



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Escola Tècnica Superior  
d'Enginyeria Agronòmica i del Medi Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Agronómica  
y del Medio Natural

Detección bioinformática de nuevos agentes subvirales de  
RNA circular con ribozimas

Trabajo Fin de Grado

Grado en Biotecnología

AUTOR/A: Yerga Castelló, Víctor

Tutor/a: Forment Millet, José Javier

Cotutor/a externo: LA PEÑA DEL RIVERO, MARCOS DE

CURSO ACADÉMICO: 2021/2022

# Detección bioinformática de nuevos agentes subvirales de RNA circular con ribozimas

**Autor:** Víctor Yerga Castelló

**Director científico:** Marcos de la Peña del Rivero

**Tutor UPV:** José Javier Forment Millet

**Fecha:** Valencia, Julio de 2022.

## **Resumen:**

Las ribozimas son biomoléculas enzimáticas cuyo centro catalítico está compuesto exclusivamente de ácido ribonucleico (RNA). Las ribozimas pequeñas de autocorte son una familia de ribozimas que se caracterizan por tener en común un pequeño tamaño y hallarse dispersas en los genomas de organismos de todos los reinos, incluyendo agentes subvirales de RNA circular de plantas (viroides) y animales (virus Delta de la hepatitis humana). Las zoonosis víricas han tenido un gran impacto en la salud humana, siendo un ejemplo notable la pandemia COVID-19 actualmente en curso causada por el coronavirus SARS-CoV-2. La identificación y caracterización del viroma de la Tierra es un paso fundamental en la lucha frente a futuras pandemias virales, y el objetivo del siguiente trabajo se centró en la búsqueda bioinformática de nuevos genomas virales de RNA circular en datos de secuenciación masiva obtenidos de diversos orígenes biológicos. Para ello, se analizaron 26 colecciones de datos metagenómicos y metatranscriptómicos procedentes de la base de datos *Sequence Read Archive* mediante la herramienta bioinformática *Infernal* con la finalidad de identificar la presencia de contigs conteniendo pequeñas ribozimas pertenecientes a secuencias de virus de RNA circular tipo Delta virus. Además, se emplearon herramientas informáticas predictivas para conocer la estructura secundaria de sus RNAs y la secuencia de las proteínas que codifican. Esta búsqueda exhaustiva condujo a la identificación de 126 elementos de RNA circular en total, de los cuales 37 pertenecerían virus tipo Delta, 39 a agentes tipo viroide y 50 en los que se agruparían elementos nuevos de genoma circular con ribozimas aún por caracterizar.

## **Palabras clave:**

Ribozima, circRNA, virus de la hepatitis Delta, viroide

# Bioinformatic detection of new circular RNA subviral agents with ribozymes

**Author:** Víctor Yerga Castelló

**Scientific director:** Marcos de la Peña del Rivero

**UPV Tutor:** José Javier Forment Millet

**Date:** Valencia, July 2022

## **Summary:**

Ribozymes are enzymatic biomolecules whose catalytic center is composed exclusively of ribonucleic acid (RNA). Small self-cleaving ribozymes are a family of ribozymes characterized by a common small size and dispersed in the genomes of organisms from all kingdoms, including circular RNA subviral agents from plants (viroids) and animals (human hepatitis Delta virus). Viral zoonoses have had a major impact on human health, a notable example being the currently ongoing COVID-19 pandemic caused by the SARS-CoV-2 coronavirus. The identification and characterization of the Earth virome is a fundamental step in the fight against future viral pandemics, and the aim of the following work focused on the bioinformatic search for novel circular RNA viral genomes in massive sequencing data obtained from diverse biological sources. For this purpose, 26 metagenomic and metatranscriptomic data collections from the Sequence Read Archive database were analyzed using the Infernal bioinformatics tool in order to identify the presence of contigs containing small ribozymes belonging to Delta virus-type circular RNA virus sequences. In addition, predictive computational tools were used to know the secondary structure of their RNAs and the sequence of the proteins they encode. This exhaustive search led to the identification of 126 circular RNA elements in total, of which 37 belong to Delta-like viruses, 39 to viroid-like agents and 50 in which new circular genomes with ribozymes that have not yet been characterized are grouped.

## **Keywords:**

Ribozyme, circRNA, human hepatitis Delta virus, viroid

## **Agradecimientos**

Quisiera darle las gracias a Marcos de la Peña por abrirme las puertas para trabajar con él y darme la oportunidad de realizar este proyecto, que no habría salido adelante de no ser por su ayuda incondicional, por todos sus consejos y explicaciones en los cuales me ha transmitido su pasión por la materia y por el cual ha hecho de la elaboración de este trabajo algo mucho más ameno y entretenido.

También agradecer a mi profesor Javier Forment, por el cual tengo muy buenos recuerdos de su asignatura *Análisis masivo de datos biológicos* en el que fue clave para mi elección del ámbito temático por el cual realizaría posteriormente mi trabajo de fin de grado. Además, le agradezco haber sido mi tutor UPV y por ayudarme a resolver ciertos problemas.

Por último, querría dar las gracias a mis padres, Jota y Sílvia, por todo el cariño, apoyo moral y recursos que me han ofrecido y gracias a los cuales me han ayudado a lograr a desarrollar mi carrera profesional.

# Índice de contenido

<b>1. Introducción</b> .....	<b>1</b>
1.1. Ribozimas.....	1
1.1.1. Ribozimas pequeñas de autocorte .....	2
1.2. RNAs circulares .....	6
1.2.1. Agentes infecciosos con RNA Circular .....	6
<b>2. Objetivo</b> .....	<b>9</b>
<b>3. Material y métodos</b> .....	<b>10</b>
3.1. Datos de secuencias utilizados.....	10
3.2. Búsqueda de genomas subvirales de RNA circular con el software <i>Infernal</i> .....	12
3.2.1. Construcción de un modelo de covarianza con <i>cmbuild</i> .....	12
3.2.2. Calibración del modelo con <i>cmcalibrate</i> .....	12
3.2.3. Búsqueda en los conjuntos de datos con <i>cmsearch</i> .....	12
3.3. Herramientas informáticas predictivas.....	13
3.3.1. Predicción de estructuras secundarias de RNA .....	13
3.3.2. Predicción de marcos de lectura abiertos .....	13
3.3.3. Predicción de la estructura tridimensional proteica .....	13
3.4. Alineación de secuencias y creación de árbol filogenético .....	13
<b>4. Resultados</b> .....	<b>14</b>
4.1. Virus Delta .....	14
4.2. Virus Épsilon.....	18
4.3. Virus Zeta .....	22
4.4. Virus Eta.....	26
4.5. Nuevas entidades de RNA circular con 2 ribozimas .....	27
4.5.1. Agentes HPR-HHR.....	27
4.4.2. Agentes HPR – DVR .....	30
4.4.3. Agentes HHR – DVR .....	30
4.4.4. Agentes HPR – HPR.....	31
4.4.5. TwR-TwR.....	33
4.4.6. TwR-HHR y TwR-HPR .....	34
4.5. Estudio de la estructura tridimensional de las proteínas codificadas por agentes circulares mediante <i>AlphaFold2</i> .....	35
<b>5. Discusión de resultados y conclusión</b> .....	<b>36</b>
<b>6. Referencias</b> .....	<b>38</b>
<b>ANEXO I</b> .....	<b>41</b>
<b>ANEXO II</b> .....	<b>48</b>
<b>ANEXO III</b> .....	<b>61</b>
<b>ANEXO IV</b> .....	<b>81</b>
<b>ANEXO V</b> .....	<b>106</b>

ANEXO VI .....	107
ANEXO VII .....	110

## Índice de figuras

<b>Figura 1.</b> Representación de la estructura de la ribozima hammerhead (HHR). .....	3
<b>Figura 2.</b> Representación de la estructura de la ribozima tipo HDV. ....	4
<b>Figura 3.</b> Secuencia consenso y modelo de estructura secundaria de la ribozima twister. ....	5
<b>Figura 4.</b> Representación esquemática de la ribozima Hairpin (HPR). ....	6
<b>Figura 5.</b> RNA circular infeccioso de plantas angiospermas y satélites virales de RNA circular. ....	7
<b>Figura 6.</b> Ejemplos del RNA circular de virus tipo HDV y de Virus Épsilon. ....	8
<b>Figura 7.</b> Ejemplo de estructuras del genoma de los virus Delta, Épsilon y Zeta. Árbol filogenético de virus Delta. ....	9
<b>Figura 8.</b> Información del hit de un contig con una ribozima tipo HDV detectado por Infernal en SRR12709032, su secuencia y alineamiento emdiante ClustalX con otra secuencia de un virus Delta encontrado en el mismo organismo. ....	15
<b>Figura 9.</b> Información del hit del contig de un virus Delta encontrado en SRR17497039 y árbol filogenético ....	16
<b>Figura 10.</b> Ejemplo de estructura secundaria predicha de uno de los contigs pertenecientes a posibles virus tipo Delta del conjunto de datos SRR6943202 .....	17
<b>Figura 11.</b> Ejemplo de estructura secundaria con forma de varilla predicha de uno de los contigs pertenecientes a posibles virus tipo Épsilon del conjunto de datos SRR6943136. ....	19
<b>Figura 12.</b> Estructura secundaria predicha del contig NODE_149688_length_274_cov_2.328358, perteneciente a un posible virus Épsilon del conjunto de datos SRR17133691. ....	21
<b>Figura 13.</b> Estructura secundaria predicha del contig NODE_16970_length_1022_cov_5.370917, encontrado en SRR17497127 y que podría pertenecer a un virus Épsilon. ....	22
<b>Figura 14.</b> Ejemplo de estructura secundaria predicha de uno de los contigs pertenecientes a posibles virus Zeta del conjunto de datos SRR6973851. ....	24
<b>Figura 15.</b> Estructura secundaria predicha del contig NODE_133762_length_445_cov_24.771505 procedente del conjunto de datos SRR6943136, el cual pertenece a un posible virus Eta. ....	27
<b>Figura 16.</b> Estructuras tridimensionales destacadas predichas por AlphaFold2 de los nuevos elementos de RNA circular. ....	35

## Índice de tablas

<b>Tabla 1.</b> Conjuntos de datos empleados en el que se indica su código en el SRA, el estudio y la muestra de donde procede.....	10
<b>Tabla 2.</b> Contigs pertenecientes a posibles virus tipo Delta encontrados en el conjunto de datos de muestra de humedales (SRR6943202).....	16
<b>Tabla 3.</b> Contigs con características de virus tipo Delta encontrados en el conjunto de datos de una muestra de humedales (SRR6943136).....	17
<b>Tabla 4.</b> Contig encontrado en SRR6943202 perteneciente a un posible virus Épsilon. ....	18
<b>Tabla 5.</b> Contigs con características de virus Épsilon encontrados en el conjunto de datos de una muestra de humedales (SRR6943136).....	18
<b>Tabla 6.</b> Contigs con características de virus Épsilon encontrados en el conjunto de datos de una muestra de humedales (SRR6943113).....	19
<b>Tabla 7.</b> Contigs con características de virus Épsilon encontrados en el conjunto de datos de sedimentos de agua dulce SRR6973926.....	20
<b>Tabla 8.</b> Contig posiblemente perteneciente a un virus Épsilon encontrados en SRR17133691 .....	21
<b>Tabla 9.</b> Contig posiblemente perteneciente a un virus Épsilon encontrados en SRR13675528. ....	21
<b>Tabla 10.</b> Contig de un posible virus Épsilon encontrados en SRR17497127. ....	22
<b>Tabla 11.</b> Contigs con características de virus Zeta encontrados en el conjunto de datos de sedimentos de agua dulce SRR6973851 .....	23
<b>Tabla 12.</b> Contigs con características de virus Zeta encontrados en el conjunto de datos procedente de un biorreactor anaeróbico (SRR6962709). ....	24
<b>Tabla 13.</b> Contigs con características de virus Zeta encontrados en el conjunto de datos de muestra de humedales (SRR6943202).....	24
<b>Tabla 14.</b> Contigs encontrados en SRR13675528 .....	26
<b>Tabla 15.</b> Contig procedente de un posible virus Eta detectado en el conjunto de datos de muestra de humedales (SRR6943136).....	27
<b>Tabla 16.</b> Contigs de tipo HPR-HHR detectados en SRR6943202, SRR6943136, SRR6943113, SRR6973926 y SRR6973851.....	28
<b>Tabla 17.</b> Contig con hits HPR-DVR encontrado en SRR694313 .....	30
<b>Tabla 18.</b> Contigs de tipo HHR-DVR detectados en SRR6943136 y SRR6973926. ....	31
<b>Tabla 19.</b> Listado de contigs de tipo HPR-HPR y los conjuntos de datos respectivos en los que fueron encontrados.....	31
<b>Tabla 20.</b> Lista de contigs tipo TwR-TwR detectados en SRR6943136 y SRR6943113.....	34
<b>Tabla 21.</b> Contig TwR-HHR y TwR-HPR encontrado en SRR6943113.....	34

## Abreviaturas

<b>aa</b>	Aminoácidos
<b>HDV</b>	Virus Delta de la hepatitis humana
<b>DVR</b>	Ribozima virus Delta de la hepatitis humana
<b>HHR</b>	Ribozima Hammerhead
<b>hhrbz_dv4</b>	Ribozima Hammerhead tipo III
<b>HPR</b>	Ribozima Hairpin
<b>ORF</b>	Marco de lectura abierto (Open Reading Frame)
<b>SRA</b>	Sequence Read Archive
<b>TwR</b>	Ribozima Twister



# 1.Introducción

## 1.1. Ribozimas

Las ribozimas son biomoléculas enzimáticas cuyo centro catalítico está compuesto exclusivamente de ácido ribonucleico (RNA). La primera ribozima descubierta fue una del tipo intrón del grupo I, descrito hace 40 años por parte de Cech y su grupo, en el que describieron la actividad catalítica de un intrón de RNA ribosomal que le confería la capacidad de autocorte, también llamado actividad en *cis* (Kruger et al., 1982). Dicho descubrimiento supuso un hito que rompía con uno de los dogmas centrales de la biología molecular al conocerse que las proteínas no eran las únicas biomoléculas capaces de catalizar reacciones químicas. Un año más tarde, se demostró que la actividad catalítica en *trans* de la ribonucleasa P (RNAasa P) de *E. coli* residía en su subunidad ribonucleica (Guerrier-Takada et al., 1983). La RNAasa P es una enzima ribonucleasa compuesta por dos subunidades, una proteica y la otra ribonucleica, esta última encargada de catalizar el corte de una secuencia precursora de tRNA para dar lugar a un tRNA maduro. Además, se observó que la RNAasa P permanecía intacta al finalizar la reacción, es decir, presentaba capacidad de recambio, una característica similar a las enzimas, lo que hizo que estas moléculas se les denominará con el término “ribozimas” para diferenciarlas de las enzimas proteicas tradicionales.

Entre otros avances, el descubrimiento de las ribozimas permitió establecer la hipótesis del mundo prebiótico de RNA (Gilbert, 1986), la cual afirma que las primeras formas de vida primitivas eran RNAs encargados tanto de la función de catalizar reacciones químicas como de almacenamiento de información genética necesaria para su replicación. Posteriormente, el papel de estas funciones sería reemplazado por las proteínas y el DNA respectivamente, quedando el RNA como intermediario entre estos dos. Esta hipótesis fue reforzándose conforme se fue observando que las ribozimas tenían más protagonismo en la naturaleza de lo que se pensaba, ya que se fue descubriendo que están implicadas, por ejemplo, en el autoprocesamiento de intrones de grupo II (Peebles et al., 1986), en la catálisis del enlace peptídico de los ribosomas (Nissen et al., 2000), y en el spliceosoma eucariótico (Valadkhan y Manely, 2001).

Hoy en día se han identificado múltiples clases de ribozimas naturales y se han descrito sus estructuras con alta resolución y además de sus mecanismos de acción (Müller et al., 2016), e incluso se han diseñado ribozimas artificiales con interés por sus aplicaciones biotecnológicas. Normalmente las ribozimas se clasifican en función de su tamaño y mecanismo de acción.

En cuanto al mecanismo de acción de las ribozimas hay que destacar que, a excepción del ribosoma que cataliza la formación de enlaces peptídicos para la síntesis de proteínas, todas las ribozimas naturales descritas hasta ahora se encargan de realizar la catálisis de rotura y/o ligación del enlace fosfodiéster del ácido ribonucleico (Müller et al., 2016). En función del mecanismo de acción, por tanto, se pueden clasificar la mayoría de las ribozimas naturales en dos tipos: las de corte y ligación (ribozimas de *splicing*); y las de corte (ribozimas de corte), cuya actividad puede ser tanto en *cis* como en *trans*. (Serganov & Patel, 2007). Esta catálisis es posible gracias a la complementariedad de bases y por las interacciones secundarias y terciarias existentes en la estructura de RNA de la propia ribozima, de manera que su estructura permite la conformación de un núcleo activo y la especificidad del autocorte en un sitio concreto de la molécula (Doherty & Doudna, 2001).

Las ribozimas también se pueden clasificar en función del tamaño. Por lo general las ribozimas de gran tamaño (150-3000 nucleótidos) catalizan reacciones que generan productos con hidroxilos en 3' y fosfatos en 5'. Por otro lado, las de pequeño tamaño (35-150 nucleótidos) catalizan reacciones que generan productos con un fosfato cíclico en 2'-3' y un hidroxilo en el extremo 5' (Tanner, 1999). En ambos casos se el número neto de enlaces se conserva durante la reacción.

### 1.1.1. Ribozimas pequeñas de autocorte

Las ribozimas pequeñas de autocorte es una familia de ribozimas que se caracterizan por tener en común su pequeño tamaño (entre los 35 y los 150 nucleótidos), realizar la escisión entre dos nucleótidos concretos; utilizar el mismo mecanismo para producir la rotura; y encontrarse de forma natural en algunos genomas y secuencias de RNA, en el que suelen estar involucradas en el proceso de replicación de estos (Puerta-Fernández et al., 2003).

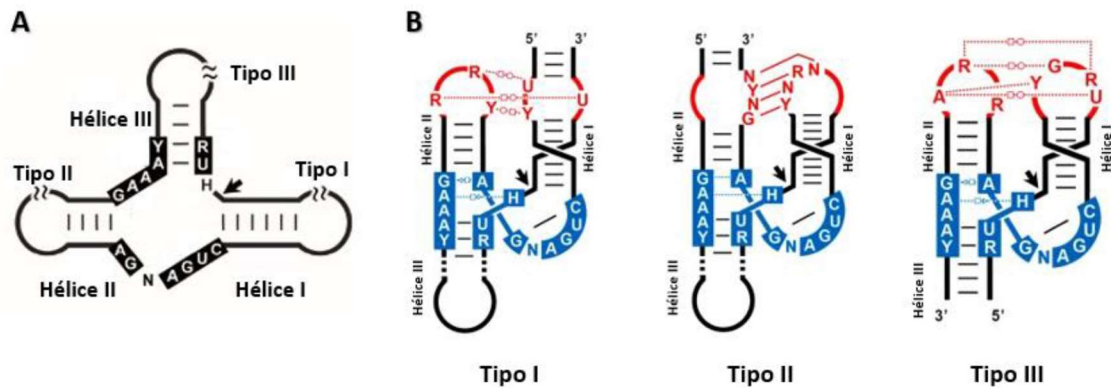
Dentro de esta familia de ribozimas se han descubierto hasta el momento nueve clases distintas: las ribozimas de cabeza de martillo (*hammerhead* o HHR) (Prody et al. 1986), ribozimas del virus Delta de la hepatitis humana (DVR) (Sharmeen, et al., 1988), ribozimas de horquilla (*hairpin* o HPR) (Buzayan et al. 1986), ribozima del satélite varkud (VS) (Saville & Collins, 1990), ribozima de la glucosamina-6-fosfato sintasa (glmS) (Winkler et al. 2004), *twister* (TwR) (Roth et al., 2014), y las ribozimas *twister sister* (TwSR), *hatchet* y *pistol* (Weinberg et al., 2015). A su vez esta familia de ribozimas se puede dividir en dos grupos, dependiendo de si su centro activo se forma por una hélice irregular abierta (como las ribozimas *hammerhead*, *hairpin*, VS, TwS, *hatchet* y *pistol*) o por un doble pseudonudo (DVR, glmS y Tw) (Lilley, 2019).

