201/eid2204.151812>
- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., & Rohwer, F. (2003). Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *Journal of Bacteriology*, 185(20), 6220–6223. <https://doi.org/10.1128/JB.185.20.6220-6223.2003>
- Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O. J., & Ian Lipkina, W. (2015). Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio*, 6(5). <https://doi.org/10.1128/mBio.01491-15>
- Burián, Z., Szabó, H., Székely, G., Gyurkovits, K., Pankovics, P., Farkas, T., & Reuter, G. (2011). Detection and follow-up of torque teno midi virus (“small anelloviruses”) in nasopharyngeal aspirates and three other human body fluids in children. *Archives of Virology*, 156(9), 1537–1541. <https://doi.org/10.1007/s00705-011-1021-0>
- Burra, P., Masier, A., Boldrin, C., Calistri, A., Andreoli, E., Senzolo, M., ... Palù, G. (2008). Torque Teno virus: Any pathological role in liver transplanted

- patients? *Transplant International*, 21(10), 972–979. <https://doi.org/10.1111/j.1432-2277.2008.00714.x>
- Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge – Accurate paired shotgun read merging via overlap. *PLoS ONE*, 12(10), 1–15. <https://doi.org/10.1371/journal.pone.0185056>
- Carlson, C. J. (2020). From PREDICT to prevention, one pandemic later. *The Lancet Microbe*, 1(1), e6–e7. [https://doi.org/10.1016/S2666-5247\(20\)30002-1](https://doi.org/10.1016/S2666-5247(20)30002-1)
- Carroll, D., Daszak, P., Wolfe, N. D., Gao, G. F., Morel, C. M., Morzaria, S., ... Mazet, J. A. K. (2018). The Global Virome Project. *Science*, 359(6378), 872–874. <https://doi.org/10.1126/science.aap7463>
- Castillo, D. J., Rifkin, R. F., Cowan, D. A., & Potgieter, M. (2019). The healthy human blood microbiome: Fact or fiction? *Frontiers in Cellular and Infection Microbiology*, 9, 1–12. <https://doi.org/10.3389/fcimb.2019.00148>
- Chandriani, S., Skewes-Cox, P., Zhong, W., Ganem, D. E., Divers, T. J., Van Blaricum, A. J., ... Kistler, A. L. (2013). Identification of a previously undescribed divergent virus from the Flaviviridae family in an outbreak of equine serum hepatitis. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15). <https://doi.org/10.1073/pnas.1219217110>
- Chang, C. M., Stapleton, J. T., Klinzman, D., McLinden, J. H., Purdue, M. P., Katki, H. A., & Engels, E. A. (2014). GBV-C Infection and Risk of NHL among U.S. Adults. *Cancer Research*, 74(19), 5553–5560. <https://doi.org/10.1158/0008-5472.CAN-14-0209>
- Charest, A. J., Plummer, J. D., Long, S. C., Carducci, A., Verani, M., & Sidhu, J. P. S. (2015). Global occurrence of Torque teno virus in water systems. *Journal of Water and Health*, 13(3), 777–789. <https://doi.org/10.2166/wh.2015.254>
- Chehoud, C., Dryga, A., Hwang, Y., Nagy-Szakal, D., Hollister, E. B., Luna, R. A., ... Bushman, F. D. (2016). Transfer of Viral Communities between Human Individuals during Fecal Microbiota Transplantation. *MBio*, 7(2). <https://doi.org/10.1128/mBio.00322-16>
- Cheval, J., Sauvage, V., Frangeul, L., Dacheux, L., Guigon, G., Dumey, N., ... Eloit, M. (2011). Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *Journal of Clinical Microbiology*, 49(9), 3268–3275. <https://doi.org/10.1128/JCM.00850-11>
- Chivero, E. T., & Stapleton, J. T. (2015). Tropism of human pegivirus (formerly known as GB virus C/hepatitis G virus) and host immunomodulation: Insights into a highly successful viral infection, 96(7), 1521–1532.

- <https://doi.org/10.1099/vir.0.000086.Tropism>
- Choi, Y. K. (2021). Emerging and re-emerging fatal viral diseases. *Experimental and Molecular Medicine*, 53(5), 711–712. <https://doi.org/10.1038/s12276-021-00608-9>
- Cibulski, S. P., Teixeira, T. F., Lima, F. E. de S., do Santos, H. F., Franco, A. C., & Roehe, P. M. (2014). A novel Anelloviridae species detected in Tadarida brasiliensis bats: First sequence of a chiropteran Anellovirus. *Genome Announcements*, 2(5), 2509. <https://doi.org/10.1128/genomeA.01028-14>
- Conceição-Neto, N., Zeller, M., Lefrère, H., De Bruyn, P., Beller, L., Deboutte, W., ... Matthijssens, J. (2015). Modular approach to customise sample preparation procedures for viral metagenomics: A reproducible protocol for virome analysis. *Scientific Reports*, 5. <https://doi.org/10.1038/srep16532>
- Cotten, M., Oude Munnink, B., Canuti, M., Deijs, M., Watson, S. J., Kellam, P., & Van Der Hoek, L. (2014). Full genome virus detection in fecal samples using sensitive nucleic acid preparation, deep sequencing, and a novel iterative sequence classification algorithm. *PLoS ONE*, 9(4). <https://doi.org/10.1371/journal.pone.0093269>
- Cotten, M., Watson, S. J., Kellam, P., Al-Rabeeah, A. A., Makhdoom, H. Q., Assiri, A., ... Memish, Z. A. (2013). Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study. *The Lancet*, 382(9909), 1993–2002. [https://doi.org/10.1016/S0140-6736\(13\)61887-5](https://doi.org/10.1016/S0140-6736(13)61887-5)
- Danovaro, R., Corinaldesi, C., Dell'Anno, A., Fuhrman, J. A., Middelburg, J. J., Noble, R. T., & Suttle, C. A. (2011). Marine viruses and global climate change. *FEMS Microbiology Reviews*, 35(6), 993–1034. <https://doi.org/10.1111/j.1574-6976.2010.00258.x>
- de Souza, W. M., Fumagalli, M. J., de Araujo, J., Sabino-Santos, G., Maia, F. G. M., Romeiro, M. F., ... Figueiredo, L. T. M. (2018). Discovery of novel anelloviruses in small mammals expands the host range and diversity of the Anelloviridae. *Virology*, 514, 9–17. <https://doi.org/10.1016/j.virol.2017.11.001>
- De Vlamincq, I., Khush, K. K., Strehl, C., Kohli, B., Luikart, H., Neff, N. F., ... Quake, S. R. (2013). Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell*, 155(5), 1178. <https://doi.org/10.1016/j.cell.2013.10.034>
- Drexler, J. F., Corman, V. M., Müller, M. A., Lukashev, A. N., Gmyl, A., Coutard, B., ... Drosten, C. (2013). Evidence for Novel Hepaciviruses in Rodents. *PLoS Pathogens*, 9(6). <https://doi.org/10.1371/journal.ppat.1003438>

- du Plessis, L., McCrone, J. T., Zarebski, A. E., Hill, V., Ruis, C., Gutierrez, B., ... Pybus, O. G. (2021). Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science*, *371*(6530), 708–712. <https://doi.org/10.1126/science.abf2946>
- Dudas, G., Carvalho, L. M., Bedford, T., Tatem, A. J., Baele, G., Faria, N. R., ... Rambaut, A. (2017). Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature*, *544*(7650), 309–315. <https://doi.org/10.1038/nature22040>
- Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*, 1–19. <https://doi.org/10.1186/1471-2105-5-113>
- Eibach, D., Hogan, B., Sarpong, N., Winter, D., Struck, N. S., Adu-Sarkodie, Y., ... Cadar, D. (2019). Viral metagenomics revealed novel betatorquevirus species in pediatric inpatients with encephalitis/meningoencephalitis from Ghana. *Scientific Reports*, *9*(1). <https://doi.org/10.1038/s41598-019-38975-z>
- Erlwein, O., Robinson, M. J., Dustan, S., Weber, J., Kaye, S., & McClure, M. O. (2011). DNA Extraction Columns Contaminated with Murine Sequences. *PLoS ONE*, *6*(8), e23484. <https://doi.org/10.1371/journal.pone.0023484>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Fahsbender, E., Burns, J. M., Kim, S., Kraberger, S., Frankfurter, G., Eilers, A. A., ... Varsani, A. (2017). Diverse and highly recombinant anelloviruses associated with Weddell seals in Antarctica. *Virus Evolution*, *3*(1), 1–11. <https://doi.org/10.1093/ve/vex017>
- Fancello, L., Monteil, S., Popgeorgiev, N., Rivet, R., Gouriet, F., Fournier, P. E., ... Desnues, C. (2014). Viral communities associated with human pericardial fluids in idiopathic pericarditis. *PLoS ONE*, *9*(4). <https://doi.org/10.1371/journal.pone.0093367>
- Feng, Y., Zhao, W., Feng, Y., Dai, J., Li, Z., Zhang, X., ... Xia, X. (2011). A novel genotype of GB virus C: Its identification and predominance among injecting drug users in Yunnan, China. *PLoS ONE*, *6*(10). <https://doi.org/10.1371/journal.pone.0021151>
- Fernández-Ruiz, M., Forque, L., Albert, E., Redondo, N., Giménez, E., López-Medrano, F., ... Navarro, D. (2021). Human pegivirus type 1 infection in kidney transplant recipients: Replication kinetics and clinical correlates. *Transplant Infectious Disease*. <https://doi.org/10.1111/tid.13771>
- Focosi, D., Antonelli, G., Pistello, M., & Maggi, F. (2016). Torquetenovirus: the

- human virome from bench to bedside. *Clinical Microbiology and Infection*. Elsevier B.V. <https://doi.org/10.1016/j.cmi.2016.04.007>
- Focosi, D., Spezia, P. G., Macera, L., Salvadori, S., Navarro, D., Lanza, M., ... Maggi, F. (2020). Assessment of prevalence and load of torquetenovirus viraemia in a large cohort of healthy blood donors. *Clinical Microbiology and Infection*, 26(10), 1406–1410. <https://doi.org/10.1016/j.cmi.2020.01.011>
- Focosi, Daniele, Macera, L., Pistello, M., & Maggi, F. (2014). Torque Teno Virus Viremia Correlates With Intensity of Maintenance Immunosuppression in Adult Orthotopic Liver Transplant. *The Journal of Infectious Diseases*, 210(4), 667–668. <https://doi.org/10.1093/infdis/jiu209>
- Focosi, Daniele, Maggi, F., Albani, M., Macera, L., Ricci, V., Gragnani, S., ... Petrini, M. (2010). Torquetenovirus viremia kinetics after autologous stem cell transplantation are predictable and may serve as a surrogate marker of functional immune reconstitution. *Journal of Clinical Virology*, 47(2), 189–192. <https://doi.org/10.1016/j.jcv.2009.11.027>
- Forns, X., Fernández-Llama, P., Costa, J., López-Labrador, F. X., Ampurdanés, S., Olmedo, E., ... Rodés, J. (1997). Hepatitis G virus infection in a haemodialysis unit: Prevalence and clinical implications. *Nephrology Dialysis Transplantation*, 12(5), 956–960. <https://doi.org/10.1093/ndt/12.5.956>
- Foulongne, V., Sauvage, V., Hebert, C., Dereure, O., Cheval, J., Gouilh, M. A., ... Eloit, M. (2012). Human Skin Microbiota: High Diversity of DNA Viruses Identified on the Human Skin by High Throughput Sequencing. *PLoS ONE*, 7(6), e38499. <https://doi.org/10.1371/journal.pone.0038499>
- French, R. K., & Holmes, E. C. (2020). An Ecosystems Perspective on Virus Evolution and Emergence. *Trends in Microbiology*, 28(3), 165–175. <https://doi.org/10.1016/j.tim.2019.10.010>
- Frutos, R., Gavotte, L., & Devaux, C. A. (2021). Understanding the origin of COVID-19 requires to change the paradigm on zoonotic emergence from the spillover to the circulation model. *Infection, Genetics and Evolution*, 9, 104812. <https://doi.org/10.1016/j.meegid.2021.104812>
- Frye, B. C., Bierbaum, S., Falcone, V., Köhler, T. C., Gasplmayr, M., Hettich, I., ... Müller-Quernheim, J. (2019). Kinetics of torque teno virus-DNA plasma load predict rejection in lung transplant recipients. *Transplantation*, 103(4), 815–822. <https://doi.org/10.1097/TP.0000000000002436>
- Furuta, R. A., Sakamoto, H., Kuroishi, A., Yasiui, K., Matsukura, H., & Hirayama, F. (2015). Metagenomic profiling of the viromes of plasma collected from blood donors with elevated serum alanine aminotransferase levels. *Transfusion*, 55(8), 1889–1899. <https://doi.org/10.1111/trf.13057>

- Garretto, A., Thomas-White, K., Wolfe, A. J., & Putonti, C. (2018). Detecting viral genomes in the female urinary microbiome. *Journal of General Virology*, *99*(8), 1141–1146. <https://doi.org/10.1099/jgv.0.001097>
- Ghai, R. R., Sibley, S. D., Lauck, M., Dinis, J. M., Bailey, A. L., Chapman, C. A., ... Goldberg, T. L. (2013). Deep sequencing identifies two genotypes and high viral genetic diversity of human pegivirus (GB virus C) in rural Ugandan patients. *Journal of General Virology*, *94*(12), 2670–2678. <https://doi.org/10.1099/vir.0.055509-0>
- Gibbs, M. J., Armstrong, J. S., & Gibbs, A. J. (2000). Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*, *16*(7), 573–582. <https://doi.org/10.1093/bioinformatics/16.7.573>
- Gire, S. K., Goba, A., Andersen, K. G., Sealfon, R. S. G., Park, D. J., Kanneh, L., ... Sabeti, P. C. (2014). Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*, *345*(6202), 1369–1372. <https://doi.org/10.1126/science.1259657>
- Gonzales-Gustavson, E., Timoneda, N., Fernandez-Cassi, X., Caballero, A., Abril, J. F., Buti, M., ... Girones, R. (2017). Identification of sapovirus GV.2, astrovirus VA3 and novel anelloviruses in serum from patients with acute hepatitis of unknown aetiology. *PLoS ONE*, *12*(10), 1–17. <https://doi.org/10.1371/journal.pone.0185911>
- Górski, A., Wazna, E., Dabrowska, B.-W., Dabrowska, K., Kinga, S.-J., & Miedzybrodzki, R. (2006). Bacteriophage translocation. *FEMS Immunology & Medical Microbiology*, *46*(3), 313–319. <https://doi.org/10.1111/j.1574-695X.2006.00044.x>
- Greenhalgh, S., Schmidt, R., & Day, T. (2019). Fighting the Public Health Burden of AIDS with the Human Pegivirus. *American Journal of Epidemiology*, *188*(9), 1586–1594. <https://doi.org/10.1093/aje/kwz139>
- Gregory, A. C., Zablocki, O., Zayed, A. A., Howell, A., Bolduc, B., & Sullivan, M. B. (2020). The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. *Cell Host & Microbe*, *28*(5), 724–740.e8. <https://doi.org/10.1016/j.chom.2020.08.003>
- Greninger, A. L., Naccache, S. N., Federman, S., Yu, G., Mbala, P., Bres, V., ... Chiu, C. Y. (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Medicine*, *7*(1), 1–13. <https://doi.org/10.1186/s13073-015-0220-9>
- Grubaugh, N. D., Ladner, J. T., Kraemer, M. U. G., Dudas, G., Tan, A. L., Gangavarapu, K., ... Andersen, K. G. (2017). Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature*, *546*(7658), 401–405. <https://doi.org/10.1038/nature22400>

- Hall, R. J., Wang, J., Todd, A. K., Bissielo, A. B., Yen, S., Strydom, H., ... Peacey, M. (2014). Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *Journal of Virological Methods*, *195*, 194–204. <https://doi.org/10.1016/j.jviromet.2013.08.035>
- Haloschan, M., Bettesch, R., Görzer, I., Weseslindtner, L., Kundi, M., & Puchhammer-Stöckl, E. (2014). TTV DNA plasma load and its association with age, gender, and HCMV IgG serostatus in healthy adults. *Age*, *36*(5). <https://doi.org/10.1007/s11357-014-9716-2>
- Hannigan, G. D., Meisel, J. S., Tyldsley, A. S., Zheng, Q., Hodkinson, B. P., SanMiguel, A. J., ... Grice, E. A. (2015). The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome. *MBio*, *6*(5). <https://doi.org/10.1128/mBio.01578-15>
- Harvey, E., & Holmes, E. C. (2022). Diversity and evolution of the animal virome. *Nature Reviews Microbiology*. <https://doi.org/10.1038/s41579-021-00665-x>
- Haynes, M., & Rohwer, F. (2011). The Human Virome. In *Metagenomics of the Human Body* (pp. 63–77). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-7089-3_4
- Hill, V., Ruis, C., Bajaj, S., Pybus, O. G., & Kraemer, M. U. G. (2021). Progress and challenges in virus genomic epidemiology. *Trends in Parasitology*, *37*(12), 1038–1049. <https://doi.org/10.1016/j.pt.2021.08.007>
- Holmes, E. C. (2011). What Does Virus Evolution Tell Us about Virus Origins? *Journal of Virology*, *85*(11), 5247–5251. <https://doi.org/10.1128/jvi.02203-10>
- Hrazdilová, K., Slaninková, E., Brožová, K., Modrý, D., Vodička, R., & Celer, V. (2016). New species of Torque Teno miniviruses infecting gorillas and chimpanzees. *Virology*, *487*, 207–214. <https://doi.org/10.1016/j.virol.2015.10.016>
- Hsiao, K. L., Wang, L. Y., Lin, C. L., & Liu, H. F. (2016). New phylogenetic groups of torque teno virus identified in eastern Taiwan indigenes. *PLoS ONE*, *11*(2). <https://doi.org/10.1371/journal.pone.0149901>
- Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–214. <https://doi.org/10.1038/nature11234>
- Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267. <https://doi.org/10.1093/molbev/msj030>
- Huson, D. H., & Mitra, S. (2012). Introduction to the Analysis of Environmental Sequences: Metagenomics with MEGAN (pp. 415–429).

- https://doi.org/10.1007/978-1-61779-585-5_17
- Jonas, O., & Seifman, R. (2019). Do we need a Global Virome Project? *The Lancet Global Health*, 7(10), e1314–e1316. [https://doi.org/10.1016/S2214-109X\(19\)30335-3](https://doi.org/10.1016/S2214-109X(19)30335-3)
- Jordier, F., Deligny, M. L., Barré, R., Robert, C., Galicher, V., Uch, R., ... Biagini, P. (2019). Human pegivirus isolates characterized by deep sequencing from hepatitis C virus-RNA and human immunodeficiency virus-RNA–positive blood donations, France. *Journal of Medical Virology*, 91(1), 38–44. <https://doi.org/10.1002/jmv.25290>
- Jurasz, H., Pawłowski, T., & Perlejewski, K. (2021). Contamination Issue in Viral Metagenomics: Problems, Solutions, and Clinical Perspectives. *Frontiers in Microbiology*, 12. <https://doi.org/10.3389/fmicb.2021.745076>
- Kaczorowska, J., & Van Der Hoek, L. (2020). Human anelloviruses: Diverse, omnipresent and commensal members of the virome. *FEMS Microbiology Reviews*, 44(3), 305–313. <https://doi.org/10.1093/femsre/fuaa007>
- Kakkola, L., Bondén, H., Hedman, L., Kivi, N., Moisala, S., Julin, J., ... Söderlund-Venermo, M. (2008). Expression of all six human Torque teno virus (TTV) proteins in bacteria and in insect cells, and analysis of their IgG responses. *Virology*, 382(2), 182–189. <https://doi.org/10.1016/j.virol.2008.09.012>
- Kapoor, A., Kumar, A., Simmonds, P., Bhuvana, N., Chauhan, L. S., Lee, B., ... Ian Lipkina, W. (2015). Virome analysis of transfusion recipients reveals a novel human virus that shares genomic features with hepaciviruses and pegiviruses. *MBio*, 6(5). <https://doi.org/10.1128/mBio.01466-15>
- Kapoor, A., Simmonds, P., Scheel, T. K. H., Hjelle, B., Cullen, J. M., Burbelo, P. D., ... Lipkin, W. I. (2013). Identification of rodent homologs of hepatitis C virus and pegiviruses. *MBio*, 4(2). <https://doi.org/10.1128/mBio.00216-13>
- Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kernbauer, E., Ding, Y., & Cadwell, K. (2014). An enteric virus can replace the beneficial function of commensal bacteria. *Nature*, 516(729), 94–98. <https://doi.org/10.1038/nature13960>
- Khan Mirzaei, M., Xue, J., Costa, R., Ru, J., Schulz, S., Taranu, Z. E., & Deng, L. (2021). Challenges of Studying the Human Virome – Relevant Emerging Technologies. *Trends in Microbiology*, 29(2), 171–181. <https://doi.org/10.1016/j.tim.2020.05.021>

- Kim, D., Song, L., Breitwieser, F. P., & Salzberg, S. L. (2016). Centrifuge: rapid and accurate classification of metagenomic sequences, version 1.0.4_beta. *BioRxiv*, 26(12), 054965. <https://doi.org/10.1101/gr.210641.116>.Freely
- Kobayashi, N. (2018). Impact of Emerging, Re-Emerging and Zoonotic Viral Infectious Diseases, in a Virologist's Perspective. *The Open Virology Journal*, 12(1), 131–133. <https://doi.org/10.2174/1874357901812010131>
- Kohl, C., Brinkmann, A., Dabrowski, P. W., Radonić, A., Nitsche, A., & Kurth, A. (2015). Protocol for metagenomic virus detection in clinical specimens. *Emerging Infectious Diseases*, 21(1), 48–57. <https://doi.org/10.3201/eid2101.140766>
- Krajden, M., Yu, A., Braybrook, H., Lai, A. S., Mak, A., Chow, R., ... Spinelli, J. J. (2010). GBV-C/hepatitis G virus infection and non-Hodgkin lymphoma: A case control study. *International Journal of Cancer*, 126(12), 2885–2892. <https://doi.org/10.1002/ijc.25035>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Kyathanahalli, C., Snedden, M., & Hirsch, E. (2021). Human Anelloviruses: Prevalence and Clinical Significance During Pregnancy. *Frontiers in Virology*, 1, 1–12. <https://doi.org/10.3389/fviro.2021.782886>
- Lam, H. M., Ratmann, O., & Boni, M. F. (2018). Improved Algorithmic Complexity for the 3SEQ Recombination Detection Algorithm. *Molecular Biology and Evolution*, 35(1), 247–251. <https://doi.org/10.1093/molbev/msx263>
- Lauck, M., Bailey, A. L., Andersen, K. G., Goldberg, T. L., Sabeti, P. C., & O'Connor, D. H. (2015). GB Virus C Coinfections in West African Ebola Patients. *Journal of Virology*, 89(4), 2425–2429. <https://doi.org/10.1128/jvi.02752-14>
- Law, J., Jovel, J., Patterson, J., Ford, G., O'keefe, S., Wang, W., ... Wong, G. K. S. (2013). Identification of Hepatotropic Viruses from Plasma Using Deep Sequencing: A Next Generation Diagnostic Tool. *PLoS ONE*, 8(4). <https://doi.org/10.1371/journal.pone.0060595>
- Li, L., Deng, X., Linsuwanon, P., Bangsberg, D., Bwana, M. B., Hunt, P., ... Delwart, E. (2013). AIDS Alters the Commensal Plasma Virome. *Journal of Virology*, 87(19), 10912–10915. <https://doi.org/10.1128/jvi.01839-13>
- Li, Y., Fang, L., Cui, S., Fu, J., Li, X., Zhang, H., ... Zhao, P. (2017). Genomic characterization of recent chicken anemia virus isolates in China.

