

Búsqueda y caracterización de RNAs catalíticos a través de aproximaciones bioinformáticas

Autor: Raquel Ceprián Mascuñano

Máster universitario en Bioinformática y Bioestadística UOC-UB.

Área del trabajo final: Bioinformática clínica.

Nombre Consultor/a: Luis Franco Serrano.

Nombre Profesor/a responsable de la asignatura: Marc Maceira Duch.

Fecha Entrega: 05/01/2021



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada 3.0 España de Creative Commons

Licencias alternativas (elegir alguna de las siguientes y sustituir la de la página anterior)

A) Creative Commons:



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada 3.0 España de Creative Commons



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-CompartirIgual 3.0 España de Creative Commons



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial 3.0 España de Creative Commons



Esta obra está sujeta a una licencia de Reconocimiento-SinObraDerivada 3.0 España de Creative Commons



Esta obra está sujeta a una licencia de Reconocimiento-CompartirIgual 3.0 España de Creative Commons



Esta obra está sujeta a una licencia de Reconocimiento 3.0 España de Creative Commons

B) GNU Free Documentation License (GNU FDL)

Copyright © 2021 Raquel Ceprián Mascuñano.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free

Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

C) Copyright

© (Raquel Ceprián Mascuñano)

Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Búsqueda y caracterización de RNAs catalíticos a través de aproximaciones bioinformáticas</i>
Nombre del autor:	<i>Raquel Ceprián Mascuñano</i>
Nombre del consultor/a:	<i>Luis Franco Serrano</i>
Nombre del PRA:	<i>Marc Maceira Duch</i>
Fecha de entrega (mm/aaaa):	01/2021
Titulación::	<i>Máster universitario en Bioinformática y Bioestadística UOC-UB</i>
Área del Trabajo Final:	<i>Bioinformática clínica.</i>
Idioma del trabajo:	<i>Castellano</i>
Palabras clave	<i>RNA catalítico, ribozima, genómica comparada.</i>

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.*

El descubrimiento de que los RNAs son capaces de catalizar reacciones bioquímicas tuvo lugar en los experimentos del grupo de Cech, en el año 1981. Este grupo de investigación demostró el autoprosesamiento no enzimático del intrón del RNA ribosomal de *Tetrahymena*. Entre las ribozimas naturales descritas desde entonces, destacan la familia de las pequeñas ribozimas de autocorte, siendo la más común y sencilla la de cabeza de martillo (o "hammerhead") además de la ribozima del virus de la hepatitis delta, en las cuales se centrará este estudio.

El estudio a realizar se centrará fundamentalmente en el análisis de diversos motivos y secuencias de RNAs tanto catalíticos como circulares, y siguiendo fundamentalmente aproximaciones bioinformáticas. Dichas aproximaciones consistirán por un lado en el análisis y búsqueda de motivos autocatalíticos en genomas virales y eucarióticos utilizando software específico basado en estructuras de RNA (RNAmotif) como de genómica comparada (BLAST).

En los análisis realizados se llegó a la conclusión de que en diversos genomas tanto de tipo viral como animal, existen unas nuevas variantes de ribozimas hammerhead y del virus de la hepatitis delta, hasta ahora no descritas en la bibliografía. Además, se encontró un caso muy particular de ribozimas hammerhead ubicadas en el intrón del gen LASP del mosquito *A. albopictus* que parecen desempeñar un papel biológico conservado.

Abstract (in English, 250 words or less):

The discovery that RNAs are capable of catalyzing biochemical reactions took place in the experiments of Cech's group, in 1981. This research group demonstrated the non-enzymatic autoprocessing of the intron of Tetrahymena ribosomal RNA. Among the natural ribozymes described since then, we can highlight the family of small self-cleaving ribozymes, such as the hammerhead ribozyme, as the most common and simple example, or the hepatitis delta virus ribozyme, in which will focus this study.

The study to be carried out will mainly focus on the analysis of various motifs and sequences of catalytic and circular RNAs, and fundamentally following bioinformatic approaches. These approaches will consist on the one hand in the analysis and search for autocatalytic motifs in viral and eukaryotic genomes using specific software based on RNA structures (RNAmotif) and comparative genomics (BLAST).

From these analyses, we found in different viral and animal genomes the existence of new variants of the hammerhead and the hepatitis delta virus ribozymes, which have not been previously described in the literature. Furthermore, a very particular case of hammerhead ribozymes located in the intron of the LASP gene of the mosquito *A. albopictus* was also found, which appear to play a conserved biological role.

Índice

1. Introducción	1
1.1 Contexto y justificación del Trabajo	1
1.1.1. RNAs circulares (circRNA).	1
1.1.2. Ribozimas	6
1.1.3. Ribozimas pequeñas de autocorte	7
1.1.3.1. Ribozima Hammerhead (HHR).....	8
1.1.3.2. Ribozima del Virus de la Hepatitis Delta (HDV).....	9
1.2 Objetivos del Trabajo	11
1.3 Enfoque y método seguido.....	12
1.4 Planificación del Trabajo	13
1.5 Breve resumen de productos obtenidos	13
1.6 Breve descripción de los otros capítulos de la memoria.	14
2. Materiales y métodos.....	16
2.1. Búsqueda bioinformática de motivos ribozimáticos en los circRNAs tipo HDV de tritón y pez	16
2.2. Búsqueda bioinformática de ribozimas en genomas de termitas ...	18
2.3. Búsqueda de HHRs en otros invertebrados.	22
2.4. Búsqueda de bioinformática de HHRs en genomas de <i>A. aegypti</i> y <i>A. albopictus</i>	25
2.5. Caracterización bioinformática y molecular de ribozimas tipo III en el intrón del gen LASP de <i>A. albopictus</i>	26
3. Resultados.	27
3.1. Caracterización bioinformática de las secuencias de diversos circRNAs tipo hepatitis descritos en diversos animales.....	27

3.2. Búsquedas bioinformáticas de otros circRNAs en las bases de datos.....	29
3.3. Búsquedas bioinformáticas en genomas eucarióticos de invertebrados de las ribozimas detectados en circRNAs descritos.....	31
3.4. Caracterización de ribozimas HHRs tipo III en el intrón del gen LASP de <i>A. albopictus</i>	35
4. Conclusiones.....	39
5. Glosario.....	42
6. Bibliografía.....	45

Lista de figuras

Figura 1. Biogénesis de RNA largos no codificantes. _____	1
Figura 2. Representación esquemática de las 3 clases de patógenos subvirales y sus propiedades. _____	3
Figura 3. Reacción de transesterificación interna catalizada por las hammerhead, hairpin, HDV, VS y glmS. _____	7
Figura 4. Representación esquemática de las tres posibles estructuras de ribozimas de hammerhead (HHR) (Tipos I, II y III). _____	8
Figura 5. Familia HDV de ribozimas autodescindibles. _____	10
Figura 6. Alineamiento de las secuencias de pez y tritón. _____	16
Figura 7. Posible estructura secundaria I de la ribozima pez HDV-like y tritón HDV-like. _____	17
Figura 8. Posible estructura secundaria II de la ribozima pez HDV-like y tritón HDV-like. _____	18
Figura 9. Descripción del script usado para las búsquedas de HHRs en genomas de termitas. _____	19
Figura 10. Ejecución de la orden para la búsqueda de la estructura de la HHR en termitas. _____	20
Figura 11. Resultados de la búsqueda en RNAmotif. _____	21
Figura 12. Alineamiento en ClustalX de <i>Coptotermes formosanus</i> , <i>Cryptotermes secundus</i> y <i>Zootermopsis nevadensis</i> . _____	21
Figura 13. Descripción del script usado para las búsquedas de HHRs en genomas de himenópteros e invertebrados. _____	22
Figura 14. Ejecución de la orden para la búsqueda de la estructura de la HHR tipo III en himenópteros. _____	23
Figura 15. Resultados de la búsqueda de la estructura de la HHR tipo III en himenópteros. _____	23
Figura 16. Resultados de la búsqueda de la estructura de la HHRs tipo III en invertebrados. _____	24
Figura 17. Alineamiento en ClustalX de genomas de invertebrados. _____	24
Figura 18. Descripción del script usado para las búsquedas de HHRs en genomas de <i>Aedes</i> . _____	25

Figura 19. Ejecución de la orden para la búsqueda de la estructura de la HHR en <i>Aedes</i> .	26
Figura 20. Búsqueda en BLASTn con la secuencia molde usada en <i>A. albopictus</i> .	27
Figura 21. Organización del genoma del HDV humano y los circRNA de tipo HDV.	28
Figura 22. Estructura secundaria de una secuencia de ribozima antigenómica humana del virus de la hepatitis Delta (HDVR).	29
Figura 24. Estructura secundaria de las ribozimas de cabeza de martillo (HHR) antigenómicas (izquierda) y genómicas.	31
Figura 25. Resultados de la búsqueda en rnamotif en <i>Aedes</i> .	31
Figura 26. Alineamiento en ClustalX de <i>A. aegypti</i> y <i>A. albopictus</i> .	32
Figura 27. Alineamiento final de <i>Aedes</i> .	32
Figura 28. Estructura secundaria de las ribozimas de cabeza de martillo (HHR) tipo III presentes en los genomas de <i>A. albopictus</i> y <i>A. aegypti</i> .	34
Figura 29. Estructura secundaria de las ribozima de cabeza de martillo (HHR) tipo I presentes en los genomas de <i>A. albopictus</i> del grupo II.	34
Figura 30. Secuencia consenso de una familia min-HHR encontrada en repeticiones PLE del mosquito <i>A. aegypti</i> .	35
Figura 31. Análisis en BLAST de <i>A. albopictus</i> .	36
Figura 32. Intrón del gen LASP en <i>A. albopictus</i> . La flecha naranja indica la búsqueda realizada (query).	37
Figura 33. Ejemplo de las ribozimas HHR tipo III presentes en el intrón de LASP.	38

1.Introducción

1.1 Contexto y justificación del Trabajo

1.1.1. RNAs circulares (circRNA).

A diferencia de los DNAs circulares, que son moléculas ubicuas en genomas de bacterias, fagos y orgánulos, los RNAs circulares o circRNAs se han considerado como rarezas biológicas hasta muy recientemente. En la actualidad, se han descrito cinco tipos generales de circRNAs, incluyendo a los viroides y otros agentes subvirales de plantas, los intermediarios en reacciones de procesamiento en RNAs, o los productos del backsplicing del espliceosoma de pre-mRNAs nucleares en eucariotas entre otros. La mayoría de circRNAs se forman a través de enlaces 5' – 3', a diferencia de los RNAs "lariat" que se forman a través de enlaces 2' – 5' formados por un ataque nucleofílico en las ramificaciones durante el empalme (1).

Como se muestra en la Figura 1F los circRNAs génicos son generados a partir de exones empalmados. Los intrones flanqueantes de los exones de circRNAs a menudo contienen elementos repetidos Alu en direcciones opuestas (2-4) pero algunas proteínas que se unen a los intrones flanqueantes también pueden regular la formación de circRNAs (5,6).

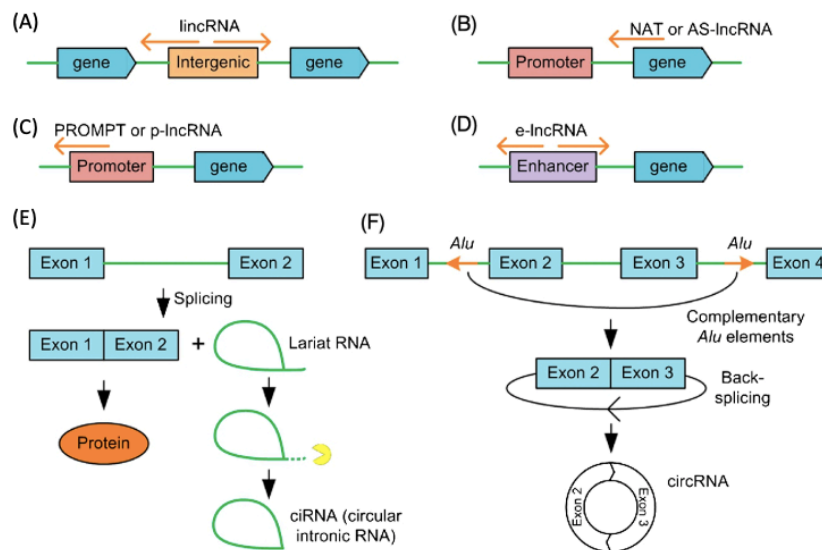


Figura 1. Biogénesis de RNA largos no codificantes. (A) Biogénesis de RNA intergénico largo / interviniendo (lincRNA). (B) Biogénesis de NAT (transcripción antisentido natural) o AS-lincRNA (lincRNA antisentido). (C) Biogénesis de PROMPT (Promoter upstream transcript) o

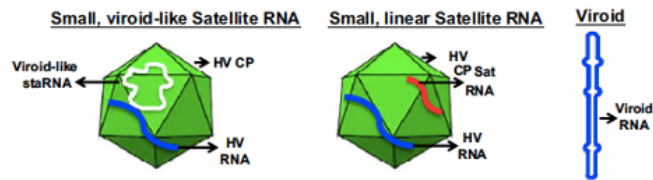
promoter lncRNA (p-lncRNA). (D) Biogénesis del lncRNA potenciador (e-lncRNA). (E) La biogénesis del RNA intrónico circular (ciRNA). (F) Biogénesis del RNA circular (circRNA) que se origina a partir de exones empalmados en sentido inverso. (7)

Los circRNAs prevalecen biológicamente porque tienen distintas propiedades que les pueden ser ventajosas. Por ejemplo, los circRNAs pueden servir como plantillas para la replicación del círculo rodante, que es una forma eficiente de generar copias a partir de un circRNA específico. Además tienen la capacidad de reorganizar la información genética presente en el genoma de DNA. También los circRNAs son muy estables debido a su inaccesibilidad por parte de las exonucleasas. Por último, también pueden crear una restricción en el plegamiento del RNA, lo que puede ser beneficioso en ciertos contextos (1).

Entre los diversos tipos de circRNA conocidos, los primeros en ser descritos fue un grupo de circRNAs infecciosos como son los viroides y satélites virales de plantas, y el virus de la hepatitis delta humana (HDV). Estos agentes tienen pequeños genomas de RNA circular que les permiten la replicación a través del mecanismo de círculo rodante en la que pueden crear múltiples copias genómicas a partir de un solo evento de iniciación (8-10).

1.1.1.1. Satélites virales y viroides.

Un grupo de circRNAs no codificantes y altamente estructurados, como los circRNAs satélites de virus (satRNA) y los viroides, son los patógenos más pequeños conocidos. Son agentes subvirales de plantas que han evolucionado para manipular los sistemas celulares de los huéspedes que invaden (11-13) (Figura 2). Los circRNAs satélites y los virus satélites se diferencian entre sí en que estos últimos codifican una proteína que empaqueta el RNA del virus satélite del mismo nombre (14,15). Es decir, los virus satélite son satRNAs que codifican y están encapsulados en sus propias proteínas de la cápsida (CP)(16).



	<u>Viroid-like satRNA</u>	satRNA lineal	Viroide
Huésped	Plantas	Plantas	Plantas
Virus auxiliar	Ej. SNMV	Ej. CMV	Ninguno
Genoma	ssRNA	ssRNA	ssRNA
Replicación	Replicación en virus auxiliar	Replicación en virus auxiliar	Polimerasa celular
Sitio de replicación	Citoplasma	Núcleo/citoplasma	Núcleo
Capacidad de codificar	Ninguna	Ninguna	Ninguna
Encapsidación	Proteína de la cápsida en virus auxiliar	Proteína de la cápsida en virus auxiliar	Ninguna

Figura 2. Representación esquemática de las 3 clases de patógenos subvirales y sus propiedades. HV, helper virus; SNMV, *Solanum Odiflorum Mottle Virus*; CMV, *Cucumber Mosaic Virus*; ssRNA, RNA monocatenario; CP, proteína de la cápsida. (Adaptado de A.L.N.Rao & Kriton Kalantidis, 2015).

Los satélites virales se definen como virus que tienen un ciclo de vida dependiente de un virus auxiliar, pero que carecen de una larga secuencia de nucleótidos homólogos con el virus auxiliar y es prescindible para la proliferación de tal virus (17-20).

Estos elementos infecciosos, se encuentran tanto en células de organismos eucariotas como de procariontes. En primer lugar, pueden desempeñar un papel importante en la dinámica de población de virus y sus huéspedes, y los virus satélites pueden limitar en gran medida la proliferación de sus virus auxiliares. Por ejemplo, la presencia de islas de patogenicidad de *Staphylococcus aureus* (SaPIs), es un tipo de virus satélite que reduce la proliferación de fagos (21,22). Debido al papel crucial de los virus en la formación de las comunidades microbianas (23), los virus satélite pueden ser en sí mismos un factor clave de la estructura y función de la comunidad microbiana. En segundo lugar, los virus satélite pueden tener un papel importante en la virulencia al controlar los síntomas inducidos por sus virus auxiliares o al codificar genes de virulencia relevantes.