Durante años estas ribozimas se han considerado remanentes del mundo de RNA con una presencia restringida fundamentalmente a genomas de RNA circular de ciertos agentes infecciosos subvirales como viroides y satélites virales de plantas. No obstante, hace unos años se descubrió que las ribozimas pequeñas de autocorte se hallan en realidad dispersas en los genomas de organismos de todos los reinos, desde virus y procariontes hasta eucariotas, formando parte de elementos genéticos transponibles como los retrotransposones, y especialmente las ribozimas HHR, DVR y TwR (Hammann et al., 2012). Más recientemente, se han encontrado multitud de ribozimas *hairpin* o HPR en nuevos RNAs circulares distribuidos en diferentes contextos genéticos (Weinberg, et al., 2021).

#### 1.1.1.1. Ribozimas *hammerhead* (HHR)

La ribozima de cabeza de martillo, también llamada *hammerhead* o HHR, fue la primera ribozima descrita en la literatura dentro de las nueve clases de ribozimas pequeñas de autocorte. Se descubrieron por primera vez como un motivo catalítico en el RNA de algunos virus satélites de plantas (Prody et al., 1986) y en viroides (Hutchins et al., 1986), participando en el procesamiento de los transcritos multiméricos resultantes de la replicación por círculo rodante típica de estos agentes infecciosos. Nuevas HHRs fueron encontradas posteriormente en regiones repetitivas de genomas de animales como tritones (Epstein et al., 1987), tremátodos (Ferbeyre et al., 1998) o grillos (Rojas et al., 2000), y más adelante se descubrió que en realidad se hallan distribuidas en organismos de todos los reinos de la vida (De la Peña y García-Robles, 2010a), incluyendo el genoma del ser humano (De la Peña y García-Robles, 2010b), como parte de diversos intrones génicos (García-Robles et al., 2012), convirtiendo a la HHR en una de las ribozimas pequeñas de autocorte con mayor presencia en la biosfera.

La ribozima *hammerhead* debe su nombre a la similitud de su estructura secundaria con la morfología de un tiburón de cabeza de martillo. Dicha estructura está comprendida por tres hélices, denominadas I, II y III, las cuales rodean a un centro catalítico compuesto de 15 nucleótidos conservados que son esenciales para que se produzca la catálisis. Estas hélices, a su vez, están cerradas por unos bucles terminales, los bucles 1, 2 y 3, cada uno contenido en su respectiva hélice con la misma enumeración (Figura 1A). Se ha descrito una interacción terciaria del bucle 1 con el bucle 2, que aumenta la cinética de la reacción de autocorte varios órdenes de magnitud, demostrando su implicación en la catálisis de la reacción (De la Peña et al. 2003). La reorganización del centro activo provocada por las mencionadas interacciones terciarias favorece la adopción de una estructura en forma de letra Y, confirmado tras la resolución de su estructura cristalográfica en 2006 (Martick et al., 2006). Por último, esta ribozima existe de forma natural en tres diferentes topologías dependiendo de qué hélice presente los extremos 3' y 5' del RNA: tipo I, II y III (Figura 1B)



**Figura 1. Representación de la estructura de la ribozima hammerhead (HHR).** Los nucleótidos altamente conservados se muestran resaltados (de color negro en A de azul en B). La letra R indica una bases púricas y la letra Y una base pirimidínica. (A) Representación clásica de la estructura secundaria de HHR formada por las tres hélices (I, II y III) que rodean el centro catalítico compuesto por 15 nucleótidos. (B) Representación esquemática de la estructura terciaria de los tres tipos de HHR naturales (I, II y III). Las interacciones terciarias que se producen entre la hélice I y la hélice II aparecen de color rojo. Las interacciones de Watson-Crick se indican con líneas continuas mientras que las líneas punteadas indican apareamientos no Watson-Crick. Imagen modificada de De la Peña et al. (2017).

En cuanto a las funciones biológicas de las ribozimas HHR, hoy en día no son del todo claras, aunque en eucariotas parecen estar íntimamente ligadas a diversos tipos de elementos transponibles, tanto autónomos, como por ejemplo los elementos tipo Penélope (Cervera y De la Peña, 2014), como no autónomos, como por ejemplo los retrozimas de animales (Cervera y De la Peña, 2020) y plantas (Cervera et al., 2016). También se han hallado otras HHRs altamente conservadas dentro de regiones no codificantes de genes de vertebrados superiores, en intrones (De la Peña y García-Robles, 2010b) y dentro de 3' UTRs (Martick et al., 2008). Análisis bioinformáticos sugieren que estas ribozimas intrónicas podrían ofrecer una nueva forma de regulación genética adquirida a través de la domesticación de las ribozimas de retroelementos (García-Robles et al., 2012). La característica común en todas las HHRs es la capacidad de autocorte durante su síntesis, algo que en procariontes podría usarse en el procesamiento de RNAs policistrónicos, la generación de pequeños RNA regulatorios o en la generación de nuevas pautas de lectura mediante la actividad de múltiples HHRs (Hamman et al., 2012).

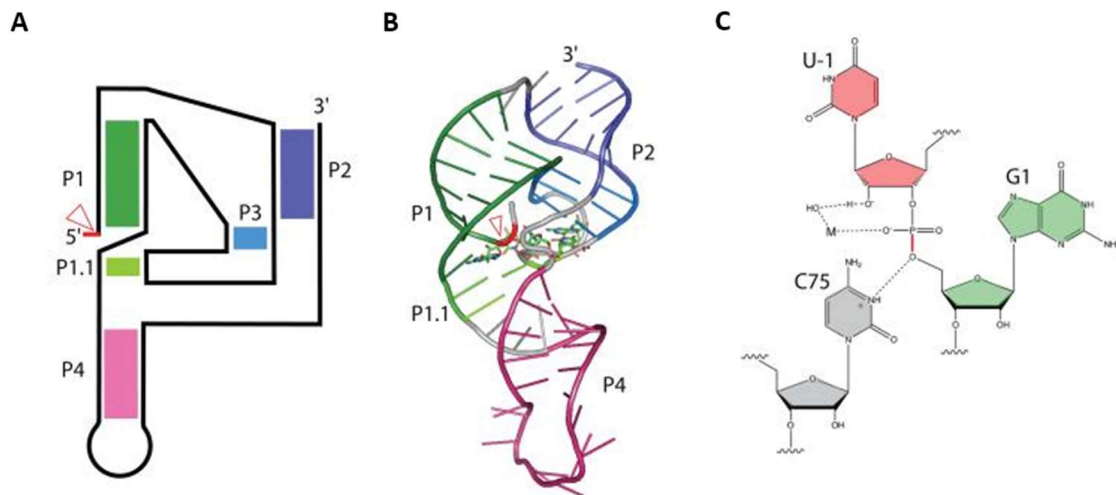
### 1.1.1.2. Ribozimas tipo virus Delta de la hepatitis humana (DVR)

El virus Delta de la hepatitis humana (HDV) es un virus satélite del virus de la Hepatitis B humana (HBV), y que se caracteriza por tener un genoma de RNA de pequeño tamaño (~1700 nucleótidos), circular y monocatenario, además de la presencia de ribozimas de autocorte en ambas polaridades del RNA genómico (Been y Wickham 1997). En el apartado 1.2.1.2 se hablará con más profundidad del HDV. Las ribozimas del HDV (DVR) asisten en el procesamiento de las copias multiméricas del genoma viral, producto de la replicación mediante el mecanismo de replicación de círculo rodante.

En 2006 se descubrió experimentalmente la presencia de una ribozima tipo HDV conservada en el intrón del gen CPEB3 de mamíferos (Salehi-Ashtiani et al., 2006), aunque su función es desconocida. Más adelante, mediante análisis bioinformáticos, se descubrió una gama de ribozimas muy similares a la DVR en multitud de genomas de organismos muy diversos en el árbol de la vida (Webb et al., 2009), las cuales a menudo se localizaban en las regiones 5' UTR de retrotransposones tipo LINE, indicando una posible implicación de estas ribozimas en el procesamiento de elementos móviles (Eickbush y Eickbush, 2010). Estudios del procesamiento del extremo 5' de retrotransposones LINE sugieren que estas ribozimas tipo HDV actuarían en distintos niveles durante la retrotransposición (Ruminski et al., 2011), ya sea procesando el extremo 5' del RNA del transposón, o actuando como promotores en la

traducción del marco de lectura abierto (ORF) del retroelemento, y facilitando su inserción en un nuevo locus genómico (Bibillo *et al.*, 2004; Ruminski *et al.*, 2011).

En cuanto a su estructura secundaria, la ribozima tipo HDV se compone por cinco regiones tipo hélice (P1, P1.1, P2, P3, y P4), formando un doble pseudonudo entre P1 y P2 (Figura 2).



**Figura 2. Representación de la estructura de la ribozima tipo HDV.** Los colores verdes oscuro, verde claro, morado, azul y rosa indican las regiones helicoidales P1, P1.1, P2, P3, y P4 respectivamente. **(A)** Forma esquemática la estructura canónica de las ribozimas tipo HDV. El triángulo rojo indica el lugar de escisión, mientras que las bandas de colores marcan cada hélice. **(B)** Representación de la ribozima HDV mínima cristalizada, los colores indican los diferentes tipos regiones helicoidales, y en rojo se indica la región de la escisión. **(C)** Modelo del sitio de autocorte de la ribozima HDV en el que se resaltan las interacciones de los nucleótidos esenciales en la reacción. Modificado de Jiménez *et al.*, 2015.

### 1.1.1.3. Ribozimas twister (TwR)

La ribozima twister (TwR) se descubrió en 2014 mediante análisis bioinformáticos como un motivo de RNA autocatalítico altamente conservado en genomas de procariontes y de algunos eucariotes (Roth *et al.*, 2014). La estructura mínima de esta ribozima consta de una gran hélice interrumpida por bucles, los bucles L1, L2 y L3, los cuales están separados por tres regiones apareadas (hélices P1, P2 y P4), estando el punto de corte localizado en el bucle L1. Además, la mayoría de los motivos encontrados presentaban una hélice extra, la P3 en el bucle L2, y, en un porcentaje menor, la hélice P5 o P0. (Figura 3). De forma similar a lo descrito para la ribozima HHR, dependiendo de qué hélice contenga los extremos 5' y 3' del RNA, TwR puede presentar tres conformaciones posibles, denominadas tipo P1, tipo P3 y tipo P5. La denominación "twister" es debido a análisis de esta ribozima en el que predominaban una estructura terciaria en forma del jeroglífico egipcio "twisted flax" debido a la existencia de interacciones terciarias entre los nucleótidos de los bucles L1, L2 y L3, que permitían la formación de pseudonudos.

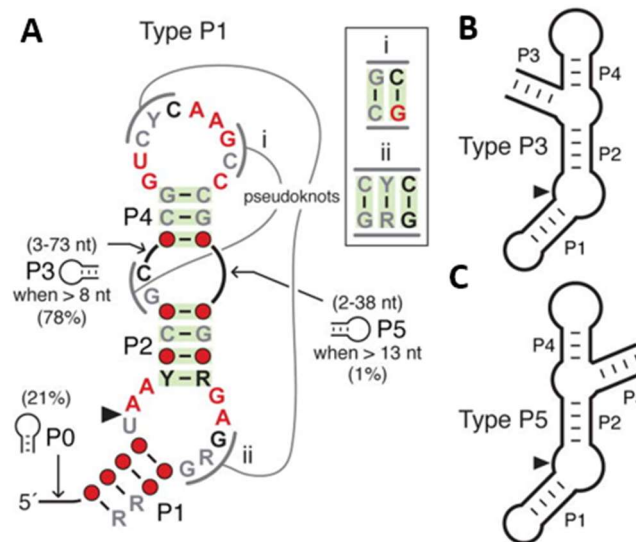
### 1.1.1.4. Ribozimas Hairpin (HPR)

La ribozima Hairpin (HPR) fue descubierta por primera vez como un motivo catalítico en la polaridad negativa del RNA del satélite viral del virus *Tobacco ringspot* (sTRSV) (Feldstein *et al.*, 1989; Hampel y Tritz, 1989) participando en reacciones de escisión y ligación en el proceso de replicación. Además del satélite sTRSV, también se han encontrado ejemplos de ribozimas HPR en otros virus satélites de RNA como el *Chicory yellow mottle* (sChYMV) y el *Arabidopsis mosaic* (sARmV) (Rubino *et al.*, 1990; Kaper *et al.*, 1988). Hay que destacar que se ha visto que en estos satélites que poseen una ribozima HPR siempre tienen asociada una HHR en el RNA de la cadena opuesta, que se encargaría del procesamiento de los RNAs multiméricos de la polaridad positiva durante el proceso de replicación por círculo rodante simétrico. Tanto HHR como HPR participan en procesos de escisión-ligación, no obstante, se ha determinado que

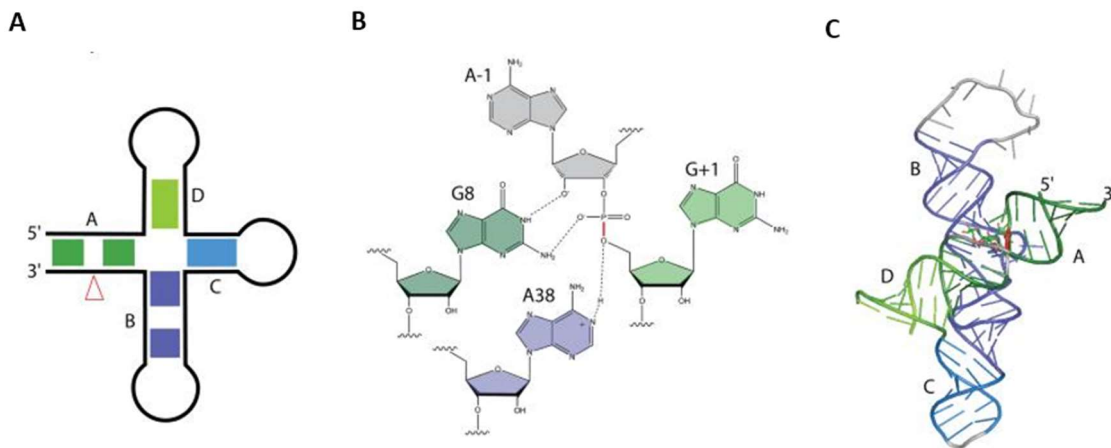
la HPR tiene una mejor actividad ligasa tanto *in vivo* como *in vitro*, mientras que la ribozima HHR favorece la escisión sobre la ligación de productos (Fedor, 2000).

A diferencia de las ribozimas HHR, DVR y TwR, los cuales se encuentran ampliamente distribuidos en genomas de diversos organismos, la HPR parece ser una ribozima exclusiva de satélites virales de RNA circular, ya que únicamente se ha descrito en las cadenas de polaridad negativa de tres RNAs satélites de nepovirus.

La estructura de la HPR se organiza en torno a la unión de 4 hélices, los cuales están nombrados de forma secuencial en el sentido de las agujas del reloj: A, B, C y D (Figura 4A). Las hélices A y B forman el centro activo de la ribozima gracias a la presencia de bucles que contienen todos los nucleótidos conservados esenciales para la actividad catalítica (Jiménez *et al.*, 2015). Además, en la hélice A se encuentra el sitio de rotura y unión de la ribozima (Figura 4B). Las hélices C y D tienen una función estructural y de aporte estabilidad a la ribozima. Este último hecho se ha confirmado al estudiar la estructura tridimensional mediante cristalografía de la HPR en dos estados distintos, con y sin las hélices C y D (Figura 4C) (Grum-Tokars, *et al.*, 2003), (Rupert y Ferré-D'Amare 2001). Pese a que la forma mínima sin las hélices C y D es activa, su plegamiento parece requerir dos o tres órdenes de magnitud superiores de  $Mg^{2+}$  respecto a la forma completa de la HPR, encontrándose el equilibrio interno entre la escisión y la ligación desplazándolo hacia la ligación (Fedor, 1999).



**Figura 3. Secuencia consenso y modelo de estructura secundaria de la ribozima twister. (A)** Modelo de consenso detallado basado en 2690 ribozimas twister representadas en su configuración de tipo P1, en el que la cadena de RNA comienza y termina en la base de la hélice P1. El triángulo de color negro indica el sitio de escisión. Los nucleótidos de color gris, negro y rojo designan una conservación de al menos 75, 90 y 97 %, respectivamente; las posiciones en las que la identidad de los nucleótidos está menos conservada se representan mediante círculos. El sombreado verde denota pares de bases pronosticados respaldados por covariación natural. Las notaciones i y ii identifican los pseudonudos predichos. Los números entre paréntesis significan las longitudes variables de las secuencias enlazadoras de las hélices indicados, así como el porcentaje de ribozimas en el que las pueden presentar. R e Y denotan purina y pirimidina, respectivamente. **(B)** y **(C)** Estructura consenso en conformación P3 y P5 respectivamente. Modificado de Roth *et al.*, 2014.



**Figura 4. Representación esquemática de la ribozima Hairpin (HPR).** Los colores verde oscuro, morado, azul y verde claro representan las hélices A, B, C y D respectivamente. **(A)** Forma esquemática la estructura canónica de las HPR. El triángulo rojo indica el lugar de escisión, mientras que las bandas de colores marcan cada hélice. **(B)** Modelo del sitio de autocorte de la HPR donde se resaltan las interacciones de los nucleótidos clave en la reacción. **(C)** Representación de la HPR mínima cristalizada, en rojo se indica la región de la escisión.