- Frontiers in Microbiology*, 8, 1–9.
<https://doi.org/10.3389/fmicb.2017.00401>
- Liang, G., & Bushman, F. D. (2021). The human virome: assembly, composition and host interactions. *Nature Reviews Microbiology*, 19(8), 514–527.
<https://doi.org/10.1038/s41579-021-00536-5>
- Liang, G., Conrad, M. A., Kelsen, J. R., Kessler, L. R., Breton, J., Albenberg, L. G., ... Bushman, F. D. (2020). Dynamics of the Stool Virome in Very Early-Onset Inflammatory Bowel Disease. *Journal of Crohn's and Colitis*, 14(11), 1600–1610. <https://doi.org/10.1093/ecco-jcc/jjaa094>
- Liang, G., Zhao, C., Zhang, H., Mattei, L., Sherrill-Mix, S., Bittinger, K., ... Bushman, F. D. (2020). The stepwise assembly of the neonatal virome is modulated by breastfeeding. *Nature*, 581(7809), 470–474.
<https://doi.org/10.1038/s41586-020-2192-1>
- Lim, E. S., Zhou, Y., Zhao, G., Bauer, I. K., Droit, L., Ndao, I. M., ... Holtz, L. R. (2015). Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nature Medicine*, 21(10), 1228–1234.
<https://doi.org/10.1038/nm.3950>
- Linnen, J., Wages, J., Zhang-Keck, Z.-Y., Fry, K. E., Krawczynski, K. Z., Alter, H., ... Kim, J. P. (1996). Molecular Cloning and Disease Association of Hepatitis G Virus: A Transfusion-Transmissible Agent. *Science*, 271(5248), 505–508.
<https://doi.org/10.1126/science.271.5248.505>
- Lipkin, W. Ian, & Anthony, S. J. (2015). Virus hunting. *Virology*. Academic Press Inc. <https://doi.org/10.1016/j.virol.2015.02.006>
- Lipkin, Walter Ian, & Firth, C. (2013). Viral surveillance and discovery. *Current Opinion in Virology*, 3(2), 199–204.
<https://doi.org/10.1016/j.coviro.2013.03.010>
- Lombardi, V. C., Ruscetti, F. W., Das Gupta, J., Pfof, M. A., Hagen, K. S., Peterson, D. L., ... Mikovits, J. A. (2009). Detection of an Infectious Retrovirus, XMRV, in Blood Cells of Patients with Chronic Fatigue Syndrome. *Science*, 326(5952), 585–589.
<https://doi.org/10.1126/science.1179052>
- Ma, Y., Madupu, R., Karaoz, U., Nossa, C. W., Yang, L., Yooseph, S., ... Pei, Z. (2014). Human Papillomavirus Community in Healthy Persons, Defined by Metagenomics Analysis of Human Microbiome Project Shotgun Sequencing Data Sets. *Journal of Virology*, 88(9), 4786–4797.
<https://doi.org/10.1128/JVI.00093-14>
- Maggi, F., Andreoli, E., Lanini, L., Fornai, C., Vatteroni, M., Pistello, M., ... Bendinelli, M. (2005). Relationships between total plasma load of torquetenovirus (TTV) and TTV genogroups carried. *Journal of Clinical Microbiology*, 43(9), 4807–4810.

- <https://doi.org/10.1128/JCM.43.9.4807-4810.2005>
- Magouras, I., Brookes, V. J., Jori, F., Martin, A., Pfeiffer, D. U., & Dürr, S. (2020). Emerging Zoonotic Diseases: Should We Rethink the Animal–Human Interface? *Frontiers in Veterinary Science*, *7*, 1–6. <https://doi.org/10.3389/fvets.2020.582743>
- Maidana-Giret, M. T., Silva, T. M., Sauer, M. M., Tomiyama, H., Levi, J. E., Bassichetto, K. C., ... Kallas, E. G. (2009). GB virus type C infection modulates T-cell activation independently of HIV-1 viral load. *Aids*, *23*(17), 2277–2287. <https://doi.org/10.1097/QAD.0b013e32832d7a11>
- Maio, F., Helderma, T. A., Arroyo-Mateos, M., van der Wolf, M., Boeren, S., Prins, M., & van den Burg, H. A. (2020). Identification of Tomato Proteins That Interact With Replication Initiator Protein (Rep) of the Geminivirus TYLCV. *Frontiers in Plant Science*, *11*. <https://doi.org/10.3389/fpls.2020.01069>
- Marano, G., Franchini, M., Farina, B., Piccinini, V., Pupella, S., Vaglio, S., ... Liunbruno, G. M. (2017). The human pegivirus: A new name for an “ancient” virus. Can transfusion medicine come up with something new? *Acta Virologica*, *61*(4), 401–412. https://doi.org/10.4149/av_2017_402
- Martí, J. M. (2019). Recentrifuge: Robust comparative analysis and contamination removal for metagenomics. *PLOS Computational Biology*, *15*(4), e1006967-. Retrieved from <https://doi.org/10.1371/journal.pcbi.1006967>
- Martí, J. M., & Garay, C. P. (2019). Not just BLAST nt: WGS database joins the party. *BioRxiv*, 1–15. <https://doi.org/10.1101/653592>
- Martin, D., & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, *16*(6), 562–563. <https://doi.org/10.1093/bioinformatics/16.6.562>
- Martin, D.P., Posada, D., Crandall, K. A., & Williamson, C. (2005). A Modified Bootscan Algorithm for Automated Identification of Recombinant Sequences and Recombination Breakpoints. *AIDS Research and Human Retroviruses*, *21*(1), 98–102. <https://doi.org/10.1089/aid.2005.21.98>
- Martin, Darren P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(1), 1–5. <https://doi.org/10.1093/ve/vev003>
- Miao, Z., Gao, L., Song, Y., Yang, M., Zhang, M., Lou, J., ... Xia, X. (2017). Prevalence and clinical impact of human pegivirus-1 infection in HIV-1-infected individuals in Yunnan, China. *Viruses*, *9*(2), 1–13. <https://doi.org/10.3390/v9020028>
- Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S. A., Wu, G. D., ... Bushman, F. D. (2011). The human gut virome: Inter-individual variation and dynamic

- response to diet. *Genome Research*, 21(10), 1616–1625. <https://doi.org/10.1101/gr.122705.111>
- Mohr, E. L., & Stapleton, J. T. (2009). GB virus type C interactions with HIV: the role of envelope glycoproteins. *Journal of Viral Hepatitis*, 16(11), 757–768. <https://doi.org/10.1111/j.1365-2893.2009.01194.x>
- Morrison, D. A. (2005). The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny.—Marco Salemi and Anne-Mieke Vandamme (editors). 2003. Cambridge University Press, Cambridge, UK. *Systematic Biology*, 54(6), 984–986. <https://doi.org/10.1080/10635150500354811>
- Moustafa, A., Xie, C., Kirkness, E., Biggs, W., Wong, E., Turpaz, Y., ... Telenti, A. (2017). The blood DNA virome in 8,000 humans. *PLOS Pathogens*, 13(3), e1006292. <https://doi.org/10.1371/journal.ppat.1006292>
- Muerhoff, A. S., Tillmann, H. L., Manns, M. P., Dawson, G. J., & Desai, S. M. (2003). GB virus C genotype determination in GB virus-C/HIV co-infected individuals. *Journal of Medical Virology*, 70(1), 141–149. <https://doi.org/10.1002/jmv.10375>
- Muhire, B. M., Varsani, A., & Martin, D. P. (2014). SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. *PLoS ONE*, 9(9). <https://doi.org/10.1371/journal.pone.0108277>
- Naccache, S. N., Greninger, A. L., Lee, D., Coffey, L. L., Phan, T., Rein-Weston, A., ... Chiu, C. Y. (2013). The Perils of Pathogen Discovery: Origin of a Novel Parvovirus-Like Hybrid Genome Traced to Nucleic Acid Extraction Spin Columns. *Journal of Virology*, 87(22), 11966–11977. <https://doi.org/10.1128/jvi.02323-13>
- Neibecker, M., Schwarze-Zander, C., Rockstroh, J. K., Spengler, U., & Blackard, J. T. (2011). Evidence for extensive genotypic diversity and recombination of GB virus C (GBV-C) in Germany. *Journal of Medical Virology*, 83(4), 685–694. <https://doi.org/10.1002/jmv.22029>
- Nelson, J. R. (2014). Random-Primed, Phi29 DNA Polymerase-Based Whole Genome Amplification. *Current Protocols in Molecular Biology*, 105(1), 15.13.1-15.13.16. <https://doi.org/https://doi.org/10.1002/0471142727.mb1513s105>
- Ng, T. F. F., Dill, J. A., Camus, A. C., Delwart, E., & Van Meir, E. G. (2017). Two new species of betatorqueviruses identified in a human melanoma that metastasized to the brain. *Oncotarget*, 8(62), 105800–105808. <https://doi.org/10.18632/oncotarget.22400>
- Ng, T. F. F., Wheeler, E., Greig, D., Waltzek, T. B., Gulland, F., & Breitbart, M. (2011). Metagenomic identification of a novel anellovirus in Pacific harbor seal (*Phoca vitulina richardsii*) lung samples and its detection in samples from multiple years. *Journal of General Virology*, 92(6), 1318–

1323. <https://doi.org/10.1099/vir.0.029678-0>
- Ngoi, C. N., Siqueira, J., Li, L., Deng, X., Mugo, P., Graham, S. M., ... Delwart, E. (2016). The plasma virome of febrile adult kenyans shows frequent parvovirus B19 infections and a novel arbovirus (Kadapiro virus). *Journal of General Virology*, 97(12), 3359–3367. <https://doi.org/10.1099/jgv.0.000644>
- Ngoi, C. N., Siqueira, J., Li, L., Deng, X., Mugo, P., Graham, S. M., ... Delwart, E. (2017). Corrigendum: The plasma virome of febrile adult Kenyans shows frequent parvovirus B19 infections and a novel arbovirus (Kadapiro virus). *Journal of General Virology*, 98(3), 517–517. <https://doi.org/10.1099/jgv.0.000762>
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Ninomiya, M., Nishizawa, T., Takahashi, M., Lorenzo, F. R., Shimosegawa, T., & Okamoto, H. (2007). Identification and genomic characterization of a novel human torque teno virus of 3.2 kb. *Journal of General Virology*, 88(7), 1939–1944. <https://doi.org/10.1099/vir.0.82895-0>
- Ninomiya, M., Takahashi, M., Hoshino, Y., Ichiyama, K., Simmonds, P., & Okamoto, H. (2009). Analysis of the entire genomes of Torque teno midi virus variants in chimpanzees: Infrequent cross-species infection between humans and chimpanzees. *Journal of General Virology*, 90(2), 347–358. <https://doi.org/10.1099/vir.0.007385-0>
- Nishizawa, T., Okamoto, H., Konishi, K., Yoshizawa, H., Miyakawa, Y., & Mayumi, M. (1997). A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochemical and Biophysical Research Communications*, 241(1), 92–97. <https://doi.org/10.1006/bbrc.1997.7765>
- Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., & Koopmans, M. P. G. (2018). Overview of virus metagenomic classification methods and their biological applications. *Frontiers in Microbiology*. Frontiers Media S.A. <https://doi.org/10.3389/fmicb.2018.00749>
- Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., ... Pevzner, P. (2013). *Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads In: Research in Computational Molecular Biology. RECOMB 2013. Lecture Notes in Computer Science*. Retrieved from <http://cab.spbu.ru/software/spades/>
- Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). MetaSPAdes: A new versatile metagenomic assembler. *Genome Research*, 27(5), 824–

834. <https://doi.org/10.1101/gr.213959.116>
- Okamoto, H., Takahashi, M., Nishizawa, T., Tawara, A., Fukai, K., Muramatsu, U., ... Yoshikawa, A. (2002). Genomic characterization of TT viruses (TTVs) in pigs, cats and dogs and their relatedness with species-specific TTVs in primates and tupaías. *Journal of General Virology*, *83*(6), 1291–1297. <https://doi.org/10.1099/0022-1317-83-6-1291>
- Padidam, M., Sawyer, S., & Fauquet, C. M. (1999). Possible Emergence of New Geminiviruses by Frequent Recombination. *Virology*, *265*(2), 218–225. <https://doi.org/10.1006/viro.1999.0056>
- Pan, S., Yu, T., Wang, Y., Lu, R., Wang, H., Xie, Y., & Feng, X. (2018). Identification of a torque teno mini virus (TTMV) in Hodgkin's lymphoma patients. *Frontiers in Microbiology*, *9*, 1–8. <https://doi.org/10.3389/fmicb.2018.01680>
- Paprotka, T., Delviks-Frankenberry, K. A., Cingöz, O., Martinez, A., Kung, H.-J., Tepper, C. G., ... Pathak, V. K. (2011). Recombinant Origin of the Retrovirus XMRV. *Science*, *333*(6038), 97–101. <https://doi.org/10.1126/science.1205292>
- Parras-Moltó, M., Rodríguez-Galet, A., Suárez-Rodríguez, P., & López-Bueno, A. (2018). Evaluation of bias induced by viral enrichment and random amplification protocols in metagenomic surveys of saliva DNA viruses. *Microbiome*, *6*(1). <https://doi.org/10.1186/s40168-018-0507-3>
- Parreira, R., Branco, C., Piedade, J., & Esteves, A. (2012). GB virus C (GBV-C) evolutionary patterns revealed by analyses of reference genomes, E2 and NS5B sequences amplified from viral strains circulating in the Lisbon area (Portugal). *Infection, Genetics and Evolution*, *12*(1), 86–93. <https://doi.org/10.1016/j.meegid.2011.10.011>
- Pavesi, A. (2001). Origin and evolution of GBV-C/hepatitis G virus and relationships with ancient human migrations. *Journal of Molecular Evolution*, *53*(2), 104–113. <https://doi.org/10.1007/s002390010198>
- Pérez-Brocal, V., & Moya, A. (2018). The analysis of the oral DNA virome reveals which viruses are widespread and rare among healthy young adults in Valencia (Spain). *PLoS ONE*, *13*(2). <https://doi.org/10.1371/journal.pone.0191867>
- Pérez-Losada, M., Arenas, M., Galán, J. C., Palero, F., & González-Candelas, F. (2015). Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. *Infection, Genetics and Evolution*, *30*, 296–307. <https://doi.org/10.1016/j.meegid.2014.12.022>
- Phan, T. G., Desnues, C., Switzer, W. M., Djoko, C. F., Schneider, B. S., Deng, X., & Delwart, E. (2015). Absence of giant blood Marseille-like virus DNA detection by polymerase chain reaction in plasma from healthy US blood

- donors and serum from multiply transfused patients from Cameroon. *Transfusion*, 55(6), 1256–1262. <https://doi.org/10.1111/trf.12997>
- Popgeorgiev, N., Boyer, M., Fancello, L., Monteil, S., Robert, C., Rivet, R., ... Desnues, C. (2013). Marseillevirus-like virus recovered from blood donated by asymptomatic humans. *Journal of Infectious Diseases*, 208(7), 1042–1050. <https://doi.org/10.1093/infdis/jit292>
- Popgeorgiev, N., Temmam, S., Raoult, D., & Desnues, C. (2013). Describing the silent human virome with an emphasis on giant viruses. *Intervirology*, 56(6), 395–412. <https://doi.org/10.1159/000354561>
- Porter, A. F., Cobbin, J., Li, C. X., Eden, J. S., & Holmes, E. C. (2021). Metagenomic identification of viral sequences in laboratory reagents. *Viruses*, 13(11), 1–13. <https://doi.org/10.3390/v13112122>
- Posada, D., & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences*, 98(24), 13757–13762. <https://doi.org/10.1073/pnas.241370698>
- Pride, D. T., Salzman, J., Haynes, M., Rohwer, F., Davis-Long, C., White, R. A., ... Relman, D. A. (2012). Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *The ISME Journal*, 6(5), 915–926. <https://doi.org/10.1038/ismej.2011.169>
- Rascovan, N., Duraisamy, R., & Desnues, C. (2016). Metagenomics and the Human Virome in Asymptomatic Individuals. *Annual Review of Microbiology*. Annual Reviews Inc. <https://doi.org/10.1146/annurev-micro-102215-095431>
- Reshetnyak, Vasiliy I., Maev, I. V., Burmistrov, A. I., Chekmazov, I. A., & Karlovich, T. I. (2020). Torque teno virus in liver diseases: On the way towards unity of view. *World Journal of Gastroenterology*, 26(15), 1691–1707. <https://doi.org/10.3748/WJG.V26.I15.1691>
- Reshetnyak, Vasiliy Ivanovich, Karlovich, T. I., & Ilchenko, L. U. (2008). Hepatitis G virus. *World Journal of Gastroenterology*, 14(30), 4725–4734. <https://doi.org/10.3748/wjg.14.4725>
- Reyes, A., Haynes, M., Hanson, N., Angly, F. E., Heath, A. C., Rohwer, F., & Gordon, J. I. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*, 466(7304), 334–338. <https://doi.org/10.1038/nature09199>
- Reyes, A., Semenkovich, N. P., Whiteson, K., Rohwer, F., & Gordon, J. I. (2012). Going viral: next-generation sequencing applied to phage populations in the human gut. *Nature Reviews Microbiology*, 10(9), 607–617. <https://doi.org/10.1038/nrmicro2853>
- Reyes, G. R., & Kim, J. P. (1991). Sequence-independent, single-primer

- amplification (SISPA) of complex DNA populations. *Molecular and Cellular Probes*, 5(6), 473–481. [https://doi.org/10.1016/S0890-8508\(05\)80020-9](https://doi.org/10.1016/S0890-8508(05)80020-9)
- Roberts, M., Dobson, A., Restif, O., & Wells, K. (2021). Challenges in modelling the dynamics of infectious diseases at the wildlife-human interface. *Epidemics*, 37(November), 100523. <https://doi.org/10.1016/j.epidem.2021.100523>
- Roossinck, M. J. (2015). Plants, viruses and the environment: Ecology and mutualism. *Virology*, 479–480, 271–277. <https://doi.org/10.1016/j.virol.2015.03.041>
- Rosario, K., & Breitbart, M. (2011). Exploring the viral world through metagenomics. *Current Opinion in Virology*, 1(4), 289–297. <https://doi.org/10.1016/j.coviro.2011.06.004>
- Rosseel, T., Ozhelvaci, O., Freimanis, G., & Van Borm, S. (2015). Evaluation of convenient pretreatment protocols for RNA virus metagenomics in serum and tissue samples. *Journal of Virological Methods*, 222, 72–80. <https://doi.org/10.1016/j.jviromet.2015.05.010>
- Rupasinghe, R., Chomel, B. B., & Martínez-López, B. (2022). Climate change and zoonoses: A review of the current status, knowledge gaps, and future trends. *Acta Tropica*, 226, 106225. <https://doi.org/10.1016/j.actatropica.2021.106225>
- Sabir, J. S. M., Lam, T. T.-Y., Ahmed, M. M. M., Li, L., Shen, Y., E. M. Abo-Aba, S., ... Guan, Y. (2016). Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science*, 351(6268), 81–84. <https://doi.org/10.1126/science.aac8608>
- Santiago-Rodriguez, T. M., & Hollister, E. B. (2020). Potential Applications of Human Viral Metagenomics and Reference Materials: Considerations for Current and Future Viruses. *Applied and Environmental Microbiology*, 86(22), 1–12. <https://doi.org/10.1128/AEM.01794-20>
- Sathiamoorthy, S., Malott, R. J., Gisonni-Lex, L., & Ng, S. H. S. (2018). Selection and evaluation of an efficient method for the recovery of viral nucleic acids from complex biologicals. *Npj Vaccines*, 3(1), 31. <https://doi.org/10.1038/s41541-018-0067-3>
- Sauvage, V., & Eloit, M. (2016). Viral metagenomics and blood safety. *Transfusion Clinique et Biologique*, 23(1), 28–38. <https://doi.org/10.1016/j.tracli.2015.12.002>
- Sauvage, Virginie, Livartowski, A., Boizeau, L., Servant-Delmas, A., Lionnet, F., Lefrère, J. J., & Laperche, S. (2014). No evidence of marseillevirus-like virus presence in blood donors and recipients of multiple blood transfusions. *Journal of Infectious Diseases*, 210(12), 2017–2018.

- <https://doi.org/10.1093/infdis/jiu443>
- Schwarze-Zander, C., Blackard, J. T., & Rockstroh, J. K. (2012). Role of GB virus C in modulating HIV disease. *Expert Review of Anti-Infective Therapy*, *10*(5), 563–572. <https://doi.org/10.1586/eri.12.37>
- Schwarze-Zander, C., Blackard, J. T., Zheng, H., Addo, M. M., Lin, W., Robbins, G. K., ... Chung, R. T. (2006). GB virus C (GBV-C) infection in hepatitis C virus (HCV)/HIV-coinfected patients receiving HCV treatment: Importance of the GBV-C genotype. *Journal of Infectious Diseases*, *194*(4), 410–419. <https://doi.org/10.1086/505713>
- Segura-Wang, M., Görzer, I., Jaksch, P., & Puchhammer-Stöckl, E. (2018). Temporal dynamics of the lung and plasma viromes in lung transplant recipients. *PLoS ONE*, *13*(7). <https://doi.org/10.1371/journal.pone.0200428>
- Seppey, M., Manni, M., & Zdobnov, E. M. (2020). LEMMI: a continuous benchmarking platform for metagenomics classifiers. *Genome Research*, *30*(8), 1208–1216. <https://doi.org/10.1101/gr.260398.119>
- Sharp, P. M., & Simmonds, P. (2011). Evaluating the evidence for virus/host co-evolution. *Current Opinion in Virology*, *1*(5), 436–441. <https://doi.org/10.1016/j.coviro.2011.10.018>
- Shi, C., Liu, Y., Hu, X., Xiong, J., Zhang, B., & Yuan, Z. (2015). A metagenomic survey of viral abundance and diversity in mosquitoes from hubei province. *PLoS ONE*, *10*(6), 1–15. <https://doi.org/10.1371/journal.pone.0129845>
- Shi, M., Zhang, Y.-Z., & Holmes, E. C. (2018). Meta-transcriptomics and the evolutionary biology of RNA viruses. *Virus Research*, *243*, 83–90. <https://doi.org/10.1016/j.virusres.2017.10.016>
- Shimodaira, H. (2002). An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology*, *51*(3), 492–508. <https://doi.org/10.1080/10635150290069913>
- Shkoporov, A. N., & Hill, C. (2019). Bacteriophages of the Human Gut: The “Known Unknown” of the Microbiome. *Cell Host & Microbe*, *25*(2), 195–209. <https://doi.org/10.1016/j.chom.2019.01.017>
- Shoaib, M., Baconnais, S., Mechold, U., Le Cam, E., Lipinski, M., & Ogryzko, V. (2008). Multiple displacement amplification for complex mixtures of DNA fragments. *BMC Genomics*, *9*, 1–14. <https://doi.org/10.1186/1471-2164-9-415>
- Shreiner, A. B., Kao, J. Y., & Young, V. B. (2015). The gut microbiome in health and in disease. *Current Opinion in Gastroenterology*, *31*(1), 69–75. <https://doi.org/10.1097/MOG.000000000000139>
- Sibley, S. D., Lauck, M., Bailey, A. L., Hyeroba, D., Tumukunde, A., Weny, G., ...