Por ejemplo, el virus de la hepatitis B (HBV) causa un importante problema de salud a escala mundial. Entre los pacientes con infección crónica de HBV,

existen muchos pacientes coinfectados con el virus de la hepatitis delta (HDV), un virus satélite que necesita al HBV para su propagación (24).

Los viroides fueron los primeros circRNAs no codificantes descritos en biología (25). Con el creciente conjunto de circRNAs descubiertos en todos los dominios de la vida, ya no se pueden considerar como una rareza (1), pero siguen siendo únicos en varios aspectos, incluida la autonomía de replicación en plantas hospedantes y su capacidad de inducir patogenicidad (26).

Los viroides no codifican para ninguna proteína, ni se encapsidan ni requieren de un virus auxiliar para infectar. Poseen un genoma monocatenario de circRNA compuesto entre 200-400 nucleótidos (nt), el cual es altamente autocomplementario, formando diversas regiones de doble cadena de RNA (dsRNA). Tiene capacidad para infectar a plantas superiores y en algunos casos inducir enfermedades (27,8). Este genoma de circRNA no codificante lleva toda la información para la especificidad del huésped, el sistema de replicación, la propagación sistémica, la localización subcelular y la interacción con los factores del hospedador. Por ello, todo su ciclo biológico de infección, replicación, movimiento intra e intercelular, superar la defensa del huésped y la inducción de síntomas son el resultado de interacciones con factores del huésped (11,13,28). Los viroides se clasifican en 2 familias; *Pospiviroidae* y *Avsunviroidae*. La familia *Pospiviroidae* tiene 26 miembros, cuya especie tipo es el viroide del tubérculo fusiforme de la patata (PSTVd), el primer viroide descrito (29,30). Estos poseen una región central conservada (CCR) y una región terminal conservada (TCR) o una horquilla terminal conservada (TCH) (8). La secuencia de CCR y la presencia o ausencia de TCR y TCH sirven para distribuir a los miembros de esta familia en cinco géneros (31). La familia *Avsunviroidae* está compuesta por al menos cuatro viroides, cuya especie tipo es el viroide del manchado solar del aguacate (ASBVd), y que se caracterizan por no contener motivos conservados. No obstante, estos viroides tienen la capacidad para formar estructuras de ribozimas de cabeza de martillo, también llamadas hammerhead o HHR (una clase particular de RNAs catalíticos) cuya función es la autoescisión de sus intermediarios replicativos multiméricos de ambas polaridades (8,32).

1.1.1.2. Retroelementos de RNA circular.

Los elementos transponibles (TEs) son secuencias de DNA que tienen la capacidad de integrarse en otra parte del genoma. Con algunas excepciones, los TEs han sido identificados en todos los genomas de eucariotas secuenciados hasta la fecha (33). Existen dos clases de TEs: Retrotransposones (clase I), que se transponen vía RNA como intermediario y los transposones de DNA (clase II) que se transponen directamente como DNA, sin RNAs intermediarios (34).

Existen tres ordenes principales de retrotransposones: con repeticiones terminales largas (LTRs), elementos intercalados largos (LINEs) y elementos intercalados cortos (SINEs). Los retrotransposones se propagan mediante un mecanismo de amplificación vía copia y pega que les permite acumularse en el DNA, dando lugar a la mayor parte de las repeticiones en los genomas eucariotas. Los LINEs son transcritos por la RNA polimerasa Pol II dando lugar a mRNAs de varios miles de pares de bases (35). Por ejemplo, la transcripción del LINE-1 (L1), que es la única clase activa de retrotransposones autónomos en humanos, contiene dos marcos abiertos de lectura (ORFs), los cuales son cruciales para la retrotransposición: ORF1 codifica para una proteína de unión al RNA y el ORF2 codifica para la proteína reverso transcriptasa (RT) y endonucleasa (36).

Los SINE son elementos transcritos generalmente por la RNA polimerasa Pol III, y constituyen una familia de retrotransposones no autónomos que no codifican ninguna proteína, pero que se retrotransponen secuestrando las actividades de RT y endonucleasa de una proteína asociada codificada por otros retrotransposones autónomos (37).

Recientemente, se ha encontrado que diversos retrotransposones contienen en su secuencia RNAs catalíticos o ribozimas de autocorte similares a los descritos en patógenos subvirales de plantas y animales. Así por ejemplo, se ha visto que existen unas variantes mínimas atípicas de la HHR en retrotransposones de la familia Penélope (PLEs) (38) y Terminon (39), mientras algunas copias altamente conservadas de HHRs en intrones de genes de amniotas que parece ser que juegan un papel en la biogénesis del mRNA (40). Más recientemente, se ha descubierto que las HHRs presentes en los genomas de multitud de plantas angiospermas y de animales, están involucradas en el procesamiento de nuevas familias de retrotransposones no autónomos, llamadas *retrozimas*, un

retroelemento con ribozimas HHR, las cuales se propagan a través de circRNAs (41,64).

1.1.2. Ribozimas

Los ribozimas, son enzimas cuyos centros catalíticos están compuestos íntegramente de RNA y además no requieren proteínas para la catálisis. El descubrimiento de los ribozimas en la década de los años 80 proporcionó una explicación al enigma sobre cual pudo ser la primera molécula en originarse, el DNA o las proteínas, siendo los RNA catalíticos la respuesta mas plausible (42). Quizás el ribozima mas importante en biología sea el ribosoma, ya que el RNA 23S parece comprender todo el sitio activo (43,44). Las estructuras de los ribosomas de *Escherichia coli* y *Thermus thermophilus* 70S han proporcionado muchos conocimientos estructurales nuevos acerca del RNA (42)

La catálisis del RNA en biología es considerada como una característica inusual de los RNAs (44). Originalmente se creía que todos los ribozimas eran metaloenzimas los cuales requerían de Mg^{2+} y otros iones metálicos divalentes para realizar el proceso de plegado y catálisis. Se había propuesto un mecanismo de dos metales (46) en el que los iones de Mg^{2+} hidratados desempeñaban el papel de ácido-base generales. Esta predicción parece haber sido correcta para diversos ribozimas (47), pero no se puede generalizar a todos los sistemas de ribozimas.

En muchos casos, el RNA no es un un andamio pasivo para la unión de iones metálicos, sino que es un participante activo en la catálisis ácido-base en el sentido de que los grupos funcionales de nucleótidos imitan los roles que los aminoácidos juegan en los sitios activos de las enzimas proteicas (42).

Como consecuencia, varios de los RNA pequeños que se autocortan no requieren estrictamente iones metálicos divalentes para la catálisis (48) y aún no se han observado iones metálicos divalentes en el sitio activo de la peptidil transferasa del ribosoma. Las pequeñas ribozimas llevan a cabo la escisión o ligación del enlace fosfodiéster de un sitio específico. Todas estas ribozimas se basan en una combinación de estrategias catalíticas para completar la autoescisión en un sitio activo formado por las estructuras secundarias y terciarias únicas de cada familia. La escisión implica un ataque nucleófilico por un oxígeno 2' en un enlace fosfodiéster adyacente, generando un producto de

fosfato cíclico 2'-3' y 5'-hidroxilo (Figura 3). Las estructuras cristalinas han sido esenciales para esclarecer cómo los grupos funcionales participan en el mecanismo catalítico utilizado por cada familia de ribozimas. Los entornos electrónicos locales dentro de los sitios activos cambian los valores de pKa, lo que permite que las nucleobases y los cofactores participen en la catálisis ácido-base general (49).

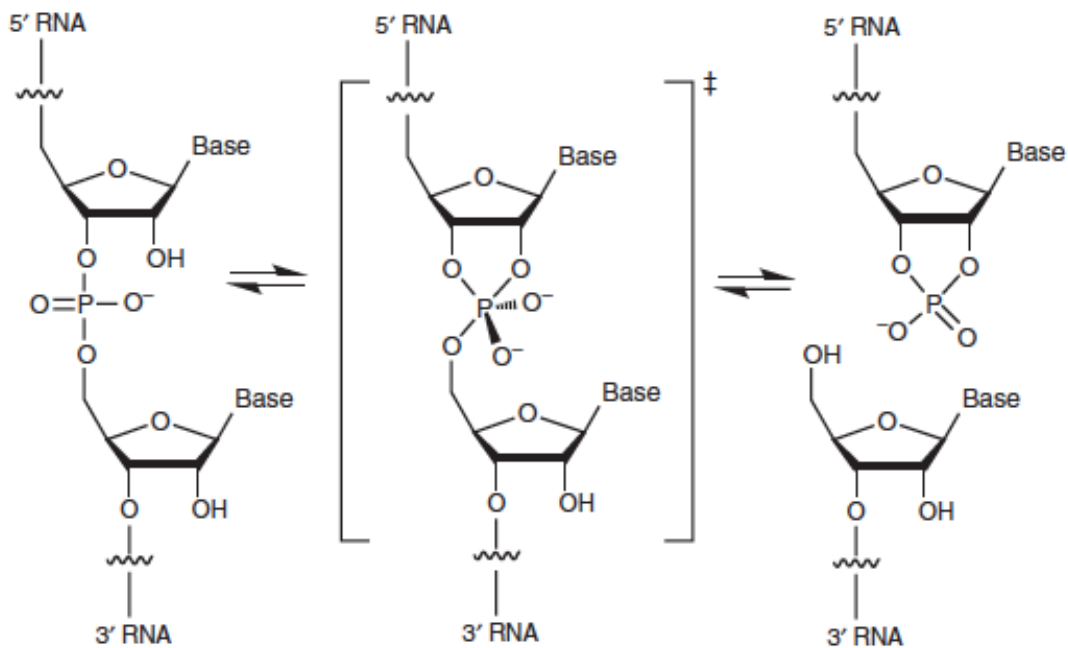


Figura 3. Reacción de transesterificación interna catalizada por las hammerhead, hairpin, HDV, VS y glmS. La reacción de escisión concertada transcurre sin intermediarios. Las ribozimas cabeza de martillo, horquilla y VS también catalizan la reacción de unión (49).

1.1.3. Ribozimas pequeñas de autocorte.

Las ribozimas pequeñas de autocorte varían entre 50 y 150 nucleótidos de longitud (49). Hasta el momento se conocen 9 clases de ribozimas pequeñas que de forma natural se autocortan y son: hammerhead (49,50) hairpin (51), hepatitis δ o delta humana (52), satélite Varkud (53), GImS (54), twister (55), twister sister, hatchet y pistol (56).

El estudio de ribozimas pequeñas ha proporcionado una plataforma para descubrir nuevos RNAs catalíticos y nuevas funciones para el RNA no codificante (ncRNA), además de ayudar en el diseño de nuevas moléculas para la biología sintética (57).

Este estudio, se va a centrar en la ribozimas hammerhead y la ribozima del virus de la hepatitis delta humana.

1.1.3.1. Ribozima Hammerhead (HHR).

La ribozima hammerhead (HHR) es el miembro mejor estudiado de las familias de las ribozimas. Tiene un núcleo conservado de 15 nucleótidos rodeado por 3 hélices (hélices I, II y III) las cuales adoptan un pliegue en forma de letra γ , donde la hélice I interactúa con la II, a través de interacciones terciarias características, que son necesarias para un autocorte eficiente (49,58,59). Existen 3 formas características de esta ribozima, denominadas tipo I, II o III, dependiendo de la hélice abierta del motivo (Figura 4).

Esta ribozima, fue descubierta en viroides y satRNAs, donde median el procesamiento de hebras multiméricas generadas durante el proceso de replicación (48,60). Más recientemente se han identificado en multitud de genomas de especies de fagos, procariontas y eucariotas como animales y plantas (61-64), incluido humanos (65).

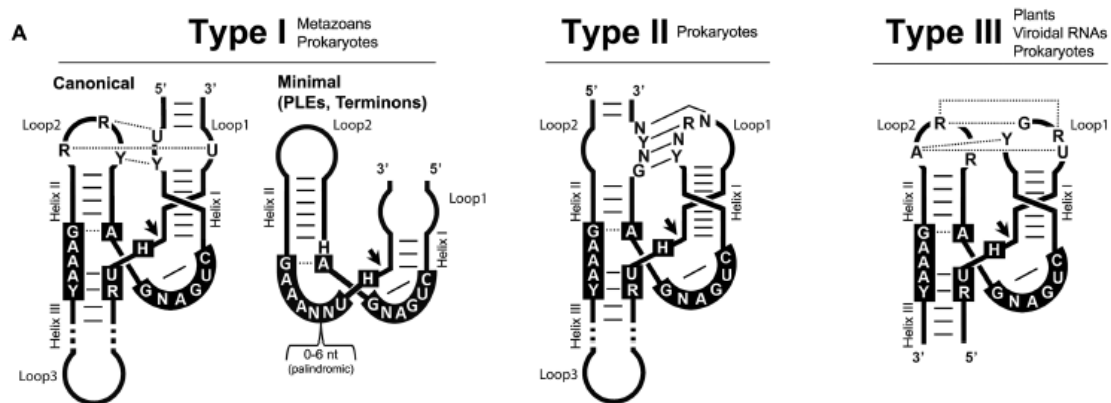


Figura 4. Representación esquemática de las tres posibles estructuras de ribozimas de hammerhead (HHR) (Tipos I, II y III). Se indican los nucleótidos más frecuentes en el núcleo catalítico (cajas negras) y en las interacciones bucle-bucle. Las líneas punteadas y continuas se refieren a pares de bases no canónicos y de Watson-Crick. Las flechas negras indican el sitio de autocorte. Los tres tipos de HHR se han encontrado en genomas de procariontas / fagos, mientras que los genomas de metazoos y plantas muestran principalmente motivos de tipo I y III, respectivamente (64).

1.1.3.2. Ribozima del Virus de la Hepatitis Delta (HDV).

El genoma de circRNA del virus de la hepatitis delta (HDV) depende de la coinfección con el virus de la hepatitis B (HBV) (67), pero se replica gracias a la maquinaria de celular del huésped a través de un mecanismo de círculo rodante (68). El genoma de HDV codifica dos pequeñas ribozimas (en el genoma y el antígeno del virus) relacionadas pero distintas, que se descubrieron por primera vez a través de su función en el procesamiento de copias concatémicas del genoma del circRNA viral durante su replicación (69-73).

Este ribozima tiene una longitud mínima de aproximadamente 85 nucleótidos y cataliza la escisión de su propia cadena principal de fosfodiéster en cis, dando como resultado la formación de productos de 5'-OH y 2', 3'-P cíclico.

La arquitectura general de la ribozima HDV consiste en un doble pseudonudo anidado que abarca las hélices P1, P1.1, P2, P3 y P4 (74-76) (Figura 5). El autocorte de las ribozimas HDVR genómicas y antigenómicas, que comparten aproximadamente el 75% de la secuencia, es algo esencial en el paso en el que se produce la replicación de círculo rodante del HDV (69,70).

Curiosamente, un ribozima muy similar al HDVR se descubrió en el propio genoma humano a través de experimentos de selección in vitro de una biblioteca genómica. El análisis de secuencia mostró que este nuevo ribozima está altamente conservado en multitud de mamíferos (77). Posteriormente, y utilizando aproximaciones bioinformáticas basadas en su estructura secundaria anidada de doble pseudonudo, se descubrió que, de forma similar a lo descrito para la HHR, existe toda una familia de ribozimas similares a la HDVR en todas las ramas del árbol filogenético de la vida (75).

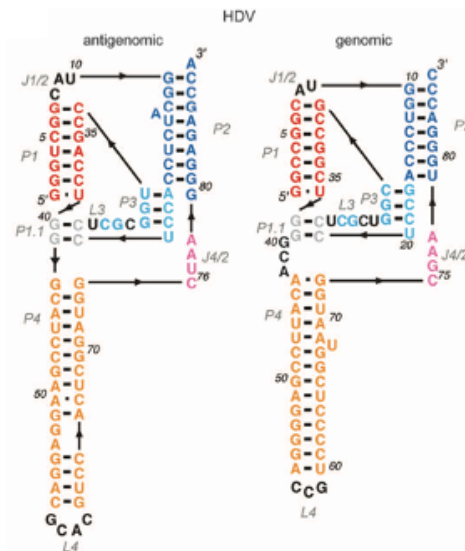


Figura 5. Familia HDV de ribozimas autodescindibles. Estructuras secundarias de las ribozimas HDV antigenómicas y genómicas (76).

1.1.4. Justificación del trabajo.

El principal tema de estudio se centrará en la detección de nuevos RNAs catalíticos o ribozimas codificados tanto en determinados agentes infecciosos de RNA como en genomas eucarióticos. A pesar de su elevado interés básico y biotecnológico como tijeras moleculares, los ribozimas de autocorte son un conjunto de RNAs que todavía no han sido estudiados con la suficiente profundidad. En el laboratorio donde se realizará el estudio se ha encontrado que la familia de las pequeñas ribozimas de autocorte no son una simple curiosidad de algunos agentes infecciosos de plantas, sino que se encuentran distribuidas en multitud de genomas de toda la escala biológica (78), donde desempeñan diversas funciones, desde procesos de retrotransposición de RNAs circulares en plantas (40) y metazoos inferiores (64), hasta control del splicing de anti-oncogenes en mamíferos (63,39).