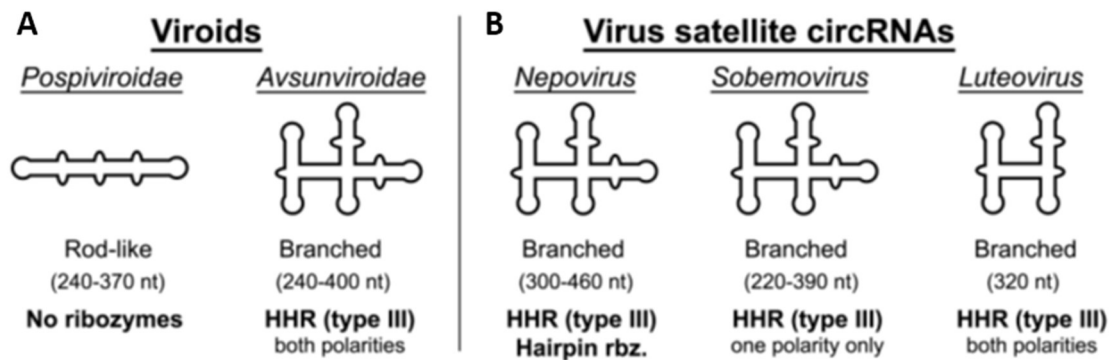
## 1.2. RNAs circulares

Los RNAs circulares hace referencia a aquellas moléculas de RNA cerradas de forma covalente en sus extremos. A diferencia de los DNAs circulares que se encuentran en presentes en muchos ámbitos de la vida, los RNAs circulares siempre han sido tratados como elementos peculiares. Esto ha ido cambiando a lo largo de los años con el descubrimiento de nuevos ejemplos de RNAs circulares de diverso origen. Hoy en día se han descrito varios tipos de RNAs circulares, algunos de los cuales ya se han mencionados anteriormente como agentes infecciosos con genomas de RNA circular, los intrones de grupos I y II autocortantes, los RNAs circulares intermediarios de retrotransposones no autónomos con ribozimas (retrozimas) y los RNAs circulares intrónicos/exónicos (Lasda y Parker, 2014). Los RNA circulares aparecen muy frecuentemente en la mayor parte de los transcriptomas celulares, y muestran propiedades distintas a las del RNA lineal, como lo es su alta estabilidad frente a exonucleasas o su capacidad de realizar una replicación eficiente mediante el mecanismo de círculo rodante.

### 1.2.1. Agentes infecciosos con RNA Circular

#### 1.2.1.1. Viroides y virus satélites de plantas

Los primeros RNA circulares descubiertos fueron un grupo de agentes infecciosos de plantas llamados viroides y RNAs satélite de virus en los años 70 (Sanger et al., 1976; Randles et al., 1981). Estos grupos subvirales infectan algunas plantas angiospermas, ya sea de manera autónoma en el caso de los viroides o en presencia de un virus auxiliar en el caso de los satélites virales de RNA (Daròs y Flores, 1995). Todos ellos tienen un genoma de pequeño tamaño (220-400 nucleótidos) de RNA circular que se caracteriza por no codificar proteínas (Flores et al., 2015) pero sí de codificar una ribozima en una o en ambas de sus polaridades, a excepción de los miembros de la familia de viroides *Pospiviroidae*, que no codifican ninguna ribozima (Figura 5) (Jimenez et al., 2015).



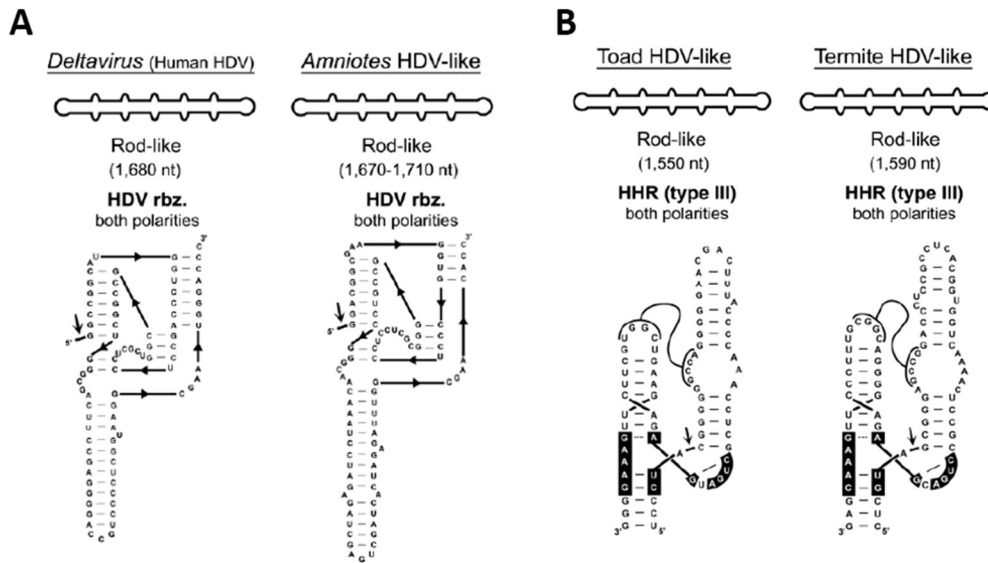
**Figura 5.** RNA circular infeccioso de plantas angiospermas como viroides (A) y satélites virales de RNA circular (B). La conformación global de RNA, tamaños típicos y clases de ribozimas se muestran indicadas en cada caso. Modificado de de la Peña et al., 2020.

### 1.2.1.2. El virus Delta de la hepatitis humana (HDV) y otros virus tipo HDV

El virus Delta hepatitis humana (HDV) es un agente infeccioso único de tipo virus satélite cuyo genoma está compuesto por un genoma de RNA circular de cadena única de un tamaño de 1680 nucleótidos, con una alta autocomplementariedad (Botelho-Souza et al. 2017). Este virus satélite codifica sólo dos proteínas, el antígeno Delta de la hepatitis pequeño y el grande (S- y L-HDAg respectivamente), ambos derivados de un único marco de lectura abierto. Además, el HDV contiene un motivo característico de RNA catalítico autocortante, la ribozima del virus de la hepatitis Delta (DVR), tanto en las polaridades genómica como antigénica (WP et al. 1988; Wu et al. 1989; Riccitelli y Lupták 2013). El HDV depende del virus de la hepatitis B (HBV) para el ensamblaje de las partículas virales, la liberación de la célula huésped y la entrada en nuevas células (Botelho-Souza et al. 2017).

El HDV era un virus que desde su descubrimiento en los años 70 solo se había detectado en humanos y el único representante del género Deltavirus. Sin embargo, datos recientes han revelado la presencia de RNAs circulares similares a HDV altamente divergentes en muestras de diversas especies de metazoos, que van desde amniotas (reptiles, aves y mamíferos) hasta anfibios (un tritón y un sapo) e invertebrados (una termita) (Wille et al. 2018; Chang et al. 2019; Hetzel et al. 2019; Paraskevopoulou et al. 2020; Bergner et al. 2021), lo que parece indicar que este tipo de virus tiene una historia evolutiva más compleja de lo que se pensaba. A diferencia del HDV humano, ninguno de los genomas de los virus tipo HDV descritos se ha encontrado asociado con un hepadnavirus co-infectante.

Estos nuevos virus tipo HDV recientemente descritos comparten muchas características con su homólogo humano. Todos ellos tienen genomas de RNA circular monocatenario de aproximadamente 1500-1700 nucleótidos, que se pliegan en estructuras tipo varillas (Figura 6A), y codifican para un antígeno Delta con 13-55% de identidad aminoácidos al humano. Sin embargo, en el caso del motivo DVR, aunque su presencia se ha reportado para los virus tipo HDV procedentes de amniotas (Wille et al. 2018; Hetzel et al. 2019; Paraskevopoulou et al. 2020), no ha sido así en el caso de los anfibios, peces e invertebrados, los cuales no se encontraron secuencias compatibles con DVRs (Chang et al. 2019).



**Figura 6.** (A) Ejemplos del RNA circular de virus tipo HDV encontrado en humanos (izquierda) y en amniotas (derecha). (B). Ejemplos de Virus Épsilon encontrados en un sapo (izquierda) y en termitas (derecha). Además, se ilustra también un ejemplo de la ribozima antígenómica en cada caso. La conformación global del RNA, los tamaños típicos y la clase de ribozimas se indica para cada ejemplo. Modificado de de la Peña et al., 2020.

### 1.2.1.3. Virus Épsilon y Zeta

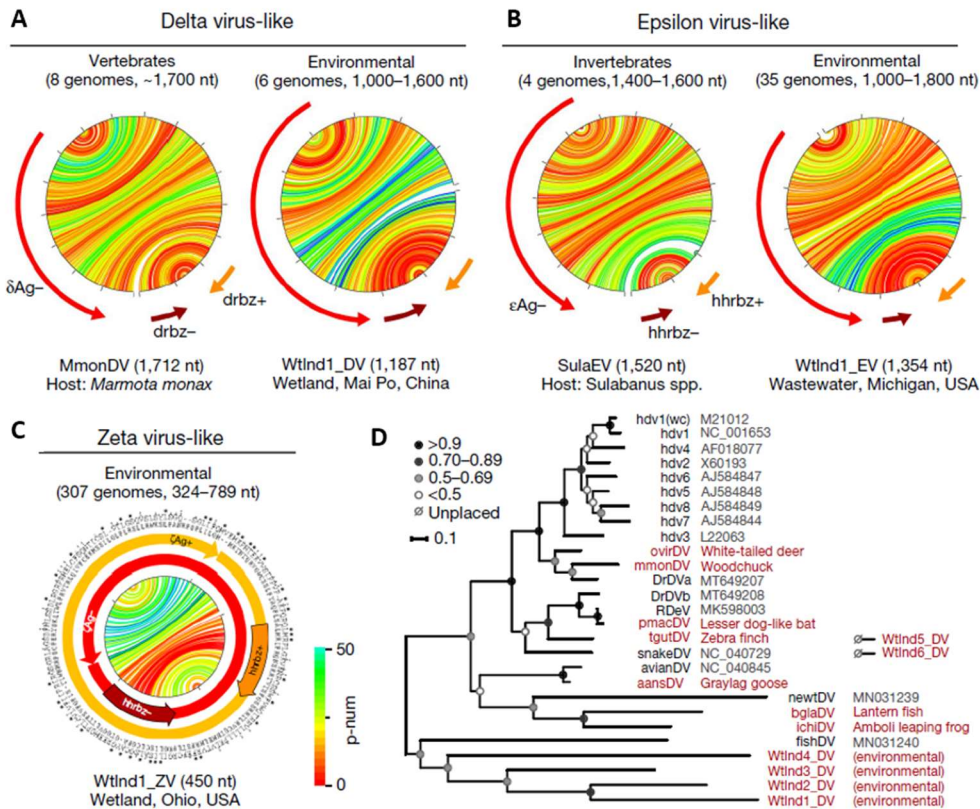
Recientemente se ha detectado mediante búsquedas bioinformáticas la presencia de secuencias conservadas de HHR en ambas polaridades de los RNAs circulares de virus tipo HDV procedentes de sapos y termitas (de la Peña et al., 2021). Estos virus fueron bautizados más tarde como virus Épsilon (Edgar et al., 2022), y se caracterizan por ser una variante de los Deltavirus que tienen HHRs de tipo III que presentan similitudes entre sí como hélices más largas de lo habitual e interacciones de pseudonudo conservadas (Figura 6B).

Más recientemente se ha publicado un estudio en el que, mediante búsqueda de secuencias de ribozimas y de la proteína del antígeno Delta, se identificaron 14 virus Delta y 39 virus Épsilon nuevos. Pero, además, también se encontraron 311 secuencias enigmáticas de RNA circular con ribozimas HHR y una sintenia similar a la de los virus Delta/Épsilon, las cuales se englobaron bajo el nombre de virus Zeta (Figura 7) (Edgar et al., 2022).

En el mencionado estudio los autores vieron que los genomas circulares de los virus Zeta eran altamente compactos, con un tamaño que oscila entre 324 y 789 nucleótidos y se predice que se pliegan en estructuras tipo varilla. Contienen una HHR en cada orientación y suelen codificar dos ORFs, uno en sentido positivo y el otro en negativo. Lo más peculiar es que ambas ORFs generalmente carecen de codones de parada y abarcan todo el genoma, produciendo potencialmente una repetición en tándem interminable del antígeno.

Los autores indicaron que las secuencias estos nuevos virus Delta, Épsilon y Zeta encontrados podrían compartir un ancestro común, esto es sugerido por el hecho de que el dominio típico de “coiled-coil” del antígeno HDV (Zuccola et al., 1998) se conserva en los antígenos de los nuevos virus Delta y Épsilon, mientras que los genomas de virus Épsilon y Zeta muestran HHRs análogas. Estos elementos abundantes podrían ayudar a resolver la cuestión de los orígenes de los agentes subvirales de RNA circulares en eucariotas superiores, considerados históricamente como fósiles moleculares del hipotético mundo de RNA prebiótico (Flores et al., 2014).





**Figura 7.** (A) Estructura del genoma del virus Delta de *Marmota monax* (MmonDV) y un genoma similar al virus Delta detectado en un conjunto de datos ambiental, cada uno conteniendo un ORF en sentido negativo del antígeno Delta ( $\delta Ag$ ); dos ribozimas Delta (drbz); y plegado característico en forma de varilla, donde cada línea muestra el apareamiento de bases predicho dentro del RNA genómico. (B) Estructura genómica para el virus tipo Epsilon de *Sulabanus* spp. (SulaEV) y de un virus tipo Épsilon procedente de un conjunto de datos ambientales, cada uno con un antígeno Épsilon ( $\epsilon Ag$ ) de sentido negativo; dos ribozimas de cabeza de martillo (hhrbz). (C) Ejemplo de la estructura del genoma compacto de un virus tipo Zeta a partir de un conjunto de datos ambientales que contiene dos antígenos zeta predichos ( $\zeta Ag^{+/-}$ ; la alineación de proteínas se muestra en los círculos exteriores) ORFs sin codones de parada; dos hhrbzs superpuestos con los ORFs; y plegado tipo varilla. (D) Árbol filogenético de máxima probabilidad de virus Delta derivado de una alineación proteica Delta-antígeno con valores bootstrap. Todavía no se han podido colocar dos virus ambientales Delta divergentes. Modificado de Edgar et al., 2022.

## 2. Objetivo

Las zoonosis se definen como cualquier enfermedad o infección transmisible de manera natural entre los animales y las personas, directa o indirectamente. Las zoonosis víricas han tenido un gran impacto en la salud humana en el siglo actual, siendo un ejemplo notable la pandemia COVID-19 actualmente en curso derivada de la enfermedad causada por el coronavirus SARS-CoV-2. Las estimaciones apuntan a  $3 \times 10^5$  especies de virus de mamíferos de las que pueden surgir enfermedades infecciosas en humanos (Anthony et al., 2013), de las cuales solo se conoce una fracción en la actualidad.

La identificación del viroma de la Tierra es un fundamental en la preparación para una próxima pandemia vírica (Johnson et al., 2020), por tanto, el objetivo del siguiente trabajo es la búsqueda de nuevos virus de RNA circular similares al HDV en conjuntos de datos medioambientales mediante herramientas bioinformáticas para así contribuir en la vigilancia global de la diversidad de virus la cual es necesaria para mejorar la predicción y prevención de futuras pandemias.

### 3. Material y métodos

#### 3.1. Datos de secuencias utilizados

La base de datos *Sequence Read Archive* (o SRA) es el principal archivo de datos de secuenciación de alto rendimiento perteneciente a los Institutos Nacionales de Salud (NIH) de los Estados Unidos, cuya función es almacenar datos de secuenciación sin procesar e información de alineación de plataformas de secuenciación de alto rendimiento, y ponerlos a disposición de la comunidad de investigación para mejorar la reproducibilidad y permitir nuevos descubrimientos mediante la comparación de conjuntos de datos (The Sequence Read Archive, 2007).

Se han seleccionado 26 conjuntos de datos metatranscriptómicos procedentes de SRA para su análisis (Tabla 1). Estos conjuntos de datos en concreto se obtuvieron a partir de los análisis con la infraestructura *Serratus* (Edgar et al., 2022), en el que se utilizó tanto el Transcriptome Shotgun Assembly (TSA) de NCBI como ensamblados de datos del SRA cuyas secuencias contenían posibles agentes tipo deltavirus por homología con el antígeno Delta.

**Tabla 1. Conjuntos de datos empleados en el que se indica su código en el SRA, el estudio y la muestra de donde procede.**

Conjunto de datos SRA	Estudio	Muestra
SRR6943113	Metatranscriptome of peatland microbial communities from Houghton, MN, USA - PEATcosm2014_Bin01_30_metaT	Metagenoma de humedales
SRR6943136	Metatranscriptome of peatland microbial communities from Houghton, MN, USA. - PEATcosm2014_Bin06_10_metaT	Metagenoma de humedales
SRR6943202	Metatranscriptome of peatland microbial communities from Houghton, MN, USA - PEATcosm2014_Bin10_10_metaT	Metagenoma de humedales
SRR6962709	Anaerobic biogas reactor microbial communities from eattle, Washington, USA - Biogas_R2-A RNA time zero metatranscriptome	Metagenoma de biorreactor
SRR6973851	Groundwater sediment microbial communities from an aquifer adjacent to Colorado River, USA - East River metaT ERMLIBT47_16_1Ra metatranscriptome	Metagenoma de sedimentos de agua dulce
SRR6973926	Groundwater sediment microbial communities from an aquifer adjacent to Colorado River, USA - East River metaT ERMLT660_1_16_1Ra metatranscriptome	Metagenoma de sedimentos de agua dulce
SRR11266555	Transcriptome analysis of developmental stages of cocoa pod borer, <i>Conopomorpha cramerella</i> : A polyphagous insect pest of economic importance in Southeast Asia.	Crisálida de <i>Conopomorpha cramerella</i>
SRR11266556	Transcriptome analysis of developmental stages of cocoa pod borer, <i>Conopomorpha cramerella</i> : A polyphagous insect pest of economic importance in Southeast Asia.	Polilla de <i>Conopomorpha cramerella</i>

SRR12709032	Brain transcriptomic responses of <i>Sceloporus jarrovii</i> to social stimuli	Cerebro de macho adulto de <i>Sceloporus jarrovii</i>
SRR12709062	Brain transcriptomic responses of <i>Sceloporus jarrovii</i> to social stimuli	Cerebro de macho adulto de <i>Sceloporus jarrovii</i>
SRR12709063	Brain transcriptomic responses of <i>Sceloporus jarrovii</i> to social stimuli	Cerebro de macho adulto de <i>Sceloporus jarrovii</i>
SRR13255034	Transcriptomic analyses of the termite, <i>Cryptotermes secundus</i> , reveal a gene network underlying a long lifespan and high fecundity [JH manipulation experiment]	Muestra genérica de <i>Cryptotermes secundus</i> , reina (tratamiento con solventes)
SRR13255047	Transcriptomic analyses of the termite, <i>Cryptotermes secundus</i> , reveal a gene network underlying a long lifespan and high fecundity [JH manipulation experiment]	Muestra genérica de <i>Cryptotermes secundus</i> , reina (tratamiento precoceno)
SRR13426401	Disentangling the aging network of a termite queen	Muestra genérica de <i>Cryptotermes secundus</i> , reina primaria, más de 5 años, ID de muestra DC24-3c
SRR13480240	<i>Myiopsitta monachus</i> Genome sequencing and assembly	RNA-seq del cerebro anterior de perico monje argentino ( <i>Myiopsitta monachus</i> ) hembra
SRR13675528	Volatile organic compounds mediated antifungal activity of <i>Pichia</i> and its effect on the metabolic profiles of fermentation communities	RNA-Seq de fermentación metagenómica: granos fermentados
SRR13756016	Metatranscriptome sequencing of hospital surfaces before and after AIONX cleanSURFACE intervention	Metagenoma hospitalario
SRR13945426	<i>Ectropis grisescens</i> isolate:XBB-Egri Genome sequencing	<i>Ectropis grisescens</i>
SRR14753726	<i>Cydia pomonella</i> Raw sequence reads	Antenas de macho adulto de <i>Cydia pomonella</i>
SRR14753728	<i>Cydia pomonella</i> Raw sequence reads	Antenas de hembra adulta de <i>Cydia pomonella</i>
SRR14753729	<i>Cydia pomonella</i> Raw sequence reads	Cabeza de macho adulto de <i>Cydia pomonella</i>
SRR15533330	Raw reads and assembly of chromosome fusion in stickleback	Transcriptoma de Individuo hembra de <i>Apeltes Quadracus</i>
SRR17133691	<i>Pinctada fucata</i> Raw sequence reads	RNA-seq of <i>Pinctada fucata</i>
SRR17497039	Meta-transcriptomic analysis of rodent animals.	Metagenoma de organismo completo, HB-AH-F-1
SRR17497127	Meta-transcriptomic analysis of rodent animals.	Metagenoma de organismo completo, MH-LN-MO-1

### 3.2. Búsqueda de genomas subvirales de RNA circular con el software Infernal

Infernal es un programa utilizado para buscar motivos de RNA estructural en bases de datos de secuencias y para realizar alineaciones de secuencias de RNA basadas tanto en secuencias como estructuras (Nawrocki y Eddy, 2013). Infernal crea un perfil a partir de una alineación de secuencias múltiples anotadas estructuralmente de una familia de RNA con un sistema de puntuación específico de posición para sustituciones, inserciones y eliminaciones. El sistema de puntuación de Infernal se caracteriza por considerar la estructura secundaria, además de la secuencia primaria de la familia que se modela. Los perfiles de Infernal son modelos probabilísticos llamados “modelos de covarianza”, un tipo especializado de gramática estocástica libre de contexto (SCFG) (Lari y Young, 1990).