- Friedrich, T. C. (2014). Discovery and characterization of distinct simian pegiviruses in three wild African old world monkey species. *PLoS ONE*, 9(6). <https://doi.org/10.1371/journal.pone.0098569>
- Simmonds, P., Adams, M. J., Benk, M., Breitbart, M., Brister, J. R., Carstens, E. B., ... Zerbini, F. M. (2017). Consensus statement: Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*, 15(3), 161–168. <https://doi.org/10.1038/nrmicro.2016.177>
- Simons, J. N., Leary, T. P., Dawson, G. J., Pilot-Matias, T. J., Muerhoff, A. S., Schlauder, G. G., ... Mushahwar, I. K. (1995). Isolation of novel virus-like sequences associated with human hepatitis. *Nature Medicine*, 1(6), 564–569. <https://doi.org/10.1038/nm0695-564>
- Smith, D. B., Basaras, M., Frost, S., Haydon, D., Cuceanu, N., Prescott, L., ... Simmonds, P. (2000). Phylogenetic analysis of GBV-C / hepatitis G virus, 769–780.
- Smith, J. (1992). Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, 34(2). <https://doi.org/10.1007/BF00182389>
- Smuts, H., Kew, M., Khan, A., & Korsman, S. (2014). Novel Hybrid Parvovirus-Like Virus, NIH-CQV/PHV, Contaminants in Silica Column-Based Nucleic Acid Extraction Kits. *Journal of Virology*, 88(2), 1398–1398. <https://doi.org/10.1128/JVI.03206-13>
- Soliman, H. K., Abouelhoda, M., El Roubi, M. N., Ahmed, O. S., Esmat, G., Hassan, Z. K., ... Zekri, A.-R. N. (2019). Whole-genome sequencing of human Pegivirus variant from an Egyptian patient co-infected with hepatitis C virus: a case report. *Virology Journal*, 16(1), 132. <https://doi.org/10.1186/s12985-019-1242-5>
- Spandole-Dinu, S., Cimponeriu, D. G., Crăciun, A. M., Radu, I., Nica, S., Toma, M., ... Nica, R. (2018). Prevalence of human anelloviruses in Romanian healthy subjects and patients with common pathologies. *BMC Infectious Diseases*, 18(1), 1–10. <https://doi.org/10.1186/s12879-018-3248-9>
- Spandole, S., Cimponeriu, D., Berca, L. M., & Mihăescu, G. (2015). Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Archives of Virology*. Springer-Verlag Wien. <https://doi.org/10.1007/s00705-015-2363-9>
- Stapleton, J. T., Fong, S., Muerhoff, A. S., Bukh, J., & Simmonds, P. (2011). The GB viruses: A review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae. *Journal of General Virology*, 92(2), 233–246. <https://doi.org/10.1099/vir.0.027490-0>
- Strassl, R., Schiemann, M., Doberer, K., Görzer, I., Puchhammer-Stöckl, E., Eskandary, F., ... Bond, G. (2018). Quantification of torque teno virus

- viremia as a prospective biomarker for infectious disease in kidney allograft recipients. *Journal of Infectious Diseases*, 218(8), 1191–1199. <https://doi.org/10.1093/infdis/jiy306>
- Stremlau, M. H., Andersen, K. G., Folarin, O. A., Grove, J. N., Odia, I., Ehiane, P. E., ... Happi, C. T. (2015). Discovery of Novel Rhabdoviruses in the Blood of Healthy Individuals from West Africa. *PLoS Neglected Tropical Diseases*, 9(3). <https://doi.org/10.1371/journal.pntd.0003631>
- Strimmer, K., & von Haeseler, A. (1997). Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences*, 94(13), 6815–6819. <https://doi.org/10.1073/pnas.94.13.6815>
- Strimmer, Korbinian, & Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1487), 137–142. <https://doi.org/10.1098/rspb.2001.1862>
- Subramanian, B., Gao, S., Lercher, M. J., Hu, S., & Chen, W. H. (2019). Evolvew v3: A webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Research*, 47, W270–W275. <https://doi.org/10.1093/nar/gkz357>
- Suttle, C. A. (2005). Viruses in the sea. *Nature*, 437(7057), 356–361. <https://doi.org/10.1038/nature04160>
- Thom, K., Morrison, C., Lewis, J. C. M., & Simmonds, P. (2003). Distribution of TT virus (TTV), TTV-like minivirus, and related viruses in humans and nonhuman primates. *Virology*, 306(2), 324–333. [https://doi.org/10.1016/S0042-6822\(02\)00049-1](https://doi.org/10.1016/S0042-6822(02)00049-1)
- Thom, K., & Petrik, J. (2007). Progression towards AIDS leads to increased torque teno virus and torque teno minivirus titers in tissues of HIV infected individuals. *Journal of Medical Virology*, 79(1), 1–7. <https://doi.org/https://doi.org/10.1002/jmv.20756>
- Tyschik, E. A., Rasskazova, A. S., Degtyareva, A. V., Rebrikov, D. V., & Sukhikh, G. T. (2018). Torque teno virus dynamics during the first year of life. *Virology Journal*, 15(1), 1–4. <https://doi.org/10.1186/s12985-018-1007-6>
- Varsani, A., Opriessnig, T., Celer, V., Maggi, F., Okamoto, H., Blomström, A.-L., ... Kraberger, S. (2021). Taxonomic update for mammalian anelloviruses (family Anelloviridae). *Archives of Virology*, 166(10), 2943–2953. <https://doi.org/10.1007/s00705-021-05192-x>
- Virgin, H. W. (2014). The virome in mammalian physiology and disease. *Cell*, 157(1), 142–150. <https://doi.org/10.1016/j.cell.2014.02.032>
- Virgin, H. W., Wherry, E. J., & Ahmed, R. (2009). Redefining Chronic Viral

- Infection. *Cell*, 138(1), 30–50. <https://doi.org/10.1016/j.cell.2009.06.036>
- Wan, Z., Liu, J., Hu, F., Shui, J., Li, L., Wang, H., ... Tang, S. (2020). Evidence that the second human pegivirus (HPgV-2) is primarily a lymphotropic virus and can replicate independent of HCV replication. *Emerging Microbes and Infections*, 9(1), 485–495. <https://doi.org/10.1080/22221751.2020.1730247>
- Worobey, M., & Holmes, E. C. (2001). Homologous recombination in GB virus C/hepatitis G virus. *Molecular Biology and Evolution*, 18(2), 254–261. <https://doi.org/10.1093/oxfordjournals.molbev.a003799>
- Wu, H., Padhi, A., Xu, J., Gong, X., & Tien, P. (2016). Evidence for within-host genetic recombination among the human pegiviral strains in HIV infected subjects. *PLoS ONE*, 11(8), 1–13. <https://doi.org/10.1371/journal.pone.0161880>
- Xu, B., Zhi, N., Hu, G., Wan, Z., Zheng, X., Liu, X., ... Young, N. S. (2013). Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing. *Proceedings of the National Academy of Sciences*, 110(25), 10264–10269. <https://doi.org/10.1073/pnas.1303744110>
- Yu, Y., Wan, Z., Wang, J.-H., Yang, X., & Zhang, C. (2022). Review of human pegivirus: Prevalence, transmission, pathogenesis, and clinical implication. *Virulence*, 13(1), 323–340. <https://doi.org/10.1080/21505594.2022.2029328>
- Zárate, S., Taboada, B., Yocupicio-Monroy, M., & Arias, C. F. (2017). Human Virome. *Archives of Medical Research*, 48(8), 701–716. <https://doi.org/10.1016/j.arcmed.2018.01.005>
- Zhang, T., Breitbart, M., Lee, W. H., Run, J.-Q., Wei, C. L., Soh, S. W. L., ... Ruan, Y. (2005). RNA Viral Community in Human Feces: Prevalence of Plant Pathogenic Viruses. *PLoS Biology*, 4(1), e3. <https://doi.org/10.1371/journal.pbio.0040003>
- Zhang, W., Li, L., Deng, X., Blümel, J., Nübling, C. M., Hunfeld, A., ... Delwart, E. (2016). Viral nucleic acids in human plasma pools. *Transfusion*, 56(9), 2248–2255. <https://doi.org/10.1111/trf.13692>
- Zhang, Y., Li, F., Shan, T. L., Deng, X., Delwart, E., & Feng, X. P. (2016). A novel species of torque teno mini virus (TTMV) in gingival tissue from chronic periodontitis patients. *Scientific Reports*, 6, 1–7. <https://doi.org/10.1038/srep26739>
- Zhang, Y. Z., Shi, M., & Holmes, E. C. (2018). Using Metagenomics to Characterize an Expanding Virosphere. *Cell*, 172(6), 1168–1172. <https://doi.org/10.1016/j.cell.2018.02.043>

Anexo

Anexo

Los documentos correspondientes al material suplementario se encuentran disponibles online en <https://github.com/mariacebria/tesis>

Tabla Suplementaria S1. Lista de secuencias/*contigs* de anelovirus detectados en el capítulo 1 mediante el análisis por SPAdes. Para cada secuencia, se indica la longitud del *contig* (en nucleótidos), la profundidad de cobertura media, el tamaño de la ORF1 putativa (en aminoácidos) y el número de acceso en GenBank. En la última columna se indica el género obtenido tras los análisis filogenéticos y en el análisis por BLAST. **Contigs* que muestran redundancia terminal y, en consecuencia, se consideran genomas completos. ***Contigs* que producen ORF1 incompletas, ya que la secuencia nucleotídica se interrumpe antes de llegar al codón de inicio/parada.

Muestra	Nombre secuencia	Tamaño del genoma (nt)	Tamaño ORF (aa)	Media de la profundidad de cobertura	Número de acceso	Género
P1	P1-c1	3687	765	240,31	MW455345	TTV
	P1-c2	3569	745	291,58	MW455346	TTV
	P1-c3	3487	757	34,01	MW455347	TTV
	P1-c4	3243	772	20,21	MW455348	TTV
	P1-c5	3154	735	96,13	MW455349	TTV
	P1-c6	3120	756	134,93	MW455350	TTV
	P1-c7	3108	775	23,52	MW455351	TTV
	P1-c8	3091	768	42,72	MW455352	TTV
	P1-c10	2987	752	10,24	MW455353	TTV
	P1-c12*	2894	675	371,79	MW455354	TTMV
	P1-c13	2955	762	27,23	MW455355	TTV
	P1-c14	2947	732	43,79	MW455356	TTV
	P1-c15*	2859	650	3300,13	MW455357	TTMV
	P1-c16*	2853	654	943,28	MW455358	TTMV
	P1-c17	2926	738	34,07	MW455359	TTV
	P1-c18	2887	739	22,51	MW455360	TTV
	P1-c19	2555	682	15,88	MW455361	TTMDV
	P1-c20**	2289	635	47,06	MW455451	TTV
	P1-c21**	2257	732	4,43	MW455452	TTV
	P1-c22**	2051	596	319,28	MW455453	TTMV
	P1-c23	3092	652	506,75	MW455362	TTMDV
	P1-c24*	2912	660	140,63	MW455363	TTMV
	P1-c25*	2897	656	297,99	MW455364	TTMV
P1-c26	2576	661	55,84	MW455365	TTMDV	
P1-c27	2572	659	2763,91	MW455366	TTMDV	
P2	P2-c2	3175	736	106,02	MW455367	TTV
	P2-c3	3169	746	78,74	MW455368	TTV
	P2-c4	3074	740	25,22	MW455369	TTV
	P2-c5	3051	739	54,75	MW455370	TTV
	P2-c6	3001	767	30,49	MW455371	TTV
	P2-c7	2904	732	70,00	MW455372	TTV
	P2-c8	2567	669	105,03	MW455373	TTMDV
	P2-c9**	2379	670	3,35	MW455454	TTV
	P2-c10**	2341	677	6,08	MW455455	TTV
	P3	P3-c1	3160	743	23,12	MW455374

Anexo

	P3-c2	3125	731	63,89	MW455375	TTV
	P3-c3	3022	771	8,99	MW455376	TTV
	P3-c4*	2885	652	21,63	MW455377	TTMV
	P3-c5	2796	665	82,01	MW455378	TTMDV
	P3-c6	3309	746	310,91	MW455379	TTV
	P3-c7	3215	756	37,26	MW455380	TTV
	P3-c8*	2989	673	1540,48	MW455381	TTMV
	P3-c9*	2810	653	256,65	MW455382	TTMV
P4	P4-c1	3709	739	75,45	MW455383	TTV
	P4-c2	2968	741	75,44	MW455384	TTV
	P4-c3	2865	732	8,04	MW455385	TTV
	P4-c4*	2777	634	57,60	MW455386	TTMV
P5	P5-c1	3235	762	92,41	MW455387	TTV
	P5-c2	2241	676	4,87	MW455388	TTMV
P6	P6-c1	3826	781	153,44	MW455389	TTV
	P6-c2	3135	786	19,98	MW455390	TTV
	P6-c3*	2839	659	378,42	MW455391	TTMV
P7	P7-c1	3328	755	72,00	MW455392	TTV
	P7-c2	3280	758	829,08	MW455393	TTV
	P7-c3	3268	676	322,59	MW455394	TTMDV
	P7-c4	3243	769	121,33	MW455395	TTV
	P7-c5	3081	767	100,97	MW455396	TTV
	P7-c6	3065	755	73,86	MW455397	TTV
	P7-c7	3042	625	1148,61	MW455398	TTMV
	P7-c8	2987	761	70,75	MW455399	TTV
	P7-c9	2958	743	455,31	MW455400	TTV
	P7-c10	2950	745	130,67	MW455401	TTV
	P7-c12*	2854	667	331,47	MW455402	TTMV
	P7-c14	2785	680	7,95	MW455403	TTMV
	P7-c16	2711	732	66,43	MW455404	TTV
	P7-c17	2606	661	321,71	MW455405	TTMDV
	P7-c18	2557	661	232,98	MW455406	TTMDV
	P7-c19	2550	662	186,35	MW455407	TTMDV
	P7-c20*	2098	653	12,34	MW455408	TTMDV
	P7-c21*	2866	653	485,29	MW455409	TTMV
	P7-c22	2915	665	1747,01	MW455410	TTMDV
	P7-c23*	2837	659	2290,58	MW455411	TTMV
	P7-c24	2634	659	392,27	MW455412	TTMDV
	P7-c25	2565	666	119,00	MW455413	TTMDV
	P8	P8-c1	3215	756	166,43	MW455414
P8-c2		3201	763	73,03	MW455415	TTV
P8-c3		3181	767	42,27	MW455416	TTV
P8-c4		3134	751	6,39	MW455417	TTV
P8-c5		3075	782	112,88	MW455418	TTV
P8-c6*		2988	668	32,44	MW455419	TTMV
P8-c8		3013	780	51,63	MW455420	TTV
P8-c9		3000	743	75,81	MW455421	TTV
P8-c10		2988	754	100,16	MW455422	TTV
P8-c13		2927	739	53,25	MW455423	TTV
P8-c16*		2810	653	139,11	MW455424	TTMV
P8-c17*		2810	642	459,02	MW455425	TTMV
P8-c18		2546	667	29,36	MW455426	TTMDV
P8-c20**		2252	613	277,18	MW455456	TTV
P8-c21**		2146	603	30,35	MW455457	TTMV
P8-c22*		2947	666	468,97	MW455427	TTMV
P8-c23*		2860	548	139,78	MW455428	TTMV
P8-c24*		2851	663	35,52	MW455429	TTMV
P8-c25*		2848	659	383,54	MW455430	TTMV
P8-c26*		2824	646	127,71	MW455431	TTMV
P9	P9-c1	3763	766	114,03	MW455432	TTV

	P9-c2	3130	733	12,94	MW455433	TTV
	P9-c3	3057	660	199,41	MW455434	TTMV
	P9-c4	2996	756	38,10	MW455435	TTV
	P9-c5	2475	659	19,67	MW455436	TTMDV
	P9-c6	2881	765	27,85	MW455437	TTV
P10	P10-c2	3717	739	155,84	MW455438	TTV
	P10-c3	3655	733	176,08	MW455439	TTV
	P10-c4	3195	752	37,47	MW455440	TTV
	P10-c5	3191	766	66,03	MW455441	TTV
	P10-c6	2961	743	29,93	MW455442	TTV
	P10-c7	2926	731	41,23	MW455443	TTV
	P10-c8*	2779	642	127,42	MW455444	TTMV
	P10-c9	2874	673	36,16	MW455445	TTMV
	P11	P11-c1	3160	642	51,72	MW455446
P11-c2		3088	786	16,31	MW455447	TTV
P12	P12-c1	3741	739	72,28	MW455448	TTV
	P12-c2	3007	757	39,57	MW455449	TTV
	P12-c3	2568	666	13,63	MW455450	TTMDV
	P12-c4**	1806	518	41,76	MW455458	TTV

Tabla Suplementaria S2. Lista de secuencias/*contigs* de anelovirus detectados en el capítulo 2 mediante el análisis metaSPAdes. Para cada secuencia, los nombres en la primera y segunda columna se refieren a la mezcla (SP) y al nombre del *contig* (C), respectivamente. Se indican la longitud del *contig* (en nucleótidos), la profundidad de cobertura media, el tamaño deducido de la ORF1 putativa (en aminoácidos) y el número de acceso en GenBank. En la última columna se indica el género obtenido tras los análisis filogenéticos y en el análisis por BLAST. **Contigs* que muestran redundancia terminal y, en consecuencia, se consideran genomas completos. ***Contigs* que producen ORF1 incompletas, ya que la secuencia nucleotídica se interrumpe antes de llegar al codón de inicio/parada.

Muestra	Nombre secuencia	Tamaño del genoma (nt)	Tamaño ORF (aa)	Media de la profundidad de cobertura	Número de acceso	Género
SP1	SP1_C1	3223	659	1404,01	MZ285962	TTMDV
	SP1_C2*	2988	668	57636,33	MZ286226	TTMV
	SP1_C3	2961	666	326,79	MZ285963	TTMDV
	SP1_C4	2930	739	461,20	MZ285964	TTV
	SP1_C5	2922	765	10,24	MZ285965	TTV
	SP1_C6*	2837	423	4666,41	MZ286227	TTMV
	SP1_C7**	2424	514	497,78	MZ285966	TTV
	SP1_C8**	1768	513	28,30	MZ285967	TTV
SP2	SP2_C1	3809	761	111786,54	MZ285968	TTV
	SP2_C2	3668	641	5032,80	MZ285969	TTV
	SP2_C3	3408	749	577,90	MZ285970	TTV
	SP2_C4	2988	768	25,65	MZ285971	TTV
	SP2_C5	2971	745	14,86	MZ285972	TTV

Anexo

	SP3_C1	3195	719	882,86	MZ285973	TTV
	SP3_C2	3087	780	557,20	MZ285974	TTV
	SP3_C3	3070	756	2507,19	MZ285975	TTV
	SP3_C4	3012	766	415,67	MZ285976	TTV
	SP3_C5	3004	773	129,71	MZ285977	TTV
	SP3_C6	2976	750	35,57	MZ285978	TTV
	SP3_C7	2972	745	394,64	MZ285979	TTV
	SP3_C8	2968	767	288,15	MZ285980	TTV
SP3	SP3_C9	2961	769	188,65	MZ285981	TTV
	SP3_C10	2943	771	237,14	MZ285982	TTV
	SP3_C11	2904	642	122,48	MZ285983	TTV
	SP3_C12	2872	762	1366,74	MZ285984	TTV
	SP3_C13	2867	760	3571,83	MZ285985	TTV
	SP3_C14*	2808	664	845,04	MZ286228	TTMV
	SP3_C15	2842	663	169,80	MZ285986	TTMDV
	SP3_C16	2759	759	6,35	MZ285987	TTV
	SP3_C17	2542	732	14,93	MZ285988	TTV
	SP3_C18	3033	645	1302,16	MZ285989	TTV
SP4	SP4_C1*	2879	659	3249,84	MZ286229	TTMV
	SP4_C2*	2848	661	1723,93	MZ286230	TTMV
SP6	SP6_C1**	3607	508	12,46	MZ285990	TTV
	SP6_C2*	2925	678	4300,08	MZ286231	TTMV
	SP6_C4*	2810	653	66200,71	MZ286232	TTMV
	SP6_C5	2817	669	99,23	MZ285991	TTMV
	SP6_C6	3585	767	17,25	MZ285992	TTV
	SP6_C7*	2937	668	13199,22	MZ286233	TTMV
SP7	SP7_C1	3603	773	105,35	MZ285993	TTV
	SP7_C2	3167	644	93,61	MZ285994	TTV
	SP7_C3	3006	767	31,82	MZ285995	TTV
	SP7_C4*	2876	655	2468,99	MZ286234	TTMV
	SP7_C5*	2810	642	4733,40	MZ286235	TTMV
	SP7_C6	2751	507	103,53	MZ285996	TTMDV
	SP7_C7**	1849	549	12,61	MZ285997	TTMV
	SP7_C8**	1760	520	26,22	MZ285998	TTMDV
	SP7_C10	2475	662	39,79	MZ285999	TTMDV
SP8	SP8_C1	3185	660	10486,68	MZ286000	TTMDV
	SP8_C2*	2839	653	3603,48	MZ286236	TTMV
	SP8_C3	2779	667	72,46	MZ286001	TTMDV
SP9	SP9_C1*	2876	655	20188,03	MZ286237	TTMV
SP10	SP10_C1	3643	641	308,14	MZ286002	TTV
	SP10_C2	3175	644	150,22	MZ286003	TTV
	SP10_C3	2971	661	14257,13	MZ286004	TTMV
	SP10_C4	2786	661	441,50	MZ286005	TTMDV
SP11	SP11_C1	3649	735	73,43	MZ286006	TTV
	SP11_C2	2817	643	1571,38	MZ286007	TTMV
SP12	SP12_C1	3566	648	848,34	MZ286008	TTV
	SP12_C2	3515	698	5664,27	MZ286009	TTV
SP13	SP13_C1	3229	781	576,80	MZ286010	TTV
	SP13_C2*	3268	662	24083,26	MZ286238	TTMDV
	SP13_C3	3165	758	6448,77	MZ286011	TTV
	SP13_C4	3006	766	17,90	MZ286012	TTV
SP14	SP14_C1	3288	765	2181,54	MZ286013	TTV
	SP14_C2	3029	765	19130,38	MZ286014	TTV
	SP14_C3*	2906	667	9893,72	MZ286239	TTMV
	SP14_C4*	2862	652	4351,38	MZ286240	TTMV
	SP14_C5**	1531	509	50,42	MZ286015	TTV
SP15	SP15_C1	3811	766	4852,22	MZ286016	TTV
	SP15_C2	3806	749	4958,55	MZ286017	TTV
	SP15_C3	3145	759	395,66	MZ286018	TTV
	SP15_C4*	2948	668	17764,23	MZ286241	TTMV

