

UNIVERSIDAD DE LA REPUBLICA
PROGRAMA DE DESARROLLO DE LAS CIENCIAS BASICAS
AREA BIOINFORMATICA

Tesis presentada para acceder a la Maestría en Bioinformática

Análisis del perfil de expresión de microARNs en el
estadio Juvenil de *Fasciola hepatica* e identificación de
genes de las Vías de Regulación por ARN pequeños en



Santiago Fontenla Martínez

Departamento de Genética, Facultad de Medicina.

Universidad de la República, UDELAR.

Orientadores

Dr. José Tort

Dr. Guilherme Correa de Oliveira

Tribunal

Dr. Fernando Alvarez Valin

Dr. Alfonso Cayota

Dr. Andres Iriarte

AGRADECIMIENTOS

INDICE

RESUMEN.....	5
INTRODUCCIÓN	7
1 <i>Fasciola hepatica</i> como problema sanitario.....	7
1.1 <i>Un ciclo de vida complejo que debe estar finamente regulado</i>	8
1.2 <i>Variación de la expresión génica durante el estadio juvenil</i>	10
2 Los ARNs pequeños como reguladores generales de la transcripción	11
2.1 <i>Descubrimiento de las vías de ARN pequeños</i>	11
2.2 <i>Interferencia por ARNdc en platelmintos</i>	14
3 Vías mediadas por ARNs pequeños en <i>Caenorhabditis elegans</i>	14
3.1 <i>Dicer es la proteína central de las vías de síntesis de ARN pequeños</i>	15
3.2 <i>Las Argonautas guían el silenciamiento</i>	16
3.3 <i>El complejo efector RISC</i>	17
3.4 <i>Vía de ARNi</i>	18
3.5 <i>Vía de microARNs</i>	20
3.6 <i>Vía de síntesis de endoARNs</i>	22
3.7 <i>Silenciamiento nuclear y remodelación de la cromatina</i>	24
3.8 <i>Biogénesis de piRNAs</i>	24
4 Biogénesis de ARNs pequeños en platelmintos	26
HIPÓTESIS	28
OBJETIVOS	28
MATERIALES Y MÉTODOS	29
1 Análisis de las Vías de ARN pequeño en platelmintos	29
1.1 <i>Identificación de proteínas participantes de las vías de ARNi y generación de repositorios</i>	29
1.2 <i>Búsqueda de homólogos por BLAST directo y recíproco</i>	32
1.3 <i>Predicción de dominios proteicos</i>	32
1.4 <i>Identificación de homólogos usando como query proteínas de platelmintos</i>	33
1.5 <i>Confirmación por BLAST con los borradores genómicos</i>	33
1.6 <i>Alineamiento de secuencias, curado manual y construcción de árboles filogenéticos</i>	33

2	Identificación de miARN en el estadio juvenil	34
2.1	<i>Obtención de muestras biológicas</i>	34
2.2	<i>Procesamiento de los datos y filtrado</i>	34
2.3	<i>Análisis de los datos y clasificación de los miARNs</i>	35
2.4	<i>Distribución filogenética y conservación de las familias de miARNs en el estadio juvenil de F. hepatica</i>	36
2.5	<i>Ensamblaje de novo del genoma de F. hepatica</i>	36
2.6	<i>Identificación de miARNs noveles</i>	37
2.7	<i>Identificación de secuencias derivadas de tARNs</i>	37
	RESULTADOS	38
	Capítulo 1. Análisis de vías mediadas por ARN pequeños	38
1	<i>Vía de generación de miARNs</i>	39
1.1	<i>El complejo microprocesador</i>	42
1.2	<i>Búsqueda de homólogos de Dicer</i>	46
1.3	<i>Identificación de Argonautas en platelmintos</i>	51
1.4	<i>Localización genómica de las FL-Agos</i>	53
	Capítulo 2 miARNs expresados en el estadio NEJ de F. hepatica	57
2	<i>Artículo: el miRNoma del juvenil de F. hepatica</i>	59
	DISCUSIÓN	72
	<i>Identificación de proteínas de la vía de miARNs en platelmintos</i>	72
	<i>Los platelmintos presentan más de una proteína de tipo Dicer</i>	74
	<i>En platelmintos hay dos subgrupos de Argonautas</i>	75
	<i>Los microARN del estadio NEJ de F. hepatica</i>	77
	<i>Los platelmintos presentan una reducción de su miRNoma</i>	77
	<i>La menor conservación de la secuencia de los miARNs de platelmintos sugiere una alta tasa mutacional</i>	78
	<i>Secuencias derivadas de ARNt integran la población de ARN pequeños</i>	79
	CONCLUSIÓN	80
	BIBLIOGRAFÍA	81
	APENDICE	92

RESUMEN

El advenimiento de las nuevas tecnologías de secuenciación masiva ha permitido conocer los genomas y transcriptomas de varios platelmintos parásitos responsables de enfermedades relevantes tanto en la salud humana como animal. Esto ha revelado aspectos relevantes de la biología parasitaria, pero sin embargo, buena parte de las secuencias generadas aún no han sido exploradas por lo que es necesario realizar estudios que permitan su identificación. La Bioinformática ofrece métodos rápidos, económicos y robustos para el análisis de este tipo de datos.

Los ARN pequeños han surgido como reguladores relevantes de la expresión génica durante el desarrollo en distintos organismos. Dada la disponibilidad de datos en platelmintos parásitos en el presente trabajo evaluamos la conservación de las vías de silenciamiento génico en los platelmintos, focalizando en la regulación mediada por microARNs. Hallamos que la vía de silenciamiento por microARNs esta conservada en todos los platelmintos pero presenta características diferentes. Identificamos variantes exclusivas de platelmintos en *Dicers* y *Argonautas* dos de los principales componentes de las vías mediadas por ARN pequeños. En particular detectamos un nuevo grupo de proteínas *Dicer* con variantes en su estructura proteica entre distintos linajes de platelmintos. Por otro lado detectamos un nuevo grupo de proteínas *Argonauta* exclusivo de platelmintos.

Por otro lado, tomando como modelo el trematodo *Fasciola hepatica*, estudiamos el perfil de expresión de ARNs pequeños en el estadio invasivo al hospedero mamífero y lo comparamos con los datos previamente reportados en adultos. Identificamos 40 microARNs distintos (incluyendo algunos noveles) ampliando la cantidad de familias de miARNs conocidas en este trematodo. Si bien la mayoría de los miARNs identificados pertenecen a familias ancestrales conservadas entre organismos protostomados y metazoarios,

observamos una reducción en el miRNoma dada la ausencia de varias otras familias conservadas en metazoarios.

Notablemente hallamos una menor conservación de secuencia en los miARNs maduros entre especies de platelmintos en comparación con otros metazoarios, sugiriendo que la tasa mutacional en los platelmintos parásitos podría ser alta.

Por otro lado detectamos una población de ARN pequeños de 32 a 33 nucleótidos que fueron identificadas como secuencias derivadas de ARN de transferencia, la mayoría correspondientes a mitades 5' de tARNs (tRNA halves). Se ha detectado que tARN halves son reguladores de la expresión génica en otros organismos, por lo tanto podrían intervenir en la regulación de la expresión génica junto con los miARNs en este parásito.

Los resultados obtenidos en este trabajo muestran que los platelmintos presentan varias particularidades en los mecanismos de regulación mediados por ARN pequeños en comparación con otros organismos. Estas singularidades requieren un estudio más detallado y pueden aportar al conocimiento de los mecanismos de regulación en general y potencialmente ser un blanco relevante para el diseño de nuevas drogas o estrategias diagnósticas.

INTRODUCCIÓN

1 *Fasciola hepatica* como problema sanitario

Fasciola hepatica pertenece a la clase Trematoda del phylum platelmintos (que deriva de las palabras griegas ‘platys’ que significa *plano* y ‘helmintos’ que significa *gusano*). Junto a *Fasciola gigantica* producen la fasciolosis, una enfermedad de distribución mundial que afecta principalmente al ganado pero también al hombre (http://www.who.int/foodborne_trematode_infections/fascioliasis/en/). La fasciolosis es una zoonosis, lo que significa que la enfermedad puede ser transmitida de animales a humanos. Ambas especies de *Fasciola* se superponen en varias áreas de África y Asia, mientras que *F. hepatica* se distribuye predominantemente en áreas templadas y resulta el causante de la enfermedad en las Américas, Europa y Oceanía (Mas-Coma, Bargues, and Valero 2005) (Figura 1).

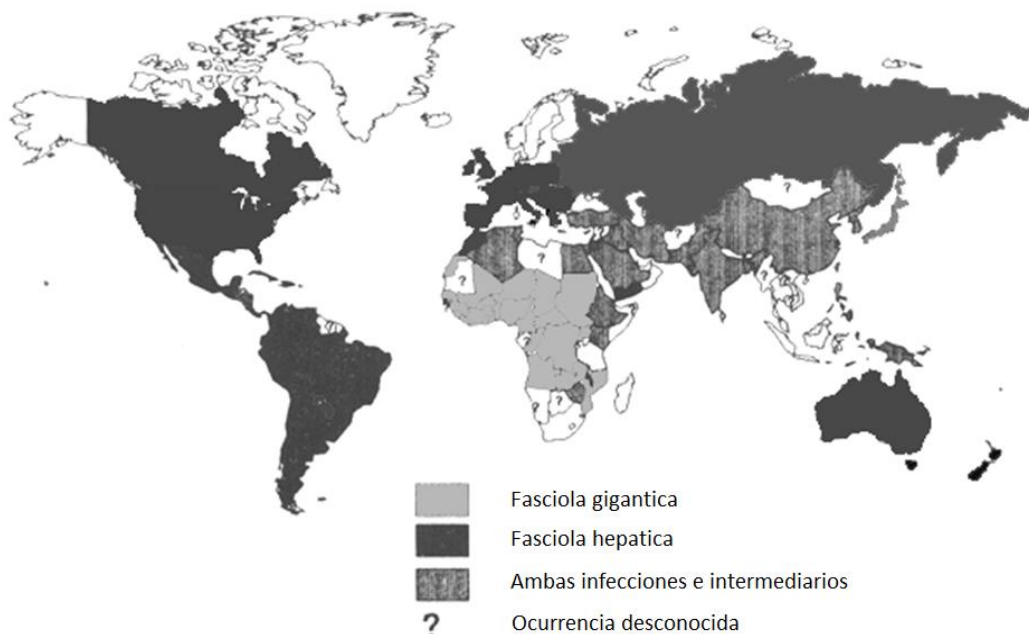


Figura 1. Distribución de *Fasciola* spp.

F. hepatica se distribuye predominantemente en áreas templadas mientras que *F. gigantica* predomina en áreas tropicales de África y Asia.