#### 3.2.1. Construcción de un modelo de covarianza con *cmbuild*

Se comienza proporcionando a Infernal un archivo con alineaciones múltiples de secuencias de RNA de ribozimas de autocorte. Infernal requiere que esté en el formato *Stockholm* y que incluya una anotación de estructura secundaria de consenso. Para ello se emplea el comando *cmbuild* que crea un modelo de covarianza a partir del archivo introducido con la alineación y lo guarda en un archivo de extensión “.cm” que es tratado como modelos compilados de la alineación de ribozimas.

#### 3.2.2. Calibración del modelo con *cmcalibrate*

El siguiente paso es calibrar el modelo que se ha obtenido en el paso anterior. Este paso debe realizarse antes de usar el modelo para búsquedas en la base de datos con el comando *cmsearch*. En este paso, los parámetros estadísticos necesarios para informar los *E-values* (valores esperados) se estiman y almacenan en el archivo CM. Cuando *cmsearch* se utiliza posteriormente para una búsqueda en la base de datos y se encuentra un resultado con una puntuación  $x$ , el *E-value* de ese resultado es el número de resultados que se espera que obtengan una puntuación  $x$  o más simplemente por casualidad (dado el tamaño de la búsqueda que se esté realizando).

#### 3.2.3. Búsqueda en los conjuntos de datos con *cmsearch*

A continuación, se ha usado el modelo ya calibrado de nuestras ribozimas para buscar homólogos en una base de datos mediante el comando *cmsearch*. La base de datos empleada en este caso ha consistido en los datos listados en el apartado 3. 1.. Se ha ejecutado *cmsearch* de manera que el programa devuelva un archivo de extensión “.tblout” que consiste en una tabla que recopila todos los *hits* con ribozimas de autocorte que se ha encontrado en un determinado conjunto de datos, donde en cada fila informa del nombre del *contig* que ha dado positivo para un determinado acierto o *hit*, el nombre del tipo ribozima con el que se ha producido el hit, posición de la secuencia, *E-value*, etc.

La información del *E-value* es muy importante ya que es lo que va a determinar la veracidad del hit. Es la significancia estadística del acierto: el número de aciertos que esperaríamos obtener en una base de datos de este tamaño (medido por el número total de nucleótidos) si la base de datos contuviera solo secuencias aleatorias no homólogas.

Por lo general se ha considerado un hit como altamente significativo en aquellos casos que tengan un *E-value* menor o igual a  $\sim 10^{-5}$ , con la excepción de los hits para las ribozimas tipo DVR y hairpin que se han considerado como veraces con valores menores o iguales a 0,1 y 1 respectivamente, esto es debido a que dichas ribozimas tienen una alta variabilidad. Además de tener un *hit* suficientemente bajo para una ribozima, para que un *contig* sea considerado como un posible candidato a un agente subviral, se comprobó si dicho *contig* presentaba ribozimas duplicadas en ambas polaridades y/o circularidad en su secuencia (repeticiones en los extremos 5' y 3' producto del ensamblado de una molécula circular).

### **3.3. Herramientas informáticas predictivas**

#### **3.3.1. Predicción de estructuras secundarias de RNA**

Se ha utilizado la página web RNAfold WebServer para predecir estructuras secundarias de secuencias de RNA monocatenario. Está basado en ViennaRNA Package 2.4.18, que consta de una biblioteca de código C y varios programas independientes para la predicción y comparación de estructuras secundarias de RNA (Lorenz et al., 2011). Se ha usado la opción del algoritmo de mínima energía libre el cual produce una única estructura óptima.

También se ha empleado el servidor web UNAFold (*Unified Nucleic Acid Folding*) para la obtención de las estructuras secundarias de secuencias de RNA en formato de gráfico circular. El paquete de software UNAFold es una colección integrada de programas que simulan vías de plegamiento, hibridación y fusión para una o dos secuencias de ácido nucleico monocatenario (Markham y Zucker, 2008)

#### **3.3.2. Predicción de marcos de lectura abiertos**

Se usó la herramienta ORFfinder de la página web del NCBI para busca marcos de lectura abiertos (ORF) en la secuencia de DNA que ingresa. El programa devuelve el rango de cada ORF, junto con su traducción de proteínas.

También se usó *getorf* de EMBOSS, un programa de línea de comandos que al introducir un archivo input fasta con secuencias de DNA devuelve un archivo output fasta con secuencias de aminoácidos de las posible ORFs codificadas por el mismo.

#### **3.3.3. Predicción de la estructura tridimensional proteica**

Se ha empleado AlphaFold 2, desarrollado por la empresa DeepMind, para predecir la estructura 3D de una proteína dada a partir de su secuencia de aminoácidos. (Jumper et al., 2021; Varadi et al., 2021). AlphaFold 2 deduce la estructura tridimensional de las proteínas con la comparación, mediante aprendizaje profundo haciendo uso de una red neuronal artificial, de estructuras y secuencias de aminoácidos ya conocidas, y a partir de esa comparación es capaz de predecir la forma de proteínas cuya estructura se desconoce, pero de las que se sabe su secuencia. La mayoría de las predicciones de AlphaFold 2 arrojan resultados muy similares a los que se obtienen a partir de las observaciones mediante cristalografía de rayos X. (Callaway, 2020).

### **3.4. Alineación de secuencias y creación de árbol filogenético**

Se ha usado ClustalX 2.1 para la alineación de secuencias. ClustalX es una interfaz de Windows para el programa de alineación de secuencias múltiples ClustalW. Proporciona un entorno integrado para realizar múltiples alineaciones de secuencias y perfiles y analizar los resultados (Larkin et al., 2007).

En cuanto a la creación del árbol filogenético se ha empleado el programa MEGA X: *Molecular Evolutionary Genetics Analysis* versión X (Kumar et al., 2018). MEGA es un programa para realizar análisis estadísticos de la evolución molecular y para construir árboles filogenéticos. Incluye muchos métodos y herramientas sofisticados para la filogenómica y la filomedicina.

## 4. Resultados

### 4.1. Virus Delta

Nuestros análisis bioinformáticos con los datos TSA detectaron un contig con un primer *hit* duplicado de ribozimas tipo DVR con unos *E-values* muy bajos en datos de un experimento con el reptil *Sceloporus jarrovi* (SRR12709032), que indicaban que podría pertenecer al genoma de un virus tipo Delta (Figura 8A). En un conjunto de datos limpio y ya ensamblado de otro experimento realizado con el mismo organismo se encontró también la secuencia de un virus tipo Delta (Figura 8B). Se alineó mediante Clustal X 2.1 esta secuencia con el contig encontrado en el conjunto de datos SRR12709032 y su reversa complementaria, donde se confirmó que secuencias pertenecen al mismo virus. Además, cabe destacar que se detectaron dos sustituciones de base en el alineamiento, lo cual es indicativo de que eran dos cepas distintas del mismo virus (Figura 8C). Dichos genomas de RNA circular y con ribozimas codificaban un homólogo del antígeno Delta del virus HDV humano.

En el conjunto de datos SRR17497039, procedente de un experimento con animales roedores, se encontró un nuevo *hit* para la ribozima del virus Delta con un *E-value* muy bajo en el contig NODE\_58632\_length\_829\_cov\_3.743386 (Figura 9A). No se llegó a detectar circularidad en dicho contig. Al realizar un árbol filogenético con la secuencia casi completa del antígeno delta codificado de este contig junto con otros virus Delta se ha visto que se encuentra cercano filogenéticamente con un virus Delta de marmota (Figura 9B). Esto concuerda con el hecho de que el conjunto de datos proviene de muestra de órganos del organismo *Marmota himalayana*, un mamífero muy cercano a *Marmota marmota* en el que se encontró recientemente un ejemplo de virus tipo Delta (Edgar et al. 2022). Aunque muy probablemente se trata de una secuencia procedente de un virus tipo Delta, ésta está incompleta (longitud del contig no circular de solo 829 nucleótidos), lo que explica porque no se ha encontrado un *hit* duplicado para la ribozima HDV. No podemos saber su secuencia completa con la información que nos da el ensamblaje del conjunto de datos, debido a su baja cobertura.

Por otro lado, en un conjunto de datos procedente de un experimento metagenómico de humedales (SRR6943202) se encontraron siete contigs cada uno de ellos con un *hit* duplicado para la ribozima DVR y con una longitud de secuencia similar a la de un virus tipo Delta (Tabla 2). Se ha introducido la secuencia de estos contigs con el programa getorf y solo tres de ellos han resultado tener un ORFs con un tamaño mayor a 100 aa, lo cual no es lo habitual en virus tipo Delta, y podrían considerarse como nuevos tipos de virus Delta no canónicos. En otro set de datos proveniente del mismo experimento que el anterior (SRR6943136) se encontraron también cinco contigs con características de virus Delta (circularidad, *hit* de ribozimas duplicado, tamaño de secuencia cercano) (Tabla 3). Además, las estructuras secundarias predichas en todos estos contigs también tenían la forma de varilla característica de dichos virus (Figura 10 y ANEXO I).

**A**

```
Query:          DVR4 [CLEN=86]
Hit scores:
rank  E-value  score  bias  sequence                               start  end  mdl  trunc  gc  description
-----
(1)  2.6e-05  54.6   0.0   NODE_85216_length_453_cov_2.031579    238    331  +   cm      no 0.68  -
(2)  0.00052  47.8   0.0   NODE_85216_length_453_cov_2.031579    453    368  -   cm      5  0.65  -
```

**B**

```
>NODE_85216_length_453_cov_2.031579

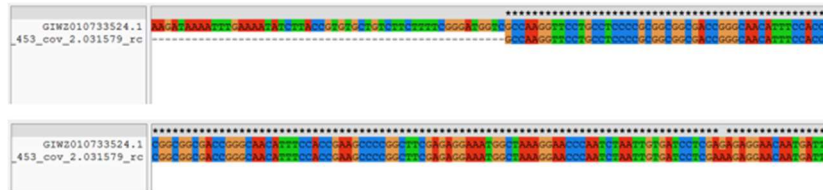
GGAGTGAACCCACTCTGGTTCTTGGGGCCCCGCCGCCGGAGAAATCCCTTCGGACTGAACCTGCAGGGAGTTCCAGTGACTGAAGGACCACCCGGGGA
CCTTGGACATAAGAGGGGACCGACAGTTCGAAAAGGGGACTGAGAATTTGGGGACTCGGAGGGTGGAAAACAGGGGTAAAGAGGTAATAAGAAAAATT
GAAAATCTTACCTTTGTGTCTGTTTCTTCAATTCGGGCCGCCAAGGAAGACTCACCTCCACGTGGTGGCCCTGGGCATCCACCCGAGGCCCGCA
GCTCTGTAGAGGGATGGCTGAGGAGGCTCAATCATTGTTCTCTTTCGAGGATCAACAATTAGATTGGTCTCTTATGCCATTTCTCTCGAAGCCGGG
GCTTCGGTGGAAATGTTGCCGGTCCGCCGCCGGGGAGGAGGAACCTTGCC

>NODE_85216_length_453_cov_2.031579_rc

GCCAAGGTTCTGCTCCCGCGCGGCCGACCCGGGCAACATTTCCACCAGAGCCCGGCTTCGAGAGGAAATGGCTAAAGGAACCCAAATCTAATTGTGAT
CCTCGAAAAGGGAACAATGATTGAGACCTCTCCAGCCATCCCTACGAGAGCTCGGGCCCTCGAGTGGATGCCAGGCCGCCACCCAGGTGGAGGTGAGG
TCTTCTTGGCGGCCGAAGAATGAAGAAACAGACACAAAAGGTAAGATTTTCAATTTCTTATACCTCTTTACCCCTTTGTTTCCACCTCCGAGTC
CCCAAATTTCTAGTCCCTTTTCAAGAACTGTCGGTCCCTCTTATGTCCAAGGTCGCCGGTGGTCTTCAGTCACTGGAACCTCCCGAGGTTCACTCC
AAGGGATTTTCTCGCGCGGCCGGGCCCCAAGAACCAGGAGTGGGTTCACTCC

>GIWZ010733524.1 TSA: Sceloporus jarrovi TRINITY_DN69151_c0_g1_i1, transcribed RNA sequence

GGGGGAAAGTGGGAAAAGGATAAATAAGATAAAATTTGAAATATCTTACCGTGTCTGCTCTTCTTTTCGGGATGGTCCCAAGGTTCTGCTCCCTCC
CGCGGC GCGGACCGGGCAACATTTCCACCAGAGCCCGGCTTCGAGAGGAAATGGCTAAAGGAACCCAAATCTAATTGTATCTCGAGAGGGAACAATG
ATTGAGACCTCTCAGCCATCCCTTACGAGAGCTCGGGCCCTCGAGTGGATGCCAGGCCGCCACACGTGGAGGTGAGGCTCTTCTGGCGGCCCGAA
GAATGAAGAAACAGACACAAAAGGTAAGATTTTCAATTTCTTATTACCTCTTTACCCCTTTGTTTCCACCTCCGAGTCCCAAAATCTCGATCCCTC
CTCAGAACTGTCGGTCCCTCTTATGTCCAAGGTCGCCGGTGGTCTTCAGTCACTGGAACCTCCGAGGTTCAAGTCCGAAAGGATTTTCTCGCCGG
GCGGGGCCCCAAGAACCAGGAGTGGGTTCACTCCCGAGCCATTCACCTCCTAGAAAGGCCCGAGTACTGCTGTCAGGACCGCGCGGAGCGCCAT
CAGCTGTCCTGGTCCCGGCCGAGCATCTCCCTCTCTTTGATAGTTGATTTCCCTTGTGTTAAGAGACTTGTACTCTTTGACTCTCCCGTTC
TCTCTGGAAGATGTGTTCCCTCTGCGCCGCTGCGTTTTCTCGGGGGTCCCGGACATCCATGGGAGATTCCTCTTTACCTTCTTGGGGTCACT
CCCCGTCCGCGGCCGGTTCCTTGTATAATGCCAAGGATGTTACCAGGCCAAGGATGTCGGCTCGAACTTACGTTTCCGTCGCCTTAGCTTCGCCAA
CTTCTCCTAGTCTCCGAATCTCTGCGTGGTGGCGCCCATTTTTCAAGGATCGAGCTCGACTCTCTTCTTGGGCGTCCCGGCTCTTGGAGATC
TTCTCGGGAGCGGTCGCCATCTCTCTTCTTACTTTTATTGAGAAATAGAGAGAGAGGAAGAGCGCACCAACTCCAGAGGTTCCAAAGCGCCGGAGGAC
GGTCCAGACAGGAAGCGAGATCAACCTTCAGAAAAGGGAACACCCCGAGCGAGAGGAGCCATGGAGGAGATTGTGCGCGAGCTGC AAGAGCACTGGA
AAAGTTAAAGGTACAGGCCGCGCATAGGCTCTCAACACCTGGCTTATCGGAATTTCCGGGCCCGGTGACTAGTGGAGTAAGACCTCCAGAGGGGAGGA
TGGATCGACAGATTCATCGGGGACCTTACGATCCGCTCGGATCCAGGATTTGAAACACCTCTGCGCCGGAATGGGGAGATCAGTACGAGAGGCA
ATCTCGATTTACTCGGAAACCTTATCGATGAGGAGGAGGAAGTCTCGAGAGAGGCCACGCACTGCGCGGCCGCCCACTCCGGAAGCGGAC
GGGAGCGACTCTAGAGGCCGCCGGGAGTTCACCTCTACTGGGAGGCGTCTCCCGACCGTGGGATCCAGGAAAAGAACCCAGCGGAGCGAGG
CTTACAGCCCGAGGGGACAGGAC
```

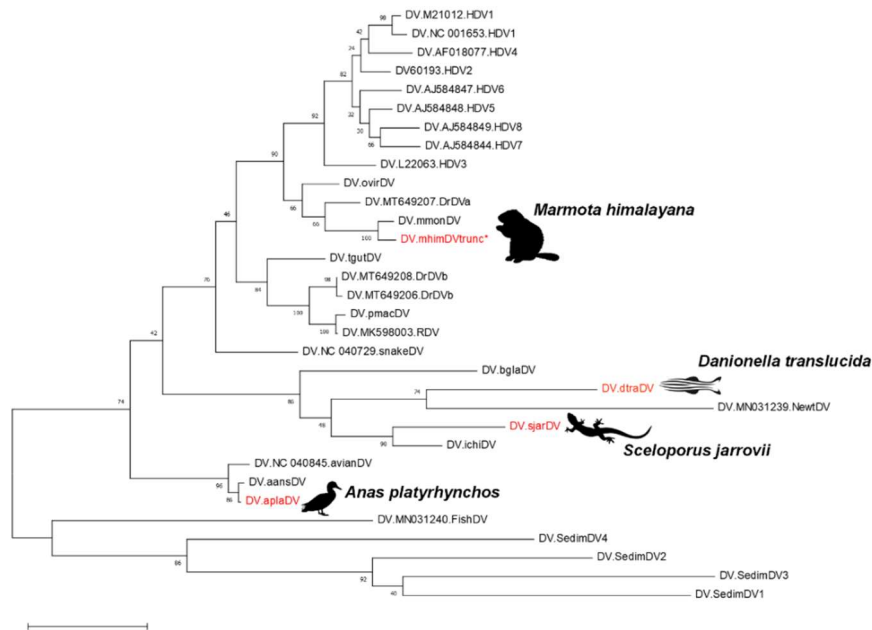
**C**

**Figura 8. (A)** Información del hit de un contig (*NODE\_85216\_length\_453\_cov\_2.031579*) con una ribozima tipo DVR detectado el conjunto de datos SRR12709032. **(B)** Secuencias en formato FASTA del contig anterior (arriba), de su reversa complementaria (centro) y del contig GIWZ010733524.1 TSA correspondiente al genoma de un virus tipo Delta detectado en otro conjunto de datos (abajo). **(C)** Alineamiento realizado con ClustalX 2.1, en la imagen de arriba se ve un fragmento de la parte lateral izquierda del alineamiento de la reversa complementaria de *NODE\_85216\_length\_453\_cov\_2.031579* con GIWZ010733524.1 TSA mientras que la de abajo se ve un fragmento de la parte central. Se puede observar un alineamiento casi perfecto en la parte central de GIWZ010733524.1 TSA, por lo que se puede afirmar que ambas secuencias pertenecen al mismo virus, teniendo GIWZ010733524.1 TSA la secuencia completa del mismo. Nótese que se observa un mismatch en la imagen de abajo que indica una mutación por sustitución de base.