	SP15_C5	2947	743	72,34	MZ286019	TTV
	SP15_C6	2776	639	788,89	MZ286020	TTMV
	SP15_C7	2514	643	1483,92	MZ286021	TTMV
	SP16_C1*	2859	671	4200,25	MZ286242	TTMV
	SP16_C2	3918	741	6798,99	MZ286022	TTV
	SP16_C3	3443	738	35,10	MZ286023	TTV
	SP16_C4	3230	658	201,25	MZ286024	TTMDV
	SP16_C5	3174	743	747,08	MZ286025	TTV
	SP16_C6	3157	661	3879,41	MZ286026	TTMDV
	SP16_C7	3134	745	1446,57	MZ286027	TTV
	SP16_C8	3026	773	237,92	MZ286028	TTV
	SP16_C9	3007	769	133,29	MZ286029	TTV
SP16	SP16_C10*	2890	653	14019,40	MZ286243	TTMV
	SP16_C11*	2879	656	608,66	MZ286244	TTMV
	SP16_C12	2928	735	206,85	MZ286030	TTV
	SP16_C13	2917	767	14,73	MZ286031	TTV
	SP16_C14	2871	668	313,03	MZ286032	TTMDV
	SP16_C15*	2871	647	2750,23	MZ286245	TTMV
	SP16_C16	2830	664	382,00	MZ286033	TTMV
	SP16_C17	2623	655	206,41	MZ286034	TTMV
	SP16_C18	2351	670	13,82	MZ286035	TTV
	SP16_C19	2127	647	5,31	MZ286036	TTMV
	SP18_C1	3816	771	167,45	MZ286037	TTV
SP18	SP18_C2	3732	749	1107,90	MZ286038	TTV
	SP18_C3**	2155	578	5,34	MZ286039	TTV
	SP18_C4**	2072	613	3405,47	MZ286040	TTMDV
SP19	SP19_C1*	2961	681	124699,42	MZ286246	TTMV
	SP19_C2	2837	658	1067,60	MZ286041	TTMDV
	SP20_C1	3370	767	231,25	MZ286042	TTV
SP20	SP20_C2	3184	764	460,63	MZ286043	TTV
	SP20_C3	2893	664	2226,38	MZ286044	TTMV
	SP20_C4	2889	645	26,91	MZ286045	TTV
	SP20_C5*	2713	658	273,98	MZ286247	TTMV
	SP21_C1	3507	742	123,02	MZ286046	TTV
SP21	SP21_C2	2988	759	320,02	MZ286047	TTV
	SP21_C3	2890	648	109216,26	MZ286048	TTMV
	SP22_C1	3501	765	22,29	MZ286049	TTV
	SP22_C2	3241	667	6933,94	MZ286050	TTMDV
SP22	SP22_C3	3015	767	16,45	MZ286051	TTV
	SP22_C4*	2825	688	2145,83	MZ286248	TTMV
	SP22_C5	2858	658	1080,10	MZ286052	TTMV
	SP22_C6	2818	643	4027,17	MZ286053	TTMV
	SP22_C7**	2673	759	8,90	MZ286054	TTV
	SP22_C8	2471	667	73,50	MZ286055	TTMDV
	SP23_C1	3866	773	2197,07	MZ286056	TTV
SP23	SP23_C2	3515	762	60,92	MZ286057	TTV
	SP23_C3*	2860	664	5323,36	MZ286249	TTMV
	SP23_C4	2879	636	5605,43	MZ286058	TTMV
	SP24_C1	3766	644	266,18	MZ286059	TTV
	SP24_C2	3537	756	1461,93	MZ286060	TTV
SP24	SP24_C3	3445	735	655,25	MZ286061	TTV
	SP24_C4	3034	726	88,28	MZ286062	TTV
	SP24_C5	3006	782	229,97	MZ286063	TTV
	SP24_C6	2954	748	7,36	MZ286064	TTV
	SP24_C7*	2852	663	6386,53	MZ286250	TTMV
	SP24_C8	2680	653	12,37	MZ286065	TTMV
	SP25_C1	3750	773	114,53	MZ286066	TTV
SP25	SP25_C2	2644	670	1353,52	MZ286067	TTMDV
	SP25_C3	2640	664	333,58	MZ286068	TTMDV
	SP25_C4	2501	658	158,87	MZ286069	TTMDV

Anexo

SP26	SP26_C1	3872	766	2176,36	MZ286070	TTV
	SP26_C2	2796	647	1289,66	MZ286071	TTMDV
	SP26_C3	2634	668	22,17	MZ286072	TTMV
SP27	SP27_C1	3400	786	185,02	MZ286073	TTV
	SP27_C2*	2988	668	438,45	MZ286251	TTMV
	SP27_C3	3026	768	145,83	MZ286074	TTV
	SP27_C4	3015	780	193,07	MZ286075	TTV
	SP27_C5	2957	759	37,24	MZ286076	TTV
	SP27_C6	2936	766	10,62	MZ286077	TTV
	SP27_C7	2878	742	106,60	MZ286078	TTV
SP28	SP28_C1*	2832	662	8111,81	MZ286252	TTMV
	SP28_C2	3550	556	27,42	MZ286079	TTV
	SP28_C3	3483	765	27,95	MZ286080	TTV
	SP28_C4*	2920	680	1365,27	MZ286253	TTMV
	SP28_C5	2544	671	108,27	MZ286081	TTMDV
	SP28_C6	2543	659	8352,31	MZ286082	TTMV
	SP28_C7	2534	670	281,98	MZ286083	TTMDV
SP29	SP29_C1	2803	668	144,61	MZ286084	TTMV
	SP29_C2	1802	540	53,58	MZ286085	TTMDV
SP30	SP30_C1	3770	772	1042,27	MZ286086	TTV
	SP30_C2*	2978	673	48406,19	MZ286254	TTMV
	SP30_C3	2953	660	1042,45	MZ286087	TTMDV
	SP30_C4*	2832	657	433,06	MZ286255	TTMV
	SP30_C5	2871	659	286,66	MZ286088	TTMV
	SP30_C6*	2811	642	484,99	MZ286256	TTMV
	SP30_C7*	2762	628	565,14	MZ286257	TTMV
	SP30_C8	2582	670	68,01	MZ286089	TTMDV
	SP30_C9	2550	666	259,84	MZ286090	TTMDV
	SP30_C10	2460	662	49,97	MZ286091	TTMDV
SP31	SP31_C1	3733	621	577,26	MZ286092	TTV
	SP31_C2	3690	660	4647,80	MZ286093	TTV
	SP31_C3	3122	732	694,21	MZ286094	TTV
	SP31_C4*	2952	675	25284,50	MZ286258	TTMV
	SP31_C5	2842	663	18027,70	MZ286095	TTMDV
	SP31_C6	2818	661	181,43	MZ286096	TTMDV
SP32	SP32_C1	3823	780	79,15	MZ286097	TTV
SP33	SP33_C1	3686	644	2830,19	MZ286098	TTV
	SP33_C2	3574	767	2390,81	MZ286099	TTV
	SP33_C3	3546	753	219,24	MZ286100	TTV
	SP33_C4	3172	744	214,99	MZ286101	TTV
	SP33_C5	3136	789	625,14	MZ286102	TTV
	SP33_C6	2157	511	44,18	MZ286103	TTV
	SP33_C7	2688	763	4,55	MZ286104	TTV
SP34	SP34_C1*	2821	658	46784,26	MZ286259	TTMV
	SP35_C1	2983	768	13,85	MZ286105	TTV
	SP35_C2*	2928	677	17165,56	MZ286260	TTMV
	SP35_C3	2976	735	1081,14	MZ286106	TTV
	SP35_C4*	2912	660	16452,07	MZ286261	TTMV
	SP35_C5	2933	731	9724,29	MZ286107	TTV
	SP35_C6*	2849	652	43126,03	MZ286262	TTMV
	SP35_C8**	2058	568	1245,28	MZ286108	TTMV
	SP35_C9*	3667	644	2931,01	MZ286109	TTV
SP36	SP36_C1	3716	720	118,15	MZ286110	TTV
	SP36_C2*	2878	657	14977,41	MZ286263	TTMV
	SP36_C3*	2851	663	6745,50	MZ286264	TTMV
	SP36_C4	2889	667	31,62	MZ286111	TTMV
	SP36_C5	2860	665	39799,43	MZ286112	TTMDV
SP37	SP37_C1	3818	747	3813,24	MZ286113	TTV
	SP37_C2	3187	782	1892,73	MZ286114	TTV
	SP37_C3	3113	756	685,70	MZ286115	TTV

	SP37_C4	2961	654	25799,63	MZ286116	TTMDV
	SP37_C5*	2848	656	17031,36	MZ286265	TTMV
	SP37_C6*	2802	645	45042,49	MZ286266	TTMV
	SP37_C7	2592	667	1232,15	MZ286117	TTMDV
SP38	SP38_C1	3237	780	421,45	MZ286118	TTV
	SP38_C2*	2823	659	35768,98	MZ286267	TTMV
	SP38_C3	2617	693	20,00	MZ286119	TTV
	SP38_C4**	1778	533	33,78	MZ286120	TTV
SP39	SP39_C1	3630	773	1268,27	MZ286121	TTV
	SP39_C2	3503	644	20,12	MZ286122	TTV
	SP39_C3	3103	782	274,99	MZ286123	TTV
SP40	SP40_C1	3513	751	16645,54	MZ286124	TTV
	SP40_C2	3171	767	335,37	MZ286125	TTV
	SP40_C3	3169	734	1684,56	MZ286126	TTV
	SP40_C4	2993	655	103703,61	MZ286127	TTMV
SP41	SP41_C1	3623	786	1409,47	MZ286128	TTV
	SP41_C2*	2965	665	10182,35	MZ286268	TTMV
	SP41_C3	3013	739	118,27	MZ286129	TTV
	SP41_C4*	2888	653	27901,24	MZ286269	TTMV
	SP41_C5*	2781	647	24788,76	MZ286270	TTMV
SP42	SP42_C1	3193	750	26,96	MZ286130	TTV
	SP42_C2*	2855	675	4401,99	MZ286271	TTMV
	SP42_C3	2545	732	8,90	MZ286131	TTV
SP43	SP43_C1	3719	773	122,66	MZ286132	TTV
	SP43_C2	2969	679	569,14	MZ286133	TTMV
	SP43_C3	2923	672	1109,29	MZ286134	TTMV
	SP43_C4*	2849	664	37268,94	MZ286272	TTMV
	SP43_C5*	2834	658	7485,03	MZ286273	TTMV
	SP43_C6*	2840	662	2375,47	MZ286274	TTMV
	SP43_C7	2854	755	21,35	MZ286135	TTV
	SP43_C8*	2762	628	2638,79	MZ286275	TTMV
	SP43_C9	2652	671	991,04	MZ286136	TTMV
	SP43_C10	2575	667	2197,62	MZ286137	TTMDV
	SP43_C11	2557	664	723,04	MZ286138	TTMDV
	SP43_C12	2538	666	483,32	MZ286139	TTMDV
SP44	SP44_C1	3787	739	283,01	MZ286140	TTV
	SP44_C2	3479	733	1199,62	MZ286141	TTV
	SP44_C3	2936	744	55,59	MZ286142	TTV
SP45	SP45_C1	3605	702	2865,62	MZ286143	TTV
	SP45_C2	3146	762	906,52	MZ286144	TTV
	SP45_C3	2978	765	20,49	MZ286145	TTV
	SP45_C4	2873	756	432,81	MZ286146	TTV
	SP45_C5	2822	662	29153,25	MZ286147	TTMDV
	SP45_C6	2816	642	695,10	MZ286148	TTMV
	SP45_C7	2645	659	5136,11	MZ286149	TTMDV
	SP45_C8	2517	661	3213,26	MZ286150	TTMDV
	SP45_C9	2895	739	5,01	MZ286151	TTV
	SP45_C11*	2876	655	4731,90	MZ286276	TTMV
	SP45_C12*	2818	642	712,16	MZ286277	TTMV
	SP46	SP46_C1	3705	756	843,23	MZ286152
SP46_C2*		2908	675	17198,90	MZ286278	TTMV
SP46_C3		2963	658	2622,96	MZ286153	TTMDV
SP46_C4		2600	662	925,58	MZ286154	TTMDV
SP46_C5**		1866	508	9,59	MZ286155	TTMV
SP47	SP47_C1	3996	739	1980,00	MZ286156	TTV
	SP47_C2	3428	760	2658,26	MZ286157	TTV
	SP47_C3	3353	733	964,86	MZ286158	TTV
	SP47_C4	3226	667	7582,78	MZ286159	TTMDV
	SP47_C5	3076	767	399,69	MZ286160	TTV
	SP47_C6	3072	773	119,61	MZ286161	TTV

Anexo

	SP47_C7	3013	780	64,40	MZ286162	TTV
	SP47_C8	2906	663	4461,35	MZ286163	TTMV
	SP47_C9	2517	663	107,42	MZ286164	TTMDV
	SP47_C10	2510	672	56,57	MZ286165	TTMDV
	SP47_C11**	2462	539	94,55	MZ286166	TTV
	SP47_C12**	2461	542	579,45	MZ286167	TTV
	SP47_C13	2923	756	6,94	MZ286168	TTV
SP48	SP48_C1	3545	765	256,12	MZ286169	TTV
	SP48_C2	3367	742	12,22	MZ286170	TTV
	SP48_C3*	2962	663	246,47	MZ286279	TTMV
	SP48_C4	2938	678	347,71	MZ286171	TTMV
	SP48_C5	2921	662	4484,32	MZ286172	TTMDV
	SP48_C6	2783	671	192,10	MZ286173	TTMDV
	SP48_C7**	2755	618	32708,82	MZ286174	TTMV
	SP48_C8	2539	670	423,30	MZ286175	TTMDV
SP49	SP49_C1	3275	655	14389,18	MZ286176	TTMDV
	SP49_C2	3181	661	35126,91	MZ286177	TTMDV
	SP49_C3	3034	767	723,28	MZ286178	TTV
	SP49_C4	3034	765	164,80	MZ286179	TTV
	SP49_C5*	2859	652	15617,88	MZ286280	TTMV
		SP49_C6	2829	639	24,12	MZ286180
SP51	SP51_C1	3738	741	210,00	MZ286181	TTV
	SP51_C2	3651	780	448,24	MZ286182	TTV
	SP51_C3	3621	661	1705,61	MZ286183	TTV
	SP51_C4	3537	752	99,41	MZ286184	TTV
	SP51_C5	3113	648	26,06	MZ286185	TTV
	SP51_C6*	2963	674	11701,43	MZ286281	TTMV
	SP51_C7*	2958	667	5539,36	MZ286282	TTMV
	SP51_C8*	2837	659	44151,18	MZ286283	TTMV
	SP51_C9*	2907	675	2239,29	MZ286284	TTMV
	SP51_C10	2946	645	55,73	MZ286186	TTV
	SP51_C11	2942	680	3452,89	MZ286187	TTMV
	SP51_C12	2932	739	23,17	MZ286188	TTV
	SP51_C13	2905	676	11111,00	MZ286189	TTMDV
	SP51_C14	2880	653	6491,73	MZ286190	TTMV
	SP51_C15	2850	666	3852,73	MZ286191	TTMDV
	SP51_C16	2820	663	1654,50	MZ286192	TTMV
	SP51_C17	2775	645	960,81	MZ286193	TTMV
	SP51_C18	2675	671	596,39	MZ286194	TTMV
	SP51_C19	2596	670	357,82	MZ286195	TTMV
	SP51_C20	2581	663	522,33	MZ286196	TTMV
	SP51_C21	2554	664	251,83	MZ286197	TTMDV
	SP51_C22	2546	662	661,04	MZ286198	TTMDV
	SP51_C23	2506	655	2011,83	MZ286199	TTMDV
	SP51_C24	2492	634	29,65	MZ286200	TTMV
	SP51_C25	2463	660	45,97	MZ286201	TTMDV
	SP51_C26	2463	670	24,54	MZ286202	TTMDV
	SP51_C27	2443	662	16,32	MZ286203	TTMDV
		SP51_C28**	2367	503	1283,05	MZ286204
SP52	SP52_C1	3778	739	3782,62	MZ286205	TTV
	SP52_C2	3552	760	349,39	MZ286206	TTV
	SP52_C3*	2963	681	12308,94	MZ286285	TTMV
SP53	SP53_C1	3016	660	30801,45	MZ286207	TTMV
SP54	SP54_C1	3762	767	187,90	MZ286208	TTV
SP55	SP55_C1	3813	769	65,97	MZ286209	TTV
	SP55_C2	2052	601	10,07	MZ286210	TTMDV
SP56	SP56_C1	3109	764	13,03	MZ286211	TTV
	SP56_C2	2900	667	184,66	MZ286212	TTMV
	SP56_C3*	2837	663	8941,90	MZ286286	TTMV
	SP56_C4*	2778	644	1558,48	MZ286287	TTMV

	SP56_C5**	1947	598	8,55	MZ286213	TTV
	SP56_C6*	2898	668	1236,75	MZ286288	TTMV
	SP56_C7	2831	661	1798,55	MZ286214	TTMDV
SP57	SP57_C1	3809	779	95,68	MZ286215	TTV
	SP57_C2*	2847	662	5216,81	MZ286289	TTMV
SP58	SP58_C1	3703	765	10,58	MZ286216	TTV
	SP58_C2*	2834	658	591,29	MZ286290	TTMV
	SP58_C3	2818	661	2304,33	MZ286217	TTMDV
SP59	SP59_C1	3161	756	17,54	MZ286218	TTV
	SP59_C2*	2824	665	339,40	MZ286291	TTMV
	SP59_C3	2861	653	154,59	MZ286219	TTMV
SP60	SP60_C1	3532	765	26,70	MZ286220	TTV
	SP60_C2	3160	763	22,33	MZ286221	TTV
	SP60_C3	2959	735	28,54	MZ286222	TTV
	SP60_C4*	2889	662	6310,69	MZ286292	TTMV
	SP60_C5*	2834	649	7103,48	MZ286293	TTMV
	SP60_C6	2861	670	2349,91	MZ286223	TTMDV
	SP60_C7	2849	657	158,19	MZ286224	TTMV
	SP60_C8	2772	627	73,62	MZ286225	TTMV

Tabla Suplementaria S3. Lista de aislados de HPgV descargados del GenBank y descritos en el capítulo 2. Se indica el número de acceso, el nombre del aislado y el origen. *Aislados descritos en el capítulo 2.

Número de acceso	Aislado	Origen	Número de acceso	Aislado	Origen
D90600	T110	Japón	KP710601	280-10	Camerún
AF121950	Iowan	EEUU	KP710603	920-49	Camerún
U44402	PNF2161	EEUU	KP710606	8013815	Sudáfrica
D87262	GS185	Japón	KP710599	740-14	Camerún
D87709	K1737	Japón	KP710604	833-62	Camerún
D87710	K1741	Japón	KP710600	62-11	Camerún
D87711	K1789	Japón	MN215902	P25	Brasil
D87712	K1916	Japón	AB013500	HGV-GA128	Ghana
D87714	K1668	Japón	AB021287	HGV-MY14	Myanmar
D90601	T230	Japón	AB018667	HGV-VT48	Vietnam
AB008342	HGV-IM71	Japón	AB013501	HGV-BL230	Bolivia
LT009478	89859249	Reino Unido	U94695	HGVCN	China
LT009479	89859262	Reino Unido	MH053119	JD2B1I	Francia
LT009483	67845223	Democrática del Congo	MN215911	P55	Brasil
U36380		Este de África	MH053120	JD2B3C	Francia
D87255	HGV-Iw	Japón	MN551064	NIG-N3559	Niger
AF081782	pHGVqz	China	MN215903	P26	Brasil
MK234885	EG	Egipto	MN215898	P21	Brasil
JN127373	UU1	EEUU	MH053121	JD2B8C	Francia
U63715	GBV-C(EA)	Este de África	AY949771	D50	Sudáfrica
LT009485	56330265	Reino Unido	MH746815	GBV-C-ZI-180501	China
LT009487	89860237	Reino Unido	D87715	K1775	Japón
LT009481	56330227	Sudáfrica	MN215910	P35	Brasil
LT009484	56330260	Reino Unido	U75356	HGVC964	China
LT009480	56330218	Reino Unido	MN215907	P32	Brasil
LT009482	56330281	Australia	MN215900	P23	Brasil
LT009486	56330292	Reino Unido	MH053117	JD2B9I	Francia
LT009494	56330286	Reino Unido	U45966	R10291	EEUU
KP259281	H40048	Dinamarca	MN215897	P13	Brasil

AF031827	HGV-T55875	EEUU	MK684252	BRA-2015-17	Brasil
LT009488	56330272	República Democrática del Congo	MH053118	JD2B15C	Francia
LT009490	56330276	República Democrática del Congo	MN551063	MAU-23	Mauritania
AB008336	F1029	Japón	MN551061	BUR-19955	Burundi
KM670110	EM110	Sierra Leona	AF104403	European	Francia
KM670096	G3796	Sierra Leona	MH179063	LE-1	Canadá
KM670097	G3795	Sierra Leona	AF006500	HGV-GD (Guangdong)	Hong Kong
KM670099	G3765.2	Sierra Leona	AB003291	CG12LC	Japón
MN551062	MAL-13279	Mali	AB003289	CG01BD	Japón
KM670109	G3808	Sierra Leona	AB003290	CG07BD	Japón
MK291245	H.sapiens-wt/EEUU/GBVC-SM2	EEUU	MH053116	JD2B7I	Francia
MH053115	JD2B2C	Francia	AB003288	BG1HC	Japón
MK291244	H.sapiens-wt/EEUU/GBVC-SM3	EEUU	AB003292	G05BD	Japón
LT009489	56330229	Reino Unido	AB003293	G13HC	Japón
AY196904	765	EEUU	MW467971	P9-c1	España
AF309966	PEI	Germany	MZ420565	HPgV-SP1*	España
HQ331233	KY117	China	MZ420566	HPgV-SP3*	España
HQ331235	DH028	China	MZ420567	HPgV-SP4*	España
KM670098	G3821	Sierra Leona	MZ420568	HPgV-SP11*	España
KM670102	G3825	Sierra Leona	MZ420569	HPgV-SP13*	España
KM670106	G3819	Sierra Leona	MZ420570	HPgV-SP19*	España
KM670108	EM119	Sierra Leona	MZ420571	HPgV-SP20*	España
KM670100	G3845	Sierra Leona	MZ420572	HPgV-SP29*	España
KM670101	G3850	Sierra Leona	MZ420573	HPgV-SP30*	España
KM670107	G3826	Sierra Leona	MZ420574	HPgV-SP33*	España
KC618401	Hu_21	Uganda	MZ420575	HPgV-SP38*	España
KC618398	Hu_5	Uganda	MZ420576	HPgV-SP44*	España
KC618400	Hu_20	Uganda	MZ420577	HPgV-SP49*	España
KC618399	Hu_6	Uganda	MZ420578	HPgV-SP53-C1*	España
KP710602	469-66	Camerún	MZ420579	HPgV-SP53-C2*	España
KP710605	CHU2727	Camerún	MZ420580	HPgV-SP56*	España
D87708	K606	Japón	MZ420581	HPgV-SP58*	España
KP710598	263-26	Camerún			

Tabla Suplementaria S4: Lista de *contigs* mayores de 1,5 kb detectados con el análisis de los datos de metaSPAdes que pertenecen a la familia *Anelloviridae*. Para cada secuencia, en la tabla aparecen los siguientes datos: en la primera columna, el nombre de la muestra en la que se han descrito, en la segunda columna, el nombre del *contig*. En las columnas siguientes, aparece la longitud del *contig* en nucleótidos, el tamaño en aminoácidos de la ORF1 putativa, la media de la profundidad de cobertura que presenta la secuencia y el número de acceso en GenBank. En la última columna aparece el género al que pertenece cada *contig*, en base a los resultados obtenidos del análisis de las filogenias y del análisis por BLAST. **Contigs* que presentan redundancia terminal y que, consecuentemente, se consideran genomas completos.

