



Universidad de Valladolid

Máster

Gestión Forestal basada en Ciencia de Datos
Forest Management Based on Data Science
(DATAFOREST)

Metaviroma de bosques mediterráneos en decaimiento

Alumno: Pedro Luis Pérez Alonso
Directores: Dr. D. Sergio Díez Hermano
Dr. D. Julio Díez Casero

Septiembre 2023



ÍNDICE

RESUMEN	4
ABSTRACT	4
ABREVIATURAS	5
1.- INTRODUCCIÓN.....	6
1. Los virus en el ecosistema forestal	6
2. Virus forestales con capacidad “cross-kingdom”	6
3. Identificación bioinformática de virus mediante RdRp	7
2.- OBJETIVOS	11
3.- MATERIALES Y MÉTODOS	11
3.1. Muestreo.....	11
3.2. Aislamiento de hongos.....	12
3.3. Extracción y secuenciación de RNA.....	12
3.4. Análisis bioinformático.....	12
4.- RESULTADOS	13
4.1. Taxonomía y distribución de las RdRp	13
4.2. Filogenia de las RdRp	16
4.3. RdRps divergentes.....	17
5.- DISCUSIÓN	20
6.- CONCLUSIÓN	23
7.- AGRADECIMIENTOS.....	24
8.- BIBLIOGRAFÍA.....	25
TABLAS SUPLEMENTARIAS.....	31
CÓDIGO	32

RESUMEN

Los bosques son ecosistemas con complejas interacciones entre los organismos que los habitan, y los virus desarrollan un papel importante que pueden modificar estas relaciones. Entre ellos cabe destacar a los virus RNA, por su gran capacidad de moverse entre organismos de reinos diferentes (capacidad “cross-kingdom”). No obstante, la imagen actual del viroma de los bosques es muy incompleta, dada la escasez de estudios sobre el tema. Por tanto, el objetivo de este trabajo es describir la diversidad de virus RNA, centrándose en los bosques mediterráneos. Para ello, se recogieron muestras de árboles, hongos e insectos procedentes de 3 hábitats forestales de Castilla y León (*Castanea sativa*, *Pinus radiata* y *Quercus ilex*). Se asignó identidad a los virus a partir de la secuencia de aminoácidos de la RNA polimerasa dependiente de RNA (RdRp) y el programa “Palmscan”. De esta forma se han identificado 159 RdRps diferentes, el 54,7% en *Quercus ilex*, 26,4% en *Pinus radiata* y 18,9% en *Castanea sativa*. Las familias víricas más abundantes fueron *Botourmiaviridae*, *Partitiviridae*, *Narnaviridae* y *Mitoviridae*. Algunas de las familias identificadas, como *Botourmiaviridae* y *Tombusviridae*, podrían tener potencial “cross-kingdom”, ya que se encuentran presentes en diferentes reinos. No se pudo asignar categoría taxonómica a 26 de las 159 RdRps, lo que puede suponer una oportunidad para describir especies víricas desconocidas hasta el momento. Conocer la diversidad de la virosfera forestal permitirá identificar qué virus son más susceptibles de saltar las barreras entre reinos, así como aquellos que puedan ocasionar enfermedades en los ecosistemas en que se encuentran.

Palabras clave: cross-kingdom, virus, bosque mediterráneo, RdRp

ABSTRACT

Forests are ecosystems with complex interactions between organisms living there and viruses play an important role in modifying those relations. RNA viruses stand out among them, as they are particularly capable of jumping between organisms of different kingdoms (also known as cross-kingdom capacity). However, current knowledge on the forest virome is very incomplete due to the lack of studies in this area. Therefore, the objective of this work is to describe the RNA virus diversity focusing on Mediterranean forests. For this reason, trees, fungi, and insects' samples were collected from 3 different habitats in Castilla y León (*Castanea sativa*, *Pinus radiata* and *Quercus ilex*). Viral identity was assigned based on the amino acid sequence of the RNA-dependent RNA polymerase (RdRp) and the program “Palmscan”. Through this method we have been able to identify 159 different RdRPs, 54.7% from *Q. ilex*, 26.4% from *P. radiata* and 18.9% from *C. sativa*. Most abundant viral families were *Botourmiaviridae*, *Partitiviridae*, *Narnaviridae* and *Mitoviridae*. Some of the identified families such as *Botourmiaviridae* and *Tombusviridae* may have cross-kingdom potential since they were found in different kingdoms. It wasn't possible to assign taxonomic category to 26 of the 159 RdRps. This is an opportunity to describe previously unknown viral species. Understanding the diversity of the forest virosphere will allow us to identify which viruses are more susceptible to overcome the barriers between kingdoms, as well as those which may cause diseases in the ecosystems in which they are found.

Key words: cross-kingdom, virus, mediterranean forest, RdRp

ABREVIATURAS

- Asn: Asparagina
- BLAST: Basic Local Alignment Search Tool
- *C. sativa*: *Castanea sativa*
- DNA: Ácido desoxirribonucleico
- dNTP: trifosfato de desoxirribonucleótido
- ds: bicatenario
- HMM: Hidden Markov Models
- IA: Inteligencia artificial
- Mg: Magnesio
- Mn: Manganeso
- MP: Movement Protein
- mRNA: RNA mensajero
- NJ: Neighbour Joining
- NTPs: Trifosfato de ribonucleótido
- OTU: Unidad taxonómica operativa
- *P.radiata*: *Pinus radiata*
- PDA: Agar de patata y destroxa
- *Q.ilex*: *Quercus ilex*
- RdRp: RNA polimerasa dependiente de RNA
- RNA: Ácido ribonucleico
- SRA: Sequence Read Archive
- Ss: monocatenario

1.- INTRODUCCIÓN

1. Los virus en el ecosistema forestal

Los ecosistemas forestales desempeñan un papel fundamental en la provisión de diversos beneficios y servicios. Algunos de ellos son tangibles, como los recursos madereros, la micología, la caza, la piscicultura y los frutos del bosque. Sin embargo, también proporcionan beneficios intangibles, como la captación de carbono, la protección de los recursos hídricos y del suelo, el uso recreativo, los valores socioculturales y la fijación de la población en áreas rurales (Martin-García y Diez, 2020). Además de estos beneficios, las especies arbóreas sustentan ecosistemas completos, aumentan la biodiversidad y desempeñan una función vital para numerosos organismos (Ellison et al., 2005).

A pesar de la creciente conciencia ambiental y la atención prestada a la delicada situación de los bosques, la superficie forestal sigue disminuyendo a nivel global, siendo los bosques nativos los más afectados. Se estima que la presión que ejercen los humanos sobre los bosques es la causante de la desaparición de más de la mitad de estos en los últimos 6000 años, principalmente para usar los terrenos como áreas agrícolas o usar la madera como fuente de combustible (Roberts et al., 2018). Además, la acción del cambio climático, que altera las condiciones del hábitat forestal, y la globalización, que permite la entrada de nuevos patógenos con los que los bosques no han tenido la oportunidad de coevolucionar, han transformado nuestros ecosistemas naturales hasta su forma actual. Por ello las Naciones Unidas proclamó el año 2020 como el Año de la Sanidad Vegetal, con intención de concienciar sobre la importancia de proteger la salud de los ecosistemas vegetales y prevenir la propagación de plagas y enfermedades.

En los últimos años, el conocimiento sobre patógenos forestales se ha enriquecido con estudios sobre la diversidad de la virosfera de los bosques. Los virus son organismos infecciosos que necesitan de células vivas para poder reproducirse. Por ende, cualquier integrante biótico de un ecosistema forestal puede ser portador y verse afectado por los virus. Existen virus de plantas y de todos sus asociados, como animales, protistas, bacterias y hongos. Cada integrante presenta un viroma (su comunidad vírica), que en conjunto forman el metaviroma o virosfera del ecosistema donde habitan. Actualmente se conocen 64 especies víricas que afectan a plantas forestales de bosques templados y urbanos (Rumbou et al., 2021). También se han detectado algunos virus en insectos lepidópteros como *Lymantria dispar*, *Thaumetopoea pityocampa* y *Leucoma salicis*, en coleópteros como el escarabajo de la corteza *Ips typographus* y en himenópteros como *Neodiprion sertifer* (Pilarska et al., 2018). Los virus capaces de infectar hongos se denominan micovirus y poseen en su mayoría un genoma de RNA. Algunos de estos micovirus afectan al comportamiento de los propios hongos patógenos a los que infectan haciendo que sean mucho menos virulentos (hipovirulencia) al inhibir su crecimiento o alterar la producción de micotoxinas (Munkvold, 2017).

2. Virus forestales con capacidad “cross-kingdom”

Los flujos de virus que se dan entre las diversas especies y reinos ejercen un impacto en el equilibrio que existe en los bosques y puede llegar a afectar al ecosistema global (Balique et al., 2015). A los virus con capacidad de infectar gran diversidad de organismos e incluso saltar las barreras entre reinos biológicos, se los denomina virus inter-reino o “cross-kingdom”. En los ecosistemas forestales, esto implica principalmente a plantas, hongos e insectos (Lefeuvre et al., 2019). Estos virus “cross-kingdom” no son pocos, por

ejemplo, muchas de las especies fúngicas que se conocen son portadoras de algún virus vegetal (Cao et al., 2022).

Los saltos virales entre especies pueden alterar características esenciales de los organismos, como el nivel de patogenicidad ejercido sobre el huésped. Además de las conocidas vías de transmisión de virus de RNA entre animales, como el virus Nipah o los coronavirus, existen numerosos ejemplos de transmisión de virus entre especies y reinos (Afelt et al., 2018). Cada vez se presentan más evidencias de que tanto micovirus como fitovirus pueden afectar a un espectro de organismos más amplio de lo esperado. Los insectos son uno de los vectores más frecuentes de transmisión de los fitovirus a otros organismos. Se han documentado dos mecanismos al respecto, uno en el que el virus no atraviesa ninguna barrera del cuerpo del insecto (transmisión no persistente) y otro en el que se distribuye por su sistema circulatorio y se acumula en las glándulas salivares (transmisión circulatoria). En este caso, el fitovirus puede replicarse en el propio insecto e invadir diversos tejidos. Si se puede replicar en un insecto, el fitovirus también se puede considerar virus del propio insecto (Balique et al., 2015).

Se han encontrado homologías estructurales y genómicas entre fitovirus de la familia *Secoviridae* como el *Cowpea Mosaic virus* y picornavirus de animales como los poliovirus. Estas evidencias señalan a un posible ancestro común. Los fitovirus que se transmiten de manera circulatoria se clasifican en familias que contienen virus que infectan animales. La familia *Reoviridae* infecta plantas, insectos y animales y uno de sus géneros, *Rotavirus*, es conocido por ser uno de los causantes de gastroenteritis en humanos. Sabiendo que la distancia filogenética entre insectos y plantas es mucho mayor que la existente entre insectos y vertebrados, cobra fuerza la hipótesis de que algunos virus vegetales podrían llegar a infectar, incluso, a los seres humanos, directa o indirectamente por medio de otros organismos (Balique et al., 2015).

En cuanto a los micovirus, algunos géneros pertenecientes a las familias *Totiviridae*, *Partitiviridae*, *Endornaviridae* y *Chrysoviridae* son virus vegetales que se replican en las células meristemáticas y son capaces de infectar hongos (H. Liu et al., 2012). El género *Genomovirus* constituye el único ejemplo conocido de virus DNA (circular) capaz de replicarse tanto en insectos como en hongos (S. Liu et al., 2016; Nerva et al., 2017).

Desde el punto de vista mecanístico, existen unas proteínas exclusivas de virus vegetales denominadas “Movement Proteins” (MP) que están presente solo en ciertos virus. Estas MP permiten la difusión de las partículas víricas entre células atravesando la pared celular. Los micovirus no tienen estas MP porque no suelen presentar una fase extracelular. Una actividad sinérgica entre los micovirus y las MP permiten que el micovirus use las MP para distribuirse por toda la planta. Una coinfección entre dos virus permite que se puedan usar las estructuras que crea el otro para diseminarse por el organismo. Esto permite que los virus puedan infectar organismos que sin ayuda no podrían infectar (Bian et al., 2020). La presencia de MPs en el genoma de los fitovirus, por tanto, podría interpretarse como una señal de la capacidad del virus de favorecer la ocurrencia de eventos “cross-kingdom”.

3. Identificación bioinformática de virus mediante RdRp

En 1917 David Baltimore creó una clasificación para agrupar los virus en familias según su tipo de genoma (Baltimore, 1971) atendiendo si este era DNA o RNA, si es monocatenario (“single stranded” o ss) o bicatenario (“double stranded” o ds), su polaridad (+, -) y por último si usa retrotranscriptasa (Figura 1). Según la hipótesis ampliamente aceptada, el ciclo de replicación basado únicamente en RNA es anterior a la replicación basada en el DNA y a la transcripción inversa (Koonin et al., 2015). Por ende, estos virus

para replicarse dependen de una RNA-polimerasa dependiente de RNA (RdRp) que permite catalizar la síntesis de una nueva hebra de RNA a partir de otra hebra de RNA que actúa como molde.