Por ello, se procederá al estudio de estos ribozimas mediante programas bioinformáticos específicos, los cuales ayudarán a caracterizar tales ribozimas y poder saber cual es su papel principal en estos organismos.

1.2 Objetivos del Trabajo

1.1.1. Objetivo general.

El objetivo principal del estudio es un análisis y búsqueda de motivos autocatalíticos en genomas eucarióticos utilizando software específico basado en estructuras de RNA y de genómica comparada.

1.1.2. Objetivos específicos.

Como objetivos específicos tenemos:

1. Caracterización bioinformática de las secuencias de diversos circRNAs tipo hepatitis descritos recientemente en la literatura.
2. Búsquedas bioinformáticas de otros circRNAs en las bases de datos.
3. Búsquedas bioinformáticas en genomas eucarióticos de los ribozimas detectados en circRNAs descritos.
4. Caracterización molecular de la actividad de los ribozimas detectados.

1.3 Enfoque y método seguido

La detección bioinformática de pequeños motivos (<100nt) de RNAs estructurados y funcionales codificados en genomas compuestos por miles de millones de pares de bases es una labor especialmente complicada, debido sobre todo a la gran cantidad de falsos positivos que se llegan a obtener. Este estudio se centrará en la búsqueda de dos tipos de pequeños ribozimas (cabeza de martillo y Hepatitis delta) que inicialmente se detectó en diversos RNAs circulares de tipo Hepatitis delta descritos por diversos grupos (79-82).

El trabajo consistirá en comprobar si dichas nuevas ribozimas existen en diversos genomas eucarióticos utilizando aproximaciones bioinformáticas basadas en búsquedas por estructura, así como su correspondiente filtrado mediante genómica comprada. De esta forma, los motivos así hallados tendrán una alta probabilidad de existir como ribozimas reales, y para su confirmación, se podrán ensayar *in vitro* para determinar su actividad.

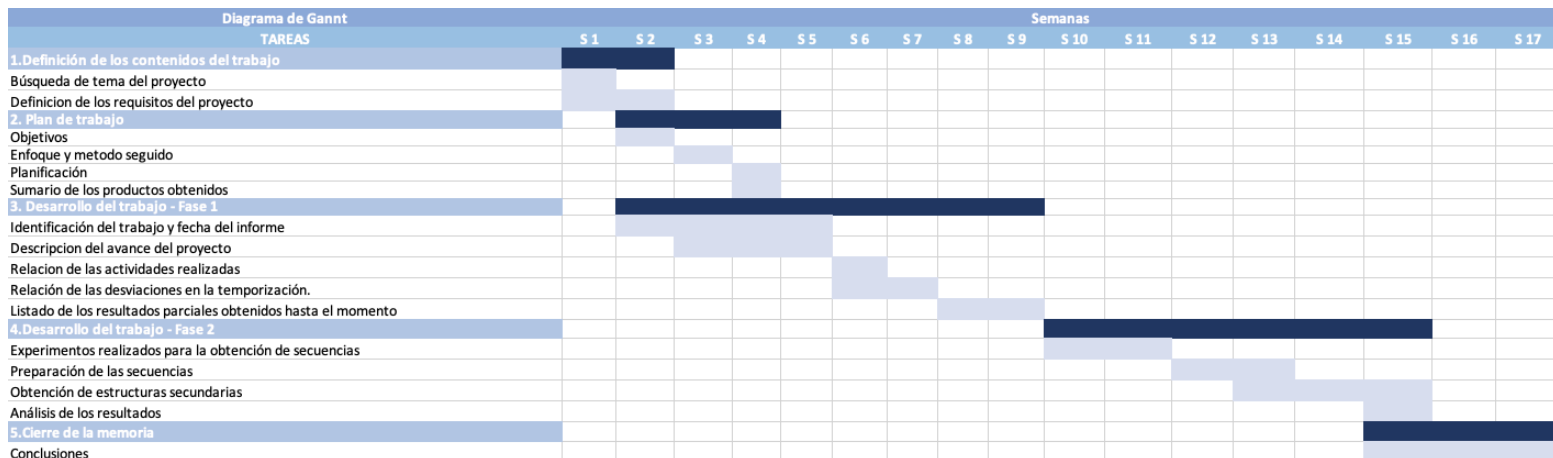
Para generar las secuencias de interés, se procede a usar el programa ClustalX. A finales de la década de 1990, ClustalW y ClustalX eran los programas de alineación múltiple más utilizados. Alineaban conjuntos de datos de tamaño mediano muy rápidamente y era fáciles de usar. Las alineaciones tenían una calidad suficiente para no requerir edición o ajuste manual con mucha frecuencia (83). En el laboratorio se decantó por el uso de este programa porque ofrece alineamientos con los que es posible identificar con facilidad las zonas que se han conservado o aquellas que han variado durante la evolución de la secuencias.

Para la visualización y el estudio estructural de las ribozimas, se utilizó RNAfold y Rnaviz. El primer programa indicado predice estructuras secundarias de mínima energía libre a partir de secuencias de RNA monocatenario. Respecto al programa RnaViz su función es producir dibujos de estructuras secundarias de calidad de moléculas de RNA, obtenidas del formato ct que se trata de la estructura mínima de energía libre (MFE) que proporciona RNAfold. Este programa permite cambiar la estructura presentada fácilmente. Las secuencias se pueden enumerar automáticamente y se pueden usar varios tipos de etiquetas para anotar bases o áreas concretas.

En cuanto a la búsqueda de otros circRNAs y motivos de RNA catalítico, se usó la herramienta BLAST sobre las bases de datos de NCBI, que compara una secuencia problema (query sequence) con todas las secuencias depositadas en la base de datos. Además de encontrar secuencias idénticas o con similitud local, BLAST hace un alineamiento con la secuencia problema y calcula la significancia estadística de los resultados.

También se trabajó la web <https://www.ncbi.nlm.nih.gov> en la sección Assembly para descargarnos los genomas de interés para las búsquedas de nuevos motivos en los genomas usando el programa RNAmotif. El programa RNAmotif busca en una base de datos las secuencias que coincidan con un motivo previamente definido que describe la estructura secundaria buscada. Cada coincidencia significa que la secuencia dada es capaz de adoptar la estructura secundaria dada, pero no pretende ser predictiva (84).

1.4 Planificación del Trabajo



1.5 Breve resumen de productos obtenidos

En nuestro caso de estudio, se pretende caracterizar los motivos de autocorte presentes en RNAs circulares tipo HDV en diversos animales.

Se encontró la presencia de motivos tipo HHR en agentes tipo HDV de sapos y termitas.

Los motivos encontrados, se utilizarán para definir nuevas búsquedas a partir de su secuencia primaria en diversos genomas y bases de datos, encontrando nuevas HHRs en genomas de termitas.

Por otro lado, búsquedas de estructuras secundarias tipo HHR permitió encontrar en otra especie animal, en concreto en *A. albopictus* o más comúnmente conocido como mosquito tigre, la presencia de HHRs en intrones génicos. Además, estas HHRs aparecieron en tándem (~300 pb) y en ambas polaridades del intrón, lo que sorprendentemente recuerda a un viroide de plantas.

Por último, tanto *A. albopictus* como *A. aegypti* (conocido como mosquito del dengue o de la fiebre amarilla) contienen múltiples ribozimas HHR del tipo III, pero que en realidad serían resultado de artefactos bioinformáticos debido a la presencia de miles de motivos HHR de tipo I previamente descritos en estos genomas.

1.6 Breve descripción de los otros capítulos de la memoria.

2. Materiales y métodos.

2.1. Búsqueda bioinformática de motivos ribozimáticos en los circRNAs tipo HDV de tritón y pez

2.2. Búsqueda bioinformática de ribozimas en genomas de termitas

2.3. Búsqueda de HHRs en otros invertebrados.

2.4. Búsqueda bioinformática de HHRs en genomas de *A. aegypti* y *A. albopictus*

2.5. Caracterización bioinformática y molecular de ribozimas tipo III en el intrón del gen LASP de *A. albopictus*.

3. Resultados.

3.1. Caracterización bioinformática de las secuencias de diversos circRNAs tipo hepatitis descritos en diversos animales.

3.2. Búsquedas bioinformáticas de otros circRNAs en las bases de datos.

3.3. Búsquedas bioinformáticas en genomas eucarióticos de invertebrados de las ribozimas detectados en circRNAs descritos.

3.4. Caracterización de ribozimas HHRs tipo III en el intrón del gen LASP de

A. albopictus.

4. Conclusiones.

5. Glosario.

2. Materiales y métodos.

En nuestro grupo de investigación se han realizado múltiples búsquedas bioinformáticas para realizar un análisis a posteriori de la viabilidad y funcionalidad de nuestras ribozimas de interés y saber de esta manera si tienen un papel funcional en los organismos estudiados.

2.1. Búsqueda bioinformática de motivos ribozimáticos en los circRNAs tipo HDV de tritón y pez

En primer lugar, para comenzar la búsqueda las ribozimas en los circRNAs tipo HDV descritos en tritón y pez, se procedió a la búsqueda de la secuencia completa del genoma del virus en ambas especies mediante la herramienta BLAST. El tamaño del circRNAs tipo HDV de pez es de 1,606 nt y en el caso de tritón tiene un tamaño algo mayor, 1,735 nt. Una vez obtenidas las secuencias de ambos agentes, se procedió a delimitar el principio y el final usando como referente el estudio realizado por Paraskevopoulou et al, (85).

Tras la selección de las secuencias de estudio, se usa el programa ClustalX para generar un alineamiento de las secuencias (Figura 6) y ver donde tienen más similitudes entre ellas y así poder detectar si existe una conservación de las secuencias.

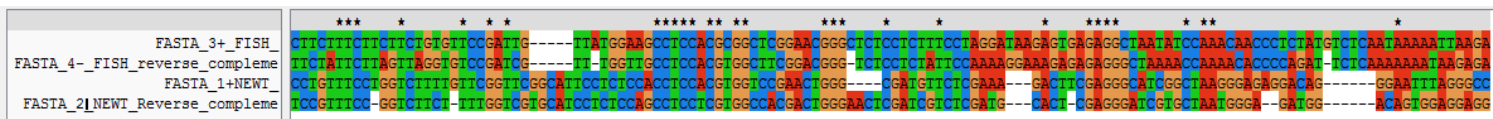


Figura 6. Alineamiento de las secuencias de pez y tritón. En la primera línea se muestra la polaridad positiva de la región de interés del genoma del circRNA tipo HDV de pez. En la segunda línea se presenta la polaridad negativa del mismo. En la tercera y cuarta línea se muestra la secuencia positiva y negativa del agente de tritón, respectivamente.

Una vez generado el alineamiento, se crea un archivo en formato .fasta que seguidamente se usará para generar posibles estructuras secundarias de la

ribozima de estudio en el servidor RNAfold, el cual mostró posibles formas de las ribozimas (Figura 7).

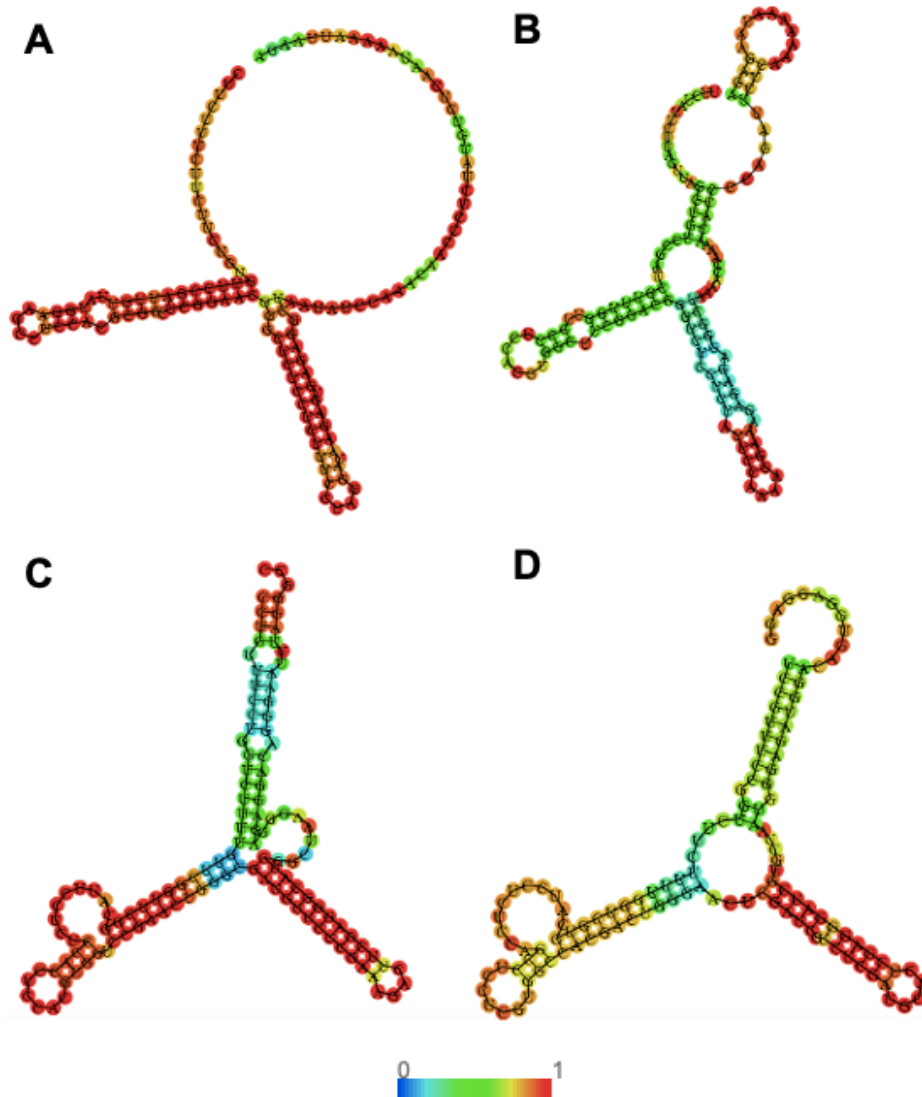


Figura 7. Posible estructura secundaria I de la ribozima pez HDV-like y tritón HDV-like. (A) Representación de la secuencia positiva pez; (B) Representación de la secuencia negativa pez; (C) Representación de la secuencia positiva tritón; (D) Representación de la secuencia negativa tritón. Los colores representan las probabilidades de emparejamiento de bases. Para las regiones no emparejadas, el color denota la probabilidad de no estar emparejado, en caso de encontrar el color rojo o naranja eso indica alta probabilidad de emparejamiento entre si.

Tras la creación de las posibles estructuras secundarias se pueden generar otro formato de estructuras con el programa RnaViz, el cual permite manualmente darle otra presentación a las estructuras calculadas. Para poder realizar este paso se obtuvieron los archivos en formato .ct del RNAfold.

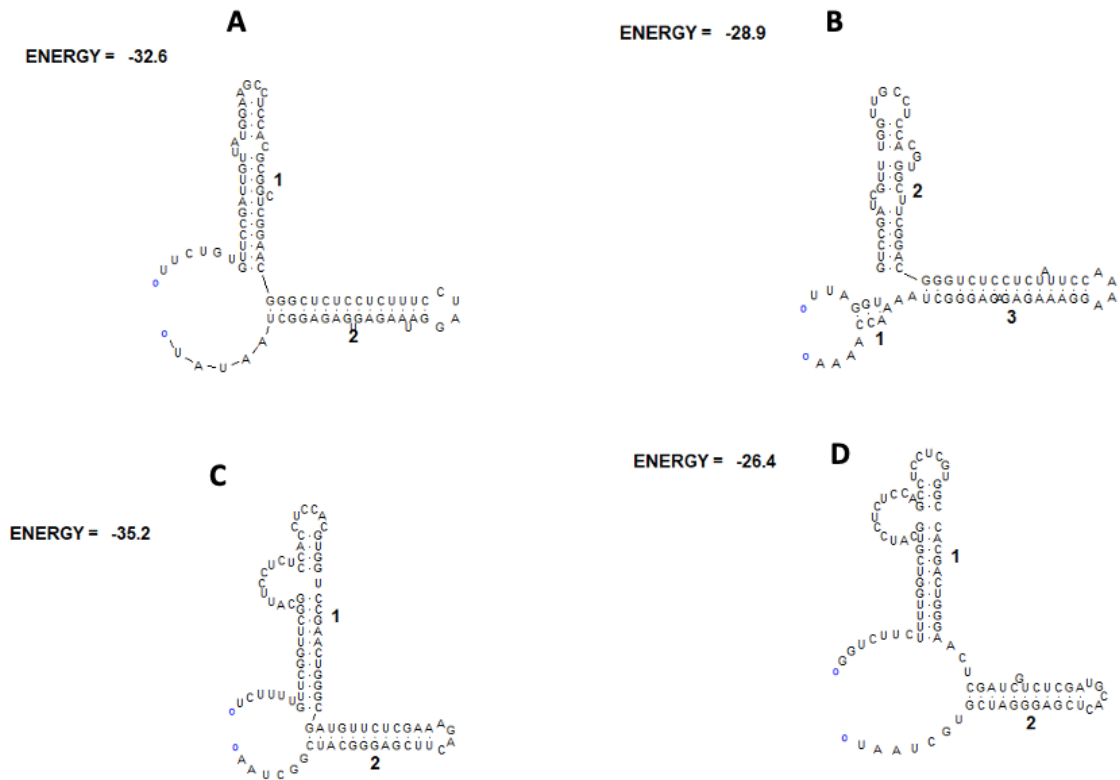


Figura 8. Posible estructura secundaria II de la ribozima pez HDV-like y tritón HDV-like. (A) Representación de la secuencia positiva pez; (B) Representación de la secuencia negativa pez; (C) Representación de la secuencia positiva tritón; (D) Representación de la secuencia negativa tritón.