F. hepatica infecta a más de 300 millones de bovinos y 250 millones de ovinos a nivel global, produciendo pérdidas económicas estimadas en 3 billones de dólares anualmente debido a la reducción en la productividad de los animales infectados (Mas-Coma, Bargues, and Valero 2005; Robinson and Dalton 2009). En nuestro país entre el 50 y 60% del ganado bovino y ovino está infectado (Acosta D 1991; Sanchis, Suarez, et al. 2015; Sanchis, Hillyer, et al. 2015). El impacto económico que produce esta parasitosis es importante si se tiene en cuenta que se pierden anualmente 6 millones de dólares solo a causa del decomiso de hígados (según los datos de la Auditoría de Calidad de la Cadena Cárnica de 2007-2008). A esto se le deben sumar las pérdidas económicas ocasionadas por anemia, disminución en la producción de leche, reducción en la fertilidad y tratamiento de los animales infectados (Sanchis, Hillyer, et al. 2015; Sariözkan and YalÇin 2011).

Adicionalmente a la afectación de especies productivas, se estima que entre 2.4 y 17 millones de personas están infectados en el mundo y que 91 millones están en riesgo (Keiser and Utzinger 2005; Keiser and Utzinger 2009). Las áreas rurales de las regiones Andinas de Perú y Bolivia son las regiones más afectadas en el mundo con rangos de prevalencia de entre 6 y 68% (Marcos, Terashima, and Gotuzzo 2008; Parkinson, O'Neill, and Dalton 2007). Si bien la enfermedad es tratable con antihelmínticos como el triclabendazol, ya han sido descritas cepas resistentes a esta droga en varios países como Australia, Irlanda, Reino Unido, Países Bajos y España.

1.1 Un ciclo de vida complejo que debe estar finamente regulado

El ciclo de vida de *F. hepatica* comienza cuando los huevos, liberados al ambiente en las heces contaminadas del hospedero definitivo, alcanzan un espejo de agua fresca donde se desarrollan a miracidios. Los miracidios liberados del huevo infectan caracoles del género *Lymnea*, que son el hospedero intermediario (Figura 2).

Dentro del caracol los miracidios se transforman en esporocistos, y estos, por amplificación asexual dan lugar a redias (una forma intermedia, que permanece parasita del caracol) y cercarías. Las cercarías se liberan del caracol y se enquistan en plantas acuáticas. Estas formas quísticas de resistencia denominadas metacercarias, son las que infectan al hospedero definitivo al ser ingeridas junto a la vegetación (o vegetales infectados en el caso humano) (Keiser and Utzinger 2005).

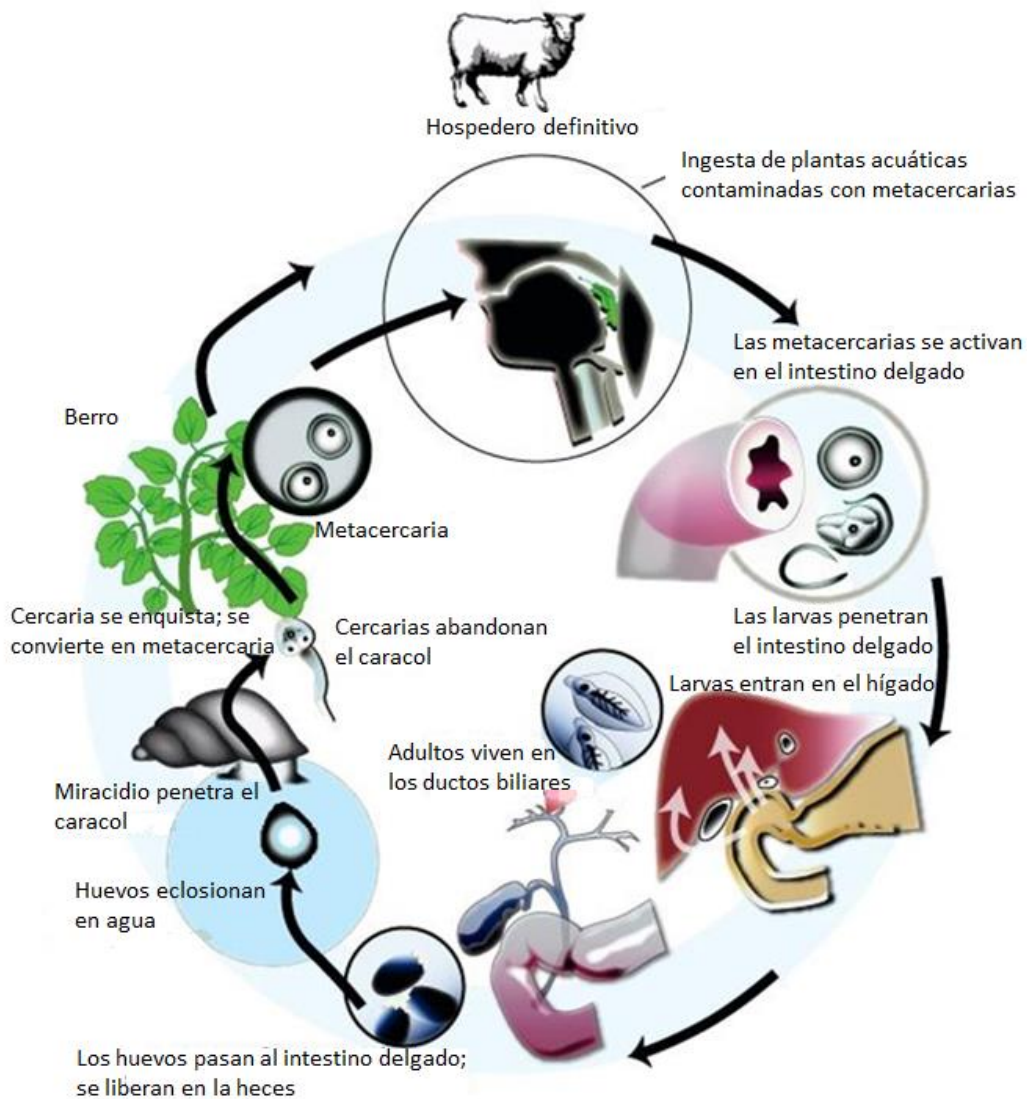


Figura 2. Ciclo de vida de *F. hepatica*

Una vez en el tubo digestivo del hospedero definitivo diversos factores (las enzimas digestivas, las sales biliares, el potencial redox, el pH, la temperatura) activan a la metacercaria, que se desenquista a una forma juvenil (NEJ por sus siglas en inglés). Los juveniles atraviesan la pared duodenal y migran en la cavidad peritoneal hasta alcanzar y penetrar el hígado. Las formas inmaduras que migran por el parénquima hepático generan lesiones importantes que son la causa principal de la enfermedad aguda. El destino de esta migración son los canalículos biliares mayores donde el parásito madura sexualmente alcanzando el estadio adulto y genera miles de huevos diarios.

1.2 Variación de la expresión génica durante el estadio juvenil

A diferencia de lo que sucede con el parásito adulto que sobrevive en el ambiente inmunológicamente seguro de los canalículos biliares, los NEJ son el primer contacto del parásito con el huésped, y a lo largo del proceso de invasión se ven enfrentados a distintos microambientes, interactuando directamente con diversos tejidos y sistemas del huésped. En nuestro grupo se demostró que este proceso está acompañado de una expresión diferencial de genes, con algunos de ellos de expresión exclusiva en los estadios juveniles (Cancela et al. 2008; Corvo et al. 2009; Cancela et al. 2010). Estas variaciones en la expresión de genes entre los diferentes estadios, deben estar finamente reguladas. Como forma de estudiar algunos de estos mecanismos nuestro grupo apostó al silenciamiento génico mediado por ARN pequeños (Dell'Oca et al. 2014).

2 Los ARNs pequeños como reguladores generales de la transcripción

El descubrimiento de las vías de silenciamiento génico mediada por ARN pequeños transformó la forma en que entendemos la regulación de la expresión génica. Los principales mediadores del silenciamiento en estas vías son ARN pequeños de 20-30 nucleótidos de los cuales los más conocidos son microARNs (miARNs), ARN pequeños interferentes (siARNs) y ARNs de unión a proteínas Piwi (piARNs).

Otros grupos de ARN pequeños con capacidad de regular de la expresión génica han sido identificados recientemente, como los tARN halves derivados de tARNs (Thompson and Parker 2009; Garcia-Silva et al. 2010), heterochromatin small interfering RNA (hc-siARN), trans acting siRNA (ta-siARN), natural antisense short interfering RNA (nat-siARN) y DNA damage-induced small RNA (qiARN)(Choudhuri 2009). Aunque poco se sabe de estos nuevos ARN pequeños su descubrimiento plantea nuevas interrogantes sobre los sistemas de regulación génica en los seres vivos.

2.1 Descubrimiento de las vías de ARN pequeños

Las vías de regulación mediadas por ARN pequeños han sido estudiadas principalmente en el nematodo *Caenorhabditis elegans*, primer organismo en que se describió la interferencia de ARN (ARNi) por medio de un ARN doble cadena exógeno (ARNdc), y en que se identificó por primera vez un miARN, lin-4 (Lee 1993). Los primeros indicios fueron publicados por el grupo de Fire en 1991, donde estudiando el silenciamiento por ARNs exógenos sospecho de la existencia de una nucleasa de ARN doble cadena (Fire et al. 1991) (Figura 3).

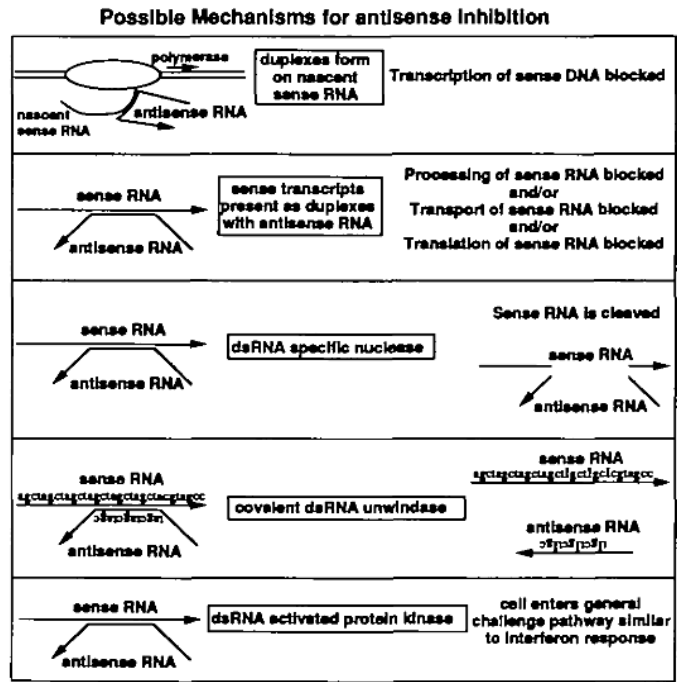


Figura 3. Primeros indicios del mecanismo de silenciamiento mediado por ARN en *C. elegans* (Tomado de Fire et al. 1991)

Posteriormente el mismo grupo describió la vía de silenciamiento mediada por ARN doble cadena que hoy conocemos como ARNi (Fire et al. 1998) (Figura 4).