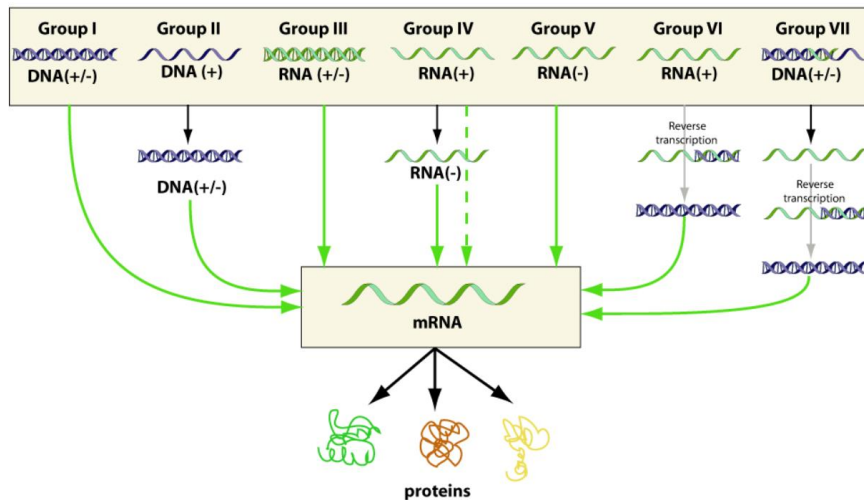


Figura 1. Clasificación de Baltimore. Se muestra el proceso de cada grupo, con su tipo de material genético, que conduce a la síntesis del mRNA necesario para traducirse en proteína. Obtenido de (Baltimore classification ~ ViralZone, s. f.)

La RdRp es la única proteína viral distintiva que se conserva universalmente en todos los virus de RNA, lo que permite reconstruir filogenias a partir de motivos conservados en su secuencia de aminoácidos. La RdRp está lejanamente relacionadas con otras enzimas como la DNA polimerasa dependiente de RNA o las transcriptasas inversas. La única familia de RdRps que se conoce en células eucariotas se encarga de la amplificación de RNA pequeño de interferencia (siRNA) (Koonin et al., 2015).

La alta conservación que existe en todas las RdRps permite que se use como método de identificación y clasificación de virus de RNA. La estructura de la RdRp (Figura 2) se asemeja a una mano y presenta 3 subdominios denominados dedos, palma y pulgar. La palma es la subunidad que mejor se encuentra conservada en todas las RdRps. Está formada por una hoja beta antiparalela de cuatro hebras con dos hélices alfa en la que encontramos 3 motivos. El motivo A y C están yuxtapuestos espacialmente y unen Mg^{2+} o Mn^{2+} que son esenciales para la catálisis. El motivo B es una larga hélice α que contiene un residuo de asparagina (Asn) que permite discriminar entre dNTPs y NTPs determinando si sintetiza DNA o RNA. Estos 3 motivos son indispensables para que la polimerasa pueda realizar correctamente su función. En la palma también existen otros dos motivos, nombrados D y E, pero al no encontrarse en todos los virus de RNA, a efectos de clasificación solo se usan los motivos A, B y C (Charon et al., 2022; Gorbalenya et al., 2002).

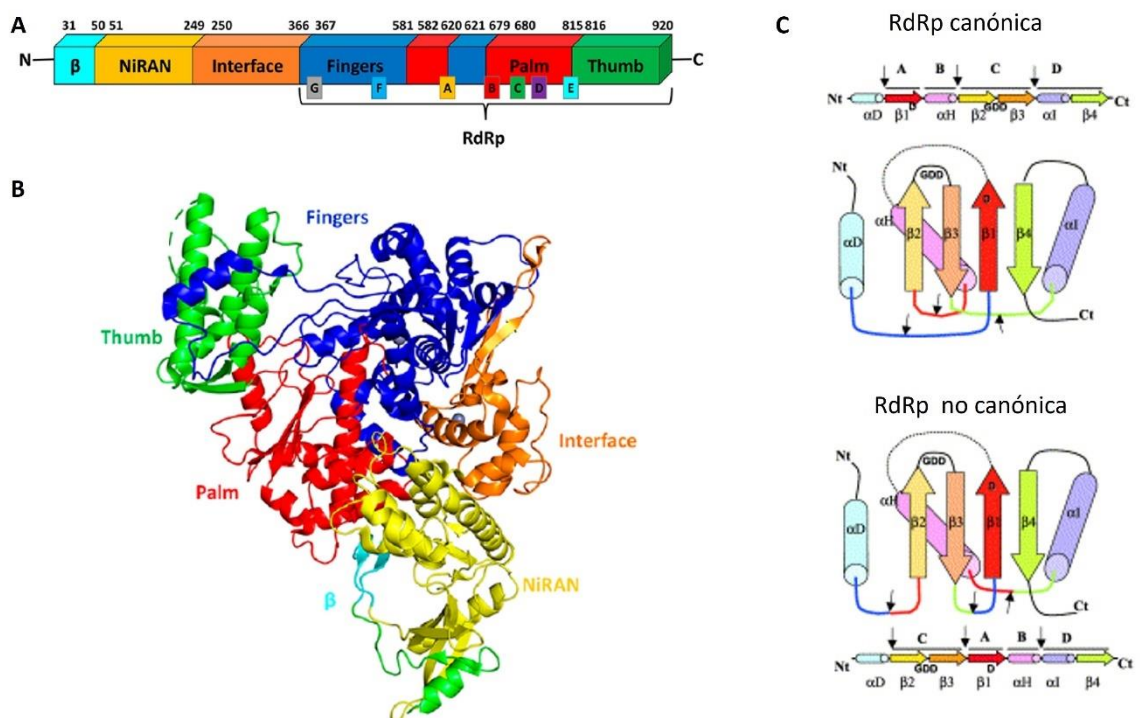


Figura 2. Estructura de la RdRp. A) Estructura lineal de las regiones y motivos conservados SARS-CoV-2 RdRp. B) Estructura 3D, los diferentes colores resaltan cada dominio. C) Organización lineal y terciaria de los motivos de secuencia y de los principales elementos de la estructura secundaria del dominio de la palma de una RdRp canónica (arriba) y una RdRp no canónica (abajo). Modificado de (Gorbalenya et al., 2002; Wakchaure et al., 2020).

El análisis bioinformático de estas secuencias ha experimentado un gran avance tras la irrupción de nuevas herramientas basadas en la Inteligencia Artificial (IA) y el “machine learning” (Lefin et al., 2023). Estos algoritmos están diseñados para detectar patrones en los datos permitiendo predecir y clasificar los nuevos datos que se van introduciendo. El entrenamiento de estos algoritmos requiere de una gran cantidad de ejemplos y su auge en el área de la genética es innegable, gracias a la generación masiva de datos mediante herramientas de secuenciación de nueva generación.

El proyecto Serratus es una de estas aplicaciones que usan “machine learning” para descubrir nuevos virus a partir de los motivos conservados en la secuencia primaria de las RdRp (Edgar et al., 2022). Mediante el análisis de 5.7 millones de experimentos disponibles en las bases de datos genómicas, lo que suponía 10.2 petabases de datos de secuenciación, lograron identificar 400 mil posibles nuevos virus (Figura 3). También permitió confirmar que el uso de la RdRp era idóneo, tanto como marcador evolutivo, como para asignar una identidad a los virus. El proyecto Serratus se basó en programas como PalmScan, que también usa “machine learning” y está entrenada para identificar virus mediante las regiones conservadas de la RdRp (Babaian & Edgar, 2022). Basados en esta filosofía han surgido nuevos algoritmos, como RdRp-scan (Charon et al., 2022), que permiten identificar secuencias de RdRp divergentes o menos convencionales. Este tipo de algoritmos permiten obtener mejores resultados que las técnicas de alineamiento o reconocimiento de patrones habituales como pueden ser BLAST (“Basic Local Alignment Search Tool”) (McGinnis & Madden, 2004) o HMM (“Hidden Markov Models”) (Yoon, 2009).

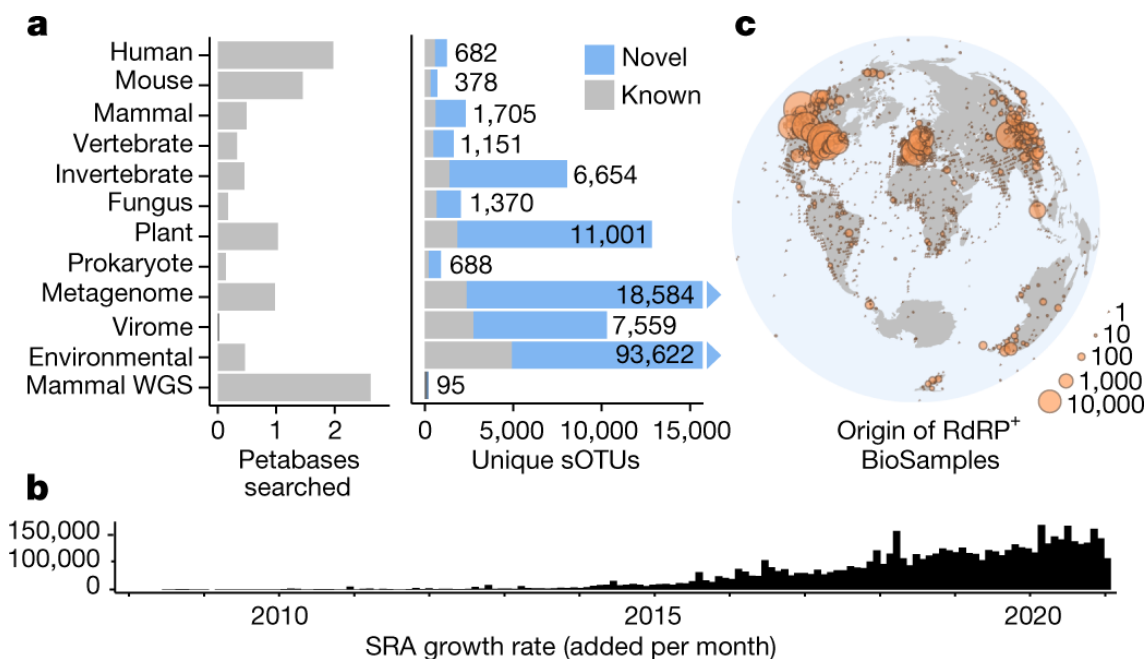


Figura 3. RdRps descubiertas con Serratus. A) Cantidad de datos analizados en las secuencias de la base de datos pública Sequence Read Archive (SRA) en la búsqueda de RdRps virales, agrupadas por taxonomía de las muestras y número de Unidades taxónomicas operativas (OTUs) conocidas y descubiertas. B) Fecha de publicación de los estudios que se analizaron en el proyecto Serratus. C) Procedencia de las muestras analizadas. Obtenido de (Edgar et al., 2022)

Todos estos aspectos subrayan las facilidades que existen hoy en día para estudiar la virosfera de los bosques, que ha sido tradicionalmente olvidada o excesivamente compartimentalizada. Por ello, en este trabajo, a partir del análisis de las RdRps, se busca describir la mayor diversidad posible de virus RNA con potencial de saltar entre reinos dentro de los ecosistemas forestales mediterráneos. Esto permitirá no solo entender cuál puede ser el impacto de los virus “cross-kingdom” en la salud de los bosques, sino también mejorar la comprensión sobre la propagación de enfermedades infecciosas emergentes, que suponen un desafío en la ciencia y la salud mundial actuales.

2.- OBJETIVOS

El objetivo general de este trabajo es describir la diversidad de virus RNA en ecosistemas forestales mediterráneos. Se articula en dos objetivos específicos:

- Objetivo específico 1: Reconstruir una filogenia de virus en bosques mediterráneos
- Objetivo específico 2: Estimar la proporción de virus desconocidos en bosques mediterráneos

3.- MATERIALES Y MÉTODOS

3.1. Muestreo

Se delimitaron áreas de muestreo en 3 hábitats diferentes de bosque mediterráneo (Figura 4):

- Bosque de frondosas perennes: *Quercus ilex* de La Alamedilla (Salamanca)
- Bosque de frondosas caducifolias: *Castanea sativa* de Las Médulas (León)
- Bosque de coníferas: *Pinus radiata* de Villafranca del Bierzo (León)

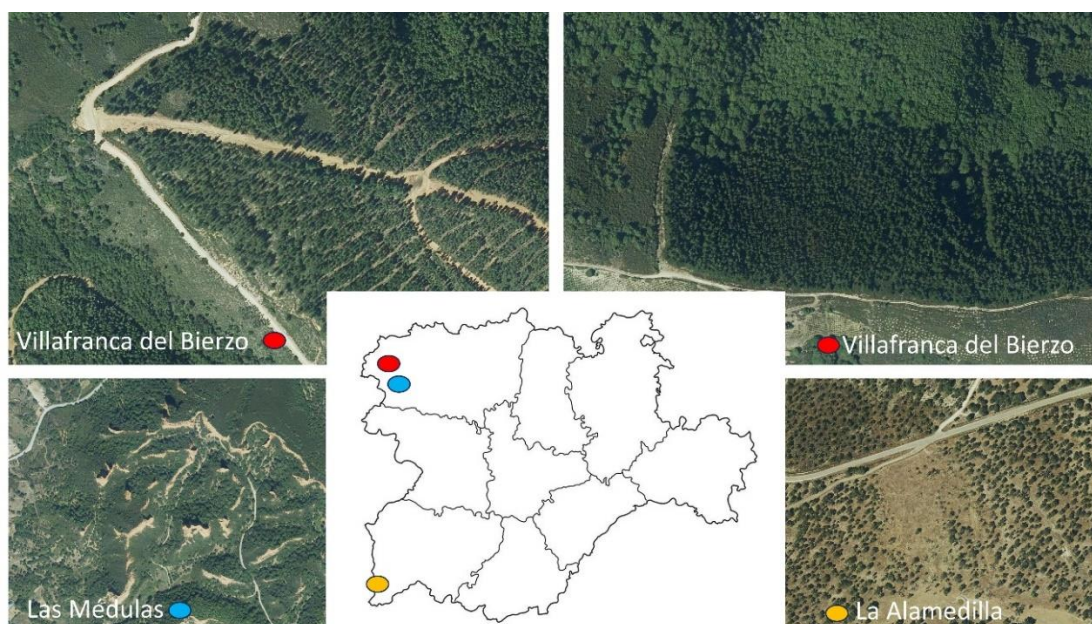


Figura 4. Zonas de muestreo. En rojo, Villafranca del Bierzo, donde se obtuvo las muestras de *Pinus radiata* (fotos superiores, dos sitios de muestreo). En azul, Las Médulas, donde se obtuvieron las muestras de *Castanea sativa* (inferior izquierda). En naranja, La Alamedilla, donde se obtuvo las muestras de *Quercus ilex* (inferior derecha). Ortofotos obtenidas de (Centro de Descargas del CNIG (IGN), s. f.)