Tras la creación de las últimas posibles estructuras secundarias, se procedió a generar la caracterización de la ribozima de forma manual teniendo como guía con las estructuras obtenidas.

A continuación, se procedió a seguir con búsquedas bioinformáticas, pero en otras especies y otros motivos.

2.2. Búsqueda bioinformática de ribozimas en genomas de termitas

Para las búsqueda de ribozimas como las HHRs en genomas, en primer lugar se necesita la secuencia del genoma completo de los organismos de estudio. Las búsquedas de tales genomas se realizaron mediante la herramienta BLAST,

de las cuales obtuvimos 3 especies: *Coptotermes formosanus*, *Cryptotermes secundus* y *Zootermopsis nevadensis*.

Tras la obtención de los genomas de interés se realizaron las búsquedas de ribozimas mediante el programa bioinformático RNAmotif. Este programa lo vamos a ejecutar desde el terminal de Linux y para realizar esta búsqueda es necesario generar en primer lugar un script en formato .descr con el siguiente contenido:

```
parms
    wc += gu;

#type III
descr
    ss (len=20)                # secuencia extra
    h5 (len=4, seq="u$")      # stem III (ends with U16)
    ss (len=1, seq="h")      # cleavage site (anything but G)
    h5 (minlen=4, maxlen=7)  # stem I
    ss (minlen=20, maxlen=40) # loop 1
    h3                        # stem I
    ss (minlen=7, maxlen=8, seq="^cuganga") # core
    h5 (minlen=4, maxlen=8)  # stem II
    ss (minlen=4, maxlen=10) # loop 2
    h3                        # stem II
    ss (len=3, seq="gaa$")   # core2
    h3 (seq="^a")           # stem III
    ss (len=20)
```

Figura 9. Descripción del script usado para las búsquedas de HHRs en genomas de termitas.

Lo que la Figura 9 representa es en primer lugar que las bases puedan emparejarse en una hélice estándar de Watson/Crick (W / C), es decir, A:U, G:C, C:G y U:A, lo que corresponde a la descripción en el script "wc +=gu". Seguidamente se procede a mostrar que la secuencia inicial tenga una longitud de 20 nucleótidos exactamente, indicado con "ss". A continuación, comienza la hélice III de nuestra HHR de interés, se muestra con "h5". Indicamos que la hélice debe de tener una longitud de 4 nucleótidos y que debe de terminar por el nucleótido "U". Continuando el recorrido de la hélice encontramos el primer nucleótido conservado, por ello mostramos que su longitud sea de 1 y que los

nucleótidos que puede contener son: A, U o C (indicado en el script por una “h”, usando el código de nomenclatura de la IUPAC).

En seguida encontramos la hélice I, que puede tener una longitud entre 4 y 7 nucleótidos y tras esta hélice aparece un bucle con un tamaño de 20-40 nt. Al finalizar este gran bucle cerramos la hélice I, denotado por “h3”, y aparece la tercera combinación de nucleótidos conservados. En este caso indicamos que nuestro bucle empiece por la secuencia: “CUGANGA”, la “n” indica que puede aparecer en esa posición cualquier nucleótido.

Una vez indicada la secuencia de la caja conservada se continua a crear la hélice II la cual debe contener una longitud de 4-8 nt y un segundo bucle que posea de 4-10 nt y se cierra la hélice 2. Para finalizar se indica que la ultima secuencia de cierre de la ribozima debe contener los nucleótidos “GAA” y a continuación se cierra la hélice III obligando a que el ultimo nucleótido que contenga sea la “A” para así poder aparear con el nucleótido inicial “U” y cerrar la ribozima.

Tras la creación del script deseado, mediante el terminal accedemos al programa RNAmotif y a continuación ejecutamos las órdenes para realizar la búsqueda indicada en nuestros genomas de interés.

```
./rnamotif -N 75000000 -descr HDV_HHR_Coptotermes.descr /media/Genomas_termitas/*.fa |
./rmprune > termitas.out

HDV_HHR_Coptotermes.descr: complete descr length: min/max= 99/140
```

Figura 10. Ejecución de la orden para la búsqueda de la estructura de la HHR en termitas.

En la Figura 10, tras acceder al programa rnamotif, se indica inicialmente la longitud máxima que puede tener la búsqueda de secuencias, a continuación se añade nuestro archivo creado anteriormente en formato *.descr* y accedemos a la carpeta donde tenemos guardados los genomas de estudio. En este caso vamos a ejecutar “rmprune”, el cual tiene la función de tomar la salida de rnamotif y elimina las soluciones descomprimidas, dejando solo esa solución con la raíz más larga posible. Por ultimo indicamos nuestro nombre del archivo y su formato final, *.out*.

Tras la finalización de la búsqueda (Figura 11) se comprueba si hemos obtenido resultados en nuestra búsqueda:

```

root@ubuntuM0151:/home/student/Desktop/Raquel/rnamotif-master/src# more termitas.out
#RM scored
#RM descr ss h5 ss h5 ss h3 ss h5 ss h3 ss h3 ss
#RM dfile HDV_HHR_Coptotermes.descr
>BLKM0100078.1 Coptotermes formosanus DNA, fragScaff_scaffold_89, whole genome shotgun sequence
BLKM0100078.1 0.000 0 76571 105 agacacagtaactaactggt gtgt t gggtc tacattcctttgactgaacc
gacc ctgatgat ggca cctcga tgct gaa acat gcacggggttttgtataata
>BLKM01000115.1 Coptotermes formosanus DNA, fragScaff_scaffold_42, whole genome shotgun sequence
BLKM01000115.1 0.000 0 379371 104 tgcacaatacctagactagt gtgt t gggat tgtattactttgttctaaaaa
attcc ctgaggac ggca tgcgc tgcc gaa acat gtaggagtttaactggct
>BLKM01000119.1 Coptotermes formosanus DNA, fragScaff_scaffold_152, whole genome shotgun sequence
BLKM01000119.1 0.000 1 43179 110 nnnnnngtgcctctgaa acgt c ggagg tttttaaaaaattgacaccac
ttcc ctgaggac ggagaca gagtg tgtctct gaa acgt cggagtttttaaacaa
tgg
>BLKM01000119.1 Coptotermes formosanus DNA, fragScaff_scaffold_152, whole genome shotgun sequence
BLKM01000119.1 0.000 1 43113 110 agacagagtgtctctgaa acgt c ggagg tttttaaacatggacaccac
ttcc ctgaggat ggagacg gagtg tgtctct gaa acgt cggatgtttttcaaac
ctg
>BLKM01000129.1 Coptotermes formosanus DNA, fragScaff_scaffold_53, whole genome shotgun sequence
BLKM01000129.1 0.000 0 472293 117 tcaaaaatccttgaaaaaat tggt t ataa
tagattagtttctttgtcaataaacacaatatc ttat ctgaaga tcaa catggctt ttgg gaa aata gatcaacgga
gactgcatgt

```

Figura 11. Resultados de la búsqueda en RNAmotif.

Como se puede observar rnamotif ha encontrado secuencias compatibles con las indicadas en nuestro script.

Finalizada la búsqueda, se procede a realizar cambiar el formato del archivo resultante termitas.out a termitas.fa, para así poder generar un alineamiento en ClustalX para así poder proceder a la caracterización de la ribozima y confirmar que tenemos la presencia de ella en estas 3 especies.

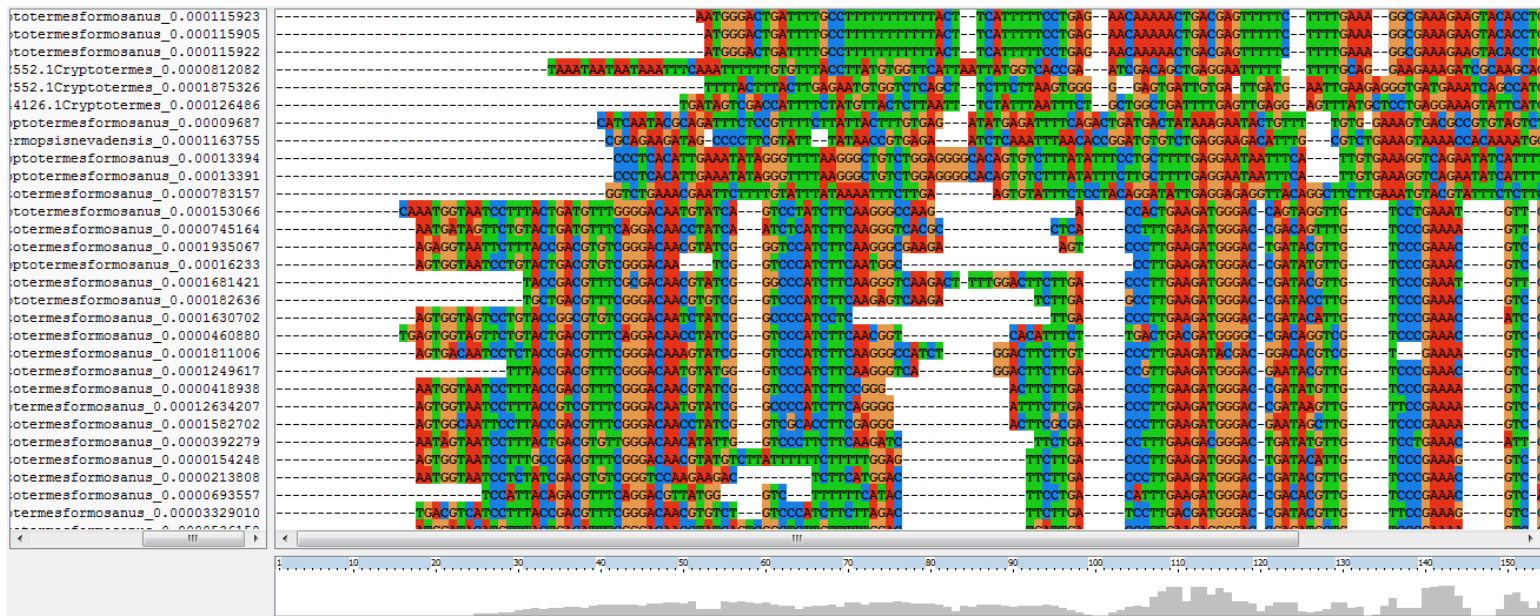


Figura 12. Alineamiento en ClustalX de *Coptotermes formosanus*, *Cryptotermes secundus* y *Zootermopsis nevadensis*.

2.3. Búsqueda de HHRs en otros invertebrados.

Como continuación de la búsqueda de HHRs tipo III en este caso se centra en invertebrados. Se realizó una búsqueda inicialmente en himenópteros, entre los cuales encontramos avispas, abejas, hormigas y moscas sierra. Los genomas de estas especies se obtuvieron mediante la búsqueda en BLAST en las secuencias de RefSeq. Siguiendo con el programa RNAmotif utilizamos un script diferente que en las búsquedas de termitas. En este caso se han realizado modificaciones en la hélice III, donde el primer nucleótido conservado "U", se indica con una "r" delante, lo cual significa que pueda aparecer cualquier purina ("A" o "G"), además de su longitud ampliando el rango entre 3-15 nucleótidos. El bucle 1 se le indica que tenga un rango de 3-20 nucleótidos. Con el resto de hélices y bucles se han hecho ligeras modificaciones y de nuevo para cerrar la hélice III se indica que acabe con una "A" y "h" lo que significa que pueda contener "A", "U" o "C".

```
parms
wc += gu;

#type III
descr

  ss (len=20)
  h5 (minlen=3, maxlen=15, seq="ru$")          # stem III (ends with U16)
  ss (len=1, seq="h")                          # cleavage site (anything but G)
  h5 (minlen=4, maxlen=8)                      # stem I
  ss (minlen=3, maxlen=20)                     # loop 1
  h3                                            # stem I
  ss (minlen=7, maxlen=8, seq="^cuganga")     # core
  h5 (minlen=4, maxlen=7)                      # stem II
  ss (minlen=3, maxlen=15)                     # loop 2
  h3                                            # stem II
  ss (minlen=3, maxlen=4, seq="gaa$")         # core2
  h3 (seq="^ah")                               # stem III
  ss (len=20)
```

Figura 13. Descripción del script usado para las búsquedas de HHRs en genomas de himenópteros e invertebrados.


```
./rnamotif -N 7500000 -descr III_hymenoptera.descr /media/Genomas_himenopteros/*.fa |
./rmprune > himenopteros.out
III_hymenoptera.descr: complete descr lenght: min/max= 79/148
```

Figura 14. Ejecución de la orden para la búsqueda de la estructura de la HHR tipo III en himenópteros.

Tras la ejecución del programa (Figura 14) obtuvimos los resultados mostrados en la Figura 15.

```
root@ubuntuM0151:/home/student/Desktop/Raquel/rnamotif-master/src# more himenopteros.out
#RM scored
#RM descr ss h5 ss h5 ss h3 ss h5 ss h3 ss h3 ss
#RM dfile III_hymenoptera.descr
>gnl|Acep_1.0|scaffold00001
gnl|Acep_1.0|scaffold00001 0.000 0 4009931 104 tcaatctaaaattcattcca ggcgt c taaggt
agtccccacactacatc gcctta ctgatga gttt agtgtg ggac gaa acgct acgaccgacca
aaaagtttc
>gnl|Acep_1.0|scaffold00002
gnl|Acep_1.0|scaffold00002 0.000 0 48825 109 ttggcgggaaggaagcgctt tgt t tcat
agggaaggcacgaaggta atgg ctgaaga cgaatg aaggcgaacacca catttg cgaa atg ctgatg
gctttgtgcaagta
>gnl|Acep_1.0|scaffold00002
gnl|Acep_1.0|scaffold00002 0.000 0 6989218 84 aagtatatatttgcacatcat tat t tttt acat agga
ctgatgat gatt aagtag agtc gaa atg ttcctctctcaatttatat
>gnl|Acep_1.0|scaffold00002
gnl|Acep_1.0|scaffold00002 0.000 0 11431662 95 gttttaaatgtttcaacttt gtat t tttaa
ttgtatttttat ttgaa ctgatgaa gattt tgt gagtt gaa atgt tttatcaagttatgtatg
t
>gnl|Acep_1.0|scaffold00002
gnl|Acep_1.0|scaffold00002 0.000 0 12748452 96 gttattcacatgaacgcata aat a agatt
ttaatgaaaattgtaac gattt ctgatga aatt aga aatt agaa att atgattaaaaacaataa
gt
>gnl|Acep_1.0|scaffold00002
gnl|Acep_1.0|scaffold00002 0.000 1 13949571 84 gtggaaaatgaagaagagat aat t gagga caa ttctt
ctgacgaa tttta taac tgga agaa att tatagtattgatgacaaaaat
>gnl|Acep_1.0|scaffold00002
gnl|Acep_1.0|scaffold00002 0.000 1 12749069 96 gttgttattcatgaagcata aat a agattgta
atgaaaactgt tacagttt ctgatga aatt aga aatt ggaa att atgacagaataaattat
at
```

Figura 15. Resultados de la búsqueda de la estructura de la HHR tipo III en himenópteros.