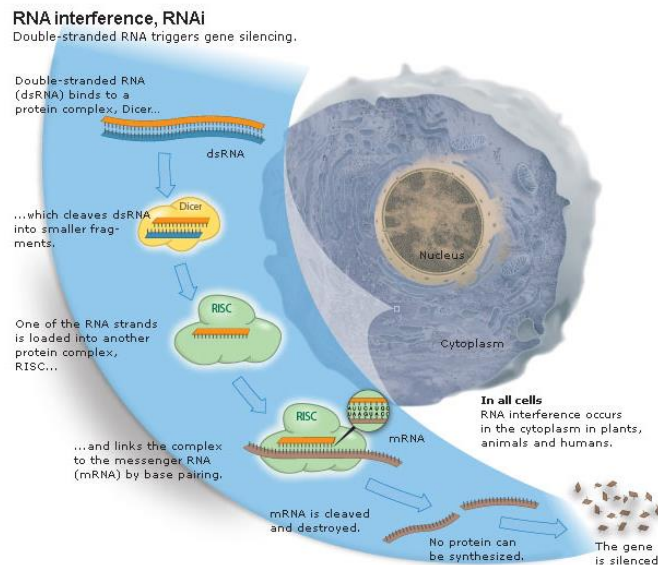


Figura 4. Esquema del mecanismo de RNAi descrito en *C. elegans*

En el año 1999 se reportan los primeros genes participantes de las vías de ARN pequeños por la identificación de cepas mutantes de *C. elegans* resistentes a la interferencia (Tabara et al. 1999; Ketting et al. 1999). A partir de entonces, rápidamente comenzaron a identificarse miembros de estas vías en otros organismos tan diversos como humanos e insectos.

En 2001, Grishok y col. (Grishok et al. 2001) identificaron el ortólogo en *C. elegans* de Dicer, la proteína central de las vías de ARN pequeñas, que fuera reportada ese mismo año en *Drosophila* (Bernstein et al. 2001), y construyen el primer modelo de las vías. En este modelo ya se diferencian las tres grandes vías de generación de ARN pequeños, la vía de microARNs, la vía de ARNi y la vía de ARNs endógenos aún poco conocida en ese momento (Figura 5).

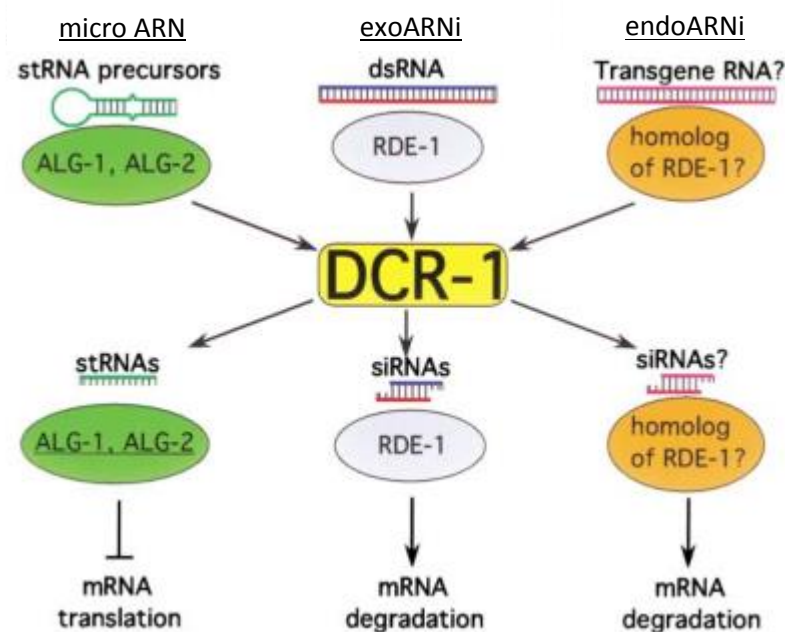


Figura 5. Primer modelo de las vías de ARN pequeños en *C. elegans*.

Se diferencian las tres grandes vías de generación de ARN pequeños, la vía de microARNs, la vía de ARNi y la vía de ARNs endógenos. Tomado y modificado de Grishok et al, 2001.

Estos descubrimientos ofrecían una vía sencilla de estudiar la función génica en organismos donde no existía información genómica para generar silenciamiento por knockout. Esto condujo a estudiar la existencia y aplicación de estas vías de ARN pequeños en otros organismos no modelo, por su gran potencial como herramienta para el estudio de la función génica (ARNi) y por su rol en la regulación de la expresión génica (miARNs).

2.2 Interferencia por ARNdc en platelmintos

El silenciamiento génico ha sido demostrado en varios platelmintos, y ha resultado fundamental en el estudio de la función génica en planarias (Sánchez Alvarado and Newmark 1999). En platelmintos parásitos esta herramienta se ha desarrollado principalmente en los *Schistosomas* (Krautz-Peterson et al. 2010; Sripta et al. 2011; Newmark et al. 2003). El grupo del cual formo parte a sido pionero en esta área, primero logrando el silenciamiento del gen reportero luciferasa en *S. mansoni* y *F. hepatica* (Rinaldi et al. 2008), y recientemente optimizando el silenciamiento usando como modelo el gen de la aminopeptidasa de leucina (LAP) en el estadio juvenil de *F. hepatica* (Dell'Oca et al. 2014). Estos resultados demuestran la existencia de una vía de silenciamiento por ARN exógeno, en estos organismos, y abrieron la puerta a pensar en que posiblemente otros mecanismos mediados por ARN pequeños pudieran estar teniendo lugar en la regulación normal en el ciclo de *F. hepatica*.

3 Vías mediadas por ARNs pequeños en *Caenorhabditis elegans*

C. elegans es el organismo donde mejor se han descrito las vías de regulación mediadas por ARNs pequeños y es la referencia más utilizada por los trabajos que investigan las vías de ARN pequeños en organismos no modelo.

Los primeros resultados de genes asociados a las vías surgieron a partir del estudio en *C. elegans* que presentaban mutaciones que los hacían resistente a la interferencia por ARN.

Así se identificaron proteínas como las RDE (**R**nai **D**Efficient) o las MUT (**M**UTator) (Tabara et al. 1999; Ketting et al. 1999), las que fueron numeradas consecutivamente a medida que se fueron describiendo. Sin embargo en la mayoría de los casos, los genes codificantes de estas proteínas no forman parte de una misma familia génica o no cumplen funciones similares. Por ejemplo, mientras que RDE-1 es una proteína de la familia Argonauta, RDE-4 es una proteína de unión a ARN doble cadena (ARNdc). Esta nomenclatura puede resultar algo confusa, y por lo tanto, en las descripciones de las vías que haremos más adelante indicaremos la función asociada a los nombres de los genes en aquellos casos que sea relevante.

3.1 *Dicer es la proteína central de las vías de síntesis de ARN pequeños*

Dicer fue identificada como una ribonucleasa que funciona al inicio de la ARNi (Bernstein et al. 2001). Pertenece a la familia de las enzimas Ribonucleasas de tipo III (RNasa III), un grupo de enzimas que tienen especificidad por ARNdc. La familia de las RNasa III se divide en tres clases estructurales. Dicer pertenece a la clase III y contiene un dominio DEXH-box, un dominio ARN helicasa, dos dominios Ribonucleasa III, un dominio PAZ y un dominio de unión a ARNdc (DSRM o dsRBD) (Figura 6).

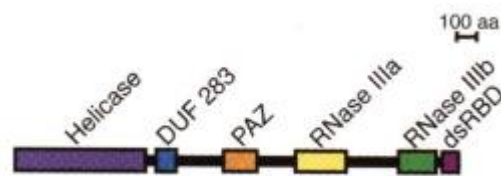


Figura 6. Estructura de Dicer. Tomado de MacRae et al. 2006.

Las Dicer tienen un centro de procesamiento de ARN. Dentro de este centro, cada dominio Ribonucleasa III cliva una de las hebras del ARNdc. La dimerización intramolecular de los dominios Ribonucleasa III, en cooperación con los dominios de unión a ARN, PAZ y dsRBD, resultan en un centro de procesamiento que genera ARNdc de aproximadamente 21 nucleótidos (nt), con un desfase de dos nucleótidos en 3' (Jaronczyk, Carmichael, and Hobman 2005).

3.2 Las Argonautas guían el silenciamiento

Las proteínas Argonauta forman una superfamilia altamente conservada que se encuentran en diversos organismos desde archaeas a humanos (Jaronczyk, Carmichael, and Hobman 2005). En humanos existen 8 genes Argonauta, 5 en *Drosophila* (Jaronczyk, Carmichael, and Hobman 2005; Hutvagner and Simard 2008), sin embargo el genoma de *C. elegans* codifica 26 de estas proteínas (Hutvagner and Simard 2008). Aparentemente, en algunos Nematodos las Argonauta cumplen funciones altamente especializadas lo que explicaría esta amplificación (Dalzell et al. 2011; Hutvagner and Simard 2008).

Las Argonautas contienen dos dominios funcionales, un dominio PAZ y un dominio C-terminal, PIWI. El dominio PAZ se une al desfase de 2 nucleótidos en 3' de los ARNdc pequeños generados por Dicer y el dominio catalítico PIWI cliva la hebra complementaria o pasajera de los ARNdc pequeños (Hutvagner and Simard 2008). Esto genera el ARN pequeño maduro (siARN o miARN) que se mantiene asociado a la proteína Argonauta, y junto a otros factores forman el complejo efector RISC (RNA induced Silencing Complex) que silencia la traducción del ARNm (Grishok 2005; Sylvia E J Fischer 2010).

3.3 El complejo efector RISC

El complejo RISC está formado mínimamente por una proteína Argonauta asociada a un miARN o siARN. Sin embargo, se han identificado otros factores que forman parte del complejo RISC, principalmente por ensayos de purificación (Caudy et al. 2002)(Caudy et al. 2003).

Otros componentes identificados asociados a RISC son la enzima TSN-1 (Tudor Staphylococcal Nuclease) que tiene cinco dominios homólogos a la nucleasa microcoala y podría ser la enzima exonucleasa de RISC (Grishok 2005), una proteína de unión a ARN llamada VIG-1 (Vasa Interonic Gene) (Caudy et al. 2002; Chan et al. 2008) y dos proteínas que interactúan directamente con las proteínas ALG y que están localizadas en los cuerpos P llamadas AIN-1 y 2 (Kuzuoglu-Öztürk et al. 2012).