En cada uno de ellos, se obtuvieron muestras de hoja, rama, corteza y madera de 16 árboles, seleccionados al azar en transectos lineales (La Alamedilla y Villafranca) o circulares (Las Médulas). También se recogieron de 5 a 6 individuos vivos de diversas familias de insectos comunes (hormigas, saltamontes, escarabajos). Las muestras se almacenaron a 4°C hasta su procesado, entre 1 y 2 días después de su recolección. El muestreo fue llevado a cabo por el equipo del profesor Julio Diez Casero (Universidad de Valladolid) con anterioridad a este trabajo, en los meses de junio y julio de 2021.

3.2. Aislamiento de hongos

Parte del material vegetal recogido se almacenó a -80°C , mientras que otra parte se troceó y esterilizó y se dispuso en placas de Petri con medio PDA + ampicilina ($100\ \mu\text{g/ml}$). Las placas se cultivaron en estufa a 36°C durante 3 días. Se aislaron colonias fúngicas visualmente diferentes y se almacenaron a -80°C . Los insectos se lavaron en una solución Ringer y se almacenaron a -80°C . $50\ \mu\text{l}$ de la solución de lavado se extendieron en placas con PDA + ampicilina y se cultivaron en estufa a 36°C durante 3 días. Se aislaron colonias fúngicas visualmente diferentes y se almacenaron a -80°C . El aislamiento de hongos fue llevado a cabo por el equipo del profesor Julio Diez Casero (Universidad de Valladolid) con anterioridad a este trabajo.

3.3. Extracción y secuenciación de RNA

Las muestras se pulverizaron y se extrajo RNA total con el kit Spectrum Plant (Sigma Aldrich, MO, USA). Todas las extracciones superaron un control de calidad mediante Qubit (RIN > 7). Las extracciones se combinaron por sitio de muestreo y tipo de muestra (árboles, hongos aislados de árbol, insectos y hongos aislados de insectos), haciendo un total de 12 muestras finales que se enviaron para su secuenciación a Macrogen (Corea del Sur). La secuenciación se llevó a cabo en equipos Illumina MiSeq mediante librerías TruSeq y depleción de RNA ribosomal, con lecturas pareadas de 150 pb. La extracción de RNA fue llevada a cabo por el equipo del profesor Julio Diez Casero (Universidad de Valladolid) con anterioridad a este trabajo.

3.4. Análisis bioinformático

La detección de RdRp se llevó a cabo mediante el programa PalmScan (v2), un software especializado en encontrar los motivos conservados A, B, C de la secuencia de aminoácidos de las polimerasas virales en datos brutos de secuenciación. Fue llevada a cabo por el profesor Marcos de la Peña (CSIC-Universidad Politécnica de Valencia). Se usaron los parámetros por defecto del programa.

El análisis de las RdRps obtenidas se realizó con el lenguaje de programación R (v 4.2.3). El código empleado en los análisis puede consultarse en los materiales suplementarios (Código). Como bases de datos de referencia de las RdRps se usaron las últimas versiones de PALMdb (2021-03-14) (Edgar et al., 2022) y RdRp-scan (Charon et al., 2022). Para la asignación de la taxonomía se usó el programa de alineamiento acelerado de secuencias DIAMOND (v2.1.8). Se fijaron diferentes porcentajes de identidad según el nivel de resolución taxonómica de la asignación: especie ($> 90\%$), género ($> 70\%$), y familia ($> 30\%$).

El dendrograma se realizó mediante alineamiento múltiple de la secuencia completa de aminoácidos de las RdRps detectadas, con el programa Clustal Omega (Larkin et al., 2007). La construcción del árbol se basó en el método "Neighbor-Joining" (NJ). Para su representación gráfica se utilizó el paquete de R ggtree (Yu, 2022). A aquellas RdRps cuyos motivos A, B y C no alineaban con ninguna RdRp conocida en las bases de datos, se les asignó identidad manualmente, mediante la comparación visual con las secuencias consenso de los diferentes órdenes y filos publicados en RdRp-scan.

4.- RESULTADOS

4.1. Taxonomía y distribución de las RdRp

Se identificaron un total de 159 RdRps entre todas las muestras analizadas. De ellas, se consiguió asignar grupo taxonómico a 133 (83%), con diferentes niveles de resolución: 91 a nivel de familia, 36 a nivel de género y 6 a nivel de especie (Tabla 1). El hábitat en el que se encontró más RdRps fue el de *Quercus ilex* (56,4%), seguido de *Pinus radiata* (28,6%) y *Castanea Sativa* (15%). El tipo de muestra en el que se identificaron más RdRps correspondió con las muestras de árbol (48,1%) seguidas de artrópodos (21,1%), hongos de artrópodos (15,8%) y hongos de árbol (15%) (Figura 5 y Tabla 2).

Tabla 1. Identidad taxonómica de las RdRp.

	Tipo de genoma		
	ssRNA(+)	ssRNA(-)	dsRNA
Familia (-viridae)	Astro- Botourmia- Dicistro- Endorna- Fusari- Ifla- Kita- Mito- Narna- Noda- Polycipi- Seco- Solemo- Tombus- Virga-	Chu-	Chryso- Partiti- Polymyco- Reo- Toti-
Género (-virus)	Alphanoda- Hypo- Mito- Ourmia-		Alphachryso- Betapartiti- Botybirna-
Especie	Black queen cell virus Deformed wing virus Lake Sinai Virus Lake Sinai Virus 1 Hubei picorna-like virus 15		Ustilaginoidea virens partitivirus

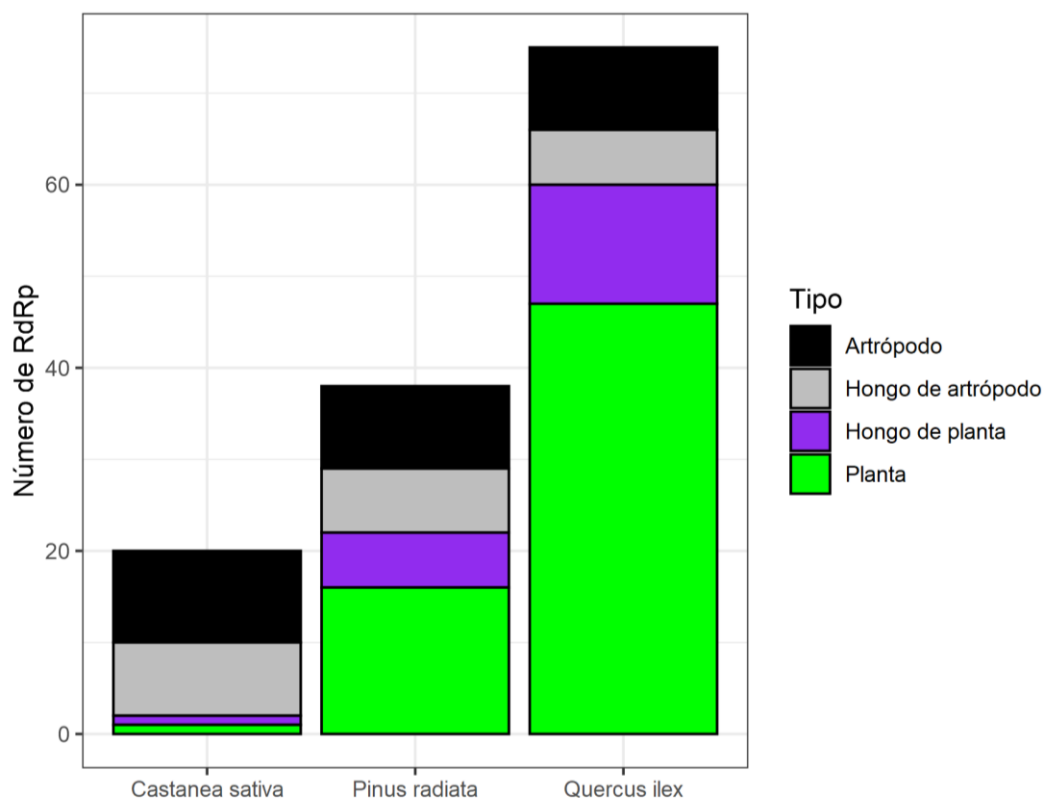


Figura 5. Recuento de las RdRps según el hábitat y el tipo de muestra. En la Tabla Suplementaria 1 (Tabla S1) se encuentra desglosada el recuento de las RdRps identificadas por hábitat y por tipo de muestra.

Tabla 2. Recuentos totales y porcentajes de RdRps. La columna “#RdRp” contiene el número de RdRps con identidad asignada para ese hábitat y tipo de muestra. La columna “Porcentaje” contiene porcentajes relativos al número total de RdRps encontradas. En la Tabla Suplementaria 2 (Tabla S2) se puede consultar el recuento de todas las RdRps incluyendo las que no se pudo asignar una identidad.

Hábitat	Tipo de muestra	#RdRp	Porcentaje
<i>Castanea sativa</i>	Artrópodo	10	7,5
	Hongo artrópodo	8	6,0
	Hongo planta	1	0,8
	Planta	1	0,8
<i>Pinus radiata</i>	Artrópodo	9	6,8
	Hongo artrópodo	7	5,3
	Hongo planta	6	4,5
	Planta	16	12,0
<i>Quercus ilex</i>	Artrópodo	9	6,8
	Hongo artrópodo	6	4,5
	Hongo planta	13	9,8
	Planta	47	35,3
Total	Total	133	100,0

Las familias víricas más numerosas fueron *Botourmiaviridae* (36,1%), *Partitiviridae* (12%), *Narnaviridae* (10,5%), *Mitoviridae* (7,5%) y *Tombusviridae* (6%). Gran parte de los virus identificados en hongos pertenecieron a dichas familias (Figura 6). En cuanto al tipo de genoma, los virus ssRNA(+) fueron los más representados (80,5 %), seguidos de los dsRNA (18,8%) y un único representante ssRNA(-), *Chuviridae*, encontrado en insectos del hábitat del pino.

La distribución de las RdRp permite observar que los tipos de muestras poseen perfiles víricos distintos, como ocurre en muestras de insectos y hongos (Figura 6). En este sentido, se aprecia cómo algunas de las familias son exclusivas de hongos (*Mitoviridae*) o de insectos (*Dicistroviridae*, *Solemoviridae*, *Chrysoviridae* e *Iflaviridae*). En las muestras de insectos, las familias *Dicistroviridae* y *Solemoviridae* se encontraron en los 3 hábitats muestreados, la familia *Iflaviridae* se encontró en los hábitats de *C. sativa* y *P. radiata*, mientras que el resto de las familias identificadas estuvieron presentes en un único hábitat.

Por el contrario, las muestras de plantas y hongos poseen un perfil bastante similar entre ellas. En las muestras de planta y hongo de planta procedentes del hábitat de *C. sativa* solo se ha conseguido identificar una familia vírica. En la Figura 6 puede observarse que la familia *Endornaviridae* se encontró solamente en las muestras de planta de los hábitats de *Q. ilex* y *P. radiata*, mientras que las familias de fitovirus *Virgaviridae* y *Secoviridae* sólo se han encontrado en *Q. ilex*.

En hongos aislados de insectos, se ha identificado a las familias *Botourmiaviridae* y *Mitoviridae* en todos los hábitats analizados. Las familias *Partitiviridae* y *Totiviridae* solo se encontraron en el hábitat de *Q. ilex* y *C. sativa*. Todas las familias de virus identificadas en plantas se han identificado en insectos salvo *Fusariviridae* e *Hypoviridae* (solamente en hongos de planta), así como *Polymycoviridae* y *Chrysoviridae* (hongos de insectos). Esta última familia también se ha identificado en muestras de insectos, pero perteneciente a otro hábitat.