****Contigs** que presentan ORF1 incompletas, debido a que la secuencia nucleotídica se interrumpe antes del codón de inicio/parada.

Muestra	Nombre secuencia	Tamaño del genoma (nt)	Tamaño ORF (aa)	Media de la profundidad de cobertura	Número de acceso	Género
SPI1	SPI1_C1	3774	768	3685,65	MZ824758	TTV
	SPI1_C2	3774	768	3599,63	MZ824759	TTV
	SPI1_C3	3774	768	3498,63	MZ824760	TTV
	SPI1_C4	3723	745	2023,56	MZ824761	TTV
	SPI1_C5	3238	641	5961,36	MZ824762	TTV
	SPI1_C6	3236	765	18981,07	MZ824763	TTV
	SPI1_C7	3205	756	18213,75	MZ824764	TTV
	SPI1_C9	3169	782	3948,51	MZ824765	TTV
	SPI1_C10	3113	731	5581,94	MZ824766	TTV
	SPI1_C12*	2949	675	9276,04	MZ825091	TTMV
	SPI1_C13*	3009	676	6630,37	MZ825092	TTMV
	SPI1_C14	3051	761	2472,31	MZ824767	TTV
	SPI1_C15*	2989	673	5396,26	MZ825093	TTMV
	SPI1_C16	3041	732	1669,48	MZ824768	TTV
	SPI1_C17	3017	744	559,50	MZ824769	TTV
	SPI1_C18*	2865	667	51451,11	MZ825094	TTMV
	SPI1_C19*	2904	668	3193,62	MZ825095	TTMV
	SPI1_C20*	2810	653	20974,72	MZ825096	TTMV
	SPI1_C21*	2810	470	13540,91	MZ825097	TTMV
	SPI1_C22*	2869	667	2491,56	MZ825098	TTMV
	SPI1_C24*	2839	651	7247,22	MZ825099	TTMV
	SPI1_C25	2894	742	908,29	MZ824770	TTV
	SPI1_C26*	2829	661	12530,89	MZ825100	TTMV
	SPI1_C28	2852	667	4292,44	MZ824771	TTMDV
	SPI1_C29	2842	662	984,29	MZ824772	TTMDV
	SPI1_C30	2839	658	13431,96	MZ824773	TTMDV
	SPI1_C31	2818	785	1024,98	MZ824774	TTV
	SPI1_C32*	2816	647	170810,86	MZ825101	TTV
SPI2	SPI2_C2	3857	773	77220,46	MZ824878	TTV
	SPI2_C3	2940	739	3348,03	MZ824879	TTV
	SPI2_C4	2841	646	109886,96	MZ824880	TTMDV
	SPI2_C5	2837	647	81834,90	MZ824881	TTMV
SPI3	SPI3_C1*	2810	653	4771,39	MZ825130	TTMV
SPI4	SPI4_C1	3607	782	35868,72	MZ824970	TTV
	SPI4_C2	3551	768	203524,64	MZ824971	TTV
	SPI4_C3	2908	735	13051,16	MZ824972	TTV
	SPI4_C4*	2841	668	5314,65	MZ825141	TTMV
SPI4_C5	2884	732	20747,34	MZ824973	TTV	
SPI5	SPI5_C1*	2782	642	53452,71	MZ825146	TTMV
SPI6	SPI6_C1	3793	763	24377,50	MZ825001	TTV
	SPI6_C2	3307	771	3352,55	MZ825002	TTV
SPI7	SPI7_C1	2857	546	125200,26	MZ825003	TTMDV
	SPI7_C2	2502	656	125200,26	MZ825004	TTMDV
	SPI7_C3	3282	720	86811,23	MZ825005	TTV
	SPI7_C4	3089	765	103582,31	MZ825006	TTV
	SPI7_C5	3042	758	8184,54	MZ825007	TTV
	SPI7_C6	2981	765	10837,62	MZ825008	TTV
	SPI7_C7	2948	673	132677,60	MZ825009	TTMV
	SPI7_C8	2941	732	122454,58	MZ825010	TTV
	SPI7_C9	2917	739	100235,15	MZ825011	TTV
	SPI7_C10	2912	749	6897,90	MZ825012	TTV
SPI8	SPI8_C1*	3860	760	52978,14	MZ825148	TTV

Anexo

	SPI8_C2*	3726	645	75539,14	MZ825149	TTV
	SPI8_C3	3708	734	54188,50	MZ825013	TTV
	SPI8_C4	3015	771	2960,60	MZ825014	TTV
	SPI8_C5	2798	572	2704,55	MZ825015	TTMDV
	SPI9_C1	3711	644	52978,73	MZ825016	TTV
	SPI9_C2	3249	765	6636,10	MZ825017	TTV
	SPI9_C3	3216	742	5373,75	MZ825018	TTV
	SPI9_C4	3210	743	5865,09	MZ825019	TTV
	SPI9_C5	3106	768	12233,97	MZ825020	TTV
	SPI9_C6	3033	765	4067,50	MZ825021	TTV
	SPI9_C7	3032	750	27691,40	MZ825022	TTV
	SPI9_C8	2958	786	3691,05	MZ825023	TTV
	SPI9_C9*	2896	520	7458,40	MZ825150	TTMV
	SPI9_C10*	2895	673	41025,16	MZ825151	TTMV
	SPI9_C11*	2842	659	18593,57	MZ825152	TTMV
	SPI9_C12*	2842	661	1031,36	MZ825153	TTMDV
	SPI9_C13	2820	653	8789,48	MZ825024	TTMV
	SPI9_C14	2766	642	23763,10	MZ825025	TTMV
SPI9	SPI9_C15**	2658	749	5378,96	MZ825026	TTV
	SPI9_C16	2651	655	33,80	MZ825027	TTV
	SPI9_C17	2647	669	121228,62	MZ825028	TTMDV
	SPI9_C18**	2647	753	1128,45	MZ825029	TTV
	SPI9_C19	2613	663	215,57	MZ825030	TTMV
	SPI9_C20	2598	677	53192,50	MZ825031	TTMDV
	SPI9_C21	2552	667	8567,43	MZ825032	TTMDV
	SPI9_C22	2527	667	74606,90	MZ825033	TTMDV
	SPI9_C23	2504	588	17993,84	MZ825034	TTV
	SPI9_C24	2504	588	7815,10	MZ825035	TTV
	SPI9_C25	2503	661	13714,79	MZ825036	TTMDV
	SPI9_C26**	2253	659	4814,93	MZ825037	TTV
	SPI9_C29**	1795	569	6537,68	MZ825038	TTV
	SPI9_C30**	1758	531	1263,85	MZ825039	TTV
	SPI9_C31**	1752	529	248,34	MZ825040	TTV
SPI10	SPI10_C1	2945	643	370504,36	MZ824775	TTMV
	SPI10_C2	2793	667	379461,32	MZ824776	TTMDV
	SPI11_C1	3176	757	78,55	MZ824777	TTV
	SPI11_C2	3158	750	68,99	MZ824778	TTV
SPI11	SPI11_C3*	2944	678	77280,23	MZ825088	TTMV
	SPI11_C4*	2912	660	22886,46	MZ825089	TTMV
	SPI11_C5*	2837	416	1377,45	MZ825090	TTMV
	SPI12_C1	3899	771	67419,54	MZ824779	TTV
	SPI12_C2	3548	761	47494,74	MZ824780	TTV
	SPI12_C3	3334	647	16748,84	MZ824781	TTV
	SPI12_C4*	2988	668	4547,77	MZ825102	TTMV
	SPI12_C5	2986	754	454,59	MZ824782	TTV
	SPI12_C6*	2920	680	17276,60	MZ825103	TTMV
	SPI12_C7	2932	673	20,28	MZ824783	TTMV
SPI12	SPI12_C8	2925	731	2035,12	MZ824784	TTV
	SPI12_C9	2791	667	3521,33	MZ824785	TTMDV
	SPI12_C10	2596	678	49309,28	MZ824786	TTMDV
	SPI12_C11	2514	659	280,96	MZ824787	TTMDV
	SPI12_C12	2513	657	56740,45	MZ824788	TTMDV
	SPI12_C13	2498	662	524,39	MZ824789	TTMDV
	SPI12_C14	2481	659	110088,56	MZ824790	TTMDV
	SPI12_C15	2458	647	103966,00	MZ824791	TTMDV
SPI14	SPI14_C1	3692	641	35181,46	MZ824792	TTV
	SPI14_C2	3412	754	187191,07	MZ824793	TTV
SPI15	SPI15_C2*	2902	586	16467,54	MZ825104	TTMV
	SPI15_C3	2676	783	17,25	MZ824794	TTV
SPI16	SPI16_C1	3791	641	92600,52	MZ824795	TTV

	SPI16_C2	3081	786	7113,40	MZ824796	TTV
	SPI16_C3	2983	768	165624,99	MZ824797	TTV
	SPI16_C4	2942	744	13451,55	MZ824798	TTV
	SPI16_C5	2878	677	3884,33	MZ824799	TTMDV
	SPI17_C1	3526	752	13373,44	MZ824800	TTV
	SPI17_C2	3498	739	4445,99	MZ824801	TTV
	SPI17_C3	3405	773	12046,68	MZ824802	TTV
	SPI17_C4	3219	765	4303,91	MZ824803	TTV
	SPI17_C5	3159	769	9457,13	MZ824804	TTV
	SPI17_C6	3116	773	3911,75	MZ824805	TTV
	SPI17_C8	3093	655	5986,17	MZ824806	TTMDV
	SPI17_C9	3084	641	48772,31	MZ824807	TTV
	SPI17_C10	3075	532	46837,03	MZ824808	TTV
	SPI17_C11	2994	670	606,86	MZ824809	TTV
	SPI17_C12	2975	757	3348,70	MZ824810	TTV
	SPI17_C13*	2906	667	12269,38	MZ825105	TTMV
	SPI17_C14	2955	745	367,29	MZ824811	TTV
	SPI17_C15*	2887	660	5306,38	MZ825106	TTMV
	SPI17_C16	2935	731	21877,65	MZ824812	TTV
	SPI17_C17	2932	759	5254,73	MZ824813	TTV
	SPI17_C18	2926	746	101347,99	MZ824814	TTV
	SPI17_C19	2919	765	619,32	MZ824815	TTV
	SPI17_C20*	2863	739	36847,04	MZ825107	TTMV
	SPI17_C21	2905	670	4680,91	MZ824816	TTMV
	SPI17_C22	2901	742	576,96	MZ824817	TTV
	SPI17_C23	2898	650	10345,12	MZ824818	TTMV
	SPI17_C24*	2843	643	1465,74	MZ825108	TTMV
	SPI17_C25	2894	765	2384,26	MZ824819	TTV
	SPI17_C26	2888	744	2092,23	MZ824820	TTV
	SPI17_C27*	2830	645	2842,13	MZ825109	TTMV
	SPI17_C28	2884	674	14035,38	MZ824821	TTMDV
	SPI17_C29	2882	765	1558,69	MZ824822	TTV
	SPI17_C30	2853	735	2399,19	MZ824823	TTV
	SPI17_C32	2832	653	587,54	MZ824824	TTMV
	SPI17_C33	2797	664	13847,29	MZ824825	TTMDV
	SPI17_C34	2797	642	589,27	MZ824826	TTMV
	SPI17_C35	2770	671	4524,29	MZ824827	TTMDV
	SPI17_C36	2768	664	622,20	MZ824828	TTMV
	SPI17_C37	2727	607	438,60	MZ824829	TTV
	SPI17_C39	2689	680	204,47	MZ824830	TTMV
	SPI17_C40	2664	765	387,86	MZ824831	TTV
	SPI17_C41	2657	635	2939,76	MZ824832	TTMV
	SPI17_C42	2618	658	538,88	MZ824833	TTMV
	SPI17_C44	2567	669	523,81	MZ824834	TTMDV
	SPI17_C45	2545	662	6882,66	MZ824835	TTMDV
	SPI17_C46	2543	661	2204,08	MZ824836	TTMDV
	SPI17_C47	2531	667	1230,80	MZ824837	TTMDV
	SPI17_C48	2527	663	8272,30	MZ824838	TTMDV
	SPI17_C49	2514	654	3185,99	MZ824839	TTMDV
	SPI17_C50	2511	669	32460,78	MZ824840	TTMDV
	SPI17_C51	2509	659	2436,87	MZ824841	TTMDV
	SPI17_C52	2502	667	23640,37	MZ824842	TTMDV
	SPI17_C53	2498	660	3463,57	MZ824843	TTMDV
	SPI17_C54	2478	665	9381,34	MZ824844	TTMDV
	SPI17_C55	2452	658	7093,57	MZ824845	TTMV
	SPI17_C57	2341	659	1547,97	MZ824846	TTMV
	SPI17_C58	2126	571	817,19	MZ824847	TTMDV
	SPI17_C60**	1872	603	9738,55	MZ824848	TTV
	SPI18_C1	3856	761	14134,27	MZ824849	TTV
	SPI18_C2	3841	779	93235,28	MZ824850	TTV

Anexo

	SPI18_C4	3282	779	11180,41	MZ824851	TTV
	SPI18_C6	3167	736	13421,59	MZ824852	TTV
	SPI19_C1	3258	754	6536,28	MZ824853	TTV
	SPI19_C2	3160	763	5965,25	MZ824854	TTV
	SPI19_C3	3116	648	4816,18	MZ824855	TTV
	SPI19_C4	3093	661	6327,53	MZ824856	TTV
	SPI19_C5	3084	786	2388,75	MZ824857	TTV
	SPI19_C6	3072	736	57249,45	MZ824858	TTV
	SPI19_C7	3061	546	9494,69	MZ824859	TTV
	SPI19_C8	3043	773	25002,97	MZ824860	TTV
	SPI19_C9	3039	779	2959,61	MZ824861	TTV
	SPI19_C11	3015	750	7198,81	MZ824862	TTV
	SPI19_C12	2977	746	29607,12	MZ824863	TTV
	SPI19_C13	2977	761	8579,24	MZ824864	TTV
	SPI19_C14	2953	621	10013,60	MZ824865	TTV
SPI19	SPI19_C15*	2893	664	2603,91	MZ825110	TTMV
	SPI19_C16	2931	765	944,00	MZ824866	TTV
	SPI19_C17	2928	743	669,37	MZ824867	TTV
	SPI19_C18	2912	745	471,69	MZ824868	TTV
	SPI19_C19	2876	733	17750,18	MZ824869	TTV
	SPI19_C20	2871	743	1393,05	MZ824870	TTV
	SPI19_C21	2848	735	20,66	MZ824871	TTV
	SPI19_C22	2828	662	2428,55	MZ824872	TTMDV
	SPI19_C23	2828	734	899,23	MZ824873	TTV
	SPI19_C24*	2767	639	17281,41	MZ825111	TTMV
	SPI19_C25*	2783	642	35542,70	MZ825112	TTMV
	SPI19_C26	2808	756	7117,49	MZ824874	TTV
	SPI19_C27	2782	753	1725,06	MZ824875	TTV
	SPI19_C28**	2562	729	7367,77	MZ824876	TTV
	SPI19_C29	2549	744	79,34	MZ824877	TTV
SPI20	SPI20_C1	3783	762	107466,57	MZ824882	TTV
	SPI20_C2	3636	641	3145,35	MZ824883	TTV
	SPI21_C1	2932	664	44713,01	MZ824884	TTMV
SPI21	SPI21_C2*	2832	658	132520,91	MZ825113	TTMV
	SPI21_C3**	2093	522	21709,01	MZ824885	TTMDV
SPI22	SPI22_C1	3171	733	681,50	MZ824886	TTV
	SPI22_C2*	2029	491	4367,91	MZ825114	TTMDV
	SPI23_C1	3547	742	7857,45	MZ824887	TTV
	SPI23_C2	3169	750	1136,87	MZ824888	TTV
	SPI23_C4	3135	739	3559,78	MZ824889	TTV
	SPI23_C5	3109	644	404,30	MZ824890	TTV
	SPI23_C6*	2827	662	43002,39	MZ825115	TTMDV
	SPI23_C7*	2988	425	104719,21	MZ825116	TTMV
	SPI23_C8*	2962	663	12344,69	MZ825117	TTMV
	SPI23_C9	3008	733	1802,54	MZ824891	TTV
	SPI23_C10	2986	731	5457,88	MZ824892	TTV
	SPI23_C11	2981	767	332,94	MZ824893	TTV
	SPI23_C12	2963	749	2930,06	MZ824894	TTV
SPI23	SPI23_C13*	2888	673	11466,34	MZ825118	TTMV
	SPI23_C14*	2876	655	33172,16	MZ825119	TTMV
	SPI23_C15*	2839	664	31328,33	MZ825120	TTMV
	SPI23_C16	2910	808	112,14	MZ824895	TTV
	SPI23_C17	2909	680	587,87	MZ824896	TTMV
	SPI23_C18	2892	653	34196,99	MZ824897	TTMV
	SPI23_C19	2889	666	10127,43	MZ824898	TTMDV
	SPI23_C20	2880	742	6980,07	MZ824899	TTV
	SPI23_C21	2872	670	6527,69	MZ824900	TTMDV
	SPI23_C22*	2781	647	104897,80	MZ825121	TTMV
	SPI23_C23	2800	664	179,70	MZ824901	TTMV
	SPI23_C24	2771	658	12354,45	MZ824902	TTMV

	SPI23_C25	2684	751	686,51	MZ824903	TTV
	SPI23_C26	2680	743	2674,65	MZ824904	TTV
	SPI23_C27	2644	662	27754,17	MZ824905	TTMDV
	SPI23_C28	2619	665	3854,05	MZ824906	TTMV
	SPI23_C29	2570	677	3976,33	MZ824907	TTMDV
	SPI23_C30	2557	667	7280,33	MZ824908	TTMDV
	SPI23_C31	2540	661	1541,22	MZ824909	TTMDV
	SPI23_C33	2539	672	2054,61	MZ824910	TTMDV
	SPI23_C34	2521	670	962,94	MZ824911	TTMDV
	SPI23_C35	2493	667	2896,06	MZ824912	TTMDV
	SPI23_C36**	2238	721	27976,94	MZ824913	TTV
	SPI25_C1	3219	767	492,96	MZ824914	TTV
	SPI25_C2	3126	641	223,59	MZ824915	TTV
	SPI25_C3	3003	769	146,44	MZ824916	TTV
	SPI25_C4	2948	739	85,08	MZ824917	TTV
	SPI25_C5	2935	740	72,13	MZ824918	TTV
	SPI25_C6	2927	658	431979,25	MZ824919	TTMV
	SPI25_C7	2913	745	57,49	MZ824920	TTV
	SPI25_C8	2894	732	129,45	MZ824921	TTV
	SPI25_C9**	2542	539	1503,95	MZ824922	TTV
	SPI25_C10**	2486	542	682,07	MZ824923	TTV
	SPI26_C1	3661	644	251976,90	MZ824924	TTV
	SPI26_C2	2943	641	6848,47	MZ824925	TTV
	SPI27_C1	3761	739	305527,97	MZ824926	TTV
	SPI27_C2*	2910	665	9741,01	MZ825122	TTMV
	SPI27_C3*	2849	657	44476,58	MZ825123	TTMV
	SPI27_C4	2868	663	4703,97	MZ824927	TTMV
	SPI28_C1	3451	768	11330,08	MZ824928	TTV
	SPI28_C2*	2949	675	18731,53	MZ825124	TTMV
	SPI29_C1*	3699	648	415686,22	MZ825125	TTV
	SPI29_C2	3249	661	114993,04	MZ824929	TTV
	SPI29_C3*	2972	670	129265,33	MZ825126	TTMV
	SPI29_C4*	2952	673	60843,55	MZ825127	TTMV
	SPI29_C5*	2936	661	14724,74	MZ825128	TTMV
	SPI29_C6*	2802	655	47690,18	MZ825129	TTMV
	SPI30_C1	2837	665	240,87	MZ824930	TTMV
	SPI30_C2*	2781	647	723513,40	MZ825131	TTMV
	SPI32_C1	3410	760	46475,00	MZ824931	TTV
	SPI32_C2	3340	648	152541,07	MZ824932	TTV
	SPI32_C3	3144	748	8933,32	MZ824933	TTV
	SPI32_C4	3133	641	146103,53	MZ824934	TTV
	SPI32_C5*	2989	673	1278,80	MZ825132	TTMV
	SPI32_C6*	2984	687	4125,23	MZ825133	TTMV
	SPI32_C7	3008	786	123,04	MZ824935	TTV
	SPI32_C8	3005	786	1316,67	MZ824936	TTV
	SPI32_C9	2927	761	28121,39	MZ824937	TTV
	SPI32_C10	2900	732	1373,91	MZ824938	TTV
	SPI32_C11	2896	659	52686,73	MZ824939	TTMV
	SPI32_C12	2868	661	27935,97	MZ824940	TTMDV
	SPI32_C13**	2198	605	1208,88	MZ824941	TTV
	SPI32_C14**	1959	621	46211,44	MZ824942	TTV
	SPI32_C15**	1956	620	8702,60	MZ824943	TTV
	SPI32_C16**	1796	547	18358,17	MZ824944	TTMV
	SPI34_C1*	5852	666	39934,18	MZ825134	TTMDV
	SPI34_C2	3799	752	17888,09	MZ824945	TTV
	SPI34_C3*	2970	677	102569,02	MZ825135	TTMV
	SPI34_C4*	2968	678	252394,69	MZ825136	TTMV
	SPI34_C5*	2958	667	26157,86	MZ825137	TTMV
	SPI34_C6	2868	668	40697,85	MZ824946	TTMDV
	SPI35_C1	3677	509	356532,68	MZ824947	TTV

Anexo

	SPI35_C2	3398	751	5822,88	MZ824948	TTV
	SPI35_C3	3398	770	985,65	MZ824949	TTV
	SPI35_C4	3283	771	17376,61	MZ824950	TTV
	SPI35_C5	3249	735	35470,73	MZ824951	TTV
	SPI35_C6	3180	733	60396,88	MZ824952	TTV
	SPI35_C7	3085	786	3224,04	MZ824953	TTV
	SPI35_C8	2992	735	13210,54	MZ824954	TTV
	SPI35_C9	2949	763	41354,85	MZ824955	TTV
	SPI35_C10	2937	762	6581,28	MZ824956	TTV
	SPI35_C11	2895	671	91817,87	MZ824957	TTMDV
	SPI35_C12	2862	741	28,59	MZ824958	TTV
	SPI35_C13	2748	664	15970,35	MZ824959	TTMDV
	SPI35_C14**	2560	654	190472,67	MZ824960	TTMDV
SPI36	SPI36_C1	3684	742	216,80	MZ824961	TTV
	SPI36_C2*	2920	676	100363,80	MZ825138	TTMV
	SPI36_C3*	2832	658	151,24	MZ825139	TTMV
	SPI36_C4*	2810	653	24849,01	MZ825140	TTMV
SPI37	SPI37_C1	3835	757	353877,88	MZ824962	TTV
SPI38	SPI38_C1	3771	648	40217,98	MZ824963	TTV
	SPI38_C2	3703	751	680508,86	MZ824964	TTV
	SPI38_C3	3674	782	65578,95	MZ824965	TTV
	SPI38_C4	3411	735	377,06	MZ824966	TTV
	SPI38_C5	2967	765	24639,55	MZ824967	TTV
SPI39	SPI39_C1	3861	771	215906,80	MZ824968	TTV
	SPI39_C2	3750	742	163818,66	MZ824969	TTV
SPI41	SPI41_C1	3788	763	220242,68	MZ824974	TTMDV
	SPI41_C2*	3737	767	93991,81	MZ825142	TTV
	SPI41_C3*	3154	742	118237,61	MZ825143	TTV
	SPI41_C4**	2863	640	13457,12	MZ824975	TTMV
	SPI41_C5	2602	676	10139,84	MZ824976	TTMDV
	SPI41_C7	2559	501	64168,75	MZ824977	TTMDV
	SPI41_C8	2551	661	146771,28	MZ824978	TTMDV
	SPI41_C9	2547	654	3955,40	MZ824979	TTMDV
	SPI41_C11	2484	657	49649,74	MZ824980	TTV
SPI43	SPI43_C1	3771	760	557879,98	MZ824981	TTV
SPI44	SPI44_C1*	2848	662	98154,17	MZ825144	TTMV
	SPI44_C3	2672	674	126771,69	MZ824982	TTV
SPI45	SPI45_C1	3633	641	422,84	MZ824983	TTV
	SPI45_C2	2970	780	97321,42	MZ824984	TTV
	SPI45_C3	2834	642	649581,22	MZ824985	TTMV
	SPI45_C4	2834	653	87147,36	MZ824986	TTMV
	SPI45_C5	2566	672	810,19	MZ824987	TTMDV
	SPI45_C6**	1828	566	6644,95	MZ824988	TTV
	SPI45_C7**	1825	565	13797,21	MZ824989	TTV
SPI47	SPI47_C1*	2965	655	314569,01	MZ825145	TTMV
SPI48	SPI48_C1	3171	641	83045,37	MZ824990	TTV
SPI49	SPI49_C1	3575	782	7813,12	MZ824991	TTV
	SPI49_C2	2920	670	91645,49	MZ824992	TTMDV
SPI50	SPI50_C1	3144	739	8165,71	MZ824993	TTV
	SPI50_C2	3061	765	7477,88	MZ824994	TTV
	SPI50_C3	3043	750	59259,91	MZ824995	TTV
	SPI50_C4*	2957	584	200827,84	MZ825147	TTMV
	SPI50_C5	3005	784	14523,97	MZ824996	TTV
	SPI50_C6	2987	766	12102,40	MZ824997	TTV
	SPI50_C7	2970	757	46180,93	MZ824998	TTV
	SPI50_C8	2914	731	4317,65	MZ824999	TTV
	SPI50_C10**	2513	768	15207,98	MZ825000	TTV

Tabla Suplementaria S5. Lista de aislados de anelovirus descargados de GenBank empleados para el análisis filogenético en el capítulo 1. Se indica el número de acceso, el nombre del aislado y el género de anelovirus al que pertenecen. *Los que cumplían los criterios de demarcación de especies fueron elegidos para los posteriores análisis filogenéticos, incluyendo las secuencias descritas en el capítulo 1.