El silenciamiento de la traducción puede ocurrir por dos mecanismos. En la vía exógena o de ARNi la complementariedad perfecta entre el siARN y su ARNm blanco produce la degradación del ARNm blanco mediada por el dominio PIWI de la proteína Argonauta del complejo RISC (o en este caso siRISC) (Figura 7).

En el caso de los miARNs o miRISC la unión es más inespecífica y puede haber múltiples mARN blanco para un mismo miARN. Básicamente, la especificidad del blanco de un miARN está dada por la complementariedad de secuencia entre la región semilla (nucleótidos 2-8) del miARN y la región 3' no traducida (3' UTR) del ARN mensajero blanco. En este caso el mecanismo efector sería el bloqueo de la traducción y la desestabilización del ARN mensajero (Figura 7).

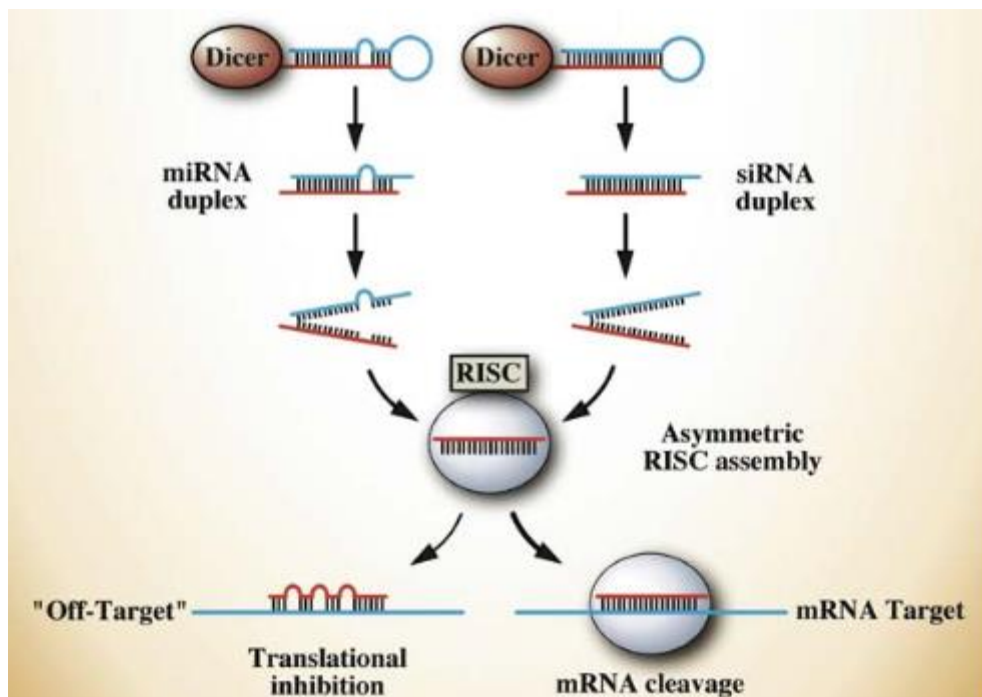


Figura 7. Diferencias en el mecanismo de silenciamiento mediado por siARN o miARN.

3.4 Vía de ARNi

En varios organismos la introducción de ARNdc exógeno puede inducir el silenciamiento post-transcripcional de un gen específico. Esta respuesta es mediada por la vía de ARN interferente (también llamada vía de ARN exógeno). La vía de ARNi como ya se dijo, fue la primera descubierta, esta conservada a lo largo de la evolución y posiblemente surgió como un mecanismo de defensa contra ARNdc virales u otros parásitos genéticos (Zamore and Haley 2005).

En *C. elegans* los ARNdc largos son transportados dentro de las células y procesados en siARNs de aproximadamente 22 nt. La incorporación de ARNdc está mediada por unas proteínas de canal transmembrana llamadas SID-1 y SID-2 (van Roessel and Brand 2004)(Figura 8).

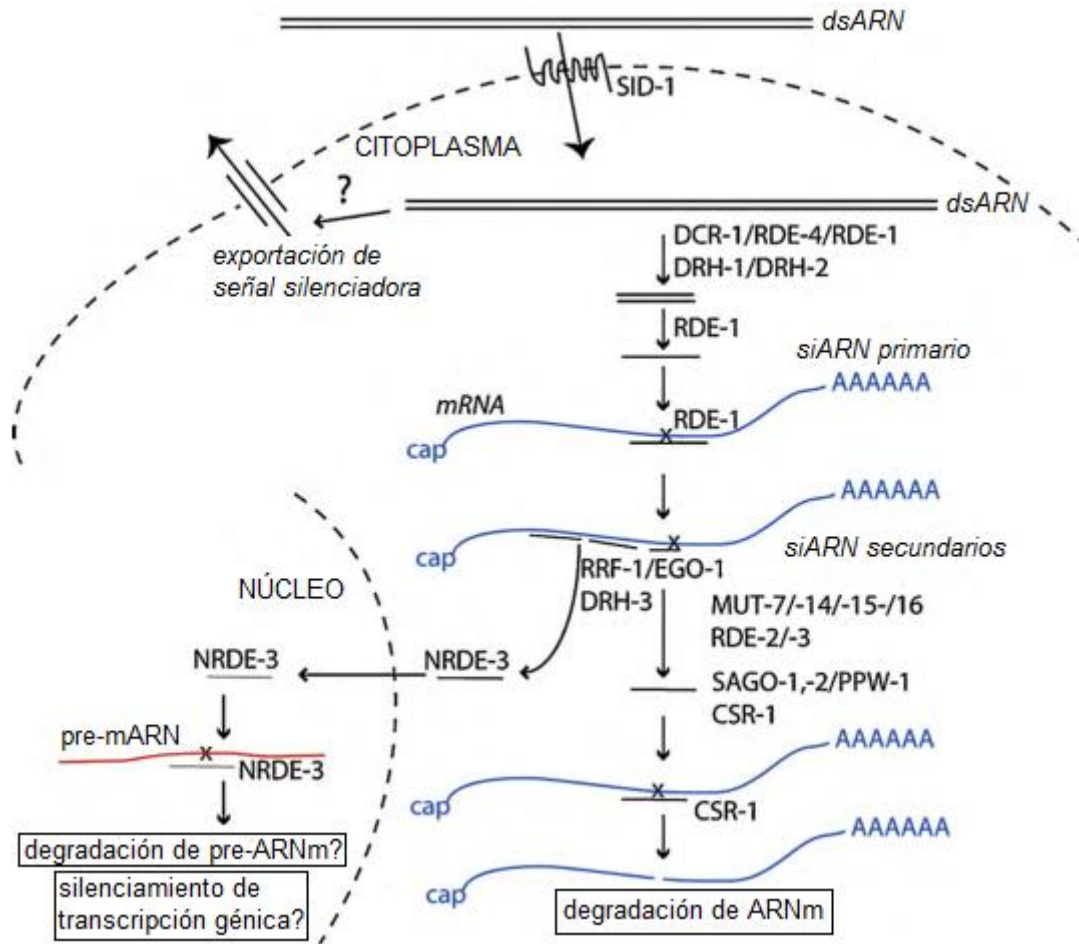


Figura 8. Vía de ARNi en *C. elegans*.

Los ARNdc son procesados por un complejo formado por RDE-4, DCR-1, DRH-1/2 y RDE-1. RDE-1 junto con el siARN maduro silencian la traducción de su ARNm blanco. A su vez, el siARN unido al ARNm es usado como plantilla para producir nuevos siARNs produciendo la amplificación de la señal interferente (Figura tomada y modificada de Fischer 2010).

En el citoplasma el ARNdc exógeno es reconocido por RDE-4, una proteína de unión a ARNdc largo. RDE-4 unido a ARNdc promueve la formación de un complejo que procesa el ARNdc largo en siARNs maduros. Este complejo está compuesto por la enzima RNasa III DCR-1 (el ortólogo en *C. elegans* del gen Dicer descrito en *Drosophila*) que cliva procesivamente el ARNdc.

Dos proteínas con dominio DexH-box/Helicasa (DRH-1 y DRH-2) y una proteína Argonauta (RDE-1) también forman parte de este complejo (Tabara et al. 2002). RDE-1 se une a los siARNs estabilizándolos y cliva la hebra complementaria o pasajera. Los siRNA maduros permanecen unidos a RDE-1 formando el complejo RISC.

En algunos organismos como las levaduras, las plantas y *C.elegans* ocurre un fenómeno de amplificación de la señal interferente (Jamalkandi and Masoudi-Nejad 2011). La unión de siARN al ARNm es usado como plantilla para la síntesis de ARN, por unas enzimas llamadas en conjunto ARN polimerasas dependientes de ARN (RdRP, RNA dependent RNA Polimerases) (RRF-1 y EGO-1 en la Figura 8). Estas enzimas extienden el siRNA por su extremo 3' tomando como molde al ARNm. Esto regenera un ARNdc que es procesado por un grupo de proteínas Argonauta secundarias (SAGOs, secondary argonautes), que resulta en nuevos siARN, amplificando la señal silenciadora (Fig.8) Estos ARNdc generados por la amplificación pueden ser exportados a otras células, permitiendo que el silenciamiento génico se extienda a otros tejidos (Fischer 2010) (Jamalkandi and Masoudi-Nejad 2011).

3.5 Vía de microARNs

Los primeros miARNs descritos fueron *lin-4* y *let-7* por su papel como reguladores del desarrollo en *C. elegans*. *Lin-4* parece ser relevante en la transición del estadio larval uno al dos, mientras que *let-7* controla el pasaje del estadio larval tardío a adulto (Lee 1993; Reinhart et al. 2000).

Los miARNs son secuencias cortas de 21 a 23 nucleótidos de largo que regulan negativamente la traducción de genes codificantes de proteínas. Esto produce la degradación o el bloqueo de la traducción del mensajero. La presencia de miARNs ha sido descrita en organismos tan diversos como plantas, metazoarios y virus de eucariotas (Tarver et al. 2013; Wheeler et al. 2009).

Los locus de los miARNs se encuentran ubicados generalmente en regiones intergénicas y cuentan con las secuencias regulatorias reconocidas por la ARN polimerasa II, por lo que una vez transcritos se les agrega el capuchón 5' y la cola poly-A. Posteriormente los transcriptos primarios, llamados pri-miARNs, son procesados dentro del núcleo a pre-miARNs por el complejo microprocesador formado por una proteína de unión a ARNdc llamada Pasha (PASH-1) y una proteína con actividad RNasaIII llamada Drosha (DRSH-1) (Figura 9).