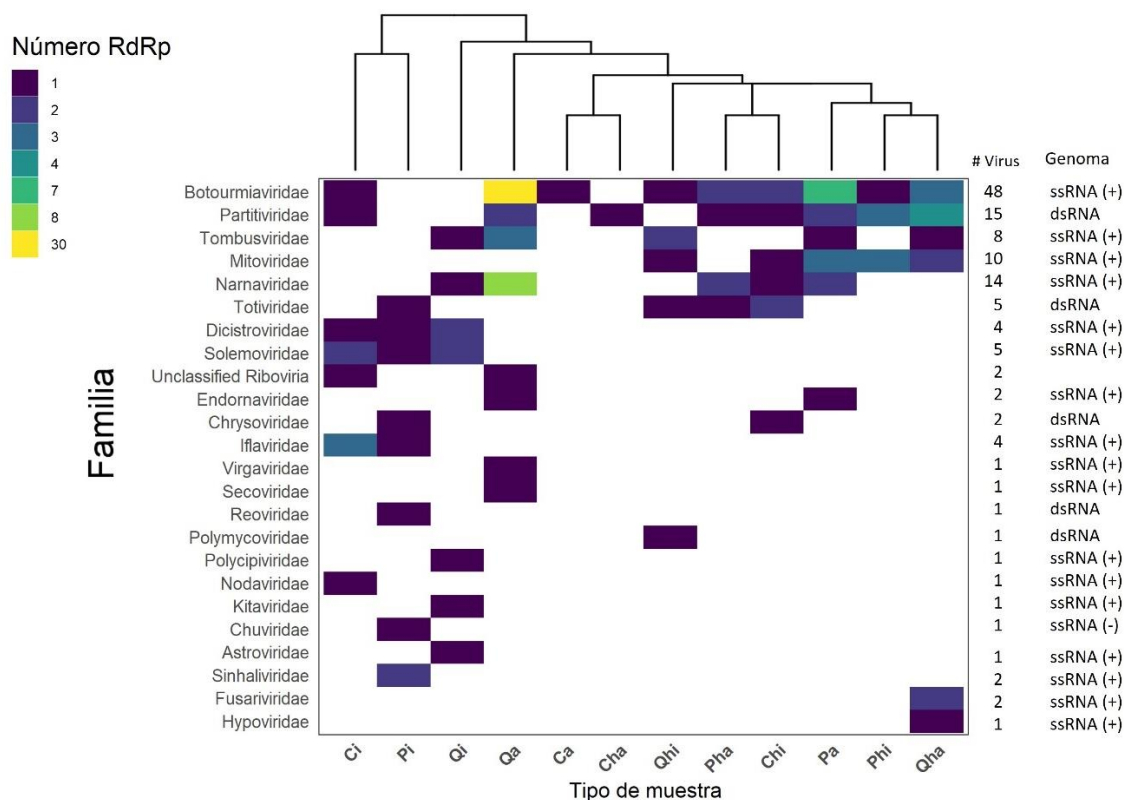


Figura 6. Distribución de las RdRp en función de la familia vírica y el tipo de muestra. El tipo de muestra (abajo) está abreviada con una mayúscula inicial que representa el hábitat (C: *C. sativa*, P: *P. radiata* y Q: *Q. ilex*) y las minúsculas representan el tipo de muestra (a: árbol, ha: hongo de árbol, hi: hongo de insecto, i: insecto). El árbol (arriba) agrupa los diferentes grupos de muestras según la similitud de sus perfiles de presencia y abundancia víricas. A la derecha se indica el número de RdRps por familia y el tipo de genoma.

4.2. Filogenia de las RdRp

El alineamiento múltiple de las secuencias de las RdRps dio lugar al dendrograma que se observa en la Figura 7. En él se aprecia cómo las RdRps procedentes de muestras de planta y de insecto forman dos grupos bien diferenciados, mientras que las RdRps de hongos forman agrupaciones más dispersas en subgrupos heterogéneos. Dentro de las RdRps identificadas en insectos se encuentran el género *Sinavirus* con dos especies, *Lake sinai virus 1* y *2*; las familias *Dicistroviridae* e *Iflaviridae*, al que pertenece otra de las RdRps identificadas con la especie “*Deformed wing virus*”; la especie “*Black queen cell virus*”, que es un *Triatovirus* y la familia *Polycipiviridae*. Podemos ver que solo hay una RdRp que pertenece a planta y otra a hongo de insectos entre las anteriores. Las secuencias de las RdRp que aparecen aisladas, como ocurre con la familia *Botourmiaviridae* y *Narnaviridae*, se parecen más entre ellas que entre miembros de sus respectivas familias, posiblemente por ser de insectos.

También se puede observar que existen algunos nodos con RdRps procedentes de muestras diversas, especialmente en el caso de los *Botourmiaviridae*, *Narnaviridae*, *Partitiviridae* y *Tombusviridae*. Todo ello podría considerarse una evidencia preliminar de virus con capacidad “cross-kingdom”. Además, en los casos en los que aparecen emparejadas dos RdRp procedentes de hongos de planta y hongos de insecto como ocurre con algunos *Betapartivirus* y *Narnaviridae*, se ha podido comprobar que sus secuencias poseen una identidad del 100%. En todos los demás casos la homología en los motivos conservados de la RdRp siempre fue mayor al 50%. Otros nodos solo presentan muestras de hongos y plantas, como la familia *Mitoviridae*; el género

Betapartitivirus, que ha sido identificado exclusivamente en hongos y el género *Ourmiavirus*, que solo se ha identificado en planta.

Las RdRps asignadas manualmente se concentran principalmente en dos grupos. En el primero, alinean junto a RdRps identificadas como *Reoviridae*, *Chuviridae* y *Totiviridae*, de muestras de hongos e insectos. En otro nodo del mismo grupo se sitúan junto a la familia *Fusariviridae* y el género *Hypovirus*, que son conocidos por su capacidad de producir hipovirulencia. En el segundo grupo, comparten nodo con las familias *Kitaviridae*, *Endornaviridae*, *Virgaviridae* y las especies *Lake sinai virus*. Las RdRps asignadas manualmente que aparecen fuera de estos grupos presentan el orden o filo de las RdRps que las rodean.

4.3. RdRps divergentes

Del total de 159 RdRps detectadas, a 26 no se les pudo asignar identidad a partir de las bases de datos disponibles. A estas RdRps se les asignó una taxonomía tentativa manualmente de forma visual, comparando sus motivos A, B y C con las secuencias consenso de cada grupo. La mayor resolución taxonómica alcanzada fue de Orden (19/26), seguida por Filo (5/26) (Tabla 3). A 2 de las RdRps no se les pudo asignar relación taxonómica de ningún tipo porque no presentaban similitud con ninguna secuencia consenso.

Comparando esta asignación con la agrupación del dendrograma, en aquellas RdRps que se agrupan juntas no es posible decidir con fiabilidad su clasificación más probable, ya que presentan diversidad de taxonomías asignadas. En cambio, para las que aparecen aisladas, en concreto del orden *Wolframvirales* y *Ourmiavirales*, sí podría considerarse una asignación correcta, ya que las próximas a ellas en el dendrograma pertenecen a esos mismos órdenes. En la Tabla 3 puede consultarse en detalle la decisión sobre cada asignación hecha manualmente (columna “Comentarios”).

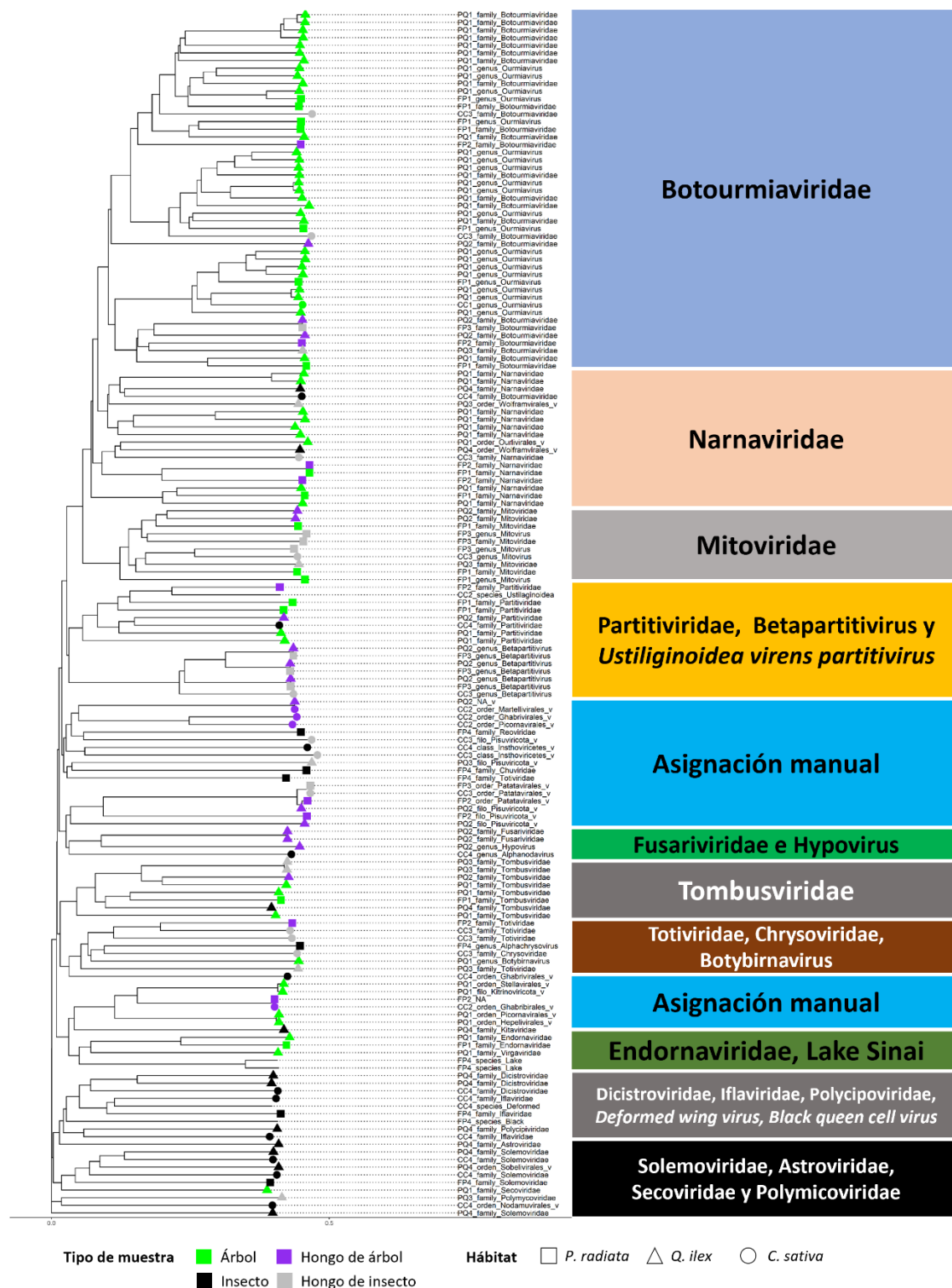


Figura 7. Dendrograma obtenido por “Neighbour Joining”. Los colores indican el tipo de muestra (verde: árbol, morado: hongo de árbol, gris: hongo de insecto y negro: insecto) y la forma el hábitat (círculo: *Castanea sativa*, triángulo: *Quercus ilex* y cuadrado: *Pinus radiata*). A la derecha se muestran los principales géneros y familias.

Tabla 3. RdRps divergentes asignadas de manera visual. En la primera columna se encuentra el identificador de la RdRp, seguido del posible filo/orden que podrían asignarse y de los comentarios del porqué se ha asignado o no.

Muestra	Posible			Comentarios
	Dendrograma	Filo	Orden	
cc2_4862	NA	Duplornaviricota	Reoviral/Ghabrivirales	Difícil de discernir entre las dos
cc2_13850	NA	Duplornaviricota	Ghabrivirales	Para identificar la RdRp, el PalmScan la comparaba con los Duplornaviricota El motivo B presenta similitudes con los Ghabrivirales por el TTF
cc2_14464	NA	Pisuviricota	Picornavirales/Nidovirales	El motivo B presenta más similitud con los Picornavirales
cc2_60275	NA	Kitrinoviricota	Martellivirales/Hepelivirales	El motivo B presenta más similitud con los Martellivirales
cc3_77	NA	Negarviricota	Elloviricetes/Insthoviricetes	Asignado a Elloviricetes por la base de datos del RdRp-scan El motivo B presenta más similitud con Insthoviricetes. Ambos son clases
cc3_1997	NA	Pisuviricota	NA	No presenta similitud con ningún orden de los Pisuviricota
cc3_33765	NA	Pisuviricota	Patatavirales	Asignado a Patatavirales por ser el único similar
cc4_8462	NA	Duplornaviricota	Ghabrivirales	El motivo B presenta más similitud con los Ghabrivirales
cc4_25596	NA	Kitrinoviricota	Nodamuvirales/Martellivirales	Asignado a Martellivirales en el MSA. El motivo A presenta más similitud con Nodamuvirales. El motivo B es difícil de asignar
cc4_32538	NA	Negarviricota	Insthoviricetes	Las repeticiones de M en el motivo B presenta similitudes con Insthoviricetes. El motivo A es similar Es clase
fp2_22978	NA	NA	NA	No se ha podido asignar a ningún filo ni orden
fp2_44269	NA	Pisuviricota	Patatavirales	Asignado a Patatavirales por presentar SGD en el motivo B
fp2_48362	NA	Pisuviricota	NA	Difícil asignar orden. El motivo A presenta similitudes a muchos ordenes pero el B podría asignarse a todos los ordenes
fp3_3421	NA	Pisuviricota	Patatavirales	Asignado a Patatavirales. El motivo B presenta SGD y el motivo C SDD
pq1_5218	Ourlivirales/Wolframvirales	Lenarviricota	Ourlivirales	Asignado a Lenarviricota al presentar el motivo A igual que su secuencia consenso. El motivo B aparece cortado, el consenso es más grande El motivo B presenta más similitud con Ourlivirales
pq1_12440	NA	Kitrinoviricota	NA	Presenta similitudes a las secuencias consenso del filo para los motivos A y B Difícil asignar orden porque solo presenta similitud el motivo A
pq1_14086	NA	Kitrinoviricota	Hepelivirales	Asignado a Hepelivirales. Podría ser cualquier orden
pq1_43111	NA	Pisuviricota	Stellavirales	Asignado a Stellavirales por presentar GS*TT en el motivo B
pq1_55531	NA	Pisuviricota	Picornavirales	Asignado a Picornavirales por presentar TTD en el motivo B
pq2_10979	NA	Pisuviricota	NA	Difícil de asignar. El motivo B solo presenta similitud el SGS inicial
pq2_17978	NA	Pisuviricota	Patatavirales/Durnavirales	El motivo A es similar a los Durnavirales y el B a los Patatavirales
pq2_82132	NA	NA	NA	No se ha podido asignar a ningún filo ni orden
pq3_2108	NA	Pisuviricota	NA	Para identificar la RdRp, el PalmScan la comparaba con los Pisuviricota. No se observa el motivo B ni el A de este filo
pq3_4400	Wolframvirales	Lenarviricota	Wolframvirales	Al motivo B le faltan unos aminoácidos de la secuencia consenso pero presenta similitud con Wolframvirales
pq4_2732	Wolframvirales	Lenarviricota	Wolframvirales	Al motivo B le faltan unos aminoácidos de la secuencia consenso pero presenta similitud con Wolframvirales
pq4_7034	Sobelivirales	Pisuviricota	Sobelivirales	El motivo A presenta más similitud con los Sobelivirales

5.- DISCUSIÓN

El estudio de viromas ambientales ha aumentado su popularidad gracias a los últimos grandes avances que han surgido en las tecnologías de secuenciación masiva. La escala de estos análisis es muy variable, con algunos estudios centrados en una especie en particular (Faizah et al., 2020; Wu et al., 2015), en un ecosistema (Hurwitz & Sullivan, 2013) o en diferentes entornos (Daugrois et al., 2021; Ma et al., 2021). No obstante, la mayoría de los avances que hay en virología vegetal se enfocan a cultivos o árboles frutales con interés económico, lo que contrasta con escasez de estudios en el ámbito forestal (Rumbou et al., 2021). Si reducimos el área de estudio al bosque mediterráneo, siendo la cuenca mediterránea considerada un punto caliente de biodiversidad (Médail et al., 2019), apenas se encuentra información. En el presente trabajo, centrado en 3 hábitats de bosque mediterráneo del noroeste de España, se han encontrado 159 RdRps de origen viral, de las cuales 26 son desconocidas. Esto muestra el gran vacío de conocimiento que poseemos de la diversidad vírica que existe en nuestros propios bosques.