A

#target name	query name	from	seq to st	rand t	runc p	ass gc	bias	score	E-value inc
NODE_58632_length_829_cov_3.DVR4		203	117-		no	1.0.59	0.0	71.2	3e-08!

B



**Figura 9. (A)** Información del hit del contig NODE\_58632\_length\_829\_cov\_3.743386 con una ribozima tipo DVR detectado por Infernal dentro del conjunto de datos SRR17497039. **(B)** Árbol filogenético de máxima verosimilitud de virus Delta derivado del alineamiento de la proteína del antígeno delta con valores Bootstrap. El virus Delta encontrado en Marmota himalayana (DV.nhimDVtrunc\*) se encuentra muy cercano en cuanto origen filogenético respecto al virus Delta de Marmota monax (DVmmon.DV).

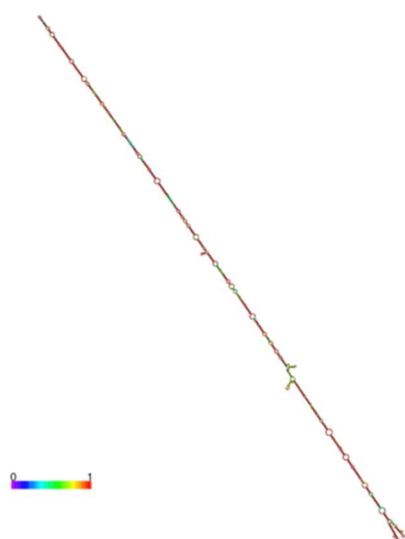
**Tabla 2. Contigs pertenecientes a posibles virus tipo Delta encontrados por Infernal en el conjunto de datos de muestra de humedales (SRR6943202).** Los contigs subrayados son aquellos que codifican ORFs con un tamaño de >100 aa. DVR4 = Ribozima tipo virus Delta de hepatitis humana.

Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
<u>NODE_25045_length_1154_cov_11.748381</u>	DVR4	217	281	+	no	0.0021
	DVR4	357	281	-	no	0.014
<u>NODE_2637_length_2831_cov_16.006889</u>	DVR4	868	946	+	no	0.26
	DVR4	338	261	-	no	0.83
<u>NODE_27947_length_1103_cov_54.734951</u>	DVR4	346	410	+	no	0.00078
	DVR4	485	412	-	no	0.081
<u>NODE_28970_length_1087_cov_18.590730</u>	DVR4	322	386	+	no	0.0032
	DVR4	460	387	-	no	0.018
<u>NODE_31272_length_1053_cov_13.721429</u>	DVR4	898	830	-	no	0.00059
	DVR4	724	797	+	no	0.025
<u>NODE_32796_length_1033_cov_8.283333</u>	DVR4	583	503	-	no	2.6e-05
	DVR4	226	311	+	no	0.026

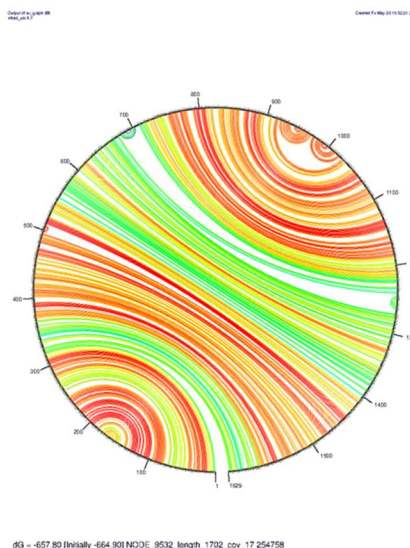


NODE_9532_length_1702_cov_17.254758	DVR4	700	615	-	no	0.00062
	DVR4	1397	1478	+	no	0.0094

A



B



**Figura 10. Ejemplo de estructura secundaria predicha de uno de los contigs pertenecientes a posibles virus tipo Delta del conjunto de datos SRR6943202. (A) Muestra la estructura secundaria MFE (energía mínima libre) del contig NODE\_9532\_length\_1702\_cov\_17.254758 mientras que (B) representa la estructura secundaria del mismo contig pero en formato de gráfico circular. Los colores de las estructuras indican las probabilidades de emparejamiento de bases.**

**Tabla 3. Contigs con características de virus tipo Delta encontrados en el conjunto de datos de una muestra de humedales (SRR6943136). Los contigs en cursiva indican que tienen ORFs con un tamaño de >100 aa.**

Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
<i>NODE_11069_length_1102_cov_19.470360</i>	DVR4	526	462	-	no	0.00054
	DVR4	387	460	+	no	0.056
<i>NODE_11442_length_1089_cov_199.755906</i>	DVR4	481	417	-	no	0.0037
	DVR4	120	180	+	no	0.58
<i>NODE_11514_length_1087_cov_6.013807</i>	DVR4	46	110	+	no	0.0018
	DVR4	184	111	-	no	0.012
<i>NODE_12695_length_1052_cov_3.222676</i>	DVR4	146	214	+	no	0.00041
	DVR4	320	247	-	no	0.018
<i>NODE_5953_length_1375_cov_4.884793</i>	DVR4	191	260	+	no	0.00072
	DVR4	533	465	-	no	0.048

## 4.2. Virus Épsilon

En el conjunto de datos SRR6943202 se encontró un contig con características de virus Épsilon (hit duplicado para la ribozima Hammerhead tipo III, tamaño de 1000-2000 nucleótidos, circular) pero no tenía ningún ORF con un tamaño relevante (>100 aminoácidos) ni tampoco una estructura secundaria con forma de varilla (ANEXO II)

**Tabla 4. Contig encontrado en SRR6943202 perteneciente a un posible virus Épsilon. hhrbz\_dv4 = Ribozima Hammerhead tipo III.**

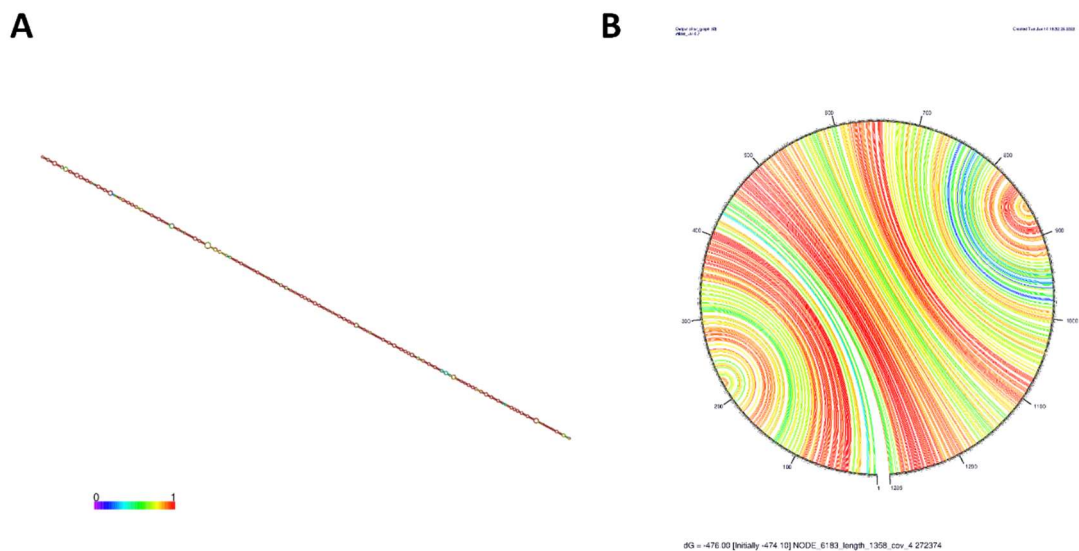
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
NODE_22667_length_1203_cov_19.969912	hhrbz_dv4	432	385	-	no	6.5e-05
	hhrbz_dv4	9	60	+	no	7.1e-05

En el conjunto de datos SRR6943136 se encontraron 13 contigs con características de virus Epsilon, de los cuales 8 presentaban ORFs de un tamaño de >100 aminoácidos (Tabla 5) y 7 de ellos una estructura secundaria predicha con forma de varilla. (Figura 11 y ANEXO II)

**Tabla 5. Contigs con características de virus Épsilon encontrados en el conjunto de datos de una muestra de humedales (SRR6943136). Los contigs subrayados indican que tienen ORFs de un tamaño de >100 aminoácidos. Los contigs resaltados en amarillo indican estructura secundaria con forma de varilla.**

Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
<b>NODE_11411_length_1090_cov_14.008850</b>	hhrbz_dv4	563	616	+	no	0.00042
	hhrbz_dv4	981	909	-	no	0.0011
NODE_11896_length_1074_cov_89.564436	hhrbz_dv4	513	565	+	no	4.1e-05
	hhrbz_dv4	113	62	-	no	8.6e-05
<u>NODE_12112_length_1068_cov_9.292462</u>	hhrbz_dv4	774	723	-	no	5.2e-05
	hhrbz_dv4	584	633	+	no	0.00026
<u>NODE_12119_length_1068_cov_6.289447</u>	hhrbz_dv4	123	174	+	no	2.1e-06
	hhrbz_dv4	357	308	-	no	3,00E-06
NODE_12393_length_1060_cov_3.627153	hhrbz_dv4	387	435	+	no	3.7e-06
	hhrbz_dv4	841	806	-	no	0.012
<b>NODE_13708_length_1022_cov_16.571128</b>	hhrbz_dv4	65	143	+	no	9.1e-15
	hhrbz_dv4	430	354	-	no	2.7e-10
NODE_14088_length_1012_cov_7.358892	hhrbz_dv4	293	340	+	no	2.5e-06
	hhrbz_dv4	461	413	-	no	8.5e-06
<u>NODE_14349_length_1006_cov_67.151125</u>	hhrbz_dv4	547	495	-	no	8,00E-05
	hhrbz_dv4	381	432	+	no	0.00013
<b>NODE_2689_length_1827_cov_8.539339</b>	hhrbz_dv4	1663	1576	-	no	1.4e-12
	hhrbz_dv4	1226	1308	+	no	1.4e-12

<b>NODE_29706_length_767_cov_353.659942</b>	hhrbz_dv4	118	195	+	no	6.1e-12
	hhrbz_dv4	472	399	-	no	3.9e-09
<b>NODE_5526_length_1414_cov_64.239374</b>	hhrbz_dv4	851	768	-	no	7.1e-14
	hhrbz_dv4	587	673	+	no	5.2e-13
<b>NODE_6183_length_1358_cov_4.272374</b>	hhrbz_dv4	479	369	-	no	2,00E-12
	hhrbz_dv4	1	72	+	5	1.3e-05
<b>NODE_9094_length_1184_cov_15.320432</b>	hhrbz_dv4	146	35	-	no	1.1e-11
	hhrbz_dv4	960	1074	+	no	1.4e-11



**Figura 11. Ejemplo de estructura secundaria con forma de varilla predicha de uno de los contigs pertenecientes a posibles virus tipo Épsilon del conjunto de datos SRR6943136. (A) Muestra la estructura secundaria MFE (energía mínima libre) del contig NODE\_6183\_length\_1358\_cov\_4.272374 mientras que (B) representa la estructura secundaria del mismo contig pero en formato de gráfico circular. Los colores de las estructuras indican las probabilidades de emparejamiento de bases.**

En el conjunto de datos SRR6943113, procedente de un experimento de metagenómica de humedales, se encontraron 5 contigs con características de virus Épsilon, aunque se caracterizan por tener un tamaño de secuencia menor de lo habitual en estos virus. De los 5 contigs, 4 han obtenido una estructura predicha con forma de varilla (ANEXO II) y 3 codifican una o varias ORFs de un tamaño >100 aminoácidos (Tabla 6).

**Tabla 6. Contigs con características de virus Épsilon encontrados en el conjunto de datos de una muestra de humedales (SRR6943113). El subrayado indica los contigs con tamaño de >100 aminoácidos. Los contigs resaltados en amarillo indican estructura secundaria con forma de varilla.**

Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
<u>NODE_10732_length_1074_cov_42.591409</u>	hhrbz_dv4	709	657	-	no	2.4e-05
	hhrbz_dv4	108	159	+	no	5.1e-05

<u>NODE_17963_length_887_cov_7.835381</u>	hhrbz_dv4	298	241	-	no	1e-11
	hhrbz_dv4	611	669	+	no	2.7e-10
<u>NODE_19406_length_861_cov_369.751269</u>	hhrbz_dv4	274	216	-	no	8.1e-11
	hhrbz_dv4	509	566	+	no	1e-09
<u>NODE_19819_length_855_cov_4.262148</u>	hhrbz_dv4	125	204	+	no	1.6e-14
	hhrbz_dv4	346	266	-	no	6.5e-13
<u>NODE_22304_length_818_cov_3.903356</u>	hhrbz_dv4	384	438	+	no	1e-12
	hhrbz_dv4	93	36	-	no	2.4e-08

Por otro lado, también se encontraron 4 contigs con características de virus Épsilon en un conjunto de datos procedente de un experimento metatranscriptómico en sedimentos de agua dulce (SRR6973926) (Tabla 7). Todos ellos codifican ORFs con un tamaño de >100 aminoácidos y tienen una estructura con forma de varilla (ANEXO II). De los 4 contigs, 3 de ellos tienen una longitud de secuencia más corta de lo esperado, similar a lo visto en el caso anterior.

**Tabla 7. Contigs con características de virus Épsilon encontrados en el conjunto de datos de sedimentos de agua dulce SRR6973926.** El subrayado indica los contigs con tamaño de >100 aminoácidos. El resaltado en amarillo indica estructura secundaria con forma de varilla.

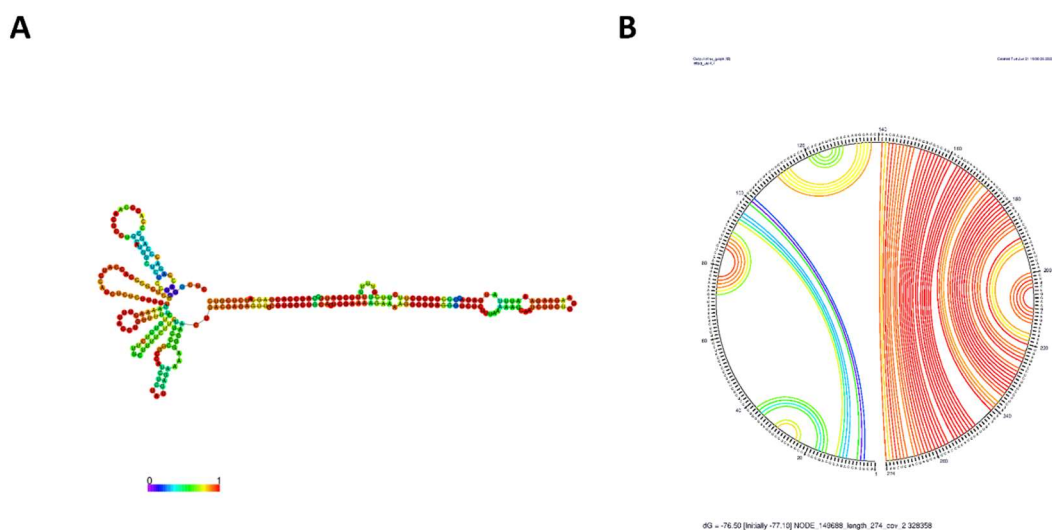
Contig	Hit ribozima	Desde	Hacia	Cadena	Truncado	E-value
	hhrbz_dv4	600	546	-	no	1.3e-12
<u>NODE_14755_length_939_cov_2.953811/1-887</u>	hhrbz_dv4	1	30	+	5'	3.1e-05
	hhrbz_dv4	862	887	+	3'	0.7
<u>NODE_19461_length_848_cov_3.348387/1-775</u>	hhrbz_dv4	206	263	+	no	7.1e-13
	hhrbz_dv4	695	640	-	no	9.4e-13
<u>NODE_20553_length_830_cov_4.393659/1-757</u>	hhrbz_dv4	682	737	+	no	3.9e-12
	hhrbz_dv4	441	382	-	no	7.1e-11
<u>NODE_2384_length_1801_cov_13.152199/1-1728</u>	hhrbz_dv4	1205	1304	+	no	1.6e-12
	hhrbz_dv4	189	96	-	no	3e-05

En el conjunto de datos SRR17133691, procedente de un experimento con el molusco *Pinctada fucata*, se ha encontrado un hit duplicado para la ribozima Hammerhead tipo III para ambos sentidos de la cadena (tabla 8), aunque la ribozima de la cadena negativa está truncada y tiene un E-value muy

elevado en comparación al del hit del otro sentido de la cadena, además de que no se ha encontrado circularidad en su secuencia. La estructura secundaria predicha no tiene forma de varilla (Figura 12 y ANEXO II), aunque esto sería debido a que la secuencia esté incompleta. Probablemente se trate de un virus Épsilon en el que las lecturas que hay en el conjunto de datos son demasiado cortas y tienen muy poca cobertura como para poder captar la secuencia entera del virus. El hecho de que el contig estuviera incompleto explicaría su pequeño tamaño (272 nucleótidos), y también explicaría que haya un hit truncado en la cadena negativa y que por tanto el E-value sea alto.

**Tabla 8. Contig posiblemente perteneciente a un virus Épsilon en SRR17133691.**

Contig	Hit ribozima	Desde	Hacia	Cadena	Truncado	E-value
NODE_149688_length_274_cov_2.328358	hhrbz_dv4	105	171	+	no	7.7e-08
	hhrbz_dv4	274	244	-	5'	4.6



**Figura 12. Estructura secundaria predicha del contig NODE\_149688\_length\_274\_cov\_2.328358, perteneciente a un posible virus Épsilon en SRR17133691. (A) Muestra la estructura secundaria MFE (energía mínima libre) mientras que (B) representa la estructura secundaria del mismo contig pero en formato de gráfico circular. Los colores de las estructuras indican las probabilidades de emparejamiento de bases.**

En un conjunto de datos de un RNA-Seq de granos fermentados (SRR13675528) se ha detectado un contig con características de virus Épsilon, el contig tiene tamaño de 2699 nucleótidos, pero después de circularizar manualmente su secuencia el tamaño se ha reducido drásticamente hasta tener un tamaño de 1842 nt, tamaño que entra dentro de lo normal en un virus Épsilon (Tabla 9). Tiene estructura con forma de varilla (ANEXO II) y codifica para ORFs de tamaño considerablemente largo.

