



Universidad Nacional Autónoma de México

Facultad de Medicina

Instituto de Investigaciones Biomédicas

---

---

**“Identificación de RNA  
no codificante en el genoma de  
*Taenia solium*”**

**T E S I S**  
QUE PARA OBTENER EL TÍTULO DE  
LICENCIADO EN INVESTIGACIÓN BIOMÉDICA BÁSICA  
P R E S E N T A :

**Homero Garcés Flores**

Director de Tesis  
Dr. Marco Antonio José Valenzuela

Facultad de Medicina



Ciudad Universitaria  
México D.F. Febrero 2012



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

DEDICADO A

*...mi Madre...*

*...mi Padre...*

*...mi Hermana...*

*...mi Hermano...*

*...y al amor de mi vida...*

## AGRADECIMIENTOS

### *Dentro de la Universidad:*

A Marco José Valenzuela, por aceptarme en su grupo de Biología Teórica y darme la oportunidad de trabajar en varios proyectos de lo más interesantes;

A Tzipe Govezensky, por su incansable apoyo y discusión de aspectos estadísticos a lo largo de este y otros trabajos;

A Juan Román Bobadilla, por su asesoría y apoyo informático;

A Martha Cariño, por su apoyo en el acceso y localización de material bibliográfico a lo largo de la licenciatura;

Mención especial a Maria Petra Muñoz “Mary” de la biblioteca y Alicia, que siempre me dan palabras alentadoras para no dejar de perseverar en esta carrera científica.;

A mis compañeros y amigos de la carrera, desde la camada original con la que ingresé, hasta la que terminó por adoptarme: Miguel “ñoñoxcuintle”, Maritrini, Paco, Charis, a Eda que siempre me ha motivado a utilizar al máximo mi “ñoñez”, hasta mis carnalazos Octavio, David Fajardo y Hugo Cruces con quienes he crecido y vivido muy buenos momentos. También a “iGilberto “, David Heras, Leti y Agustín.

A los miembros del grupo de Biología Teórica con quienes he convivido de forma alegre y en los seminarios discutido ferozmente;

A mi queridísima casa de estudios por la formación recibida y abrirme las puertas a un universo intelectual y humano. En ella encontré personas valiosísimas cuya amistad y amor atesoro.

### *A mi familia:*

En especial con todo mi amor a mi madre, que libra una batalla contra esa enfermedad que nos hizo pasar momentos difíciles, pero que por fortuna la va superando y nos deja largo camino por recorrer juntos.

A mi padre por estar siempre a mi lado y apoyarme siempre que lo he necesitado. Por su buen corazón hacia las criaturas necesitadas de ayuda.

A mi hermana Lissete, a quien ¡jamás termino de comprender! Y que alguna vez sentí temor de perderla, por lo que agradezco a la vida que esté acompañándome en estos momentos tan importantes.

A mi hermano Toño por su preocupación de mí persona y sus oraciones en mi ayuda. Va para mis sobrinos Toño y Luis, así como mi cuñada Lucero.

A Kary, por el amor que me infunde a diario y lo hermoso que es vivir juntos. Su apoyo y comprensión me han alentado a perseguir metas y anhelos en todos los ámbitos de mi vida. También por hacerme ver sin concesión mis errores y al mismo tiempo motivar mi autoconfianza y potenciar mis virtudes.

En extensión a mi familia adoptiva, Susana Sol y Enrique Arredondo, así como Alan y Enrique con sus respectivas familias, con la esperanza de que algún día sus “chamacosis agudas” consigan redactar sus propios agradecimientos.

A mi tío Martín, quien siempre ha mostrado un interés sincero en mi progreso y las ideas que tengo, así como al tío Gabriel por sus anotaciones siempre pertinentes y con ese toque tan ácido que les imprime.

De igual forma agradezco a mis cuates y amigos “foráneos” de la carrera Mario y Nidia. A Ibrahim y Edrey Romero, así como Eric Aguilar por seguir juntos en este camino compartir esta dualidad de profesionista/músico.

Y por último pero no menos importante, a mis compañeros en esta otra mitad de mi vida que es la música, esa sangre “exosomática” que tanto llena al ser humano. Especialmente a Fabián Jacome que torció mi existencia al enseñarme a tocar la guitarra. A Bruno, Fernando, Juan Carlos y Rubber Beto, Vicente “Chentufas”, Oskar “Fantasma”, con quienes he roqueado ya en tres bandas y hemos dado vida a algunas de mis composiciones. Todos ellos han preguntado: ¿para cuando la tesis?, bueno... aquí está y ahora se los agradezco.

Mil disculpas a quienes no mencioné y que están pendientes de mí. No es falta de gratitud, sino de espacio y tiempo. Gracias a todos.

# ÍNDICE

I. INTRODUCCIÓN .....	1
I. I. TAENIA SOLIUM .....	1
I. I. I. Historia de Taenia solium. ....	1
I. I. II. Morfología del Parásito .....	4
I. I. II. I. Huevecillo .....	4
I. I. II. II. Larva (Cisticerco) .....	5
I. I. II. III. Adulto .....	6
I. I. III. Ciclo de Vida .....	8
I. I. IV. Enfermedades causadas por la Taenia solium .....	9
I. I. IV. I. Teniasis .....	9
I. I. IV. II. Cisticercosis .....	9
I. I. IV. III. Neurocisticercosis .....	11
I. I. IV. IV. Diagnóstico .....	12
I. I. IV. V. Tratamiento .....	13
I. I. V. Estadísticas .....	15
I. I. VI. Evolución .....	16
I. I. VII. Secuenciación del genoma de Taenia solium .....	18
I. I. VII. I. Características del genoma de Taenia solium .....	18
I. I. VIII. Parásitos y carcinogénesis .....	19
I. II. RNA NO-CODIFICANTE (ncRNA) .....	21
I. II. I. El RNA salta a la escena .....	21
I. II. II. RNA no codificante .....	22
I. II. III. Clasificación de los RNA no codificantes .....	23
I. II. III. I. RNA Pequeño Nucleolar (snoRNA, small nucleolar RNA) .....	24
I. II. III. II. siRNA .....	25
I. II. III. III. MicroRNA (miRNA) .....	26
I. II. III. IV. ncRNA largo (lncRNA) .....	28
I. II. III. V. RNA de interacción con proteínas Piwi (piRNA) .....	29
I. II. III. VI. Riboswitches .....	30
I. II. IV. ncRNA y enfermedad .....	31
I. II. V. ncRNA y complejidad biológica .....	32
I. II. VI. Análisis computacional de los ncRNA .....	35
I. II. VII. Bases de datos de ncRNA .....	36
II. PLANTEAMIENTO DEL PROBLEMA .....	39
III. OBJETIVOS .....	39
IV. JUSTIFICACIÓN .....	40
V. MATERIAL Y MÉTODOS .....	41
VI. RESULTADOS .....	46
NONCODE .....	52
NONCODE vs EST Larva .....	52
NONCODE vs EST adulto .....	54
NONCODE vs EST Japonés (Full-length) .....	56
NONCODE vs 454AllContigs2010 .....	58
HAIRPINS .....	62
Hairpins vs EST Larva .....	62
Hairpins vs EST Adulto .....	63
Hairpins vs EST Japón (full-length) .....	64
Hairpins vs 454AllContigs2010 .....	65
VII. DISCUSIÓN .....	67
VIII. CONCLUSIONES .....	70
IX. PERSPECTIVAS .....	71
X. BIBLIOGRAFÍA .....	72
XI. APÉNDICE A .....	77
XII. APÉNDICE B .....	81

## RESUMEN

Descrita por Carlos Linneo en el siglo XVIII, y a pesar de haber sido erradicada en Europa hace más de un siglo, la *Taenia solium* sigue aquejando al género humano, infligiendo enfermedad y muerte, principalmente en países en desarrollo de Latinoamérica, África y el sureste asiático. Afortunadamente, científicos mexicanos han hecho grandes aportaciones para el conocimiento del parásito y la enfermedad que ocasiona. Entre ellas, se han descrito los tratamientos antiparasitarios más ampliamente usados a nivel mundial, así como algunos de los métodos de diagnóstico más sensibles y específicos, así como varias versiones de vacunas en contra del cisticerco de la *Taenia solium*.

En la UNAM se emprendió un proyecto dedicado a obtener la secuencia del genoma de *Taenia solium*, involucrando para ello a varios investigadores de distintas dependencias. El conjunto de secuencias generadas hasta ahora son analizadas con el fin de encontrar genes de interés y en su caso, únicos a *Taenia solium*, además del parecido con otras especies que pueda ayudar a dilucidar con mejor detalle el proceso infeccioso por este organismo.

Los análisis hechos hasta ahora por el consorcio del genoma de *T. solium* se han enfocado a proteínas, por lo que decidimos complementar el estudio al buscar secuencias de RNA no codificante (*ncRNA* en inglés). Esta clase de RNA, relativamente nueva, se ha posicionado rápidamente como elemento regulador de varios procesos biológicos y se ha convertido un blanco de estudio importante en distintas enfermedades.

Este trabajo presenta los resultados que obtuvimos de buscar secuencias de ncRNA dentro del genoma de *Taenia solium* mediante la herramienta computacional *BLASTn* como estudio previo a análisis experimentales que puedan llevarse a cabo en este organismo enfocados a ncRNA. Se trabajó con secuencias proporcionadas por el consorcio del genoma de *Taenia solium* -EST's (cDNA) de larva y adulto así como borradores del genoma provenientes de ensayos de pirosecuenciación- contra dos bases de datos especializadas en secuencias de ncRNA; NONCODE y miRBase. Obtuvimos alineamientos con puntuaciones significativas, además de corroborar hallazgos previos de secuencias de ncRNA en *Taenia solium* (*U5/U6*). Tanto las secuencias de ncRNA

como el contig de *Taenia* que presentaron similitudes fueron comparadas contra las bases Gene Ontology y NCBI nr/nt para obtener mayor información de ellas y ver que nuevos datos sobre este organismo podríamos obtener.



# **I. INTRODUCCIÓN**

## **I. I *Taenia Solium***

La *Taenia solium* es un parásito platelminto cuya larva (cisticerco) puede invadir varios tejidos en el humano, ocasionando la enfermedad llamada **Cisticercosis** (CC), mientras que la forma adulta conocida comúnmente como “*solitaria*”, se fija en la pared del intestino, padecimiento llamado **teniasis**<sup>1</sup>. Se encuentra asociada normalmente a países en desarrollo, en alto grado de pobreza, pero actualmente está emergiendo nuevamente en países desarrollados debido a la inmigración. Tiene un alto costo económico por los tratamientos médicos y por las pérdidas en la producción de alimento al sacrificarse los animales infectados, acentuando el rezago de los países afectados por este organismo [1]. A continuación narramos algo de su historia.

### ***I. I. I. Historia de Taenia solium***<sup>2</sup>.

Informes acerca de *Taenia solium* se conocen desde la antigüedad. Los griegos Hipócrates, Aristóteles y Teofrasto llamaron platelmintos a los gusanos responsables, por su parecido con cintas o listones, mientras que los romanos Celso, Plinio el Viejo y Galeno los llamaron *lumbricus latus*, que quiere decir “gusano ancho”. Los griegos sabían ya de la existencia de la cisticercosis porcina, aunque ignoraban la relación de ésta con la teniasis humana. La medicina árabe, con Serapión a la cabeza, creía que cada proglótido era un gusano diferente. Los musulmanes le impusieron el nombre de “cucurbitineos”, no sólo por su parecido con las semillas de calabaza, sino también porque éstas constituyeron uno de los remedios más antiguos contra la teniasis, todavía en uso.

Se atribuye a Arnau de Vilanova, a comienzos del siglo XIV, la primera descripción de la especie. Varios siglos pasaron antes de tener completa la anatomía de *T. solium*. Carlos Linneo incluyó la especie *Taenia solium* en la décima edición de su *Systema Naturae* en 1758 [2].

---

<sup>1</sup> Se creía que sólo existía un parásito adulto por individuo, llevando una vida “solitaria”.

<sup>2</sup> La mayor parte de la información de esta sección proviene de [1] y [2]

La primera referencia de un caso de cisticercosis humana la dio Johannes Udalric Rumler en 1558, quien la atribuyó a un tumor en la duramadre de un epiléptico. Un siglo después, Domenico Panarolus en 1652 observó quistes parecidos en el cuerpo calloso del cerebro de otro epiléptico. En 1683, Tyson descubrió y describió la cabeza de las tenias y Redi publicó ilustraciones del escólex de las tenias de perros y gatos. Fue en 1697 cuando Marcello Malpighi descubrió el origen animal de estos quistes y describió también el escólex [1, 2].

En 1784, Johann August Ephraim Goeze, examinando los cisticercos de puerco, descubrió su naturaleza helmíntica, y dos años después, P.C. Werner redescubrió la cisticercosis humana en la autopsia de un soldado al encontrar dos quistes en el músculo pectoral que se asemejaban a los observados en la cisticercosis porcina. Es así como a finales del siglo XVIII se conocían ya plenamente la teniasis y la cisticercosis, pero se desconocía la relación entre ambos al ignorarse el ciclo biológico del parásito.

Goeze en 1784 y Dujardin en 1845 notaron similitudes en la forma del escólex del gusano adulto con el del cisticerco y sospecharon una conexión entre ambos. En 1835, Karl Theodor Ernst Von Siebold descubrió que los huevecillos de *Taenia* contienen una oncósfera, una cavidad que encierra un embrión con seis ganchos pequeños [2].

El desarrollo de cisticercos en cerdos quedó demostrado en 1853, cuando Pierre-Joseph Van Beneden alimentó a un cerdo con huevecillos de *Taenia solium*, y encontró cisticercos durante la necropsia. Dos años después, Friedrich Kuchenmeister demostró que las tenias se desarrollan a partir de cisticercos, al introducir secretamente cisticercos en la dieta de un preso condenado a muerte. En el intestino del cadáver del reo observó las tenias adultas. En diciembre de 1854, Aloys Humbert se provocó una infección al ingerir trece cisticercos y en marzo del año siguiente comenzó a expulsar segmentos de *T. solium*. En 1933, K. Yoshino se propuso experimentar en sí mismo el curso de la infección ingiriendo cisticercos y describiendo durante dos años la expulsión de proglótidos [2].

Respecto de la neurocisticercosis, en 1934 se pudo establecer el tiempo de aparición de los síntomas neurológicos de la enfermedad. A un gran número de soldados británicos que habían estado en la India se les diagnosticó Cisticercosis y fueron sometidos a un estricto tratamiento médico. Las crisis convulsivas, principal

síntoma, tardaron en promedio dos años en manifestarse, aunque hubo casos en que la epilepsia tardó hasta 20 o más años en presentarse desde el regreso de las tropas. La sintomatología se asoció con la muerte de los cisticercos y se definieron dos aspectos importantes de la enfermedad: su duración y su permanencia en el individuo por varios años sin ocasionar síntomas neurológicos [1, 2].

De esta manera fue como el ciclo de vida de *T. solium* fue dilucidado poco a poco, así como la descripción clínica de las enfermedades causadas. Es de interés notar que en México el primer informe realizado acerca de cisticercosis humana lo dio Ignacio Gómez Izquierdo en 1901 al publicar la descripción del caso de una paciente cubana que murió en un asilo psiquiátrico, la autopsia mostraba múltiples cisticercos. En ese entonces el autor mismo se planteaba que no habría manera de diagnosticar esta enfermedad y tampoco de tratarla [1, 2]. Felizmente en la actualidad ya no es así, los métodos de diagnóstico son bastante confiables y los tratamientos muy eficaces, con la esperanza de mejorarlos todavía más.

Es sorprendente que la infección por *Taenia solium*, que se conoce desde hace siglos, y que fue erradicada en Europa por primera vez a comienzos del siglo XX mediante el desarrollo social y programas de inspección-confiscación de animales infectados, siga persistiendo, sobre todo en América Latina, África y Asia donde es un problema de salud pública y económico. De hecho, hay una reemergencia de la enfermedad en países desarrollados de Europa y en Estados Unidos, principalmente a través de los trabajadores migrantes infectados provenientes de regiones con una alta prevalencia de la enfermedad [2].

Para terminar, podemos decir que mucho del esfuerzo científico reciente se ha concentrado en el mejoramiento del diagnóstico y de las terapias, así como un aumento en la comprensión de las variables geográficas y socioeconómicas que inciden en la transmisión de la enfermedad. Los factores inmunológicos y genéticos tanto del parásito como del hospedero que están presumiblemente relacionados a la historia natural, patología y prevención de la enfermedad también son objetos de amplio estudio. Otro aspecto muy importante y del que se ignora bastante tiene que ver con la naturaleza de las moléculas y los eventos microscópicos que conducen y determinan el establecimiento del cisticercos en ciertos órganos blanco y los mecanismos que subyacen la alta especificidad con el hospedero [3].

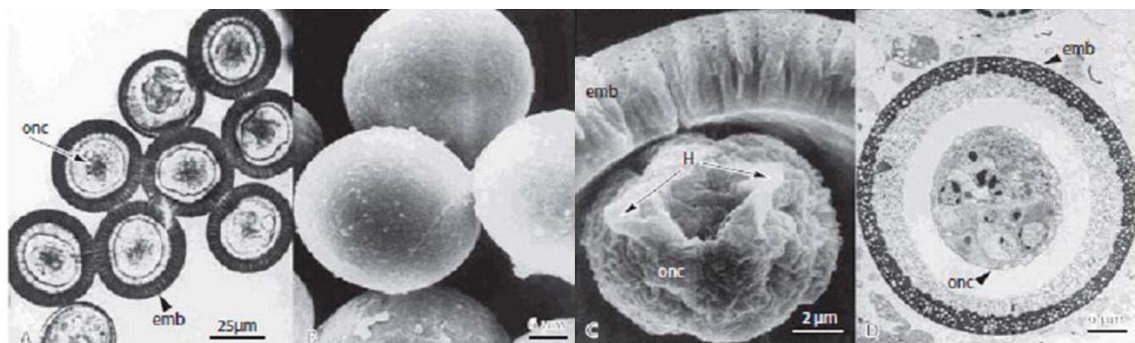
## ***I. I. II Morfología del Parásito***

### ***I. I. II. I. Huevecillo***

Invisibles al ojo humano, son esféricos y de tamaño microscópico, aproximadamente 30-45  $\mu\text{m}$  x 30-40  $\mu\text{m}$  (Fig. 1). Tienen una capa protectora denominada embrióforo alrededor de la oncósfera, que constituye la principal defensa del huevecillo contra el medio ambiente, y que está compuesta por bloques de una proteína similar a la queratina unidos por una sustancia cementosa. Esta capa tiene dos propiedades contrastantes, ya que al mismo tiempo que es muy protectora contra el medio circundante, es sensible a la acción de enzimas de origen digestivo. Es así como después de resistir tanto tiempo en el exterior, al ser ingerido por el hospedero intermediario o definitivo el embrióforo rápidamente pierde su propiedad rígida y libera la oncósfera que lleva en su interior [2].

La oncósfera contiene al embrión de *T. solium*, el cual tiene seis ganchillos que por acción de enzimas digestivas y sales biliares se activa y atraviesa la pared intestinal, entrando al torrente sanguíneo, desde donde alcanza y se desarrolla en el tejido subcutáneo, músculo esquelético, sistema nervioso central y ojos, convirtiéndose en la forma larvaria o cisticerco [2, 3].

El huevecillo es muy resistente al medio ambiente, pudiendo sobrevivir al aire libre hasta por ocho meses. Diariamente, un adulto infectado con la forma adulta de *Taenia* puede liberar cerca de 200 000 huevecillos.



**Figura 1.** Huevecillos de *Taenia Solium*.- a) Microscopía de luz de huevecillos frescos, se puede apreciar su tamaño, alrededor de 40 $\mu\text{m}$ ; b) Huevecillos vistos por microscopía electrónica de barrido; c) Microscopía electrónica de transmisión sobre corte de un huevecillo inmaduro, y otro corte de un huevecillo pero visto por microscopía electrónica de transmisión. Tomado de [4].

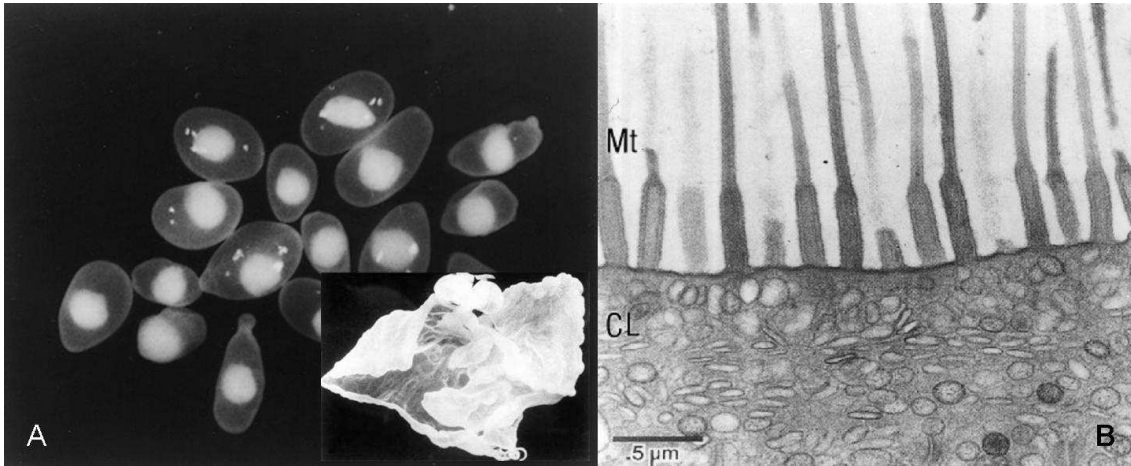
### I. I. II. II. Larva (*Cisticerco*)

Es una vesícula ovalada (en algunos casos racemosa), llena de líquido, que va de 0.5 a 2 cm de diámetro en su parte más larga y que contiene un pequeño escólex (cabeza del parásito) en su interior (Fig. 2 A). Al igual que la tenia adulta, el escólex del cisticerco posee cuatro ventosas y dos hileras de ganchillos [2, 3].

La superficie de contacto del parásito con los tejidos del hospedero es el tegumento, superficie desde la cual obtiene sus nutrientes y secreta sus desechos, y es central en el mantenimiento de la relación parásito-hospedero (Fig. 2 B). Consiste de una capa citoplásmica rellena de microvesículas y mitocondrias, su membrana exterior es aumentada por unos pliegues altamente organizados llamados microtricas, muy parecidas a las microvellosidades del intestino de vertebrados e invertebrados. La superficie de las microtricas va cubierta de glicocalix, mientras que la membrana interna del tegumento descansa sobre una membrana basal fibrosa. Debajo de la membrana basal están los citones tegumentales que conectan con el tegumento a través de extensiones citoplásmicas o procesos internunciales. Los citones tegumentales sintetizan proteínas y otros componentes superficiales que son exportadas al tegumento en la forma de vesículas que alcanzan el citoplasma distal vía procesos internunciales [2, 3].

El cisticerco puede presentar dos formas: racemosa o monovesicular (Fig. 2). La forma racemosa, que se observa normalmente en la neurocisticercosis, es grande, y contiene una vesícula multilobulada que asemeja a un racimo de uvas. Por su parte, la forma monovesicular o celulosa es pequeña, esférica u ovalada, con una vesícula translúcida a través de la cual se puede observar el escólex, esta forma es característica en el tejido muscular. Ambas formas pueden coexistir en pacientes con neurocisticercosis, y se pueden encontrar formas intermedias donde se ven principios de multilobulización [4].

Se creía que el cisticerco de *Taenia solium* era un organismo completamente distinto de la tenia, por lo que fue designado como *Cysticercus cellulosae* a principios del siglo XIX, sin embargo, esta clasificación dejó de tener validez al descubrirse que el cisticerco es la forma larvaria de *T. solium* [4].



**Figura 2.** Larva y tegumento con microtricas.- A) Larvas esféricas de *taenia solium* (cisticercos) que consisten de vesículas llenas de fluido y que albergan un escólex en su interior, el cual es morfológicamente igual que en la forma adulta. En el recuadro se muestra un cisticercos en forma racemosa. B) Micrografía electrónica de la superficie del tegumento de un cisticercos en el que se aprecian las microtricas (Mt) y la capa citoplásmica que les da soporte (CL). Tomado de [3, 4].

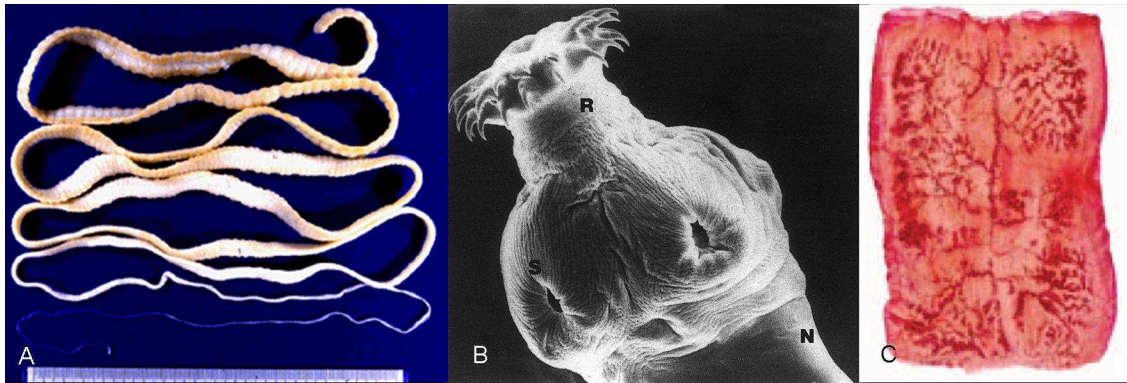
### I. I. II. III. Adulto

En esta etapa el parásito, la *tenia*, mejor conocida como “solitaria”, presenta un cuerpo largo y aplanado cuya longitud va de 2 a 7 metros y habita exclusivamente en el intestino humano (Fig. 3A). Al igual que el cisticercos, es el tegumento por donde absorbe los nutrientes del medio y secreta sus desechos, ya que carece de tubo digestivo [2, 3].

En uno de sus extremos presenta una estructura globular llamada escólex (Fig. 3B), del tamaño de una cabeza de alfiler y que posee una doble corona de ganchillos y cuatro ventosas, las cuales permiten que se fije en la pared del intestino delgado a pesar del continuo movimiento de contracción o peristalsis que hay en él. Los ganchillos son los que anclan el parásito y mediante las ventosas es como se adhiere a la pared intestinal.

A la mayor parte del cuerpo de la *tenia* se le denomina estróbilo, el cual está segmentado en estructuras llamadas proglótidos, que surgen de una región terminal situada en la parte inferior del escólex. Los proglótidos situados en la vecindad del escólex están inmaduros, carecen de órganos sexuales desarrollados, y conforme se van alejando del escólex hay una maduración progresiva de los mismos, siendo la mitad del estróbilo a partir de donde sólo hay proglótidos maduros [2, 3].

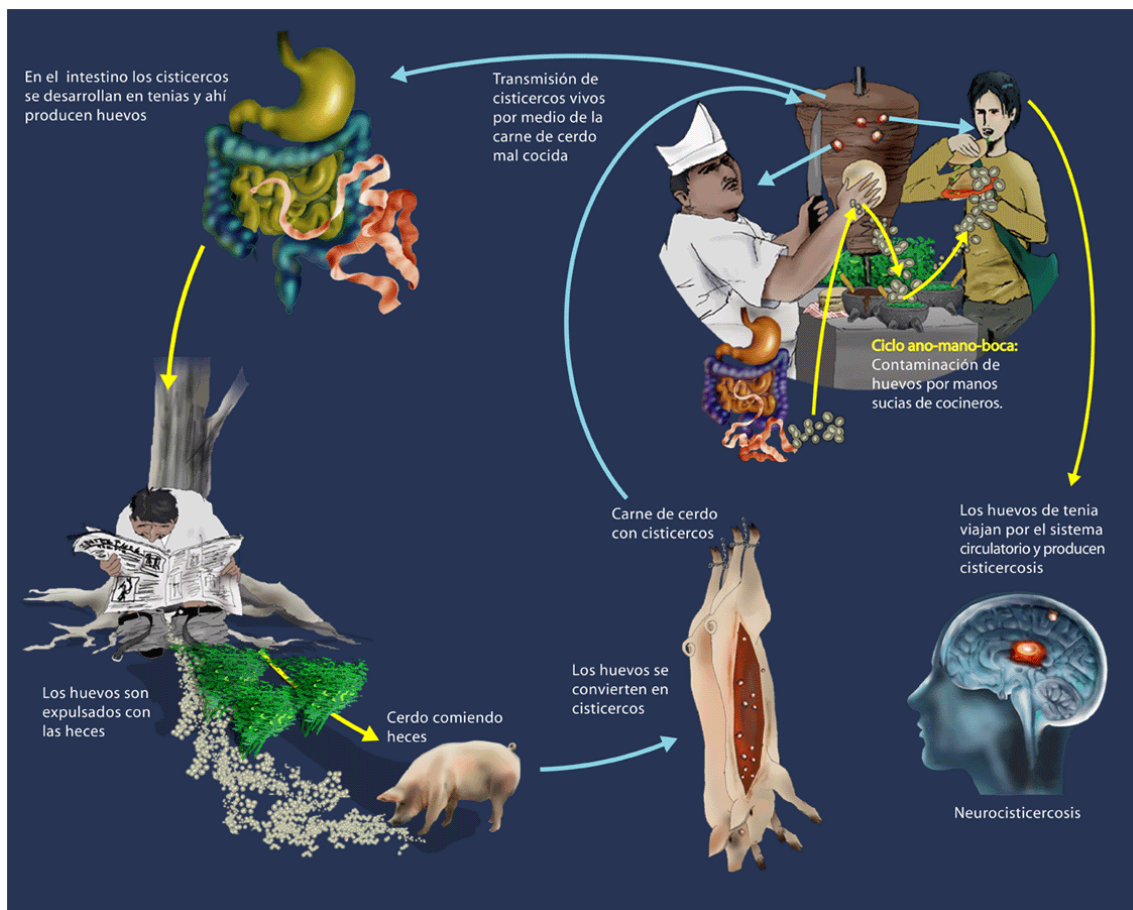
Cada proglótido maduro posee órganos sexuales masculinos y femeninos bien diferenciados, que hacen de la *tenia* un organismo hermafrodita. Los proglótidos más alejados del escólex muestran un útero “grávido” ramificado, lleno de huevecillos (alrededor de 50,000 en cada uno, Figura 3C). Cada proglótido mide entre 0.5 y 2 centímetros, y los proglótidos grávidos se desprenden de la *tenia* espontáneamente, siendo expulsados posteriormente hacia el medio exterior junto con las heces fecales del hospedero. Conforme van siendo expulsados, otros proglótidos van gestándose en el cuello de este organismo [2, 3].