El grupo en el que se han identificado más géneros, por tipo de genoma, es el de los virus ssRNA (+). Además, estos virus ssRNA (+) son el origen evolutivo de los otros dos grupos identificados, ssRNA(-) y dsRNA (Koonin et al., 2015, 2021). Dado que la RdRp es exclusiva de los virus de RNA, no se ha encontrado ninguna familia que presente su material genético en forma de DNA.

De las 133 RdRps a las que se les pudo asignar una identidad, el 56,4% pertenecieron al hábitat de *Quercus ilex*. Las muestras de *Quercus ilex* se tomaron en una dehesa, un ecosistema único de la Península Ibérica que ha sido modificado por el hombre para simultanear la obtención de recursos ganaderos y forestales no maderables. Esto hace que sea un ecosistema con una elevada biodiversidad (Moreno et al., 2016; Rodríguez-Rojo et al., 2022). Ante esta gran biodiversidad no resulta extraño que también exista una gran diversidad vírica que nunca haya sido investigada.

Algunas de las familias víricas encontradas están presentes en todos los tipos de muestra, lo que podría interpretarse como un primer indicio de su potencial “cross-kingdom”. Es el caso de los *Botourmiaviridae*, *Partitiviridae*, *Tombusviridae* y *Narnaviridae*. La familia *Botourmiaviridae* presenta muchos géneros que infectan hongos. Algunos miembros de estos géneros se han identificado en hojas de vid afectados por el oomiceto *Plasmopara viticola* (Chiapello et al., 2020). Los micovirus de esta familia pueden persistir en su huésped fúngico sin la necesidad de una cápside o MP y se cree que solo necesitan la RdRp para poder replicarse (Wang et al., 2020). También se ha descrito, en manzanos, el “*Apple ourmia-like virus 2*”, pero no se sabe con certeza si el hospedador era la planta o el hongo asociado (Wright et al., 2020). Solamente el género *Ourmiavirus* infecta plantas en exclusividad (Rastgou et al., 2009). La RdRp de este género tiene mayor similitud con la de los géneros *Mitovirus* y *Narnavirus* (familia *Narnaviridae*) y su proteína de movimiento (MP) es similar a la de la familia *Tombusviridae*. La proteína de la cubierta presenta cierta similitud con la de algunos virus vegetales o animales. Estos virus se transmiten muy fácilmente por vía mecánica y no se ha identificado vector, por lo que se piensa que podría existir una transmisión horizontal (Ayllón et al., 2020). En el presente trabajo hemos identificado en una muestra de insecto una RdRp identificada como *Boutourmiaviridae* al que no hemos podido asignar género, pero podría ser una evidencia de los insectos como vector.

La familia *Narnaviridae* presenta el genoma más simple de todos los virus RNA porque solo codifican un polipéptido, en el que se encuentra la RdRp. Dentro de esta familia, el género *Narnavirus* se replica en el citoplasma, mientras que los *Mitovirus* lo hacen en las mitocondrias. Filogenéticamente, los *Narnavirus* están mucho más relacionados con los *Ourmiavirus* (familia *Botourmiaviridae*) que con los *Mitovirus*. Aunque no se hayan

encontrado *Mitovirus* fuera de los hongos, sí que se han identificado secuencias de *Mitovirus* en mitocondrias y secuencias mitocondriales de plantas (Hong et al., 1998; Marienfeld et al., 1999), una de las RdRp que hemos asignado como *Mitovirus* es de una muestra de planta que puede ser una evidencia de esto. No se conoce función alguna, pero se piensa que pudieron ser intentos fallidos de invasión a las plantas. Algunos mitovirus pueden causar hipovirulencia (de Rezende et al., 2021; Hillman & Cai, 2013).

La familia *Partitiviridae* presenta varios géneros que tienen hospedadores característicos, que pueden ser plantas (*Deltapartivirus*), hongos (*Gammapartivirus*) y protozoos (*Cryspovirus*). En el género identificado, *Betapartivirus*, sus hospedadores pueden ser plantas y hongos. También se han identificado muchas secuencias de esta familia en artrópodos (Cross et al., 2023). Entre ellos cabe destacar el “*Galbut virus*”, común en las poblaciones salvajes de *Drosophila melanogaster* y que se puede transmitir de forma vertical (de madre a hijo). La especie *Ustilaginoidea virens partitivirus*, encontrada en el presente trabajo en una muestra de hongo de planta, es la causante de la enfermedad del “falso carbón” en el arroz (Jose et al., 2023). Puede que haya un hospedador alternativo no identificado de esta especie de virus.

Respecto a la familia *Tombusviridae*, se ha descrito tanto en animales (Yin et al., 2022) como en plantas (Lappe et al., 2022) y hongos (Botella et al., 2022). En plantas, las infecciones suelen estar limitadas al sistema radicular, pero puede infectar toda la planta gracias a sus MP (Canuti et al., 2023). Se transmite de manera mecánica, por contacto, a través de las semillas o a través de algún vector como puede ser un hongo o un insecto, dependiendo del género del virus.

En algunas familias sólo se ha detectado una RdRp en las muestras analizadas, pero se han descrito evidencias de su capacidad para infectar diversos reinos. Es el caso de los *Totiviridae*, *Reoviridae*, *Chuviridae*, *Chrysoviridae*, *Astroviridae*, *Polymycoviridae* y dos ribovirus con clasificación incompleta: el género *Botybirnavirus* y la especie *Hubei picorna-like virus 15* (Ribovirus sin clasificar). La familia *Totiviridae* se asocia a infecciones latentes tanto en hongos como en protozoos. La familia *Reoviridae* tiene una distribución global e infecta muchas especies de vertebrados, incluyendo a los humanos, en los que puede producir multitud de enfermedades del tracto digestivo, respiratorio y hepatitis (Fuchs et al., 2022). Uno de los géneros más conocidos de esta familia son los Rotavirus. Esta familia de virus se transmite por vía aérea o por contacto, y algunos géneros pueden usar los insectos como vectores y hospedadores. Esta familia podría ser considerada “cross-kingdom”, ya que algunos de sus integrantes también afectan a insectos y plantas. Los virus que infectan estos dos últimos muestran una mayor uniformidad y una variación antigénica mucho menor que los que infectan vertebrados (Fuchs et al., 2022).

La familia *Chuviridae*, exclusiva de insectos, es la única familia que pertenece al grupo V de la clasificación de Baltimore por ser un RNA monocatenario negativo [ssRNA (-)]. La familia *Chrysoviridae* infecta hongos, plantas e insectos (Kotta-Loizou et al., 2020). La familia *Astroviridae* presenta el género *Mamastrovirus*, que infecta mamíferos (incluidos los humanos), y el *Avastrovirus*, que infecta a aves; produciendo en los dos casos gastroenteritis agudas. En la actualidad se siguen describiendo nuevos hospedadores de esta familia (Fernández-Correa et al., 2019), por lo que no resultaría raro que nunca se haya descrito en insectos, que es en el tipo de muestra donde se ha encontrado esta RdRp. La familia *Polymycoviridae* afecta hongos en los que causa hipovirulencia o hipersensibilidad a antifúngicos o bacterias, incluso es capaz de modular el metabolismo del carbono, nitrógeno y hierro (Kotta-Loizou & Coutts, 2022). El género *Botybirnavirus* (*Orthornavirae*) tiene como hospedadores naturales a los hongos (Hough et al., 2023), sin embargo, esta RdRp se ha encontrado en una muestra perteneciente a planta, por lo que podría ser indicativo de alguna capacidad “cross-kingdom”. En cambio, sí se ha descrito que el *Hubei picorna-like virus 15* tiene un porcentaje de identidad aminoacídica

del 68 % con el *Aphid Lethal Paralysis Virus* (Shi et al., 2016), e incluso se ha encontrado presente en las abejas melíferas en España (Granberg et al., 2013).

En cuanto al resto de virus identificados, cabe destacar un grupo de familias que comparten el uso de vectores. La familia *Kitaviridae* infecta plantas y presenta algunos géneros que usan insectos como vectores (Ramos-González et al., 2022) e incluso algunos producen MP, como los *Blunervirus* (Morozov et al., 2020), *Cilevirus* (Leastro et al., 2021) e *Higrevirus* (Atabekova et al., 2022) . La familia *Virgaviridae* también expresa estas MP y alguno de sus géneros usa como vectores los insectos (Adams et al., 2017). En cambio, los *Endornaviridae*, cercanos filogenéticamente, no las presenta e infecta plantas y hongos. Son virus ssRNA (+) que no presentan una verdadera cápside ya que el genoma se encapsula junto a una replicasa viral. Esto llevó a la confusión de asignarlos inicialmente a la categoría III de Baltimore (dsRNA), ya que se acumulaban intermediarios en forma de dsRNA en las células infectadas (Adams et al., 2017; Valverde et al., 2019). La familia *Solemoviridae* infecta plantas, pero se conoce que una de sus maneras de transmisión es el uso de insectos como vectores (Sömera et al., 2021), que puede ser una de las razones por la que hemos encontrado esta familia solo en muestras de insectos. Otras familias, como la *Secoviridae*, también son importantes patógenos de las plantas y su transmisión está mediada por insectos y nematodos (Fuchs et al., 2022).

Con respecto a las 26 RdRps divergentes a las que no se les pudo asignar identidad con las bases de datos, estas se agruparon en el dendrograma en dos nodos bien diferenciados y cercanos entre sí. Si bien podría tratarse de RdRps de nuevos virus no descritos anteriormente, no es posible descartar que correspondan con RdRps no canónicas, en las que los motivos A, B y C podrían aparecer desordenados, o incluso tratarse de RdRps de los propios hospedadores.

Por último, cabe señalar algunas limitaciones del presente estudio. En primer lugar, muchas secuencias presentes en las bases de datos víricas no tienen una asignación taxonómica o esta es incompleta, y la alta variación genética de los virus hace difícil conseguir una asignación precisa a nivel de género o especie. Es necesaria una actualización recurrente de las bases de datos de las RdRps para facilitar la asignación de otras nuevas. En segundo lugar, aunque algunas de las familias víricas descritas pueden ser “cross-kingdom” porque contienen integrantes que infectan diversos reinos, lo más habitual es que dichos virus, de forma individual, se especialicen en hospedadores de un único reino, como ocurre, por ejemplo, con la familia *Totiviridae*, en la que unos géneros infectan hongos y otros protozoos. De igual modo, el análisis *in silico* realizado no permite verificar si el hecho de que algunas RdRps se hayan identificado en organismos de reinos diferentes a los esperados (por ejemplo, fitovirus encontrados en muestras de hongos) implica que el hospedador es susceptible al virus, o este solo actúa como vector o portador asintomático. Además, muchos de los fitovirus que se conocen han sido descritos principalmente en cultivos. En el área forestal hay aún muy pocos estudios al respecto, por lo que pueden existir hospedadores no descritos para estos virus que ya se conocían.

6.- CONCLUSIÓN

En el presente trabajo se pretendía describir la diversidad de virus RNA de tres ecosistemas de bosque mediterráneo. Con respecto al primer objetivo, relativo a caracterizar una filogenia de virus ya conocidos y registrados en bases de datos, se han encontrado 133 RdRps repartidas en 20 familias, 7 géneros y 6 especies. La distribución de los virus por ecosistema fue 54,7% en *Quercus ilex*, 26,4% en *Pinus radiata* y 18,9% en *Castanea sativa*. Con respecto al segundo objetivo, relativo a estimar la proporción de virus ARN desconocidos o no registrados en bases de datos, se han encontrado 26 RdRps (16%) sin identidad en las bases de datos actuales, aunque no se puede descartar que puedan pertenecer a los propios organismos hospedadores. Esto demuestra que en los bosques existe una gran diversidad vírica que apenas se ha estudiado. Además, muchos de estos virus se han descrito en la literatura como capaces de saltar a otros reinos. Conocer la diversidad vírica no solo permitirá actuar frente a las diferentes enfermedades que pueden causar, también servirá para prever qué virus son más susceptibles de presentar potencial “cross-kingdom” y poder evitar una difusión descontrolada que pueda ocasionar nuevas pandemias.