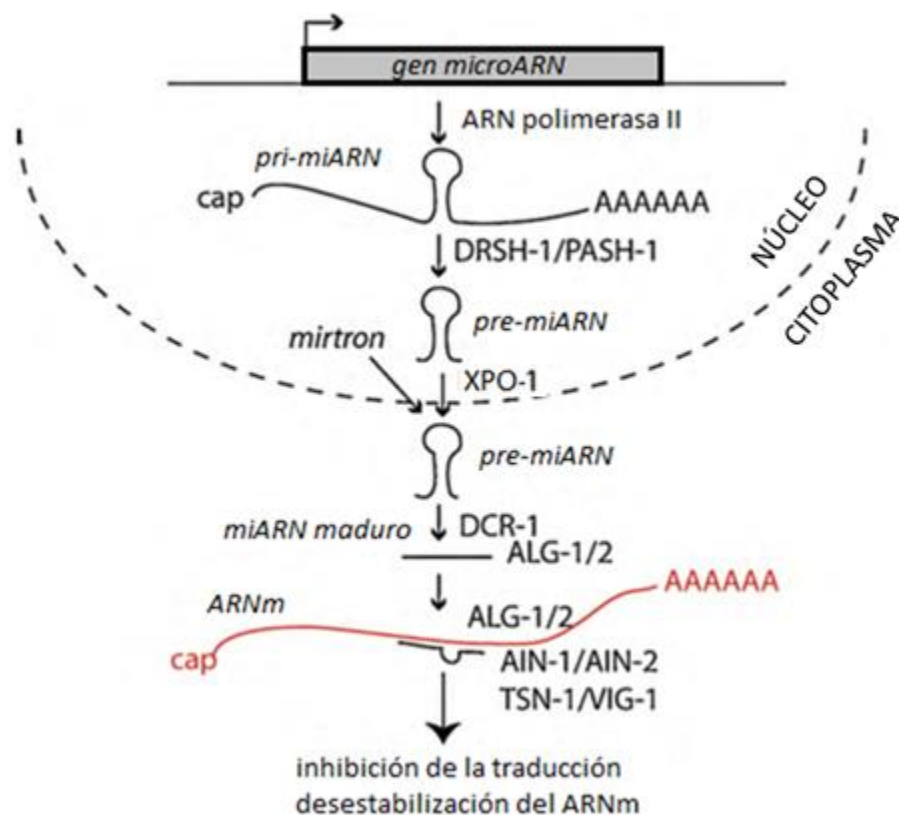


Figura 9. Vía de síntesis de miARNs en *C.elegans*.

Los pri-miARNs son procesados por el complejo microprocesador, los pre-miARNs son exportados al citoplasma donde son procesados por Dicer y ALG-1/2 en miARN maduros. El complejo miRISC inhibe la traducción y desestabiliza el ARNm (tomado y modificado de Fischer et al. 2010)

Los pre-miARNs tienen una estructura en horquilla de aproximadamente 70 nucleótidos de largo y son exportados al citoplasma desde el núcleo por una canal de membrana llamado exportina (XPO-1) (Büssing et al. 2010). A continuación, los pri-miRNAs son procesados por la enzima RNasaIII Dicer junto con las proteínas argonautas ALG (Argonaute Like

Gene), ALG-1/2 cumplen una función similar a la de RDE-1 en la vía de RNAi separando la hebra madura de la complementaria, generando un miARN maduro. La hebra madura junto ALG-1 y ALG-2 forman el complejo miRISC (Figura 9).

Los miARNs también pueden estar codificados en regiones intrónicas de genes. A este tipo de miARNs se los llama mirtrons. Son procesados dentro del núcleo por la maquinaria de splicing sin recurrir al complejo Microprocesador y son exportados al citoplasma como otros pre-miARNs.

3.6 Vía de síntesis de endoARNs

Además de las vías de miARNs y siARNs existe una vía de interferencia dependiente de ARNdc largos endógenos derivados de ARN antisentido de secuencias codificantes y elementos transponibles, que parecen regular genes expresados en la línea germinal (Pavelec et al. 2009). Si bien esta vía aún no ha sido tan bien estudiada como las anteriores, la vía de ARNi endógeno presenta muchas similitudes con la vía de ARNi exógeno.

La vía endógena de ARNi comparte a las proteínas de unión a ARNdc largo, RDE-4, y a la RNasaIII Dicer (DCR-1) (Duchaine et al. 2006) con la vía exógena. ERGO-1 es la argonauta de esta vía y podría cumplir una función similar a RDE-1 en la vía exógena (Figura 10)(Fischer 2010).

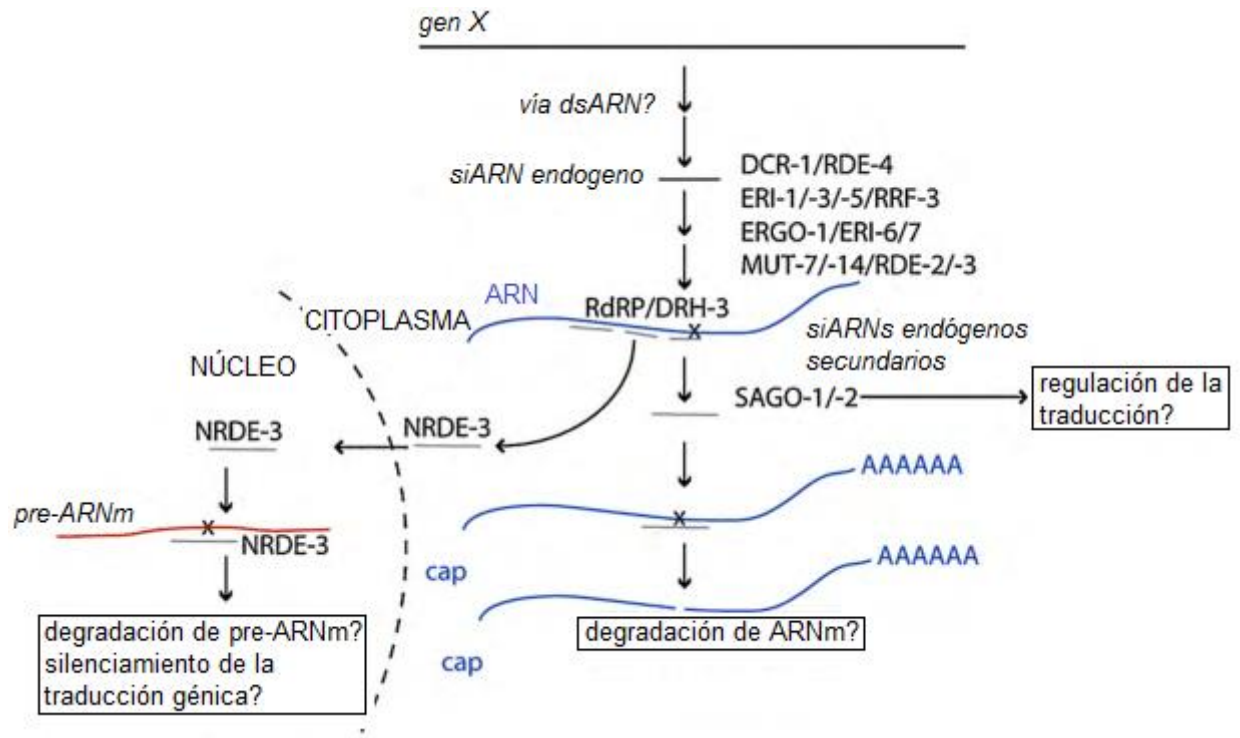


Figura 10. Vía de ARNi endógeno.

El complejo ERI/Dicer es el complejo efector de esta vía. Como en la vía exógena de ARNi, en la vía endógena también puede ocurrir una amplificación de la señal interferente.

Al igual que en la vía exógena de ARNi los ARNm blancos de los siARN endógenos son degradados. Participa un complejo efector tipo RISC denominado complejo ERI/Dicer (Pavelec et al. 2009). Esta formado por una proteína con dominio Helicasa DRH-3 que esta en complejo con Dicer y es esencial para el funcionamiento de la vía endógena (Duchaine et al. 2006). A su vez, DRH-3 y Dicer están en complejo con una ARN polimerasa dependiente de ARN, llamada RRF-3, y con otros factores llamados en conjunto ERI (Enhanced RNAInterference), nombrados así porque su silenciamiento potencia la interferencia mediada por ARNdc exógeno. Esos resultados hicieron suponer que existe una situación de competencia por factores comunes entre las vías de ARNi exógena y endógena (Duchaine et al. 2006; Fischer 2010). Como en la vía exógena de ARNi, en la vía endógena también puede ocurrir una amplificación de la señal interferente.

3.7 Silenciamiento nuclear y remodelación de la cromatina

Una tercera vía derivada es el silenciamiento de la transcripción a nivel nuclear mediado por la interacción con otra proteína argonauta (NRDE-3). Esta transporta los siARNs secundarios al núcleo, donde interactúa con otros factores produciendo el silenciamiento génico por remodelación de la cromatina (Figuras 8 y 10). Entre los factores remodeladores que intervienen en el silenciamiento mediado por ARN pequeños están HPL-1 (un homólogo de la proteína de heterocromatina HP-1), y enzimas modificadoras de la cromatina como la metil transferasa PRMT-5, ZFP-1 (una proteína con dominio dedos de zinc) y proteínas del complejo Polycomb. Estas modificaciones epigenéticas mantienen la señal silenciadora por varias generaciones luego de la señal inductiva (Grishok 2005; Fischer 2010).

3.8 Biogénesis de piRNAs

En paralelo a los ARN pequeños de 21-23 nt se descubrieron otros ARN mayores de entre 26-30 nt expresados fundamentalmente en las células de la línea germinal de la mayoría de los metazoarios. Estos ARNs fueron denominados piARNs, pues interactúan con proteínas Piwi y parecen ser fundamentales para silenciar la expresión de elementos transponibles. En la línea germinal la función de los piARNs unidos a las Piwi (perteneciente a la familia de las Argonautas) previene la acumulación de cambios en el genoma en las siguientes generaciones.

En *Drosophila melanogaster* hay 3 proteínas Piwi: Piwi, Aubergine (Aub) y Ago-3. Este tipo de proteínas parecen ser esenciales para la fertilidad por pruebas hechas en mutantes de diferentes especies. Los piARNs derivan de clusters de elementos repetidos de la heterocromatina telomérica y pericentromérica, y son transcritos por la ARNpol II y transportados al citoplasma. Los precursores de los piARNs son procesados por unas proteínas de la membrana mitocondrial llamadas Zucchini (Zuc) junto a Gasz, que a su vez

reclutan a la DEAD-box helicasa Armi y a Piwi/Aub para la formación del complejo piARN: Piwi/Aub. Adicionalmente existe una vía de amplificación secundaria llamada ciclo ping-pong, y que se basa en la preferencia de Piwi/Aub por piARNs con un U en el extremo 5' y que la secuencia sea antisentido a la secuencia del transposon. Estos piARNs son capaces de unirse por complementariedad a los transcriptos derivados de transposones y Piwi/Aub clivandolos en la posición 10 u 11 del piARN, generando un piARN que tiene el mismo sentido que el transposon y es incorporado en la Ago-3 que a su vez se va a unir al precursor de los piRNAs para generar nuevos piRNAs antisentido.