Puesto que los resultados no eran muy concluyentes se procedió a realizar una búsqueda mas generalizada en invertebrados, conservando el mismo script y haciendo esta nueva búsqueda en 157 genomas de invertebrados obtenidos también de BLAST de RefSeq.

```
root@ubuntuM0151:/home/student/Desktop/Raquel/rnamotif-master/src# more insects.out
#RM scored
#RM descr ss h5 ss h5 ss h3 ss h5 ss h3 ss h3 ss
#RM dfile III_hymenoptera.descr
>NT_033778.4 Drosophila melanogaster chromosome 2R
NT_033778.4 0.000 0 4392238 97 cctgttgatcctatgccagt gat t gcgg ggtgttccca ttgc ctgacga caga
gtgctcgcgagaaa tttg gaa att cgtatatagtttaaggctat
```

```

>NT_033778.4 Drosophila melanogaster chromosome 2R
NT_033778.4      0.000 0 5290181   87 catccaagcactgattaaa tggt a gtagc gcgt gttat ctgaaga aatgc
ggct gcatt gaa atg cgcttcgatctgcacaaaacc
>NT_033778.4 Drosophila melanogaster chromosome 2R
NT_033778.4      0.000 0 10480780  97 atgaagggcacatgggcaca tat a ttta gatcatcctg tgag ctgatga
tagct aagtaagcttc agtta tgaa ata ctttaaaattccttcgattc
>NT_033778.4 Drosophila melanogaster chromosome 2R
NT_033778.4      0.000 1 14378217 100 ctcttcgcacattatntaaa aat t agtt tccactttaac ggct ctgagga
ggac aataggttttcgcaa gttt tgaa att ccgaaacgaaccgcactcgt
>NT_033778.4 Drosophila melanogaster chromosome 2R
NT_033778.4      0.000 1 2148490   87 catccaagcactgattaaa tggt a gtagc gcgt gttat ctgaaga aatgc
ggct gcatt gaa atg cgcttcgatcttcacaaaacc
>NT_037436.4 Drosophila melanogaster chromosome 3L
NT_037436.4      0.000 0 24106338  87 catccaagcactgattaaa tggt a gtagc gcgt gttat ctgaaga aatgc
ggct gcatt gaa atg cgcttcgatctgcacaaaacc

```

Figura 16. Resultados de la búsqueda de la estructura de la HHRs tipo III en invertebrados.

Tras el análisis de los resultados obtenidos y hacer el alineamiento en Clustal X se observa la presencia de HHRs tipo III, aunque la mayoría de ella están truncadas, excepto en una especie de invertebrados que si que se observaban completas.

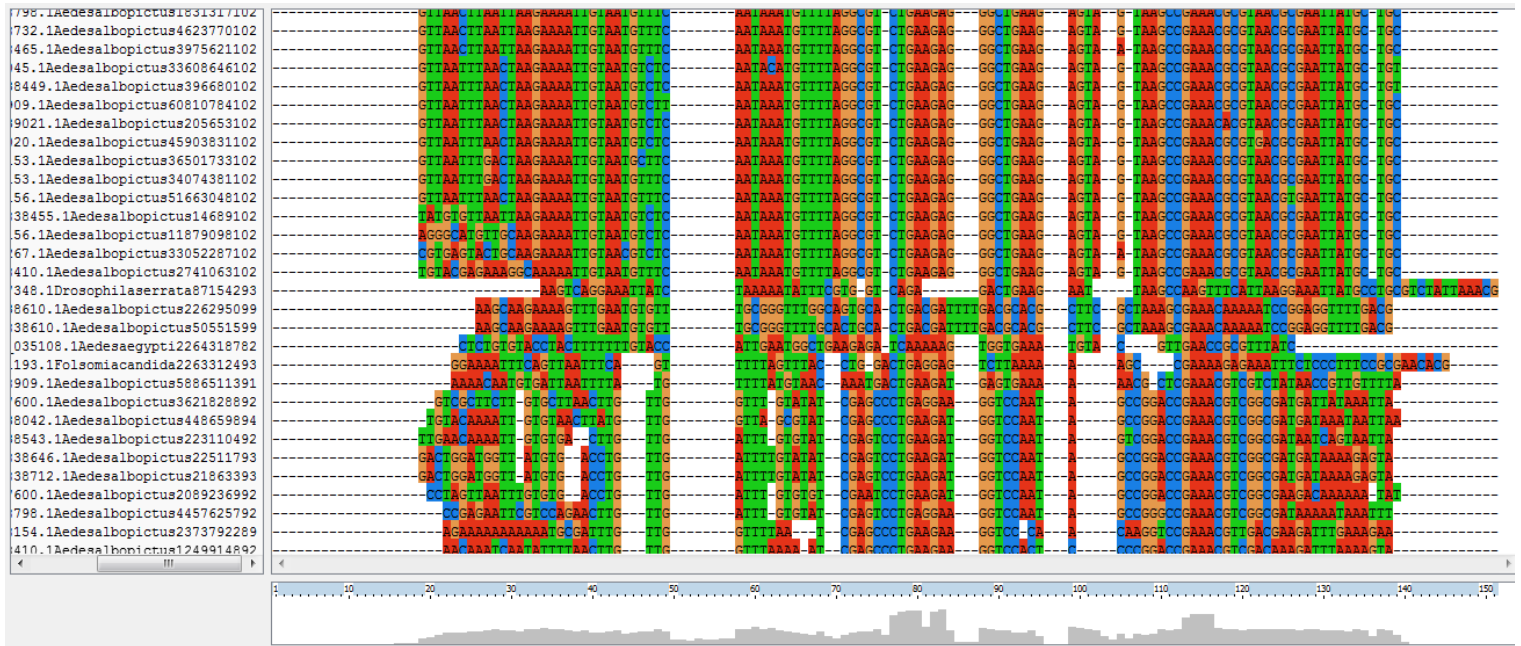


Figura 17. Alineamiento en ClustalX de genomas de invertebrados.

Esta familia de invertebrados en la cual se observa la HHR completa es Aedes, entre la que encontramos la especie *A. albopictus* y *A. aegypti* (Figura 17).

2.4. Búsqueda de bioinformática de HHRs en genomas de *A. aegypti* y *A. albopictus*.

Para proceder con la continuación de la búsqueda, se generó un nuevo script para poder trabajar nuevamente con RNAmotif. En este caso, se modificaron los parámetros para adaptarlo a la nueva ribozima de ambas especies de Aedes (*A. aegypti* y *A. albopictus*).

```
parms

wc += gu;

#type III

descr

ss (len=20)                # secuencia extra
h5 (len=4, seq="u$")      # stem III (ends with U16)
ss (len=1, seq="h")       # cleavage site (anything but G)
h5 (minlen=5, maxlen=7)   # stem I
ss (minlen=7, maxlen=15)  # loop 1
h3                         # stem I
ss (minlen=7, maxlen=8, seq="^cuganga") #core
h5 (minlen=5, maxlen=7)   # stem II
ss (minlen=3, maxlen=7)   # loop 2
h3                         # stem II
ss (len=3, seq="gaa$")    # core2
h3 (seq="^a")             # stem III
ss (len=20)
```

Figura 18. Descripción del script usado para las búsquedas de HHRs en genomas de Aedes.

Como se puede observar las nuevas modificaciones (Figura 18) aparecen en la longitud de las hélices y bucles. Los nucleótidos conservados no se modifica la descripción porque al ser una HHR tipo III estos deben de ser iguales en la ribozima aunque sean especies diferentes. Por ejemplo, en la hélice I (stem I) la longitud debe de ser entre 5-7 nt, y el bucle (loop 1) que viene a continuación es algo más pequeño que en el caso anterior, se indica que tenga entre 7-15 nucleótidos. Por ultimo se modifican los parámetros de la hélice II (stem II) y el bucle 2 (loop 2), indicando que la hélice II posea una longitud de 5-7 nt y el bucle 2 de 3-7 nucleótidos respectivamente.

En *Aedes*, estas ribozimas, tienen una estructura característica en cada especie, pero se unifica el script para obtener un resultado más genérico, aunque luego a la hora de caracterizar la ribozima no supone ningún problema, ya que al tener las secuencias y su entry correspondiente se puede seleccionar la especie de interés para construirla.

Una vez modificado el script correctamente procedemos a ejecutar las órdenes (Figura 19). Como se observa se ha cambiado el archivo `.descr` y la carpeta en la que debe realizar la búsqueda de los genomas (`Genomes_aedes`), donde se encuentran los genomas de ambas especies.

```
./rnamotif -N 75000000 -descr HDV_HHR_Mosquito_aegypti.descr  
/media/Genomes_aedes/*/*.fna | ./rmprune > mosquito.out  
HDV_HHR_Mosquito_aegypti.descr: complete descr lenght: min/max = 89/110
```

Figura 19. Ejecución de la orden para la búsqueda de la estructura de la HHR en *Aedes*.

2.5. Caracterización bioinformática y molecular de ribozimas tipo III en el intrón del gen LASP de *A. albopictus*.

Por último, siguiendo con *A. albopictus*, tras realizar las búsquedas correspondientes en BLAST aparecieron miles de secuencias truncadas en 5' como se indicó anteriormente. Tras estudiarlo con más detalle se observó que realmente correspondían con las de tipo I encontradas en elementos PLE.

De las ribozimas encontradas de tipo III se detectó un pequeño grupo que no tenían la inserción final en la caja conservada CUGANGA(N). Tras un análisis BLAST se encontró que en *A. albopictus* en realidad existían tres pares de estas HHRs muy cercanas entre si. Dichas copias se vio que mapeaban en intrones que pertenecían al mismo gen LASP que se encuentra en tres copias distintas (dos copias normales del gen y otro que aparenta ser un pseudogen). Además, también entre las dos parejas de HHRs había una tercera en la polaridad opuesta (Figura 20).

Figura 20. Búsqueda en BLASTn con la secuencia usada en *A. albopictus*.

3. Resultados.

3.1. Caracterización bioinformática de las secuencias de diversos circRNAs tipo hepatitis descritos en diversos animales.

Para poder confirmar la presencia o ausencia de motivos de autoescisión en estos circRNAs de tipo HDV, se realizaron inicialmente búsquedas basadas en homología. La búsqueda se centró en las regiones que poseían aproximadamente 400 pb, las cuales siguen inmediatamente las secuencias codificantes del antígeno delta (Figura 21). Las búsquedas realizadas en BLAST mostraron que una porción en la región 5' de la secuencia de tipo HDV del tritón claramente mapeaba con el extremo 5' del motivo antigenómico del HDVR (ribozima del virus de la hepatitis Delta) humano, lo que sugiere una sólida la presencia de motivos similares a HDVR en la secuencia de tritón (Figura 22). Las alineaciones de secuencia seguidas de inspección manual de las regiones de interés de la secuencia de tipo HDV de tritón confirmaron la presencia de los HDVR en los sentidos genómico o antigenómico (Figura 22B). Un análisis de secuencia similar reveló la presencia de motivos HDVR análogos en el circRNA de tipo HDV de pez (Figura 22C).

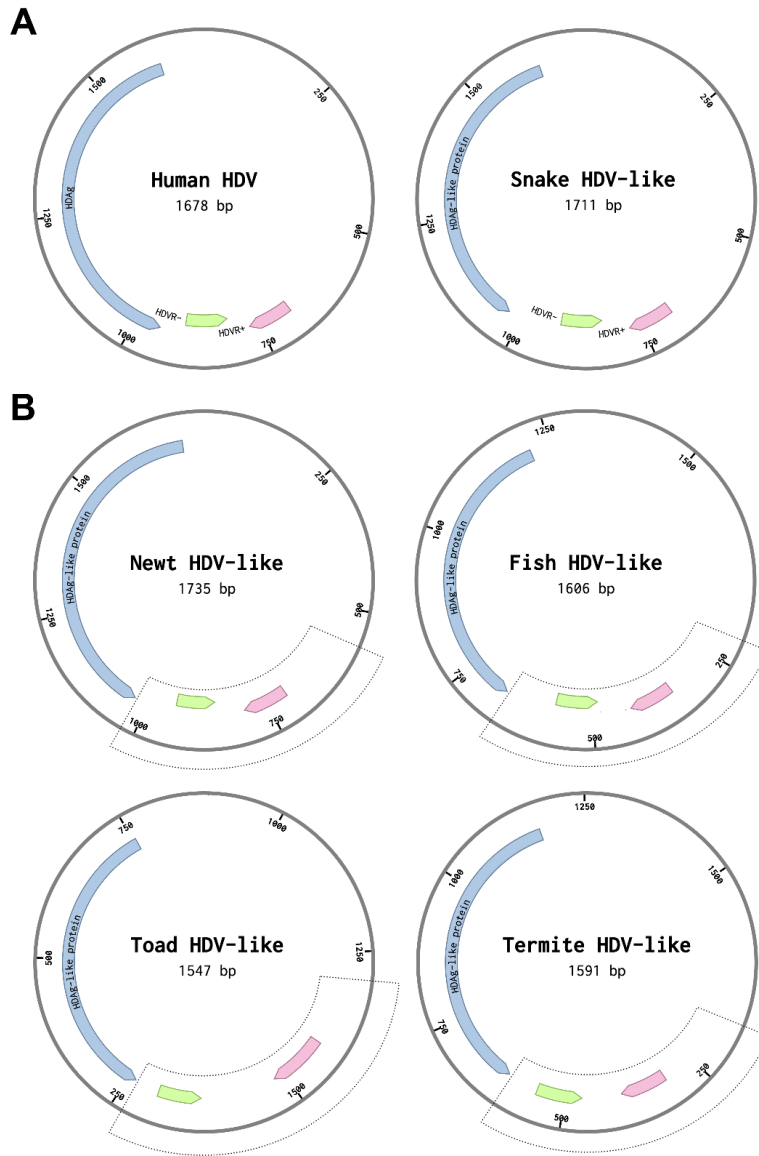


Figura 21. (A) Organización del genoma del HDV humano y los circRNA de tipo HDV de serpiente boa constrictor. Se indican las regiones de ribozima de HDV codificadoras de antígenos delta, genómicas y antigenómicas (en colores azul, rosa y verde, respectivamente). (B) circRNA de tipo HDV de tritón, pez, sapo y termitas. Las regiones utilizadas para el análisis de secuencia están marcadas en cuadros de línea de puntos. Se indican las regiones correspondientes a los motivos de autoescisión genómica y antigenómica.

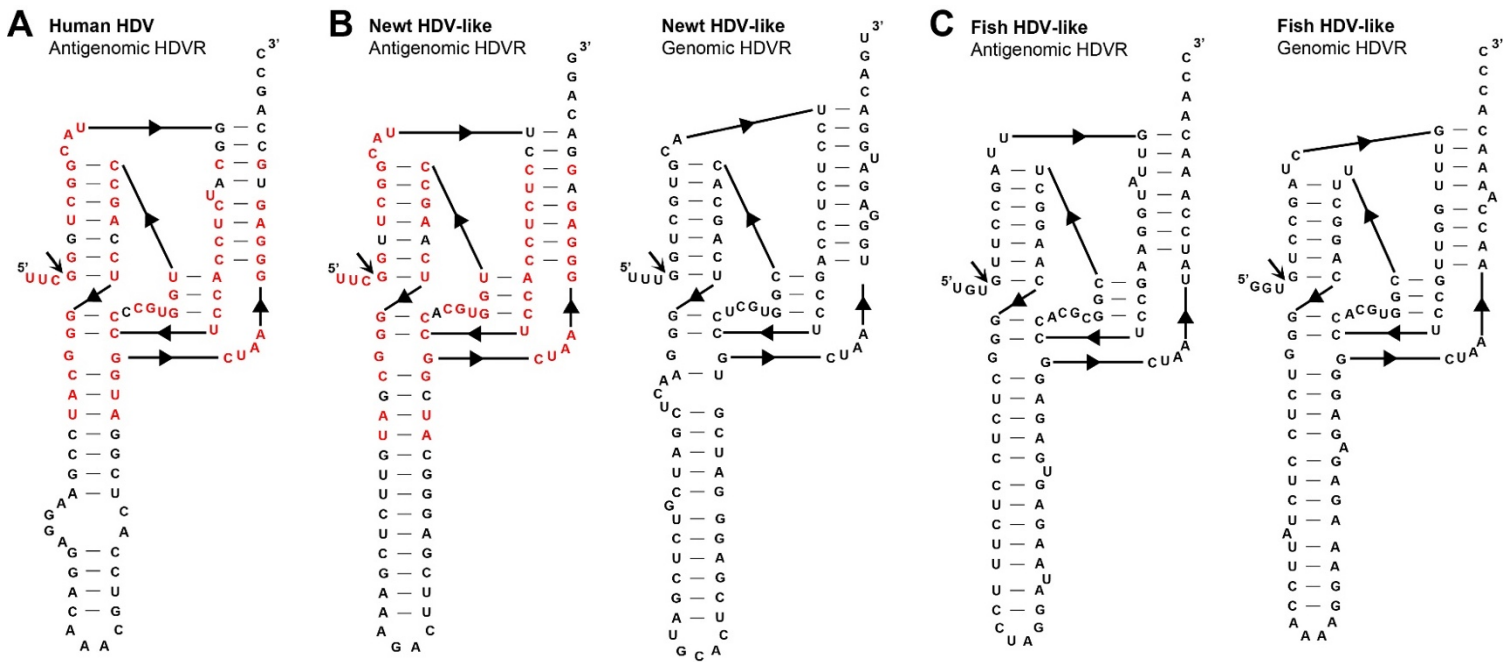


Figura 22. (A) Estructura secundaria de la secuencia de ribozima antigenómica humana del virus de la hepatitis delta (HDVR). La estructura secundaria de los HDVR antigenómicos (izquierda) y genómicos (derecha) presentes en genomas similares al HDV de (B) tritón y (C) pez. La similitud de secuencia entre los HDVR antigenómicos del HDV humano y el del tipo tritón HDV se resalta en los nucleótidos rojos.