**Figura 3.** *Taenia solium*, escólex y proglótido.- A) La forma adulta tiene una longitud aproximada de hasta 10 metros; B) Escólex de *Taenia solium* en el cual se aprecia el rostelo (R) que contiene dos hileras de ganchillos que le permiten anclarse a la pared intestinal del hospedero, las ventosas (S) que complementan la acción del rostelo, y el cuello (N) a partir del cual comenzará el estróbilo. C) Proglótido grávido en el que puede observarse el conducto uterino central que llega a tener 12 ramificaciones, en éstas se albergan ambos órganos genitales, masculino y femenino. Éste proglótido está repleto de huevos y listo a ser expulsado del intestino en la defecación. Tomado de; A: [www.islabit.com/wp-content/imagenes/solitaria.gif](http://www.islabit.com/wp-content/imagenes/solitaria.gif), B y C de [3] y [4], respectivamente.

### I. I. III. Ciclo de Vida

*Taenia solium* tiene al humano como hospedero definitivo y al puerco como intermediario (aunque se han encontrado perros y otros animales portando cisticercos de *Ts*). En lugares donde convivan cerdos con humanos en condiciones de poca higiene, una persona infectada con la *tenia* adulta en su intestino libera miles de huevecillos diariamente en sus heces. Éstos pueden ser ingeridos por el puerco, donde eclosionan a la forma larvaria y se fijan en distintos órganos (causando la cisticercosis porcina). Esta carne infectada puede ser consumida por el humano nuevamente, causando la teniasis y cerrando así el ciclo de vida de este parásito. Cuando se ingieren únicamente los huevecillos (por mala higiene o comer verduras contaminadas), se da la cisticercosis (Fig. 4) [2, 3].



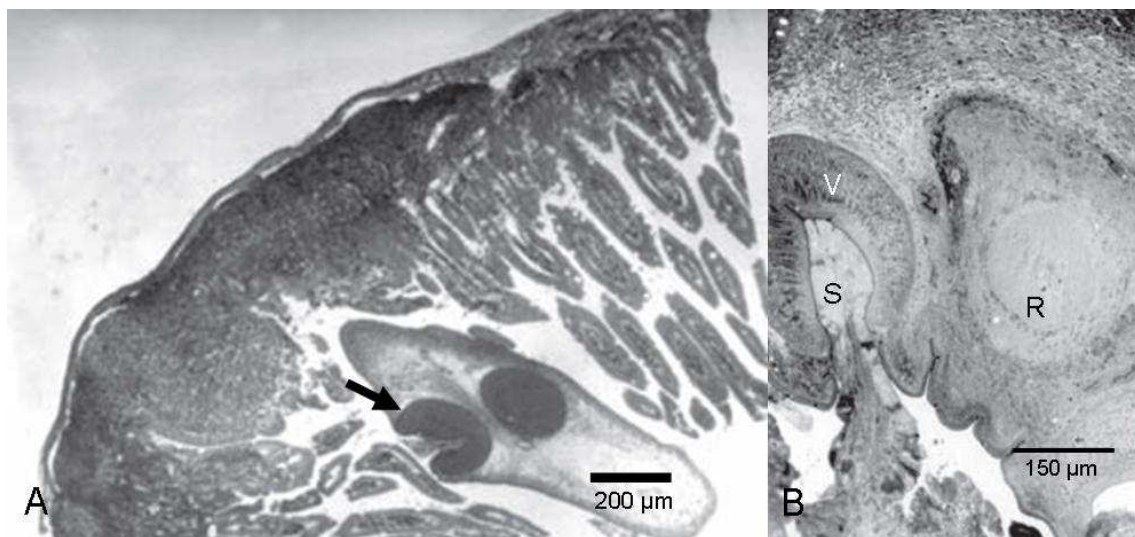
**Figura 4.** Ciclo de Vida de *Taenia solium*.- Los humanos se infectan por ingerir comida contaminada con huevecillos, o por autoinfección (un humano con *Ts* adulta puede consumir accidentalmente huevecillos expulsados en sus heces o que los proglótidos sean acarreados al estómago por peristalsis reversa). Una vez que se ingieren huevecillos, las oncosferas se activan en el intestino, invaden la pared intestinal, y migran hacia músculos estriados, así como cerebro, hígado y otros tejidos, donde se desarrollan en cisticercos. En zonas rurales donde hay deposición al aire libre, los puercos pueden infectarse al alimentarse de heces contaminadas, mientras que los humanos ingieren carne de puerco mal cocida que contiene cisticercos, lo que resulta nuevamente en una infección por la forma adulta de *Ts*. La *tenia* adulta llega a medir de 2 a 10 metros de longitud y produce poco menos de 1000 proglótidos con miles de huevecillos diariamente, pudiendo sobrevivir en el intestino delgado por años. [Obtenida de CistiMex <http://www-lab.biomedicas.unam.mx/cistimex/index.html>]



## I. I. IV. Enfermedades causadas por la *Taenia solium*

### I. I. IV. I. Teniasis

Como se vio en el ciclo de vida, esta infección por *Taenia solium* se presenta cuando el humano ingiere cisticercos vivos presentes en la carne cruda o insuficientemente cocida proveniente de un cerdo con cisticercosis, o consume verduras y frutas mal lavadas que provienen de zonas de cultivo regadas con aguas negras contaminadas por proglótidos y huevecillos de este organismo. Las enzimas gástricas e intestinales así como las sales biliares del hospedero inducen la evaginación del escólex de la larva, el cual se fija en la pared intestinal (Fig. 5). Una vez anclado, el parásito crece y se diferencia hasta convertirse en una *tenia* adulta productora de proglótidos grávidos [4].

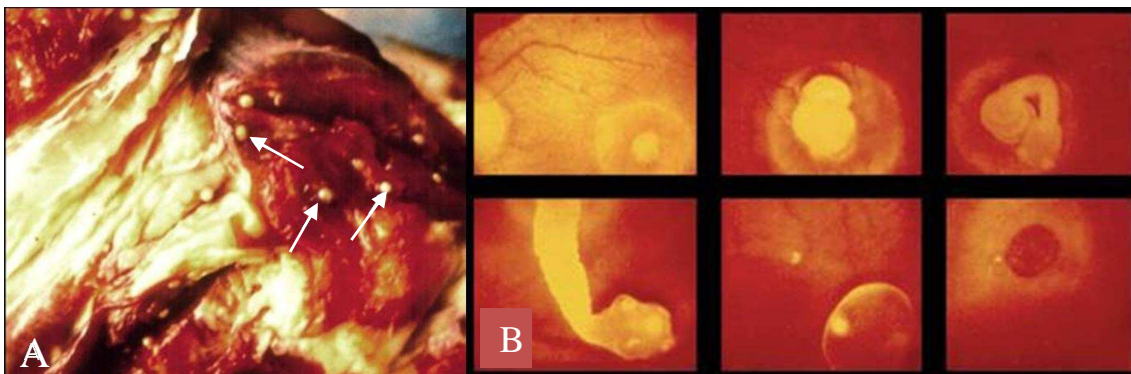


**Figura 5.** Teniasis.- A) Micrografía de luz de un corte de *Taenia solium* en intestino de hamster, en la que se observa el escólex, llamando la atención una ventosa contraída en contacto con la pared intestinal (flecha). B) Micrografía de luz en la que se puede observar una ventosa (V) rodeando una vellosidad intestinal (S), se ve también parte del roseto (R). Tomado de [4].

### I. I. IV. II. Cisticercosis

La cisticercosis (CC), tanto humana como porcina, se da cuando los huevecillos de *T. solium* llegan al intestino delgado y comienzan a evaginar en respuesta a las enzimas digestivas y sales biliares, infiltrándose en el torrente sanguíneo para fijarse en varios tejidos como el músculo esquelético, tejido subcutáneo, sistema nervioso central y el ojo (Fig. 6) [2, 3].

Es de resaltar que esta forma de la infección no es mortal ni sintomática, ya que el parásito tiene una manera bastante eficaz de evadir e inhibir la respuesta inmune y de la cual se desconoce mucho como es que lo logra\*. De esta manera, la mayoría de las personas que padecen cisticercosis lo ignoran y pueden pasar sus días sin que esto perjudique su vida [2, 3].



**Figura 6.** Cisticercosis.- A) Carne de puerco con cisticercosis, en la que se pueden observar claramente varios cisticercos de forma monovesicular (esferas blancas señaladas con flecha). B) Cisticercosis ocular. En la serie de recuadros, se observa la vesícula que contiene al cisticerco dentro de la cámara ocular posterior (Sup.Izq.), en los tres recuadros siguientes se observa la evaginación de la larva y su paso a la cámara anterior, se puede ver claramente el rostelo y las cuatro ventosas (Inf.Izq), en el recuadro siguiente (Inf.Central) se ve como la larva invagina nuevamente en su vesícula, y en el último recuadro se observa el orificio dejado por el paso del cisticerco entre las cámaras. Tomado de [2].

En el caso del puerco pasa lo mismo, con la gran diferencia que todos los animales para consumo humano tienen que pasar por un proceso de inspección, en el que estar infectado con cisticercos causa la confiscación y sacrificio del animal afectado sin que pueda llegar al mercado. Esta medida sanitaria, en el caso de México y otros países subdesarrollados es uno de los puntos frágiles en la prevención de la cisticercosis, debido a que muchos productores evitan la inspección al mandar a los animales infectados hacia rastros clandestinos donde no hay medidas de seguridad y cuya carne logra ser puesta a la venta junto con carne no infectada [3].

\* Una excepción es la forma ocular, que si presenta sintomatología debido a la molestia que ocasiona, lo que lleva a un rápido diagnóstico. También la hay cuando la cisticercosis se localiza en corazón, aunque son casos poco frecuentes.

### *I. I. IV. III. Neurocisticercosis*

Esta forma de la enfermedad se da cuando los huevecillos, al pasar al estado larvario en el intestino delgado después de evaginar, se infiltran al torrente sanguíneo y logran llegar hasta el sistema nervioso central (Fig. 7). Este padecimiento es la enfermedad neurológica más grave causada por algún parásito [2].

El cuadro clínico de la NCC es heterogéneo y no-específico. Cerca del 40% de los casos son subclínicos, es decir, no presentan sintomatología alguna. En los casos sintomáticos, puede estar presente casi cualquier sintomatología neurológica: desde síntomas leves como dolores de cabeza, mareos, convulsiones ocasionales, hasta cuadros neurológicos bastante severos con hipertensión intracraneal o demencia. Parece que el pleiomorfismo de la NCC resulta de factores del parásito (localización, tamaño y número), y del hospedero (grado de reacción inmune e inflamación desarrollada) lo que hace imposible el diagnóstico basado solamente en la clínica [3].

El síntoma más común es la epilepsia, usualmente en la forma de convulsiones parciales con o sin generalización secundaria, lo cual ocurre principalmente cuando el cisticerco se ubica en el área parenquimal del cerebro. La presión intracraneal, relacionada a la perturbación de la circulación normal del fluido cerebroespinal y/o la obstrucción de su absorción, también es frecuente, principalmente cuando el cisticerco radica en el área subaracnoidea. En este caso, dos mecanismos patofisiológicos distintos son responsables de las manifestaciones clínicas: la masa del cisticerco que comprime la estructuras vecinas y la inflamación del espacio subaracnoideo (aracnoiditis) [2, 3].

Parece haber diferencias en la CC y NCC respecto a la edad y género. En niños, la localización ocular es la más frecuente. En el caso de género, parece que en las mujeres hay una mayor titulación de anticuerpos en contra del cisticerco que en los hombres, además que en el caso de infección difusa de cisticerco, la respuesta inflamatoria aguda es bastante severa en las mujeres (Síndrome de Encefalitis), que es visible incluso en estudios de resonancia magnética [3]. Asimismo, la NCC es la más severa de las formas de la infección por cisticercos, ya que a pesar de ser no muy frecuente, puede causar la muerte del que la padece.



**Figura 7.** Neurocisticercosis.- A) Sección cerebral que muestra múltiples cisticercos parenquimatosos subaracnoideos en forma vesicular. B) Micrografía de luz de corte histológico de cisticercos parenquimatosos en la que se puede observar la corona de ganchillos del escólex. Tomado de; A de [3]; B de Fleury, A., et al., II. "Cisticercosis en el ser humano" en Cisticercosis: guía para profesionales de la salud, FCE 2006.

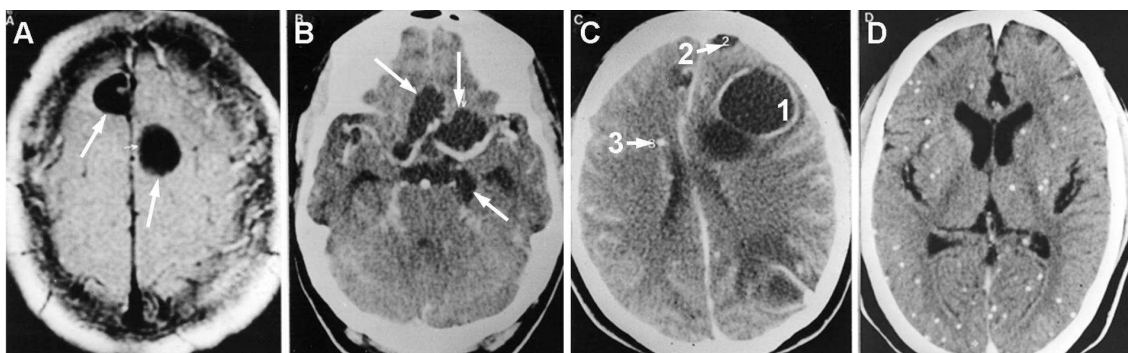
#### I. I. IV. IV. Diagnóstico

Hay dos herramientas de diagnóstico de infección por *Taenia solium*: imagenología e inmunitaria. En el caso de la imagen, hay dos métodos que son el estándar de diagnóstico: tomografía computarizada (TC) y resonancia magnética (RM) (Fig. 8). Ambas permiten definir el número, estadio, localización y extensión de las lesiones cerebrales y músculo-esqueléticas. La tomografía es muy útil para el estudio de lesiones supratentoriales parenquimatosas, permite la visualización de formas activas e inactivas de NCC, y proporciona su orientación y localización. En contraste, la resonancia permite una mejor visualización de las lesiones en la base del cerebro, en el tallo cerebral, en los ventrículos, en la médula espinal y en la fosa posterior, además de dar una evaluación precisa de la reacción inflamatoria alrededor del parásito. Su resolución de alto contraste permite el reconocimiento de cisticercos no visualizados por la tomografía [2, 3].

En el inmunodiagnóstico hay también dos pruebas estándar: ELISA (el más usado), y la inmunoelectrotransferencia. Ambos métodos se basan en la detección de anticuerpos anticisticercos presentes en el suero cuando hay infección por *T. solium*. Estos métodos en general son de bajo costo en comparación con los de imagen. La inmunoelectrotransferencia es una técnica de alta especificidad y sensibilidad, mientras que la ELISA es mucho más eficiente y fácil de realizar.

En ambas pruebas inmunológicas como lo único que se detecta es la presencia de anticuerpos anticisticercos, sólo indica si se ha estado en contacto con el parásito, lo que impide distinguir una infección activa de una resuelta [2, 3]. Para distinguir la presencia de una infección activa existe una prueba que mide un antígeno presente en la superficie y en secreciones de cisticercos viables (denominado HP10)\* [5]

El número de proteínas usadas en la detección por anticuerpos y el desarrollo en técnicas de detección de antígenos y de serodiagnóstico sigue aumentando [6], tanto así que un organismo gubernamental tan importante como es el Centro para el control y prevención de enfermedades de Estados Unidos (*Centers for Disease Control and Prevention, CDC*) participó en el desarrollo y evaluación de un método inmunocromatográfico para la detección de individuos con taeniasis y cisticercosis por *T. solium* [7]. Todo esto refleja el esfuerzo actual para mejorar sustancialmente el diagnóstico temprano de infección, que es uno de los puntos fundamentales en la lucha contra este parásito.



**Figura 8.** Imagenología de neurocisticercosis.- A) Imágenes por resonancia magnética nuclear muestran dos cisticercos vesiculares (flechas). B) Tomografía computarizada (TC) con aumento de contraste muestran tres cisticercos subaracnoideos de gran tamaño (flechas). C) TC muestra un cisticercos coloidal (1) rodeado por un edema perilesional, se observan también un céstodo vesicular (2) y una calcificación cerebral (3). D) TC muestra múltiples calcificaciones cerebrales parenquimatosas que corresponden a parásitos muertos (puntos blancos dispersos por todo el tejido). Tomado de [3].

#### I. I. IV. V. Tratamiento

Antes de 1975, el tratamiento de la neurocisticercosis estaba limitado al removimiento quirúrgico del parásito, lo que no siempre era posible, o a la colocación de una válvula para desviar la circulación del líquido cefalorraquídeo hacia el tórax o el abdomen cuando un cisticercos obstruía su circulación, fue entonces que aparecieron medicamentos que atacaban y destruían al parásito. Actualmente se emplean dos

\* Esta prueba mide la reacción de anticuerpos monoclonales contra un epítipo de carbohidratos repetitivos presente en glicoproteínas adherentes a lectina-lentina de *T. solium*.

fármacos, el albendazol y praziquantel; aproximadamente un 85% de los cisticercos parenquimatosos son destruidos en un solo curso del tratamiento con albendazol, mientras que 75% son destruidos en un curso con praziquantel, y más del 95% en el uso secuencial de un fármaco después del otro [2]<sup>1</sup>.

El principal problema con el tratamiento químico de la NCC es el incremento de la reacción inflamatoria alrededor de la ubicación del cisticerco después de su destrucción repentina. Este incremento inflamatorio está asociado frecuentemente con un deterioro clínico temporal y podría ser la causa de fibrosis en las leptomeninges y de una hidrocefalia obstructiva consecuente. Por esta razón, de manera conjunta con los fármacos se recomienda administrar antiinflamatorios (corticosteroides), por lo que es necesario encontrar nuevos tratamientos que aminoren la respuesta inflamatoria sin dejar de destruir al parásito [2, 3].

En general se pueden enunciar cuatro puntos para el control de la transmisión de la Cisticercosis:

- 1) Tratar a los portadores del parásito adulto, ya que se ha visto es el principal factor de riesgo para la transmisión de la cisticercosis en el ambiente cercano, además de que es más fácil tratar a un portador que modificar el manejo del drenaje y la infraestructura de irrigación de los países en desarrollo [1].
- 2) Ofrecer un tratamiento cestocida contra el parásito adulto intestinal a toda la población. Estudios en Ecuador y México han mostrado una disminución en la prevalencia de la teniasis mediante el tratamiento masivo [1].
- 3) Proveer educación para la salud. Mediante esta se pretende modificar los conocimientos, actitudes y prácticas de la comunidad con la meta de evitar nuevas infecciones en humanos y cerdos. Prácticas de este tipo en una comunidad del estado de Morelos lograron que en 6 meses después de la intervención no se encontrara ningún cerdo con cisticercosis [1].
- 4) Manejo adecuado de los cerdos, al mantenerlos confinados y sin acceso a la basura y heces fecales, además de buenas prácticas de inspección de estos animales para su ingreso a rastros [1].

---

<sup>1</sup> Ver Apéndice B para una información más detallada del tratamiento por albendazol y praziquantel.

## I. I. V. Estadísticas

Basado en datos del año de 1992 se estimó que había 50 millones de casos de cisticercosis/teniasis en todo el mundo, y que la neurocisticercosis había causado cerca de 50 000 muertes [8]. Sin embargo, una estimación verdaderamente confiable de la prevalencia real en humanos está muy limitada debido a que la cisticercosis no es una enfermedad reportable a pesar de varios esfuerzos por cambiar esta situación. Este parásito es endémico en Brasil, Colombia, Ecuador, Guatemala, México, Perú, India, China, Nueva Guinea (Figura 9) [8].

En el mismo año de 1992, se hizo una encuesta serológica que estableció una seropositividad en población abierta de 1.2%, siendo más frecuente en mujeres y niños que tenían condiciones socioeconómicas, de higiene personal y de vivienda bajas, así como las que vivían en regiones geográficas rurales [1].



**Figura 9.** Distribución de la Cisticercosis a Nivel Mundial.- Se nota su alta prevalencia en Latinoamérica, África y Asia, aunque el reciente aumento en el número de casos en países desarrollados la hace una patología reemergente (debido a que ya había sido erradicada en Europa hace más de un siglo) y de dimensiones globales. Tomado de [2].

La visión actual que se tiene de la enfermedad es que a pesar de las estrictas medidas de higiene implementadas en los países “desarrollados”, hay un aumento reciente en el número de casos de cisticercosis debido a la migración de personas infectadas provenientes de países endémicos para este parásito [9]. A su vez, la situación en países en vías de desarrollo presenta posiciones encontradas; hay quien proclama una disminución importante en el número de casos de NCC desde hace 25

años [10], mientras que otros afirman lo contrario, indicando además que la prevalencia real sigue subestimada, sobre todo en América latina [11].

### **I. I. VI. Evolución**

Las especies del género *Taenia* fueron de los primeros helmintos reconocidos en el humano, remontándose hasta los griegos, y actualmente hay descritas unas a 40 especies, de las cuales el ciclo de vida es conocido para treinta y tres [12].

Parece que la diversidad de estas especies se ha dado en base a cambios en la “elección” por parte del parásito del hospedero carnívoro definitivo. Por tanto, hay cambios entre hospederos filogenéticamente no relacionados (entre cánidos, félidos y hiénidos por ejemplo) que históricamente han explotado los mismos recursos alimenticios (la presa). Así, la distribución de hospederos para la mayoría de las especies de *Taenia* es una consecuencia de la ecología y hábitos alimenticios históricos de los hospederos [12].

Hay opiniones encontradas acerca de cómo la *tenia* estableció ese vínculo parásito-hospedero con el ser humano. Por un lado se propone que en el comienzo de la crianza de animales, circulaban en ciclos sinantrópicos entre perros y rumiantes que la *tenia* colonizó primeramente, de donde pasó a seres humanos. En otra hipótesis contrastante, es el humano quien sería el portador histórico, en específico de *Taenia solium*, mucho antes de la domesticación de animales, y partir del cual *Taenia solium* terminó adaptándose al puerco como su hospedero intermediario con el advenimiento de su domesticación y crianza [12].

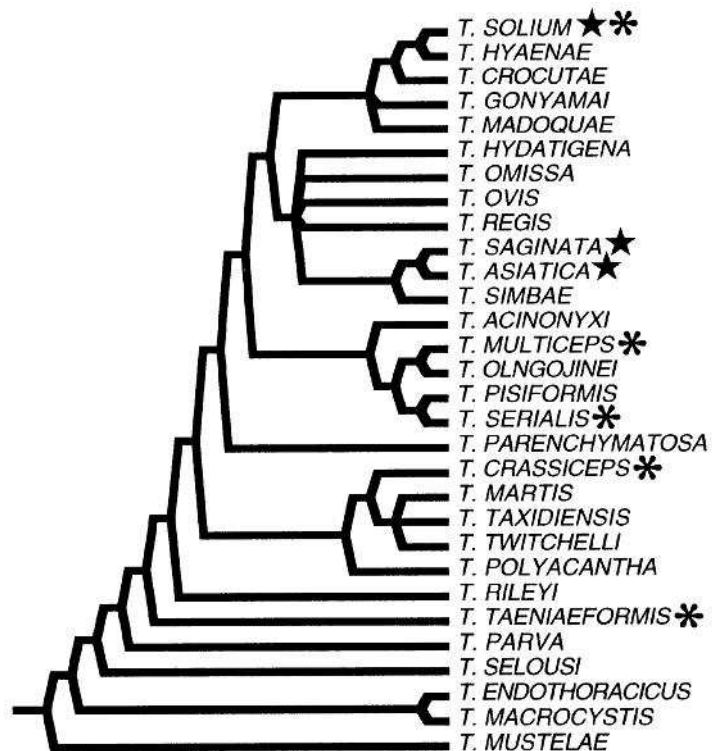
Pues bien, en estudios con datos filogenéticos, geográficos, ecológicos y moleculares, la segunda idea parece ser la más acertada, indicando una asociación entre la *tenia* y los homínidos que se desarrolló antes del origen de los humanos modernos en África. Interesantemente, los estudios muestran que el origen de la relación hospedero-parásito que involucra a la *tenia* y a los humanos podría haber surgido a través de asociaciones directas presa-predador entre los homínidos y los bóvidos matados por predadores carnívoros como los félidos, cánidos y hiénidos en las sabanas del África Subsahariana durante el Pleistoceno o mucho antes. Esto es consistente con las siguientes observaciones: 1) una de las especies hermanas de *T. solium* es *T. hyaenae*,



que es un ténido que circula entre hienas y antílopes; y 2) la especie hermana de *T. saginata*+*T. asiatica* es *T. simbae*, un parásito típico en leones y antílopes en África (Fig. 10). Así aparece un contexto ecológico y biogeográfico para la adquisición de la *tenia* por homínidos, ya que un cambio de en la dieta de herbívoro a carnívoro por nuestros ancestros habría dado como consecuencia el compartir los parásitos dentro de una “comunidad” de carnívoros y su presa bóvida [12].

Consecuentemente, los humanos representarían la fuente de ténidos que después se establecieron en ganado y cerdos como hospederos intermediarios. En este aspecto, la adquisición de *Taenia* por animales domésticos a partir de humanos ha ocurrido independiente al parecer en tres ocasiones distintas, representadas ahora por *T. saginata* en vacas, y los linajes separados *T. asiatica* y *T. solium* en puercos. Es así como las asociaciones ecológicas establecidas por nuestros ancestros han tenido consecuencias en la distribución de patógenos y parásitos que continúan ejerciendo substanciales repercusiones médicas y veterinarias para los humanos contemporáneos [12].

**Figura 10.** Evolución de *Taenia Solium*.- Filogenia hipotética de las especies de *taenia* basadas en análisis de 30 especies para los que se tienen datos morfológicos completos de céstodo y adulto, así como información detallada sobre su historia de vida. Se resaltan aquellas especies que específicamente habitan dentro del hospedero como gusanos adultos estrobilados en humano (estrellas), y de aquellas especies que se saben causan cisticercosis o coenurosis en el humano (asteriscos). Solamente *Taenia solium* puede causar tanto teniasis como cisticercosis en el humano. La especie hermana de *T. solium*, *T. hyaenae*, es un taenido que circula entre hienas y antílopes, mientras que la especie más cercana a *T. saginata* y *T. asiatica* es *T. simbae*, un parásito típico de leones y antílopes en África. Esto podría apuntar al origen la asociación específica entre *T. solium* y homínidos a los albores del hombre moderno al cambiar su condición de hervívoro a carnívoro y compartir parásitos con otros carnívoros y sus presas. Tomado de [12].



### ***I. I. VII. Secuenciación del genoma de Taenia solium***

En 2004, se creó en la UNAM un consorcio con investigadores de distintas dependencias\* de la misma con la finalidad de obtener la secuencia del genoma de *Taenia solium*. El proyecto se planteó como un intento de comprender mejor la fisiopatología del parásito, búsqueda de blancos terapéuticos y desarrollo de vacunas en contra del mismo, así como elucidar su origen, evolución, y la especificidad hacia el humano como hospedero final [13]. Este consorcio (denominado IMPULSA03) surgió como parte de un conjunto de megaproyectos emprendidos por la UNAM, y es el mayor proyecto genómico que se ha realizado dentro de la universidad hasta la fecha.

Se ha empleado una estrategia híbrida durante la dilucidación del genoma; por un lado se emplea la técnica de Sanger de secuenciación capilar que genera fragmentos de 800 a 1000 nucleótidos, y se complementa con el novedoso método de pirosecuenciación 454, que reduce los tiempos y costos de secuenciación pero que produce secuencias de menor tamaño y más difíciles de armar.

El laboratorio de Biología Teórica del Dr. Marco José en el Instituto de Investigaciones Biomédicas, se integró al consorcio con la propuesta de buscar secuencias de RNA no codificante (abreviadas en lo sucesivo como “ncRNA”, traducción del inglés *ncRNA*, *non-coding RNA*), debido a la creciente evidencia de la importancia de éstas en la regulación genética de organismos eucariontes [14].

Es de mencionar que hasta el mes de octubre de 2009 había veintidós proyectos de secuenciación genómica de parásitos helmintos que infectan al humano (o biológicamente muy cercanos), de los cuales cuatro ya están en etapa de “borrador” y los dieciocho restantes siguen en progreso\* [15].

#### ***I. I. VII. I. Características del genoma de Taenia solium***

El tamaño del genoma calculado mediante citofluorometría es de 270 MB [13], mientras que el ensamblado “*in silico*” del genoma es de 130 MB (esta diferencia se

---

\* Centro de Ciencias Genómicas, Instituto de Biotecnología, Facultad de Medicina, Facultad de Ciencias e Instituto de Investigaciones Biomédicas

\* *Trichinella spiralis*, *Schistosoma mansoni*, *Schistosoma japonicum* y *Taenia solium*

debe a que el ensamble computacional es haploide, mientras que los 270 MB corresponden al genoma diploide) y la densidad codificante parece ser relativamente alta: un ~ 9% codifica proteínas. El número de intrones y su tamaño no presentan diferencias respecto de la distribución promedio [13]. Su contenido de secuencias repetitivas es de ~ 7%, llamando la atención que una repetición en *tandem* de 53 pares de bases y diferentes repeticiones de tetranucleótidos representan el 0.5% y 4.5% del genoma, respectivamente [16].