Número de acceso	Nombre del aislado	Género	Número de acceso	Nombre del aislado	Género
AY823989	3h	TTV	MH017578	BNI-700831-G4-SER28	TTV
AY823988	2h	TTV	MH017577	BNI-700831-G4-SERO	TTV
MK452767	N119	TTV	MH017576	BNI-700742-G4-CSF	TTV
MK452766	M265	TTV	MH017575	BNI-700639-G4-SERO	TTV
MK690196	TTVMY01	TTV	MH017574	BNI-700639-G4-CSF	TTV
MK820646	BD67	TTV	MH017573	BNI-700620-G4-SERO	TTV
MK820645	FL08	TTV	MH017572	BNI-700620-G4-CSF	TTV
MK820644	FL100*	TTV	MH017571	BNI-700611-G4-CSF	TTV
FJ392117	TW53A37*	TTV	MH017570	BNI-700199-G4-CSF	TTV
FJ392116	TW53A36	TTV	LC385663	KD5	TTV
FJ392115	TW53A31	TTV	LC385662	KD1	TTV
FJ392114	TW53A30*	TTV	AF261761	PMV*	TTV
FJ392112	TW53A39	TTV	AF435014	KAV*	TTV
FJ392111	TW53A35	TTV	AB041007	VT416	TTV
FJ392109	TW53A32	TTV	AF348409	TCHN-B*	TTV
FJ392108	TW53A29	TTV	AF345526	TCHN-A*	TTV
FJ392107	TW53A27	TTV	AF345523	TCHN-C1*	TTV
FJ392105	TW53A25	TTV	AF345524	TCHN-D1*	TTV
KJ082064	TTV-Hebei-1	TTV	AB017613	TUS01*	TTV
HM224451	L1984	TTV	AB041958	Mf-TTV3*	TTV
FJ426280	SIA109	TTV	AB064607	JT34F*	TTV
KT163917	P19-3	TTV	AB064605	CT44F*	TTV
KT163916	P19-2	TTV	AB064598	CT43F*	TTV
KT163915	P19-1	TTV	AB064595	CT23F*	TTV
KT163912	P18-1	TTV	AB060597	SAa-01*	TTV
KT163911	P17	TTV	AB060594	SAa-10*	TTV
KT163906	P15-4	TTV	AB054647	Kt-08F*	TTV
KT163905	P15-3	TTV	AB049608	s-TTV_CH71*	TTV
KT163904	P15-2	TTV	AB049607	s-TTV_CH65-2*	TTV
KT163903	P15-1	TTV	AB041959	Mf-TTV9*	TTV
KT163902	P14-3	TTV	AB041957	Pt-TTV6*	TTV
KT163901	P14-2	TTV	AB037926	s-TTV_CH65-1*	TTV
KT163900	P14-1	TTV	AB028668	TJN01*	TTV
KT163897	P13-2	TTV	AB038621	TTVyon-KC009*	TTV
KT163896	P13-1	TTV	MN116509	TTVMY02	TTV
KT163893	P10-1	TTV	KP036971	TTV-2013	TTV

KT163887	P9-2	TTV	KJ194503	TTV_Amsterdam_1994	TTV
KT163886	P9-1	TTV	KJ194502	TTV_Amsterdam_1995	TTV
KT163885	P8	TTV	DQ187006	TTV9-BM1C-18*	TTV
KT163880	P4-1	TTV	AX025830	TTV17-SENV-G*	TTV
KT163879	P3	TTV	AX025718	TTV18-SENV-C*	TTV
KT163875	P1-1	TTV	AX174942	TTV22-svi-1*	TTV
DQ361268	ViPi04*	TTV	KU243129	TTMV-204*	TTMV
DQ187005	BM1C-10	TTV	MH017549	BNI-700835-G1-CSF	TTMV
DQ187004	BM1C-16	TTV	MK212031	vzttmv4*	TTMV
DQ187002	BM1B-14	TTV	MK212030	vzttmv3*	TTMV
DQ187001	BM1B-11	TTV	MK212029	vzttmv2	TTMV
DQ187000	BM1B-07	TTV	MK139485	vzttmv1	TTMV
DQ186999	BM1B-05	TTV	KY462770	SHA*	TTMV
DQ186998	BM1A-13	TTV	KY462769	SH-A*	TTMV
DQ186997	BM1A-09	TTV	KY462768	SH-C9	TTMV
DQ186996	BM1A-02	TTV	KY462767	SHC11	TTMV
DQ186995	P605	TTV	KY462766	SH-C8	TTMV
DQ186994	P601	TTV	KY462765	SH-C7	TTMV
AJ620235	tth23	TTV	KY462764	SH-C6	TTMV
AJ620233	tth19	TTV	KY462763	SH-C5	TTMV
AJ620232	tth13	TTV	KY462762	SH-C4	TTMV
AJ620231	tth8	TTV	KY462761	SH-C3	TTMV
AJ620229	tth29	TTV	KY462760	SH-C2	TTMV
AJ620228	tth14	TTV	KY462759	SH-C1	TTMV
AJ620227	tth5	TTV	KY462758	SHC	TTMV
AJ620226	tth4	TTV	KY462757	SHC01	TTMV
AJ620225	tth31	TTV	KX810064	Emory2*	TTMV
AJ620224	tth27	TTV	KX810063	Emory1*	TTMV
AJ620223	tth26	TTV	KU041847	222*	TTMV
AJ620222	tth25	TTV	JX134046	TTMV_LY3*	TTMV
AJ620221	tth17	TTV	JX134045	TTMV_LY2	TTMV
AJ620220	tth16	TTV	JX134044	TTMV_LY1*	TTMV
AJ620219	tth9	TTV	EF538883	LIL-y4*	TTMV
AJ620217	tth21	TTV	MH017569	BNI-701098-G3-SERO	TTMV
AJ620218	tth3	TTV	MH017568	BNI-701098-G3-CSF	TTMV
AJ620216	tth20*	TTV	MH017567	BNI-700907-G3-SER28	TTMV
AJ620215	tth18	TTV	MH017565	BNI-700886-G3-SERO	TTMV
AJ620213	tth10	TTV	MH017564	BNI-700886-G3-CSF	TTMV
AJ620212	tth6	TTV	MH017563	BNI-700835-G3-CSF	TTMV
AB011493	TX011(O)	TTV	MH017562	BNI-701800-G2-CSF*	TTMV
AB011494	TY96117	TTV	MH017561	BNI-701157-G2-SERO	TTMV
AB011491	G88801	TTV	MH017560	BNI-701157-G2-CSF	TTMV
AB011490	G105001	TTV	MH017559	BNI-700886-G2-SERO	TTMV
AB011489	G104901	TTV	MH017558	BNI-700886-G2-CSF	TTMV
AB011488	G102001	TTV	MH017557	BNI-700835-G2-SER28	TTMV
AB011487	G103301	TTV	MH017556	BNI-700835-G2-SERO	TTMV
AB011486	G97801	TTV	MH017555	BNI-701800-G1-SER	TTMV

AB064606	JT33F	TTV	MH017554	BNI-701800-G1-CSF	TTMV
AB064604	CT39F	TTV	MH017553	BNI-701196-G1-CSF	TTMV
AB064603	JT41F	TTV	MH017552	BNI-700886-G1-CSF	TTMV
AB064602	JT19F	TTV	MH017551	BNI-700835-G1-SER28	TTMV
AB064601	JT14F	TTV	MH017550	BNI-700835-G1-SERO	TTMV
AB064600	JT05F	TTV	MH017549	BNI-700835-G1-CSF(2)	TTMV
AB064599	JT03F	TTV	MH017548	BNI-700831-G1-CSF*	TTMV
AB064597	CT30F	TTV	MH017547	BNI-700684-G1-CSF	TTMV
AB064596	CT25F	TTV	MH017546	BNI-700620-G1-CSF	TTMV
AB060593	SAa-38	TTV	AB038625	TLMV-CLC062	TTMV
AB054648	23	TTV	EF538882	LIL-y3*	TTMV
AB050448	TYM9*	TTV	EF538881	LIL-y2*	TTMV
EU305676	LTT10	TTV	EF538880	LIL-y1*	TTMV
EU305675	LTT7	TTV	KM259874	ALH8*	TTMV
EU305674	LTT6	TTV	KM259873	ALA22*	TTMV
EF538879	CSC5	TTV	AF291073	PB4TL*	TTMV
AF351132	TWH	TTV	AB038631	TLMV-NLC030*	TTMV
AF298585	P/1C1	TTV	AB041963	Pt-TTV8-II*	TTMV
AF116842	BDH1	TTV	AB041962	TGP96*	TTMV
AF247138	T3PB	TTV	AB038630	TLMV-NLC026*	TTMV
AF247137	TUPB	TTV	AB038629	TLMV-NLC023*	TTMV
AF122920	US35	TTV	AB038627	TLMV-CLC156*	TTMV
AF122917	JA4	TTV	AB026931	TLMV-CBD279*	TTMV
AF122916	JA1	TTV	AB026929	TLMV-CBD203*	TTMV
AF122915	JA9	TTV	KF764702	D50	TTMV
AF122914	JA20	TTV	KF764701	D11*	TTMV
AF345529	TCHN-G2	TTV	KY856742	zhenjiang*	TTMV
AF345528	TCHN-F	TTV	AB026930	TLMV-CBD231	TTMV
AF345527	TCHN-C2*	TTV	AB038628	TLMV-CLC205	TTMV
AF345525	TCHN-D2	TTV	AB038626	TLMV-CLC138	TTMV
AF345522	TCHN-E	TTV	KT163888	P9-3*	TTMV
AF345521	TCHN-G1	TTV	MH017566	BNI-700907-G3-SERO*	TTMV
AY026466	L02	TTV	LC381845	KS025*	TTMV
AY026465	L01	TTV	KT163876	P1-2*	TTMV
AF254410	TTV021	TTV	KT163914	P18-3	TTMV
AF122913	GH1	TTV	KT163898	P13-3*	TTMV
AF129887	TTVCHN2	TTV	KT163910	P15-8*	TTMV
AF079173	TTVCHN1	TTV	KT027936	Cpz1cl2*	TTMV
AB060595	SAj-30	TTV	KT027938	Cpz2cl2*	TTMV
AB060596	SAf-09	TTV	KT027941	GorMcl4*	TTMV
AB041821		TTV	AB303563	MDJN52	TTMDV
AB030489	JaBD98	TTV	AB449064	Pt-TTMDV225-2	TTMDV
AB030488	JaBD89	TTV	AB449063	Pt-TTMDV225-1	TTMDV
AB030487	JaCHCTC19	TTV	AB303565	MDJN91	TTMDV
AB030486	JaCHC2c	TTV	AB303557	MDJHem8-2	TTMDV
AB028669	TJN02	TTV	AB303556	MDJHem8-1	TTMDV
AB038624	TTVyon-KC197	TTV	AB303555	MDJHem5*	TTMDV

AB038623	TTVyon-KC186	TTV	AB449062	Pt-TTMDV210*	TTMDV
AB038622	TTVyon-LC011	TTV	AB303566	MDJN97*	TTMDV
AB038620	TTVsan-S039	TTV	AB303564	MDJN69*	TTMDV
AB038619	TTVsan-IR1031	TTV	AB303562	MDJN51*	TTMDV
AB038340	TRM1	TTV	AB303561	MDJN47*	TTMDV
AB026347	TP1-3	TTV	AB303560	MDJN14*	TTMDV
AB026346	TK16	TTV	AB303559	MDJN2*	TTMDV
AB026345	TRM1	TTV	AB303558	MDJN1*	TTMDV
AB008394	TA278*	TTV	AB303554	MDJHem3-2*	TTMDV
AB025946	TTV_SANBAN*	TTV	AB303553	MDJHem3-1*	TTMDV
AY666122	HEL32*	TTV	AB303552	MDJHem2*	TTMDV
GU797360	BIS8-17	TTV	AB290917	MD1-032	TTMDV
MH017587	BNI-701800-G4-SER	TTV	EF538876	6PoSMA*	TTMDV
MH017586	BNI-701098-G4-CSF	TTV	EF538875	2PoSMA*	TTMDV
MH017585	BNI-700907-G4-SER28	TTV	AB290918	MD1-073*	TTMDV
MH017584	BNI-700907-G4-SERO	TTV	AB290919	MD2-013*	TTMDV
MH017583	BNI-700907-G4-CSF	TTV	KT163882	P6*	TTMDV
MH017582	BNI-700886-G4-CSF	TTV	AY622909	SAV2	TTMDV
MH017581	BNI-700835-G4-SER28	TTV	AY622908	SAV1	TTMDV
MH017579	BNI-700835-G4-CSF	TTV	KT163891	P9-6	TTMDV

Tabla Suplementaria S6. Lista de aislados de anelovirus descargados de GenBank empleados para el análisis filogenético en el capítulo 2. En la tabla, se indica el número de acceso, el nombre del aislado, el género de anelovirus al que pertenecen y el origen. La última columna muestra el número de secuencias del capítulo 2 que se agrupan con cada aislado en particular. *Aislados identificados como posibles nuevas especies en el capítulo 1.

Número de acceso	Nombre del aislado	Género	Origen	Número de secuencias del capítulo 2 que agrupan con esta especie
KJ082064	TTV-Hebei-1	TTV	China	1
AB025946	TTV19-SANBAN	TTV	Japón	3
AY666122	HEL32	TTV	Finlandia	13
AF261761	PMV	TTV	Reino Unido	1
AF435014	KAV	TTV	Alemania	1
AB041007	VT416	TTV	Japón	4
AF348409	TCHN-B	TTV	China	2
AF345526	TCHN-A	TTV	China	15
AF345523	TCHN-C1	TTV	China	5
AB041958	Mf-TTV3	TTV	Japón (Macaco)	0
AB064607	JT34F	TTV	Japón	1
AB060597	SAa-01	TTV	Japón	25
AB060594	SAa-10	TTV	Japón	6
AB049608	s-TTV_CH71	TTV	Japón (Chimpancé)	0

AB041959	Mf-TTV9	TTV	Japón (Macaco)	0
AB041957	Pt-TTV6	TTV	Japón (Chimpancé)	0
AB037926	s-TTV_CH65-1	TTV	Japón (Chimpancé)	0
AB028668	TJN01	TTV	Japón	2
AB038621	TTVyon-KC009	TTV	Japón	18
DQ187006	TTV9-BM1C-18	TTV	EEUU	0
AX025830	TTV17-SENV-G	TTV	Italia	0
AX025718	TTV18-SENV-C	TTV	Italia	20
KP296857	VWP00522.2	TTV	Saint Kitts and Nevis (Mono verde africano)	0
KP296856	VGA00120.1	TTV	Gambia (Mono verde africano)	0
KP296853	VGA00123.3	TTV	Gambia (Mono verde africano)	0
KP296854	VGA00154.2	TTV	Gambia (Mono verde africano)	0
MW455346	P1-c2*	TTV	España	0
MW455455	P2-c10*	TTV	España	10
MW455379	P3-c6*	TTV	España	0
MW455393	P7-c2*	TTV	España	5
MW455435	P9-c4*	TTV	España	2
MW455417	P8-c4*	TTV	España	3
MK212031	vztm4	TTMV	Vietnam	0
KY462770	SHA	TTMV	China	0
KX810064	Emory2	TTMV	EEUU	1
KX810063	Emory1	TTMV	EEUU	0
KU041847	222	TTMV	China	4
JX134046	TTMV_LY3	TTMV	Francia	6
JX134044	TTMV_LY1	TTMV	Francia	0
MH017563	BNI-700835-G3-CSF	TTMV	Ghana	0
MH017546	BNI-700620-G1-CSF	TTMV	Ghana	1
AB038625	TLMV-CLC062	TTMV	Japón	2
EF538882	LIL-y3	TTMV	Francia	1
EF538881	LIL-y2	TTMV	Francia	2
EF538880	LIL-y1	TTMV	Francia	3
KM259874	ALH8	TTMV	España	0
KM259873	ALA22	TTMV	España	0
AF291073	PB4TL	TTMV	Francia	0
AB041963	Pt-TTV8-II	TTMV	¿?(Chimpancé)	0
AB041962	TGP96	TTMV	Japón	0
AB038630	TLMV-NLC026	TTMV	Japón	0
AB038629	TLMV-NLC023	TTMV	Japón	6
AB038627	TLMV-CLC156	TTMV	Japón	1
AB026931	TLMV-CBD279	TTMV	Japón	10
AB026929	TLMV-CBD203	TTMV	Japón	0
KF764701	D11	TTMV	Países Bajos	0
KY856742	zhenjiang	TTMV	China	0
LC381845	KS025	TTMV	Japón	0
MH648907	ctcf040	TTMV	EEUU	3
MH648989	ctga035	TTMV	EEUU	6
MH648910	ctcd026	TTMV	EEUU	0
MH649141	ctbc019	TTMV	EEUU	2
MH648986	ctbd020	TTMV	EEUU	0
MH648912	ctea38	TTMV	EEUU	0
MH649017	ctbb016	TTMV	EEUU	4
MH649029	ctbi042	TTMV	EEUU	0
MH649114	ctbf050	TTMV	EEUU	1
MH648966	ctei055	TTMV	EEUU	0
MH649209	ctbg056	TTMV	EEUU	1
MH648982	ctbf014	TTMV	EEUU	5
KT027936	Cpz1cl2	TTMV	República Checa (Chimpancé)	0

KT027938	Cpz2cl2	TTMV	República Checa (Chimpancé)	0
KT027941	GorMcl4	TTMV	República Checa (Gorila)	0
MW455358	P1-c16*	TTMV	España	0
MW455453	P1-c22*	TTMV	España	0
MW455381	P3-c8*	TTMV	España	4
MW455388	P5-c2*	TTMV	España	0
MW455402	P7-c12*	TTMV	España	0
MW455403	P7-c14*	TTMV	España	1
MW455409	P7-c21*	TTMV	España	2
MW455424	P8-c16*	TTMV	España	3
MW455427	P8-c22*	TTMV	España	0
MW455431	P8-c26*	TTMV	España	1
MW455445	P10-c9*	TTMV	España	1
AB449062	Pt-TTMDV210	TTMDV	Japón (Chimpancé)	1
AB303566	MDJN97	TTMDV	Japón	2
AB303564	MDJN69	TTMDV	Japón	2
AB303562	MDJN51	TTMDV	Japón	3
AB303561	MDJN47	TTMDV	Japón	1
AB303560	MDJN14	TTMDV	Japón	0
AB303559	MDJN2	TTMDV	Japón	0
AB303558	MDJN1	TTMDV	Japón	8
AB303554	MDJHem3-2	TTMDV	Japón	0
AB303553	MDJHem3-1	TTMDV	Japón	3
AB303552	MDJHem2	TTMDV	Japón	1
AB290917	MD1-032	TTMDV	Japón	2
EF538876	6PoSMA	TTMDV	Francia	0
EF538875	2PoSMA	TTMDV	Francia	0
AB290919	MD2-013	TTMDV	Japón	0
MW455361	P1-c19*	TTMDV	España	1
MW455362	P1-c23*	TTMDV	España	0
MW455378	P3-c5*	TTMDV	España	0
MW455394	P7-c3*	TTMDV	España	1
MW455406	P7-c18*	TTMDV	España	5
MW455410	P7-c22*	TTMDV	España	3
MW455412	P7-c24*	TTMDV	España	2
MW455436	P9-c5*	TTMDV	España	1
MW455450	P12-c3*	TTMDV	España	1

Tabla Suplementaria S7: Lista de genomas de anelovirus descargadas de GenBank para hacer el análisis filogenético en el capítulo 3. Para cada aislado, aparece su número de acceso en GenBank, su nombre correspondiente, el género al que pertenece y su lugar de procedencia. Los aislados marcados con un asterisco son los que se han identificado en el capítulo 2 como potenciales nuevas especies, incluyendo las especies obtenidas tras la reevaluación de los datos del capítulo 1. En la última columna se encuentra indicado el número de secuencias de este estudio que agrupan con cada aislado de referencia.