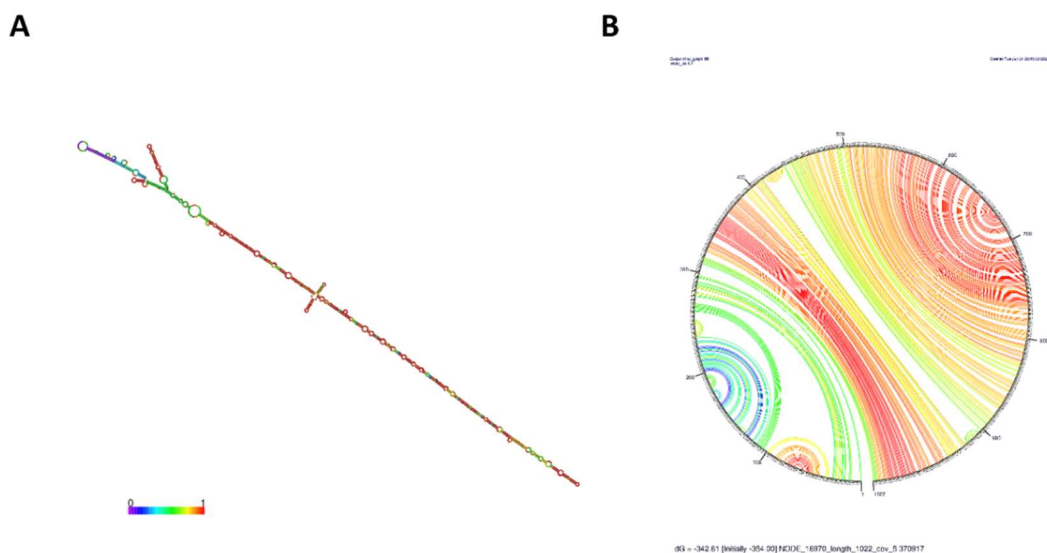
**Tabla 9. Contig posiblemente perteneciente a un virus Épsilon encontrados en SRR13675528. El resaltado amarillo y el subrayado indica estructura en forma de varilla y ORFs codificados de tamaño de >100 aminoácidos respectivamente.**

Contig	Hit ribozima	Desde	Hacia	Cadena	Truncado	E-value
<b>NODE_4922_length_2699_cov_14.473343_1842_nt</b>	hhrbz_dv4	1828	1903	+	no	3.7e-11
	hhrbz_dv4	276	207	-	no	7.1e-10

En SRR17497127, un conjunto de datos de una muestra de heces del escorpión *Hystrix brachyura* también se ha hallado un contig particular, en el cual no se ha encontrado circularidad y solo tiene solo un hit para la ribozima Hammerhead tipo III (en lugar de dos), pero dicho hit posee un E-value muy bajo y una longitud típica de virus Épsilon (tabla 10). La estructura predicha tiene una forma de varilla (Figura 13 y ANEXO II) y se ha visto que codifica para ORFs de tamaño relativamente grande. Muy seguramente se trate de un virus Épsilon de hongos que se encontraban en las heces del animal. Este genoma estaría incompleto (baja cobertura) y le faltaría la otra ribozima Hammerhead. Se intentó obtener la secuencia completa mediante el programa Infernal para buscar esta ribozima, usando directamente un archivo fasta de las lecturas del conjunto de datos en lugar de emplear un archivo con los contigs ensamblados, para de esta forma intentar hacer un ensamblaje manual, pero no se pudo obtener este segundo motivo.

**Tabla 10. Contig de un posible virus Épsilon encontrados por Infernal en SRR17497127. Solo se ha encontrado un hit para la ribozima Hammerhead en este caso.**

Contig	Hit ribozima	Desde	Hacia	Cadena	Truncado	E-value
NODE_16970_length_1022_cov_5.370917	hhrbz_dv4	285	229	-	no	8e-08



**Figura 13. Estructura secundaria predicha del contig NODE\_16970\_length\_1022\_cov\_5.370917, encontrado en SRR17497127 y que podría pertenecer a un virus Épsilon. (A) Muestra la estructura secundaria MFE (energía mínima libre) mientras que (B) representa la estructura secundaria del mismo contig pero en formato de gráfico circular. Los colores de las estructuras indican las probabilidades de emparejamiento de bases.**

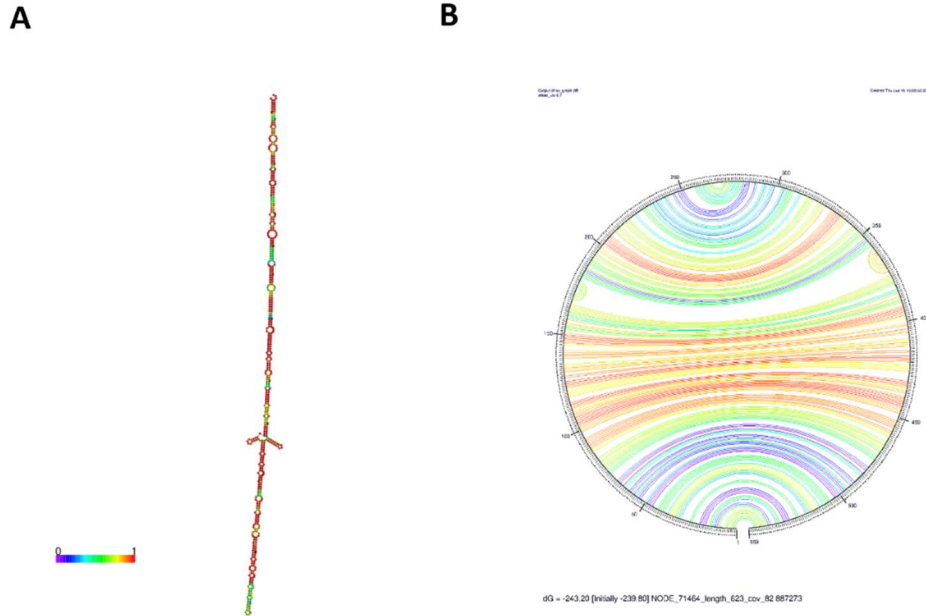
### 4.3. Virus Zeta

En el conjunto de datos SRR6973851, procedente también del ya mencionado experimento metagenómico en sedimentos de agua dulce, se encontraron 8 contigs con unas características particulares del Virus Zeta: dos ribozimas HHR codificados en ambos sentidos de la cadena y una longitud aproximada de entre 300 y 700 nucleótidos (tabla 11). Se caracterizan también por tener una estructura secundaria de varilla, lo cual se da para todos los contigs candidatos a posibles virus Zeta encontrados en este conjunto de datos (figura 14 y ANEXO III). Los virus Zeta tienen la particularidad de codificar habitualmente una ORF larga en cada sentido de la cadena, lo cual se ha verificado que ocurre en estos contigs.

**Tabla 11. Contigs con características de virus Zeta encontrados por Infernal en el conjunto de datos de sedimentos de agua dulce SRR6973851. El resaltado amarillo indica que el contig posee una estructura en forma de varilla mientras que el subrayado se pone para aquellos contigs con al menos dos ORFs predichas, una en cada sentido de la cadena.**

Contig	Ribozima name	Desde	Hasta	Cadena	Truncado	E-value
<u>NODE_124506_length_502_cov_2.344988</u>	hhrbz_dv4	323	272	-	no	1.8e-06
	hhrbz_dv4	124	175	+	no	4.1e-06
<u>NODE_42734_length_763_cov_3.975362</u>	hhrbz_dv4	648	591	-	no	2.6e-11
	hhrbz_dv4	250	305	+	no	3.8e-11
<u>NODE_63026_length_655_cov_6.341924</u>	hhrbz_dv4	579	526	-	no	1.8e-06
	hhrbz_dv4	217	270	+	no	2.9e-06
<u>NODE_71464_length_623_cov_82.887273</u>	hhrbz_dv4	388	319	-	no	3.1e-17
	hhrbz_dv4	153	217	+	no	1.4e-12
<u>NODE_73555_length_616_cov_10.287293</u>	hhrbz_dv4	293	371	+	no	1.4e-13
	hhrbz_dv4	136	59	-	no	6.9e-13
<u>NODE_79273_length_598_cov_4.539048</u>	hhrbz_dv4	436	511	+	no	2.8e-15
	hhrbz_dv4	279	205	-	no	2.7e-13
<u>NODE_86913_length_577_cov_8.492063</u>	hhrbz_dv4	136	85	-	no	1.2e-05
	hhrbz_dv4	423	474	+	no	0.00015
<u>NODE_90441_length_568_cov_9.684848</u>	hhrbz_dv4	11	62	+	no	3.1e-06
	hhrbz_dv4	213	162	-	no	7e-06

También hemos encontrado 3 contigs de este tipo en un conjunto de datos procedente de un experimento de metatranscriptómica realizado en un biorreactor anaeróbico (SRR6962709) (tabla 12) y otros 24 en el ya mencionado SRR6943202 (tabla 13). En los datos de la muestra del biorreactor todos los contigs presentaban las características de los virus Zeta, con la excepción de NODE\_96369\_length\_541\_cov\_2.215812\_1 el cual solo se ha predicho una ORF larga codificada en la cadena positiva del genoma. En cuanto al conjunto de datos SRR6943202, todos los 24 contigs presentaban estructura de varilla (ANEXO III), pero solo 11 codificaban una ORF larga en ambos sentidos de la secuencia.



**Figura 14. Ejemplo de estructura secundaria predicha de uno de los contigs pertenecientes a posibles virus Zeta del conjunto de datos SRR6973851. (A)** Muestra la estructura secundaria MFE (energía mínima libre) del contig NODE\_71464\_length\_623\_cov\_82.887273 mientras que **(B)** representa la estructura secundaria del mismo contig pero en formato de gráfico circular. Los colores de las estructuras indican las probabilidades de emparejamiento de bases.

**Tabla 12. Contigs con características de virus Zeta encontrados en el conjunto de datos procedente de un biorreactor anaeróbico (SRR6962709).** El resaltado amarillo indica que el contig posee una estructura en forma de varilla mientras que el subrayado se pone para aquellos contigs con al menos dos ORFs predichas, una para cada sentido de la cadena.

Contig	Ribozima name	Desde	Hasta	Cadena	Truncado	E-value inc
<u>NODE_65432_length_673_cov_9.173</u> 333	hhrbz_dv4	393	337	-	no	1.9e-12
	hhrbz_dv4	1	55	+	5'	1e-06
<u>NODE_67120_length_664_cov_1.774</u> 958	hhrbz_dv4	24	115	+	no	5.2e-13
	hhrbz_dv4	386	297	-	no	3.2e-11
<u>NODE_96369_length_541_cov_2.215</u> 812	hhrbz_dv4	475	420	-	no	1.4e-15
	hhrbz_dv4	263	316	+	no	1.7e-09

**Tabla 13. Contigs con características de virus Zeta encontrados en una muestra de humedales (SRR6943202).** El resaltado amarillo indica que el contig posee una estructura en forma de varilla mientras que el subrayado se pone para aquellos contigs con al menos dos ORFs predichas, una para cada sentido de la cadena.

Contig	Ribozima name	Desde	Hasta	Cadena	Truncado	E-value
<u>NODE_102143_length_637_cov_2.443262</u>	hhrbz_dv4	10	88	+	no	7.6e-13
	hhrbz_dv4	356	277	-	no	5.7e-09



NODE_104248_length_631_cov_20.985663	hhrbz_dv4	541	455	-	no	1.6e-13
	hhrbz_dv4	201	283	+	no	5.3e-09
NODE_104289_length_631_cov_4.933692	hhrbz_dv4	49	135	+	no	1.4e-13
	hhrbz_dv4	389	307	-	no	5.3e-12
NODE_105466_length_628_cov_4.679279	hhrbz_dv4	409	314	-	no	1.5e-10
	hhrbz_dv4	15	111	+	no	7.3e-09
NODE_106584_length_625_cov_28.670290	hhrbz_dv4	46	138	+	no	2.5e-12
	hhrbz_dv4	387	296	-	no	3.8e-08
NODE_110258_length_616_cov_6.598527	hhrbz_dv4	212	300	+	no	4.7e-10
	hhrbz_dv4	543	503	-	5'	2.2e-06
NODE_112899_length_610_cov_4.029795	hhrbz_dv4	218	126	-	no	2.9e-09
	hhrbz_dv4	395	487	+	no	8.3e-08
NODE_113263_length_609_cov_6.585821	hhrbz_dv4	417	330	-	no	8.6e-11
	hhrbz_dv4	47	133	+	no	2.1e-10
NODE_116781_length_601_cov_9.829545	hhrbz_dv4	270	188	-	no	1.5e-11
	hhrbz_dv4	452	528	+	3'	5.3e-05
NODE_120123_length_594_cov_3.842610	hhrbz_dv4	281	197	-	no	5,00E-13
	hhrbz_dv4	487	552	+	3'	0.006
NODE_120981_length_592_cov_17.516378	hhrbz_dv4	459	405	-	no	4.2e-13
	hhrbz_dv4	137	192	+	no	4.3e-12
NODE_121038_length_592_cov_5.169557	hhrbz_dv4	240	161	-	no	3.5e-12
	hhrbz_dv4	399	476	+	no	1.8e-11
NODE_121464_length_591_cov_447.318533	hhrbz_dv4	232	315	+	no	2.7e-10
	hhrbz_dv4	90	3	-	no	1.6e-08
NODE_123910_length_586_cov_25.407407	hhrbz_dv4	179	257	+	no	3.9e-14
	hhrbz_dv4	513	455	-	5'	6.1e-05
NODE_131855_length_571_cov_36.146586	hhrbz_dv4	171	245	+	no	6.8e-11
	hhrbz_dv4	498	440	-	5'	0.00012
NODE_133521_length_568_cov_61.167677	hhrbz_dv4	238	327	+	no	2.3e-10
	hhrbz_dv4	93	5	-	no	3.2e-10
NODE_133728_length_568_cov_2.612121	hhrbz_dv4	369	314	-	no	1.7e-12
	hhrbz_dv4	114	171	+	no	1.8e-10
NODE_144063_length_550_cov_23.601677	hhrbz_dv4	167	255	+	no	1,00E-10
	hhrbz_dv4	477	421	-	5'	0.0027

NODE_157751_length_529_cov_21.563596	hhrbz_dv4	282	228	-	no	8.3e-13
	hhrbz_dv4	40	92	+	no	4.7e-10
NODE_157766_length_529_cov_10.728070	hhrbz_dv4	254	201	-	no	1.8e-11
	hhrbz_dv4	428	456	+	3'	0.025
	hhrbz_dv4	1	28	+	5'	0.33
NODE_157924_length_529_cov_3.087719	hhrbz_dv4	186	266	+	no	4.5e-14
	hhrbz_dv4	456	410	-	5'	1.1e-05
NODE_196087_length_482_cov_90.403423	hhrbz_dv4	85	137	+	no	2.5e-05
	hhrbz_dv4	302	251	-	no	4.1e-05
NODE_198943_length_479_cov_58.192118	hhrbz_dv4	187	269	+	no	1.4e-09
	hhrbz_dv4	118	34	-	no	8.7e-08
NODE_270739_length_421_cov_6178.17816	hhrbz_dv4	344	290	-	no	2.5e-13
	hhrbz_dv4	204	255	+	no	6.7e-06
NODE_301491_length_403_cov_12.151515	hhrbz_dv4	170	224	+	no	1.3e-13
	hhrbz_dv4	312	257	-	no	1.1e-10

En el conjunto de datos SRR13675528 obtenido de granos fermentados también se han encontrado tres contigs con características de virus Zeta (Tabla 14), pero al estudiar su estructura dos de ellos no cumplen con la estructura de varilla típica de los virus Zeta (ANEXO III) y en ninguno se ha detectado las típicas ORFs largas.

**Tabla 14. Contigs encontrados en SRR13675528.** El contig NODE\_15680\_length\_1022\_cov\_20.053741 al circularizar su secuencia la longitud de esta baja de 1022 a 662 por lo que se cataloga como virus Zeta. El resaltado amarillo indica estructura con forma de varilla.

Contig	Ribozima name	Desde	Hasta	Cadena	Truncado	E-value
NODE_27971_length_611_cov_19.237918	hhrbz_dv4	315	263	-	no	3,00E-05
	hhrbz_dv4	81	132	+	no	0.00014
NODE_41491_length_449_cov_279.364362	hhrbz_dv4	206	154	-	no	2.9e-05
	hhrbz_dv4	205	256	+	no	0.00078
NODE_15680_length_1022_cov_20.053741 662nt	hhrbz_dv4	376	304	-	no	8.9e-12
	hhrbz_dv4	2	71	+	no	2.3e-11

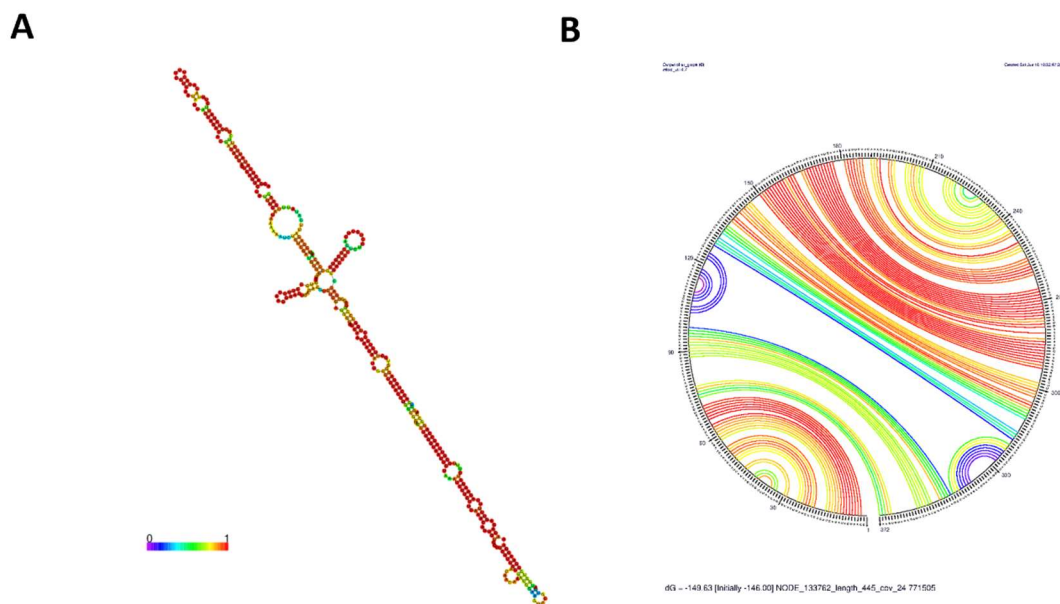
#### 4.4. Virus Eta

En el conjunto de datos del metagenoma de humedales SRR6943136 se encontró un contig que posiblemente corresponda un tipo de virus al que hemos denominado virus Eta (tabla 15). Este tipo de virus están caracterizados por sus estructuras de RNA circular de unos 300-700 nucleótidos con 2

ribozimas DVR y 2 ORFs largas codificadas en ambos sentidos de la cadena, además de la conformación con forma de varilla. En el contig hallado en SRR6943136 se ha confirmado que su estructura secundaria tiene forma de casi varilla (Figura 15), pero solo codifica una ORF larga en el sentido positivo de la cadena.