Respecto del porcentaje de secuencia codificante de proteínas, se estima que posee cerca de 11,000 genes, entre los cuales se han identificado más de 6,000 en bases de datos especializadas, y varios se expresan abundantemente tanto en adulto como en larva. Cerca de 2,000 genes tienen una similitud significativa con secuencias de la base SwissProt y cerca de un 27 % de los genes no tienen una secuencia similar en las bases SwissProt y TREMBL [16].

Las secuencias producidas cubren más del 99% del genoma, por lo actualmente el proyecto se encuentra en la etapa de ensamble de contigs. Las secuencias EST generadas durante la secuenciación están disponibles a través de GenBank o de la página web del consorcio\*.

### ***I. I. VIII. Parásitos y carcinogénesis.***

Hallazgos recientes han indicado una asociación entre la infección por parásitos helmintos y la aparición de carcinogénesis. Aunque sólo en dos helmintos, *Schistosoma haematobium* y *Opisthorcis viverrini*, se ha demostrado que son carcinogénicos para el humano, otros han sido implicados en facilitar la transformación maligna [17]. Los mecanismos conocidos de cáncer inducido por helmintos incluyen inflamación crónica, modulación del sistema inmune del hospedero, inhibición de comunicación intracelular, disrupción de las rutas de proliferación-antiproliferación, inducción de inestabilidad genómica y estimulación de progenie de células troncales malignas [17]. Aproximadamente el 16 % de todos los casos de cáncer a nivel mundial son atribuibles a agentes patógenos, incluyendo a *Schistosoma* y *Fasciola hepatica* [17].

---

\* <http://bioinformatica.biomedicas.unam.mx/taenia/>.

Para poder sobrevivir largos periodos en un medio agresivo y desventajoso, los parásitos helmintos secretan varios factores solubles que interactúan con la célula hospedero. Por tanto, es posible que algunas de estas moléculas puedan modificar la homeostasis de la célula e incrementen el riesgo de una transformación maligna [18]. Una molécula de RNA, secretada por el cisticerco de *Taenia solium*, tiene la capacidad de transformar células embrionarias de hamster sirio *in vitro* [18]. Esta molécula también induce daño cromosómico en linfocitos humanos cultivados, lo que sugiere que este factor podría inducir inestabilidad genética en individuos infectados [18]. Se desconoce el mecanismo por el cual esta molécula induce el daño genético y la transformación de la célula hospedero; sin embargo, una interacción directa entre el RNA secretado y el DNA genómico del hospedero podría relacionarse con los efectos observados en otras moléculas de RNA de bajo peso molecular como U5, capaces de inducir aberraciones cromosómicas y transformar células murinas *in vitro* [18].

El hecho de que una molécula de RNA secretada por *T. solium* pudiera relacionarse con carcinogénesis nos obliga a prestar más atención a esta clase de secuencias. Hay una gran cantidad de estudios en otros organismos donde se asocia RNA y cáncer, en particular, los genes de ncRNA conocidos como *microRNA*.

Por eso consideramos apropiado buscar similitudes entre las secuencias de ncRNA disponibles y aquellas del genoma de *T. solium* que fueron generadas por el consorcio del genoma de *Taenia solium* en un intento por encontrar mecanismos moleculares relacionados respecto de otras especies que no hayan sido identificados hasta ahora. Razón por la cuál a continuación se hace una revisión del tema de ncRNA.

\* \* \*



Consortio del Genoma de *Taenia solium*

## **I. II RNA no-codificante (ncRNA)**

### ***I. II. I. El RNA salta a la escena***

El DNA había sido considerado por mucho tiempo como la molécula de la vida, desde comienzos del siglo veinte cuando se descubrió que era causante de la transformación de bacterias inocuas en mortales, hasta 2001 cuando se concluye el proyecto del genoma humano, pasando por 1953, cuando se describe su estructura tridimensional y a partir de ella se deduce el mecanismo de la herencia. En contraparte, se veía al RNA como un producto intermediario entre el DNA y las proteínas debido al descubrimiento del RNA mensajero y el RNA ribosomal en los años sesenta. Esto fue reflejado en el famoso y hoy extinto “dogma central de la biología molecular”, que se convirtió el punto de partida conceptual de toda la biología molecular del siglo pasado.

Esta visión DNA → mRNA → Proteína, pronto sufrió del falsacionismo al descubrirse los retrovirus, que trasgredían la visión unidireccional del flujo de la información biológica al invertir el sentido de la misma, generando una cadena de DNA a partir de RNA mediante una enzima hasta entonces desconocida pero teóricamente posible que fue bautizada como transcriptasa reversa [19].

Después no sólo se cuestionó su lugar como punto de partida, sino su presencia misma durante la historia de la vida en la tierra. La hipótesis del “mundo de RNA” propone la aparición primigenia de cadenas de RNA con poder catalítico, autoorganizadas y lo más sorprendente, autorreplicantes [20, 21]. En algún momento de la evolución, el DNA se presentó como la mejor opción para ser el depositario de la información genética, preservando tal función hasta ahora.

Cuando se descifró el código genético, el RNA emergió como el depositario de éste mediante el RNA de transferencia (tRNA), descifrando los codones en el mRNA y uniendo los aminoácidos correspondientes dentro del ribosoma [19]. Este hallazgo apoya la idea del mundo de RNA, y al mismo tiempo cuestiona su único papel como intermediario entre DNA y proteínas. Actualmente se sabe que el ribosoma es en verdad una ribozima[22].

Además, los estudios de regulación genética abundaban sobre asociaciones entre proteínas y secuencias específicas de DNA, lo que hizo dar más énfasis a proteínas y el

DNA como elementos reguladores en todos los organismos\*. El descubrimiento del RNA *lin-4*, que regula la traducción de la proteína *lin-14* al interactuar con su RNA, mostró que la regulación genética no era exclusiva de proteínas y DNA [23].

Finalmente, los estudios hechos a la gran cantidad de genomas secuenciados también han modificado la visión del DNA como simple contenedor de secuencias codificantes de proteína, rRNA y tRNA. Ahora sabemos que los genomas contienen, además, una gran cantidad de secuencias repetidas, pseudogenes, transposones, intrones y fósiles virales [19].

Todo lo anterior muestra como el RNA ha ido cobrando cada vez mayor importancia dentro de la biología, siendo el tema de ncRNA el más novedoso y que abordamos en esta sección.

## ***I. II. II. RNA no codificante***

El término de RNA no codificante (*ncRNA*) se emplea para denominar aquellas secuencias de RNA que no codifican una proteína, y que son distintos a los RNA de transferencia y ribosomal. El adjetivo no codificante no significa que tales secuencias carezcan de información o función, al contrario, estudios recientes indican que tienen la capacidad de regular varios procesos moleculares y celulares como la transcripción, traducción, metilación del DNA, impronta, arquitectura de la cromatina, regulación epigenética, etc. [14]. Se ha encontrado que alteraciones en estas secuencias están presentes en enfermedades crónico degenerativas o complejas cáncer [24, 25].

Las secuencias de RNA no codificante fueron descubiertas paulatinamente y de manera independiente desde los años noventa. En 1990, se encontró que al intentar sobreexpresar el gen de la chalcona sintasa en petunias mediante la introducción de copias de su secuencia, se daba una “co-supresión” del gen introducido y del gen endógeno [26]. En 1992 se describió algo parecido en *Neurospora crassa* aunque en ese momento no se relacionó con lo hallado en petunia<sup>^</sup> [27]. En 1993 se descubre *lin-4*, un microRNA que regula negativamente el nivel de la proteína *lin-14* [23]. En ese mismo

---

\* Recordemos la famosa frase de Jacques Monod “*lo que es verdad para Escherichia coli, también es verdad para un elefante*”

<sup>^</sup> En ambos, la cosupresión se da a través de una disminución en los niveles de mRNA del gen, aunque se ignoraba como sucedía.

año, se descubre el gen *Xist*, un RNA de ~17 kb que participa en la inactivación del cromosoma X en hembras y se expresa únicamente en el cromosoma silenciado [28]. Después, en 1998 se describe en *Caenorhabditis elegans* el silenciamiento específico de genes mediante la introducción de RNA de doble cadena<sup>♦</sup> [29]. Y recientemente, en 2007, se reportó el RNA de 2.2 Kb denominado *HOTAIR*, que a pesar de residir en el locus *HOXC*, reprime la transcripción en *trans* a través de 40 Kb del locus *HOXD* [30]. Lo anterior es para eucariontes, pero la aparición de ncRNA ha sido tan inesperada que incluso en organismos procariontes cuyos genomas han sido estudiados extensivamente como *Escherichia coli* se encontraron secuencias de esta clase en el 2001 [31].

Es importante mencionar que los ncRNA han aparecido dentro de un contexto que cuestiona la definición del concepto “gen”, término acuñado en 1909 por el botánico danés Wilhelm Johansen que lo definía como la unidad asociada con un rasgo hereditario, que después fue asociado a una enzima por Beadle y Tatum mediante su hipótesis “un gen, una enzima”, y que ahora está siendo llevada hasta límites un tanto difusos al definírsele como una unidad transcripcional [32].

La cantidad de información que se ha obtenido sobre los ncRNA; como mecanismo de acción, estructura secundaria, motivos en su secuencia, función, etc., han permitido conceptualizarlos y clasificarlos de una mejor manera. A continuación elaboramos sobre esto último.

### **I. II. III Clasificación de los RNA no codificantes**

Los RNA no codificantes se dividen en\*:

- |  |                                     |
|--|-------------------------------------|
| a) Pequeños Nucleolares ( <i>snoRNA</i> ),         | d) RNA de interacción con proteínas |
| b) Pequeños RNA de Interferencia ( <i>siRNA</i> ), | Piwi ( <i>piRNA</i> ),              |
| c) MicroRNAs ( <i>miRNA</i> ),                     | e) ncRNA largos ( <i>lncRNA</i> )   |
|  | f) Riboswitches                     |

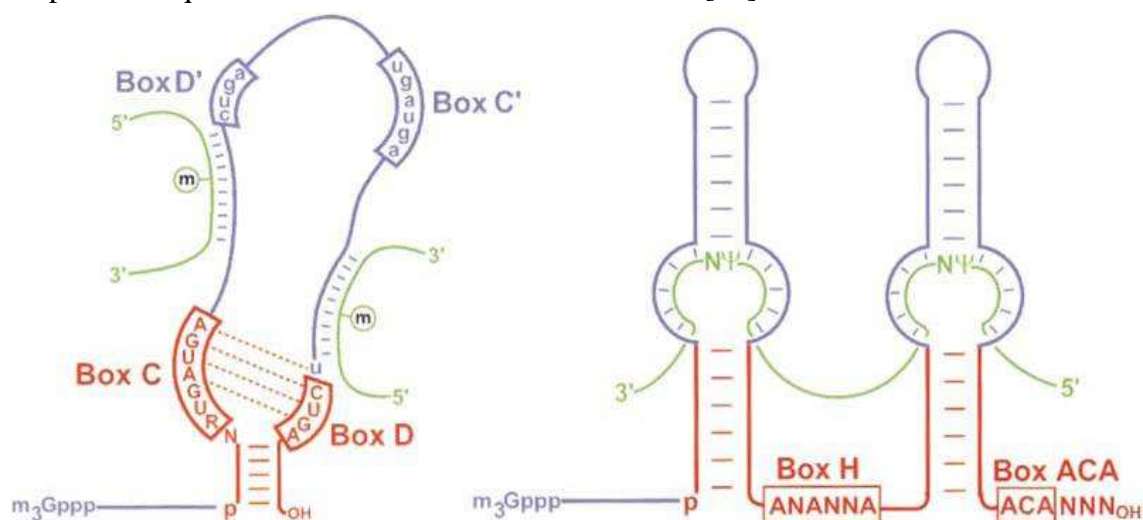
---

<sup>♦</sup> La descripción de la interferencia de RNA fue reportada en 1998 por Mello y Fire, y la importancia que ha demostrado tener su estudio les llevó a recibir el premio Nobel en 2006.

\* Esta clasificación nos permitirá relacionar rápidamente los resultados de *BLASTn* sobre el genoma de *T. solium* con lo que se sabe sobre ncRNA, ayudando a la anotación de las secuencias que presenten similitud con las bases de datos de ncRNA NONCODE y miRBase.

### I. II. III. I. RNA Pequeño Nucleolar (snoRNA, small nucleolar RNA)

Los snoRNA guían la modificación del rRNA durante la biogénesis del ribosoma, aunque pueden tener como blanco otros RNA. Hay dos clases; la caja C/D, que guía la reacción 2'-O-ribose-methylation, y la caja H/ACA, que guía la pseudouridilación de los RNA blanco [33]. Aunque los snoRNAs involucrados en la biogénesis del ribosoma están localizados en el nucleolo (dónde se descubrió y de allí su nombre), también se localizan en los cuerpos de Cajal, por lo que también se conocen como scaRNA (Small Cajal body RNA) [34]. De hecho, su nombre ya resulta anacrónico debido a que tienen un amplio rango de localización celular además del nucleolo, así como de otras funciones y blancos [35]. Proviene de los intrones, miden entre de 60 y 300 nucleótidos y algunos exhiben especificidad de tejido y regulación durante el desarrollo o impronta. Como ejemplo de disfunción está el caso del snoRNA HBII-52, asociado al splicing aberrante del gen del receptor de serotonina 5-HT(2C)R en pacientes que sufren del Síndrome de Prader-Willi [25].

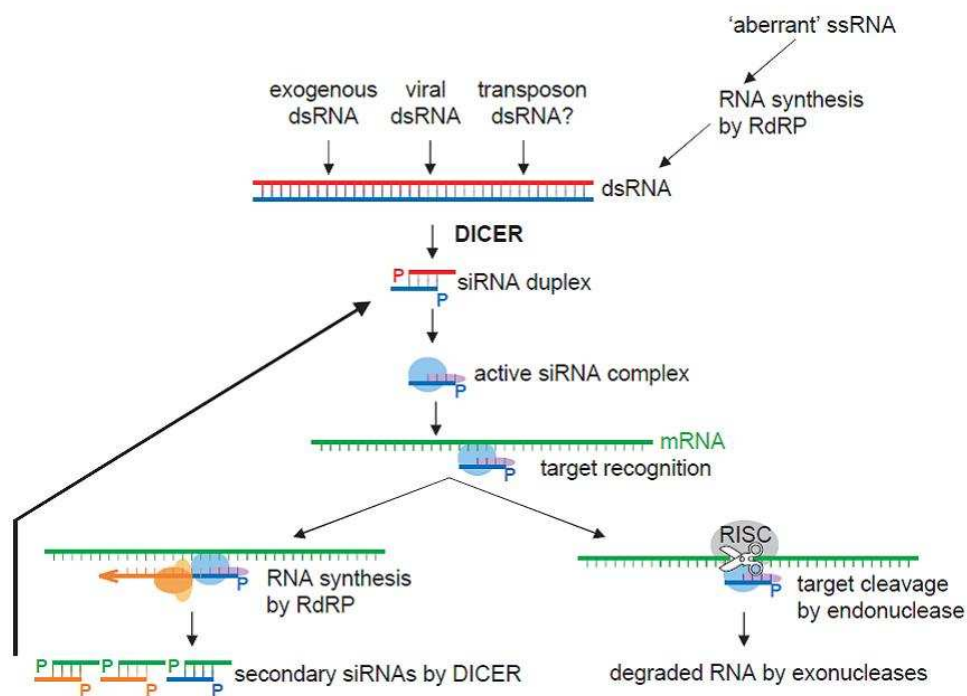


**Figura 11.** Estructura y función de los snoRNA.- En los snoRNA que guían la 2'-O-metilación, los motivos de las cajas C y D, y el tallo que se forma entre las terminales 5',3' constituyen un motivo estructural denominado "kink turn" (K-turn) que es reconocido específicamente por la proteína snoRNP de 15.5kDa. Las cajas C' y D' representan copias internas, y a menudo imperfectas, de las cajas C y D. Las líneas punteadas indican nucleótidos que interactúan en las cajas C y D. Los snoRNAs que guían la pseudouridilación se pliegan en estructuras "hairpina-bisagra-hairpina-cola" y contienen las cajas H y ACA. La caja C/D de los snoRNA's metiladores y su RNA sustrato forman un par de doble hélices de 10 a 21 bases de largo en el cual el residuo a modificar se posiciona a 5 nucleótidos exactos "upstream" de la caja D o D'. La hairpina 5' y/o 3' de la caja H/ACA en los snoRNAs que uridilan contienen un loop interno, llamado el "bolsillo" de pseudouridilación, que forma dos dúplex cortos (3-10 bp) con nucleótidos que flanquean a la base de uridina sustrato que está localizada aproximadamente a 15 nucleótidos de la caja H/ACA del snoRNA. Aunque cada caja snoRNA podría llevar a cabo dos reacciones de modificación, a excepción de unos cuantos, la mayoría de los snoRNAs poseen un solo dominio funcional de 2'-O-metilación o pseudouridilación. Los motivos centrales de los snoRNA que son esenciales y suficientes para el correcto procesamiento y acumulación nucleolar de snoRNAs están resaltados en rojo. Las regiones que no contribuyen a la estabilidad metabólica del snoRNA se marcan en azul. Los snoRNAs transcritos independientemente por la RNA pol II contienen secuencias líder 5' y llevan la clásica estructura de trimetilguanosina "cap". Los RNAs sustrato se encuentran en verde. Se marcan también los nucleótidos destinados para la pseudouridilación ( $\Psi$ ) y la 2'-O-metilación (una "m" con círculo). Tomado de [33].



### I. II. III. II. siRNA

Son secuencias de 22 nucleótidos de longitud cuya función es silenciar transcritos homólogos de mayor tamaño, fenómeno conocido como interferencia de RNA (RNAi). Un siRNA se deriva de un precursor de RNA de doble cadena de mayor longitud que es cortado por la enzima Dicer (una RNAasa tipo III multidominio), e incorporado mediante la proteína Argonauta al complejo de silenciamiento inducido por RNA (RISC). El siRNA guía el complejo RISC hacia el mRNA blanco, que es cortado por la proteína Argonauta (Fig. 12) [36].



**Figura 12.** La producción de siRNAs a partir del precursor de RNA de doble cadena se da mediante la enzima Dicer, este precursor puede provenir de varias fuentes, como una infección viral, transposones dentro del genoma, e incluso del medio ambiente o de transcritos aberrantes. Estas secuencias se incorporan al complejo RISC, mediante el cual se corta el mRNA con secuencia complementaria al siRNA. Además, puede haber una amplificación de siRNAs secundarios por una polimerasa de RNA dependiente de RNA y Dicer. Tomado de [37]

Una característica impresionante de la interferencia de RNA es su potencia. A excepción de insectos y vertebrados, unas cuantas moléculas de RNA de doble cadena (dsRNA) por célula pueden inducir una respuesta y, en plantas y nemátodos, inducir un silenciamiento sistémico en todo el organismo. Esto se logra por la síntesis de siRNAs secundarios mediante la polimerasa de RNA-dependiente de RNA (RdRP) [36].

Los siRNAs cumplen un papel protector ante infecciones virales. Estudios en plantas reportan una mayor sensibilidad a infecciones virales cuando se inhibe la interferencia por RNA. Esta protección se da cuando el sistema de interferencia

responde al “monitoreo” de los niveles de RNA dentro de la célula en el momento que se hallara un mRNA con una concentración mayor a la normal, situación muy común en la replicación intracelular de los virus [38].

Además de silenciar genes específicos, los siRNAs participan en la formación de heterocromatina mediante el complejo de silenciamiento transcripcional inducido por RNA (RITS, *RNA Induced Transcriptional Silencing*). El reconocimiento de transcritos nacientes mediante siRNA y RITS promueve la metilación de la histona H3 en la lisina nueve por metiltransferasas de histona, que conduce al reclutamiento de la proteína Swi6 y la subsiguiente compactación de la cromatina [36].

Las proteínas Ago, Dicer, Piwi y RdRP están conservadas en todos los linajes eucariontes principales, mientras que los procariontes no tienen homólogos de éstas (a excepción de Ago, y parecen tener un sistema análogo de interferencia que evolucionó de manera independiente) [39, 40]. Lo anterior hace pensar que el último ancestro común de eucariontes poseía ya un sistema de interferencia de RNA [40].

### ***I. II. III. III. MicroRNA (miRNA)***

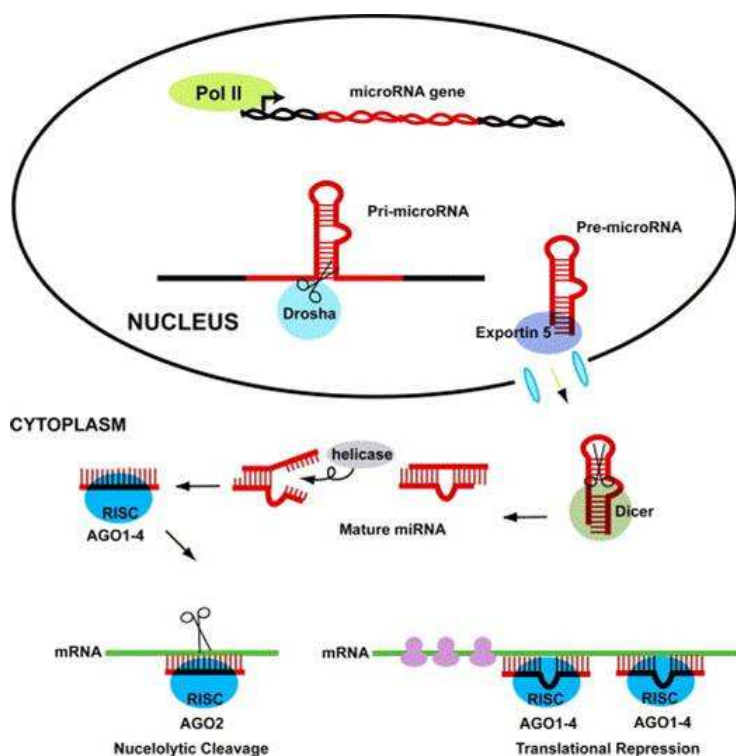
Son genes que al final de una serie de modificaciones generan secuencias de 22 nucleótidos de longitud que regulan la expresión de genes específicos mediante inhibición de la traducción, y en algunos casos la degradación del mRNA blanco (Figura 13). La represión de la traducción se da por un apareamiento imperfecto con el mRNA blanco en su región 3' no traducible (UTR), empleando para ello al complejo de silenciamiento inducido por RNA (RISC), pero se ignora el mecanismo y momento exacto en el que se da la represión de la traducción. La degradación se da cuando hay un apareamiento perfecto entre el miRNA y su RNA blanco [36].

La RNA polimerasa II es quien transcribe los genes de miRNA. Los transcritos contienen el residuo de 7-metilguanosina en su extremo 5' (cap) y están poliadenilados en su región 3'. Generalmente se localizan en regiones intergénicas e intrones, en menor grado en exones, y unos cuantos en zona repetidas [41, 42]. Los intergénicos pueden ser monocistrónicos (una unidad transcripcional por gen), o policistrónicos (más de un gen por unidad transcripcional), teniendo ambos casos secuencias promotoras propias [41].

Cuando se localizan en intrones, pueden ser igualmente monocistrónicos o policistrónicos, y a menudo el intrón corresponde exactamente en tamaño al precursor del miRNA. Los miRNAs “exónicos” a menudo entrelazan un exón y un intrón en genes no codificantes cuyos transcritos son parecidos a mRNA\* [41].

Los miRNA participan en la regulación de procesos biológicos, que incluyen diferenciación, proliferación celular, apoptosis, angiogénesis, etc. Son fundamentales en el desarrollo, ya que animales que son incapaces de producir miRNAs no sobreviven o no pueden reproducirse [43]. Defectos en los miRNA han sido implicados en varias enfermedades, como el síndrome de Tourette, síndrome de DiGeorge, degeneración neuronal y sobre todo en tumores y cáncer, en el que hay una cantidad inmensa de estudios [44, 45].

Debido a lo conservado de los miRNA en eucariontes, se propone que este sistema de regulación evolucionó justo antes de la divergencia entre plantas y metazoarios, y que apareció después del sistema de interferencia de RNA [40]. Se estima además que un tercio de los genes en animales están controlados por miRNA, lo que da mayor regulación de la expresión génica aparte de las secuencias promotoras, potenciadoras, etc. debido a la especificidad por su mRNA blanco [43].



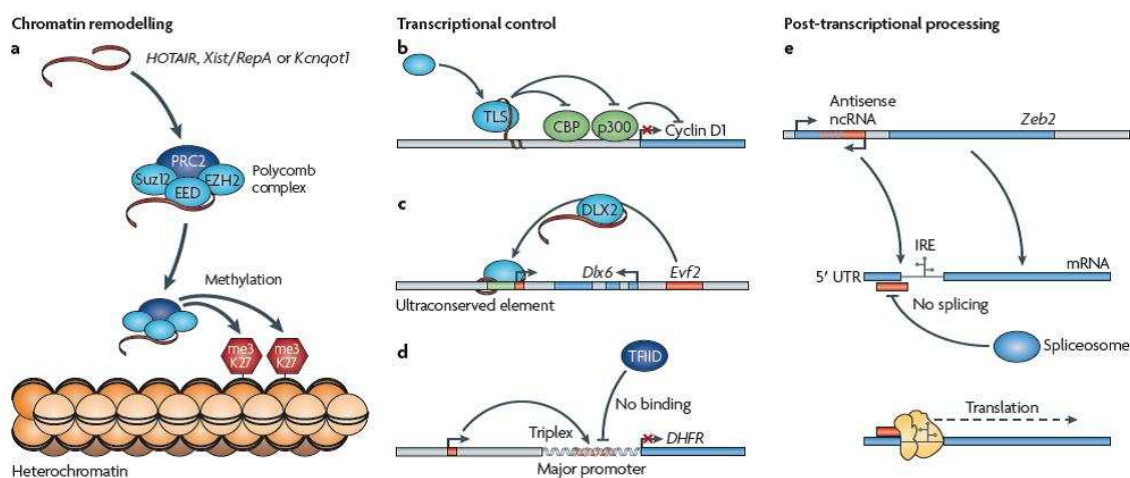
**Figura 13.** Biogénesis de los MicroRNA. El transcrito del gen de miRNA generado por la RNA polimerasa II es procesado por la enzima Drosha en el núcleo. El precursor resultante, “pre-mirRNA”, se exporta al citoplasma (al parecer por la Exportina 5) para que sea procesado por Dicer y genere los oligonucleótidos característicos de 22 bases. De manera coordinada el mirRNA maduro se transfiere al complejo de silenciamiento RISC y es desenrollado por una helicasa. Aquellos mRNA blanco que sean perfectamente complementarios al mirRNA son degradados, mientras que mRNA blanco con apareamiento imperfecto se suprime su traducción. Tomado de [46]

\* Conocidos como” mRNA-like”, son genes cuyos transcritos poseen exones, están poliadenilados, pero no codifican proteínas y además su función es desconocida.

## I. II. III. IV. ncRNA largo (lncRNA)

Los lncRNA son transcritos que van de 300 hasta 10,000 nucleótidos de longitud. La forma en que actúan es amplia, siendo la modificación de la cromatina la más común, pero también regulan la transcripción y las modificaciones post-transcripcionales (Fig. 14) [47]. Son regulados durante el desarrollo, muestran expresión tejido-específica, localización en compartimientos subcelulares y asociación a enfermedades [48]. Algunos lncRNA presentan evidencia de evolución selectiva [48].

El primer lncRNA caracterizado fue *XIST*, involucrado en el silenciamiento en *cis* del cromosoma X en mamíferos [28]. El más reciente, denominado *HOTAIR*, se localiza en el clúster *HOXC*, y reprime la transcripción en *trans* del clúster *HOXD* [30]. Otro ejemplo es *Air* (de 100 Kb), asociado a un loci improntado en ratón, donde silencia en *cis* una región de 400 Kb que contiene a tres genes. Se espera aumente el número de lncRNA conforme se estudien más transcritos [48].