3.2. Búsquedas bioinformáticas de otros circRNAs en las bases de datos.

Un análisis de secuencia previo de las mismas regiones de 400 nt de los circRNAs de tipo HDV de termitas no reveló ningún motivo HDVR. No obstante, los alineamientos de secuencia de las regiones genómicas y antigenómicas de termitas y sapo (Figura 24B) revelaron sorprendentemente la presencia de los motivos de secuencia conservados compatibles con el motivo HHR (Figura 24B). Las secuencias de los motivos HHR detectadas previamente en los circRNAs de

tipo HDV de sapo y termitas se utilizaron como consultas para búsquedas basadas en homología en bases de datos genómicas. No se obtuvieron coincidencias claras usando ninguno de los motivos HHR de sapo. Sin embargo, las búsquedas BLAST con el HHR antigenómico del genoma similar a la HHR de las termitas, revelaron la presencia de dos secuencias HHR muy similares en el genoma del ADN de la termita *C.formosanus* (Figura 24C, izquierda). De manera similar, las búsquedas BLAST con el motivo HHR genómico de tipo HDV de la termita identificaron otro par de motivos HHR en el genoma de la termita *C.secundus* (Figura 24C, derecha). Las búsquedas basadas en estructuras de HHR genómicos con scripts basados en esta nueva topología de tipo III y las interacciones terciarias conservadas no revelaron motivos de ribozimas claros en ninguno de los genomas de estas termitas ni en ningún otro insecto relacionado (como diversos himenópteros, hemípteros y una cucaracha).

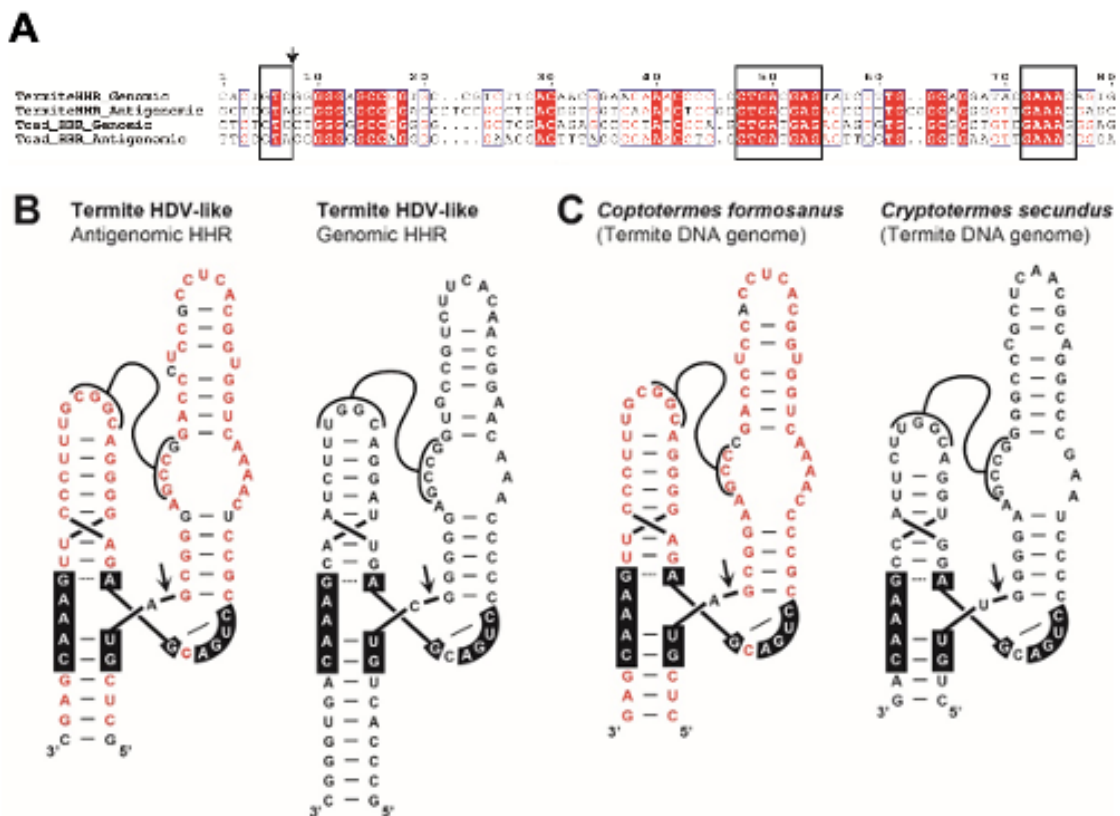


Figura 23. Estructura secundaria de las ribozimas de cabeza de martillo (HHR) antigenómicas (izquierda) y genómicas (derecha) presentes en los genomas similares al HDV de (A) una termita. (B) Ejemplos de HHR similares codificados en los genomas de ADN de las especies *C. formosanus* y *C. secundus*. La similitud de secuencia entre el HHR antigenómico del RNA de tipo HDV de termitas y un motivo HHR codificado por el genoma de ADN en la termita *C. formosanus* se resalta en los nucleótidos rojos.

3.3. Búsquedas bioinformáticas en genomas eucarióticos de invertebrados de las ribozimas detectados en circRNAs descritos.

Tras la comprobación de que nuestro script (Figura 25) ha funcionado y obtenemos los resultados deseados, se genera el archivo mosquito.out y procedemos a modificarlo a mosquito.fa para generar un alineamiento en ClustalX.

```
root@ubuntuM0151:/home/student/Desktop/Raquel/rnamotif-master/src# more mosquito.out
#RM scored
#RM descr ss h5 ss h5 ss h3 ss h5 ss h3 ss h3 ss
#RM dfile HDV_HHR_Mosquito_aegypti.descr
>NC_035109.1 Aedes aegypti strain LVP_AGWG chromosome 3, AaegL5.0 Primary Assembly, whole genome
shotgun sequence
NC_035109.1    0.000 1 20601723   99 atcagtttagtttaggtttc gtat a agttg agtctagtaca cgact
ctgaagat ggcctta cagt ttaggtc gaa atac gcgtatctgtcaaaggatac
>NW_021837489.1 Aedes albopictus isolate FPA unplaced genomic scaffold, Aalbo_primary.1
scaffold_14, whole genome shotgun sequence
NW_021837489.1    0.000 1 40690503   102 accacttgtaatcttcctca gtgt c tgtta tcagttatagataac tgaca
ctgaggaa gattac aagtg gtagtc gaa atac gcgtatctgtcaaaggataa
>NW_021837600.1 Aedes albopictus isolate FPA unplaced genomic scaffold, Aalbo_primary.1
scaffold_15, whole genome shotgun sequence
NW_021837600.1    0.000 0 25219910   96 aagtgtagtcgaaatcgc gtat c tgtcag tggataa ctgaca
ctgaggaa gattac aagtg gtagtc gaa atac gcgtatctgtcaaaggataa
```

Figura 24. Resultados de la búsqueda en rnamotif en *Aedes*.

Tras generar el alineamiento (Figura 26), se observan que aparecen varios grupos dependiendo de los nucleótidos que tienen en común, indicando que esas secuencias están conservadas.



Figura 25. Alineamiento en ClustalX de *A. aegypti* y *A. albopictus*.

Al tener ya ambas especies alineadas, se hace una limpieza de las secuencias, ya que hay muchas que no terminan de alinear correctamente y hace que el resto de secuencias aparezcan en posiciones desplazadas. Como resultado final del alineamiento se obtiene la Figura 27:

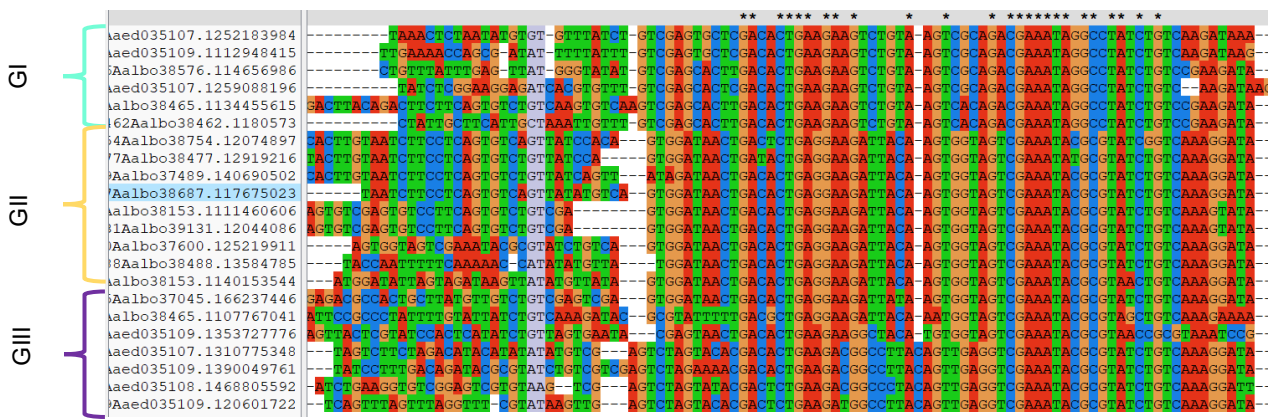


Figura 26. Alineamiento final de Aedes.

En el alineamiento, se observan 3 grupos diferenciados, marcados en la figura 27 por las llaves.

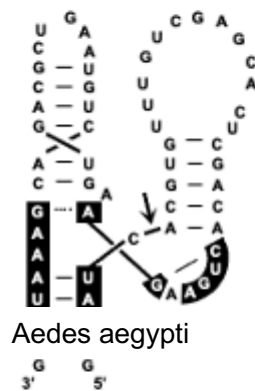
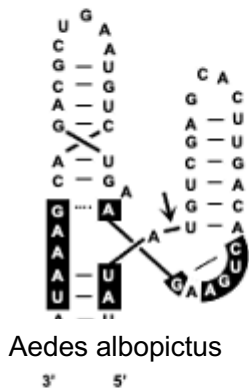
Este agrupamiento permite definir la caracterización de las ribozimas de interés y poder estudiar si corresponden a las de tipo III.

En el alineamiento de la figura 27, se observan 3 grupos diferenciados.

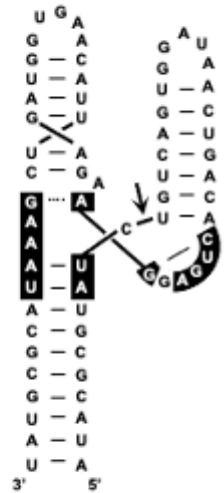
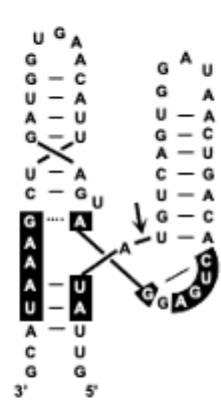
Este agrupamiento permite definir la caracterización de las ribozimas de interés y poder estudiar si corresponden a las de tipo III.

Tras generar la caracterización se observa que es posible formar las ribozimas HHR de tipo III pero que realmente la conformación real de estas es de tipo I. Es decir, bioinformáticamente se pueden crear pero biológicamente su conformación sería de tipo I que es realmente cuando podría realizar la función de autoescisión.

A GRUPO I



B GRUPO II



C GRUPO III

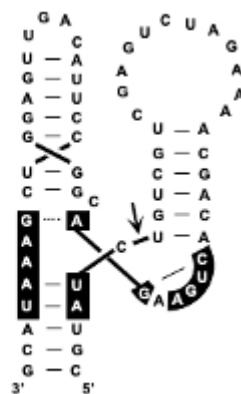
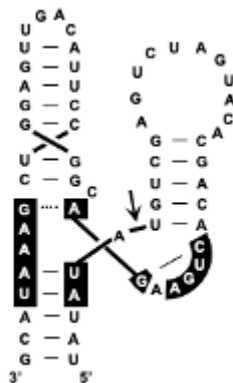


Figura 27. Estructura secundaria de las ribozimas de cabeza de martillo (HHR) tipo III presentes en los genomas de *A. albopictus* y *A. aegypti*. (A) Representación de las HHR tipo III en genomas de ambas especies pertenecientes al grupo I. (B) Ejemplo de HHR tipo III de *A. albopictus* de secuencias seleccionadas del grupo II y (C) ejemplo de HHR tipo III de *A. aegypti* de secuencias seleccionadas del grupo III.

Como se puede observar los bucles principales en las dos especies difieren bastante. En el caso de *A. albopictus* nos encontramos con un bucle pequeño que contienen entre 5-4 nt, pero en el caso de *A. aegypti* se observa que el bucle II contiene entre 10-11 nt, esto hace que su hélice en si sea más corta pero estable igualmente.

La figura 29, muestra cual seria el modelo real de la HHR tipo I con la secuencia elegida de *A. albopictus*.

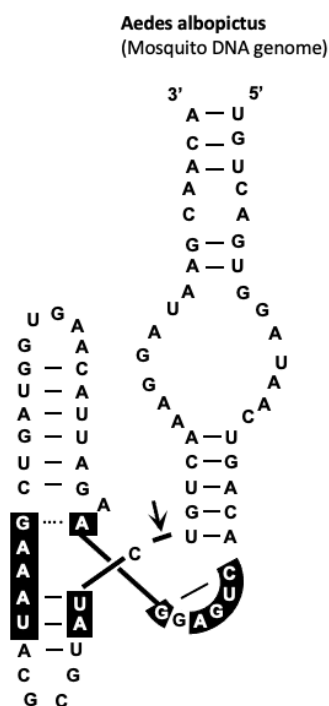


Figura 28. Estructura secundaria de las ribozima de cabeza de martillo (HHR) tipo I presentes en los genomas de *A. albopictus* del grupo II. Los nucleótidos marcados en negro muestran la conservación de ellos.

En este caso la conformación de la ribozima de tipo I, es totalmente diferente a las de tipo III. Nos encontramos con un bucle central formado entre 14-18 nt y luego dos más pequeños que contienen entre 4-5 nt. Como en los casos

anteriores nos encontramos regiones conservadas presentes en todas las ribozimas de tipo I.

3.4. Caracterización de ribozimas HHRs tipo III en el intrón del gen LASP de *A. albopictus*.

Como última parte de las búsquedas, como se comentó en el punto 2.4, se encontró que *A. aegypti* y *A. albopictus* contenían en su genoma la ribozima HHR tipo III según las búsquedas realizadas con RNAmotif.

Puesto que se observa claramente que *A. albopictus* posee múltiples repeticiones se procede a escoger una de las secuencias encontradas por el programa RNAmotif y realizar una búsqueda en BLASTN. Tras la búsqueda de BLAST se observa que aparecen truncadas en 5'.

Tras seguir estudiando estas ribozimas truncadas y realizar su caracterización se observa que en realidad pertenecen a las ribozimas HHRs pero tipo I.

Estas ribozimas tipo I se observó que estaban asociadas a elementos Penélope (PLE), ya encontradas previamente en *A. aegypti* (37) (Figura 30).

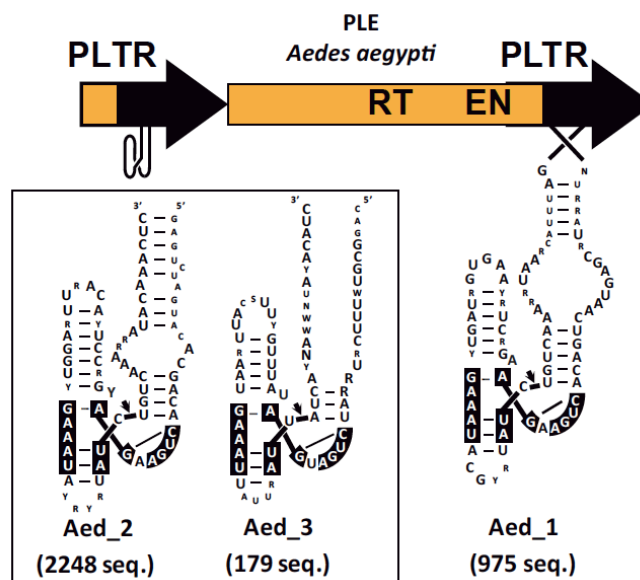


Figura 29. Secuencia consenso de una familia min-HHR encontrada en repeticiones PLE del mosquito *A. aegypti* (37).

Además de las ribozimas de tipo III encontradas, una de ellas no contenía inserción en la caja conservada CTGANGA (N) y al realizar el análisis en BLAST se descubrió que pertenecían al grupo I de ribozimas intrónicas.

En el análisis de BLAST aparecen tres pares de copias relativamente cercanas entre si (Figura 31).