El mecanismo descrito en *Drosophila* es similar al de mamíferos pero distinto al descrito en *C. elegans*, donde este tipo ARN pequeño es llamado 21U por tener un largo de 21 nt. y preferentemente un U en el extremo 5'. Los piARN maduros se unen a PRG-1 (Piwi Related Gene) una proteína de la familia Argonauta que si bien son esenciales para la estabilidad de los piARN maduros no parecen participar en su biogénesis. Igual que en *D. melanogaster* los genes de los piARN son transcriptos por la ARN pol II pero los transcriptos primarios tienen un largo menor, de 28 a 29 nt. La transcripción de los genes de piARN está regulada por un motivo llamado Ruby al que se unen un grupo de reguladores de la transcripción que en conjunto se denominan proteínas Forkhead. Si bien se han descrito otros factores reguladores e intermediarios, aun no se conocen las proteínas involucradas en la maduración de los precursores. Si bien en *C. elegans* existe una vía de amplificación de piARN esta es diferente a la descrita en *D. melanogaster* y mamíferos e involucra un proceso similar al descrito para la biogénesis de siARNs secundarios (Weick and Miska 2014).

4 Biogénesis de ARNs pequeños en platelmintos

Los estudios de silenciamiento por RNAi en planarias, *S. mansoni* y otros trematodos demostraron que algunas de estas vías están presentes en platelmintos, pero se conoce poco de los genes responsables de las mismas. Por otro lado diversos estudios han demostrado la presencia de miRNA como reguladores tanto en platelmintos de vida libre como *Schmidtea mediterranea* (Palakodeti, Smielewska, and Graveley 2006; Lu et al. 2009), como en las distintas clases parasitas: monogénea como *Gyrodactilus salar* (Fromm et al. 2013), en los cestodos del género *Echinococcus spp.* (Cucher et al. 2011; Bai et al. 2014; Cucher et al. 2015), *Hymenolepis microstoma* (Jin et al. 2013) y *Taenia saginata* (Ai et al. 2012), y en los trematodos incluyendo las dos especies principales de *Schistosoma spp.* (Simões et al. 2011; de Souza Gomes et al. 2011; Cai et al. 2013; Sun et al. 2014), *Clonorchis sinensis* (Xu et al. 2010), *Eurytrema pancreaticum* (Xu et al. 2013) y en el estadio adulto de *F. hepatica* (Xu et al. 2012; Fromm et al. 2015).

Si bien los ARNs pequeños han sido descritos en varios gusanos planos, las vías de síntesis de ARNs pequeños son aún poco conocidas en estos organismos. Se han descrito apenas algunos ortólogos de proteínas claves como Dicer, Drosha, TSN y exportina en *S. mansoni* (Gao et al. 2014; Krautz-peterson and Skelly 2008; Gomes et al. 2009), SID-1 en *S. mansoni* y *japonicum* (Krautz-Peterson et al. 2010), y proteínas de la familia Argonauta en *S. mansoni*, *S. japonicum*, *S. haematobium*, *E. granulosus*, *E. multilocularis*, *H. microstoma*, *T. solium* y *S. mediterranea* (Gomes et al. 2009; Palakodeti et al. 2008; Zheng 2012; Skinner et al. 2014; Luo et al. 2010). En estos reportes se ha detectado, a su vez, la ausencia de proteínas similares a Piwi tanto en cestodos como en trematodos, lo que se correlaciona con la ausencia de piARNs en estos parásitos (Fontenla et al. 2015). Esto plantea la interrogante de cómo se mantiene la estabilidad genómica y si hay alguna otra clase de ARN pequeño que substituya la función de los piARNs en estos organismos (Skinner et al. 2014).

Basándose en estos antecedentes nos planteamos como objetivos analizar los miRNA presentes en las formas juveniles de *F. hepatica*. Adicionalmente, puesto que existía información transcripcional de calidad de varios platelmintos (y posteriormente información genómica) nos planteamos buscar por homología todos los genes asociados a vías mediadas por ARN pequeños en platelmintos.

HIPÓTESIS

Las vías de síntesis de ARN pequeños deben estar conservadas en platelmintos y estos a su vez deben cumplir un papel esencial en la regulación de la expresión génica.

OBJETIVOS

A partir de la hipótesis de trabajo nos planteamos dos objetivos:

- 1) Conocer las vías de síntesis de ARN pequeños presentes en nuestro modelo, y comparar con las existentes en todos los platelmintos, tanto en los de vida libre como en ecto y endoparásitos.
- 2) Identificar los miARN expresados en el estadio juvenil de *F. hepatica* y compararlos con los descritos en el estadio adulto.

MATERIALES Y MÉTODOS

1 Análisis de las Vías de ARN pequeño en platelmintos

1.1 Identificación de proteínas participantes de las vías de ARNi y generación de repositorios

Se hizo una búsqueda bibliográfica y se generó una lista de 62 genes identificados (Tabla 1) por participar en las vías de biosíntesis de ARNi en *C. elegans*. Las secuencias fueron descargadas de <http://www.wormbase.org/> versión WS244. A su vez se generó un repositorio local con los proteomas y genomas de 8 especies de trematodos, 12 especies de cestodos, 2 especies de Monogenea y 1 especie de Turbellaria. En este trabajo reportaremos los resultados obtenidos en algunas especies (Tabla 2). La Figura 11 muestra la estrategia seguida para cumplir con el objetivo 1.

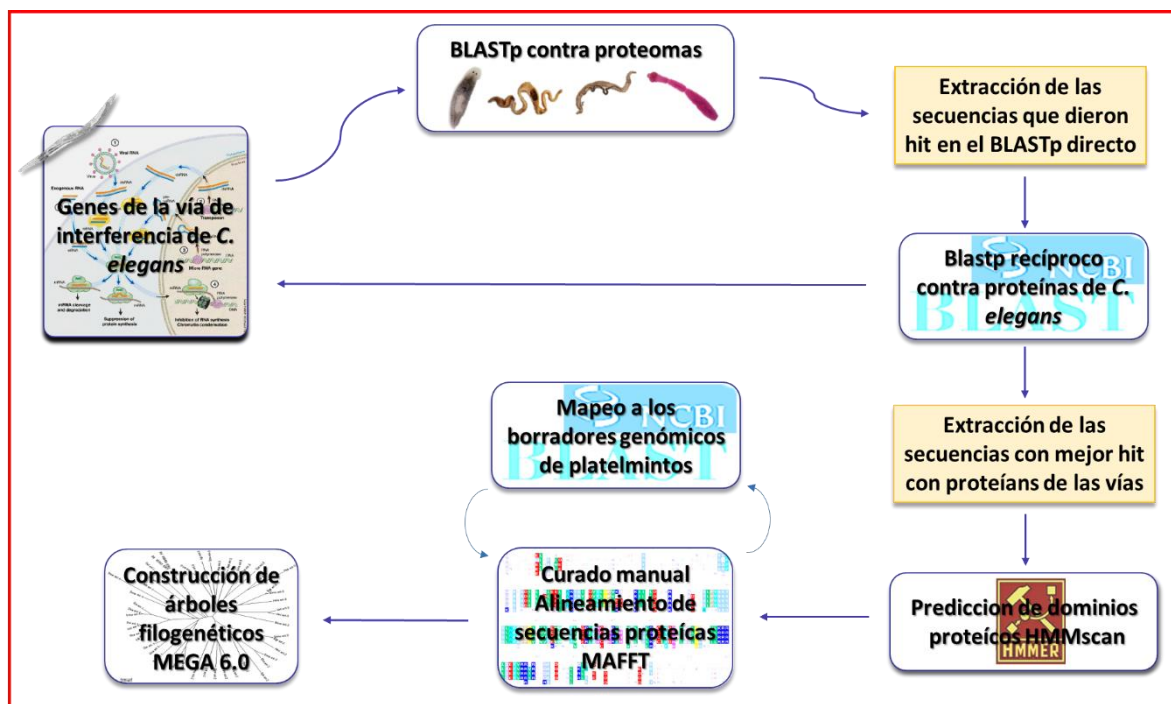


Figura 11. Estrategia seguida para cumplir con el objetivo 1.

Tabla 1 Genes de la vías de biosíntesis de ARN pequeños de *C. elegans*.

En cada caso se indican el nombre del gen, las vías y/o complejos de que forma parte, el nombre de la proteína, función y referencia bibliográfica.