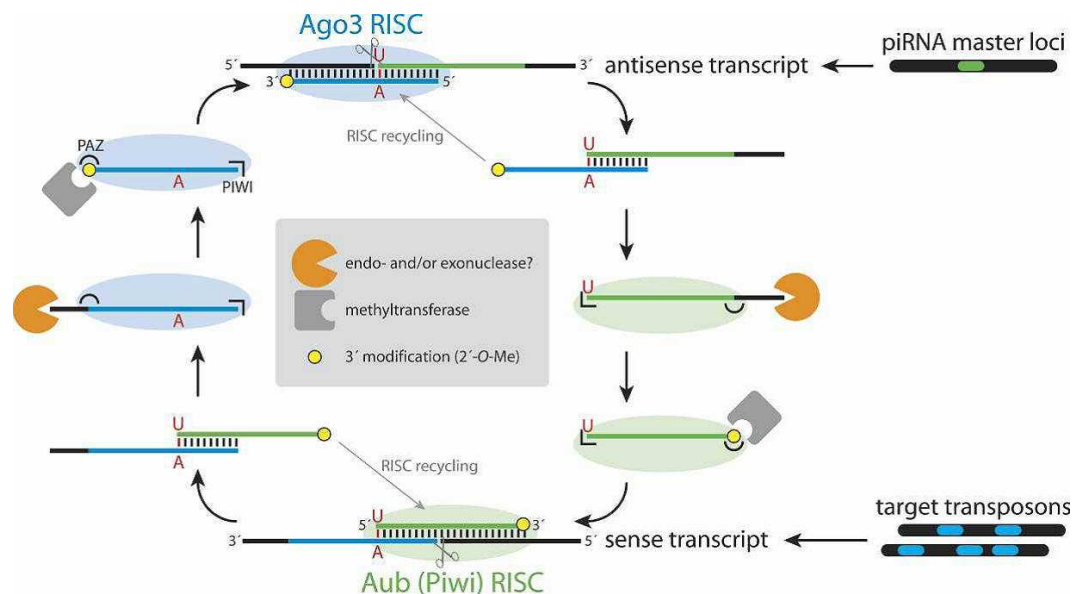


**Figura 14.** Función de los lncRNA. . a) Los lncRNA reclutan complejos modificadores de cromatina a loci específicos. *HOTAIR*, *Xist* (o *RepA*) y *Kcnqot1*, guían al complejo represor “Polycomb” (*PRC2*) hacia el locus *HoxD*, el cromosoma X, o el dominio *Kcnq1*, respectivamente. El *PRC2* trimetila los residuos de lisina 27 (*me3K27*) de las histonas H3, induciendo la formación de heterocromatina y reprimiendo así la expresión génica. b) lncRNAs anclados al gen de ciclina D reclutan a la proteína de unión a RNA (*TLS*) para modular la actividad de acetiltransferasa de histona de la proteína de unión a *CREB* (*CBP*) y *p300*, reprimiendo la transcripción. c) Un potenciador ultraconservado se transcribe como lncRNA, *Evf2*, que actúa como co-activador del factor de transcripción *DLX2*, regulando la transcripción de *Dlx6*. d) Un lncRNA transcrito del promotor menor de *DHFR* forma una triple hélice con el promotor mayor, impidiendo la unión del factor de transcripción *TFIID*, silenciando el gen de *DHFR*. e) Un ncRNA antisentido oculta el sitio de splicing 5' del mRNA *Zeb2* (de la homeocaja dedo de zinc) del spliceosoma, reteniendo el intrón. El ribosoma reconoce un sitio interno de entrada al ribosoma (*IRE*) en el intrón retenido, lo que resulta en una traducción y expresión eficiente de *Zeb2*. Tomado de [47]

### I. II. III. V. RNA de interacción con proteínas Piwi (piRNA)

Estas secuencias, de 25 a 27 nt, e interactúan con proteínas *Piwi*, una subfamilia de las proteínas Argonauta. Se encuentran activas en células germinales y silencian genes mediante la formación de heterocromatina por metiltransferasas, inhibiendo además la actividad de elementos genéticos móviles, contrarrestando su actividad mutagénica [49]. Son procesadas secuencialmente después de la transcripción mediante ribonucleasas distintas a *Dicer* y se acoplan directamente a la proteína *Piwi* o *Aub* dentro del complejo RISC, igual que los siRNA y miRNA (Fig. 15) [50].

La actividad represiva en células germinales ha sido llamado el “sistema inmune” del genoma, ya que identifican a las secuencias genéticas móviles que atentan contra la integridad de la información del genoma [49]. Se desconoce cómo se inactivan los transposones en adulto ya que no se han detectado piRNA en células somáticas, pero parece que la formación de heterocromatina mediada por piRNA durante la embriogénesis se mantiene a lo largo de la vida del organismo, quizá mediante los siRNA, ya que comparten el complejo RISC [51].



**Figura 15** Biogénesis de piRNA.- (Inferior) Los transcritos sentido provenientes de transposones, son cortados por Piwi o Aub y cargadas dentro de RISC con un piRNA que sirve de guía. El transcrito cortado no solo es degradado, se usa para programar a Ago3 en RISC. (Superior) Este complejo a su vez corta los transcritos antisentido que se originan del loci que funge como control maestro. Nuevamente el RNA cortado sirve para programar a PIWI o AUB dentro del complejo RISC. Por tanto, los transcritos sentido y antisentido alimentan un ciclo de amplificación en el que los extremos 5' son definidos por la forma de cortar de RISC. Se piensa que los extremos 3' son acortados mediante endonucleasa y/o exonucleasa hasta el tamaño adecuado que se acomode entre los dominios PAZ y PIWI. El extremo 3' es modificado subsecuentemente (2'-O-Me) por una metiltransferasa llamada Pimet/DmHen1 en *Drosophila*. Tomado de [50].

## I. II. III. VI. Riboswitches

Estas secuencias actúan como sensores de metabolitos específicos, que al unirse a ellos son capaces de encender/apagar la maquinaria transcripcional de genes asociados al metabolismo de la molécula que reconocen. Son de una afinidad y especificidad excepcionales, encontrándose ampliamente distribuidos y conservados en procariontes, e incluso en algunos eucariontes [52].

El mecanismo por el cual regulan la expresión génica involucra la formación de estructuras alternativas de su secuencia. En la conformación represora, provocan la terminación prematura de la transcripción al formarse un tallo de terminación cuando el metabolito se une, imposibilitando la elongación del transcrito. En el caso de la traducción, la unión del metabolito enmascara las secuencias Shine-Dalgarno dentro de la estructura secundaria, lo que impide la unión del ribosoma, inhibiendo el comienzo de la síntesis de la proteína.

La diversidad de genomas que contienen riboswitches y la variedad de mecanismos moleculares de regulación, junto con el hecho de que la interacción directa con sus moléculas efectoras no requiere de cofactores y es de una altísima afinidad y especificidad, sugiere que los riboswitches representan uno de los sistemas regulatorios más remotos [53].

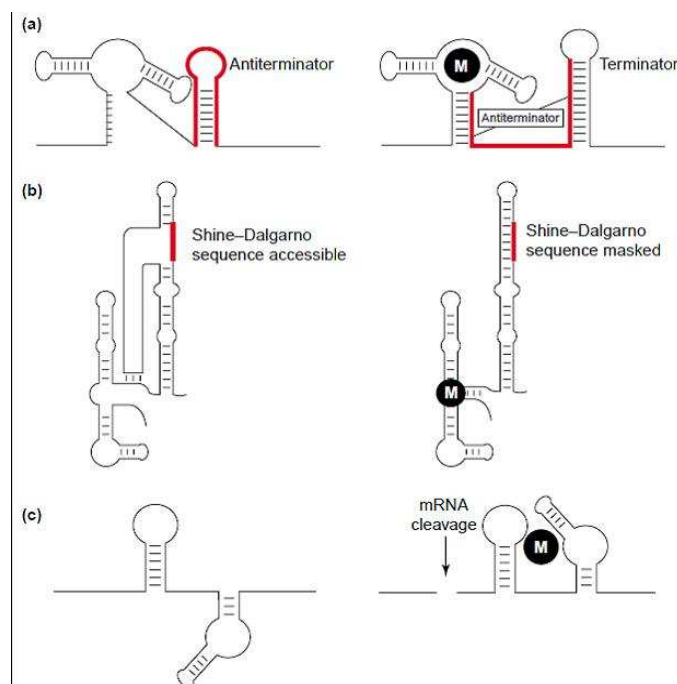


Figura 18 *Mecanismos de función de los riboswitches.*- A) Término de la transcripción inducido por la unión de un metabolito (M) a un RNA naciente, como ha sido observado en el riboswitch de guanina. B) Inicio modulado de la traducción mediante el enmascaramiento de la secuencia Shine-Dalgarno inducido por la unión de un metabolito, como el caso el riboswitch de pirofosfato de tiamina. C) Procesamiento del RNA mediado por un auto-corte dependiente de la unión del metabolito, como se observa en el riboswitch de Glucosamina-6-fosfato. Tomado de [52].

A la luz de la clasificación anterior, vemos que el RNA no es sólo un intermediario entre el DNA y las proteínas, sino un actor de gran importancia dentro de diversos fenómenos al interior de la célula. Contribuye a la organización estructural y dinámica de los cromosomas, procesos epigenéticos, compensación de dosis y la impronta genómica en animales [54]. También participa dentro del sistema inmune (principalmente los miRNA), al mediar el desarrollo y diferenciación de células B y T, la proliferación de neutrófilos y monocitos, el switch de anticuerpos y la liberación de mediadores inflamatorios [55, 56]. Además, otras evidencias sugieren que el RNA podría suplementar la señalización endocrina y paracrina llevada a cabo por hormonas, metabolitos y proteínas [57].

#### ***I. II. IV. ncRNA y enfermedad***

Todas las clases anteriores de ncRNA tienen relación con alguna enfermedad. Algunos ejemplos son; la hipoplasia del cartílago del cabello, ataxia espinocerebelar tipo 8, síndrome de DiGeorge, autismo, Alzheimer, esquizofrenia, enfermedades de origen inmunitario, diabetes y cáncer [24]. Mencionemos dos ejemplos.

***Antígeno Prostático.*** Este antígeno, empleado rutinariamente en análisis de diagnóstico para cáncer de próstata, es una molécula de RNA. Se sobreexpresa en la mayoría de los cánceres de próstata, y se expresa a un nivel mínimo en tejido prostático no maligno. No se expresa en otros tejidos, normales o malignos [58]. Sufre de modificaciones pos-transcripcionales, ya que el gen mide cerca de 20 kb pero sólo se detectan transcritos de 0.6, 2 y 4 kb. Se encuentra en el cromosoma 9 (9q21-22), y fue descrito por primera vez en 1999 [59]. Otro gen de RNA involucrado en cáncer de próstata es PCGEM1, que también se sobreexpresa en cáncer de próstata, y se encuentra en el cromosoma 2 [60]. Paradójicamente, no se sabe el papel que juegan estas dos moléculas en el cáncer de próstata a pesar de estar claramente sobreexpresados.

***Diabetes.*** Los miRNA han sido relacionados con diabetes. Datos recientes sugieren que participan en la secreción de insulina, el desarrollo de las isletas pancreáticas, la diferenciación de las células beta, y que indirectamente controlan el metabolismo de lípidos y glucosa, además de estar involucrados en complicaciones secundarias de este padecimiento [61]. Por ejemplo, el knockout de *Dicer 1*, produce defectos en todo el linaje pancreático en ratones, mientras que el miRNA mir-375, muy

abundante en células beta e isletas pancreáticas, regula negativamente la secreción de insulina estimulada por glucosa, ya que la sobreexpresión de este RNA inhibe completamente la secreción de insulina [61].

### ***I. II. V. ncRNA y complejidad biológica***

En procariontes, la mayor parte del genoma codifica proteínas, mientras que en eucariontes multicelulares el porcentaje de estas secuencias es más reducido. Así, al observar varias especies, parece existir una transición cualitativa del contenido de la carga transcripcional entre procariontes y eucariontes [25]. Por otro lado, el proyecto del genoma humano reveló que sólo un ~1.2% del genoma se traduce a proteínas (corresponde a exones), mientras que un 93% del genoma se transcribe activamente, por lo que se plantea que la mayor parte de la carga transcripcional podría consistir de ncRNA [19, 62].

Debido a lo anterior, los ncRNA podrían ser indicadores del nivel de complejidad de las especies. Así, organismos “inferiores” como las levaduras contienen un número menor de estas secuencias en comparación de los organismos multicelulares pequeños como los gusanos, y éstos a su vez contienen menos cantidad que uno “superior” como un mamífero. Es el humano, quien tiene el mayor número de secuencias ncRNA de todos los organismos analizados hasta ahora (Fig. 19) [14].

Mattick, uno de los principales promotores de los ncRNA, plantea que en procariontes el número de proteínas reguladoras escala casi exponencialmente respecto al tamaño del genoma, lo que impone un límite a la complejidad de estos organismos [14]. Por su parte, los eucariontes sobrepasarían este límite al emplear RNA como una solución regulatoria en forma “digital”, concertada con la evolución de una infraestructura proteica necesaria para reconocer y actuar sobre este tipo de señales [14]. Esta hipótesis (controversial a causa de lo arbitrario de la definición de complejidad<sup>1</sup>), ilustra algo importante (aun ignorando lo antropocéntrico de la idea); que hay un aumento en el número de secuencias de ncRNA a medida que nos desplazamos por las distintas ramas del árbol filogenético de todos los seres vivos.

---

<sup>1</sup> Este mismo término es empleado para describir varios fenómenos o cualidades biológicas.



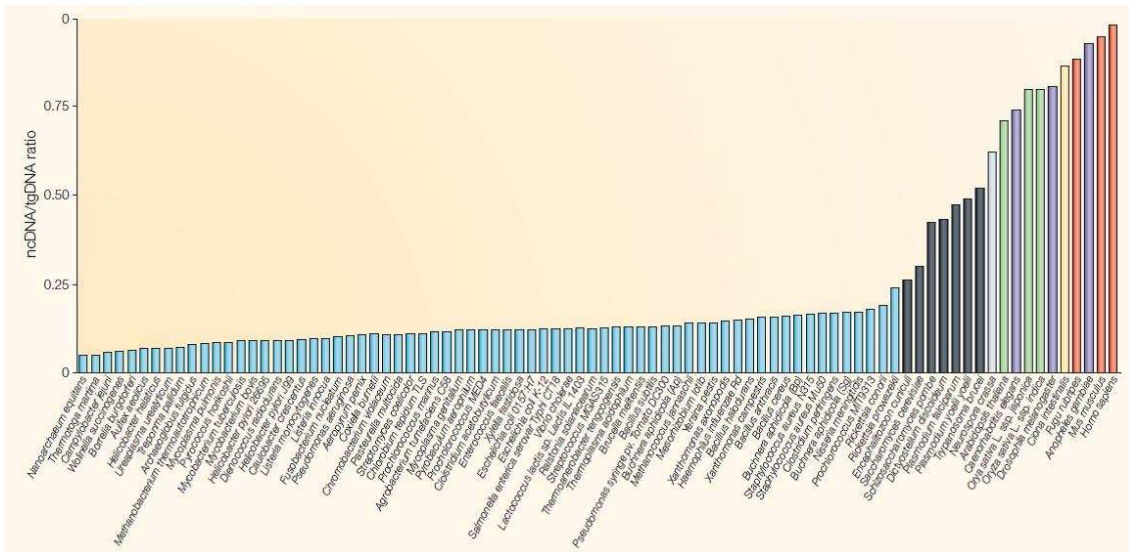


Figura 19 La proporción de DNA no-codificante respecto del DNA codificante de proteína se incrementa como función de la complejidad del desarrollo. Los procariontes contienen menos de un 25% de DNA no-codificante, los eucariontes simples contienen entre 25 y 50%, mientras que organismos más complejos como hongos, plantas y animales tienen más de 50%, llegando hasta 98.5% de DNA no-codificante en humanos. El azul representa procariontes (Archea y Bacteria), el negro eucariontes simples, *Neurospora crassa* en gris, plantas (verde), invertebrados no cordados (nemátodos, insectos) en púrpura, *Ciona intestinalis* (urocordado) en amarillo, y vertebrados (rojo). ncDNA, DNA no-codificante; tgDNA, DNA genómico total. Tomado de [14].

La forma en que actuaría la regulación “digital” mediante ncRNA sería:

- 1) Los genes eucariontes expresarían dos tipos de información en forma paralela; las proteínas y señales de RNA.
- 2) Estas secuencias de RNA estarían sujetas a una selección que las forzaría a expandirse en organismos complejos.
- 3) Algunos ncRNA serían procesados pos-transcripcionalmente, generando un mayor número de señales que actuarían sobre distintos blancos en varias redes de regulación.
- 4) Al mandar señales epigenéticas “corriente abajo” (*downstream*) en el sistema, a la larga podrían fijar trayectorias en el desarrollo, al ser guiadas o activadas por señales ambientales que proporcionarían un contexto de expresión y así corregir la activación estocástica del programa de desarrollo.
- 5) Gracias a la especificidad hacia su secuencia blanco (DNA o RNA), serían captadas por proteínas y/o complejos que reconocieran la estructura secundaria o tridimensional de las interacciones y actuar en consecuencia, permitiendo una regulación exacta de la expresión génica y de procesos epigenéticos.

Mattick propone entonces que el ncRNA es el causante de toda la variedad taxonómica que se dio desde la explosión del Cámbrico (conocida como radiación de los Metazoarios), en donde surgieron los ancestros de todas las *phylas* que existen actualmente (Fig. 20) [14]. En el caso de humanos, los ncRNA representarían quizá una de las claves de su complejidad, al menos así lo sugiere un estudio que muestra como un RNA (HAR1F) está involucrado en el desarrollo del sistema nervioso y cuyo locus genómico es una de las regiones que han evolucionado muy rápidamente en humanos respecto a los demás primates [63]. Siendo así, tanto el programa del desarrollo como la diferencia fenotípica entre especies e individuos está influenciada, si no es que fundamentalmente controlada, por el repertorio de ncRNA [25].

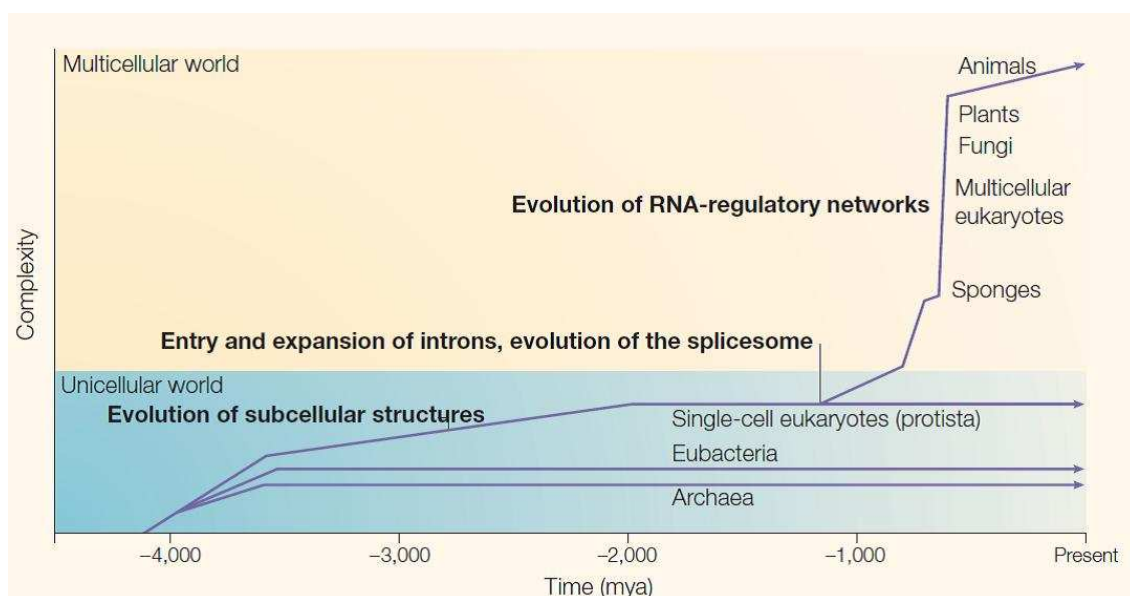


Figura 20 **Historia biológica simplificada de la Tierra.**- Esta gráfica intenta ser un bosquejo. Algunas fechas siguen en debate y la abcisa de complejidad (Complexity) tiene una escala arbitraria. La vida aparece en la Tierra hace 4,200 millones de años aproximadamente, y consiste (o se diversifica rápidamente) en tres reinos: Eukarya, Bacteria y Archea. La vida se mantiene como unicelular (o en pequeñas colonias) por al menos otros 3,000 millones de años. Se piensa que el ancestro común de animales, plantas y hongos surgió hace 1,200 millones de años, más o menos el tiempo también en el que se cree se dio la relación endosimbiótica que dio origen a las mitocondrias. A partir de este punto, la incidencia de intrones y de otras secuencias de DNA no codificantes correlacionan con la complejidad del organismo. Hace 520 millones de años, en el periodo Cámbrico, evolucionó la vida animal compleja en el evento mejor conocido como la “radiación de los metazoarios”, en la que ancestros reconocibles de todas las “phyla” modernas aparecieron y cuyo registro fósil se encuentra en un solo estrato rocoso. ¿Qué tuvo la aparición de organismos multicelulares complejos por tanto tiempo? ¿Fueron factores ambientales o bioquímicos (como la concentración de oxígeno y metabolismos energético asociado a éste)? ¿O era acaso debido a sistemas operativos genéticos “primitivos”? Tomado de [14].

## ***I. II. VI. Análisis computacional de los ncRNA***

Hay varios programas y métodos para tratar de identificar ncRNA's en secuencias genómicas, y operan principalmente bajo dos formas; i) basados en homología a ncRNA's ya reportados, y ii) los que están basados en la predicción *ab initio* de estructura secundaria característica del RNA bajo ciertos parámetros físico-químicos y de emparejamiento de bases nucleotídicas. Estos métodos han arrojado una vasta cantidad de secuencias posibles que están en etapa de verificación experimental. El estudio computacional de ncRNA es una de las áreas más activas en lo que se llama Genómica del RNA (*RNomics*), siendo una de las fronteras de la era pos-genómica que intenta resolver la estructura y función del ncRNA a escala genómica [64].

La predicción de genes putativos de ncRNA ha sido muy intensa en el caso de microRNAs y siRNAs, cuya secuencia presenta patrones reconocibles que son empleados al analizar secuencias genómicas. Es deseable que cada miRNA o siRNA candidato predecido computacionalmente necesite de la validación experimental [65]. Los enfoques empleados son:

- 1) Usar información de estructura secundaria, ya que la presencia de estructura “*fold-back*” es una característica esencial de los pre-miRNAs.
- 2) Se sobreestima la conservación de la secuencia así como de la estructura en la distinción de candidatos de miRNA de hairpinas genómicas irrelevantes.
- 3) Se mide la estabilidad termodinámica de las hairpinas y del complejo miRNA-mRNA.

Como ya se mencionó, las señales de RNA son transmitidas principalmente mediante interacciones específicas con otros RNA y DNA. Este tipo de interacciones posiblemente tengan estructuras secundarias y/o tridimensionales en forma de triple hélice o de otro orden que no obedecen las reglas canónicas de emparejamiento de bases, lo cual necesitaría el desarrollo de nuevos algoritmos capaces de detectar este tipo de interacciones partiendo simplemente de las secuencias genómicas [14].

En este trabajo empleamos el enfoque de búsqueda de homología a ncRNA's ya reportados al ejecutar el programa *BLAST* sobre las secuencias de *T. solium* disponibles.

## ***I. II. VII. Bases de datos de ncRNA***

Existen varias bases de datos dedicadas exclusivamente a la clasificación de secuencias de ncRNA. Las siguientes cuatro son las principales; i) NONCODE, ii) miRBase, iii) RNADB, y iv) Rfam. Cada una se especializa en cierto tipo de ncRNA y tiene diversas características de anotación. Enunciaremos las características de cada una de ellas a continuación, con especial énfasis en NONCODE y miRBase, que fueron empleadas para la comparación con las secuencias de *T. solium*.

### *NONCODE:*

- Contiene casi todos los tipos de ncRNA;
- Toda la información relacionada (función, localización celular, información cromosomal, etc.) ha sido curada manualmente mediante literatura, y cerca del 80% de las entradas están basadas en datos experimentales;
- Tiene un sistema de clasificación original basado en el proceso y función celular de cada secuencia denominado PfClass (Process Function Class) que sirve para integrar la información a sistemas de clasificación existentes;
- Asimismo tiene un sistema de anotación que indica si el ncRNA en cuestión está asociado específicamente a género, tejido, etc., y si está relacionado con tumores, enfermedades, etc.;
- Es fácil de usar y permite recabar fácilmente información adicional como los elementos regulatorios en la zona adyacente de la secuencia, su estructura secundaria, publicaciones relacionadas y cualquier otra información disponible, además de poder obtener las secuencias en formato FASTA desde su sitio en internet. Fue creada en China, se actualiza cada seis meses, y es de acceso libre a través de <http://noncode.bioinfo.org.cn> [66].

### *miRBase*

Esta base fue creada por el Instituto Sanger en el Reino Unido. Se especializa en los miRNA's, y actualmente contiene 5071 miRNA loci de 58 especies, expresando 5922 secuencias de miRNA maduras y distintas. Tiene dos conjuntos de secuencias, una consiste de los precursores de miRNA, por lo que casi todas las secuencias tienen un

tamaño aproximado de 70 nucleótidos. El otro conjunto contiene solamente las secuencias del miRNA maduro y funcional, por lo que el tamaño de cada secuencia va de 19 a 22 nucleótidos. Esta base también permite obtener los archivos con todas las secuencias en formato FASTA desde su servidor en internet, el cual es <http://microrna.sanger.ac.uk/sequences/> [67].

*RNAdb*. La base RNAdb se caracteriza por que sólo se concentra en predicciones computacionales a partir de transcriptomas de humano y ratón, por lo que consta de una gran cantidad de ncRNA putativos (sobre todo miRNA's), que están en etapa de comprobación experimental o indicio de una función comprobable [68] \*.

*Rfam*. Es una colección de familias de secuencias de RNA representadas por múltiples alineamientos de secuencia y modelos de covarianza. Esta base no sólo contiene ncRNA, sino también RNA estructural (transferencia y ribosomales), además de que su enfoque es la anotación de nuevos candidatos de familias de RNA ya conocidas y el programa en el que corre trabaja particularmente sobre secuencias de genomas completos [69].

Recordemos que estas bases no son las únicas que hay pero si las más representativas dentro del estudio del ncRNA. Cada año surgen nuevas bases de datos, desde las generales hasta las muy especializadas, y dependiendo de su utilidad, popularidad y derechos de uso, se hacen referencia obligada, van integrándose unas con otras, se privatizan o desaparecen. Cada año la revista *Nucleic Acids Research* edita un número especial dedicado a describir (en voz de sus creadores) todas las bases disponibles, presentando las nuevas y enumerando las actualizaciones de las ya existentes.

Es así como el RNA se ha posicionado nuevamente como actor principal en muchos procesos biológicos de gran importancia, por lo que fue nominada como el avance del año por la revista *Science* [70]. En consecuencia, hay un aumento considerable en el número de bases de datos referentes a sus secuencias, blancos, interacciones, asociaciones con enfermedades etc.<sup>‡</sup>. Incluso la enciclopedia en línea

---

\* La base RNAdb actualmente está en la versión 3 y ahora se llama fRNAdb (de “functional RNA database”) y fusiona a todas las bases de ncRNA (*Nucl. Acids Res.* 2009 37: D89-D92). Esta base no fué usada, tampoco Rfam.

‡ Basta echar un vistazo a la página de la revista *Nucleic Acids Research* en su edición de bases de datos para darse cuenta de ello (*Nucl.AcidsRes.*2009 37 Database Issue) [http://nar.oxfordjournals.org/content/vol37/suppl\\_1/index.dtl](http://nar.oxfordjournals.org/content/vol37/suppl_1/index.dtl)

*Wikipedia* tiene el proyecto *RNA WikiProject* que es parte de un *WikiProyecto* más amplio de biología molecular y celular (asociado a la base de Rfam), con cerca de 600 artículos describiendo distintas familias de RNA\*. Aunque este último no es de un nivel de especialización tan elevado, es de gran importancia debido a que hace interactuar de una manera más fluida y asequible la información valiosa sobre los ncRNA, siendo reconocida en *journals* tan especializados como *RNA* [71].

El siguiente recuadro presenta las cuatro bases de datos principales de ncRNA.

Nombre	Especialidad	Número Secuencias	Comentarios
NONCODE	ncRNA de toda clase	5,339	Curada manualmente mediante literatura que respalde experimentalmente la existencia de cada secuencia. Sistema de clasificación que proporciona información biológica.
miRBase	microRNA y sus precursores	15,172	Curada manualmente mediante literatura. Proporciona putativos "targets" sobre los que actúan los miRNAs (predicción computacional)
RNAdb	ncRNA mamífero ( <i>Homo sapiens</i> , <i>Mus musculus</i> )	>10,000	La mayor parte de las secuencias son de humano o ratón y son predicciones computacionales.
Rfam	ncRNA toda clase Además de RNA estructural.	379 (familias)	Clasifica los ncRNA en familias mediante perfiles de estructura secundaria y alineamientos múltiples mediante gramáticas estocásticas libres de contexto.

Para terminar esta breve revisión de ncRNA. sólo resta decir que la existencia de secuencias de RNA con función reguladora fue predicha hace casi 50 años por Jacques y Monod al hablar de genes estructurales (que producen proteínas) y genes reguladores (como lacI), cuya complementariedad de bases le permitiría interactuar específicamente con otras secuencias de ácido nucléico, modulando su expresión [72].

\* \* \*

\* [http://en.wikipedia.org/wiki/Wikipedia:WikiProject\\_RNA](http://en.wikipedia.org/wiki/Wikipedia:WikiProject_RNA)

## II. PLANTEAMIENTO DEL PROBLEMA

En base a la revisión que sobre *Taenia solium* y ncRNA se hizo en la sección anterior, cabe la siguiente pregunta de investigación: ¿Qué genes de ncRNA se encuentran dentro del genoma de *T. solium*? Dados los distintos conjuntos de secuencias producidos y disponibles por el consorcio del genoma de *T. solium*, es posible realizar en ellas una búsqueda *in silico* de secuencias similares a ncRNA ya reportadas y que se encuentran depositadas en bases de datos especializadas. El encontrar ncRNA en un metazoario como *T. solium* podría contribuir a entender su origen, evolución, y especialmente, componentes de los procesos moleculares que participan en la infección humana por este organismo.

## III. OBJETIVOS

- Identificar, dentro del genoma de *T. solium*, secuencias similares a aquellas conocidas como ncRNA, mediante alineamientos *in silico* de secuencias a través de la herramienta informática *BLAST*.
- Comparar cada uno de los distintos lotes de secuencias proporcionados por el consorcio del genoma de *T. solium* contra dos bases de datos de ncRNA; NONCODE y miRBase.
- En caso afirmativo, determinar qué tipo de secuencias de ncRNA son y cuál es su función biológica de acuerdo a la anotación existente, así como la información que pudiera existir sobre la secuencia de *T. solium* que alineó con la de ncRNA. Ambos casos mediante comparación contra las bases de NCBI y Gene Ontology (GO).

## IV. JUSTIFICACIÓN

En México, el parásito *Taenia solium* persiste como un tema de salud pública. Aunque ha sido objeto de estudio desde hace siglos y es de los organismos que más se han estudiado, se ignoran todavía aspectos moleculares de su biología y de los mecanismos involucrados en la infección por este organismo. Como estudio complementario dentro del Consorcio del Genoma de *T. solium*, la finalidad de este trabajo radica en ser un primer esfuerzo por conocer más, desde la perspectiva del ncRNA, la biología de este organismo y así complementar el conocimiento derivado del estudio de proteínas y secuencias de DNA.