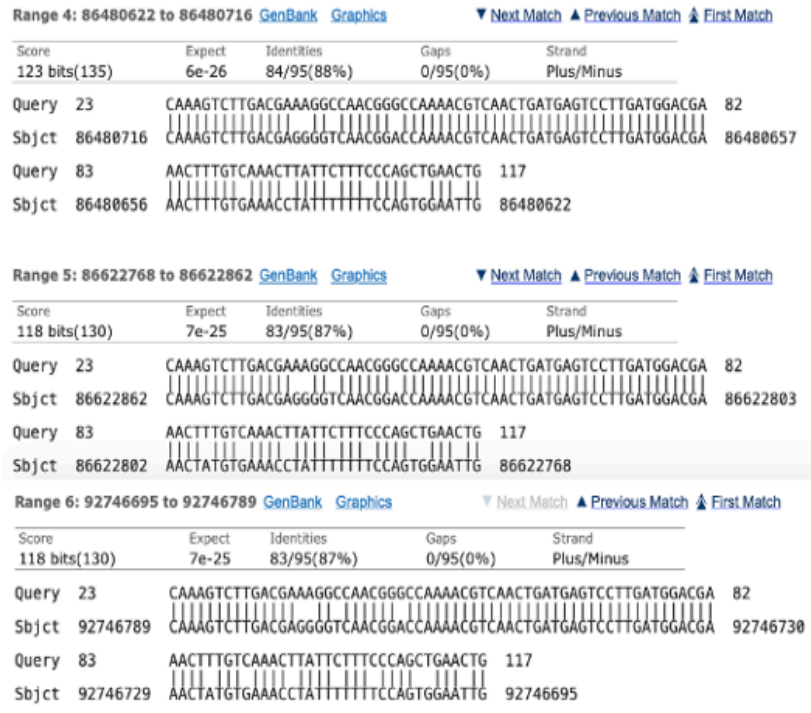
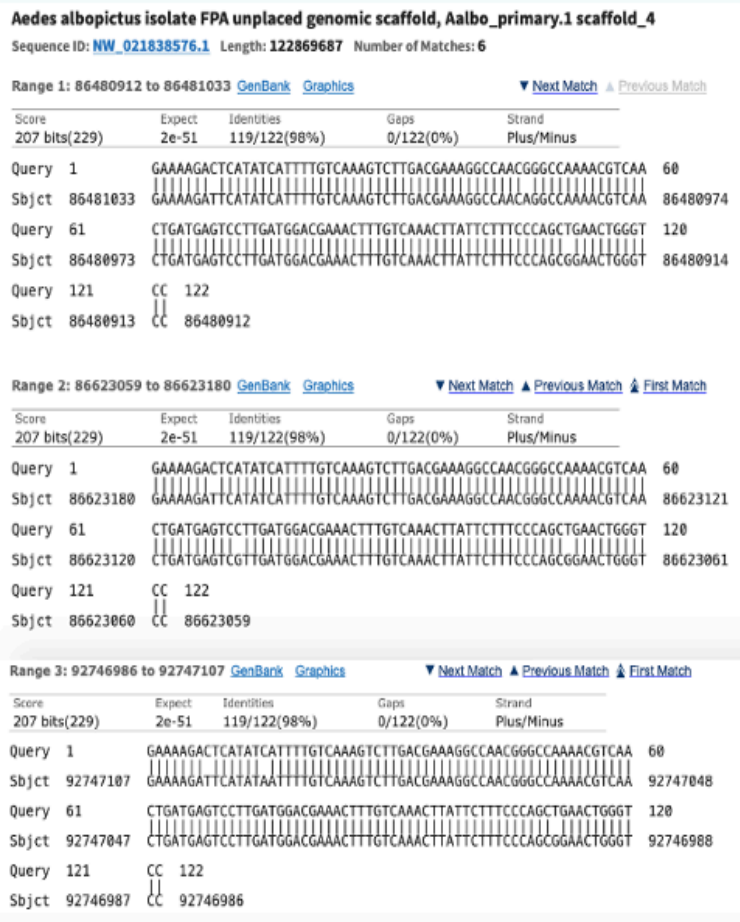


Figura 30. Análisis en BLAST de *A. albopictus*. Representación de la búsqueda en BLAST de los tres pares de copias de la secuencia del genoma de *A. albopictus*.

Una vez observadas estas secuencias se accede a Graphics y tras esto se encuentra que aparecen en el intrón del gen LASP (dos copias normales del gen y otro aparentemente parece ser un pseudogen) (Figura 32).

Aedes albopictus isolate FPA unplaced genomic scaffold, Aalbo_primary.1 scaffold_4, whole genome shotgun sequence

NCBI Reference Sequence: NW_021838576.1

[GenBank](#) [FASTA](#)

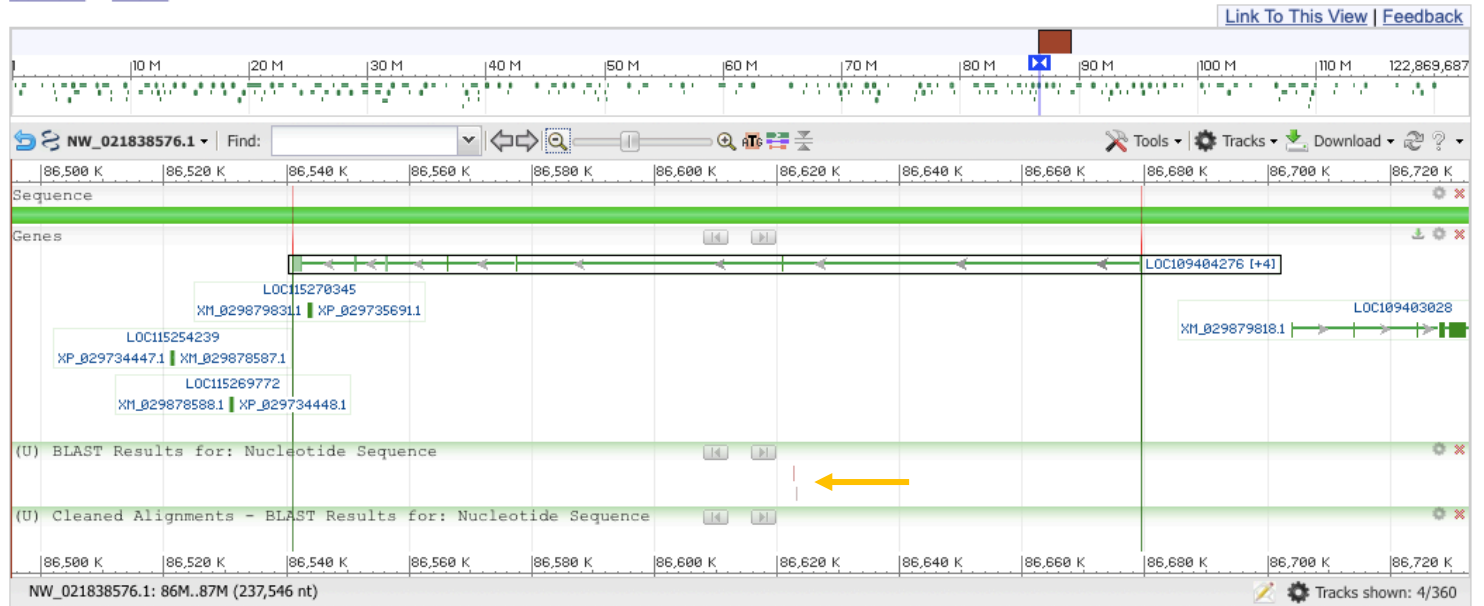


Figura 31. Intrón del gen LASP en *A. albopictus*. La flecha naranja indica la búsqueda realizada (query).

Continuando con el análisis, se observa que en realidad entre las dos parejas de HHRs había una tercera en la polaridad opuesta.

A. albopictus (296 nt, polaridad positiva)



A. albopictus (296 nt, polaridad negativa)

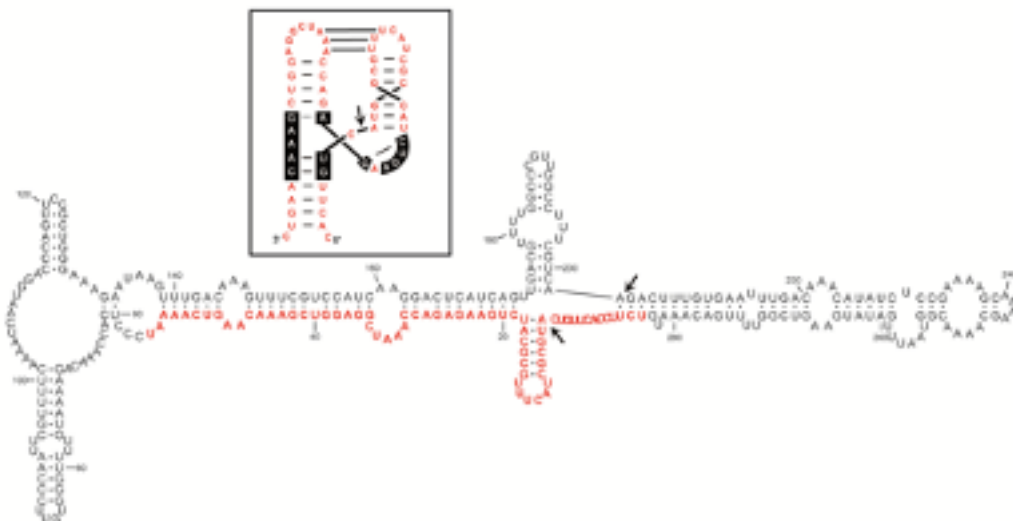


Figura 32. Ejemplo de las ribozimas HHR tipo III presentes en el intrón de LASP.

4. Conclusiones

En este estudio hemos descrito la presencia de varias clases de ribozimas de autocorte (HHR y HDVR), tanto en novedosos RNAs circulares tipo HDV detectados en animales, como en diversos genomas de DNA. Para ello se siguieron diversas aproximaciones bioinformáticas que fueron desde programas de alineamiento de secuencias primarias, búsquedas por homología de secuencia en bases de datos (BLAST) o búsquedas de estructuras secundarias predefinidas (RNAmotif).

Los resultados obtenidos nos indicaron que nuevas ribozimas HDVR y HHR son motivos que están codificadas en genomas de circRNAs tipo hepatitis delta de vertebrados (tritón y pez) e invertebrados (termitas), así como en los propios genomas de DNA de diversos invertebrados (termitas y mosquitos). Ya se ha descrito previamente que muchos de estos motivos genómicos suelen estar implicados en diversas familias de retroelementos, que en algunos casos se expresan como abundantes RNAs lineales y también circulares. Podemos asumir que estas poblaciones de RNAs conteniendo diversos ribozimas podrían ser una fuente para la aparición de nuevos agentes infecciosos de circRNA, ya sea como satélites virales (a través de la encapsidación del virus del RNA que contiene ribozimas durante la infección celular) o como agentes autónomos (a través de eventos de duplicación de ribozimas de una polaridad a la otra). La presencia de HHRs relacionadas filogenéticamente tanto en agentes similares al HDV de termitas como en los propios genomas de termitas, respaldarían nuestra hipótesis, aunque no se puede descartar la posibilidad opuesta (que los circRNA infecciosos con ribozimas fueran la fuente de motivos genómicos de autocorte). Por otro lado, el agente HDV humano y los nuevos agentes tipo HDV de amniotas (reptiles, aves y mamíferos), peces y tritones contienen todos la clásica ribozima pequeña de HDVR. Todos estos pequeños motivos de ribozimas (HHR, HDVR, HPR, etc) no muestran ninguna secuencia o relación estructural entre ellos, y por esa razón, la presencia de HHRs de tipo III conservadas en los agentes de tipo HDV de especies de sapo y termita ha sido un resultado sorprendente e inesperado. Sin embargo, aunque estas HHR en agentes tipo HDV pueden considerarse motivos análogos a las HHR de tipo III de agentes subvirales de

plantas (viroides y satélites virales) y de retrozimas, su topología única y sus interacciones terciarias tan particulares (un pseudonudo con interacciones de tipo Watson-Crick) no sugieren relación evolutiva entre ellos. Lo más probable es que estas HHRs excepcionales en circRNAs tipo HDV representen un caso de convergencia con las HHRs encontradas en circRNAs de plantas.

Los datos nos indican que, desde un punto de vista evolutivo, tanto HHRs como HDVRs proporcionarían la función de autocorte esencial para la replicación por el mecanismo de círculo rodante de los circRNAs tipo HDV, y se podrían considerar como motivos intercambiables entre miembros de esta familia. Trabajos anteriores indican que el agente HDV humano se puede circularizar en las células hospedadoras después de ser procesado por diversos motivos de autocorte (ya sean HDVRs o HHRs) con tal de que sean capaces de producir extremos 5'-OH y 2'-3'-fosfato cíclico (86).

Debe señalarse, sin embargo, que tanto las ribozimas tipo HDVR como las HHR se encuentran en regiones genómicas muy similares de los circRNA tipo HDV (Fig. 26). Esta disposición conservada, junto con la homología de secuencia entre las ribozimas genómicas y antigenómicas para cada genoma tipo HDV, indica que la actividad de autocorte estaría regulada por las estructuras que compiten en 3', tal y como se propuso previamente para el agente HDV humano (87). Los circRNA similares a HDV recientemente descubiertos en especies de vertebrados e invertebrados han abierto una interesante puerta con respecto a si estos nuevos elementos son satélites de otros virus, son novedosos agentes similares a viroides o simplemente elementos móviles originados a partir de genomas celulares. La investigación futura nos ayudará a comprender la biología y las relaciones evolutivas de esta familia única de agentes infecciosos mínimos de RNA circular.

Además por otro lado, en este trabajo también se caracterizaron las ribozimas HHRs en las especies de *Aedes*, que en un principio, y basado en lo encontrado en termitas, se buscaron con la topología tipo III. Los scripts desarrollados fueron para buscar intencionadamente este tipo de ribozima, pero finalmente tras varias comparativas se pudo concluir que en realidad estos motivos correspondían a algunas de las numerosas HHRs de tipo I presentes en estos genomas. En el trabajo realizado por Cervera y De la Peña en 2014 (37), ya se había descrito la existencias de multitud de ribozimas HHR tipo I en el genoma de *A. aegypti*

relacionadas con retrotransposones Penélope, y en el presente trabajo se ha terminado de caracterizar dichas ribozimas en *A. albopictus* como se muestra en la figura 30.

Por último, las ribozimas presentes en los intrones de las copias del gen LASP de *A. albopictus* (dos copias normales del gen y otro que aparenta ser un pseudogen), sugieren que estas HHRs intrónicas están conservadas y cumplirían una función biológica similar a la propuesta para las HHRs intrónicas humanas (63). Para confirmar que esta ribozima es 100% funcional, en un futuro estudio se analizará en el laboratorio para ver confirmar este hecho o si simplemente es un fósil que aun queda en este genoma y que terminará perdiéndose a lo largo de la evolución.

5. Glosario

- RNA: acrónimo del inglés *ribonucleic acid*, en castellano ácido ribonucleico. Es un ácido nucleico formado por una cadena de ribonucleótidos
- circRNA: acrónimo del inglés *circular ribonucleic acid*, en castellano ácido ribonucleico circular. Es un tipo de RNA monocatenario que a diferencia del RNA lineal, forma un bucle continuo cerrado covalentemente
- RNAm: acrónimo del inglés *messenger ribonucleic acid*, en castellano ácido ribonucleico mensajero. Se trata de una molécula de RNA monocatenaria que corresponde a la secuencia genética de un gen y es leída por un ribosoma en el proceso de síntesis de una proteína.
- DNA: acrónimo del inglés *deoxyribonucleic acid*, en castellano ácido desoxirribonucleico. Es una molécula compuesta por dos cadenas de polinucleótidos que se enrollan entre sí para formar una doble hélice la cual porta información genética esencial.
- ssRNA: acrónimo del inglés *single strand ribonucleic acid*, en castellano ácido ribonucleico de cadena simple. Es un virus que porta RNA de cadena sencilla de sentido positivo como material genético y es capaz de infectar a las células con producción de nuevos virus. Esto se debe a que el RNA del virus tiene la misma secuencia de bases que los RNAm.
- HDV: acrónimo del inglés, *Hepatitis Delta Virus*, en castellano Virus de la Hepatitis Delta. El agente viral productor de este proceso requiere de la membrana fabricada por el VHB (virus de la Hepatitis B). En consecuencia, la hepatitis B tiene una mayor virulencia.
- satRNA: acrónimo del inglés, *satellite ribonucleic acid*, en castellano ácido ribonucleico satélite. Se trata de genomas de RNA (500-200 nt) que necesitan de otros virus para multiplicarse, denominado "helper".