GENE	PATHWAY#	COMPLEX*	PROTEIN NAME	FUNCTION	REFERENCE
AIN-1	MIR	risc	Alg-1 INteracting protein-1/GW182	Deadenilation and decapping	(Kuzuoglu-Öztürk et al. 2012)
AIN-2	MIR	risc	Alg-1 INteracting protein-2/GW182		(L. Zhang et al. 2007)
ALG-1	MIR/ChR	risc	Argonaute Like Gene-1	miRISC Argonaute	(Grishok et al. 2001)
ALG-2	MIR	risc	Argonaute Like Gene-2	miRISC Argonaute	(Grishok et al. 2001)
CGH-1	MIR	risc	Conserved Germline Helicase	DEAD box-helicase	(Hammell et al. 2009)
CSR-1	EndR/ExoR	secS	Chromosome Segregation and Rnai deficient	Chromosome segregation and endo-RNAi	(Claycomb et al. 2009)
DCR-1	ALL	dicer	Dicer-1	RNase III	(Duchaine et al. 2006)
DRH-1	EndR/ExoR/ViR	dicer	Dicer Related Helicase-1	DEAD box-helicase	(Tabara et al. 2002)
DRH-2	EndR/ExoR/ViR	dicer	Dicer Related Helicase-2	DEAD box-helicase	(Duchaine et al. 2006)
DRH-3	EndR/ViR	rdrp	Dicer Related Helicase-3	Helicase of ERI complex	(Gu et al. 2009)
DRSH-1	MIR	mpc	Drosha-1	Nuclear RNase III	(Newman and Hammond 2010)
EGO-1	ExoR/MSUD	rdrp	Enhancer of Glp-One	RdRP in germline	(Maida and Masutomi 2011)
ERGO-1	EndR/ViR	eco	Endogenous Rnai deficient arGONaute	ERI complex argonaute	(Jamalkandi and Masoudi-Nejad 2011)
ERI-1	EndR	eco	Enhanced Rnai-1	Exonuclease	(Pavelec et al. 2009)
ERI-3	EndR	eco	Enhanced Rnai-3	siRNase	(Pavelec et al. 2009)
ERI-5	EndR	eco	Enhanced Rnai-5	Tudor-domain containing protein	(Pavelec et al. 2009)
ERI-6	Nreg	eco	Enhanced Rnai-6	Helicase	(Fischer et al. 2008)
ERI-7	Nreg	eco	Enhanced Rnai-7	Helicase	(Fischer et al. 2008)
F58G1.1	GL&SR	2ryago	Worm ArGONaute 4	Secondary Argonaute	(Hutvagner and Simard 2008)
GFL-1	ChR		human GAS41 Like	Chromatin remodeling	(Kim et al. 2005)
GLD-1	ExoR/MIR	risc	defective in Germ Line Development	RNA binding protein	(Fischer 2010)
HEN-1	ChR		HEsitation behavior	RNA methylation	(Montgomery et al. 2012)
HPL-1	ChR		Heterochromatin Protein Like -1	HP1 homolog	(Jamalkandi and Masoudi-Nejad 2011)
MES-2	ChR		Maternal Effect Sterile-2	Polycomb group protein	(Gaydos et al. 2012)
MES-3	ChR		Maternal Effect Sterile-3	Polycomb group protein	(Gaydos et al. 2012)
MES-6	ChR		Maternal Effect Sterile-6	Polycomb group protein	(Gaydos et al. 2012)
MRG-1	ChR		Mortality factor Related Gene	Chromodomain-containig protein	(Fischer 2010)
MUT-14	ExoR/TsG	secS	MUTator-14	DEAD box-helicase(part of RISC complex)	(C. Zhang et al. 2011)
MUT-16	EndR/ExoR	secS	MUTator-16	Transposon silencing	(C. Zhang et al. 2011)
MUT-7	EndR/ExoR	dicer	MUTator-7	Helicase and exonuclease	(Tops et al. 2005)
NHL-2	MIR	risc	NHL domain containing	RING (E3 activity), ZINC finger motif (interact with E2 enzymes), NHL domain	(Hammell et al. 2009)
NRDE-3	ChR		NRDE-3	Argonaute that transports siRNA into the nucleus	(Guang et al. 2008)
PASH-1	MIR	mpc	Pasha-1	dsRNA binding protein	(Newman and Hammond 2010)
PIR-1	EndR/ExoR/ViR	rdrp	Phosphatase Interacting with Rna/rnp	dual specificity Phosphatase	(Duchaine et al. 2006)
PPW-1	ExoR/ViR	2ryago	PAZ/PIWI domain containing-1	Secondary argonaute	(Hutvagner and Simard 2008)

PPW-2	ChR	2ryago	PAZ/PIWI domain containing-2	Secondary argonaute	(Hutvagner and Simard 2008)
PRG-1	PW	2ryago	Piwi fruitfly Related Gene-1	Germline maintenance	(Das et al. 2008)
PRG-2	PW	2ryago	Piwi fruitfly Related Gene-2	Germline maintenance	(Das et al. 2008)
PRMT-5	ChR		PRotein arginine Methyl Transferase	Chromatin remodeling	(Bedford and Richard 2005)
RDE-1	ExoR/ViR/ChR	dicer	RNAi Deficient-1	Argonaute	(Tabara et al. 2002)
RDE-2	ChR	secS	RNAi Deficient-2	In complex with MUT-7	(Tops et al. 2005)
RDE-3	ExoR/ViR	secS	RNAi Deficient-3	Polymerase beta nucleotidyl transferase	(C. Zhang et al. 2011)
RDE-4	ExoR/ViR/ChR	dicer	RNAi Deficient-4	dsRNA binding protein	(Tabara et al. 2002)
RHA-1	MSUD		Rna HelicAse	Implicated in cromatin silencing	(Grishok 2005)
RRF-1	ExoR/ViR/ChR	rdrp	Rna Dependent polymerase Family	RdRP in somatic line	(Maida and Masutomi 2011)
RRF-2	GL&SR	rdrp	Rna Dependent polymerase Family		(Maida and Masutomi 2011)
RRF-3	GL&SR	rdrp	Rna Dependent polymerase Family	endoRdRP	(Pavelec et al. 2009)
RSD-2	RUS	sist	Rna Spreading Defective-2	Spreading of silencing signal	(Jamalkandi and Masoudi-Nejad 2011)
RSD-3	RUS	sist	Rna Spreading Defective-3	Spreading of silencing signal	(Jamalkandi and Masoudi-Nejad 2011)
RSD-6	RUS	sist	Rna Spreading Defective-6	Spreading of silencing signal	(Jamalkandi and Masoudi-Nejad 2011)
SAGO-1	ExoR/ViR	2ryago	Synthetic secondary siRNA-deficient ArGOnaute mutant-1	Secondary argonaute	(Hutvagner and Simard 2008)
SAGO-2	ExoR/ViR	2ryago	Synthetic secondary siRNA-deficient ArGOnaute mutant-2	Secondary argonaute	(Hutvagner and Simard 2008)
SID-1	RUS	sist	Systemic rna Interference Defective-1	Transmembrane protein dsRNA uptake	(van Roessel and Brand 2004)
SID-2	RUS	sist	Systemic rna Interference Defective-2	Transmembrane protein dsRNA uptake	(van Roessel and Brand 2004)
SMG-2			Suppresor with Morphological effect on Genitalia-2	Initiates and maintains secondary RNA	(Mango 2001)
SMG-5			Suppresor with Morphological effect on Genitalia-5	Initiates and maintains secondary RNA	(Mango 2001)
SMG-6			Suppresor with Morphological effect on Genitalia-6	Initiates and maintains secondary RNA	(Mango 2001)
TSN-1	MIR/ExoR	risc	Tudor Similar Nuclease	Homologue to micrococcal nuclease	(Caudy et al. 2003)
VIG-1	MIR/ExoR	risc	Vasa Intronic Gene-1	dsRNA binding protein	(Caudy et al. 2002)
XPO-1	MIR		eXPortin	nuclear exporter protein	(Büssing et al. 2010)
XRN-1	MIR/ExoR	risc	XRN ribonuclease related	Exonuclease	(Orban and Izaurralde 2005)
ZFP-1	ChR		Zinc Finger Protein	Chromatin remodeling	(Kim et al. 2005)

PATHWAY DEFINITIONS

EndR=Endogenous RNAi
 ExoR=Exogenous RNAi
 ViR=Viral induced RNAi
 GL&SR=Germline & Somatic exoRNAi
 Nreg=Negative regulator of RNAi
 MIR=miRNA
 PW=Piwi
 ChR=Chromatin Remodeling
 TSG=Transposon Silencing in Germline
 RUS=RNA uptake & spreading
 MSUD=Meiotic Silencing of Unsynapsed chromosomes

*COMPLEX DEFINITIONS

risc=RNA Induced Silencing Complex
 sist=Systemic RNA interference
 dicer=Dicer Complex
 mpc=Microprocessor Complex
 eco=ERI Complex
 2ryago= Secondary Argonautes
 rdrp=RNA dependent RNA polymerases
 secS=Secondary RNAi Silencing

Tabla 2. Platelmintos analizados.

Entre paréntesis se indica la abreviatura utilizada para cada especie. Los proteomas fueron descargados de <http://parasite.wormbase.org/>, salvo *O.viverrini* (descargado de http://bioinfosecond.vet.unimelb.edu.au/Opisthorchis_viverrini/OV_PRJNA222628_proteins.fa.gz)

Trematodos	Cestodos	Turbelaria
<i>Clonorchis sinensis</i> (Csi)	<i>Echinococcus granulosus</i> (Egr)	<i>Schmidtea mediterranea</i> (Sme)
<i>Opisthorchis viverrini</i> (Ovi)	<i>Echinococcus multilocularis</i> (Emu)	
<i>Fasciola hepatica</i> PRJNA179522 (FhA)	<i>Hymenolepis microstoma</i> (Hmi)	
<i>Fasciola hepatica</i> PRJEB6687 (FhB)	<i>Mesocestoides corti</i> (Mco)	
<i>Schistosoma haematobium</i> (Sha)	<i>Taenia solium</i> (Tso)	
<i>Schistosoma japonicum</i> (Sja)		
<i>Schistosoma mansoni</i> (Sma)		

1.2 Búsqueda de homólogos por BLAST directo y recíproco

Las secuencias de *C. elegans* se usaron para interrogar los proteomas de platelmintos por BLASTp utilizando como cut-off un e-value de E^{-05} dejando los demás parámetros por defecto. A su vez, los hits generados en el BLASTp directo fueron utilizados como query en el BLASTp recíproco para interrogar el proteoma de *C. elegans* usando idénticos parámetros a los usados para realizar el BLASTp directo, excepto que en este caso para cada set de homólogos putativos solo se retuvieron aquellas secuencias cuyo mejor hit fue la secuencia de *C. elegans* usada en el BLAST directo (algunos de los scripts utilizados están adjuntos en la sección Apéndice).

1.3 Predicción de dominios proteicos

Para confirmar que las secuencias identificadas en platelmintos eran homólogos funcionales de las secuencias de *C. elegans*, se realizó la predicción de dominios conservados de las secuencias de platelmintos con HMMScan (Johnson, Eddy, and Portugaly 2010) con un cutoff de $1 E^{-03}$. Se espera encontrar un perfil de dominios similar al de la proteína query de *C. elegans*.

1.4 Identificación de homólogos usando como query proteínas de platelmintos

En aquellos casos en que no se hallaron homólogos en alguna de las especies de platelmintos pero si en otras se realizó una segunda búsqueda por BLASTp usando como query algunos de estos hits primarios. En los casos en los que hubo secuencias de distintas especies disponibles para usar como query se compararon los tamaños y dominios proteicos de cada secuencia, seleccionándose las más completas como nuevos blancos de búsqueda. Con ellas se repitió la búsqueda por BLASTp utilizando los mismos parámetros a los descriptos anteriormente, y se realizó la predicción de dominios conservados con HMMScan de las secuencias obtenidas que se cotejaron con los descriptos en *C. elegans*.

1.5 Confirmación por BLAST con los borradores genómicos.

A partir de la disponibilidad más reciente de los borradores genómicos de varias especies de platelmintos en el sitio Wormbase-Parasite (parasite.wormbase.org), se buscó confirmar los resultados realizando las búsquedas por BLAST contra los genomas disponibles. Esto permitió confirmar varias de las asignaciones realizadas y mostro que algunas proteínas que aparecían como truncas o que resultaban en múltiples hits correspondían muchas veces a hits simples sobre regiones con anotación incorrecta (con exones faltantes o donde se predicen dos o más genes adyacentes cuando en realidad corresponden a un mismo transcripto), en estos casos se realizó la reanotación de dichos genes.