7.- AGRADECIMIENTOS

En primer lugar, quiero agradecer a mis tutores. A Sergio por toda la ayuda que me ha dado, con el que he podido aprender mucho. A Julio por permitirme realizar este trabajo en su grupo de investigación. También agradecer a todos los compañeros del grupo de investigación que fueron partícipes de este proyecto de una manera directa o indirecta. A Margarita, Natalia y Pilar de las que he aprendido mucho durante el máster y sin su ayuda no hubiera llegado a realizar este trabajo.

A mi familia por ser un apoyo incondicional durante el desarrollo del trabajo. Os mencionaría de uno en uno, pero no acabaríamos, pero quiero agradecer especialmente a mis hermanos pequeños por hacer menos monótona la vida dando guerra de vez en cuando.

A mis amigas que son un pilar fundamental en mi vida. A Clara y Mónica con las que siempre es un placer tomarse un café o ir a comer para celebrar algo... las veces que haga falta. A Anaïs y Elena por todos esos momentos hablando de cualquier cosa o jugando a cualquier cosa por la noche que me ayudan tanto a desconectar del ajetreo del día (aunque muchas veces acabemos en la Grieta del Invocador "tilteándonos" más). A Agus, Alex, Jessi y Vanessa con los que estudié en Toledo. Aunque ya no sea posible vernos todos los días y estemos cada uno en una punta del país, siempre hace ilusión volver a verlos y hablar de vez en cuando.

A todas las personas a las que no he podido mencionar expresamente, GRACIAS.

8.- BIBLIOGRAFÍA

- Adams, M. J., Adkins, S., Bragard, C., Gilmer, D., Li, D., MacFarlane, S. A., Wong, S. M., Melcher, U., Ratti, C., & Ryu, K. H. (2017). ICTV virus taxonomy profile: Virgaviridae. *Journal of General Virology*, 98(8), 1999-2000. <https://doi.org/10.1099/JGV.0.000884>
- Afelt, A., Frutos, R., & Devaux, C. (2018). Bats, coronaviruses, and deforestation: Toward the emergence of novel infectious diseases? *Frontiers in Microbiology*, 9(APR), 352620. <https://doi.org/10.3389/FMICB.2018.00702>
- Atabekova, A. K., Lazareva, E. A., Lezzhov, A. A., Solovieva, A. D., Golyshev, S. A., Skulachev, B. I., Solovyev, I. D., Savitsky, A. P., Heinlein, M., Morozov, S. Y., & Solovyev, A. G. (2022). Interaction between Movement Proteins of Hibiscus green spot virus. *Viruses*, 14(12). <https://doi.org/10.3390/V14122742>
- Ayllón, M. A., Turina, M., Xie, J., Nerva, L., Marzano, S. Y. L., Donaire, L., & Jiang, D. (2020). ICTV Virus Taxonomy Profile: Botourmiaviridae. *Journal of General Virology*, 101(5), 454-455. <https://doi.org/10.1099/JGV.0.001409>
- Babaian, A., & Edgar, R. (2022). Ribovirus classification by a polymerase barcode sequence. *PeerJ*, 10, e14055. <https://doi.org/10.7717/PEERJ.14055>
- Balique, F., Lecoq, H., Raoult, D., & Colson, P. (2015). Can Plant Viruses Cross the Kingdom Border and Be Pathogenic to Humans? *Viruses*, 7(4), 2074. <https://doi.org/10.3390/V7042074>
- Baltimore classification ~ ViralZone*. (s. f.). Recuperado 17 de julio de 2023, de <https://viralzone.expasy.org/254>
- Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological Reviews*, 35(3), 235. <https://doi.org/10.1128/BR.35.3.235-241.1971>
- Bian, R., Andika, I. B., Pang, T., Lian, Z., Wei, S., Niu, E., Wu, Y., Kondo, H., Liu, X., & Sun, L. (2020). Facilitative and synergistic interactions between fungal and plant viruses. *Proceedings of the National Academy of Sciences of the United States of America*, 117(7), 3779-3788. <https://doi.org/10.1073/PNAS.1915996117>
- Botella, L., Jung, M. H., Rost, M., & Jung, T. (2022). Natural Populations from the Phytophthora palustris Complex Show a High Diversity and Abundance of ssRNA and dsRNA Viruses. *Journal of Fungi*, 8(11). <https://doi.org/10.3390/JOF8111118>
- Canuti, M., Rodrigues, B., Lang, A. S., Dufour, S. C., & Verhoeven, J. T. P. (2023). Novel Divergent Members of the Kitrinoviricota Discovered through Metagenomics in the Intestinal Contents of Red-Backed Voles (*Clethrionomys gapperi*). *International Journal of Molecular Sciences*, 24(1). <https://doi.org/10.3390/IJMS24010131>
- Cao, X., Liu, J., Pang, J., Kondo, H., Chi, S., Zhang, J., Sun, L., & Andika, I. B. (2022). Common but Nonpersistent Acquisitions of Plant Viruses by Plant-Associated Fungi. *Viruses*, 14(10), 2279. <https://doi.org/10.3390/V14102279>
- Centro de Descargas del CNIG (IGN)*. (s. f.). Recuperado 25 de agosto de 2023, de <https://centrodedescargas.cnig.es/CentroDescargas/index.jsp>
- Charon, J., Buchmann, J. P., Sadiq, S., & Holmes, E. C. (2022). RdRp-scan: A bioinformatic resource to identify and annotate divergent RNA viruses in metagenomic sequence data. *Virus Evolution*, 8(2). <https://doi.org/10.1093/VE/VEAC082>

- Chiapello, M., Rodríguez-Romero, J., Ayllón, M. A., & Turina, M. (2020). Analysis of the virome associated to grapevine downy mildew lesions reveals new mycovirus lineages. *Virus evolution*, 6(2). <https://doi.org/10.1093/VE/VEAA058>
- Cross, S. T., Brehm, A. L., Dunham, T. J., Rodgers, C. P., Keene, A. H., Borlee, G. I., & Stenglein, M. D. (2023). Galbut Virus Infection Minimally Influences *Drosophila melanogaster* Fitness Traits in a Strain and Sex-Dependent Manner. *Viruses*, 15(2). <https://doi.org/10.3390/V15020539>
- Daugrois, J. H., Filloux, D., Julian, C., Claude, L., Ferdinand, R., Fernandez, E., Fontes, H., Rott, P. C., & Roumagnac, P. (2021). Comparison of the virome of quarantined sugarcane varieties and the virome of grasses growing near the quarantine station. *Viruses*, 13(5). <https://doi.org/10.3390/V13050922>
- de Rezende, R. R., de Oliveira Souza, F., Leal, L. L., Morgan, T., Bermudez, J. M. M., Cascardo, R. de S., de Barros, D. R., & Alfenas-Zerbini, P. (2021). Characterization of a new mitovirus infecting the phytopathogenic fungus *Microdochium albescens*. *Archives of Virology*, 166(12), 3455-3459. <https://doi.org/10.1007/S00705-021-05242-4>
- Edgar, R. C., Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., Lohr, D., Novakovsky, G., Buchfink, B., Al-Shayeb, B., Banfield, J. F., de la Peña, M., Korobeynikov, A., Chikhi, R., & Babaian, A. (2022). Petabase-scale sequence alignment catalyses viral discovery. *Nature* 2022 602:7895, 602(7895), 142-147. <https://doi.org/10.1038/s41586-021-04332-2>
- El comercio global de plantas amenaza la salud de los bosques.* (s. f.). Recuperado 11 de julio de 2023, de <https://theconversation.com/el-comercio-global-de-plantas-amenaza-la-salud-de-los-bosques-130834>
- Ellison, A. M., Bank, M. S., Clinton, B. D., Colburn, E. A., Elliott, K., Ford, C. R., Foster, D. R., Kloeppe, B. D., Knoepp, J. D., Lovett, G. M., Mohan, J., Orwig, D. A., Rodenhouse, N. L., Sobczak, W. V., Stinson, K. A., Stone, J. K., Swan, C. M., Thompson, J., Von Holle, B., & Webster, J. R. (2005). Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. *Frontiers in Ecology and the Environment*, 3(9), 479-486. <https://doi.org/10.1890/1540-9295>
- Faizah, A. N., Kobayashi, D., Isawa, H., Amoa-Bosompem, M., Murota, K., Higa, Y., Futami, K., Shimada, S., Kim, K. S., Itokawa, K., Watanabe, M., Tsuda, Y., Minakawa, N., Miura, K., Hirayama, K., & Sawabe, K. (2020). Deciphering the Virome of *Culex vishnui* Subgroup Mosquitoes, the Major Vectors of Japanese Encephalitis, in Japan. *Viruses*, 12(3), 264. <https://doi.org/10.3390/V12030264>
- Fuchs, M., Hily, J. M., Petrzik, K., Sanfaçon, H., Thompson, J. R., van der Vlugt, R., Wetzel, T., & Ictv Report Consortium. (2022). ICTV Virus Taxonomy Profile: Secoviridae 2022. *The Journal of general virology*, 103(12). <https://doi.org/10.1099/JGV.0.001807>
- Gorbalenya, A. E., Pringle, F. M., Zeddarn, J. L., Luke, B. T., Cameron, C. E., Kalmakoff, J., Hanzlik, T. N., Gordon, K. H. J., & Ward, V. K. (2002). The Palm Subdomain-based Active Site is Internally Permuted in Viral RNA-dependent RNA Polymerases of an Ancient Lineage. *Journal of Molecular Biology*, 324(1), 47. [https://doi.org/10.1016/S0022-2836\(02\)01033-1](https://doi.org/10.1016/S0022-2836(02)01033-1)
- Granberg, F., Vicente-Rubiano, M., Rubio-Guerri, C., Karlsson, O. E., Kukielka, D., Belák, S., & Sánchez-Vizcaíno, J. M. (2013). Metagenomic Detection of Viral Pathogens in Spanish Honeybees: Co-Infection by Aphid Lethal Paralysis, Israel Acute Paralysis

- and Lake Sinai Viruses. *PLOS ONE*, 8(2), e57459. <https://doi.org/10.1371/JOURNAL.PONE.0057459>
- Hillman, B. I., & Cai, G. (2013). The Family Narnaviridae: Simplest of RNA Viruses. *Advances in Virus Research*, 86, 149-176. <https://doi.org/10.1016/B978-0-12-394315-6.00006-4>
- Hong, Y., Cole, T. E., Brasier, C. M., & Buck, K. W. (1998). Evolutionary Relationships among Putative RNA-Dependent RNA Polymerases Encoded by a Mitochondrial Virus-like RNA in the Dutch Elm Disease Fungus, *Ophiostoma novo-ulmi*, by Other Viruses and Virus-like RNAs and by the Arabidopsis Mitochondrial Genome. *Virology*, 246(1), 158-169. <https://doi.org/10.1006/VIRO.1998.9178>
- Hough, B., Steenkamp, E., Wingfield, B., & Read, D. (2023). Fungal Viruses Unveiled: A Comprehensive Review of Mycoviruses. *Viruses*, 15(5). <https://doi.org/10.3390/V15051202>
- Hurwitz, B. L., & Sullivan, M. B. (2013). The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology. *PLOS ONE*, 8(2), e57355. <https://doi.org/10.1371/JOURNAL.PONE.0057355>
- Jose, R. C., Kanchal, T., Louis, B., Talukdar, N. C., & Chowdhury, D. (2023). Grain Characteristics, Moisture, and Specific Peptides Produced by *Ustilagoidea vires* Contribute to False Smut Disease in Rice (*Oryza sativa* L.). *Biomolecules*, 13(4). <https://doi.org/10.3390/BIOM13040669>
- Koonin, E. V., Dolja, V. V., & Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology*, 479, 2. <https://doi.org/10.1016/J.VIROL.2015.02.039>
- Koonin, E. V., Krupovic, M., & Agol, V. I. (2021). The Baltimore Classification of Viruses 50 Years Later: How Does It Stand in the Light of Virus Evolution? *Microbiology and Molecular Biology Reviews*, 85(3). <https://doi.org/10.1128/mubr.00053-21>
- Kotta-Loizou, I., & Coutts, R. H. A. (2022). ICTV Virus Taxonomy Profile: Polymycoviridae 2022. *Journal of General Virology*, 103(5). <https://doi.org/10.1099/JGV.0.001747>
- Lappe, R. R., Elmore, M. G., Lozier, Z. R., Jander, G., Miller, W. A., & Whitham, S. A. (2022). Metagenomic identification of novel viruses of maize and teosinte in North America. *BMC Genomics*, 23(1). <https://doi.org/10.1186/S12864-022-09001-W>
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), 2947-2948. <https://doi.org/10.1093/BIOINFORMATICS/BTM404>
- Leastro, M. O., Freitas-Astúa, J., Kitajima, E. W., Pallás, V., & Sánchez-Navarro, J. A. (2021). Unravelling the involvement of cilevirus p32 protein in the viral transport. *Scientific Reports 2021 11:1*, 11(1), 1-18. <https://doi.org/10.1038/s41598-021-82453-4>
- Lefevre, P., Martin, D. P., Elena, S. F., Shepherd, D. N., Roumagnac, P., & Varsani, A. (2019). Evolution and ecology of plant viruses. *Nature Reviews Microbiology 2019 17:10*, 17(10), 632-644. <https://doi.org/10.1038/s41579-019-0232-3>
- Lefin, N., Herrera-Belén, L., Farias, J. G., & Beltrán, J. F. (2023). Review and perspective on bioinformatics tools using machine learning and deep learning for predicting antiviral peptides. *Molecular diversity*. <https://doi.org/10.1007/S11030-023-10718-3>