Al revisar la literatura, se percata la ausencia de estudios previos sobre ncRNA's en *T. solium*. Además, la anotación hecha hasta ahora por el consorcio del genoma de *T. solium* se ha enfocado a proteínas, por lo que este trabajo se dedica exclusivamente a la búsqueda de ncRNA.

La disponibilidad del genoma de *T. solium*, junto con la existencia de bases de datos especializada en ncRNA, y las facilidades de cómputo intensivo dentro de la universidad, hacen posible la búsqueda *in silico* de esta clase de secuencias dentro del genoma. Aunque esta metodología no es suficiente para concluir que una secuencia tenga una función biológica real, si es necesaria, debido a que en muchos casos la identificación *in silico* precede la identificación *in vivo*. Además, representa una ventaja ya que mejora el diseño de experimentos al delimitar de una mejor manera el área de investigación, y en consecuencia, reducir el tiempo y costo de un proyecto. Igualmente, recordemos que esta metodología está bien cimentada y se usa de manera rutinaria en la anotación de genomas recién secuenciados.

Se espera que los resultados encontrados en esta tesis sean útiles en el diseño de experimentos que permitan descubrir otras moléculas de ncRNA propias de *T. solium* desconocidas hasta la fecha.



## V. MATERIAL Y MÉTODOS

**Software de alineamiento de secuencias.** Para el alineamiento de secuencias se empleó el programa *BLAST* (versión 2.2.18) para sistema operativo Linux instalado en equipo de cómputo intensivo del Grupo de Biología Teórica del Instituto de Investigaciones Biomédicas. Se trabajó remotamente mediante aplicación SSH (*putty*) en entorno de Windows. Dentro del programa *BLAST*, se utilizó la subrutina *BLASTn* para realizar los alineamientos de dos bases de ncRNA contra cuatro lotes de secuencias proporcionados por el Consorcio del Genoma de *Taenia solium* (CGTs).

**Bases de datos de ncRNA.** Descargadas de su respectivo sitio web, éstas fueron:

- NONCODE versión 1.0\* (compilado de ncRNA de varias clases, China)
- Hairpin versión 16 (precursores de miRNA, Instituto Sanger)

**Secuencias de *Taenia solium*.** Los archivos de secuencias *T. solium* fueron proporcionado por el CGTs en formato FASTA. A continuación se enumera cada uno de ellos junto con una breve descripción.

<i>Nombre archivo</i>	<i>Procedencia</i>	<i>Descripción</i>
1. EST larva	(CGTs UNAM)	Lote de secuencias procedentes de EST's <sup>▲</sup> de estado larvario. Consta de 9,157 secuencias, con una longitud promedio de 380 nt.
2. EST adulto	(CGTs UNAM)	Lote de secuencias procedentes de EST's de estado adulto. Consta de 14,113 secuencias, con una longitud promedio de 488 nt.
3. Full-length	(Ensamble Japonés)	Lote de secuencias procedentes de EST, realizado en Japón en colaboración con Raul Bobes del IIB. Consta de 12,260 secuencias, con una longitud promedio de 1,029 nt.

---

\* Aunque actualmente está ya en versión 2.0, simplemente anexa otras bases de RNA, por lo que decidimos no usarla para la comparación. No obstante, la anotación de las secuencias sigue intacta en esta segunda versión.

▲ *Expressed Sequence Tag* (Marcador de secuencia expresada)

4. 454AllContigs2010 (CGTs UNAM) Ensamble computacional de la pirosecuenciación 454. Consta de 41,280 secuencias, con una longitud promedio de 3038 nt.

***Procedimiento:***

- a) Crear base de datos de cada archivo de *Taenia* mediante línea de comandos para poder ser empleados por el programa *BLAST*. Por otra parte, aleatorizar secuencias de ncRNA con las cuales generar los alineamientos “control”.
- b) Correr el programa *BLAST* en su forma de comparación de nucleótidos (*BLASTn*). Cada una de las dos bases de ncRNA fue comparada contra los 4 archivos de *T. solium* (tanto normales como aleatorizadas). Se utilizó el valor *E* de las secuencias control (aleatorizados) para determinar el valor de corte para las secuencias normales. Se ordenó también a *BLASTn* presentar el resultado en formato tabular y guardarlo en un archivo con formato de texto.
- c) El resultado de cada comparación fue importado en Excel 2003, y se le dio un formato mediante herramientas de esta aplicación para poder manipular y visualizar los distintos puntajes que presenta *BLASTn* en formato de tabular.
- d) Mediante filtros de Excel, se extrajo el número de ncRNA distintos y únicos de cada archivo de resultados.
- e) Se buscó información relacionada a cada ncRNA en su respectiva base de datos de origen.
- f) Para cada secuencia de ncRNA que alineó con el genoma de *T. solium*, así como para cada contig de *Ts* con que hubo alineamiento de ncRNA, se realizaron *BLASTn* adicionales contra las bases de datos Gene Ontology y NCBI.

A continuación se muestra un diagrama de flujo en el que se representan los distintos pasos del procedimiento efectuado.

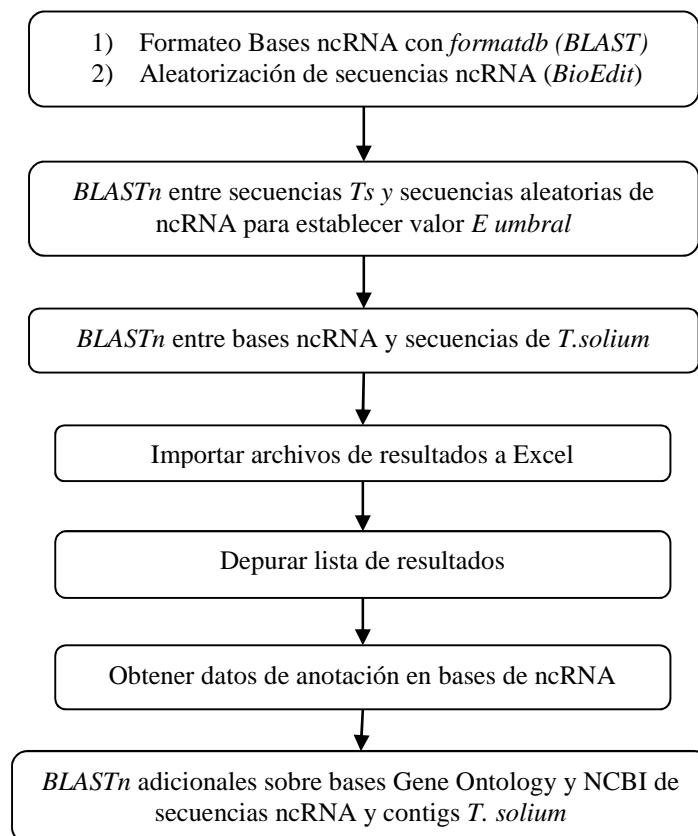


Diagrama de flujo del procedimiento llevado a cabo.

**Controles.** Una copia de las bases de miRBase y NONCODE fueron aleatorizadas mediante un comando del programa BioEdit, para romper cualquier motivo biológico dentro las secuencias. Cada una de las bases aleatorizadas resultantes fue comparada mediante *BLASTn* contra los cuatro archivos de *T. solium* para identificar la mejor puntuación que podría obtener un alineamiento entre una secuencia aleatorizada de ncRNA con una de *T. solium*. El valor *E* (*Expectation value* en inglés) usado para este ensayo de *BLASTn* fue de uno, ya que esto asegura exista uno o más alineamientos con puntajes igual o mejores que ocurran de manera aleatoria [73]. El alineamiento con mejor puntaje (valor *E* más pequeño) nos dirá a partir de qué valor nuestros alineamientos con secuencias normales serán significativos\*. A partir de esta calificación, decidimos reducir el valor *E* en cuatro órdenes de magnitud para los alineamientos contra las bases normales (no aleatorizadas). El valor de corte entonces asignado a *BLASTn* para presentar resultados fue de  $E < 10^{-9}$ . Esto último con la finalidad de tomar en cuenta el archivo de mayor tamaño de *Taenia* ( $4 \times 10^4$  en

---

\* A pesar de que no es probabilidad, el valor *E* está basado en un modelo estadístico y comúnmente se enuncia como la probabilidad de encontrar por azar el alineamiento reportado.

454Allcontigs), reduciendo así la probabilidad de que al azar, y por el tamaño del archivo contra el que se hace la búsqueda, encontremos un alineamiento aleatorio significativo.

**Depurado tabla de resultados.** Para simplificar la búsqueda de secuencias de ncRNA, sólo se tomó en cuenta el mejor alineamiento por cada secuencia de ncRNA encontrada (por ejemplo, el precursor de miRNA de *Homo sapiens* clasificado en miRBase como *hsa-mir-1300* generó alineamientos con once distintos *contigs* de *T. solium*, por lo que sólo se tomó en cuenta el de mejor puntaje). Aunque con esto pudiéramos estar descartando injustamente regiones del genoma en que se encuentran estas secuencias, se justifica porque varios de los *contigs* son redundantes todavía, algo que será mejorado en futuros ensamblajes del genoma. El procedimiento fue:

- a) Cada archivo de texto resultado de *BLASTn* fué acomodado en la aplicación *Excel* mediante el comando “Ordenar” usando las columnas “ID” y “E-value” con la finalidad de obtener el mejor alineamiento que hizo cada una de las secuencias de ncRNA contra las secuencias de *Taenia* (debido a que una sola secuencia de ncRNA puede hacer varios alineamientos con distintos fragmentos del genoma). Se copió esa fila e insertó en una nueva hoja de cálculo.
- b) La nueva tabla (resultado del procedimiento anterior), se volvió a ordenar pero ahora en base al ID Contig del genoma de *Taenia* y del valor *E* con la finalidad de mostrar aquellas secuencias de la base de datos de ncRNA que tienen nombre distinto pero que alinearon en el mismo contig del genoma de *Taenia*. Sólo se preservó la que alineó con mejor puntaje (un E-value más bajo).

**BLASTn adicionales contra las bases de datos Gene Ontology y NCBI.** Cada secuencia única de ncRNA que alineó con *T.solium* (resultado de *a* y *b* del procedimiento anterior), se usó como “sonda” (*query*) para realizar *BLASTn* versión *Web* dentro de las bases: 1) Gene Ontology (GO), 2) NCBI nt/nr y 3) NCBI EST. De igual forma, cada contig de *T. solium* con que alineó la secuencia de ncRNA fue empleada como “sonda” contra las bases de 1) GO y 2) NCBI.

**Uso de precursores de miRNA.** Aunque se pueden modificar los parámetros de BLAST para optimizar alineamientos de secuencias muy cortas y por tanto buscar directamente los miRNA “maduros”, en este trabajo se optó por emplear la secuencia

precursora del microRNA. Esto tiene cierto inconveniente, ya que se sabe que a diferencia de las secuencias “maduras”, los precursores de microRNA no son muy conservadas. No obstante, se optó por buscar estas últimas con la finalidad de encontrar aquellas que mostraran el mayor parecido tanto del precursor como de la secuencia madura<sup>1</sup>.

---

<sup>1</sup> Paralelamente se realizaron búsquedas optimizadas para microRNA maduro, encontrándose las secuencias maduras (entre otras) cuyos precursores se reportan en este trabajo.

## VI. RESULTADOS

De la base NONCODE obtuvimos 23 secuencias que produjeron distintos alineamientos distribuidos en los distintos archivos de *Taenia* como se ve en el listado al final de esta página. Al depurar la lista de resultados mediante la eliminación de alineamientos que se repiten o sobreponen en dos o más archivos de *Taenia*, obtenemos una lista de 17 secuencias únicas de NONCODE (Tabla A). Los resultados depurados de los alineamientos por cada archivo y de alineamientos contra las bases Gene Ontology y NCBI se presentan en las tablas R1A-R4A y R1B-R4B respectivamente, mientras que la tabla A muestra información adicional de anotación proveniente del sitio de NONCODE. Finalmente, tras la discusión hay una reducción más de resultados que nos deja con ocho secuencias de la base NONCODE (Tabla C).

Para el caso de los precursores de microRNAs (miRBase), se encontraron 7 alineamientos distribuidos en los distintos archivos de *Taenia* como se aprecia en el listado que sigue. Al depurar la lista de resultados mediante la eliminación de alineamientos que se repiten o sobreponen en dos o más archivos de *Taenia*, obtenemos una lista de cuatro secuencias únicas de miRBase. Los resultados depurados de los alineamientos por cada archivo y de alineamientos contra las bases Gene Ontology se presentan en las tablas R5A-R8A y R5B-R8B respectivamente, mientras que en la tabla B se muestra información adicional de anotación para cada una de ellas proveniente del sitio de miRBase.

<u>Base vs. Archivo <i>Taenia</i></u>	<u>ncRNA</u>
NONCODE vs. 454AllContigs2010.....	10
NONCODE vs. Adulto .....	4
NONCODE vs. Larva .....	5
NONCODE vs. Full-length .....	4
Hairpin vs. 454AllContigs2010 .....	3
Hairpin vs. Adulto .....	2
Hairpin vs. Larva .....	1
Hairpin vs. Full-length.....	1

ID NOCODE	E-value	% ncRNA	Class	pfClassFunction	Artículo
u1919,Taenia	5.00E-70	100.00%	snRNA U5	RNA_processing_splicing	The U5/U6 snRNA genomic repeat of <i>Taenia solium</i>
u1918,Taenia	3.00E-58	100.00%	snRNA U6	RNA_processing_splicing	The U5/U6 snRNA genomic repeat of <i>Taenia solium</i>
u4170,Echinococcus	4.00E-78	98.74%	snRNA U1	RNA_processing_splicing	Structure of the <i>Echinococcus multilocularis</i> U1 snRNA gene repeat
u3977,Schistosoma	1.00E-47	92.27%	snRNA U2	RNA_processing_splicing	<i>Schistosoma mansoni</i> : Characterization of U1, U2, and U6 small nuclear RNAs
u3976,Schistosoma	5.00E-27	89.24%	snRNA u1 snRNA	RNA_processing_splicing	<i>Schistosoma mansoni</i> : Characterization of U1, U2, and U6 small nuclear RNAs
u1468,Mus	9.00E-10	72.22%	snmRNA snmRNA	Miscfunction_snm	RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse
u3540,Erwinia	2.00E-31	63.55%	Replication Control RNAI	RNA_translation_suppression	Relatedness of Chromosomal and Plasmid DNAs of <i>Erwinia pyrifoliae</i> and <i>Erwinia amylovora</i>
u4043,Gallus	4.00E-16	54.61%	snRNA u4	RNA_processing_splicing	Structural and functional analysis of chicken U4 small nuclear RNA genes
u1139,Gorilla	9.00E-23	39%	snRNA U2	RNA_processing_splicing	Polarized junction diversification at the primate RNU2 locus and the implications for the mechanisms of concerted evolution
u1311,Arabidopsis	8.00E-15	29.20%	snoRNA snoR78	RNA_modification_methylation	Extensive characterization of methylation guide small nucleolar RNAs in <i>Arabidopsis thaliana</i>
u596,Oryza	9.00E-19	16.07%	Sn mRNA ITS1-IR36-S	Miscfunction_snm	No hay referencia en NONCODE
u298,Homo	5.00E-14	15.55%	scaRNA U90	RNA_modification_methylation	Toward a complete human genome sequence
u3429,Streptomyces	3.00E-23	2.33%	GcvB RNA gcvB	DNA_transcription_repression	Genome sequence of an industrial microorganism <i>Streptomyces avermitilis</i> : deducing the ability of producing secondary metabolites
u1741,Homo	7.00E-18	1.20%	UBE3A antisense RNA	Miscfunction_mRNAlike	The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species And as an antisense RNA for UBE3A
u1740,Homo	4.00E-18	1.06%	UBE3A antisense RNA	Miscfunction_mRNAlike	The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A
u5338,Mus	4.00E-33	0.30%	Air RNA	DNA_imprinting	The imprinted antisense RNA at the <i>Igf2r</i> locus overlaps but does not imprint <i>Mas1</i>
u2226,Microtus	3.00E-10	0.24%	XIST RNA	DNA_imprinting	Characterization of the genomic <i>Xist</i> locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence

**Tabla A.- Anotación de 17 secuencias de la base de NONCODE en el genoma de *Taenia solium*.** Se muestran en orden descendente respecto del porcentaje de la secuencia de ncRNA que alineó con *Ts* (columna %). *E* es el valor del mejor alineamiento que hizo la secuencia de ncRNA. La columna *Class* refiere la clasificación que recibe esta secuencia dentro de la base NONCODE, mientras que *pfClassFunction* indica el papel que juegan en la biología del organismo al que pertenecen. Por último, se da el título de la publicación que reporta la secuencia en cuestión (columna *Artículo*).

ID miRBase	E-value	S	% RNA	Evidencia	Artículo
sme-miR-749	3.00E-17	87.9	100%	experimental; cloned [1], 454 [2], Solexa [2]	[1] "MicroRNAs from the Planarian <i>Schmidtea mediterranea</i> : a model system for stem cell biology" [2] "High-resolution profiling and discovery of planarian small RNAs"
mdo-mir-739	4.00E-32	132	97%	not experimental	"miRNAmirer: a tool for homologous microRNA gene search"
ssc-mir-4331	4.00E-20	97.1	93%	experimental; Solexa [1]	[1] "Deciphering the porcine intestinal microRNA transcriptome"
gga-mir-3533	4.00E-19	89.8	87%	experimental; Solexa [1]	[1] "Identification of differentially expressed miRNAs in chicken lung and trachea with avian influenza virus infection by a deep sequencing approach"

**Tabla B.- Anotación de cuatro secuencias de la base de Hairpins en el genoma de *Taenia solium*.** Se muestran en orden descendente respecto del porcentaje de la secuencia de pre-miRNA que alineó con *Ts* (columna %miRNA). *E* es el valor del mejor alineamiento que hizo la secuencia de pre-miRNA. La columna *Evidencia* refiere a la forma en que fue identificada la secuencia, por homología (not experimental), por secuenciación masiva (Solexa, 454 etc.), o clonación. Al final se da la(s) referencia(s) donde proviene la secuencia en cuestión

ID NOCODE	E-value	% ncRNA	Score	Class	pfClassFunction	Artículo
u1919,Taenia	5.00E-70	100.00%	264	snRNA U5	RNA_processing_splicing	The U5/U6 snRNA genomic repeat of <i>Taenia solium</i>
u1918,Taenia	3.00E-58	100.00%	224	snRNA U6	RNA_processing_splicing	The U5/U6 snRNA genomic repeat of <i>Taenia solium</i>
u4170,Echinococcus	4.00E-78	98.74%	289	snRNA U1	RNA_processing_splicing	Structure of the <i>Echinococcus multilocularis</i> U1 snRNA gene repeat
u3977,Schistosoma	1.00E-47	92.27%	186	snRNA U2	RNA_processing_splicing	<i>Schistosoma mansoni</i> : Characterization of U1, U2, and U6 small nuclear RNAs
u3976,Schistosoma	5.00E-27	89.24%	121	snRNA u1 snRNA	RNA_processing_splicing	<i>Schistosoma mansoni</i> : Characterization of U1, U2, and U6 small nuclear RNAs
u3540,Erwinia	2.00E-31	63.55%	135	Replication Control RNAI	RNA_translation_suppression	Relatedness of Chromosomal and Plasmid DNAs of <i>Erwinia pyrifoliae</i> and <i>Erwinia amylovora</i>
u4043,Gallus	4.00E-16	54.61%	81.8	snRNA u4	RNA_processing_splicing	Structural and functional analysis of chicken U4 small nuclear RNA genes
u1139,Gorilla	9.00E-23	39%	107	snRNA U2	RNA_processing_splicing	Polarized junction diversification at the primate RNU2 locus and the implications for the mechanisms of concerted evolution

**Tabla C.- Anotación de las ocho secuencias de la base de NONCODE en el genoma de *Taenia solium* resultantes de la Discusión de resultados.** Es claro que todas excepto una pertenecen al spliceosoma (u3540), reflejando el amplio y estudio y caracterización de este sistema. [Se muestran en orden descendente respecto del porcentaje de la secuencia de ncRNA que alineó con *Ts* (columna %). *E* es el valor del mejor alineamiento que hizo la secuencia de ncRNA. La columna *Class* refiere la clasificación que recibe esta secuencia dentro de la base NONCODE, mientras que *pfClassFunction* indica el papel que juegan en la biología del organismo al que pertenecen. Por último, se da el título de la publicación que reporta la secuencia en cuestión (columna *Artículo*).]

En el caso de los controles (secuencias aleatorizadas), los dos mejores alineamientos - en porcentaje de identidad el primero y de longitud del alineamiento en el segundo-, tuvieron una identidad de 100% con longitud de 24 nucleótidos y 96% de identidad con 31nt de longitud respectivamente. Esto contrasta con los alineamientos obtenidos para las secuencias normales, ya que el de menor longitud fue de 34 nt con un 100% de identidad, apenas tres nucleótidos más largo que el mejor alineamiento aleatorio en longitud y de menor porcentaje de identidad. Asimismo no hubo demasiados “gaps” (espacios) para mejorar las puntuaciones, debido a que estas secuencias son en su mayoría de longitudes cortas (alrededor de 150 nucleótidos para la base NONCODE y 70 nt para la base de precursores de microRNA Hairpins, miRBase).

Respecto al valor *E*, el mejor alineamiento de las secuencias aleatorizadas fue de  $5 \times 10^{-05}$ , mientras que para las secuencias normales fue de  $4.00 \times 10^{-78}$ . El valor *E* de corte para reportar resultados fue de  $10^{-09}$ , ya que toma en cuenta el tamaño de la base de datos más grande (el archivo 454AllContigs con ~41,000 secuencias) lo que se consideró apropiado como margen de discriminación entre alineamientos de interés y aquellos encontrados aleatoriamente.

Cómo un parámetro adicional, se calculó el porcentaje de la secuencia de ncRNA que alineó a manera de índice de cobertura del alineamiento respecto de la



secuencia de ncRNA. Fue así como para cada secuencia (de NONCODE o Hairpins) se determinó la longitud de ésta y se calculó el porcentaje de ella que estaba contenida en el alineamiento respectivo con la secuencia de *T. solium*. Hubo casos donde la secuencia de ncRNA estaba contenida completamente (100 % para una secuencia de *Taenia solium* ya reportada por Flisser *et al.*), y otros en el que apenas llegaba a un .24 % (Microtus XIST RNA). En este aspecto, fue la base de precursores de miRNA “Hairpins” (miRBase) quien mostró los mejores resultados ya que el menor porcentaje fué de 29 % y el mayor de 100%.

En el caso de NONCODE la secuencia control aleatorizada u5338 de *Mus*, generó cientos de alineamientos con identidades de hasta 94% y longitud de hasta 139 nucleótidos con los cuatro archivos de *Taenia*. Específicamente, dos fragmentos de u5338 fueron los que alinearon innumerables veces, al hacer un *BLASTn* en el NCBI con la secuencia de cada uno de ellos, se obtuvo que eran secuencias repetitivas, en una caso combinación de G’s y A’s, y en el otro de C’s y T’s (datos no mostrados). Este hecho no se repitió con ninguna de las secuencias aleatorizadas de la base Hairpins. Como referencia, se colocaron al final de cada tabla de alineamientos que hubo con la base NONCODE.

Otras secuencias de NONCODE que alinearon, pero cuya cobertura de la secuencia era pequeña, al ser examinadas se vio que consistían de regiones repetitivas de dinucleótidos como AG’s y CT’s. Estas secuencias fueron u5338, u1740, u1741, u298, u2226.

Respecto a los *BLASTn* que se hicieron de las secuencias de ncRNA y contigs de *Taenia solium* que alinearon, contra las bases de Gene Ontology y NCBI, se puede observar que en muchos casos no se encontró una secuencia parecida dentro de la base Gene Ontology pero sí dentro de la base de NCBI nr/nt. Asimismo, a excepción de unos cuantos casos, la mayoría de los contigs de *T. solium* alinearon con secuencias de EST reportadas por el consorcio del genoma.

Otro aspecto de *BLASTn* contra GO y NCBI es que en varios casos el alineamiento que se da entre la secuencia del contig de *T. solium* y la base en cuestión, al hurgar la posición en la que se da el alineamiento, se ve que no hay parecido en las posiciones con las que alinea el ncRNA. Por ejemplo, un extremo de un contig de

*T.solium* alinea contra una secuencia de ncRNA, mientras que el otro extremo alinea contra una secuencia reportada en Gene Ontology o NCBI. De allí que en paréntesis se ponga si el alineamiento contiene a la secuencia de ncRNA en parte o completamente, o no en absoluto.

A continuación se muestran las tablas abreviadas con los resultados de BLASTn para los distintos archivos de *Taenia solium*. Para cada archivo de *T. solium* hay dos tablas, la primera (con número impar) muestra de manera resumida los resultados de BLASTn contra la base de ncRNA respectiva<sup>1</sup>. La segunda tabla (número par) muestra los resultados de BLASTn de aquellas secuencias de ncRNA y el contig de *T. solium* con que alinearon al compararlas contra las bases Gene Ontology y NCBI de una manera más escueta.

Esta es una descripción de lo que se ve en cada tabla, empezando con aquella que informa el resultado de BLAST al comparar una base de datos de ncRNA contra un archivo de *T. solium* (copiado de Tabla R1A).

NONCODE vs EST LARVA ( de Tabla R1A)									
	ID NOCODE	ID Contig <i>T.solium</i>	Id(%)	L	M	G	E-value	S	%
1	u1468,Mus	211006B_F16_TslvF07105.T3_052.ab1.fg0	86.15	65	9	0	9.00E-10	58	72%
	*u5338,Mus	230606B_G04_TslvG0433.T3_033.ab1.fg0	83.45	139	22	1	6.00E-15	85.7	

A) B) C) D) E) F) G) H) I) J)

- A) **Num (Numerador)**. Para indicar correspondencia entre cada par de tablas de resultados de *BLAST* y la posterior descripción del resultado en el texto.
- B) **ID (base de datos ncRNA)**.- Identificador de la secuencia ncRNA dentro de su base de datos correspondiente.
- C) **ID Contig**.- Identificador del contig dentro del archivo de *T. solium* correspondiente
- D) **Id (%)**. Porcentaje de identidad entre las dos secuencias alineadas
- E) **L**.- Longitud del alineamiento entre la secuencia de ncRNA *ID (base de datos ncRNA)* y la secuencia de *T. solium* “*ID Contig*”
- F) **M**.- Cantidad de bases no iguales dentro del alineamiento, pero que permiten continuar con el alineamiento.
- G) **G**.- Espacio que introduce *BLAST* para continuar un alineamiento local
- H) **E-value (Valor E)**.- Indica la “probabilidad” que de manera aleatoria se dé el alineamiento entre la secuencia “sonda” y la secuencia de la base de datos.
- I) **S**.- Es el puntaje (en bits) que da *BLAST* acerca del alineamiento.
- J) **%**.- Es el porcentaje de la secuencia ncRNA que está incluida dentro del alineamiento.

En la parte inferior de cada Tabla “xA” se encuentra el mejor resultado que obtuvieron las secuencias control (bases NONCODE y miRBase aleatorizadas). Se antepone un asterisco al ID de la secuencia así como un formato sombreado a la fila en que se encuentra.

<sup>1</sup> En el Apéndice se presentan las tablas completas generadas por *BLASTn* en formato tabular.

Ahora se describe el formato de las tablas nombradas como “xB”, que consistieron en realizar *BLAST* adicionales contra las bases de datos *online* del NCBI (*National Center for Biotechnology Information*) y **GO** (*Gene Ontology*) de :

- 1) La secuencia de ncRNA que alineó con *T.solium* (columnas D y E);
- 2) El contig de *Taenia solium* con que alineó el ncRNA del inciso anterior ( columnas F-G)

NONCODE vs GO y NCBI					Contig <i>Taenia</i> vs GO, NCBI y ESTs	
N	ID	Base	Información	Evalue	Información del Contig con que alinea	Evalue
1	u1468	GO	COX1 - Cytochrome c oxidase subunit I (protein from <i>Bos Taurus</i> )	3.6e-07	COX1 - Cytochrome c oxidase subunit I (protein from <i>Bos Taurus</i> ) si contiene tramo de u1468	3.6e-76
		NCBI	12S ribosomal RNA (parte de <i>Mus musculus</i> ATCC:CCL-1.3 clone LMTK-mitochondrion)	9e-39	Sus scrofa domestica mitochondrial DNA, complete genome (región corresponde a COX1)	0.0
		EST			LV0220025 <i>Taenia solium</i> UNAM-cd2_larva <i>Taenia solium</i> cDNA, mRNA sequence	0.0
A)	B)	C)	D)	E)	F)	G)

- A) Numerador. Para indicar correspondencia entre cada par de tablas de resultados de *BLAST* y la posterior descripción del resultado en el texto.
- B) **ID (base de datos ncRNA).**- Identificador de la secuencia ncRNA dentro de su base de datos correspondiente.
- C) **Base.-** Indica contra que base de datos (NCBI, GO u EST) se va a comparar la secuencia de ncRNA indicada en B). A partir de columna cada registro de A) y B) se divide en tres filas, una para cada base de datos.
- D) **Información.** Es la anotación que presenta la base de datos correspondiente para el mejor alineamiento que haga con la secuencia ncRNA indicada en B)
- E) **E-value.-** Muestra el valor *E* que corresponde al alineamiento entre la secuencia de ncRNA indicada en B) y la base de datos indicada en C)
- F) **Información Contig (*T.solium*) con que alinea.-** Es la anotación que presenta la base de datos indicada en C) para el mejor alineamiento que haga con el Contig de *T.solium* que alineó con la secuencia de ncRNA indicada en B).
- G) **E-value.-** Muestra el valor *E* que corresponde al alineamiento entre la secuencia del contig de *T. solium* que alineó con la secuencia ncRNA indicada en B) y la base de datos indicada en C)

## NONCODE

### NONCODE vs EST Larva.

NONCODE vs EST LARVA									
N	ID NOCODE	ID Contig <i>T.solium</i>	Id(%)	L	M	G	E-value	S	%
1	u1468,Mus	211006B_F16_TslvF07105.T3_052.ab1.fg0	86.15	65	9	0	9.00E-10	58	72%
2	u1311,Arabidopsis	092906n_J21_Ts_1_087.ab1.fg0	91.3	46	4	0	4.00E-10	60	34%
3	u596,Oryza	092906k_H18_Ts_1_073.ab1.fg0	100	45	0	0	9.00E-19	89.7	16%
4	u298,Homo	211006a_E08_TslvE0394.T3_019.ab1.fg0	100	37	0	0	5.00E-14	73.8	16%
5	u1740,Homo	031106A_M21_TslvE06132.T3_039.ab1.fg0	98	50	1	0	4.00E-18	91.7	1%
	*u3277,Agaricus	061006B_A03_TslvA0197.T3_017.ab1.fg0	100	21	0	0	4.00E-04	42.1	
	*u5338,Mus	230606B_G04_TslvG0433.T3_033.ab1.fg0	83.45	139	22	1	6.00E-15	85.7	

**Tabla R1A.** Resultado (resumido) de *BLASTn* al comparar la base de datos de ncRNA « **NONCODE** » contra el archivo de *T. solium* « **EST larva** ». En la parte inferior (sombreado) se muestra los dos mejores resultados obtenidos al ejecutar *BLASTn* entre la base NONCODE aleatorizada y el archivo de *Ts* « *EST larva* ». [ **ID\_NONCODE**=Identificador\_en\_NONCODE, **ID\_Contig**=Identificador\_contig\_*T.solium*, **ID(%)**=porcentaje\_identidad\_alineamiento, **L**=longitud\_alineamiento, **M**=mismatches, **G**=gaps, **E-value**=valor\_E, **S**= score\_(bits), **%**=porcentaje\_secuencia\_ncRNA (ver pág. 50).]