Número de acceso	Nombre del aislado	Género	Origen	Número de secuencias del Capítulo 3 que agrupan con esta especie
KJ082064	TTV-Hebei-1	TTV	China	0
AB025946	TTV19-SANBAN	TTV	Japón	7
AY666122	HEL32	TTV	Finlandia	28

AF261761	PMV	TTV	Reino Unido	0
AF435014	KAV	TTV	Alemania	0
AB041007	VT416	TTV	Japón	12
AF348409	TCHN-B	TTV	China	6
AF345526	TCHN-A	TTV	China	25
AF345523	TCHN-C1	TTV	China	13
AB041958	Mf-TTV3	TTV	Japón (macaco)	0
AB064607	JT34F	TTV	Japón	2
AB060597	SAa-01	TTV	Japón	15
AB060594	SAa-10	TTV	Japón	4
AB049608	s-TTV_CH71	TTV	Japón (Chimpancé)	0
AB041959	Mf-TTV9	TTV	Japón (macaco)	0
AB041957	Pt-TTV6	TTV	Japón (Chimpancé)	0
AB037926	s-TTV_CH65-1	TTV	Japón (Chimpancé)	0
AB028668	TJN01	TTV	Japón	3
AB038621	TTVyon-KC009	TTV	Japón	25
DQ187006	TTV9-BM1C-18	TTV	EEUU	1
AX025830	TTV17-SENV-G	TTV	Italia	2
AX025718	TTV18-SENV-C	TTV	Italia	15
KP296857	VWP00522.2	TTV	Saint Kitts and Nevis (mono verde africano)	0
KP296856	VGA00120.1	TTV	Gambia (mono verde africano)	0
KP296853	VGA00123.3	TTV	Gambia (mono verde africano)	0
KP296854	VGA00154.2	TTV	Gambia (mono verde africano)	0
MW455346	P1-c2*	TTV	España	0
MW455455	P2-c10*	TTV	España	4
MW455379	P3-c6*	TTV	España	2
MW455393	P7-c2*	TTV	España	3
MW455435	P9-c4*	TTV	España	5
MW455417	P8-c4*	TTV	España	4
MZ285970	SP2_C3*	TTV	España	1
MZ286157	SP47_C2*	TTV	España	8
MZ286030	SP16_C12*	TTV	España	1
MZ286062	SP24_C4*	TTV	España	0
MZ286110	SP36_C1*	TTV	España	1
MZ286093	SP31_C2*	TTV	España	2
MK212031	vztmv4	TTMV	Vietnam	
KY462770	SHA	TTMV	China	0
KX810064	Emory2	TTMV	EEUU	0
KX810063	Emory1	TTMV	EEUU	0
KU041847	222	TTMV	China	0
JX134046	TTMV_LY3	TTMV	Francia	2
JX134044	TTMV_LY1	TTMV	Francia	0
MH017563	BNI-700835-G3-CSF	TTMV	Ghana	0
MH017546	BNI-700620-G1-CSF	TTMV	Ghana	0
AB038625	TLMV-CLC062	TTMV	Japón	4
EF538882	LIL-y3	TTMV	Francia	1
EF538881	LIL-y2	TTMV	Francia	4
EF538880	LIL-y1	TTMV	Francia	1
KM259874	ALH8	TTMV	España	1
KM259873	ALA22	TTMV	España	0
AF291073	PB4TL	TTMV	Francia	0
AB041963	Pt-TTV8-II	TTMV	¿?(Chimpancé)	0
AB041962	TGP96	TTMV	Japón	0
AB038630	TLMV-NLC026	TTMV	Japón	0
AB038629	TLMV-NLC023	TTMV	Japón	1
AB038627	TLMV-CLC156	TTMV	Japón	1
AB026931	TLMV-CBD279	TTMV	Japón	4
AB026929	TLMV-CBD203	TTMV	Japón	0
KF764701	D11	TTMV	Países Bajos	0

Anexo

KY856742	zhenjiang	TTMV	China	0
LC381845	KS025	TTMV	Japón	0
MH648907	ctcf040	TTMV	EEUU	0
MH648989	ctga035	TTMV	EEUU	2
MH648910	ctcd026	TTMV	EEUU	0
MH649141	ctbc019	TTMV	EEUU	0
MH648986	ctbd020	TTMV	EEUU	0
MH648912	ctea38	TTMV	EEUU	0
MH649017	ctbb016	TTMV	EEUU	5
MH649029	ctbi042	TTMV	EEUU	0
MH649114	ctbf050	TTMV	EEUU	1
MH648966	ctei055	TTMV	EEUU	0
MH649209	ctbg056	TTMV	EEUU	1
MH648982	ctbf014	TTMV	EEUU	0
MW455358	P1-c16*	TTMV	España	0
MW455453	P1-c22*	TTMV	España	0
MW455381	P3-c8*	TTMV	España	5
MW455388	P5-c2*	TTMV	España	2
MW455402	P7-c12*	TTMV	España	0
MW455403	P7-c14*	TTMV	España	0
MW455409	P7-c21*	TTMV	España	2
MW455424	P8-c16*	TTMV	España	11
MW455427	P8-c22*	TTMV	España	0
MW455431	P8-c26*	TTMV	España	0
MW455445	P10-c9*	TTMV	España	3
MZ285991	SP6-C5*	TTMV	España	1
MZ286237	SP9-C1*	TTMV	España	2
MZ286004	SP10-C3*	TTMV	España	0
MZ286241	SP15-C4*	TTMV	España	0
MZ286242	SP16-C1*	TTMV	España	0
MZ286246	SP19-C1*	TTMV	España	0
MZ286044	SP20-C3*	TTMV	España	3
MZ286248	SP22-C4*	TTMV	España	0
MZ286058	SP23-C4*	TTMV	España	3
MZ286262	SP35-C6*	TTMV	España	0
MZ286108	SP35-C8*	TTMV	España	0
MZ286263	SP36-C2*	TTMV	España	1
MZ286111	SP36-C4*	TTMV	España	0
MZ286127	SP40-C4*	TTMV	España	0
MZ286268	SP41-C2*	TTMV	España	0
MZ286271	SP42-C2*	TTMV	España	5
MZ286134	SP43-C3*	TTMV	España	0
MZ286272	SP43-C4*	TTMV	España	0
MZ286278	SP46-C2*	TTMV	España	0
MZ286280	SP49-C5*	TTMV	España	0
MZ286180	SP49-C6*	TTMV	España	0
MZ286187	SP51-C11*	TTMV	España	0
MZ286194	SP51-C18*	TTMV	España	0
MZ286281	SP51-C6*	TTMV	España	4
MZ286285	SP52-C3*	TTMV	España	0
MZ286286	SP56-C3*	TTMV	España	2
AB449062	Pt-TTMDV210	TTMDV	Japón (Chimpancé)	0
AB303566	MDJN97	TTMDV	Japón	4
AB303564	MDJN69	TTMDV	Japón	2
AB303562	MDJN51	TTMDV	Japón	1
AB303561	MDJN47	TTMDV	Japón	1
AB303560	MDJN14	TTMDV	Japón	2
AB303559	MDJN2	TTMDV	Japón	0
AB303558	MDJN1	TTMDV	Japón	8
AB303554	MDJHem3-2	TTMDV	Japón	2
AB303553	MDJHem3-1	TTMDV	Japón	2
AB303552	MDJHem2	TTMDV	Japón	1
AB290917	MD1-032	TTMDV	Japón	2

EF538876	6PoSMA	TTMDV	Francia	1
EF538875	2PoSMA	TTMDV	Francia	2
AB290919	MD2-013	TTMDV	Japón	0
MW455361	P1-c19*	TTMDV	España	2
MW455362	P1-c23*	TTMDV	España	0
MW455378	P3-c5*	TTMDV	España	0
MW455394	P7-c3*	TTMDV	España	4
MW455406	P7-c18*	TTMDV	España	5
MW455410	P7-c22*	TTMDV	España	1
MW455412	P7-c24*	TTMDV	España	0
MW455436	P9-c5*	TTMDV	España	1
MW455450	P12-c3*	TTMDV	España	1
MZ285998	SP7-C8	TTMDV	España	0
MZ285999	SP7-C10*	TTMDV	España	0
MZ286000	SP8-C1*	TTMDV	España	0
MZ286041	SP19-C2*	TTMDV	España	0
MZ286069	SP25-C4*	TTMDV	España	0
MZ286071	SP26-C2*	TTMDV	España	2
MZ286081	SP28-C5*	TTMDV	España	0
MZ286087	SP30-C3*	TTMDV	España	1
MZ286096	SP31-C6*	TTMDV	España	3
MZ286116	SP37-C4*	TTMDV	España	2
MZ286117	SP37-C7*	TTMDV	España	0
MZ285986	SP3-C15*	TTMDV	España	1
MZ286165	SP47-C10*	TTMDV	España	3
MZ286172	SP48-C5*	TTMDV	España	0
MZ286176	SP49-C1*	TTMDV	España	0
MZ286199	SP51-C23*	TTMDV	España	0
MZ286217	SP58-C3*	TTMDV	España	0

Tabla Suplementaria S8. Cebadores utilizados para la amplificación por PCR y la secuenciación Sanger del gen ORF1 de los *contigs* de anelovirus. También se utilizaron cebadores internos (F2/R2) para la secuenciación cuando fue necesario. En la última columna, aparece la profundidad de cobertura que presenta el *contig*.

Nombre del aislado	Nombre del cebador	Secuencia (5'-3')	Tamaño del fragmento de la PCR (nt)	Cobertura de SPADES
P1-c2	P1-C2F	GTTATCACTTCGACCCGCCGGCAC	2427	107,43
	P1-C2R	CTCGTACTCACACTGCCAGTCCTCCTT		
	P1-C2R2	CCTCCTTAGTGGGAGGACGA		
	P1-C2F2	CCAAGGCCAGCACTGAGG		
P1-c4	P1-C4F	CAGAGCCTAACCCGAGCAGCCATG	2430	7,62
	P1-C4R	GGGGTAGTAAGGCCTGTCTGTGGTT		
	P1-C4F2	CACATTAGAAGACTGGCTGTACAC		
P1-c16	P1-C16F	GCGGTTGCGATTCTACTATTTGCCACTT	2143	249,61
	P1-C16R	CAGCTTCTGTTGTTCTTCCCATGGATA		
P1-c23	P1-C23F	CCTGACCATCCAGACAGACAATT	2059	156,52
	P1-C23R	TTCTGTTTGTAGTTCAAAGCCTGGT		
P3-c4	P3-C4F	CCTACTGCTTCACATCTTTTCTCACGGAG	2200	6,21
	P3-C4R	AGAGGTTCAGGTACAAGCCAGTCATAGT		
	P3-C4F2	GACTACCAACGTGGAACAGCAC		
P3-c5	P3-C5F	TGGTGGAGAAGAAGAAACCAACATGTACT	1993	34,38
	P3-C5R	GGCTCTACTCTAAAATTCCTGTTGTAGT		
	P3-C5F2	CCATTGCAACTGGCTCTGCAG		

P3-c6	P3-C6F	AGTCCAGACAACAATAACCCGAGAC	2350	125,10
	P3-C6R	TGCATATTCTGTTTCCAGTCTCTCT		
	P3-C6F2	CCAGCTTTTGGCAGTCAAATATAGCAA		
P3-c8	P3-C8F	TTGTGAATGTGAAAAACCACTTCAA	2214	351,31
	P3-C8R	AAAAAAAATCATCTGCCCTTTCTTG		
	P3-C8F2	GCAGCTCCAGGACTACTAAAACCAAGG		
P5-c2	P5-C2F	ACCCACACCGCTACTTTTATCTTTGAACA	2189	1,96
	P5-C2R	CCTCAAATTCTCTAGCAGTCATACGTCT		
	P5-C2R2	GGACAAATACCAACCACAAAACCTAC		
P7-c3	P7-C3F	CAACCATCCTTTTGCTCATTGCTTGCT	2269	105,41
	P7-C3R	AGGCTAAAGTTGACATAGGGTTCAGGGG		
	P7-C3F2	TCTGCAGCTGACTTTAGACACCCAACA		
P7-c14	P7-C14F	CTACCAACCATTAGCACATACAGCAAC	2211	3,22
	P7-C14R	CTTGTTTCATATTTCTCAGGAGTCATTCTCC		
	P7-C14F2	GCACAACAATATGGATATGGCACAG		
P7-c18	P7-C18F	TAATCACCAGAGACTACCAACAATGCCA	2079	91,86
	P7-C18R	CCTTGGTGGTCTGTGAAATGCTG		
	P7-C18F2	CTACTGGTTGGTTTCAACCAGAC		
P7-c20	P7-C20F	AAGTAGGTGAGCAAAGGGTAAGTTACA	2081	5,50
	P7-C20R	TGGTCACAGCGATTCGAGAATTAAGTG		
P7-c21	P7-C21F	GCTGCAACGACCACTCTACATA	2119	157,93
	P7-C21R	CAGGGGTAATCGTCTAGGTGGTC		
P7-c22	P7-C22F	CTGCTTGCCAATATCTTCTCTGTG	2121	482,31
	P7-C22R	TCACGTTCTGTCTCTTCAAAAGC		
	P7-C22F2	AACAGCCTGCACATTTAGATACCTT		
P7-c24	P7-C24F	ATTTTTCCACTTGGCCATTCTGAT	2086	150,74
	P7-C24R	TTCTGTGCTTGTTCAAATCCTGGT		
	P7-C24F2	CTGATTACATACAGAAAACCAAGCTC		
P8-c17	P8-C17F	GCAACTATTAAGTCCAGCATCTCTGCT	2109	156,85
	P8-C17R	TCAGCCTCCCAACTCTTTCCCTTATAG		
	P8-C17R2	CCAGTTTGTGGATACCAGTGCAG		
P8-c22	P8-C22F	TCCCATCTTAGATTACAAGTGGAGAA	2166	142,89
	P8-C22R	TGTCGTGGGATAGTGGTCCCATATTT		
	P8-C22F2	GGAAACAGCAATGCACAAGTTTAACC		
P8-c23	P8-C23F	AACTACTGGCAATATCCAACAAGCA	1796	53,68
	P8-C23R	TAAGTCTCTCATGGGGTGCAAAG		
P8-c25	P8-C25F	CAGTCTCAATCACCTTTAAACCAACA	2109	132,74
	P8-C25R	CATGGTTTCATGCGTCTGTTTTTAG		
	P8-C25F2	CAGAAAGCAGCACAATAACATTACATACCC		
P9-c4	P9-C4F	TCCCGCTCTGATTTCCAAAACAATCC	2442	16,43
	P9-C4R	AGGGGTAGTATGGAGTGTCTGACAGGT		
	P9-C4F2	AGACTTCTATTACCCATAATCGG		
P9-c5	P9-C5F	TTTTCTCCTGGCCATCAAGACAG	2140	9,27
	P9-C5R	GGGGTGTCTCTTTAAACAATCTAGGGG		
P10-c9	P10-C9F	GGATTCCAAAACCCGAAACAGATA	2122	15,59
	P10-C9R	TTGTTGTTCCAGGGGTGTTAGTCTT		
	P10-C9F2	GATACAATACACACATGCAGACTC		
	P10-C9R2	GTTTCATATCTTCTGATGGGTCAAAG		
P12-c3	P12-C3F	TGCTTGCCAGTATCTTCTGTGGGTCA	2149	6,16
	P12-C3R	CAGGTCTGAGAAATGCAAAGCAAGTTCA		
	P12-C3F2	TGCCCAGAAGCAACTTACACATAC		

Tabla Suplementaria S9. Cebadores utilizados para amplificación por PCR y secuenciación Sanger de la secuencia nucleotídica correspondiente a la poliproteína de HPgV. También se utilizaron cebadores internos para la secuenciación cuando fue necesario.

Muestra	Par	Nombre del primer	Secuencia (5'-3')	Tamaño del amplicón (nt)
SP3	1	SP3-1-F	CAGTCCACCGAAAGAAGGC	203
		SP3-1-R	GCGACTCCTCTGGCTTGTA	
	2	SP3-2-F	CACCGAAAGAAGGCCGGA	190
		SP3-2-R	AAAATCCAATTGATGAACACCCC	
SP4	1	SP4-1-F	TGTATCAGGCGGTTTGGC	200
		SP4-1-R	TGCACACGTTAGGTTAGGAAC	
	2	SP4-2-F	ATGACCAAGATCCGGGACTC	284
		SP4-2-R	ACCCTGGACATGTATGAATCCA	
SP13	1	SP13-1-F	CCGGGCCGATCTTAATGGT	747
		SP13-1-R	GTTTCCCGCCATGAAGTC	
	2	SP13-2-F	CCGATCTTAATGGTCGGCCT	735
		SP13-2-R	CGCCATGAAGTCGAAGACGA	
SP19	1	SP49_HPgV_9	GCAGTAGGAATGCTGTGTC	680
		SP19_HPgV_1	CTCACACTCCGCTTGGAAAT	
	2	SP19_HPgV_2	TGTCAGCTGGGCAAAAGTTC	506
		SP19_HPgV_3	GACCAATCGCAATCCACCTG	
	3	SP19_HPgV_4	TGTTTCCATTGTTGACAAGCTC	689
		SP19_HPgV_5	GCATGATCCACTCCCACAAG	
	4	SP19-1-F-int	CACCTGTGTGGCAGGGTTAC	639
		SP19-1-R-int	GGAATGGCAGAACACGAGAT	
	5	UTR-F	TTGGTAGCCACTATAGGTGGG	593
		UTR-R	CCTGATACAGTGGCCAGCAT	
SP20	1	SP20_HPgV_1	CTACGGGTTGACATGGCAGT	578
		SP49_HPgV_4	CTATTGACGAAGGGCACAGC	
	2	SP20_HPgV_3	GTGGTTGAGTGCTGTGTGATG	288
		SP20_HPgV_4	ATCCTGAAAGACACCCGATGAGA	
	3	SP20_HPgV_5	GCTCGTGTCAAGTCTTCA	758
		SP20_HPgV_6	GGTCTCCGTCCTTGATGATG	
	4	SP20_HPgV_7	TATGACGACTGCCCTACACC	1188
		SP20_HPgV_8	ACCAAGTGGTAGTGCCAGAGTT	
	5	SP20-1-F-int	TGGCAGTCCTGCTCTGTAG	509
		SP20-1-R-int	CGATGGAAAGGGAATCTGACC	
	6	SP20-2-F-ext	TGCTTGGCCTCATACTCTG	436
		SP20-2-R-ext	GATTACATCCTGAAAGACACC	
	7	SP20-3-F-int	GCTTCATTGAGTGGTAGGG	728
		SP20-3-R-int	TGGAAGAGTCTTTCCCCTA	
8	UTR-F	TTGGTAGCCACTATAGGTGGG	593	
	UTR-R	CCTGATACAGTGGCCAGCAT		
9	SP20-9-F	AGGTAGAGTTGGATGTGGCC	302	
	SP20-9-R	GACCTTGTGCCCCATGTTAC		
10	SP20-10-F	CTATGGGAGGTTTCTGGCCA	612	
	SP20-10-R	TAGTAGTAGGCCAGACCT		
11	SP20-11-F	CTCCTATACCCATGGCACCC	760	
	SP33-2-R	CCGTAGTGGCAGTGCAATTAG		
12	UTR-2-F	TAGCCACTATAGGTGGGCTTAAAG	550	
	UTR-2-R	TGATACAGTGGCCAGCAT		
13	SP20-13-F	TGACTTCCGTGGCTCTTCTG	266	
	SP20-13-R	TGTGCCCCATGTTACCATACTC		
14	SP20-14-F	CATCACACGGTGGGTCATCA	500	
	SP33-4-R	AGTGCATTAGCACCATCAGGC		
SP30	1	SP30_HPgV_1	GTTGCTATGGTCTGCTCTTC	364

	2	SP30_HPgV_2 SP20_HPgV_7 SP30_HPgV_3	GAACCTCATCACCAGTCTAGC TATGACGACTGCCCTTACACC CATGAAGTCGAAGACGAGACTG	1080	
	3	SP30-2-F-int SP30-2-R-int	CGGCTGACATCGGTGAAG GAAGTCGAAGACGAGACTGG	1046	
SP33	1	SP33-1-F SP33-1-R	CTATTGGGAGCCGCTGGTA GTGCCAGAGTTGTTAGCTGAG	290	
	2	SP33-2-F SP33-2-R	GTTACCCGAGACACAACCAA CCGTAGTGGCAGTGCAATAG	505	
	3	SP33-3-F SP33-3-R	ACACCCGCTGTGGGGTTA CCAGAGTTGTTAGCTGAGTACAAA	257	
	4	SP33-4-F SP33-4-R	AGATGGAAGCCGGCAAGGTA AGTGCAATTAGCACCATCAGGC	470	
	1	SP49_HPgV_1 SP49_HPgV_2	GGTGGGCAATGAACCTACCTTT CTCCTCCGGTCTTTGAAGAAC	505	
	2	SP49_HPgV_3 SP49_HPgV_4	AGTTCATTGACTTTGGCGGA CTATTGACGAAGGGCACAGC	729	
SP49	3	SP49_HPgV_5 SP49_HPgV_6	TATATCTGATGAAGCTGGCGGA CCCTTACACCAGTTCTGTAGA	484	
	4	SP49_HPgV_7 SP49_HPgV_8	CCAAATCCTGTCCCACTATTGC TAGCTGAGTACAAGACCAACCC	967	
	5	SP49_HPgV_9 SP49_HPgV_10	GCAGTAGGAATGCTCGTGCC GTCCTTGATGATGGAAGTGTCTT	920	
	6	SP49_int_1 SP49_int_2	TCCTGTCCCACTATTGCTGAG TACAAGACCAACCCAGGGC	954	
	SP56	1	SP56-1-F SP56-1-R	ACGTGTGCACTAGAGTGTGA TGAGAGGGGCCACTGATTTT	474
		2	SP56-2-F SP56-2-R	TTTGGAGATGGACTGAACAGC CAGTGGGTGATAGGGTCGC	403
SP58	1	SP58-1-F SP58-1-R	TCTGAAAGTGCCGTTTGAC AGCCATAGACATACTGGGGAC	405	
	2	UTR-F UTR-R	TTGGTAGCCTATAGGTGGG CCTGATACAGTGGCCAGCAT	593	
	3	SP58-3-F SP58-3-R	CCGTTTGACTTCTGGCGTGG ACAAGAGAGGGGCCACTGATT	375	
	4	UTR-2-F UTR-2-R	TAGCCACTATAGGTGGGTCTTAAG TGATACAGTGGCCAGCAT	550	

Tabla Suplementaria S10. Resultados obtenidos del análisis por Centrifuge del experimento piloto llevado a cabo en el capítulo 1. Para cada tratamiento de extracción (extracción directa vs centrifugación alta) y filtración (0,45 µm vs 1 µm), se prepararon bibliotecas de secuenciación de forma independiente para las muestras de ADN y ARN.

Tabla Suplementaria S11. Información demográfica de los donantes de plasma incluidos en cada mezcla del capítulo 1.