- Liu, H., Fu, Y., Xie, J., Cheng, J., Ghabrial, S. A., Li, G., Peng, Y., Yi, X., & Jiang, D. (2012). Evolutionary genomics of mycovirus-related dsRNA viruses reveals cross-family horizontal gene transfer and evolution of diverse viral lineages. *BMC Evolutionary Biology*, 12(1), 1-15. <https://doi.org/10.1186/1471-2148-12-91>
- Liu, S., Xie, J., Cheng, J., Li, B., Chen, T., Fu, Y., Li, G., Wang, M., Jin, H., Wan, H., & Jiang, D. (2016). Fungal DNA virus infects a mycophagous insect and utilizes it as a transmission vector. *Proceedings of the National Academy of Sciences of the United States of America*, 113(45), 12803-12808. <https://doi.org/10.1073/PNAS.1608013113>
- Ma, Y., Fort, T., Marais, A., Lefebvre, M., Theil, S., Vacher, C., & Candresse, T. (2021). Leaf-associated fungal and viral communities of wild plant populations differ between cultivated and natural ecosystems. *Plant-Environment Interactions*, 2(2), 87-99. <https://doi.org/10.1002/PEI3.10043>
- Marienfeld, J., Unseld, M., & Brennicke, A. (1999). The mitochondrial genome of Arabidopsis is composed of both native and immigrant information. *Trends in Plant Science*, 4(12), 495-502. [https://doi.org/10.1016/S1360-1385\(99\)01502-2](https://doi.org/10.1016/S1360-1385(99)01502-2)
- McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, 32(Web Server issue), W20. <https://doi.org/10.1093/NAR/GKH435>
- Médail, F., Monnet, A. C., Pavon, D., Nikolic, T., Dimopoulos, P., Bacchetta, G., Arroyo, J., Barina, Z., Albassatneh, M. C., Domina, G., Fady, B., Matevski, V., Mifsud, S., & Leriche, A. (2019). What is a tree in the mediterranean basin hotspot? A critical analysis. *Forest Ecosystems*, 6(1), 1-19. <https://doi.org/10.1186/S40663-019-0170-6>
- Moreno, G., Gonzalez-Bornay, G., Pulido, F., Lopez-Diaz, M. L., Bertomeu, M., Juárez, E., & Diaz, M. (2016). Exploring the causes of high biodiversity of Iberian dehesas: the importance of wood pastures and marginal habitats. *Agroforestry Systems*, 90(1), 87-105. <https://doi.org/10.1007/S10457-015-9817-7>
- Morozov, S. Y., Lazareva, E. A., & Solovyev, A. G. (2020). Sequence Relationships of RNA Helicases and Other Proteins Encoded by Blunervirus RNAs Highlight Recombinant Evolutionary Origin of Kitaviral Genomes. *Frontiers in Microbiology*, 11, 561092. <https://doi.org/10.3389/FMICB.2020.561092>
- Munkvold, G. P. (2017). Fusarium Species and Their Associated Mycotoxins. *Methods in molecular biology (Clifton, N.J.)*, 1542, 51-106. https://doi.org/10.1007/978-1-4939-6707-0_4
- Nerva, L., Varese, G. C., Falk, B. W., & Turina, M. (2017). Mycoviruses of an endophytic fungus can replicate in plant cells: evolutionary implications. *Scientific Reports* 2017 7:1, 7(1), 1-11. <https://doi.org/10.1038/s41598-017-02017-3>
- Pilarska, D., Georgiev, G., Dobрева, M., Takov, D., Mirchev, P., Doychev, D., Georgieva, M., Nachev, R., Dermendzhiev, P., Draganova, S., Linde, A., & Hajek, A. E. (2018). PATHOGENS AND PARASITIDS OF FOREST PEST INSECTS IN THE REGION OF FOREST PROTECTION STATION PLOVDIV DURING THE PERIOD 1990-2017. *Silva Balcanica*, 19(3).
- Ramos-González, P. L., Chabi-Jesus, C., Tassi, A. D., Calegario, R. F., Harakava, R., Nome, C. F., Kitajima, E. W., & Freitas-Astua, J. (2022). A Novel Lineage of Cile-Like Viruses Discloses the Phylogenetic Continuum Across the Family Kitaviridae. *Frontiers in Microbiology*, 13, 836076. <https://doi.org/10.3389/FMICB.2022.836076>

- Rastgou, M., Habibi, M. K., Izadpanah, K., Masenga, V., Milne, R. G., Wolf, Y. I., Koonin, E. V., & Turina, M. (2009). Molecular characterization of the plant virus genus Ourmiavirus and evidence of inter-kingdom reassortment of viral genome segments as its possible route of origin. *The Journal of general virology*, *90*(Pt 10), 2525-2535. <https://doi.org/10.1099/VIR.0.013086-0>
- Roberts, N., Fyfe, R. M., Woodbridge, J., Gaillard, M. J., Davis, B. A. S., Kaplan, J. O., Marquer, L., Mazier, F., Nielsen, A. B., Sugita, S., Trondman, A. K., & Leydet, M. (2018). Europe's lost forests: a pollen-based synthesis for the last 11,000 years. *Scientific Reports 2017 8:1*, *8*(1), 1-8. <https://doi.org/10.1038/s41598-017-18646-7>
- Rodríguez-Rojo, M. P., Roig, S., López-Carrasco, C., Redondo García, M. M., & Sánchez-Mata, D. (2022). Which Factors Favour Biodiversity in Iberian Dehesas? *Sustainability 2022*, Vol. 14, Page 2345, *14*(4), 2345. <https://doi.org/10.3390/SU14042345>
- Rumbou, A., Vainio, E. J., & Büttner, C. (2021). Towards the Forest Virome: High-Throughput Sequencing Drastically Expands Our Understanding on Virosphere in Temperate Forest Ecosystems. *Microorganisms 2021*, Vol. 9, Page 1730, *9*(8), 1730. <https://doi.org/10.3390/MICROORGANISMS9081730>
- Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., Qin, X. C., Li, J., Cao, J. P., Eden, J. S., Buchmann, J., Wang, W., Xu, J., Holmes, E. C., & Zhang, Y. Z. (2016). Redefining the invertebrate RNA virosphere. *Nature 2016 540:7634*, *540*(7634), 539-543. <https://doi.org/10.1038/nature20167>
- Sõmera, M., Fargette, D., Hébrard, E., & Sarmiento, C. (2021). ICTV Virus Taxonomy Profile: Solemoviridae 2021. *Journal of General Virology*, *102*(12). <https://doi.org/10.1099/JGV.0.001707>
- Valverde, R. A., Khalifa, M. E., Okada, R., Fukuhara, T., & Sabanadzovic, S. (2019). ICTV virus taxonomy profile: Endornaviridae. *Journal of General Virology*, *100*(8), 1024-1025. <https://doi.org/10.1099/JGV.0.001277>
- Wakchaure, P. D., Ghosh, S., & Ganguly, B. (2020). Revealing the Inhibition Mechanism of RNA-Dependent RNA Polymerase (RdRp) of SARS-CoV-2 by Remdesivir and Nucleotide Analogues: A Molecular Dynamics Simulation Study. *Journal of Physical Chemistry B*, *124*(47), 10641-10652. <https://doi.org/10.1021/ACS.JPCB.0C06747>
- Wang, Q., Mu, F., Xie, J., Cheng, J., Fu, Y., & Jiang, D. (2020). A Single ssRNA Segment Encoding RdRp Is Sufficient for Replication, Infection, and Transmission of Ourmia-Like Virus in Fungi. *Frontiers in microbiology*, *11*. <https://doi.org/10.3389/FMICB.2020.00379>
- Wright, A. A., Cross, A. R., & Harper, S. J. (2020). A bushel of viruses: Identification of seventeen novel putative viruses by RNA-seq in six apple trees. *PloS one*, *15*(1). <https://doi.org/10.1371/JOURNAL.PONE.0227669>
- Wu, Z., Yang, L., Ren, X., He, G., Zhang, J., Yang, J., Qian, Z., Dong, J., Sun, L., Zhu, Y., Du, J., Yang, F., Zhang, S., & Jin, Q. (2015). Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *The ISME Journal 2016 10:3*, *10*(3), 609-620. <https://doi.org/10.1038/ismej.2015.138>
- Yin, H. chang, Wan, D. cai, & Chen, H. yan. (2022). Metagenomic analysis of viral diversity and a novel astrovirus of forest rodent. *Virology Journal*, *19*(1), 138. <https://doi.org/10.1186/S12985-022-01847-6>

Yoon, B.-J. (2009). Hidden Markov Models and their Applications in Biological Sequence Analysis. *Current Genomics*, 10(6), 402.
<https://doi.org/10.2174/138920209789177575>

Yu, G. (2022). Data Integration, Manipulation and Visualization of Phylogenetic Trees. *Data Integration, Manipulation and Visualization of Phylogenetic Trees*.
<https://doi.org/10.1201/9781003279242>

TABLAS SUPLEMENTARIAS

Tabla S1. Desglose del recuento de RdRps identificadas según hábitat y tipo de muestras.

Taxa	Categoría	Hábitat			Tipo				Muestras
		Pinus	Quercus	Castanea	Planta	Hongo planta	Hongo artrópodo	Artrópodo	
<i>Alphachrysovirus</i>	Género	1						1	1
<i>Alphanodavirus</i>	Género			1				1	1
<i>Astroviridae</i>	Familia		1					1	1
<i>Betapartitivirus</i>	Género	3	3	1		3	4		7
<i>Black queen cell virus</i>	Especie	1						1	1
<i>Botourmiaviridae</i>	Familia	6	18	3	17	5	4	1	27
<i>Botybirnavirus</i>	Género		1		1				1
<i>Chrysoviridae</i>	Familia			1			1		1
<i>Chuviridae</i>	Familia	1						1	1
<i>Deformed wing virus</i>	Especie			1				1	1
<i>Dicistroviridae</i>	Familia		2	1				3	3
<i>Endornaviridae</i>	Familia	1	1		2				2
<i>Fusariviridae</i>	Familia		2			2			2
<i>Hubei picorna-like virus 15</i>	Especie			1				1	1
<i>Hypovirus</i>	Género		1			1			1
<i>Iflaviridae</i>	Familia	1		2				3	3
<i>Kitaviridae</i>	Familia		1					1	1
<i>Lake Sinai virus</i>	Especie	1						1	1
<i>Lake Sinai virus 1</i>	Especie	1						1	1
<i>Mitoviridae</i>	Familia	3	3		2	2	2		6
<i>Mitovirus</i>	Género	3		1	1		3		4
<i>Narnaviridae</i>	Familia	4	9	1	10	2	1	1	14
<i>Ourmiavirus</i>	Género	4	16	1	21				21
<i>Partitiviridae</i>	Familia	3	3	1	4	2		1	7
<i>Polycipiviridae</i>	Familia		1					1	1
<i>Polymycoviridae</i>	Familia		1				1		1
<i>Reoviridae</i>	Familia	1						1	1
<i>Secoviridae</i>	Familia		1		1				1
<i>Solemoviridae</i>	Familia	1	2	2				5	5
<i>Tombusviridae</i>	Familia	1	7		4	1	2	1	8
<i>Totiviridae</i>	Familia	2	1	2		1	3	1	5
<i>Ustilaginoidea virens partitivirus</i>	Especie			1		1			1
<i>Virgaviridae</i>	Familia		1		1				1
Total		38	75	20	64	20	21	28	133

Tabla S2. Recuento de RdRp por hábitat y tipo de muestra. Incluye aquellas que no se consiguió asignar una identidad.