NONCODE vs GO y NCBI					Contig <i>Taenia</i> vs GO, NCBI y ESTs	
N	Id	Base	Información	Evalue	Información del Contig con que alinea	Evalue
1	u1468	GO	COX1 - Cytochrome c oxidase subunit I (protein from <i>Bos Taurus</i> )	3.6e-07	COX1 - Cytochrome c oxidase subunit I ( <i>Bos Taurus</i> ) si contiene tramo de u1468	3.6e-76
		NCBI	12S ribosomal RNA (parte de <i>Mus musculus</i> ATCC:CCL-1.3 clone LMTK- mitochondrion)	9e-39	Sus scrofa domestica mitochondrial DNA, compl. genome (región corresponde a COX1)	0.0
		EST			LV0220025 <i>Taenia solium</i> UNAM-cd2_larva <i>Taenia solium</i> cDNA, mRNA sequence	0.0
2	u1311	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NCBI	No encuentra similitud significativa	-----	Dilepis undula 28S ribosomal RNA gene, complete sequence ( no contiene al u1311, éste cae en microsatélite)	6e-28
		EST			LV0327034 <i>Taenia solium</i> UNAM-cd2_larva <i>Taenia solium</i> cDNA, mRNA sequence	0.0
3	u596	GO	NSE_0607 hypothetical protein (protein from <i>Neorickettsia sennetsu</i> str. <i>Miyayama</i> )	0.013	No encuentra similitud significativa	-----
		NCBI	Oryza sativa rDNA ITS1-IR36-S region	9e-144	<i>Taenia solium</i> 18S ribosomal RNA gene, complete sequence (contiene fracción de u596)	1e-42
		EST			LV0314036 <i>Taenia solium</i> UNAM-cd2_larva <i>Taenia solium</i> cDNA, mRNA	0.0
4	u298	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NCBI	snoRNA-U90-related	2e-120	Zebrafish DNA sequence from clone DKEY-250D11 (u298 alinea en región GT's)	6e-23
		EST			LV0329018 <i>Taenia solium</i> UNAM-cd2_larva <i>Taenia solium</i> cDNA, mRNA	0.0
5	u1740	GO	AT4G18670 (protein from <i>Arabidopsis thaliana</i> )	0.079	No encuentra similitud significativa	-----
		NCBI	Homo sapiens clone kid4 SNURF-SNRPN mRNA, downstream untranslated exons, alternatively spliced	0.0	No encuentra similitud significativa	0.0
		EST			LV0249059 <i>Taenia solium</i> UNAM-cd2_larva <i>Taenia solium</i> cDNA, mRNA sequence.	0.0

**Tabla R1B.** Resultado resumido de *BLASTn* comparando la secuencia de NONCODE de Tabla R1A (tercera y cuarta columnas) vs las bases Gene Ontology y NCBI nr/nt. Las dos columnas de lado derecho son de *BLASTn* comparando el contig de EST de *T. solium* que mejor alineó con la secuencia de ncRNA de NONCODE contra las bases a) Gene Ontology, b) NCBI nr/nt y c) NCBI EST.

*Anotación NONCODE, Gene Ontology y NCBI.*

- 1) *u1468*. Esta secuencia está anotada como *misfunction\_smn* en NONCODE (Tabla A) y fue la que mayor porcentaje de su secuencia estuvo contenida dentro del alineamiento (72% de 90 bp). Alineó con una subunidad del citocromo C oxidasa en la base GO, mientras que para NCBI alineó contra un RNA ribosomal 12S. **Contig**. Para GO alinea contra subunidad I de citocromo C oxidasa, conteniendo un fragmento de *u1468*, mientras que para NCBI alinea contra secuencia de DNA mitocondrial, en la región de COX1.
- 2) *u1311*. Esta secuencia está anotada como snRNA en la página de NONCODE (Tabla A), involucrada en la metilación de RNA. No alineó contra ninguna secuencia de las bases GO y NCBI. **Contig**. No alinea contra GO pero si con una secuencia de RNA ribosomal 28S de la base NCBI. La región del contig que alinea con *u1311* es un microsatélite.
- 3) *u596*. Esta secuencia paradójicamente no tiene referencia de literatura en la página de NONCODE (Tabla A), como tampoco lo tiene la página del NCBI a la que tiene liga. Sólo la describen como “*misfunction\_snm*” pero nada más. Esta secuencia presenta un alineamiento poco significativo con la base GO, y la secuencia con la que alinea en NCBI es la misma que da la página de NONCODE. **Contig**. No alinea contra la base GO, pero contra NCBI alinea una secuencia reportada como el gen ribosomal 18S. Además, el fragmento de este contig que alinea con NCBI contiene solo una porción de la parte que alineó contra el gen *u596*.
- 4) *u298*. Esta secuencia está anotada como snRNA en la página de NONCODE (Tabla A), igual que *u1311*. No alinea con GO pero si con NCBI para snRNA-U90. **Contig**. No alinea con GO, con NCBI alinea con *Zebra fish* en región de GT's.
- 5) *u1740*. La base NONCODE menciona que esta secuencia participa en la generación de RNA nuclear pequeño (Tabla A), lo que es parecido a la información de la secuencia con que alineó en el NCBI, pero la base de GO encuentra un parecido poco significativo con una proteína de *Arabidopsis*. **Contig**. No presenta similitud con las bases GO y NCBI, únicamente con la secuencia de EST reportada por el consorcio en la base de EST del NCBI. Al analizar la región que alineó, se encontró que consistía de una secuencia repetitiva de C's y T's.

**NONCODE vs EST adulto.**

NONCODE vs EST Adulto									
N	ID NOCODE	ID Contig <i>T. solium</i>	Id	L	M	G	E-value	S	%
1	u1311,Arabidopsis	ljpltsa01024_P08_.g.ab1.fg0	91.3	46	4	0	7.00E-10	60	34%
2	u1741,Homo	061705c_C22_TsB1098.g1_060.ab1.fg0	96.3	54	2	0	7.00E-18	91.7	1%
3	u5338,Mus	140905C_F03_TsF03142.g1_091.ab1.fg0	93.88	98	6	0	4.00E-33	147	0.3%
4	u2226,Microtus	061705c_I06_TsE06108.g1_085.ab1.fg0	100	34	0	0	3.00E-10	67.9	0.2%
	*u3277,Agaricus	061006B_A03_TsIvA0197.T3_017.ab1.fg0	100	21	0	0	4.00E-04	42.1	
	*u5338,Mus	230606B_G04_TsIvG0433.T3_033.ab1.fg0	83.45	139	22	1	6.00E-15	85.7	

**Tabla R2A.** Resultado (resumido) de *BLASTn* al comparar la base de datos de ncRNA « **NONCODE** » contra el archivo de *T. solium* « **EST adulto** ». En la parte inferior (sombreado) se muestra los dos mejores resultados obtenidos al ejecutar *BLASTn* entre la base NONCODE aleatorizada y el archivo de *Ts* « *EST larva* ». [ **ID\_NONCODE**=Identificador\_en\_NONCODE, **ID\_Contig**=Identificador\_contig\_ *T.solium*, **ID(%)**=porcentaje\_identidad\_alineamiento, **L**=longitud\_alineamiento, **M**=mismatches, **G**=gaps, **E-value**=valor\_E, **S**= score\_(bits), **%**=porcentaje\_secuencia\_ncRNA (ver pág. 50).]

NONCODE vs GO y NCBI					Contig <i>Taenia</i> vs GO, NCBI y ESTs	
N	Id	Base	Información	Evalue	Información del Contig con que alinea	Evalue
1	u1311	GO	No encuentra similitud significativa	-----	Putative protein of unknown function identified by fungal homology comparisons and RT-PCR ( <i>Saccharomyces cerevisiae</i> ) (el alineamiento no contiene a la secuencia NONCODE)	7.3e-10
		NC BI	No encuentra similitud significativa	-----	Tetrabothrius erostris 28S ribosomal RNA gene, complete sequence (el alineamiento no contiene a u1311)	3e-167
		EST			AD0032145 <i>Taenia solium</i> UNAM-cd1_adult <i>Taenia solium</i> cDNA, mRNA sequence	0.0
2	u1741	GO	No encuentra similitud significativa	-----	Trex2 three prime repair exonuclease 2 (gene <i>Rattus norvegicus</i> ) (el alineamiento no contiene a la secuencia NONCODE)	4.3e-20
		NC BI	Homo sapiens clone kid12 SNURF-SNRPN mRNA, downstream untranslated exons, alternatively spliced	0.0	No encuentra similitud significativa	-----
		EST			AD0203023 <i>Taenia solium</i> UNAM-cd1_adult <i>Taenia solium</i> cDNA, mRNA sequence	0.0
3	u5338	GO	Muc30E "Mucin 30E" 8 (from <i>Drosophila Melanogaster</i> )	0.00016	NADH dehydrogenase (ubiquinone) activity (from <i>C. elegans</i> ) (alineamiento no contiene a la secuencia NONCODE, queda a 1b del ATG)	5.6e-22
		NC BI	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		EST			AD0136037 <i>Taenia solium</i> UNAM-cd1_adult <i>Taenia solium</i> cDNA, mRNA	0.0
4	u2226	GO	Med14 mediator complex subunit 14 (from <i>Rattus norvegicus</i> , transcription coactivator activity)	2.8e-05	No encuentra similitud significativa	-----
		NC BI	Peromyscus maniculatus bairdii BW:Br,Ts cDNA clone CCUB7237 3', mRNA	2e-112	<i>Rattus norvegicus</i> Y Chr BAC (u2226 alinea en region repetitive de GT's)	1e-11
		EST			AD0102039 <i>Taenia solium</i> UNAM-cd1_adult <i>Taenia solium</i> cDNA, mRNA sequence	0.0

**Tabla R2B.** Resultado resumido de *BLASTn* comparando la secuencia de NONCODE de Tabla R2A (tercera y cuarta columnas) vs las bases Gene Ontology y NCBI *nr/nt*. Las dos columnas de lado derecho son de *BLASTn* comparando el contig de EST de *T. solium* que mejor alineó con la secuencia de ncRNA de NONCODE contra las bases a) Gene Ontology, b) NCBI *nr/nt* y c) NCBI *EST*.

Anotación NONCODE, Gene Ontology y NCBI.

- 1) *u1311*. Esta secuencia también alineó con el archivo de EST larva, y está anotada como snRNA (Tabla A). No hace alineamiento contra las bases GO y

NCBI. **Contig.** En la base GO alineó contra una proteína putativa de función desconocida, mientras que en NCBI alineó contra un gen de RNA ribosomal 28S. En éste último, el valor E fue el más significativo de todos los alineamientos hechos por los contigs de este bloque de EST adulto ( $3e-167$ ). Al ver que la región que alinea es la misma que alineó con EST larva (bases 75 a 115), podemos ver que corresponde a un microsatélite.

2) *U1741*. Esta secuencia muestra la misma información que u1740 (Tabla A). **Contig.** A diferencia del contig que alineó con u1740, éste contig de *Taenia* alineó para GO con una exonucleasa 2 (*Trex2*), aunque la parte que alineó no contiene algún fragmento de u1741. No presenta similitud con alguna secuencia de NCBI. Al analizar la región que alineó se encontró que constaba de zona repetitiva de AG's.

3) *u5338*. Esta secuencia es conocida como AIR, y juega un papel en la impronta génica (Tabla A). La versión aleatorizada de esta secuencia genera muchos alineamientos contra todos los archivos de *Taenia* empleados en este trabajo. Aunque la longitud del alineamiento es pequeño respecto de la secuencia original (0.3%), obtuvo el valor E más bajo de todos los alineamientos contra los EST de adulto. Curiosamente, en GO hizo alineamiento con Muc30E (una glicoproteína de mucosas) con valor E pobre (.000016), mientras que para NCBI no encontró similitud alguna. **Contig.** En GO hizo alineamiento con una secuencia con actividad de NADH deshidrogenasa, pero la zona que alineó no contiene algún fragmento de u5338. Para NCBI no encontró similitud.

4) *u2226*. Esta secuencia esta anotada como impronta de DNA en NONCODE, (Tabla A), causa interés ya que está adjudicada a la secuencia de impronta XIST, responsable de la inactivación de uno de los dos cromosomas X en mamíferos, algo que se raro para un organismo hermafrodita como *Taenia*. La longitud del alineamiento corresponde a menos del 1% de la secuencia del RNA reportada. La secuencia de ncRNA de XIST alinea en GO con una subunidad de Med14 (*mediator complex* y que parece ser un coactivador de la expresión según GO). Para NCBI alinea con una secuencia de mRNA de *Peromyscus maniculatus bairdii* con un significativo valor E. **Contig.** Curiosamente, el contig de *Taenia* con que alineó XIST no hizo alineamientos con GO y NCBI, sólo con el EST reportado por el consorcio.

**NONCODE vs EST Japonés (Full-length).**

NONCODE vs EST Japón (Full-length)									
Num	ID NOCODE	ID contig <i>T. solium</i>	Id	L	M	G	E-value	S	%
1	u4170,Echinococcus	151007_utk_fl1_f0_102N01.ab1	98.09	157	3	0	4.00E-78	287	99%
2	U3977,Schistosoma	151007_utk_fl1_f0_125G13.ab1	89.94	179	11	4	1.00E-47	186	92%
3	u4043,Gallus	151007_utk_fl1_f0_087K06.ab1	88.31	77	9	0	4.00E-16	81.8	55%
4	U596,Oryza	151007_utk_fl1_f0_040L08.ab1	93.33	45	3	0	5.00E-11	65.9	16%
	*C u727,Yersinia	151007_utk_fl1_f0_100J14.ab1	96.15	26	1	0	5.00E-05	44.1	
	*C U5338,Mus	151007_utk_fl1_f0_103°22.ab1	94.87	39	2	0	3.00E-07	61.9	

**Tabla R3A.** Resultado (resumido) de *BLASTn* al comparar la base de datos de ncRNA «NONCODE» contra el archivo de *T. solium* «EST Japonés Full length». En la parte inferior (sombreado) se muestra los dos mejores resultados obtenidos al ejecutar *BLASTn* entre la base NONCODE aleatorizada y el archivo de *Ts* «EST larva». [*ID\_NONCODE*=Identificador\_en\_NONCODE, *ID\_Contig*=Identificador\_contig\_*T.solium*, *ID*(%)=porcentaje\_identidad\_alineamiento, *L*=longitud\_alineamiento, *M*=mismatches, *G*=gaps, *E-value*=valor\_*E*, *S*=score\_(bits), %=porcentaje\_secuencia\_ncRNA (ver pág. 50).]

NONCODE vs GO y NCBI					Contig <i>Taenia</i> vs GO, NCBI y ESTs	
N	Id	Base	Información	Evalue	Información del Contig con que alinea	Evalue
1	U4170	GO	No encuentra similitud significativa	-----	glutathione S-transferase mu 4 ( <i>Rattus norvegicus</i> ) (no contiene a u4170)	1.8e-08
		NC BI	<i>E.multilocularis</i> U1 small nuclear RNA gene	2e-76	<i>E.multilocularis</i> U1 small nuclear RNA gene (si contiene a u4170)	1e-69
		EST			<i>Taenia solium</i> cysticerca cDNA library (si contiene a u4170)	1e-54
2	U3977	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NC BI	<i>Schistosoma mansoni</i> U2 small nuclear RNA	6e-96	<i>Taenia asiatica</i> clone TaHC1-D8 mRNA sequence (si contiene a u3977)	0.0
		EST			AD0105064 <i>Taenia solium</i> UNAM-cd1_adult <i>Taenia solium</i> cDNA (no contiene a u3977)	0.0
3	U4043	GO	No encuentra similitud significativa	-----	Act-1, Actin, cytoplasmic 1 ( <i>C. elegans</i> ) (no contiene a u4043)	3.3e-91
		NC BI	<i>Gallus gallus</i> clone 3009 GGN130 snRNA, complete sequence	2e-66	<i>Taenia asiatica</i> clone HC10-A7 cytoskeletal actin 2 mRNA, partial cds (no contiene a u4043)	7e-161
		EST			<i>Taenia solium</i> adult full-length cDNA library	0.0
4	U596	GO	NSE_0607 hypothetical protein (protein from <i>Neorickettsia sennetsu</i> str. Miyayama)	0.013	gene_product from <i>Danio rerio</i>	3.2e-16
		NC BI	<i>Oryza sativa</i> rDNA ITS1-IR36-S region	9e-144	<i>Taenia solium</i> 18S ribosomal RNA gene, complete sequence (contiene fracción de u596)	0.0
		EST			<i>Taenia solium</i> cysticerca cDNA library <i>Taenia solium</i> cDNA, mRNA (si contiene a u596)	0.0

**Tabla R3B.** Resultado resumido de *BLASTn* comparando la secuencia de NONCODE de Tabla R3A (tercera y cuarta columnas) vs las bases Gene Ontology y NCBI nr/nt. Las dos columnas de lado derecho son de *BLASTn* comparando el contig de EST de *T. solium* que mejor alineó con la secuencia de ncRNA de NONCODE contra las bases a) Gene Ontology, b) NCBI nr/nt y c) NCBI EST.

*Anotación NONCODE, Gene Ontology y NCBI.*

- 1) u4170. Esta secuencia está anotada como snRNA u4 en NONCODE (Tabla A), involucrada en el procesamiento pos-transcripcional del RNA (“splicing”). Para GO no encuentra alineamiento, mientras que en la base NCBI alinea con la secuencia de snRNA de *E.multilocularis* a la que hace referencia la literatura de la tabla de anotación de NONCODE. *Contig.* Para GO alinea con una secuencia de glutathione



S-transferase mu4 de *Rattus norvegicus*, el fragmento que alinea no contiene a u4170. Para NCBI, el contig alinea con la secuencia de snRNA U1 de *E. multilocularis* (la misma con la que alinea la secuencia u4170). El fragmento que alinea si contiene a u4170.

- 2) u3977. Esta secuencia está anotada como snRNA u2 de *Schistosoma mansoni* en NONCODE (Tabla A), involucrada en el procesamiento pos-transcripcional del RNA (“splicing”). Para GO no encuentra alineamiento, mientras que en NCBI alinea contra la secuencia de U2 de *Schistosoma mansoni* (referenciada en NONCODE). **Contig.** Para GO no se encontró alineamiento, mientras que para la base NCBI lo hizo con una secuencia de mRNA de *Taenia asiática*, y que sí contiene a la secuencia u3977 dentro del alineamiento.
- 3) u4043. Esta secuencia está anotada como snRNA U1 en NONCODE (Tabla A), involucrada en el procesamiento pos-transcripcional del RNA (“splicing”). Para GO no hace alineamiento mientras que en NCBI alinea contra una secuencia de *Gallus gallus* identificada también como snRNA. **Contig.** Para GO alinea con una secuencia de actina (Act-1) de *C. elegans*, mientras que para NCBI alinea también contra una secuencia de actina de la especie *Taenia asiática*. En las dos bases anteriores, la zona que alineó no contiene al ncRNA u4043.
- 4) u596. Esta secuencia también alineó con el archivo de EST larva y carece de referencia de literatura en la base de NONCODE (Tabla A). **Contig.** En GO alineó contra una secuencia para la cual esta base de datos no aporta mayor información que la de ser un producto del pez *Danio rerio*. En NCBI alinea con un valor E bajísimo (prácticamente 0.0) a una secuencia de RNA ribosomal 18S igual que el contig de larva. La zona que alinea contiene una fracción de u596.

**NONCODE vs 454AllContigs2010**

NONCODE vs 454AllContigs2010									
Num	ID NOCODE	ID Contig <i>T. solium</i>	Id	L	M	G	E-value	S	%
1	u1919,Taenia	Contig194.2	100	133	0	0	5.00E-70	264	100%
2	u1918,Taenia	Contig72707.1	100	113	0	0	3.00E-58	224	100%
3	u4170,Echinococcus	Contig1914.1	98.1	158	3	0	1.00E-77	289	99%
4	u3977,Schistosoma	Contig31346.1	90	180	10	5	8.00E-45	180	93%
5	u3976,Schistosoma	Contig1760.1	85.82	141	20	0	5.00E-27	121	89%
6	u3540,Erwinia	Contig69572.1	100	68	0	0	2.00E-31	135	64%
7	u4043,Gallus	Contig18938.1	88.31	77	9	0	4.00E-15	81.8	55%
8	u1139,Gorilla	Contig67373.1	93.24	74	5	0	9.00E-23	107	39%
9	u596,Oryza	Contig72797.1	100	45	0	0	3.00E-17	89.7	16%
10	u3429,Streptomyces	Contig35492.1	95.65	69	3	0	3.00E-23	113	2%
	*C u198,Homo	Contig1161.1	93.55	31	2	0	1.00E-04	46.1	
	*C u5338,Mus	Contig24577.1	94.64	56	3	0	5.00E-14	87.7	

**Tabla R4A.** Resultado (resumido) de *BLASTn* al comparar la base de datos de ncRNA «NONCODE» contra el archivo de *T. solium* «454AllContigs2010». En la parte inferior (sombreado) se muestra los dos mejores resultados obtenidos al ejecutar *BLASTn* entre la base NONCODE aleatorizada y el archivo de *Ts* «*EST larva*». [*ID\_NONCODE*=Identificador\_en\_NONCODE, *ID\_Contig*=Identificador\_contig\_*T.solium*, *ID*(%)=porcentaje\_identidad\_alineamiento, *L*=longitud\_alineamiento, *M*=mismatches, *G*=gaps, *E-value*=valor\_*E*, *S*=score\_(bits), %=porcentaje\_secuencia\_ncRNA (ver pág. 50).]

- 1) *u1919*. Esta secuencia de *Taenia solium* está anotada como el snRNA U5 en NONCODE (Tabla A). Para la base GO no encontró secuencia similar, mientras que para NCBI alineó contra la la secuencias U5 de la cual deriva la de NONCODE. **Contig.** Para GO no encuentra secuencia parecida, mientras que para NCBI alinea nuevamente contra la secuencia U5.
- 2) *u1918*. Esta secuencia está anotada como snRNA U6 de *Taenia solium* en NONCODE (Tabla A), involucrada en el procesamiento pos-transcripcional del RNA (“splicing”). En GO no alinea con secuencia alguna, pero en NCBI alinea contra u6 snRNA de *Schistosoma mansoni* (de esta secuencia deriva la de NONCODE). **Contig.** En GO no alinea con secuencia alguna, mientras que en NCBI alinea con una entrada de U5 y U6 de *Taenia solium*. Curiosamente, el contig no alinea con un EST de *Taenia*, sino con otro de *Schistosoma mansoni*.
- 3) *u4170*. Esta secuencia de *Echinococcus* está anotada como snRNA U1 en NONCODE (Tabla A), y está involucrada en el procesamiento pos-transcripcional del RNA (“splicing”). En GO, u4170 no alinea contra ninguna secuencia, mientras que en NCBI alinea contra snRNA u1 de *Echinococcus multilocularis* (la secuencia de la que se desprende la de NONCODE). **Contig.** Para GO no encuentra secuencia similar, pero en NCBI alinea nuevamente contra snRNA u1 de *E. multilocularis*.

- 4) *u3977*. Esta secuencia (que también alineó con un EST de Japón) está anotada como snRNA u2 de *Schistosoma mansoni* en NONCODE (Tabla A), involucrada en el procesamiento pos-transcripcional del RNA (“splicing”). Para GO no encuentra alineamiento, mientras que en NCBI alinea contra la secuencia de U2 de *Schistosoma mansoni* (referenciada en NONCODE). **Contig**. Para GO no encontró secuencia parecida, mientras que para NCBI alineó con la entrada de U5 y U6 snRNA de *Taenia solium* con valor *E* altamente significativo. Esta entrada del NCBI es de donde extrae la información la base de NONCODE de u1919.
- 5) *u3976*. Esta secuencia de *Schistosoma mansoni* está anotada como snRNA u1 en NONCODE (Tabla A), involucrada en el procesamiento pos-transcripcional del RNA (“splicing”). Esta secuencia no encuentra parecida en GO, mientras que en NCBI alinea con snRNA u1 de *Schistosoma mansoni* (ella misma). **Contig**. En GO no alineó con secuencia alguna, y en NCBI alinea con el snRNA U1 de *Echinococcus multilocularis*.
- 6) *u3540*. Esta secuencia de *Erwinia amylovora* está involucrada con la represión de la traducción según NONCODE (Tabla A). En GO alinea con una proteína no caracterizada de *Gallus gallus*, mientras que en GO alinea significativamente con un vector. **Contig**. En GO alinea nuevamente con una proteína no caracterizada de *Gallus gallus*, y en NCBI alinea con el vector de clonación pRama-35 muy significativamente (un valor *E* de cero). Curiosamente, alinea contra un EST de *Gandalfus yunohana* en lugar de *Taenia solium*. Los tres alineamientos anteriores contienen a u3540.
- 7) *u4043*. Esta secuencia de *Gallus gallus* está anotada como snRNA U1 en NONCODE (Tabla A), involucrada en el procesamiento pos-transcripcional del RNA (“splicing”). Para GO no hace alineamiento mientras que en NCBI alinea contra una secuencia de *Gallus gallus* identificada también como snRNA. **Contig**. Para GO no encuentra alineamiento, mientras que en NCBI alinea contra un supercontig de *Schistosoma mansoni* (dentro de la descripción de la secuencia en el NCBI no hay referencia a función). Curiosamente, el contig no alinea con EST de *Taenia solium*, sino con uno de la babosa de mar *Aplysia kurodai*.
- 8) *u1139*. Esta secuencia de *Gorilla gorilla* está anotada como snRNA U2 en NONCODE (Tabla A). En GO alineó con una proteína de *A. thaliana* (una putativa

cinasa) pero con un valor E poco significativo. En cambio, con NCBI alineó a U2 snRNA de *Gorilla gorilla*, siendo ésta la secuencia de la que deriva la de NONCODE. **Contig.** Para GO no encontró secuencia parecida, mientras que en NCBI alineó a un pseudogen de snRNA u2 de *Rattus norvegicus*, que si contiene a u1139. Curiosamente, contra la base EST alineó con una clona de *Homo sapiens* y no contra los archivos de *Taenia solium* reportados por el consorcio.

- 9) u596. Esta secuencia de *Oryza sativa* también alineó con los archivos de EST larva y EST de Japón. Carece de referencia de literatura en la base NONCODE (Tabla A). **Contig.** Al igual que el contig del EST japonés, en GO alineó contra una secuencia para la cual esta base de datos no aporta mayor información que la de ser un producto del pez *Danio rerio*. En NCBI alinea con un valor E bajísimo (prácticamente 0.0) a una secuencia de RNA ribosomal 18S igual que el contig de larva. La zona que alinea contiene una fracción de u596.
  
- 10) u3429. Esta secuencia de *Streptomyces* está involucrada en la represión de la transcripción de acuerdo a NONCODE (Tabla A). Contrario a la anotación, esta secuencia, en la base de GO alinea con una putativa glicina deshidrogenasa de *M. tuberculosis* con un valor E en extremo significativo ( $5e-303$ ). Igualmente, para la base NCBI, alinea con otra putativa glicina deshidrogenasa de *Streptomyces avermitilis*. **Contig.** Para GO alinea con la proteína P del sistema de corte de glicina, y contiene a u3429 dentro del fragmento que alinea. En NCBI alinea con una oxidoreductasa de unión a NADP predicha para *E. coli*, y también contiene a u3429 dentro de la región que alinea. Curiosamente, alinea contra un EST de *Puccinia triticina* en lugar de *Taenia solium*.