- SaPIs: acrónimo del inglés, *Staphylococcus aureus* pathogenicity islands, en castellano islas de patogenicidad de *Staphylococcus aureus*. *S. aureus* presenta en su DNA islas de patogenicidad, las cuales se definen como una fracción de DNA que lo faculta como virulento y está contenido en plásmidos.
- HBV: acrónimo del inglés, *hepatitis B virus*, en castellano Virus de la hepatitis B. El virus de la hepatitis B es un virus pequeño de la familia Hepadnaviridae, causante de la hepatitis B.
- nt: nucleótidos. Moléculas orgánicas formadas por la unión covalente de un nucleósido (una pentosa y una base nitrogenada) y un grupo fosfato.
- dsRNA: acrónimo del inglés, *double strand ribonucleic acid*, ácido ribonucleico de cadena doble. Es un virus RNA bicatenario (genoma de doble cadena).
- PSTVd: acrónimo del inglés, *potato spindle tuber viroid*, en castellano viroide del tubérculo fusiforme de la patata.
- CCR: región central conservada.
- TCR: región terminal conservada.
- TCH: horquilla terminal conservada.
- ASBVd: acrónimo del inglés, *Avocado Sunblotch Viroid*, viroide del manchado solar del aguacate.
- TE: elemento transponible. es una secuencia de DNA que puede moverse de manera autosuficiente a diferentes partes del genoma de una célula insertándose en el lugar deseado.
- LTR: acrónimo de inglés, *Long terminal repeat*, en castellano repetición terminal largas. Se trata de una secuencia de nucleótidos característica que se encuentra en cada extremo del DNA o RNA.
- LINEs: acrónimo del inglés, *Long interspersed nuclear elements*, en castellano elementos intercalados largos. Se trata de secuencias de DNA que se “copian y pegan” en otra zona del genoma. Son autónomos y codifican su propia maquinaria de reversotranscripción.

- SINEs: acrónimo del inglés, *Short interspersed nuclear elements*, en castellano elementos intercalados cortos. Se trata de retroelementos no autónomos.
- Pol II: RNA polimerasa II. Es una enzima de eucariotas que cataliza la transcripción del DNA a precursores de RNAm, microRNAs y otros tipos de ácido ribonucleico.
- Bp: pares de bases. Es una unidad que consta de dos nucleobases unidas entre sí por enlaces de hidrógeno
- ORF: acrónimo del inglés, *open reading frame*, marcos abiertos de lectura. Se trata de cualquier región de DNA o RNA comprendida entre un codón de inicio de la traducción y un codón de parada.
- RT: acrónimo del inglés, *reverse transcriptase*, en castellano reverso transcriptasa. Es una enzima de tipo DNA polimerasa que tiene como función sintetizar DNA de doble cadena utilizando como molde RNA monocatenario
- PLEs: acrónimo del inglés, *Penelope-like elements*, en castellano elementos Penélope. Se trata de retroelementos presentes en eucariotas.
- HHR: acrónimo del inglés, *hammerhead ribozyme*, en castellano ribozima cabeza de martillo.
- VS: virus satélite Varkud.
- GlmS: proteína glucosamina-6-fosfato sintetasa.
- MFE: estructura mínima de energía libre.
- EN: endonucleasa. Enzimas que catalizan la ruptura de enlaces fosfodiéster en diferentes regiones ubicadas en el interior de una cadena polinucleotídica
- HDVR: acrónimo del inglés, *Hepatitis Delta Virus Ribozyme*, en castellano Virus de la Hepatitis Delta Ribozima.

6. Bibliografía.

1. Lasda E, Parker R. Circular RNAs: diversity of form and function. *RNA* 2014 Dec;20(12):1829-1842.
2. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 2013 Feb;19(2):141-157.
3. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell* 2014 Sep 25;159(1):134-147.
4. D Liang D, Wilusz JE. Short intronic repeat sequences facilitate circular RNA production. *Genes Dev* 2014 Oct 15;28(20):2233-2247.
5. Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* 2014 Oct 2;56(1):55-66.
6. Conn SJ, Pillman KA, Toubia J, Conn VM, Salmanidis M, Phillips CA, et al. The RNA binding protein quaking regulates formation of circRNAs. *Cell* 2015 Mar 12;160(6):1125-1134.
7. Zheng Y. Computational non-coding RNA biology . United Kingdom: Elsevier; 2018. Disponible en: <https://www.elsevier.com/books/computational-non-coding-rna-biology/zheng/978-0-12-814365-0>
8. Ding B. Viroids: self-replicating, mobile, and fast-evolving noncoding regulatory RNAs. *Wiley Interdiscip Rev RNA* 2010 Nov-Dec;1(3):362-375.
9. Flores R, Grubb D, Elleuch A, Nohales MA, Delgado S, Gago S. Rolling-circle replication of viroids, viroid-like satellite RNAs and hepatitis delta virus: variations on a theme. *RNA Biol* 2011 Mar-Apr;8(2):200-206.
10. Navarro B, Gisel A, Rodio ME, Delgado S, Flores R, Di Serio F. Viroids: how to infect a host and cause disease without encoding proteins. *Biochimie* 2012 Jul;94(7):1474-1480.
11. Flores R, Gago-Zachert S, Serra P, Sanjuan R, Elena SF. Viroids: survivors from the RNA world? *Annu Rev Microbiol* 2014;68:395-414.

12. Hu, C.C.; Hsu, Y.H.; Lin, N.S. Satellite RNAs and satellite viruses of plants. *Viruses* 2009 Oct 28, 1(3), 1325-1350.
13. Palukaitis P. What has been happening with viroids? *Virus Genes* 2014 Oct;49(2):175-184.
14. Simon A. Roossinck MJ; Havelda Z. Plant virus satellite and detective interfering RNAs: new paradigms for a new century. *Annu Rev Phytopathol* 2004 Sep;42:415-37.
15. Taliansky ME, Robinson DJ, Murrant AF. Groundnut rosette disease virus complex: biology and molecular biology. *Adv Virus Res* 2000;55:357-400. 14536-14541.
16. Rao AL, Kalantidis K. Virus-associated small satellite RNAs and viroids display similarities in their replication strategies. *Virology* 2015 May;479-480:627-636.
17. Huang YW, Hu CC, Lin NS, Hsu YH. Mimicry of molecular pretenders: the terminal structures of satellites associated with plant RNA viruses. *RNA Biol* 2010 Mar-Apr;7(2):162-171.
18. Nawaz-ul-Rehman MS, Fauquet CM. Evolution of geminiviruses and their satellites. *FEBS Lett* 2009 Jun 18;583(12):1825-1832.
19. Simon AE, Roossinck MJ, Havelda Z. Plant virus satellite and defective interfering RNAs: new paradigms for a new century. *Annu Rev Phytopathol* 2004;42:415-437.
20. Christie GE, Dokland T. Pirates of the Caudovirales. *Virology* 2012 Dec 20;434(2):210-221.
21. Ram G, Chen J, Kumar K, Ross HF, Ubeda C, Damle PK, et al. Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proc Natl Acad Sci U S A* 2012 Oct 2;109(40):16300-16305.
22. Ram G, Chen J, Ross HF, Novick RP. Precisely modulated pathogenicity island interference with late phage gene transcription. *Proc Natl Acad Sci U S A* 2014 Oct 7;111(40).
23. Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature* 1999 Jun 10;399(6736):541-548.
24. Alves C, Branco C, Cunha C. Hepatitis delta virus: a peculiar virus. *Adv Virol* 2013;2013:560105.

25. Diener TO. Potato spindle tuber virus: a plant virus with properties of a free nucleic acid. 3. Subcellular location of PSTV-RNA and the question of whether virions exist in extracts or in situ. *Virology* 1971 Jan;43(1):75-89.
26. Hadidi A, Flores R, Randles J.W., Palukaitis P. *Viroids and Satellites*. United Kingdom: Elsevier; 2017. Disponible en: <https://www.elsevier.com/books/viroids-and-satellites/hadidi/978-0-12-801498-1>
27. Diener TO. The viroid: biological oddity or evolutionary fossil? *Adv Virus Res* 2001;57:137-184.
28. Diener TO. Potato spindle tuber viroid. 8. Correlation of infectivity with a UV-absorbing component and thermal denaturation properties of the RNA. *Virology* 1972 Nov;50(2):606-609.
29. Gross HJ, Domdey H, Lossow C, Jank P, Raba M, Alberty H, et al. Nucleotide sequence and secondary structure of potato spindle tuber viroid. *Nature* 1978 May 18;273(5659):203-208.
30. Flores R, Delgado S, Gas ME, Carbonell A, Molina D, Gago S, et al. Viroids: the minimal non-coding RNAs with autonomous replication. *FEBS Lett* 2004 Jun 1;567(1):42-48.
31. Fadda Z, Daros JA, Fagoaga C, Flores R, Duran-Vila N. Eggplant latent viroid, the candidate type species for a new genus within the family Avsunviroidae (hammerhead viroids). *J Virol* 2003 Jun;77(11):6528-6532.
32. Huang CR, Burns KH, Boeke JD. Active transposition in genomes. *Annu Rev Genet* 2012;46:651-675.
33. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 2007 Dec;8(12):973-982.
34. Kroutter EN, Belancio VP, Wagstaff BJ, Roy-Engel AM. The RNA polymerase dictates ORF1 requirement and timing of LINE and SINE retrotransposition. *PLoS Genet* 2009 Apr;5(4):e1000458.
35. Ostertag EM, Kazazian HH, Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 2001;35:501-538.
36. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 2003 Sep;35(1):41-48.

37. Cervera A, De la Pena M. Eukaryotic penelope-like retroelements encode hammerhead ribozyme motifs. *Mol Biol Evol* 2014 Nov;31(11):2941-2947.
38. Arkhipova IR, Yushenova IA, Rodriguez F. Giant Reverse Transcriptase-Encoding Transposable Elements at Telomeres. *Mol Biol Evol* 2017 Sep 1;34(9):2245-2257.
39. de la Pena M, Garcia-Robles I, Cervera A. The Hammerhead Ribozyme: A Long History for a Short RNA. *Molecules* 2017 Jan 4;22(1):10.3390/molecules22010078.
40. Cervera A, Urbina D, de la Pena M. Retrozymes are a unique family of non-autonomous retrotransposons with hammerhead ribozymes that propagate in plants through circular RNAs. *Genome Biol* 2016 Jun 23;17(1):135-016-1002-4.
41. Scott WG. Ribozymes. *Curr Opin Struct Biol* 2007 Jun;17(3):280-286.
42. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. *Science* 2000 Aug 11;289(5481):920-930.
43. Steitz TA, Moore PB. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci* 2003 Aug;28(8):411-418.
44. Steitz TA, Steitz JA. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* 1993 Jul 15;90(14):6498-6502.
45. Stahley MR, Strobel SA. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* 2005 Sep 2;309(5740):1587-1590.
46. Murray JB, Seyhan AA, Walter NG, Burke JM, Scott WG. The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chem Biol* 1998 Oct;5(10):587-595.
47. Ferre´-D'Amare AR and Scott WG. Small Self-cleaving Ribozymes. *Cold Spring Harb Perspect Biol* 2010 Oct; 2(10):a003574.
48. Prody GA, Bakos JT, Buzayan JM, Schneider IR, Bruening G. Autolytic processing of dimeric plant virus satellite RNA. *Science* 1986 Mar 28;231(4745):1577-1580.
49. Hutchins CJ, Rathjen PD, Forster AC, Symons RH. Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res* 1986 May 12;14(9):3627-3640.

50. Buzayan, J.M.; Gerlach, W.L.; Bruening, G. Non-enzymatic cleavage and ligation of RNAs complementary to a plant virus satellite RNA. *Nature* 1986, 323, 349–353.
51. Kuo MY, Sharmeen L, Dinter-Gottlieb G, Taylor J. Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *J Virol* 1988 Dec;62(12):4439-4444.
52. Saville BJ, Collins RA. A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. *Cell* 1990 May 18;61(4):685-696.
53. Winkler, W.C.; Nahvi, A.; Roth, A.; Collins, J.A.; Breaker, R.R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 2004 Mar 18; 428(6980):281-6
54. Roth A, Weinberg Z, Chen AG, Kim PB, Ames TD, Breaker RR. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat Chem Biol* 2014 Jan;10(1):56-60.
55. Thompson JE, Kutateladze TG, Schuster MC, Venegas FD, Messmore JM, Raines RT. Limits to Catalysis by Ribonuclease A. *Bioorg Chem* 1995 Dec;23(4):471-481.
56. Jimenez RM, Polanco JA, Luptak A. Chemistry and Biology of Self-Cleaving Ribozymes. *Trends Biochem Sci* 2015 Nov;40(11):648-661.
57. De la Pena M, Gago S, Flores R. Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. *EMBO J* 2003 Oct 15;22(20):5561-5570.
58. Khvorova A, Lescoute A, Westhof E, Jayasena SD. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat Struct Biol* 2003 Sep;10(9):708-712.
59. Martick M, Scott WG. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* 2006 Jul 28;126(2):309-320.
60. Martick M, Horan LH, Noller HF, Scott WG. A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature* 2008 Aug 14;454(7206):899-902.
61. Flores R, Hernandez C, de la Pena M, Vera A, Daros JA. Hammerhead ribozyme structure and function in plant RNA replication. *Methods Enzymol* 2001;341:540-552.

62. Westhof E. A tale in molecular recognition: the hammerhead ribozyme. *J Mol Recognit* 2007 Jan-Feb;20(1):1-3.
63. de la Pena M, Garcia-Robles I. Intronic hammerhead ribozymes are ultraconserved in the human genome. *EMBO Rep* 2010 Sep;11(9):711-716.
64. Cervera A, de la Pena M. Small circRNAs with self-cleaving ribozymes are highly expressed in diverse metazoan transcriptomes. *Nucleic Acids Res* 2020 May 21;48(9):5054-5064.
65. Lai MM. The molecular biology of hepatitis delta virus. *Annu Rev Biochem* 1995 Jul;64:259-86.
66. Taylor JM. Hepatitis delta virus. *Intervirology* 1999;42(2-3):173-178.
67. Wu HN, Lin YJ, Lin FP, Makino S, Chang MF, Lai MM. Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proc Natl Acad Sci U S A* 1989 Mar;86(6):1831-1835.
68. Lai MM. The molecular biology of hepatitis delta virus. *Annu Rev Biochem* 1995;64:259-286.
69. Been MD. HDV ribozymes. *Curr Top Microbiol Immunol* 2006;307:47-65.
70. Flores R, Ruiz-Ruiz S, Serra P. Viroids and hepatitis delta virus. *Semin Liver Dis* 2012 Aug;32(3):201-210.
71. Taylor JM. Virology of hepatitis D virus. *Semin Liver Dis* 2012 Aug;32(3):195-200.
72. Ferre-D'Amare AR, Zhou K, Doudna JA. Crystal structure of a hepatitis delta virus ribozyme. *Nature* 1998 Oct 8;395(6702):567-574.
73. Pereira MJ, Harris DA, Rueda D, Walter NG. Reaction pathway of the trans-acting hepatitis delta virus ribozyme: a conformational change accompanies catalysis. *Biochemistry* 2002 Jan 22;41(3):730-740.
74. Shih IH, Been MD. Catalytic strategies of the hepatitis delta virus ribozymes. *Annu Rev Biochem* 2002;71:887-917.
75. Salehi-Ashtiani K, Luptak A, Litovchick A, Szostak JW. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science* 2006 Sep 22;313(5794):1788-1792.
76. Webb CH, Luptak A. HDV-like self-cleaving ribozymes. *RNA Biol* 2011 Sep-Oct;8(5):719-727.

77. Webb CH, Riccitelli NJ, Ruminski DJ, Luptak A. Widespread occurrence of self-cleaving ribozymes. *Science* 2009 Nov 13;326(5955):953.
78. de la Pena M, Garcia-Robles I. Ubiquitous presence of the hammerhead ribozyme motif along the tree of life. *RNA* 2010 Oct;16(10):1943-1950.
79. Wille M, Netter HJ, Litteljohn M, Yuen L, Shi M, Eden JS, Klaassen M, Holmes EC, Hurt AC. A divergent hepatitis D-like agent in birds. *Viruses* 2018 Dec 17; 10(12): 720.
80. Hetzel U, Szivovicza L, Smura T, Prahauer B, Vapalahti O, Kipar A, et al. Identification of a Novel Deltavirus in Boa Constrictors. *mBio* 2019 Apr 2;10(2):10.1128/mBio.00014-19.
81. Chang WS, Pettersson JH, Le Lay C, Shi M, Lo N, Wille M, et al. Novel hepatitis D-like agents in vertebrates and invertebrates. *Virus Evol* 2019 Jul 15;5(2):vez021.
82. Paraskevopoulou S, Pirzer F, Goldmann N, Schmid J, Corman VM, Gottula LT, et al. Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent *Proechimys semispinosus*. *Proc Natl Acad Sci U S A* 2020 Jul 28;117(30):17977-17983.
83. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007 Nov 1;23(21):2947-2948.
84. Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res* 2001 Nov 15;29(22):4724-4735.
85. Paraskevopoulou S, Pirzer F, Goldmann N, Schmid J, Corman VM, Gottula LT, et al. Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent *Proechimys semispinosus*. *Proc Natl Acad Sci U S A* 2020 Jul 28;117(30):17977-17983.
86. Reid CE, Lazinski DW. A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs. *Proc Natl Acad Sci U S A* 2000 Jan 4;97(1):424-429.
87. Diegelman-Parente A, Bevilacqua PC. A mechanistic framework for co-transcriptional folding of the HDV genomic ribozyme in the presence of downstream sequence. *J Mol Biol* 2002 Nov 15;324(1):1-16.