**Tabla 15. Contig procedente de un posible virus Eta detectado por Infernal en el conjunto de datos de muestra de humedales (SRR6943136). El resaltado amarillo indica estructura con forma de varilla.**

Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
NODE_133762_length_445_cov_24.771505	DVR4	132	200	+	no	0.0011
	DVR4	310	255	-	no	0.0055



**Figura 15. Estructura secundaria predicha del contig NODE\_133762\_length\_445\_cov\_24.771505 procedente del conjunto de datos SRR6943136, el cual pertenece a un posible virus Eta. (A) Muestra la estructura secundaria MFE (energía mínima libre) del contig mientras que (B) representa la estructura secundaria del mismo contig pero en formato de gráfico circular. Los colores de las estructuras indican las probabilidades de emparejamiento de bases.**

#### 4.5. Nuevas entidades de RNA circular con 2 ribozimas

En nuestra búsqueda por los conjuntos de datos se han encontrado nuevos elementos de RNA circular con ribozimas Hairpin, Twister y combinaciones de dos ribozimas distintas, los cuales no se habían caracterizado hasta en la fecha. En este trabajo se ha decidido nombrarlos usando la abreviatura de las dos ribozimas que contienen separados de un guión.

##### 4.5.1. Agentes HPR-HHR

Se han encontrado 19 contigs con ribozimas Hairpin y Hammerhead distribuidos en los conjuntos de datos SRR6943202, SRR6943136, SRR6943113, SRR6973926 y SRR6973851 (tabla 16 y ANEXO IV).

**Tabla 16. Contigs de tipo HPR-HHR detectados por Infernal en SRR6943202, SRR6943136, SRR6943113, SRR6973926 y SRR6973851. En amarillo están resaltados los contigs cuya estructura predicha por RNA fold es de varilla. Los que están subrayados son aquellos con ORFs predichas por getorf de tamaño >100 aa.**

SRR6943202						
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
<u>NODE 10790 length 1619 cov 38.0</u> <u>86028</u>	hairpin-meta	363	436	+	no	5,00E-09
	hhrbz_dv4	1015	964	-	no	2.6e-06
NODE_1293_length_3712_cov_12.06 7876	hairpin-meta	306	383	+	no	4.5e-09
	hhrbz_dv4	649	597	-	no	2.3e-07
NODE_2829_length_2753_cov_16.18 0970	hairpin-meta	612	683	+	no	1.9e-10
	hhrbz_dv4	1002	951	-	no	8.4e-08
<u>NODE 31807 length 1046 cov 7.43</u> <u>3710</u>	hairpin-meta	98	168	+	no	3.7e-10
	hhrbz_dv4	380	329	-	no	1.1e-06
<u>NODE 4605 length 2280 cov 11.58</u> <u>7676</u>	hairpin-meta	105	176	+	no	1.2e-10
	hhrbz_dv4	1132	1081	-	no	0.00026
NODE_78447_length_713_cov_13.68 9063	hairpin-meta	308	86	-	no	2,00E-08
	hhrbz_dv4	30	68	+	no	4.9e-05
<u>NODE 81354 length 702 cov 4.774</u> <u>245</u>	hairpin-meta	177	240	+	no	6.5e-08
	hhrbz_dv4	551	448	-	no	2.1e-06
SRR6943136						
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
<u>NODE 12835 length 1047 cov 5.472279</u>	hairpin-meta	355	422	+	no	2e-09
	hhrbz_dv4	636	585	-	no	7.6e-07
<u>NODE 12856 length 1046 cov 10.440904</u>	hairpin-meta	552	482	-	no	9.6e-10
	hhrbz_dv4	270	321	+	no	7.6e-07

<u>NODE_2379_length_1902_cov_10.498633</u>	hairpin-meta	1285	1213	-	no	1e-10
	hhrbz_dv4	934	986	+	no	0.0002
<u>NODE_3068_length_1742_cov_14.499700</u>	hairpin-meta	160	278	+	no	9.3e-10
	hhrbz_dv4	695	646	-	no	2.4e-05
<u>NODE_36345_length_712_cov_5.823161</u>	hairpin-meta	308	530	+	no	6e-09
	hhrbz_dv4	586	548	-	no	3.4e-05
<u>NODE_7299_length_1281_cov_22.654801</u>	hairpin-meta	464	392	-	no	2.5e-10
	hhrbz_dv4	133	184	+	no	7.6e-07
<u>NODE_7530_length_1266_cov_10.682313</u>	hairpin-meta	214	290	+	no	2e-10
	hhrbz_dv4	489	437	-	no	1.5e-06
<u>NODE_9903_length_1146_cov_7.915191</u>	hairpin-meta	622	545	-	no	2.5e-09
	hhrbz_dv4	276	328	+	no	2.5e-07
<b>SRR6943113</b>						
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E- valu e
<u>NODE_32295_length_705_cov_5.419304</u>	hairpin-meta	528	465	-	no	2e-08
	hhrbz_dv4	153	256	+	no	5e-05

SRR6973851						
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
NODE_17245_length_1075_cov_11.27245 5	hairpin-meta	485	395	-	no	4.6e-08
	hhrbz_dv4	190	241	+	no	3.5e-06
SRR6973926						
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
NODE_9244_length_1114_cov_7.308357	hairpin-meta	795	720	-	no	2.4e-08
	hhrbz_dv4	497	547	+	no	1.2e-06

#### 4.4.2. Agentes HPR – DVR

En SRR694313 se ha hallado un contig único caracterizado por tener dos hits para ribozimas, uno para Hairpin y el otro para la ribozima virus Delta de la hepatitis humana. Tiene estructura de varilla, pero no codifica un ORF de >100 aminoácidos (tabla 17 y ANEXO IV).

**Tabla 17. Contig con hits HPR-DVR encontrado en SRR694313. El resaltado amarillo indica estructura con forma de varilla.**

Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
NODE_14643_length_999_cov_6.421166	DVR4	672	729	+	no	0.0033
	hairpin-meta	203	116	-	no	3.8e-09

#### 4.4.3. Agentes HHR – DVR

Unos 4 contigs procedentes a elementos de RNA circular con ribozimas Hammerhead y de tipo virus Delta de la hepatitis humana fueron detectados en dos conjuntos de datos, SRR6943136 y SRR6973926 (Tabla 18).

**Tabla 18. Contigs de tipo HHR-DVR detectados en SRR6943136 y SRR6973926. En amarillo están resaltados los contigs cuya estructura predicha por RNA fold es de varilla. Los que están subrayados son aquellos con ORFs predichas que poseen un tamaño de tamaño >100 aa.**

SRR6943136						
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
NODE_13047_length_1040_cov_153.014478	DVR4	216	293	+	no	0.38
	hhrbz_dv4	445	391	-	no	9.1e-07
NODE_31937_length_746_cov_25.497771	DVR4	578	522	-	no	0.00018
	hhrbz_dv4	236	287	+	no	2.8e-07
SRR6973926						
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
NODE_10017_length_1082_cov_92.586720	DVR4	360	308	-	no	0.01
	hhrbz_dv4	796	848	+	no	4e-07
NODE_15011_length_933_cov_268.070930	DVR4	168	112	-	no	0.085
	hhrbz_dv4	588	639	+	no	6e-07

#### 4.4.4. Agentes HPR – HPR

Se han hallado 20 contigs con un hit duplicado para la ribozima Hairpin en los conjuntos de datos SRR6973851, SRR6943202, SRR6943136, SRR6943113 y SRR6973926 (Tabla 19 y ANEXO IV).

**Tabla 19. Listado de contigs de tipo HPR-HPR y los conjuntos de datos respectivos en los que fueron encontrados. El resaltado amarillo indica estructura con forma de varilla, el subrayado indica ORFs codificados de tamaño de >100 aminoácidos.**

SRR6973851						
Contig	Ribozima name	Desde	Hasta	Cadena	Truncado	E-value
NODE_14079_length_1158_cov_61.329954	hairpin-meta	189	346	+	no	1.7e-09
	hairpin-meta	676	597	-	no	0.057
NODE_15982_length_1106_cov_181.427880	hairpin-meta	806	690	-	no	3.6e-09
	hairpin-meta	209	276	+	no	0.0034
NODE_9976_length_1311_cov_19.197092	hairpin-meta	860	1006	+	no	9e-09
	hairpin-meta	435	367	-	no	0.19

SRR6943202						
Contig	Ribozima name	Desde	Hasta	Cadena	Truncado	E-value
NODE_12704_length_1519_cov_1 187.529737	hairpin-meta	1116	1007	-	no	6.7e-07
	hairpin-meta	658	821	+	no	0.0004
NODE_12958_length_1507_cov_1 89.006276	hairpin-meta	691	582	-	no	5.9e-07
	hairpin-meta	230	393	+	no	0.0032
NODE_14101_length_1460_cov_8. 706561	hairpin-meta	910	738	-	no	0.00023
	hairpin-meta	344	498	+	no	0.0052
NODE_16961_length_1354_cov_3 7.309133	hairpin-meta	1239	1084	-	no	9.9e-09
	hairpin-meta	476	572	+	no	0.0041
NODE_22183_length_1214_cov_1 44.185802	hairpin-meta	1026	921	-	no	6.8e-10
	hairpin-meta	450	556	+	no	2.1e-09
NODE_23178_length_1192_cov_4 8.541555	hairpin-meta	57	179	+	no	1.3e-08
	hairpin-meta	481	420	-	no	0.051
NODE_42293_length_929_cov_94. 482477	hairpin-meta	658	755	+	no	2.5e-09
	hairpin-meta	317	205	-	no	3,00E-09
SRR6943136						
Contig	Ribozima name	Desde	Hasta	Cadena	Truncado	E-value
NODE_2993_length_1757_cov_75.457838	hairpin-meta	228	47	-	no	3.3e-08
	hairpin-meta	1366	1464	+	no	0.025
NODE_4246_length_1548_cov_69.526102	hairpin-meta	1060	1205	+	no	1e-08
	hairpin-meta	195	140	-	no	0.0052
NODE_4489_length_1519_cov_218.41908 7	hairpin-meta	874	983	+	no	4.7e-07
	hairpin-meta	1332	1169	-	no	0.00028
NODE_8971_length_1190_cov_8.210385	hairpin-meta	676	554	-	no	9.3e-09
	hairpin-meta-2	233	314	+	no	0.0073



SRR6943113						
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
NODE_12279_length_1022_cov_156.232877	hairpin-meta	350	263	-	no	1.4e-06
	hairpin-meta	62	256	+	no	0.0038
SRR6973926						
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
NODE_4368_length_1452_cov_28.347353	hairpin-meta	754	695	-	no	3.6e-05
	hairpin-meta	1	62	+	5'	0.00015
NODE_5421_length_1341_cov_360.090694	hairpin-meta	199	448	+	no	8e-07
	hairpin-meta	750	681	-	no	7.3e-05
NODE_5825_length_1311_cov_133.840065	hairpin-meta	731	877	+	no	1.1e-09
	hairpin-meta	306	238	-	no	0.022
NODE_6448_length_1262_cov_446.709840	hairpin-meta	60	189	+	no	2.3e-09
	hairpin-meta	487	399	-	no	0.0022
NODE_9471_length_1105_cov_1057.118217	hairpin-meta	532	648	+	no	9.3e-10
	hairpin-meta	96	29	-	no	0.00041

#### 4.4.5. TwR-TwR

Se hallaron en total 4 contigs con un hit duplicado para la ribozima Twister, 3 de ellos en SRR6943136 y 1 en SRR6943113 (Tabla 20 y ANEXO IV).

**Tabla 20. Lista de contigs tipo TwR-TwR en SRR6943136 y SRR6943113. En amarillo están resaltados los contigs cuya estructura predicha por RNA fold es de varilla. Los que están subrayados son aquellos con ORFs predichas que poseen un tamaño de tamaño >100 aminoácidos.**

SRR6943136						
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
NODE_19577_length_897_cov_153.224515	twister-P1	358	421	+	no	3.4e-14
	twister-P1	811	730	-	no	2.2e-08
NODE_40922_length_681_cov_6.013158	twister-P1	504	583	+	no	1.4e-08
	twister-P1	113	51	-	no	6.1e-08
<u>NODE_9634_length_1159_cov_3.488029</u>	twister-P1	820	889	+	no	2.4e-10
	twister-P1	241	169	-	no	8.8e-06
SRR6943113						
Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
<u>NODE_2754_length_1773_cov_8.210000</u>	twister-P1	877	956	+	no	2.6e-11
	twister-P3	735	682	-	no	0.0011

#### 4.4.6. TwR-HHR y TwR-HPR

En SRR6943113 se ha hallado un contig que posee una combinación de ribozima Twister con Hammerhead y otro similar que combina Twister pero con Hairpin en lugar Hammerhead. Ninguno de estos dos contigs han presentado una estructura con forma de varilla ni un ORFs con un tamaño relativamente grande (tabla 21 y ANEXO IV).

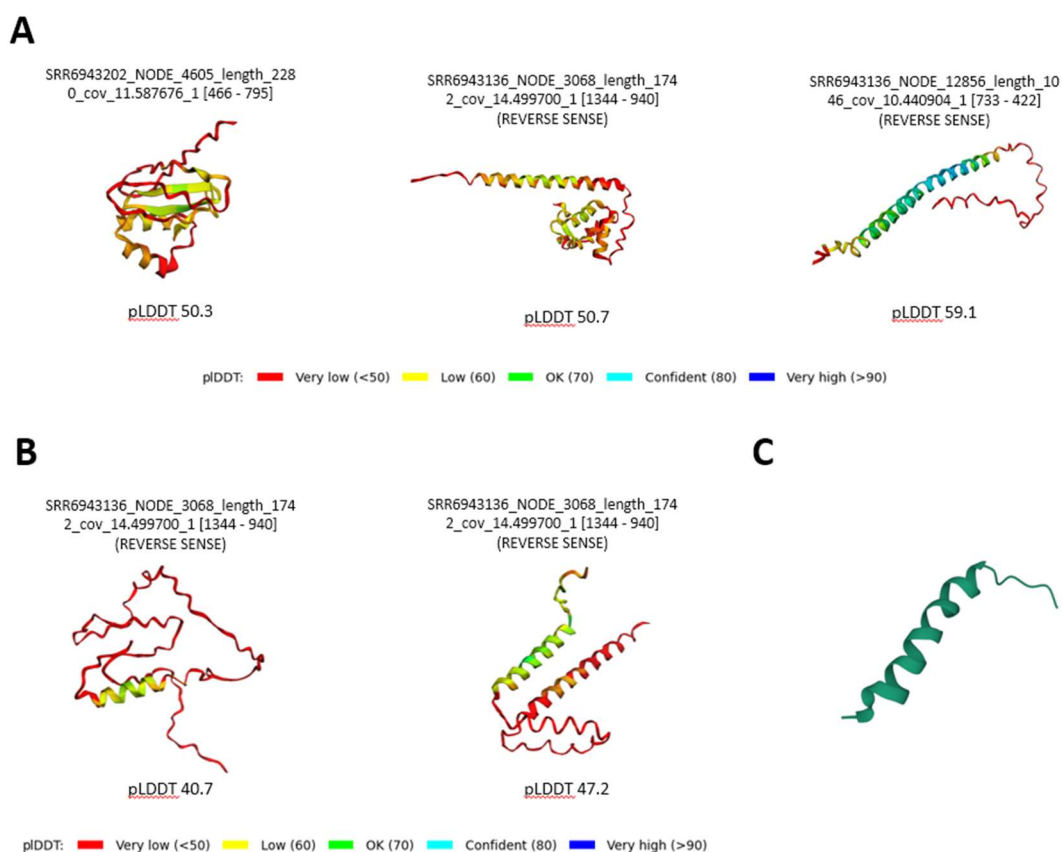
**Tabla 21. Contig TwR-HHR y TwR-HPR en el conjunto de datos SRR6943113.**

Contig	Ribozima	Desde	Hasta	Cadena	Truncado	E-value
NODE_18312_length_881_cov_6.334158	hhrbz_dv4	310	363	+	no	0.00012
	twister-P1	491	418	-	no	3.9e-12
NODE_45131_length_615_cov_134.254613	hairpin-meta	85	233	+	no	1.5e-05
	twister-P1	396	339	-	no	0.0017

#### 4.5. Estudio de la estructura tridimensional de las proteínas codificadas por agentes circulares mediante AlphaFold2

Se ha procedido a estudiar la estructura tridimensional de las proteínas codificadas de estos nuevos elementos de RNA circular mediante AlphaFold2. Para ello se han seleccionado aquellos contigs que cumpliesen el requisito de codificar al menos una ORF de un tamaño mayor a 100 aminoácidos. Se ha examinado la estructura proteica de un total 20 contigs, de los cuales 1 es de tipo HHR-DVR (ANEXO V), 13 son de tipo HPR-HPR (ANEXO VI) y 6 de tipo HPR-HHR (ANEXO VII).

AlphaFold produce una estimación por residuo de su confianza en una escala de 0 a 100. Esta medida de confianza se llama pLDDT. Solo 3 contigs de todos los analizados han dado una estructura tridimensional con un pLDDT global mayor a 50, que es cuando se puede considerar una predicción fiable (figura 16A). No obstante, se han encontrado otros dos contigs con estructuras que, pese a no llegar al pLDDT global de 50, poseen regiones con muy buena fiabilidad (pLDDT >70) (figura 16B). Además, llamó la atención la estructura de la proteína codificado por NODE\_12856\_length\_1046\_cov\_10.440904 (SRR6943136), el cual parece ser muy similar a la de la región N-terminal del antígeno Delta (figura 16C).



**Figura 16. Estructuras tridimensionales destacadas predichas por AlphaFold2 de los nuevos elementos de RNA circular.** (A) Muestra aquellas estructuras con un pLDDT global mayor a 50. (B) muestra estructuras que contienen regiones específicas con un pLDDT alto (>70). En el encabezado de las estructuras está el nombre de la proteína, en el cual se indica el conjunto de datos de donde procede, el nombre del contig, la posición dentro del contig en el que se encuentra codificada entre corchetes, y en caso de que esté codificada en la cadena reversa se indica entre paréntesis. En el pie de la estructura se indica el valor del pLDDT. Los distintos colores en los que están coloreadas las regiones de las estructuras predichas indica el valor del pLDDT de dichas regiones. En (C) se muestra la estructura tridimensional obtenida de Protein Data Bank de la región N-terminal de repeticiones de leucina del antígeno delta la hepatitis humana, el cual parece ser similar a la estructura de la proteína del contig NODE\_12856\_length\_1046\_cov\_10.440904.

## 5. Discusión de resultados y conclusión

En nuestra búsqueda bioinformática exhaustiva de los 26 conjuntos de datos del SRA hemos logrado identificar en total 126 elementos de RNA circular, de los cuales 12 pertenecerían virus tipo Delta, 25 a virus tipo Épsilon, 38 a virus Zeta, 1 a un virus Eta y 50 serían elementos nuevos que no aún no se han caracterizado. Todos estos elementos de RNA circular se hallaron distribuidos en 12 conjuntos de datos, mientras que en los 14 restantes no se descubrió nada relevante para este trabajo.

Todos los virus tipo Delta encontrados tenían estructura con forma de varilla y la mayoría de ellos contaban con ORFs codificadas de un tamaño relevante, lo cual era esperado en este tipo de virus. Los virus encontrados que no codifican estos ORFs podrían pertenecer nuevo tipo de virus tipo Delta no caracterizados llamados Virus tipo Delta no canónicos.