		NONCODE vs GO y NCBI			Contig <i>Taenia</i> vs GO, NCBI y ESTs	
N	Id	Base	Información	Evalúe	Información del Contig con que alinea	Evalúe
1	U1919	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NCBI	<i>Taenia solium</i> U6 snRNA and U5 snRNA genes, complete sequence	2e-62	<i>Taenia solium</i> U6 snRNA and U5 snRNA genes (contiene a u1919)	0.0
		EST			LV0318039 <i>Taenia solium</i> UNAM-cd2_larva <i>Taenia solium</i> cDNA, mRNA	1e-73
2	U1918	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NCBI	<i>Taenia solium</i> U6 snRNA and U5 snRNA genes, complete sequence	2e-51	<i>Taenia solium</i> U6 snRNA and U5 snRNA genes(contiene a u3979)	0.0
		EST			EG1 <i>Schistosoma mansoni</i> cDNA clone NC_EG117131e01, mRNA sequence	7e-41
3	U4170	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NCBI	<i>E.multilocularis</i> U1 small nuclear RNA gene	2e-76	<i>E.multilocularis</i> U1 small nuclear RNA gene (contiene a u4170)	1e-79
		EST			<i>Taenia solium</i> cysticercera cDNA library <i>Taenia solium</i> cDNA (contiene la mitad de u4170)	4e-59
4	U3977	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NCBI	<i>Schistosoma mansoni</i> U2 small nuclear RNA	6e-96	<i>Schistosoma mansoni</i> genome sequence supercontig (contiene a u3977)	1e-79
		EST			TSBK.R55.esd <i>Taenia solium</i> cysticercera cDNA library (contiene a u3977)	5e-83
5	U3976	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NCBI	<i>Schistosoma mansoni</i> U1 small nuclear RNA	3e-76	<i>E.multilocularis</i> U1 small nuclear RNA gene (contiene a u3976)	2e-81
		EST			LV0259054 <i>Taenia solium</i> UNAM-cd2_larva <i>Taenia solium</i> cDNA, mRNA (contiene u3976)	6e-127
6	U3540	GO	Uncharacterized protein ( <i>Gallus gallus</i> )	3.3e-12	Uncharacterized protein from <i>Gallus gallus</i> (contiene a u3540)	1.2e-15
		NCBI	Gene deletion vector pJL2, complete sequence	7e-48	Cloning vector pRama-35, complete sequence (contiene a u3540)	0.0
		EST			Gandalfus yunohana hepatopancreas library Gandalfus yunohana cDNA clone 1C10 (contiene a u3540)	1e-109
7	U4043	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NCBI	<i>Gallus gallus</i> clone 3009 GGN130 snRNA, complete sequence	2e-66	<i>Schistosoma mansoni</i> genome sequence supercontig Smp_scaff000014 (contiene u4043)	7e-32
		EST			<i>Aplysia kurodai</i> nervous system full-length enriched EST library (contiene a u4043)	9e-26
8	U1139	GO	protein from <i>Arabidopsis thaliana</i> (kinase activity)	0.00071	No encuentra similitud significativa	-----
		NCBI	<i>Gorilla gorilla</i> U2 small nuclear RNA	6e-93	rat u2 small nuclear rna pseudogene (contiene a u1139)	7e-22
		EST			NETRP2 Homo sapiens cDNA clone NETRP2003053 5' (contiene a u1139)	7e-22
9	U596	GO	NSE_0607 hypothetical protein (protein <i>Neorickettsia sennetsu</i> str. <i>Miyayama</i> )	0.013	gene_product from <i>Danio rerio</i>	2.1e-09
		NCBI	<i>Oryza sativa</i> rDNA ITS1-IR36-S region	9e-144	<i>Taenia solium</i> 18S ribosomal RNA gene, complete sequence (contiene fracción de u596)	0.0
		EST			<i>Taenia solium</i> cysticercera cDNA library <i>Taenia solium</i> cDNA, mRNA (si contiene a u596)	0.0
10	U3429	GO	Probable glycine dehydrogenase ( <i>M.tuberculosis</i> )	5e-303	glycine cleavage system P protein (contiene a u3429)	3.5e-137
		NCBI	<i>Streptomyces avermitilis</i> MA-4680 DNA, putative glycine dehydrogenase	0.0	<i>Escherichia coli</i> DH1 predicted NAD(P)-binding oxidoreductase with NAD(P)-binding Rossmann-fold domain (contiene a u3429)	0.0
		EST			PTDH <i>Puccinia tritricina</i> cDNA clone (contiene a u3429)	0.0

Tabla R4B. Resultado resumido de *BLASTn* comparando la secuencia de NONCODE de Tabla R4A (tercera y cuarta columnas) vs las bases Gene Ontology y NCBI *nr/nt*. Las dos columnas de lado derecho son de *BLASTn* comparando el contig de *T. solium* que mejor alineó con la secuencia de ncRNA de NONCODE contra las bases a) Gene Ontology, b) NCBI *nr/nt* y c) NCBI *EST*.

## HAIRPINS

### Hairpins vs EST Larva.

Hairpins (v16) vs EST larva									
N	ID miRBase	ID EST LARVA	Id	L	M	G	E-value	S	%
1	mdo-mir-739	010906A_K10_TslvA0151.T3_061.ab1.fg0	93.26	89	6	0	4.00E-32	132	97%
	*C Tni-let-7i	010906C_N06_TslvA0158.T3_088.ab1.fg0	100	21	0	0	5.00E-05	42.1	

**Tabla R5A.** Resultado (resumido) de *BLASTn* al comparar la base de datos de ncRNA « **miRBase (Hairpins)** » contra el archivo de *T. solium* « **EST larva** ». En la parte inferior (sombreado) se muestra el mejor resultado obtenido al ejecutar *BLASTn* entre la base miRBase aleatorizada y el archivo de *Taenia solium* « *EST larva* ». [ **ID\_miRBase** = Identificador\_en\_miRBase, **ID\_Contig**=Identificador\_contig *T.solium*, **ID(%)** = porcentaje identidad alineamiento, **L**=longitud\_alineamiento, **M**=mismatches, **G**=gaps, **E-value**=valor\_E, **S**= score\_(bits), **%**=porcentaje\_secuencia\_ncRNA (ver pág. 50).]

Hairpins vs GO y NCBI				Contig <i>Taenia</i> vs GO, NCBI y ESTs		
N	Id	Base	Información	Evalue	Información del Contig con que alinea	Evalue
1	mdo-mir-739	GO	No encuentra similitud significativa	-----	protein from <i>Arabidopsis thaliana</i> (no contiene al precursor del miRNA)	0.032
		NC BI	<i>Scarus ghobban voucher</i> S.gho-1-V-NBFGR-LKO 28S ribosomal RNA gene	7e-40	<i>Nippotaenia mogurndae</i> 28S ribosomal RNA gene (contiene al precursor)	7e-137
		EST			<i>Taenia solium</i> UNAM-cd2_larva <i>Taenia solium</i> cDNA	0.0

**Tabla R5B.** Resultado resumido de *BLASTn* comparando la secuencia precursora de microRNA (primeras tres columnas) vs bases Gene Ontology y NCBI *nr/nt*. Las dos columnas de lado derecho son de *BLASTn* comparando el contig de EST de *T. solium* que mejor alineó con el precursor de microRNA contra las bases a) Gene Ontology, b) NCBI *nr/nt* y c) NCBI *EST*.

1. Mdo-mir-739. Esta secuencia de *Monodelphis domestica* fue encontrada en estudios de secuenciación masiva, por lo que carece todavía de una función bien caracterizada (Tabla B). La secuencia del precursor no encuentra alguna similar en la base de GO, pero en NCBI alinea con un gen de RNA ribosomal 28S de la especie *Scarus ghobban* de manera significativa. **Contig.** El contig de *T. solium* que alinea con este precursor alinea pobremente con una proteína de *A. thaliana*, y en NCBI alinea muy significativamente contra un gen de RNA ribosomal 28S de *Nippotaenia mogurndae*

### Hairpins vs EST Adulto.

Hairpins (v16) vs EST adulto										
N	ID miRBase	ID EST ADULTO	Id	L	M	G	E-value	S	%	
1	mdo-mir-739	150605B_F05_TsF0577.b1_043.ab1.fg0	93.26	89	6	0	7.00E-32	132	97%	
2	gga-mir-3533	061705b_F05_TsC0593.g1_071.ab1.fg0	85.88	85	10	2	4.00E-19	89.8	92%	
	*C Tni-let-7i	010906C_N06_TslvA0158.T3_088.ab1.fg0	100	21	0	0	5.00E-05	42.1		

**Tabla R6A.** Resultado (resumido) de *BLASTn* al comparar la base de datos de ncRNA « **miRBase (Hairpins)** » contra el archivo de *T. solium* « **EST adulto** ». En la parte inferior (sombreado) se muestra el mejor resultado obtenido al ejecutar *BLASTn* entre la base miRBase aleatorizada y el archivo de *Taenia solium* « *EST larva* ». [ **ID\_miRBase** = Identificador\_en\_miRBase, **ID\_Contig**=Identificador\_contig *T.solium*, **ID(%)** = porcentaje identidad alineamiento, **L**=longitud\_alineamiento, **M**=mismatches, **G**=gaps, **E-value**=valor\_E, **S** = score\_(bits), **%**=porcentaje\_secuencia\_ncRNA (ver pág. 50).]

Hairpins vs GO y NCBI					Contig Taenia vs GO, NCBI y ESTs	
N	Id	Base	Información	Evalue	Información del Contig con que alinea	Evalue
1	mdo-mir-739	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NC BI	<i>Scarus ghobban voucher S.gho-1-V-NBFR-LKO 28S ribosomal RNA gene</i>	7e-40	<i>Tetrabothrius erostris 28S ribosomal RNA gene (contiene al precursor)</i>	0.0
		EST			<i>Taenia solium UNAM-cd1_adult Taenia solium cDNA (contiene al precursor)</i>	0.0
2	gga-mir-3533	GO	Actin, cytoplasmic 2" <i>Gallus gallus</i>	1.0e-09	bactin2, from <i>Danio rerio</i> (contiene al precursor)	2.8e-66
		NC BI	<i>Gallus gallus microRNA mir-3533 (MIR3533), microRNA</i>	2e-35	<i>Diphyllbothrium dendriticum clone pDidact4 actin mRNA (contiene al precursor)</i>	4e-146
		EST			<i>Taenia solium UNAM-cd1_adult Taenia solium cDNA (contiene al precursor)</i>	0.0

**Tabla R6B.** Resultado resumido de *BLASTn* comparando la secuencia precursora de microRNA (primeras tres columnas) vs bases Gene Ontology y NCBI nr/nt. Las dos columnas de lado derecho son de *BLASTn* comparando el contig de EST de *T. solium* que mejor alineó con el precursor de microRNA contra las bases a) Gene Ontology, b) NCBI nr/nt y c) NCBI EST.

1. *mdo-mir-739*. Esta secuencia de *Monodelphis domestica* fue encontrada en estudios de secuenciación masiva, por lo que carece todavía de una función bien caracterizada (Tabla B), y también aparece en EST de larva. La secuencia del precursor no encuentra alguna similar en la base de GO, pero en NCBI alinea con un gen de RNA ribosomal 28S de manera significativa. **Contig.** El contig de *T. solium* que alinea con este precursor no encuentra parecido con secuencia alguna de GO, pero en NCBI alinea con una secuencia de de RNA ribosomal 28S de *Tetrabothrius erostris* con un valor E prácticamente de cero.
2. *gga-mir-3533*. Esta secuencia de *Gallus gallus* se halló expresada diferencialmente mediante secuenciación de alto desempeño hace dos años (2009), pero no se conoce que papel desempeña (Tabla B). Esta secuencia alinea contra un gen de actina en la base GO, mientras que en la base NCBI alinea contra la secuencia de la que deriva la de miRBase. **Contig.** En la base GO alinea contra una secuencia de actina b de *Danio rerio*, mientras que en NCBI alinea muy significativamente contra otro gen de actina pero de *Diphyllbothrium dendriticum*.

**Hairpins vs EST Japón (full-length).**

Hairpins (v16) vs EST japon									
N	ID miRBase	ID EST JAPON	Id	L	M	G	E-value	S	%
1	mdo-mir-739	151007_utk_fl1_f0_130F08.ab1	93.26	89	6	0	1.00E-31	132	97%
	*C hsa-mir-28	151007_utk_fl1_f0_058F05.ab1	100	21	0	0	2.00E-04	42.1	

**Tabla R7A.** Resultado (resumido) de *BLASTn* al comparar la base de datos de ncRNA « **miRBase (Hairpins)** » contra el archivo de *T. solium* « **EST Japón (Full length)** ». En la parte inferior (sombreado) se muestra el mejor resultado obtenido al ejecutar *BLASTn* entre la base miRBase aleatorizada y el archivo de *Taenia solium* « *EST larva* ». [ **ID\_miRBase** = Identificador\_en\_miRBase, **ID\_Contig**=Identificador\_contig *T.solium*, **ID(%)** = porcentaje identidad alineamiento, **L**=longitud\_alineamiento, **M**=mismatches, **G**=gaps, **E-value**=valor\_E, **S**= score\_(bits), **%**=porcentaje\_secuencia\_ncRNA (ver pág. 50).]

Hairpins vs GO y NCBI					Contig <i>Taenia</i> vs GO, NCBI y ESTs	
N	Id	Base	Información	Evalue	Información del Contig con que alinea	Evalue
<b>1</b>	mdo-mir-739	GO	No encuentra similitud significativa	-----	Putative uncharacterized protein "Eukaryotic translation initiation factor 3subunit D" (from <i>Gallus gallus</i> ) (no contiene al precursor)	8.0e-14
		NCBI	<i>Scarus ghobban voucher</i> S.gho-1-V-NBFGR-LKO 28S ribosomal RNA gene	7e-40	<i>Acanthobothrium parviuncinatum</i> 28S large subunit ribosomal RNA gene	3e-159
		EST			<i>Taenia solium</i> adult full-length cDNA library	0.0

**Tabla R7B.** Resultado resumido de *BLASTn* comparando la secuencia precursora de microRNA (primeras tres columnas) vs bases Gene Ontology y NCBI *nr/nt*. Las dos columnas de lado derecho son de *BLASTn* comparando el contig de EST de *T. solium* que mejor alineó con el precursor de microRNA contra las bases a) Gene Ontology, b) NCBI *nr/nt* y c) NCBI *EST*.

1. *mdo-mir-739*. Esta secuencia de *Monodelphis domestica* fue encontrada en estudios de secuenciación masiva, por lo que carece todavía de una función bien caracterizada (Tabla B), y también aparece en EST de larva. La secuencia del precursor no encuentra alguna similar en la base de GO, pero en NCBI alinea con un gen de RNA ribosomal 28S de manera significativa. **Contig.** El contig de *T. solium* que alineó con este precursor encontró parecido a una secuencia de la base Gene Ontology que putativamente es un factor de inicio de la traducción en *Gallus gallus*. En la base de NCBI, alineó contra una secuencia de un gen de RNA ribosomal 28S de la subunidad grande del ribosoma en la especie *Acanthobothrium parviuncinatum*.

Solo hubo un alineamiento de esta secuencia, ningún otro precursor de miRNA alinea con este archivo de EST japonés. Esto parece deberse a que a la hora de ejecutar el *BLASTn*, éste emitía un error que reportaba no poder calcular los parámetros Karlin-alschul. Esto sucedió con muchas secuencias de EST japonés.



**Hairpins vs 454AllContigs2010.**

Hairpins (v16) vs EST 454AllContigs2010									
N	ID miRBase	ID EST 454AllContigs2010	Id	L	M	G	E-value	S	%
1	mdo-mir-739	contig37260	92.55	94	4	3	1.00E-30	132	102%
2	ssc-mir-4331	contig26401	90.54	74	6	1	4.00E-20	97.1	93%
3	sme-mir-749	contig02881	86.25	80	11	0	3.00E-17	87.9	87%
	*C dya-mir-276 <sup>a</sup>	Contig14102.1	100	24	0	0	3.00E-05	48.1	

**Tabla R8A.** Resultado (resumido) de *BLASTn* al comparar la base de datos de ncRNA « **miRBase (Hairpins)** » contra el archivo de *T. solium* « **454AllContigs2010** ». En la parte inferior (sombreado) se muestra el mejor resultado obtenido al ejecutar *BLASTn* entre la base miRBase aleatorizada y el archivo de *Taenia solium* « *EST larva* ». [ **ID\_miRBase** = Identificador\_en\_miRBase, **ID\_Contig**=Identificador\_contig *T.solium*, **ID(%)** = porcentaje identidad alineamiento, **L**=longitud\_alineamiento, **M**=mismatches, **G**=gaps, **E-value**=valor\_E, **S**= score\_(bits), **%**=porcentaje\_secuencia\_ncRNA (ver pág. 50).]

Hairpins vs GO y NCBI					Contig <i>Taenia</i> vs GO, NCBI y ESTs	
N	Id	Base	Información	Evalue	Información del Contig con que alinea	Evalue
1	mdo-mir-739	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NC BI	<i>Scarus ghobban voucher S.gho-1-V-NBFGRLKO 28S ribosomal RNA gene</i>	7e-40	<i>Zea mays 26S ribosomal RNA gene (contiene al precursor)</i>	3e-138
		EST		7e-40	<i>Zea mays cDNA clone 1563685 5' (contiene al precursor)</i>	2e-141
2	ssc-mir-4331	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NC BI	Pig DNA sequence from clone CH242-11717 on chromosome X	1e-21	Pig DNA sequence from clone CH242-127I21 on chromosome X (contiene al precursor)	3e-40
		EST		1e-21	SUSFLECK Loin Muscle Sus scrofa cDNA clone 54_E01 (contiene al precursor)	1e-38
3	sme-mir-749	GO	No encuentra similitud significativa	-----	No encuentra similitud significativa	-----
		NC BI	<i>Schistosoma mansoni</i> genome sequence supercontig	8e-15	<i>Schistosoma mansoni</i> genome sequence supercontig (contiene al precursor)	1e-22
		EST		6e-16	<i>Alvinella pompejana</i> Normalized library RN04 posterior end <i>Alvinella pompejana</i> cDNA clone	3e-14

**Tabla R8B.** Resultado resumido de *BLASTn* comparando la secuencia precursora de microRNA (primeras tres columnas) vs bases Gene Ontology y NCBI nr/nt. Las dos columnas de lado derecho son de *BLASTn* comparando el contig de *T. solium* 454AllContigs2010 que mejor alineó con el precursor de microRNA contra las bases a) Gene Ontology, b) NCBI nr/nt y c) NCBI EST.

1. *mdo-mir-739*. Esta secuencia de *Monodelphis domestica* fue encontrada en estudios de secuenciación masiva, por lo que carece todavía de una función bien caracterizada (Tabla B), y también aparece en EST de larva. La secuencia del precursor no encuentra alguna similar en la base de GO, pero en NCBI alinea con un gen de RNA ribosomal 28S de manera significativa. **Contig**. El contig de *T. solium* que alineó con este precursor no encontró secuencia parecida en la base GO, pero en NCBI alineó muy significativamente con una secuencia de RNA ribosomal 26S de *Zea mays*. De igual forma, contra la base de EST alineó con *Zea mays* y no contra un EST de *T. solium* reportado por el consorcio. El hecho de que aparezca 102% en la columna de % se debe a la inserción de gaps, ya que el dato en esta columna se

obtiene manualmente al dividir (longitud del alineamiento / longitud ncRNA), por lo que el alineamiento es más largo que la secuencia original.

2. *ssc-mir-4331*. Esta secuencia está reportada en estudios del transcriptoma de miRNA en el intestino porcino (Tabla B), careciendo todavía de una función bien caracterizada. La secuencia de este precursor no alinea con secuencia alguna de la base de Gene Ontology, pero en NCBI alinea con una clona del cromosoma X de puerco. Contig. La secuencia de *T. solium* que alineó con este precursor no alinea contra secuencia alguna de GO, mientras que en NCBI lo hace con la misma clona del cromosoma X de puerco con que alinea el precursor. Por otra parte, el contig no alinea con secuencia alguna de EST de *T. solium* reportada previamente, sino con una EST de *Sus scrofa* (puerco).
3. *sme-mir-749*. Esta secuencia fue encontrada en ensayos de secuenciación masiva en el organismo *Schmidtea mediterránea* (Tabla B). Parece que esta secuencia está involucrada en el desarrollo de este organismo en forma parecida a lo que hace *let-7* y *lin-4*, regula el avance entre estadios de larva a adulto, ¿pasará lo mismo en *Taenia* con su secuencia homóloga. Otra secuencia que aparece es ppt-MIR y parece tener un papel en el desarrollo de plantas y cuyo artículo propone que los microRNA se expandieron rápidamente durante la evolución de las primeras plantas terrestres (Tabla B).

## VII. DISCUSIÓN

Aunque el número de secuencias únicas que se obtuvieron fue de 21, un análisis detallado de la forma en que alinea cada una nos deja con la cantidad de 12 (8 de ocho NONCODE, Tabla C y cuatro de miRBase, Tabla B), además de algunas consideraciones respecto la base de datos NONCODE. De la base miRBase no se descartó ninguna secuencia.

**Secuencias que alinean a través de motivos repetitivos.** Las secuencias *u298*, *u1740*, *u1741*, *u2226* y *u5338* alinearon con los archivos de *T. solium* a través de regiones repetitivas de dinucleótidos como CT's, GT's y GA's. Esto no hubiera sido objeción si la mayor parte de cada secuencia se encontrara representada en distintos alineamientos (por la naturaleza local de *BLAST*), pero al analizar que regiones de la secuencia hacía alineamiento, se encontró que era el mismo segmento<sup>1</sup>. Esto, aunado al poco porcentaje que representaban de la secuencia (1% o menos), las descarta como posibles secuencias de ncRNA.

Lo anterior refleja simplemente la presencia de secuencias repetitivas en los genomas, por eso es necesario emplear las subrutinas de enmascaramiento de secuencias de baja complejidad incluidas dentro del programa *BLAST*<sup>2</sup>.

**Secuencias que alinean con RNA ribosomal u otros genes.** Las secuencias *u596* y *u1311*, consideradas como ncRNA en la base NONCODE, alinearon significativamente dentro de secuencias de RNA ribosomal en la base NCBI, poniendo en duda su definición como ncRNA (al analizar el alineamiento, tampoco se ve que estas secuencias alinearan con la cadena complementaria). Además, la página de NONCODE no aporta evidencia suficiente que valide su definición de estas secuencias, ni la entrada respectiva en el sitio de NCBI. Por tanto, se considera que son fragmentos de genes de RNA ribosomal y no ncRNA. De igual forma, *u1468* (*Homo sapiens*) alinea muy significativamente con la subunidad 1 de la citocromo oxidasa (COX1). La información disponible en NONCODE sólo la menciona como un candidato ncRNA a

---

<sup>1</sup> *u5338* alineó a través de dos segmentos repetitivos distintos y distantes entre sí. Para darse una idea de la cantidad de elementos repetitivos, véase la columna "Ap" en las tablas del Apéndice. Por ejemplo, *u1740* alineó 434 veces, siendo la misma zona de *u1740* quien alineó en todas esas ocasiones con distintos contigs (Tabla I).

<sup>2</sup> Se empleó la opción de NO filtrar al hacer *BLAST*. De allí los alineamientos por secuencias repetitivas.

partir de estudios de RNAómica (RNomics). Por tanto, tampoco puede considerarsele como ncRNA.

Por último, u3429 alinea pero sólo el 2.33% de su secuencia. No alinea a través de secuencias repetitivas, pero la información disponible en NONCODE simplemente habla de la secuencia de un organismo de interés industrial, y la referencia sólo habla de clústers de proteínas involucradas en metabolitos secundarios. Por tanto, no se considera como ncRNA para este trabajo.

**Consideraciones sobre la base NONCODE.** La secuencia u596, clasificada como ncRNA en NONCODE, se nota que es un fragmento de RNA ribosomal, y la información específica del archivo Genbank no da mayor información, ni siquiera una referencia, ¿entonces porque está en la base NONCODE si se supone que esta base de datos está curada manualmente y con referencias bibliográficas sólidas?

**Secuencias que alinean con genes de ncRNA catalogados como snRNA.** El que una gran parte de las secuencias encontradas en NONCODE sean de la clase snRNA, involucradas en el procesamiento postranscripcional del mRNA, refleja una vez más la vasta caracterización existente de este proceso. Encontrar las secuencias homólogas de U1, U2, U4, U5 y U6, refleja la conservación de este mecanismo tan importante para los organismos eucariontes. Al analizar la longitud de los alineamientos, el porcentaje de la secuencia de ncRNA contenida en él y los resultados de *BLAST* de los contigs contra Gene Ontology y NCBI, podemos considerar estas secuencias como válidas.

**Precursos de miRNA.** A excepción de *mdo-mir-739*, que pareciera ser también un gen de RNA ribosomal, las secuencias de *T. solium* son confiables en su semejanza a los cuatro precursores de miRNA encontrados. Al revisar información adicional sobre estos precursores, se nota que carecen de una caracterización experimental detallada, puesto que son derivados de estudios de secuenciación masiva.

Por otra parte, que sólo haya cuatro precursores de miRNA se debe a la falta de conservación en su secuencia, ya que se sabe la secuencia de los miRNA maduros están muy conservadas y existe una gran cantidad de ellos en los genomas de eucariontes (por tanto, es de esperarse que haya mucho más precursores que estos cuatro). La naturaleza

descriptiva y de primer acercamiento a ncRNA en *T. solium* de este trabajo, hizo descartar la búsqueda de miRNA maduro en favor de los precursores. No obstante, se está trabajando en este aspecto, como se está haciendo en organismos cercanos a *T. solium* como *Schistosoma mansoni*[74].

***Varias secuencias no alinean con Gene Ontology (GO) pero si NCBI.*** Esta base (GO) se caracteriza por estar enriquecida, pero esto sólo es posible en secuencias mejor caracterizadas experimentalmente (aunque no está exenta de “proteínas hipotéticas”), algo de lo que carecen muchas secuencias de ncRNA<sup>1</sup>. Igualmente, varios contigs de *T. solium* no encontraron secuencia parecida en esta base de ontologías (incluso aquellos que alineaban con rRNA o snRNA en NCBI con valores muy significativos y que son ampliamente conocidos). Desconocemos la razón de esto último.

El encontrar secuencias de ncRNA en *Taenia solium*, aun cuando se haya identificado un número reducido en comparación con otros estudios, implica que podríamos encontrar más secuencias de esta clase mediante búsquedas más refinadas, tanto por métodos bioinformáticos como experimentales.

---

<sup>1</sup> Hubo que actualizar el trabajo debido a que muchas secuencias de precursores de miRNA fueron eliminadas entre una versión a otra bajo la explicación de que eran erróneas.

## **VIII. CONCLUSIONES**

Hemos identificado bioinformáticamente en los distintos archivos disponibles del genoma de *Taenia solium* 12 secuencias (Tablas B y C) con una semejanza significativa a genes de ncRNA de las bases de datos NONCODE y miRBase.

En el caso de la base NONCODE, se hallaron secuencias similares a genes relacionados con modificaciones postranscripcionales de mRNA, así como de regulación de la transcripción. Se encontraron las distintas clases de snRNA que forman parte del spliceosoma (u1, u2, u4), de las cuales dos ya había sido reportadas anteriormente en *T. solium* (u5 y u6) [75].

En el caso de la base miRBase, se encontraron cuatro secuencias similares a precursores de miRNA. Por el momento, todas las secuencias halladas carecen de información verificable sobre su función, ya que fueron encontradas en estudios de secuenciación masiva o análisis de transcriptomas. Se espera que esto sea resuelto conforme se estudie más detalladamente estas secuencias.

La naturaleza descriptiva de este trabajo, junto con el número reducido de secuencias de ncRNA encontradas, no permite proponer respuesta alguna a las interrogantes actuales sobre la fisiología de este organismo. No obstante, este trabajo ha complementado la labor de anotación del genoma de *T. solium* llevado a cabo por el Consorcio, además de que puede ser mejorado y actualizado conforme se liberen más y mejores bases de ncRNA. Hay que resaltar también que es de los primeros a nivel genómico que se hacen en un platelminto para detectar secuencias de ncRNA (el otro estudio es en *Schistosoma mansoni* [74]), por lo que es un buen inicio para futuros trabajos bioinformáticos y experimentales.

Hay que recordar que la mejor manera de manejar esta enfermedad (aunque en la práctica parece ser muy difícil) sigue siendo el tener un mayor y mejor control sobre la calidad de la carne de cerdo que consumimos, la ingesta de vegetales limpios y desinfectados, así como el seguimiento estricto de prácticas comunes de higiene personal. *Taenia solium* sigue siendo, a la vez de un problema de salud pública, un organismo digno de estudio como modelo de infección en humano y, por qué no, una fuente de sorpresas también.

## **IX. PERSPECTIVAS**

El Proyecto del Genoma de *Taenia solium* ha avanzado bastante durante la realización de este trabajo, mejorando el ensamble de las secuencias y por ende la búsqueda de ncRNA. Podemos pasar de lo virtual (*in silico*) de la computadora, a lo experimental con miras a determinar más secuencias de esta clase y en su caso, específicas de *T. solium*, así como su función en la fisiología de este organismo.

Es así como de éste trabajo se ha desprendido un anteproyecto que plantea profundizar en el estudio de miRNA en este organismo. Específicamente, se intentará modelar una red de regulación genética por miRNA en el que se identifiquen genes “blanco” de interés inmunológico o fisiológico. Se hará uso de herramientas bioinformáticas para identificar genes candidato, experimentales para la detección de miRNA, medición de la regulación de su secuencia blanco, y de modelos matemáticos que simulen estas interacciones.