Hábitat	Tipo de muestra	#RdRp	Porcentaje
<i>Castanea sativa</i>	Artrópodo	13	8,2
	Hongo artrópodo	11	6,9
	Hongo planta	5	3,1
	Planta	1	0,6
<i>Pinus radiata</i>	Artrópodo	9	5,7
	Hongo artrópodo	8	5,0
	Hongo planta	9	5,7
	Planta	16	10,1
<i>Quercus ilex</i>	Artrópodo	11	6,9
	Hongo artrópodo	8	5,0
	Hongo planta	16	10,1
	Planta	52	32,7
Total	Total	159	100,0

CÓDIGO

Libraries

```
# Load packages
library(seqinr)
library(tidyverse)
library(data.table)
library(viridis)
```

Load data

```
# Load alignment, diamond and ids from palmdb release
trimseq <- read.fasta("palmscantrimMAFFT2.fasta", seqtype = "AA")
diamond <- read.table("test/cc_fp_pq_rdrps.pro", sep = '\t', header = F)
ids <- read.table(file = 'palmdb-main/2021-03-14/u_tax.tsv', sep = '\t', header = TRUE)
```

Assign taxonomy

```
# List of unique contig names
uniqvir <- unique(diamond$V1)

# Extract taxonomy per contig
extract_tax <- function(contig, diamond, ids, pid) {

  uvir1 <- diamond[diamond$V1 == contig & diamond$V9 >= pid, ]
  uvir2 <- ids[ids$Label %in% uvir1$V5, ]
  uvir3 <- uvir2[order(match(uvir2$Label, uvir1$V5)), ]
  return(uvir3)
}

### Assign taxonomy

## Species
# Run function and store in list
uvir <- lapply(uniqvir, function(x) extract_tax(x, diamond, ids, 90))
# Names to list entries
names(uvir) <- uniqvir
# List to data frame
species.frame <- do.call(rbind, uvir)

## Genus
uvir <- lapply(uniqvir, function(x) extract_tax(x, diamond, ids, 70))
names(uvir) <- uniqvir
genus.frame <- do.call(rbind, uvir)

## Family
uvir <- lapply(uniqvir, function(x) extract_tax(x, diamond, ids, 40))
names(uvir) <- uniqvir
fam.frame <- do.call(rbind, uvir)

## All
uvir <- lapply(uniqvir, function(x) extract_tax(x, diamond, ids, 00))
names(uvir) <- uniqvir
```



```
all_frame <- do.call(rbind, uvir)
```

```
# Those contigs without an assignation are in All
```

Extraer lista

```
# Clean data frames
```

Species

```
species <- species.frame %>% rownames_to_column() %>% rename(contig =  
rowname) %>% separate(contig,into = c("contig", "eliminate"), sep= "_cov_") %>%  
filter(species != ".") %>% select(-eliminate)
```

Genus

```
genus <- genus.frame %>% rownames_to_column() %>% rename(contig = rowname)  
%>% separate(contig,into = c("contig", "eliminate"), sep= "_cov_") %>% filter(genus !=  
".") %>% select(-eliminate)
```

Family

```
family <- fam.frame %>% rownames_to_column() %>% rename(contig = rowname)  
%>% separate(contig,into = c("contig", "eliminate"), sep= "_cov_") %>% filter(family !=  
".") %>% select(-eliminate)
```

All

```
all_v2 <- all_frame %>% rownames_to_column() %>% rename(contig = rowname) %>%  
separate(contig,into = c("contig", "eliminate"), sep= "_cov_") %>% select(-eliminate)
```

Crear lista

```
taxa_sp <- species  
gen_clean <- genus %>% filter(!contig %in% taxa_sp )  
taxa_gen <- rbind(taxa_sp, gen_clean)  
fam_clean <- family %>% filter(!contig %in% taxa_gen)  
taxa_all <- rbind(taxa_gen, fam_clean)
```

```
# Check because some contigs are selected in two categories
```

```
taxa_sp <- species  
gen_prueba <- gen_clean %>%  
  mutate(species = replace(species, species != '!', '!'))  
fam_prueba <- fam_clean %>%  
  mutate(species = replace(species, species != '!', '!')) %>%  
  mutate(genus = replace(genus, genus != '!', '!'))  
taxa_all_v2 <- rbind(taxa_sp, gen_prueba, fam_prueba)
```

```
# Select each taxa category
```

```
tax_unique <- taxa_sp %>% group_by(contig) %>% count(species) %>%  
slice(which.max(n)) %>% select(-n)
```

```
gen <- gen_prueba %>% group_by(contig) %>% count(genus, family) %>%  
slice(which.max(n)) %>% select(-n)
```

```
fam <- fam_prueba %>% group_by(contig) %>% count(family) %>% slice(which.max(n))  
%>% select(-n)  
nrow(fam)
```

```
[1] 132
```

Here it does not appear 1 contig because it has no family assigned but it has an inferior taxonomic class

```
# Check and create a label for identify the taxonomic category
t_species <- tax_unique %>% rename(taxa = species) %>% mutate(cat = "species")
t_gen <- gen %>% rename(taxa = genus) %>% mutate(cat = "genus") %>% select(-family)
t_fam <- fam %>% rename(taxa = family) %>% mutate(cat = "family")

# Select contigs that are not in the inferior category
t_fam_nr <- t_fam[!(t_fam$contig %in% t_gen$contig), ]
t_gen_nr <- t_gen[!(t_gen$contig %in% t_species$contig), ]

t_fgs <- rbind(t_species, t_gen, t_fam)
t_fgs_nr <- rbind(t_species, t_gen_nr, t_fam_nr)
unique(t_fgs$contig)
```

Compare with last taxa made

Check the other classification

```
taxaframe <- read.csv("D:/Pc/OneDrive - UVa/TFM/Datos/taxaframe.csv")
taxa_last <- taxaframe %>%
  separate(contig,into = c("contig", "eliminate"), sep= "_cov_") %>%
  rename(cat_last = eliminate) %>% rename(taxa_last = taxa)
```

Join

```
tax_join <- t_fgs %>% inner_join(taxa_last, by= ('contig' = 'contig'))
```

##write.csv(tax_join, "taxa_known.csv")

Sacar la lista de contigs que no tienen clasificación

Select names contig and shorten de names

```
names_trimseq <- as.data.table (names(trimseq)) %>% rename(contig = V1)%>%
  separate(contig,into = c("contig", "eliminate"), sep= "_cov_") %>%
  select(-eliminate)
```

```
list_unknown = list (names_trimseq, tax_join)
unknown <- list_unknown %>% reduce (left_join, by = 'contig')
```

Select the unknown (not assign contigs)

```
uk <- unknown[is.na(unknown$taxa),]
unique(uk)
```

Count Unkown and assigned

```
recuento <- t_fgs_nr
recuento$Tipo_muestra <- recuento$contig
recuento_c <- recuento %>% separate(Tipo_muestra,into = c("muestra", "eliminate"),
sep= "_NODE") %>% select(-eliminate)
recuento_final <-recuento_c %>% mutate(Muestra = case_when(startsWith(muestra,
"FP") ~ "Pino", startsWith(muestra, "PQ") ~ "Quercus ilex", startsWith(muestra, "CC") ~
"Castanea sativa")) %>% mutate(Tipo = case_when(endsWith(muestra, "1")
~"Planta",endsWith(muestra, "2") ~"Hongo planta", endsWith(muestra, "3") ~"Hongo
artropodo", endsWith(muestra, "4") ~"Artropodo" )) %>% select(-muestra)

estad <- recuento_final %>% group_by(Muestra)
```

```
# Calculate by type of sample
estad %>% group_by(Tipo)%>% count()

estad %>% group_by(taxa)%>% count()

# Calculate by habitat
estad %>% group_by(taxa) %>%count( Muestra)

estad %>% group_by(taxa) %>%count( Tipo)
```

Nombres nuevos

```
# Rename
new_names <- estad %>% separate(contig,into = c("contig", "eliminate"), sep=
"_length")%>% select(-eliminate)%>% unite(Named_contig, c(contig, cat, taxa), sep='_',
remove = FALSE)
# Obtain only the contig name with the sample and type
new_names_clean <- new_names %>% select(Named_contig, Muestra, Tipo)

# Create .csv

#write.csv(new_names, "new_names.csv")
#write.csv(new_names_clean, "new_names_clean.csv")

# only names
#write.table(new_names$Named_contig, "new_names_only.txt")
```

Heatmap familias

Heatmap familias habitat y familia changed_order

```
family_v3<-read.csv2("C:/Users/Luis/OneDrive - UVa/TFM/Identity/family_list_v2.csv")
%>% select(contig, taxa, name, Muestra, Tipo)
```

```
family_v3 <- family_v3 %>%
  group_by(contig,name) %>% summarise(total_count=n(),.groups = 'drop') %>%
  arrange(contig) %>% mutate(name = factor(name, levels= c( "Hypoviridae",
"Fusariviridae",
  "Sinhaviridae","Astroviridae",
  "Chuviridae","Kitaviridae","Nodaviridae","Polycipiviridae",
  "Polymycoviridae","Reoviridae","Secoviridae","Virgaviridae",
  "Iflaviridae", "Chrysoviridae","Endornaviridae",
  "Unclassified Riboviria","Solemoviridae","Dicistroviridae",
  "Totiviridae","Narnaviridae","Mitoviridae","Tombusviridae",
  "Partitiviridae", "Botourmiaviridae"))))
```

```
##%>%add_row(contig= "CC1", total_count = 0) %>% add_row(contig= "CC2",
total_count = 0)
```

```
contig.nn_v3 <- family_v3$contig %>%
  gsub("PQ", "Q", .) %>%
  gsub("FP", "P", .) %>%
  gsub("CC", "C", .) %>%
  gsub("1", "a", .) %>%
  gsub("2", "ha", .) %>%
  gsub("3", "hi", .) %>%
  gsub("4", "i", .)
```

```

contig.order_v3 <- c("CC4","FP4","PQ4","PQ1","CC1","CC2","PQ3","FP2","CC3","FP1",
"FP3","PQ2") %>%
  gsub("PQ", "Q", .) %>%
  gsub("FP", "P", .) %>%
  gsub("CC", "C", .) %>%
  gsub("1", "a", .) %>%
  gsub("2", "ha", .) %>%
  gsub("3", "hi", .) %>%
  gsub("4", "i", .)

family_v3$contig.nn <- factor(contig.nn_v3, levels = contig.order_v3)
family_v3_tab <- as.data.table(family_v3)

png(filename="heatmap_family.png", type="cairo",
  units="in", width=8, height=5, fontsize=12,
  res=300)

ggplot(melt(cbind(sample=rownames(family_v3_tab), family_v3)),
  aes(x = contig.nn, y = name, fill = factor(value))) +
  geom_tile()+
  theme_bw() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())+
  theme(axis.text.x = element_text(size= 9,face = "bold", angle = 45, vjust = 1,
hjust=1),axis.ticks = element_blank()) +
  ggtitle("Distribución de las familias según el tipo de muestra")+
  theme(plot.title = element_text(size=12, face="bold",
  margin = margin(10, 0, 10, 0), hjust = 0.3)) +
  theme(legend.title = element_text( size=16))+
  theme(legend.position = "left",legend.box = "vertical")+
  scale_fill_viridis(option = "D",discrete= TRUE)+
  labs (y= "Familia", x = "Tipo de muestra", fill = "Número RdRp")

dev.off()

```

Barplot

```

# Count number of sample types
list<- read.csv("new_names_clean.csv")
list_pre <- list %>% separate(Named_contig, into = c("contig", "NODE", "number", "taxa",
"name"), sep = '_') %>% select(-X,-NODE,-number)
list_type <- list_pre %>% group_by(Muestra, Tipo) %>%
summarise(total_count=n(),.groups = 'drop') %>% mutate(Muestra = recode(Muestra,
Pino = "Pinus radiata"))
#write.csv(list_type,"lista_species.csv")

pino <- filter(list_type, "Pinus radiata"== Muestra)
castanea <-filter(list_type, "Quercus ilex"== Muestra)
quercus <-filter(list_type, "Castanea sativa"== Muestra)

####
lista_group <- data.frame (x= list_type$Muestra, y= list_type$total_count, grupo=
list_type$Tipo)

barplot <- ggplot(lista_group, aes(x = x, y = y, fill = grupo)) +
  geom_bar(stat = "identity", color = "black") +
  guides(fill = guide_legend(title = "Tipo")) +

```

```

scale_fill_manual(labels = c("Artrópodo", "Hongo de artrópodo", "Hongo de planta",
"Planta"),
  values = c('black','grey', 'purple2', 'green'))+
labs(y= "Número de RdRp", x = "") +
ggtitle ("Número de RdRp por especie de árbol y tipo de muestra")+
theme(plot.title = element_text(size=8, face="bold"))+
theme_bw()
barplot

```

```

png(filename="recuento_muestras.png", type="cairo",
  units="in", width=6, height=5, fontsize=12,
  res=300)
barplot
dev.off()

```

Tree

```

# Cargar paquetes
if (!require("BiocManager", quietly = TRUE)) install.packages("BiocManager")
BiocManager::install("ggtree")
BiocManager::install("ggmsa")
library(dplyr)
library(ggtree)
library(ggmsa)
library(phytools)

# Cargar árbol
tree <- read.tree("named_tree_correct.nwk")

tree$tip.label <- gsub("_NODE_[0-9]+", "", tree$tip.label)

# Cargar csv y eliminar columna que no sirve de nada
info <- read.csv("new_names_clean.csv")%>% select(-X)
# Cambiar nombre muestra
info$Muestra[info$Muestra == "Pino"] <- "Pino"
info$Muestra[info$Muestra == "Castanea sativa"] <- "Castano"
info$Muestra[info$Muestra == "Quercus ilex"] <- "Encina"
info$Muestra <- as.factor(info$Muestra)
#Acortar nombre
info$Named_contig.nn <- gsub("_NODE_[0-9]+", "", info$Named_contig)
# https://guangchuangyu.github.io/ggtree-book/chapter-ggtree.html

# Sacar nombres del FASTA
# Leer FASTA
tree_names <- get.fasta.name("new_names.txt")
# Acortar nombre
tree_sn <- gsub("_NODE_[0-9]+", "", tree_names)
# Data frame
tree_snf <- as.data.frame(tree_sn)
# Cambiar denominación
tree_nc <- tree_snf %>% separate(tree_sn,into = c("nombre", "cat","name"), sep= "_")
tree_of <- tree_nc$nombre %>% gsub("PQ","E",.) %>%
  gsub("FP", "P", .) %>%
  gsub("CC", "C", .) %>%

```

```

gsub("1", "a", .) %>%
gsub("2", "ha", .) %>%
gsub("3", "hi", .) %>%
gsub("4", "i", .)
habitat <- substring(tree_of, 1, 1)
samplotype <- sapply(tree_of, function(x) substring(x, 2, nchar(x)))

tree_snf$habitat <- as.factor(habitat)
tree_snf$samplotype <- as.factor(samplotype)
cols <- c(ha='purple2', i='black', a='green', hi='grey')

png(filename="tree_oficial_v2.png", type="cairo",
     units="in", width=20, height=25, pointsize=12,
     res=300)

ggtree(tree) %<+% tree_snf +
  geom_tiplab(align=TRUE, offset = 0.25, ignore.negative.edge=TRUE) +

  geom_tippoint(aes(color=samplotype, shape = habitat), size = 5)+
  scale_color_manual(values=cols) +
  scale_fill_discrete(name = "Tipo de muestra", labels = c("Planta", "Hongo planta",
                                                         "Hongo insecto", "Insecto"))+

  theme_tree2() +
  xlim(0, 1.5)

dev.off()

```