En cuanto a los virus Épsilon, a diferencia de los Delta, la estructura de varilla y el tamaño de los ORFs no parecen características obligatorias en este tipo de virus ya que bastantes contigs pertenecientes a estos virus que hemos analizado no poseían estas dos características simultáneamente. Dentro de este tipo abundan los contigs que poseen una de las dos características mientras que solo unos pocos que no poseían ninguna de las dos. Los virus Épsilon encontrados comparte un tamaño similar con los vistos con los Delta, aunque se han hallado contigs pertenecientes a virus Épsilon encontrados en el conjunto de datos SRR6943113 que tenían una longitud de alrededor de 800 nucleótidos, lo cual está en el límite para ser considerados virus Épsilon, por lo que podrían tratarse de una subrama de virus Epsilon o incluso Zeta.

Otro aspecto que tienen en común Épsilon y Delta es su reservorio. Como hemos visto, el virus Delta lo hemos detectado en *Sceloporus jarrovi*, *Marmota himalayana* y en dos conjuntos de datos procedente de un humedal en Minnesota, USA (SRR6943202 y SRR6943136). El virus Épsilon, por otra parte, se ha hallado también en conjuntos de datos del mismo humedal en Minnesota (SRR6943136 y SRR6943113) y también en conjuntos de datos procedentes una muestra del Río Colorado (USA) (SRR6973926) y de granos de cereales fermentados por hongos (SRR13675528), además de en los organismos *Pinctada fucata* (SRR17133691) y *Hystrix brachyura* (SRR17497127). Por tanto, el reservorio de estos dos virus se encuentra tanto en organismos superiores como en nichos medioambientales.

Los virus Zeta que hemos hallado se caracterizan, al igual que los virus Delta, por tener toda estructura de varilla, con excepción de los virus Zeta hallados en SRR13675528 los cuales no lo tenían. También tienden a codificar para ORFs, la gran mayoría de los contigs pertenecientes a virus Zeta codifican al menos un ORF sin codón de parada, y dentro de este grupo muchos se han visto que codifican ORFs en los dos sentidos de la cadena. También llama la atención el reservorio de este tipo de virus, ya que se han encontrado solo en conjuntos de datos medioambientales (humedal de Minnesota, Río Colorado, granos de cereales fermentados, biorreactor anaeróbico). Los virus Épsilon y Delta se diferencian de los virus Zeta y Eta en que pueden infectar organismos superiores y en un mayor tamaño de secuencia. Posiblemente los virus Épsilon-Delta compartan origen evolutivo con los virus Zeta-Eta, de manera que hayan evolucionado diferenciándose y adquiriendo la capacidad de infectar organismos superiores. Esto también explicaría la presencia de viroides y RNA satélites de plantas.

Los 50 elementos nuevos de RNA circular Tienen en común todos ellos en que solo se han encontrado en conjuntos de datos medioambientales al igual que los virus Zeta y Eta, y parecen ser diversos en cuanto su estructura, longitud y tamaños de ORFs codificados. No obstante, en los elementos tipo HPR-HPR encontrados destacan por no tener en la mayoría de los casos estructuras de varilla ni ORFs codificados en su secuencia con un tamaño superior a 100 aminoácidos. Pese a eso, se les pueden considerar como un elemento replicativo autónomo ya que resulta sorprendente encontrarse en un contig un RNA circular con ribozimas codificadas en cada polaridad. Probablemente se trate de RNAs virales satélite, elementos que se aprovechan o parasitan otros virus para la entrada en la célula, replicación, y encapsidación, mediante la utilización de proteínas del huésped o del virus con el que

coinfecta. Los elementos TwR-HPR y TwR-HHR también parecen tener estas características, aunque se han encontrado muy pocos en este trabajo como para poder inferir una conclusión.

Por otra parte, en los contigs con ribozimas HPR – HHR llaman la atención el hecho de que alguno de ellos tiene una longitud de secuencia mayor de 2000 nucleótidos. Además, parece haber un patrón consistente en que aquellos que tienen un tamaño de secuencia menor de 1000 nucleótidos y los que tienen un tamaño superior a 2000 nucleótidos nunca tienen una estructura con forma de varilla, mientras que aquellos que tienen estructura con forma de varilla se encuentran en el rango de tamaño de 1000-2000 nucleótidos, aunque no siempre se cumple que el hecho de tener este rango de tamaño vaya a determinar el tener una estructura de varilla. Esta observación nos da a entender que hay varios tipos de virus distintos dentro de este tipo de elementos.

Con todo lo visto anteriormente queda patente la utilidad y el potencial que tienen las herramientas bioinformáticas para explorar, hallar y caracterizar el viroma en conjuntos de datos muy diversos procedentes de todo el mundo. El uso de infraestructuras automatizadas y con alta capacidad de procesamiento de datos pueden ser útiles para aprovechar vastas colecciones de secuencias de virus para recabar información orientar las políticas y directrices que predigan y mitiguen las pandemias emergentes, lo cual es necesario en un contexto donde se está produciendo un crecimiento de la población humana y la consecuente invasión de los hábitats de los animales que lleva a una mayor tasa de zoonosis.

## 6. Referencias

- ANTHONY, S. J. et al. (2013). A strategy to estimate unknown viral diversity in mammals. *mBio* 4, e00598-13.
- BEEN M.D.; WICKHAM, G. S. (1997). Self-cleaving ribozymes of hepatitis Delta virus RNA. *Eur. J. Biochem.*, 247:741-753.
- BERGNER, L. et al. (2021). 'Diversification of Mammalian Virus Deltaes by Host Shifting', *Proceedings of the National Academy of Sciences of the United States of America*, 118: e2019907118.
- BIBILLO, A.; EICKBUSH, T.H. (2004). End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. *J. Biol. Chem.*, 279:14945–14953.
- BOTELHO-SOUZA, L. F. et al. (2017). 'Hepatitis Delta: Virological and Clinical Aspects', *Virology Journal*, 14: 1–15.
- BUZAYAN, J., GERLACH, W., & BRUENING, G. (1986). Non-enzymatic cleavage and ligation of RNAs complementary to a plant virus satellite RNA. *Nature*, 323, 349–353.
- CALLAWAY, EWEN. (2020). «'It Will Change Everything': DeepMind's AI Makes Gigantic Leap in Solving Protein Structures». *Nature*, vol. 588, n.º 7837, pp. 203-04.
- CERVERA, A., URBINA, D., Y DE LA PEÑA, M. (2016). Retrozymes are a unique family of non-autonomous retrotransposons with hammerhead ribozymes that propagate in plants through circular RNAs. *Genome Biology*, 17(1), 135.
- CERVERA, A., Y DE LA PEÑA, M. (2020). Small circRNAs with self-cleaving ribozymes are highly expressed in diverse metazoan transcriptomes. *Nucleic Acids Research*, 48(9), 5054–5064.
- CERVERA, A.; DE LA PEÑA, M. (2014). Eukaryotic penelope-like retroelements encode hammerhead ribozyme motifs. *Mol Biol Evol.*, 31: 2941-2947.
- CHANG, W.-S. ET AL. (2019) 'Novel Hepatitis D-like Agents in Vertebrates and Invertebrates', *Virus Evolution*, 5: vez021.
- DARÒS, J.A.; FLORES, R. (1995). Identification of a retroviroid-like element from plants. *Proc. Natl. Acad. Sci. USA*, 92, 6856–6860.
- DE LA PEÑA M, CEPRIÁN R, CASEY JL, CERVERA A. (2021). Hepatitis Delta virus-like circular RNAs from diverse metazoans encode conserved hammerhead ribozymes. *Virus Evol.*; 7(1): veab016.
- DE LA PEÑA, M., & GARCÍA-ROBLES, I. C. (2017). The Hammerhead ribozyme: A long story for a short RNA. *Molecules*, 22(1): 78.
- DE LA PEÑA, M., GAGO, S., & FLORES, R. (2003). Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. *The EMBO journal*, 22(20), 5561–5570.
- DE LA PEÑA, M.; CEPRIÁN, R.; CERVERA, A. (2020). A Singular and Widespread Group of Mobile Genetic Elements: RNA Circles with Autocatalytic Ribozymes. *Cells*, 9, 2555.
- DE LA PEÑA, M.; GARCÍA-ROBLES, I. (2010A). Ubiquitous presence of the hammerhead ribozyme motif along the tree of life. *RNA*, 16: 1943-1950.
- DE LA PEÑA, M.; GARCÍA-ROBLES, I. (2010b). Intronic hammerhead ribozymes are ultraconserved in the human genome. *EMBO Rep.*, 11: 711-716.
- DOHERTY, E., & DOUDNA, J. (2001). Ribozyme structures and mechanisms. *Annu. Rev. Biophys. Biomol. Struct*, 30, 457-475.
- EDGAR, R.C., TAYLOR, J., LIN, V. et al. (2022). Petabase-scale sequence alignment catalyses viral discovery. *Nature* 602, 142–147.
- EICKBUSH, D.G.; EICKBUSH, T.H. (2010). R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Mol Cell Biol.*, 30:3142-3150.
- EPSTEIN, L.M.; GALL, J.G. (1987). Self-cleaving transcripts of satellite DNA from the newt. *Cell*, 48: 535-543
- FEDOR, M.J. (1999). Tertiary structure stabilization promotes hairpin ribozyme ligation. *Biochemistry*, 34:11040-11050
- FEDOR, M.J. (2000). Structure and function of the hairpin ribozyme. *J Mol Biol.*, 297: 269-291.
- FELDSTEIN, P.A.; BUZAYAN, J.M.; CHNG, G. (1989). Two sequences participating in the autolytic processing of satellite tobacco ringspot virus complementary RNA. *Gene*, 82: 53-61.
- FERBEYRE, G.; SMITH, J.M.; CEDERGREN, R. Schistosome satellite DNA encodes active hammerhead ribozymes. (1998). *Mol. Cell. Biol.*, 18: 3880-3888.

- FLORES, R., GAGO-ZACHERT, S., SERRA, P., SANJUÁN, R. & ELENA, S. F. Viroids: survivors from the RNA world? *Annu. Rev. Microbiol.* 68, 395–414 (2014).
- FLORES, R.; MINOIA, S.; CARBONELL, A.; GISEL, A.; DELGADO, S.; LOPEZ-CARRASCO, A.; NAVARRO, B.; DI SERIO, F. (2015). Viroids, the simplest RNA replicons: How they manipulate their hosts for being propagated and how their hosts react for containing the infection. *Virus Res.*, 209, 136–145.
- GARCÍA-ROBLES, I., SÁNCHEZ-NAVARRO, J., & DE LA PEÑA, M. (2012). Intronic hammerhead ribozymes in mRNA biogenesis. *Biol Chem.*, 393(11):1317-26.
- GILBERT, W. (1986). Origin of life: The RNA world. *Nature*, 319: 618.
- GRUM-TOKARS, V.; MILOVANOVIC, M; WEDEKIND, J. E. (2003). Crystallization and X-ray diffraction analysis of an all-RNA U39C mutant of the minimal hairpin ribozyme. *Acta Cryst.*, 59: 142–145.
- GUERRIER-TAKADA, C., GARDINER, K., MARSH, T., PACE, N., Y ALTMAN, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3 PART 2), 849–857.
- HAMMANN, C., LUPTAK, A., PERREAULT, J., & DE LA PEÑA, M. (2012). The ubiquitous hammerhead ribozyme. *RNA (New York, N.Y.)*, 18(5), 871–885.
- HAMPEL, A.; TRITZ, R. (1989). RNA catalytic properties of the minimum (-) sTRSV sequence. *Biochemistry*, 28: 4929-33.
- HETZEL, U. ET AL. (2019) 'Identification of a Novel Virus Delta in Boa Constrictors', *MBio*, 10: 1–8.
- HUTCHINS, C.J.; RATHJEN, P.D.; FORSTER, A.C.; SYMONS, R.H. (1986). Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res.*, 14:3627-3240.
- JIMENEZ, R. M.; POLANCO, J. A.; LUPTÁK, A. (2015). Chemistry and Biology of Self-Cleaving Ribozymes. *Trends Biochem Sci.*, 40: 648-661.
- JOHNSON, C. K. et al. (2020). Global shifts in mammalian population trends reveal key predictors of virus spillover risk. *Proc. R. Soc. B* 287, 20192736.
- JUMPER, J et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*.
- KAPER, J.M.; TOUSIGNANT, M.E.; STEGER, G. (1988). Nucleotide sequence predicts circularity and self-cleavage of 300-ribonucleotide satellite of arabis mosaic virus. *Biochem Biophys Res Commun.*, 154: 318-325.
- KRUGER, K., GRABOWSKI, P. J., ZAUG, A. J., SANDS, J., GOTTSCHLING, D. E., & CECH, T. R. (1982). Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, 31(1), 147–157.
- KUMAR S, STECHER G, LI M, KNYAZ C, AND TAMURA K (2018), 'MEGA X Molecular Evolutionary Genetics Analysis across computing platforms'. [Molecular Biology and Evolution](#) 35: 1547-1549.
- KUO, M. Y. ET AL. (1988) 'Characterization Self-Cleaving Antigenome Hepatitis', *Journal of Virology*, 62: 4439–44.
- LARKIN MA, BLACKSHIELDS G, BROWN NP, CHENNA R, MCGETTIGAN PA, MCWILLIAM H, VALENTIN F, WALLACE IM, WILM A, LOPEZ R, THOMPSON JD, GIBSON TJ, HIGGINS DG. (2007). [Clustal W and Clustal X version 2.0](#). *Bioinformatics*, 23, 2947-2948.
- LASDA, E.; PARKER, R. (2014). Circular RNAs: Diversity of form and function. *RNA*, 20, 1829–1842.
- LILLEY, D. (2019). Classification of the nucleolytic ribozymes based upon catalytic mechanism. *F1000Research*, 8, F1000 Faculty Rev-1462.
- LORENZ, RONNY, et al. (2011). «ViennaRNA Package 2.0». *Algorithms for Molecular Biology, BioMed Central*, vol. 6, n.º 1, p. 26.
- MARKHAM NR, ZUKER M. (2008). UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol.*; 453:3-31.
- MARTICK, M., HORAN, L. H., NOLLER, H. F., Y SCOTT, W. G. (2008). A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature*, 454(7206), 899–902.
- MARTICK, MONIKA, Y WILLIAM G. SCOTT. (2006). «Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis». *Cell*, vol. 126, n.º 2, 309-20.
- MÜLLER, S., APPEL, B., BALKE, D., HIERONYMUS, R., & NÜBEL, C. (2016). Thirty-five years of research into ribozymes and nucleic acid catalysis: where do we stand today? *F1000Research*, 5, F1000 Faculty Rev-1511.
- NAWROCKI EP, EDDY SR. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.*; 29(22):2933-5
- NISSEN, P., HANSEN, J., BAN, N., MOORE, P. B., Y STEITZ, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289(5481), 920–930.

- PARASKEVOPOULOU, S. et al. (2020). 'Mammalian Virus Delta without Hepadnavirus Coinfection in the Neotropical Rodent *Proechimys Semispinosus*', *Proceedings of the National Academy of Sciences of the United States of America*, 117: 17977–83.
- PEEBLES, C. L., PERLMAN, P. S., MECKLENBURG, K. L., PETRILLO, M. L., TABOR, J. H., JARRELL, K. A., & CHENG, H.-L. (1986). A self-splicing RNA excises an intron lariat. *Cell*, 44(2), 213–223.
- PRODY, G., BAKOS, J., BUZAYAN, J., SCHNEIDER, I., & BRUENING, G. (1986). Autolytic processing of dimeric plant virus satellite RNA. *Science*, 231: 1577–1580.
- RANGLES, J.W.; DAVIES, C.; HATTA, T.; GOULD, A.R.; FRANCKI, R.I. (1981). Studies on encapsidated viroid-like RNA I. Characterization of velvet tobacco mottle virus. *Virology*, 108, 111–122.
- RICCITELLI, N., AND LUPTÁK, A. (2013) 'HDV Family of Self-Cleaving Ribozymes', *Progress in Molecular Biology and Translational Science*, 120: 123–71.
- ROJAS, A. A.; VAZQUEZ-TELLO, A.; FERBEYRE, G.; VENANZETTI, F.; BACHMANN, L.; PAQUIN, B.; SBORDONI, V.; CEDERGREN, R. (2000). Hammerhead-mediated processing of satellite pDo500 family transcripts from Dolichopoda cave crickets. *Nucleic Acids Res.*, 28: 4037-4043.
- ROTH, A., WEINBERG, Z., CHEN, A., KIM, P., AMES, T., & BREAKER, R. (2014). A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat Chem Biol*, 10(1):56-60.
- RUBINO, L.; TOUSIGNANT, M.E.; STEGER, G.; KAPER, J.M. (1990). Nucleotide sequence and structural analysis of two satellite RNAs associated with chicory yellow mottle virus. *J Gen Virol.*, 71: 1897-1903.
- RUMINSKI, D.J.; WEBB, C.H.; RICCITELLI, N.J.; LUPTÁK, (2011). A Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis Delta virus (HDV)-like self-cleaving ribozymes. *J Biol Chem.*, 286:41286-41295.
- RUPERT, P.B.; FERRÉ-D'AMARE, A.R. (2001). Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature*, 410: 780–786.
- SALEHI-ASHTIANI, K.; LUPTÁK, A.; LITOVCHICK, A.; SZOSTAK, J.W. (2006). A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*, 313:1788-1792.
- SANGER, H.L.; KLOTZ, G.; RIESNER, D.; GROSS, H.J.; KLEINSCHMIDT, A.K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proc. Natl. Acad. Sci. USA*, 73, 3852–3856.
- SAVILLE, B., & COLLINS, R. (1990). A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora mitochondria*. *Cell*, 61(4):685-96.
- SHARMEEN, L., KUO, M., DINTER-GOTTLIEB, G., & TAYLOR, J. (1988). Antigenomic RNA of human hepatitis Delta virus can undergo self-cleavage. *J Virol.*, 62(8):2674-9.
- TANNER, N. K. (1999). Ribozymes: the characteristics and properties of catalytic RNAs. *FEMS Microbiology Reviews*, 23(3), 257–275.
- THE SEQUENCE READ ARCHIVE (SRA), 2007: *Getting Started*. <https://www.ncbi.nlm.nih.gov/sra/docs/>. Accedido 11 de mayo de 2022.
- VALADKHAN, S., & MANLEY, J. L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature*, 413, 701-707.
- VARADI, M et al. (2021). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*.
- WEBB, C.H.; RICCITELLI, N.J.; RUMINSKI, D.J.; LUPTÁK, A. (2009). Widespread occurrence of self-cleaving ribozymes. *Science*, 326:953.
- WEINBERG, Z., KIM, P., CHEN, T., LI, S., HARRIS, K., LÜNSE, C., & BREAKER, R. (2015). New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nat Chem Biol.*, 11(8):606-10.
- WILLE, M. ET AL. (2018) 'A Divergent Hepatitis D-like Agent in Birds', *Viruses*, 10: 720.
- WINKLER, W., NAHVI, A., ROTH, A., COLLINS, J., & BREAKER, R. (2004). Control of gene expression by a natural metabolite-responsive ribozyme. *Nature*, 428(6980):281-6.
- WU, H. N. et al. (1989) 'Human Hepatitis d Virus RNA Subfragments Contain an Autocleavage Activity', *Proceedings of the National Academy of Sciences of the United States of America*, 86: 1831–1835.
- ZUCCOLA, H. J., ROZZELLE, J. E., LEMON, S. M., ERICKSON, B. W. & HOGLE, J. M. (1998). Structural basis of the oligomerization of hepatitis Delta antigen. *Structure* 6, 821–830.