## **X. BIBLIOGRAFÍA**

1. Flisser, A., *III - Epidemiología*, in *Cisticercosis: guía para profesionales de la salud*. 2006, FCE.
2. Flisser, A., Vargas-Parada, L., Laclette, J.P., *Taenia solium: un parásito cosmopolita*, in *Investigación y Ciencia*. 2006. p. 24-33.
3. Sciutto, E., Fragoso, G., Fleury, A., Laclette, J. P., Sotelo, J., Aluja, A., Vargas, L., Larralde, C., *Taenia solium disease in humans and pigs: an ancient parasitosis disease rooted in developing countries and emerging as a major health problem of global dimensions*. *Microbes and Infection*, 2000. **2**(15): p. 1875-1890.
4. Willms, K., Vargas-Parada, L., Laclette, J.P., *I - Biología del Parásito.*, in *Cisticercosis: guía para profesionales de la salud* 2006, FCE.
5. Harrison, L.J.S., Joshua, G. W. P., Wright, S. H., Parkhouse, R. M. E., *Specific detection of circulating surface/secreted glycoproteins of viable cysticerci in Taenia saginata cisticercosis*. *Parasite Immunology*, 1989. **11**(4): p. 351-370.
6. Deckers, N. and P. Dorny, *Immunodiagnosis of Taenia solium taeniosis/cisticercosis*. *Trends in Parasitology*, 2010. **26**(3): p. 137-144.
7. Handali, S., Klarman, M., Gaspard, A. N., Dong, X. F., LaBorde, R., Noh, J., Lee, Y-M., Rodriguez, S., Gonzalez, A. E., Garcia, H. H., Gilman, R. H., Tsang, Victor C. W., Wilkins, P. P., *Development and Evaluation of a Magnetic Immunochromatographic Test To Detect Taenia solium, Which Causes Taeniasis and Neurocysticercosis in Humans*. *Clin. Vaccine Immunol.*, 2010. **17**(4): p. 631-637.
8. Román, G., Sotelo, J., Del Brutto, O., Flisser, A., Dumas, M., Wadia, N., Botero, D., Cruz, M., Garcia, H., de Bittencourt, P. R., Trelles, L., Arriagada, C., Lorenzana, P., Nash, T. E., and Spina-França, A., *A proposal to declare neurocysticercosis an international reportable disease*. *Bull World Health Organ*, 2000. **78**(3): p. 399-406.
9. Garcia, H.H., Moro, P. L., Schantz, P. M., *Zoonotic helminth infections of humans: echinococcosis, cysticercosis and fascioliasis*. *Current Opinion in Infectious Diseases*, 2007. **20**(5): p. 489-494 10.1097/QCO.0b013e3282a95e39.
10. Sotelo, J., Diaz-Olavarrieta, C., *Neurocysticercosis: Changes after 25 Years of Medical Therapy*. *Archives of Medical Research*, 2010. **41**(1): p. 62-63.
11. Ramírez-Zamora, A., Alarcón, T., *Management of neurocysticercosis*. *Neurological Research*, 2010. **32**: p. 229-237.
12. Hoberg, E.P., *Taenia tapeworms: their biology, evolution and socioeconomic significance*. *Microbes and Infection*, 2002. **4**(8): p. 859-866.
13. Aguilar-Díaz, H., Bobes, R. J., Carrero, J. C., Camacho-Carranza, R., Cervantes, C., Cevallos, M. A., Dávila, G., Rodríguez-Dorantes, M., Escobedo, G., Fernández, J. L., Fragoso, G., Gaytán, P., Garciarubio, A., González, V. M., González, L., José, M. V. Jiménez, L., Laclette, J. P., Landa, A., Larralde, C., Morales-Montor, J., Morett, E., Ostoa-Saloma, P., Sciutto, E., Santamaría, R. I., Soberón, X., de la Torre, P., Valdés, V., Yáñez, J., *The genome project of Taenia solium*. *Parasitology International*, 2006. **55**(Supplement 1): p. S127-S130.
14. Mattick, J.S., *RNA regulation: a new genetics?* *Nat Rev Genet*, 2004. **5**(4): p. 316-323.



15. Brindley, P.J., Mitreva, M., Ghedin, E., Lustigman, S., *Helminth Genomics: The Implications for Human Health*. PLoS Negl Trop Dis, 2009. **3**(10): p. e538.
16. Garciarribio, A., Bobes, R. J., Carrero, J. C., Cevallos, M. A., Fragoso, G., González, V. M., José, M. V., Landa, A., Larralde, C., Mendoza, L., Morales-Montor, J., Morett, E., Sciutto, E., Soberón, X., Laclette, J. P., *The Genome Project of Taenia solium*. International Journal of Infectious Diseases, 2008. **12**(Supplement 1): p. e395-e395.
17. Mayer, D.A., Fried, B., *The Role of Helminth Infections in Carcinogenesis*, in *Advances in Parasitology*, D.R. R. Muller and S.I. Hay, Editors. 2007, Academic Press. p. 239-296.
18. Herrera, L.A., Ostrosky-Wegman, P., *Do helminths play a role in carcinogenesis?* Trends in Parasitology, 2001. **17**(4): p. 172-175.
19. Lewin, B., *Genes VIII*. 2004: Pearson Education.
20. Gilbert, W., *Origin of life: The RNA world*. Nature, 1986. **319**(6055): p. 618-618.
21. Kruger, K., et al., *Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena*. Cell, 1982. **31**(1): p. 147-157.
22. Rodnina, M.V. and W. Wintermeyer, *The ribosome goes Nobel*. Trends Biochem Sci, 2010. **35**(1): p. 1-5.
23. Lee, R.C., Feinbaum, R. L., Ambros, V., *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14*. Cell, 1993. **75**(5): p. 843-854.
24. Costa, F.F., *Non-coding RNAs: New players in eukaryotic biology*. Gene, 2005. **357**(2): p. 83-94.
25. Mattick, J.S., Makunin, I. V., *Non-coding RNA*. Human Molecular Genetics, 2006. **15**(Review Issue 1): p. R17-R29.
26. Napoli, C., Lemieux, C., and Jorgensen, R., *Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans*. Plant Cell, 1990. **2**(4): p. 279-289.
27. Romano, N., Macino, G., *Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences*. Molecular Microbiology, 1992. **6**(22): p. 3343-3353.
28. Allis, C.D., Jenuwein, T. and Reinberg, D., ed. *Epigenetics*. 2007, Cold Spring Harbor Laboratory Press: New York.
29. Fire, A., Xu, SiQ., Montgomery, M. K., Kostas, S. A., Driver, S. E., Mello, C. C., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans*. Nature, 1998. **391**(6669): p. 806-811.
30. Rinn, J.L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Bruggmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., Chang, H. Y., *Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs*. Cell, 2007. **129**(7): p. 1311-1323.
31. Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E. G. H., Margalit, H., Altuvia, S., *Novel small RNA-encoding genes in the intergenic regions of Escherichia coli*. Current Biology, 2001. **11**(12): p. 941-950.
32. Pearson, H., *Genetics: What is a gene?* Nature, 2006. **441**(7092): p. 398-401.
33. Kiss, T., *Small Nucleolar RNAs: An Abundant Group of Noncoding RNAs with Diverse Cellular Functions*. Cell, 2002. **109**(2): p. 145-148.
34. Nizami, Z., Deryusheva, S., Gall, J. G., *The Cajal Body and Histone Locus Body*. Cold Spring Harbor Perspectives in Biology, 2010: p. -.

35. Matera, A.G., Terns, R. M., Terns, M. P., *Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs*. Nat Rev Mol Cell Biol, 2007. **8**(3): p. 209-220.
36. Carthew, R.W., Sontheimer, E. J., *Origins and Mechanisms of miRNAs and siRNAs*. Cell, 2009. **136**(4): p. 642-655.
37. Plasterk, R.H.A., *RNA Silencing: The Genome's Immune System*. Science, 2002. **296**(5571): p. 1263-1265.
38. Díaz-Martínez, L.A., *El ácido ribonucleico salta a la escena*. Ciencia, 2008. **59**(Octubre-Diciembre): p. 44-48.
39. Makarova, K., Wolf, Y., van der Oost, J., Koonin, E., *Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements*. Biology Direct, 2009. **4**(1): p. 29.
40. Shabalina, S.A., Koonin, E. V., *Origins and evolution of eukaryotic RNA interference*. Trends in Ecology & Evolution, 2008. **23**(10): p. 578-587.
41. Abigail F. O., J.G.P., *Genomic organization of microRNAs*. Journal of Cellular Physiology, 2010. **222**(3): p. 540-545.
42. Mourier, T., Willerslev, E., *Retrotransposons and non-protein coding RNAs*. Briefings in Functional Genomics and Proteomics, 2009. **8**(6): p. 493-501.
43. Bushati, N., Cohen, S. M., *microRNA Functions*. Annual Review of Cell and Developmental Biology, 2007. **23**(1): p. 175-205.
44. Lee, Y.S., Dutta, A., *MicroRNAs in Cancer*. Annual Review of Pathology: Mechanisms of Disease, 2009. **4**(1): p. 199-227.
45. Meola, N., Gennarino, V., Banfi, S., *microRNAs and genetic diseases*. PathoGenetics, 2009. **2**(1): p. 7.
46. Hammond, S.M., *Dicing and slicing: The core machinery of the RNA interference pathway*. FEBS Letters, 2005. **579**(26): p. 5822-5829.
47. Mercer, T.R., Dinger, M. E., Mattick, J. S., *Long non-coding RNAs: insights into functions*. Nat Rev Genet, 2009. **10**(3): p. 155-159.
48. Wilusz, J.E., Sunwoo, H., Spector, D. L., *Long noncoding RNAs: functional surprises from the RNA world*. Genes & Development, 2009. **23**(13): p. 1494-1504.
49. Aravin, A.A., Hannon, G. J., Brennecke, J., *The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race*. Science, 2007. **318**(5851): p. 761-764.
50. Hartig, J.V., Tomari, Y., Förstemann, K., *piRNAs—the ancient hunters of genome invaders*. Genes & Development, 2007. **21**(14): p. 1707-1713.
51. Malone, C.D., Hannon, G. J., *Small RNAs as Guardians of the Genome*. Cell, 2009. **136**(4): p. 656-668.
52. Soukup, J.K., Soukup, G. A., *Riboswitches exert genetic control through metabolite-induced conformational change*. Current Opinion in Structural Biology, 2004. **14**(3): p. 344-349.
53. Vitreschak, A.G., Rodionov, D. A., Mironov, A. A., Gelfand, M. S., *Riboswitches: the oldest mechanism for the regulation of gene expression?* Trends in Genetics, 2004. **20**(1): p. 44-50.
54. Amaral, P.P., Dinger, M. E., Mercer, T. R., Mattick, J. S., *The Eukaryotic Genome as an RNA Machine*. Science, 2008. **319**(5871): p. 1787-1789.
55. Baltimore, D., Boldin, M. P., O'Connell, R. M., Rao, D. S., Taganov, K. D., *MicroRNAs: new regulators of immune cell development and function*. Nat Immunol, 2008. **9**(8): p. 839-845.

56. Lindsay, M.A., *microRNAs and the immune response*. Trends in Immunology, 2008. **29**(7): p. 343-351.
57. Dinger, M.E., Mercer, T. R., Mattick, J. S., *RNAs as extracellular signaling molecules*. Journal of Molecular Endocrinology, 2008. **40**(4): p. 151-159.
58. Popa, I., Fradet, Y., Beaudry, G., Hovington, H., Beaudry, G., Tetu, B., *Identification of PCA3 (DD3) in prostatic carcinoma by in situ hybridization*. Mod Pathol, 2007. **20**(11): p. 1121-1127.
59. Bussemakers, M.J.G., van Bokhoven, A., Verhaegh, G. W., Smit, F. P., Karthaus, H. F. M., Schalken, J. A., Debruyne, F. M. J., Ru, N., Isaacs, W. B., *DD3::A New Prostate-specific Gene, Highly Overexpressed in Prostate Cancer*. Cancer Res, 1999. **59**(23): p. 5975-5979.
60. Srikantan, V., Zou, Z., Petrovics, G., Xu, L., Augustus, M., Davis, L., Livezey, J. R., Connell, T., Sesterhenn, I. A., Yoshino, K., Buzard, G. S., Mostofi, F. K., McLeod, D. G., Moul, J. W., Srivastava, S., *PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(22): p. 12216-12221.
61. Tang, X., G. Tang, and S. Özcan, *Role of microRNAs in diabetes*. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 2008. **1779**(11): p. 697-701.
62. Wong, G.K.-S., D.A. Passey, and J. Yu, *Most of the Human Genome Is Transcribed*. Genome Research, 2001. **11**(12): p. 1975-1977.
63. Pollard, K.S., et al., *An RNA gene expressed during cortical development evolved rapidly in humans*. Nature, 2006. **443**(7108): p. 167-172.
64. Zheng, L., L., Qu, L., H., *Computational RNomics: Structure identification and functional prediction of non-coding RNAs in silico*. SCIENCE CHINA Life Sciences, 2010. **53**(5): p. 548-562.
65. Berezikov, E., Cuppen, E., Plasterk, R. H. A., *Approaches to microRNA discovery*. Nat Genet, 2006. **38**: p. S2-S7.
66. Liu, C., Bai, B., Skogerbo, G., Cai, L., Deng, W., Zhang, Y., Bu, D., Zhao, Y., Chen, R., *NONCODE: an integrated knowledge database of non-coding RNAs*. Nucl. Acids Res., 2005. **33**(suppl\_1): p. D112-115.
67. Griffiths-Jones, S., Grocock, R., J., van Dongen, S., Bateman, A., Enright, A., J., *miRBase: microRNA sequences, targets and gene nomenclature*. Nucl. Acids Res., 2006. **34**(suppl\_1): p. D140-144.
68. Pang, K., C., Stephen, S., Dinger, M., E., Engstrom, P., G. Lenhard, B., Mattick, J. S., *RNAdb 2.0--an expanded database of mammalian non-coding RNAs*. Nucl. Acids Res., 2007. **35**(suppl\_1): p. D178-182.
69. Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S., R. Bateman, A., *Rfam: annotating non-coding RNAs in complete genomes*. Nucl. Acids Res., 2005. **33**(suppl\_1): p. D121-124.
70. Couzin, J., *BREAKTHROUGH OF THE YEAR: Small RNAs Make Big Splash*. Science, 2002. **298**(5602): p. 2296-2297.
71. Daub, J., Gardner, P., P.Tate, J., Ramsköld, D., Manske, M., Scott, W., G., Weinberg, Z., Griffiths-Jones, S., Bateman, A., *The RNA WikiProject: Community annotation of RNA families*. RNA, 2008. **14**(12): p. 2462-2464.
72. Eddy, S., R., *Non-coding RNA genes and the modern RNA world*. Nat Rev Genet, 2001. **2**(12): p. 919-929.
73. Mount, D., W., *Bioinformatics: Sequence and Genome Analysis*. Second ed. 2004: Cold Spring Harbor Laboratory Press.

74. Simoes, M., et al., *Identification of Schistosoma mansoni microRNAs*. BMC Genomics, 2011. **12**(1): p. 47.
75. Jean-Joseph, B., et al., *The U5/U6 snRNA Genomic Repeat of Taenia solium*. The Journal of Parasitology, 2003. **89**(2): p. 329-335.

## **XI. APÉNDICE A**

### *Abreviaturas usadas en tablas*

- ID** (primera columna).- Indica el nombre con el cual aparece la secuencia de ncRNA en su respectiva base de datos de origen;
- ID** (segunda columna).- Indica el nombre con el cual se identifica la secuencia o “contig” de *Taenia* (que archivo fue usado está indicado en el encabezado de la tabla);
- Id** (*Identity*).- Muestra el porcentaje de identidad, esto es, el número de bases idénticas entre ambas secuencias alineadas (ej, un alineamiento de 100 nucleótidos de longitud en el cual 90 bases aparecen en el mismo orden tendrá un porcentaje de identidad de 90);
- L** (*Length*).- Indica la longitud del alineamiento generado por *BLAST*;
- M** (*Mismatches*).- Es el número de bases que no son idénticas que permiten continuar un alineamiento local sin generar un “*gap*” o espacio. Son mutaciones que diferencian a dos secuencias y se clasificadas como mutación por transición (purina por purina, pirimidina por pirimidina), o mutación por transversión (purina por pirimidina o viceversa);
- G** (*Gaps*).- Es el número de espacios que se introducen en el alineamiento con la finalidad de optimizar el mismo;
- SQ** (*Start Query*).- Indica el número del nucleótido de la secuencia “sonda” (ncRNA) a partir del cual comienza el alineamiento;
- EQ** (*End Query*).- Indica el número del nucleótido de la secuencia “sonda” (ncRNA) en el que termina el alineamiento;
- SD** (*Start Database*).- La posición del nucleótido en la secuencia de la base de datos (NONCODE o miRBase-Hairpins) en el que comienza el alineamiento con la secuencia “sonda”;
- ED** (*End Database*).- Indica el número del nucleótido dentro de la base de datos (NONCODE o miRBase-Hairpins) en donde termina el alineamiento con la secuencia “sonda”;
- E-value** (*Valor E*).- Valor que da de manera aproximada que tan probable es que de manera aleatoria se dé el alineamiento entre la secuencia “sonda” y la secuencia de la base de datos;
- S** (*Score*).- Es el puntaje que genera *BLAST* de forma numérica y depende de los parámetros elegidos para calificar el alineamiento,
- Lg** (*Largo*).- Es el tamaño de la secuencia de ncRNA tal como aparece en su respectiva base de origen (dato verificado manualmente);
- %**.- Es el porcentaje de la secuencia de ncRNA que hace alineamiento con la secuencia de *T. solium* (dato generado manualmente);
- Ap**.- Es el número de veces que *BLAST* presenta el mismo alineamiento pero cambiando de posición de inicio y término (cuando el programa presenta más de dos alineamientos para un par de secuencias pero que sólo difieren en su sitio de inicio o término, este dato fue generado dentro la manipulación de los datos en Excel).

**Tabla I.- NONCODE vs EST Larva.**

NONCODE vs LARVA														
ID NOCODE	ID contig <i>T. solium</i>	Id	L	M	G	SQ	EQ	SD	ED	E-value	S	Ap	Lg	%
u1740,Homo	031106A_M21_TslvE06132.T3_039.ab1.fg0	98	50	1	0	4598	4647	14	63	4.00E-18	91.7	434	4724	1.06%
u596,Oryza	092906k_H18_Ts_1_073.ab1.fg0	100	45	0	0	3	47	90	134	9.00E-19	89.7	2	280	16.07%
u1311,Arabidopsis	092906n_J21_Ts_1_087.ab1.fg0	91.3	46	4	0	70	115	357	402	4.00E-10	60	4	137	33.58%
u298,Homo	211006a_E08_TslvE0394.T3_019.ab1.fg0	100	37	0	0	71	107	358	394	5.00E-14	73.8	9	238	15.55%
u1468,Mus	211006B_F16_TslvF07105.T3_052.ab1.fg0	86.15	65	9	0	21	85	328	392	9.00E-10	58	1	90	72.22%
*C u3277,Agaricus	061006B_A03_TslvA0197.T3_017.ab1.fg0	100	21	0	0	342	362	206	186	4.00E-04	42.1			
*C u5338,Mus	230606B_G04_TslvG0433.T3_033.ab1.fg0	83.45	139	22	1	33959	34096	163	25	6.00E-15	85.7			

**Tabla II.- NONCODE vs EST Adulto.**

NONCODE vs ADULTO														
ID NOCODE	ID contig <i>T. solium</i>	Id	L	M	G	SQ	EQ	SD	ED	E-value	S	Ap	Lg	%
u1741,Homo	061705c_C22_TsB1098.g1_060.ab1.fg0	96.3	54	2	0	2873	2926	91	38	7.00E-18	91.7	341	4489	1.20%
u2226,Microtus	061705c_I06_TsE06108.g1_085.ab1.fg0	100	34	0	0	7796	7829	624	657	3.00E-10	67.9	2	14102	0.24%
u5338,Mus	140905C_F03_TsF03142.g1_091.ab1.fg0	93.88	98	6	0	33999	34096	103	6	4.00E-33	147	706	32378	0.30%
u1311,Arabidopsis	ljpltsa01024_P08_.g.ab1.fg0	91.3	46	4	0	70	115	361	406	7.00E-10	60	14	137	33.58%
*Cu1930,Trypanosoma	061705c_E21_TsC0998.g1_038.ab1.fg0	100	19	0	0	16	34	627	645	7.00E-04	38.2			
*C u5338,Mus	ljpltsa01006_M06_.g_.019.ab1.fg0	86.67	90	12	0	34007	34096	104	15	4.00E-14	83.8			

**Tabla III NONCODE vs Japón(Full-length).**

NONCODE vs JAPON														
ID NOCODE	ID contig <i>T. solium</i>	Id	L	M	G	SQ	EQ	SD	ED	E-value	S	Ap	Lg	%
U596,Oryza	151007_utk_fl1_f0_040L08.ab1	93.33	45	3	0	3	47	829	873	5.00E-11	65.9	1	280	16.07%
u4043,Gallus	151007_utk_fl1_f0_087K06.ab1	88.31	77	9	0	3	79	67	143	4.00E-16	81.8	4	141	54.61%
u4170,Echinococcus	151007_utk_fl1_f0_102N01.ab1	98.09	157	3	0	1	157	67	223	4.00E-78	287	3	159	98.74%
U3977,Schistosoma	151007_utk_fl1_f0_125G13.ab1	89.94	179	11	4	1	172	62	240	1.00E-47	186	7	194	92.27%
*C u727,Yersinia	151007_utk_fl1_f0_100J14.ab1	96.15	26	1	0	28	53	912	937	5.00E-05	44.1			
*C U5338,Mus	151007_utk_fl1_f0_103°22.ab1	94.87	39	2	0	81438	81476	90	128	3.00E-07	61.9			

**Tabla IV.- NONCODE vs 454AllContigs.**

NONCODE vs 454ALLCONTIGS2010														
ID NOCODE	ID contig <i>T. solium</i>	Id	L	M	G	SQ	EQ	SD	ED	E-value	S	Ap	Lg	%
u3976,Schistosoma	Contig1760.1	85.82	141	20	0	1	141	9052	8912	5.00E-27	121	6	158	89.24%
u4043,Gallus	Contig18938.1	88.31	77	9	0	3	79	606	682	4.00E-15	81.8	1	141	54.61%
u4170,Echinococcus	Contig1914.1	98.1	158	3	0	1	158	1994	2151	1.00E-77	289	23	159	99.37%
u1919,Taenia	Contig194.2	100	133	0	0	1	133	2270	2138	5.00E-70	264	19	133	100.00%
u3977,Schistosoma	Contig31346.1	90	180	10	5	1	173	138	316	8.00E-45	180	19	194	92.78%
u3429,Streptomyces	Contig35492.1	95.65	69	3	0	2581	2649	706	774	3.00E-23	113	1	2964	2.33%
u1139,Gorilla	Contig67373.1	93.24	74	5	0	5	78	376	303	9.00E-23	107	1	188	39.36%
u3540,Erwinia	Contig69572.1	100	68	0	0	31	98	744	677	2.00E-31	135	1	107	63.55%
u1918,Taenia	Contig72707.1	100	113	0	0	1	113	304	416	3.00E-58	224		113	100%
u596,Oryza	Contig72797.1	100	45	0	0	3	47	188	232	3.00E-17	89.7	9	280	16.07%
*C u198,Homo	Contig1161.1	93.55	31	2	0	39	69	1938	1908	1.00E-04	46.1			
*C u5338,Mus	Contig24577.1	94.64	56	3	0	7481	7536	1097	1152	5.00E-14	87.7			

**Tabla V.- Hairpins (miRBase v16) vs EST Larva.**

HAIRPINS vs LARVA														
ID miRBase	ID EST LARVA	Id	L	M	G	SQ	EQ	SD	ED	E-value	S	Lg	%	Ap
mdo-mir-739	010906A_K10_TslvA0151.T3_061.ab1.fg0	93.26	89	6	0	2	90	442	354	4.00E-32	132	97%	96.94%	3
*C Tni-let-7i	010906C_N06_TslvA0158.T3_088.ab1.fg0	100	21	0	0	40	60	233	253	5.00E-05	42.1			

**Tabla VI.- Hairpins (miRBase v16) vs EST Adulto.**

HAIRPINS vs ADULTO														
ID miRBase	ID EST adulto	Id	L	M	G	SQ	EQ	SD	ED	E-value	S	Lg	%	Ap
mdo-mir-739	150605B_F05_TsF0577.b1_043.ab1.fg0	93.26	89	6	0	2	90	153	65	7.00E-32	132	97%	92	3
gga-mir-3533	061705b_F05_TsC0593.g1_071.ab1.fg0	85.88	85	10	2	1	84	78	161	4.00E-19	89.8	92%	92	18
*C ath-MIR865	ljpltsa01008_O07_.b_.01.ab1.fg0	100	21	0	0	62	82	68	48	2.00E-04	42.1			

**Tabla VII.- Hairpins (miRBase v16) vs Japon (Full-length).**

HAIRPINS vs JAPON														
ID miRBase	ID Japon	Id	L	M	G	SQ	EQ	SD	ED	E-value	S	Lg	%	Ap
mdo-mir-739	151007_utk_fl1_f0_130F08.ab1	93.26	89	6	0	2	90	610	698	1.00E-31	132	97%	92	3
*C hsa-mir-28	151007_utk_fl1_f0_058F05.ab1	100	21	0	0	21	41	666	686	2.00E-04	42.1			

**Tabla VIII.- Hairpins (miRBase v16) vs 454AllContigs2010.**

HAIRPINS vs 454ALLCONTIGS2010														
ID miRBase	ID 454AllContigs2010	Id	L	M	G	SQ	EQ	SD	ED	E-value	S	Lg	%	Ap
sme-mir-749	contig02881	86.25	80	11	0	4	83	151725	151804	3.00E-17	87.9	87%	92	1
mdo-mir-739	contig37260	92.55	94	4	3	1	92	55	147	1.00E-30	132	102%	92	1
ssc-mir-4331	contig26401	90.54	74	6	1	1	74	38	110	4.00E-20	97.1	93%	80	17
*C dya-mir-276 <sup>a</sup>	Contig14102.1	100	24	0	0	8	31	1556	1579	3.00E-05	48.1			



## XII. APÉNDICE B

### **Tratamiento para Cisticercosis/NeuroCisticercosis<sup>10</sup>**

#### *Albendazol*

El albendazol es un antihelmíntico de múltiples usos, sobre todo en contra de los nemátodos gastrointestinales entre los cuales se encuentra *Taenia solium*.

**Farmacocinética y farmacodinámica.** Después de la administración oral, el albendazol es absorbido en poca cantidad (menos de 5%). Además la absorción de éste es variable e irregular, aumentando por la presencia de los alimentos adiposos, y posiblemente también por las sales biliares. Su metabolito principal es el sulfóxido de albendazol, el cual se metaboliza con rapidez en el hígado y también en el intestino.

La vida media del metabolito sulfóxido de albendazol es de 8.5 horas aproximadamente. El metabolito es eliminado esencialmente por la orina. Ejerce su efecto antihelmíntico inhibiendo la polimerización de la  $\beta$ -tubulina de las células tegumentarias, depletando los niveles energéticos hasta que éstos llegan a ser insuficientes para la sobrevivencia del parásito. De este modo, el albendazol inicialmente inmoviliza y después ocasiona la muerte de los helmintos susceptibles.

**Dosis y contraindicaciones.** En el tratamiento de la Teniasis larvaria (neurocisticercosis): Se recomienda la administración de 15 mg/kg peso en dos tomas al día, conjuntamente con los alimentos por 14 a 30 días dependiendo de la respuesta. Se recomienda el uso oral o intravenoso de corticosteroides, para prevenir eventos de hipertensión cerebral, durante la primera semana de tratamiento puesto que el principal problema con el tratamiento químico de la NCC es el incremento de la respuesta inflamatoria ante la súbita destrucción del cisticerco.

Si la evidencia diagnóstica sugiere que el individuo no responde al primer esquema de tratamiento, puede responder a un segundo esquema, y éste puede resultar en la curación del paciente. Deben existir dos semanas de intervalo entre cada uno de los esquemas de tratamiento.

---

<sup>10</sup> Información obtenida del Diccionario de Especialidades Farmacéuticas, PLM, disponible desde el sitio web de la Facultad de Medicina, UNAM: <http://www.facmed.unam.mx>

### *Praziquantel*

Praziquantel es un antihelmíntico con amplio espectro de actividad frente a distintos trematodos y céstodos (*Taenia saginata*, *T. solium*, entre otros). Está indicado en el tratamiento de la cisticercosis (neurocisticercosis parenquimatosa y subaracnoidea, y cisticercosis visceral y cutánea), e infecciones por trematodos intestinales (fasciolopsiasis, heterofiasis, metagonimiasis y nanofietiasis), hepáticos (clonorquiasis, opistorquiasis y fascioliasis) y pulmonares (paragonimiasis).

**Farmacocinética y farmacodinamia en humanos:** Praziquantel penetra en el parásito actuando rápidamente (media hora aproximadamente), provocando parálisis espástica del parásito, debido a que altera la permeabilidad de la membrana al calcio, aumentando el paso de este ión al interior del verme; inhibe además la captación de glucosa del parásito, forzándolo a consumir sus propias reservas del glucógeno. Después de 5 minutos de contacto del praziquantel con los vermes, se observa al microscopio electrónico vacuación y degeneración del tegumento. Esta acción del praziquantel también se ejerce sobre los huevecillos y las larvas enquistadas de los vermes.

**Dosis y contraindicaciones:** En el tratamiento de la neurocisticercosis: La dosis diaria es de 50 mg/kg, dividida en 3 tomas durante 15 días ó 75 mg/kg divididos en tres dosis cada 2 horas, seguido de la aplicación de 10 mg de dexametasona I.M. cuatro horas después de la última dosis y en los dos días siguientes. Para la cisticercosis visceral o cutánea: 50 mg/kg divididos en tres tomas (cada 8 horas). El tratamiento requiere de 15 días como mínimo.

Los pacientes con neurocisticercosis después de completar el tratamiento con praziquantel pueden presentar fiebre, náuseas, vómito, meningitis y aumento de la presión intracraneal; síntomas asociados a la respuesta inflamatoria inducida por la destrucción del cisticerco. Para reducir esta sintomatología, se recomienda la administración de corticosteroides (prednisona, 30-40 mg/día, o dexametasona, 12-16 mg/día).

**Costos:** Se prefiere el uso de albendazol debido a que es de menor costo que el Praziquantel, así como de menor interacción medicamentosa, además de presentar una mayor eficacia<sup>11</sup>. Cabe recalcar que es muy importante el uso concomitante de corticosteroides, ya que una reacción inflamatoria desencadenada puede llegar a ser mortal para la persona.

---

<sup>11</sup> Matthaiou, D. K., G. Panos, et al. (2008). "Albendazole versus Praziquantel in the Treatment of Neurocysticercosis: A Meta-analysis of Comparative Trials." *PLoS Negl Trop Dis* 2(3): e194.