ND: No disponible

Mezcla	Sexo	Edad	Mezcla	Sexo	Edad
P1	Hombre	ND	P6	Mujer	43
	Hombre	ND		Mujer	23
	Hombre	ND		Hombre	23
	Hombre	ND		Mujer	58
	Hombre	ND		Hombre	26
	Hombre	ND		Hombre	30
	Mujer	ND		Mujer	18

	Mujer	ND		Mujer	24
	Mujer	ND		Mujer	19
	Mujer	ND		Mujer	46
	Hombre	53		Mujer	47
	Hombre	45		Mujer	59
	Hombre	43		Mujer	55
	Hombre	56		Hombre	50
P2	Hombre	36	P7	Mujer	54
	Hombre	31		Hombre	29
	Hombre	37		Mujer	55
	Hombre	19		Hombre	42
	Hombre	53		Mujer	45
	Hombre	54		Hombre	57
	Hombre	56		Hombre	57
	Hombre	44		Mujer	34
	Hombre	41		Hombre	55
	Hombre	51		Hombre	43
P3	Hombre	58	P8	Hombre	32
	Hombre	32		Hombre	46
	Hombre	54		Hombre	53
	Mujer	60		Mujer	56
	Mujer	45		Hombre	24
	Hombre	40		Hombre	63
	Mujer	21		Hombre	57
	Hombre	20		Mujer	31
	Mujer	55		Hombre	45
	Hombre	32		Mujer	23
P4	Mujer	38	P9	Mujer	43
	Hombre	18		Mujer	31
	Mujer	18		Hombre	38
	Hombre	19		Mujer	55
	Mujer	18		Hombre	57
	Mujer	19		Hombre	ND
	Mujer	18		Hombre	51
	Mujer	19		Hombre	48
	Mujer	18		Hombre	46
	Mujer	18		Hombre	51
P5	Mujer	18	P10	Mujer	55
	Mujer	18		Mujer	49
	Mujer	19		Mujer	48
	Hombre	34		Mujer	59
	Hombre	31		Mujer	47
	Mujer	26		Hombre	ND
	Mujer	43		Hombre	31
	Mujer	23		Mujer	23
	Hombre	23		Hombre	51
	Mujer	58		Mujer	24
P6	Hombre	26	P11	Hombre	42
	Hombre	30		Hombre	42
	Mujer	18		Hombre	39
	Mujer	24		Mujer	42
	Mujer	19		Hombre	54
	Mujer	46		Mujer	53
	Mujer	47		Mujer	59
	Mujer	59		Mujer	56
	Mujer	55		Mujer	42
	Hombre	50		Hombre	60
P7	Mujer	54	P12	Mujer	51
	Hombre	29		Mujer	50
	Mujer	55		Mujer	53
	Hombre	42		Mujer	52
	Mujer	45		Mujer	54
	Hombre	57		Mujer	50

Tabla Suplementaria S12. Resumen de los resultados de Centrifuge para las 12 mezclas de plasma analizadas en el capítulo 1. Para cada mezcla, se indica el número total de lecturas que han pasado los análisis de filtrado/recorte, y las lecturas clasificadas como humanas, bacterianas, anelovirus, otros virus, metazoos, *Viridiplantae*, hongos y arqueas.

Pool	Número lecturas totales	% lecturas clasificadas	Lecturas humano	Lecturas bacteria	Lecturas Anelovirus	Otros virus	Metazoa	Viridiplantae	Fungi	Archaea
P1	835.504	99,1	14.363	55.632	147.618	49.032	205.681	59.302	12.214	1.928
P2	1.948.103	96,0	55.003	237.092	126.117	24.170	544.766	270.667	69.736	2.480
P3	1.008.317	96,6	26.916	120.081	117.659	19.515	265.499	125.575	25.382	2.995
P4	1.909.824	95,7	43.516	230.218	139.537	22.437	518.405	256.947	52.663	4.445
P5	2.420.824	96,6	33.118	355.178	17.461	42.128	720.739	402.583	41.911	4.787
P6	2.706.682	93,2	48.748	328.786	7.882	5.744	1.050.021	278.520	74.294	1.996
P7	1.771.038	95,6	41.209	181.139	339.778	16.529	428.762	217.249	35.940	3.028
P8	1.712.606	95,5	43.062	191.273	188.955	32.940	458.886	199.406	42.769	4.524
P9	2.042.540	95,8	34.865	215.035	493.302	34.481	373.521	268.978	38.457	3.065
P10	1.541.190	95,7	46.055	169.011	155.854	13.926	391.019	230.288	27.651	4.127
P11	2.390.182	95,6	45.404	349.863	29.247	28.671	565.804	353.213	76.877	4.078
P12	1.935.479	94,9	47.649	224.878	60.981	23.457	554.726	318.607	64.112	4.001

Tabla Suplementaria S13. Resultados de la clasificación taxonómica bacteriana mediante Centrifuge para los controles y las muestras del capítulo 1. También se proporcionan los valores de Recentrifuge para las muestras. Las lecturas se proporcionan a nivel de filo, clase y orden.

Tabla Suplementaria S14. Resultados de la clasificación taxonómica viral mediante Centrifuge para los controles y las muestras del capítulo 1. También se proporcionan los valores del análisis mediante Recentrifuge para las muestras. Se muestra el número de lecturas totales, eucariotas, humanas, bacterianas y virales en las muestras y en las muestras controles.

Tabla Suplementaria S15. Matriz de identidades nucleotídicas por pares obtenida mediante el alineamiento de la secuencia nucleotídica de la ORF1 de los genotipos representativos de TTV descargados y las secuencias virales asignadas a este género obtenidas en el capítulo 1.

Tabla Suplementaria S16. Matriz de identidades nucleotídicas por pares obtenida mediante el alineamiento de la secuencia nucleotídica de la ORF1

de los genotipos representativos de TTMV descargados y las secuencias virales asignadas a este género obtenidas en el capítulo 1.

Tabla Suplementaria S17. Matriz de identidades nucleotídicas por pares obtenida mediante el alineamiento de la secuencia nucleotídica de la ORF1 de los genotipos representativos de TTMDV descargados y las secuencias virales asignadas a este género obtenidas en el capítulo 1.

Tabla Suplementaria S18. Información demográfica de los donantes de plasma incluidos en cada mezcla del capítulo 2.

M: mujer; H: hombre; ND: No disponible

Mezcla	Sexo	Edad	Mezcla	Sexo	Edad	Mezcla	Sexo	Edad
SP1	M	27	SP21	H	44	SP41	H	26
	H	33		M	18		M	20
	M	25		M	21		M	21
	M	41		H	32		H	22
	M	23		H	19		M	22
	M	55		H	20		M	19
	H	57		H	49		M	19
	H	28		H	59		M	21
	M	48		M	58		M	53
	M	32		H	40		M	20
M	23	M	21	H	18			
SP2	M	18	SP22	M	21	SP42	H	18
	M	22		M	24		H	21
	M	22		M	19		H	21
	M	25		H	18		M	22
	M	24		M	19		M	19
	M	21		M	19		M	20
	H	57		M	18		M	20
	M	25		H	18		M	19
	H	18		M	20		H	19
	M	18		M	57		H	54
SP3	M	35	SP23	H	18	SP43	M	41
	H	53		M	20		M	57
	H	59		H	20		H	22
	M	40		H	18		M	47
	H	42		H	18		M	43
	M	24		H	20		H	60
	H	34		H	18		M	46
	H	52		M	18		H	55
	M	52		M	18		M	19
	H	53		M	21		H	44
SP4	H	52	SP24	M	35	SP44	H	39
	M	52		H	19		H	40
	M	52		M	18		H	37
	H	53		M	19		H	32
	M	23		H	22		M	41
	H	21		H	19		M	33
	M	20		H	18		H	35
	H	29		M	20		M	48

Anexo

	M	19		M	22		H	37
	M	19		H	20		M	19
	M	18		H	21		M	57
	M	21		H	22		M	28
	H	24	SP25	M	18	SP45	H	58
	H	20		M	18		H	53
	M	18		M	18		M	61
	H	21		M	35		M	38
	M	20		M	22		M	22
	M	24		H	42		H	40
SP5	M	19	M	61		H	36	
	M	21	H	55		H	51	
	H	21	M	50		M	53	
	M	18	H	47	SP46	H	27	
	M	24	H	59		H	43	
	M	31	M	20		H	50	
	M	18	M	24		M	43	
	M	20	M	20		M	39	
M	47	M	20	H		53		
H	18	M	23	H		43		
H	21	M	27	H		41		
SP6	H	18	H	19		H	44	
	H	21	M	18	SP47	H	43	
	M	18	H	18		M	56	
	H	21	H	18		H	42	
	M	19	H	18		H	34	
	H	21	H	21		M	43	
	H	21	H	40		M	37	
	H	54	H	37		H	62	
M	48	H	43	H		40		
SP7	H	33	M	58		H	20	
	H	40	M	56		M	18	
	M	27	H	33	SP48	H	20	
	H	52	H	49		H	19	
	H	32	M	22		H	20	
	H	45	M	19		H	18	
	H	32	H	50		M	18	
	H	31	M	18		H	18	
H	40	H	26	H		18		
H	51	H	31	H		18		
SP8	M	48	M	44		H	18	
	H	54	H	20	SP49	H	21	
	H	20	M	21		H	19	
	H	53	H	54		H	19	
	H	57	H	23		H	19	
	H	41	M	38		H	18	
	H	43	H	20		H	19	
	H	61	H	19		H	44	
H	52	H	18	M		18		
SP9	H	46	H	58		M	18	
	H	50	M	26		H	54	
	H	50	M	66	SP50	H	21	
	H	42	H	58		M	18	
	H	23	H	45		H	18	
	H	46	H	36		H	18	
	M	29	H	48		M	20	
	M	48	M	18		H	19	
H	47	M	19	H		18		
M	49	M	18	H		21		
SP10	H	46	M	18		H	53	
					SP51			

	M	47		M	23		H	60
	H	33		M	23		H	39
	M	23		M	20		H	55
	H	21		H	24		H	49
SP11	M	40	SP32	H	26	SP52	H	51
	H	27		M	19		H	57
	H	37		M	23		M	19
	M	41		M	25		M	47
	M	21		M	23		H	51
	M	22		M	18		H	47
	M	20		M	49		H	39
	M	22		M	22		H	42
SP12	M	28		M	35		H	37
	M	18		M	21		H	27
	M	30		M	18		M	46
	M	18		M	20		H	25
	M	18	SP33	M	23	SP53	M	20
	H	35		M	22		M	21
	M	22		M	19		M	26
	M	19		H	31		H	33
M	18	H		66	M		24	
M	18	M		43	M		23	
		H		53	M		22	
		M		46	H		19	
SP13	M	23		H	45		M	23
	H	22		H	29		M	20
	M	41	SP34	H	22	SP54	M	26
	M	21		H	41		M	28
	M	20		M	19		M	23
	M	18		H	18		M	18
	H	22		H	52		M	18
	M	48		H	23		M	19
		M		23	M		21	
		H		23	M		19	
SP14	M	18		M	18		M	22
	M	20		M	21		H	46
	M	24		M	21		M	46
	H	36	SP35	H	21	SP55	M	40
	M	18		H	21		H	51
	H	26		H	60		H	42
	H	18		M	19		H	51
	M	32		M	19		H	43
H	20	H		46	M		54	
H	19	M		18	M		52	
		M		23	M		55	
SP15	M	19		M	19		M	27
	M	19	SP36	M	19	SP56	M	37
	H	18		M	18		M	41
	M	18		M	18		M	56
	H	26		M	19		M	24
	M	22		M	ND		H	60
		M		19	M		43	
SP16	H	37		M	20		M	38
	M	53		M	19		M	39
	H	34		H	24		H	39
	H	42	SP37	H	47	SP57	M	21
	H	38		M	48		H	18
	M	50		H	32		M	21
	H	47		M	19		H	24
	H	22		H	21		M	22
H	26							

Anexo

	M	27		M	46		H	21
	M	19		H	45		H	23
	H	22		H	39		H	22
	M	18		M	38		H	23
	H	18		M	20		H	22
SP17	H	18		H	56		H	23
	H	20		H	19		H	19
	M	23	SP38	H	20		M	19
	M	24		M	21		H	18
	M	23		H	19	SP58	M	18
	M	49		H	22		H	20
	M	19		M	18		H	18
	H	26		H	21		H	24
M	21	H		18	M		31	
H	33	H		19	H		18	
M	52	H	39		M		21	
H	30	M	18		M		46	
SP18	H	18	SP39	M	18		M	22
	H	19		H	21		M	22
	H	19		H	19	SP59	M	20
	M	19		M	18		M	55
	H	20		H	18		H	23
	M	20		H	18		M	18
	H	19		M	18		M	42
	H	19		M	18		H	20
M	19	M	19		M		54	
M	28	M	19	SP60	H		48	
M	44	M	18		M	18		
H	19	M	18		M	18		
M	20	M	18		H	30		
M	20	M	18		M	24		
H	18	M	41		H	21		
M	28	M	20		H	18		
M	31	M	22		H	18		
SP20	M	23				M	50	
	M	27						
	M	18						
	M	21						
	M	19						
	M	18						
	M	19						
	H	18						

Tabla Suplementaria S19. Resultados de la clasificación taxonómica viral mediante Centrifuge para los controles y las muestras. También se proporcionan los valores del análisis mediante Recentrifuge para las muestras. Se muestra el número de lecturas totales, eucariotas, humanas, bacterianas y virales presentes en las mezclas y en las muestras control.

Tabla Suplementaria S20. Resumen de los resultados de la clasificación taxonómica de las 60 mezclas analizadas. Para cada mezcla, se proporciona el número total de lecturas y las clasificadas por Centrifuge/Recentrifuge.

También se proporcionan las lecturas bacterianas y las clasificadas como pertenecientes a virus controles, anelovirus, HPgV y otros virus. La última columna indica el número de *contigs* de anelovirus detectados mayores de 1,5 kb.

Tabla Suplementaria S21. Resumen de los mejores resultados de BLASTp para las seis ORF putativas encontradas en la nueva secuencia de Microvirus encontrada en el capítulo 2.

ORF	Longitud (aa)	Mejor resultado (Número acceso)	Query cover (%)	Identidad (%)
ORF1	531	Major capsid protein (AXH73967)	100	60,3
ORF2	319	Replication initiator protein (AXH73970)	100	54,1
ORF3	224	DNA pilot protein (QJB19348)	85	49,5
ORF4	157	Internal scaffolding protein (AXH73969)	94	48,3
ORF5	97	Non-estructural protein (QJB19603)	80	48,7
ORF6	39	Hypothetical protein (QJB18752)	100	61,5

Tabla Suplementaria S22. Matriz de identidades nucleotídicas por pares obtenida con la ORF1 de las secuencias del género TTV pertenecientes a las especies de referencia y secuencias descritas en el capítulo 1.

Tabla Suplementaria S23. Matriz de identidades nucleotídicas por pares obtenida con la ORF1 de secuencias del género TTMV pertenecientes a las especies de referencia y secuencias descritas en el capítulo 1.

Tabla Suplementaria S24. Matriz de identidades nucleotídicas por pares obtenida con la ORF1 de secuencias del género TTMDV pertenecientes a las especies de referencia y secuencias descritas en el capítulo 1.

Tabla Suplementaria S25. Matriz de identidades nucleotídicas por pares obtenida utilizando la secuencia nucleotídica de la ORF1 de las secuencias pertenecientes al género TTV de las especies de referencia, a posibles nuevas especies obtenidas de la reevaluación de los datos del capítulo 1 y a las secuencias descritas en el capítulo 2.

Tabla Suplementaria S26. Matriz de identidades nucleotídicas por pares obtenida utilizando la secuencia nucleotídica de la ORF1 de las secuencias pertenecientes al género TTMV de las especies de referencia, a posibles

nuevas especies obtenidas de la reevaluación de los datos del capítulo 1 y a las secuencias descritas en el capítulo 2.

Tabla Suplementaria S27. Matriz de identidades nucleotídicas por pares obtenida utilizando la secuencia nucleotídica de la ORF1 de las secuencias pertenecientes al género TTMDV de las especies de referencia, a posibles nuevas especies obtenidas de la reevaluación de los datos del capítulo 1 y a las secuencias descritas en el capítulo 2.

Tabla Suplementaria S28: Resumen de los mejores resultados de BLASTp para las cinco ORFs encontradas en la secuencia del nuevo microvirus descrito en el capítulo 3.

ORF	Longitud (aa)	Mejor resultado (Número acceso)	Query cover (%)	Identidad (%)
ORF1	554	Major capsid protein (DAM92865.1)	100	89,98
ORF2	302	Minor capsid protein (DAW54564.1)	100	96,69
ORF3	339	Replication associated protein (DAF65613.1)	100	99,71
ORF4	129	Peptidase (DAF65617.1)	100	96,12
ORF5	172	Scaffold protein (DAW54566.1)	76	87,88

Tabla Suplementaria S29: Matriz de identidades nucleotídicas por pares empleando la secuencia nucleotídica de la ORF1 de las secuencias descritas en el capítulo 3 y que pertenecen al género TTV junto a las especies de referencia de este género según ICTV y las posibles nuevas especies descritas en el capítulo 2.

Tabla Suplementaria S30: Matriz de identidades nucleotídicas por pares empleando la secuencia nucleotídica de la ORF1 de las secuencias descritas en el capítulo 3 y que pertenecen al género TTMV junto a las especies de referencia de este género según ICTV y las posibles nuevas especies descritas en el capítulo 2.

Tabla Suplementaria S31: Matriz de identidades nucleotídicas por pares empleando la secuencia nucleotídica de la ORF1 de las secuencias descritas en el capítulo 3 y que pertenecen al género TTMDV junto a las especies de referencia de este género según ICTV y las posibles nuevas especies descritas en el capítulo 2.

Tabla Suplementaria S32: Resumen de los datos obtenidos en el análisis de recombinación de los genomas de anelovirus en el capítulo 3. En la tabla, para cada suceso de recombinación detectado por RDP4 (Martin et al., 2015), se muestra la secuencia recombinante, las secuencias parentales, las posiciones del punto de ruptura donde ha ocurrido el suceso de recombinación y los métodos de detección que soportan ese suceso. En la última columna, aparece el porcentaje de topologías de árbol sin resolver, procedente del análisis de los mapas de probabilidad con los que se ha contrastado la confianza de cada suceso de recombinación.

Figura Suplementaria S1. Árbol filogenético que incluye la ORF1 de las secuencias pertenecientes a la familia *Anelloviridae* descritas en el capítulo 1. Se incluyen todas las secuencias descargadas de las bases de datos y las descritas en el capítulo 1 (etiquetadas en rojo). Los rangos de valores *bootstrap* de 0,7-0,85 y 0,85-1,0 se indican con círculos azules y rojos, respectivamente. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

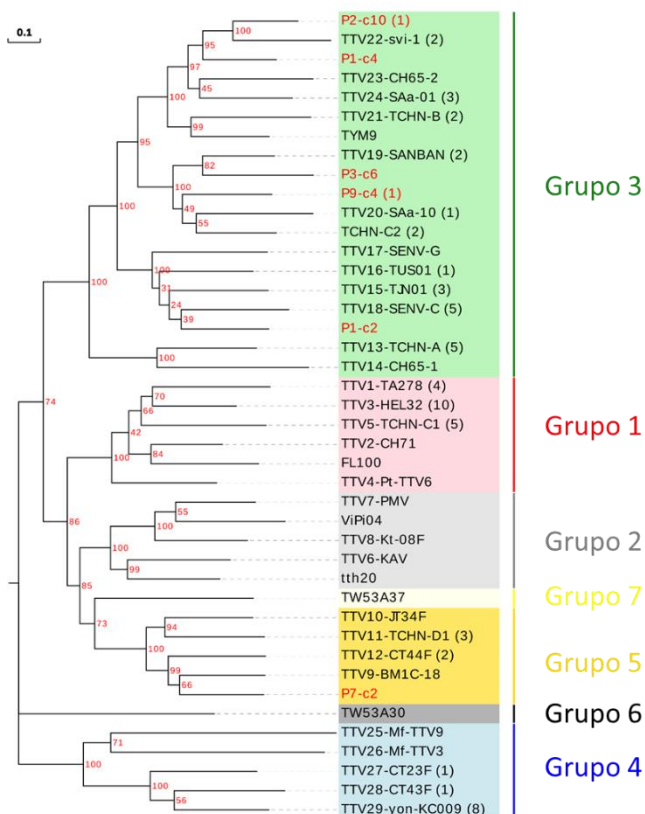


Figura Suplementaria S2. Árbol filogenético que incluye los genotipos representativos del género TTV. Las secuencias descritas en el capítulo 1 que pueden considerarse como nuevas especies también se incluyen y se etiquetan en rojo. Los números entre paréntesis indican el número de nuevas secuencias que se agrupan con un genotipo específico (todas las nuevas secuencias se muestran explícitamente en la Figura 10). Los valores de *bootstrap* se indican en rojo en cada punto de la rama. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

Figura Suplementaria S3. Árbol filogenético que incluye la ORF1 de las secuencias pertenecientes a la familia *Anelloviridae* descritas en el capítulo 2. Se incluyen todas las secuencias descargadas de las bases de datos y las descritas en el capítulo 2 (etiquetadas en rojo). Los rangos de valores *bootstrap*

de 0,7-0,85 y 0,85-1,0 se indican con círculos azules y rojos, respectivamente. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

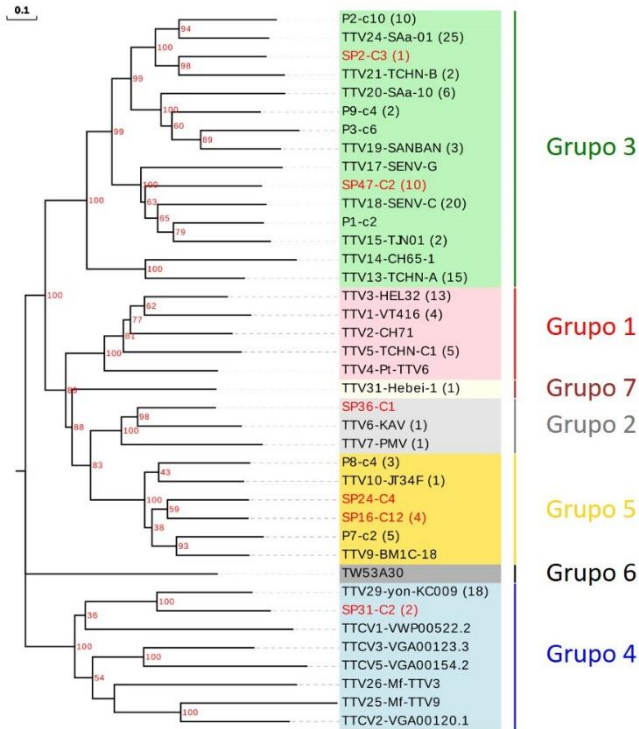


Figura Suplementaria S4. Árbol filogenético que incluye las especies de referencia del género TTV y las posibles nuevas especies descritas en el capítulo 1. Las secuencias descritas en el capítulo 2 que pueden considerarse como nuevas especies también se incluyen y se han etiquetado en rojo. Los números entre paréntesis indican el número de nuevas secuencias que se agrupan con un genotipo específico (todas las nuevas secuencias se muestran explícitamente en la Figura 13). Los valores de *bootstrap* se indican en rojo en cada punto de la rama. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

Figura Suplementaria S5. Red filogenética de HPgV obtenida mediante SplitsTree4 (GTR + Γ + I, $\Gamma = 0,7028$, I = 0,5234). Las secuencias obtenidas en el capítulo 2 están etiquetadas con el prefijo HPgV-. El resto de secuencias, obtenidas de la base de datos del NCBI, e identificadas por su número de acceso, están etiquetadas con el genotipo y/o subtipo al que pertenecen, excepto las secuencias identificadas como recombinantes, que están etiquetadas con _R. Las ramas paralelas indican historias evolutivas alternativas.

Figura Suplementaria S6. Árbol filogenético que incluye la ORF1 de las secuencias pertenecientes a la familia *Anelloviridae* descritas en el capítulo 3. Se incluyen todas las secuencias descargadas de las bases de datos, las secuencias identificadas como potenciales especies nuevas en el capítulo 2 y las descritas en el capítulo 3 (etiquetadas en rojo). Los rangos de valores *bootstrap* de 0,7-0,85 y 0,85-1,0 se indican con círculos azules y rojos, respectivamente. La barra de escala indica la distancia evolutiva en sustituciones nucleotídicas por sitio.