1.6 Alineamiento de secuencias, curado manual y construcción de árboles filogenéticos

Se realizó el alineamiento de las secuencias identificadas en platelmintos junto con la de *C. elegans* y las de otros ortólogos confirmados en otros organismos con MAFFT (Katoh and Standley 2013)(se probaron otras herramientas de alineamiento como Clustal con resultados iguales o inferiores), usando los parametros --localpair que utiliza el algoritmo de

alineamiento local Smith-Waterman y la opción --maxiterate 16, que realiza 16 ciclos de refinamiento post-alineamiento, --clustalout que genera una salida en formato clustal (.aln) y la opción --inputorder mantiene el orden de entrada de las secuencias.

Debido a que muchos de los proteomas usados son primeras versiones de genomas con anotaciones aun no curadas se usó BioEdit versión 7.2.5 (Hall 1999) para curar manualmente los alineamientos generados removiendo y/o ensamblando secuencias parciales.

Con los alineamientos obtenidos se construyeron arboles filogenéticos con MEGA 6.0 (Tamura et al. 2013) utilizando el método de Neighbor-Joining y un bootstrap de 500. Se probaron otros métodos y herramientas, como PhyML (Guindon et al. 2010), con resultados similares. Se optó por el finalmente utilizado por el menor costo computacional. El porcentaje de replicas en que los taxones asociados clusterizaron juntos en el test de bootstrap se muestran junto a las ramas.

2 Identificación de miARN en el estadio juvenil

2.1 Obtención de muestras biológicas

5000 metacercarias de *F. hepatica* fueron desenquistadas *in vitro* siguiendo protocolos estándar. Los parásitos activos emergentes fueron colectados y guardados a -80°C hasta su uso. Posteriormente se realizó la extracción de ARN con Rneasy Mini Kit (QIAGEN). El ARN fue transformado a cADN simple hebra y amplificado en 15 ciclos de PCR. El producto de amplificación fue secuenciado con tecnología Illumina.

2.2 Procesamiento de los datos y filtrado

Luego de enmascarar los adaptadores y remover los reads menores a 18 nt los reads limpios fueron sometidos a análisis de calidad con Fastx Tools (http://hannonlab.cshl.edu/fastx_toolkit/). Los reads con un 90% de bases con un Phred score

mayor a 20 fueron conservadas. Los reads redundantes fueron combinados y cuantificados. Los reads de alta calidad fueron filtrados contra datos transcriptómicos (mARNs) de platelmintos disponibles en GenBank, SchistoDB (Zerlotini et al. 2013) y HelmDB (Mangiola et al. 2013), descartando todos los reads que resultaron en matches con más del 85% de homología.

Los datos fueron posteriormente escrutados con los mismos criterios de selectividad (mismatches ≤ 3) contra otras bases de datos de ARNs no-codificantes como Rfam (Burge et al. 2013) y fRNAdb (Mituyama et al. 2009) (en ambos casos se excluyeron los miARNs de estas bases de datos) y aquellos reportados en *S. mansoni* y *S. japonicum* (Copeland et al. 2009) para identificar y remover ARNs no codificantes incluyendo rARN, tARN, snARN, snoARN y otros ncARNs.

2.3 Análisis de los datos y clasificación de los miARNs

El set de reads procesados fue usado para realizar búsquedas por homología contra la miRBase Sanger versión 20 (Kozomara and Griffiths-Jones 2014), los miARNs anotados en Rfam (Burge et al. 2013) y en ncRNA.org (Mituyama et al. 2009). Los reads también fueron comparados contra una bases de datos local generada a partir de miARNs reportados en gusanos planos que incluye miARNs de *S. japonicum* (http://omics.biosino.org:14000/kweb/sj_miRNA/) y aquellos publicados en *S. japonicum* (Huang et al. 2009), *S. mansoni* (Simões et al. 2011), *Clonorchis sinensis* (Xu et al. 2010), *F. hepatica* y *F. gigantica* (Xu et al. 2012), y el monogenea *Gyrodactilus salaris* (Bastian Fromm et al. 2013). Los reads con alta similaridad (>85%) con miARNs conocidos de otros organismos fueron clasificados dentro de la misma familia de miARNs. Los miARNs candidatos con menos de 10 reads o largo menor a 20 o mayor a 23 nucleótidos no fueron considerados. Esta búsqueda fue realizada usando miRexpress (Wang et al. 2009) y el módulo miRNA del CLC Genomics Workbench v6.5 (CLC Inc, Aarhus, Denmark) con

similares resultados en ambos procedimientos. Consecuentemente continuamos el análisis usando el dataset generado con el CLC Genomics Workbench.

2.4 Distribución filogenética y conservación de las familias de miARNs en el estadio juvenil de *F. hepatica*

La distribución filogenética de los miARNs candidatos fue evaluada comparando las entradas con anotaciones similares en miRBase (Kozomara and Griffiths-Jones 2014) y/o mirOrtho (Gerlach et al. 2009) e identificándolas como presentes o ausentes en cada taxón. Las secuencias de los diferentes precursores y miARNs maduros fueron descargados de miRBase y alineados con los miARNs correspondientes de los gusanos planos usando MAFFT (Kato and Standley 2013). Los alineamientos fueron visualizados con CLC Genomics Workbench, generando los logos consensos. La conservación fue cuantificada por la cuantificación de las sustituciones usando la interpretación más parsimoniosa siguiendo el procedimiento descrito por Wheeler et al. (Wheeler et al. 2009). En resumen, si una sustitución es compartida entre especies relacionadas o taxones hermanos (por ej. compartida entre cestodos) fueron contados como un solo cambio; si más de dos residuos diferentes estaban presentes en una posición dentro de un taxon estos fueron contados como cambios separados. Los cambios por posición fueron promediados para deuterostomados, ecdysozoas y platelmintos.

2.5 Ensamblaje de novo del genoma de *F. hepatica*

Se descargaron los reads de secuenciación genómica de *F. hepatica* depositados en el repositorio del European Bioinformatics Institute (<ftp.sra.ebi.ac.uk/vol1/fastq/ERR065/ERR065038>), los que fueron trimmeados, filtrados por calidad y ensamblados *de novo* con la plataforma CLC Genomic Workbench. El resultado del ensamblaje fue un genoma de 1.05 GB compuesto de 475283 contigs con un N25 de 2629 pb, N50 de 7125 pb y un N75 de 15174 pb. Se predijeron 39386 genes con el predictor

AUGUSTUS (Stanke and Morgenstern 2005), utilizando los genes de *S. mansoni* como modelos génicos de referencia.

Se utilizó tRNA-scanSE (Lowe and Eddy 1997) para la predicción de 4641 tARNs y RNAmmer (Lagesen et al. 2007) para la predicción de 54 rARNs.

2.6 Identificación de miARNs novedales

Para identificar miARNs novedales específicos de *F. hepatica*, los reads que no matchearon con ninguna secuencia conocida en los pasos previos y con un largo de entre 20-23 nt fueron mapeados al genoma usando miRDeep2 (Friedländer et al. 2012). Se utilizó un score de corte miRDeep2 estricto en la predicción de novedales. Solo los miARNs predichos con un score igual o mayor a 10 fueron considerados como novedales. Adicionalmente las predicciones fueron inspeccionadas manualmente para descartar posibles falsos positivos. Se siguió un procedimiento similar para confirmar los miARNs identificados por homología, este caso se utilizó un score de corte más permisivo (igual o mayor a 1), las predicciones resultantes también fueron curadas manualmente y se descartaron los falsos positivos.

2.7 Identificación de secuencias derivadas de tARNs

El análisis primario de los reads de entre 25-35 nt de largo mostro homología con tARNs. Para confirmar esto mapeamos la fracción de 25-35 nt al ensamblaje local del genoma de *F. hepatica*. Los reads que mapearon al genoma fueron mapeados a los tARNs predichos con tRNAscan-SE (Lowe and Eddy 1997) usando BLAST con un cut-off de E^{-10} . Un total de 99,014 reads mapearon con tARNs. Las secuencias derivadas de tARNs fueron mapeados a la estructura predicha por tRNAscan-SE para determinar su origen.

RESULTADOS

Capítulo 1. Análisis de vías mediadas por ARN pequeños

Las vías de silenciamiento estudiadas principalmente en *C. elegans* permitieron conocer con mayor detalle el papel de los ARN pequeños reguladores en el control post transcripcional de la expresión génica.

El conocimiento de las vías de regulación de la expresión por ARN pequeños es importante para entender como se regula la expresión y puede facilitar el desarrollo de ARNi.

Buscamos identificar las genes que participan en las vías de síntesis de ARN pequeños presentes en nuestro organismo modelo *F. hepatica* y compararlo con otros platelmintos.

Para ello se realizaron búsquedas por homología por BLAST, utilizando como query todas las proteínas participantes de las vías de generación de ARN pequeños en *C. elegans*. Se realizaron búsquedas en todos los genomas y transcriptomas de platelmintos disponibles públicamente, aunque en este trabajo solo se reportan los resultados obtenidos en 13 de ellos (ver Tabla 2).

Las secuencias de platelmintos identificadas por BLAST directo fueron usadas para interrogar el proteoma de *C. elegans* por BLAST recíproco, de modo de confirmar la homología entre pares de secuencias. Adicionalmente se predijeron los dominios proteicos funcionales de las secuencias identificadas por BLAST, verificándose que estos coincidieran con los presentes en el ortólogo de *C. elegans*.

Varias de las secuencias identificadas por este mecanismo mostraron posibles errores de anotación (secuencias parciales, varios transcriptos predichos, etc) por lo que se hizo una curación manual de las secuencias mal anotadas. Si bien se estudiaron todas las vías de generación de ARN pequeños de platelmintos, aquí reportaré solo los ortólogos identificados en la vía de generación de miARNs.

1 Vía de generación de miARNs

La vía de generación de miARNs resulta de una serie de pasos bioquímicos que transforman un transcrito primario llamado pri-miARN en un producto maduro, biológicamente activo, el miARN. Identificamos varias secuencias con buena homología a proteínas que participan en distintos pasos en la vía de regulación post-transcripcional mediada por miARNs (Figura 12, tabla 3). En todos los platelmintos la pérdida o ganancia de genes de la vía de generación de miARNs tienen características similares lo que confirma el origen común de estas especies.

Se identificaron los genes que forman el complejo microprocesador (Drosha y Pasha), que es el complejo que procesa los pri-miARNs a pre-miARNs, que luego son exportados al citoplasma (Newman and Hammond 2010). Identificamos los genes ortólogos de Xpo-1 la proteína que media el transporte al citoplasma de los pre-miARNs en *C. elegans* (Büssing et al. 2010). Se identificaron también genes ortólogos de la enzima RNasaIII Dicer, la proteína central de las vías de biogénesis de ARN pequeños que procesan los ARNdc para que puedan ser incorporados en el complejo RISC (Hutvagner and Simard 2008). Identificamos a su vez, un segundo grupo de enzimas RNasa III similares a Dicer, conservadas en todos los platelmintos al que llamamos Dicer-like y que describimos más adelante (Tabla 3) (Tabla Suplementaria 1, Apéndice).

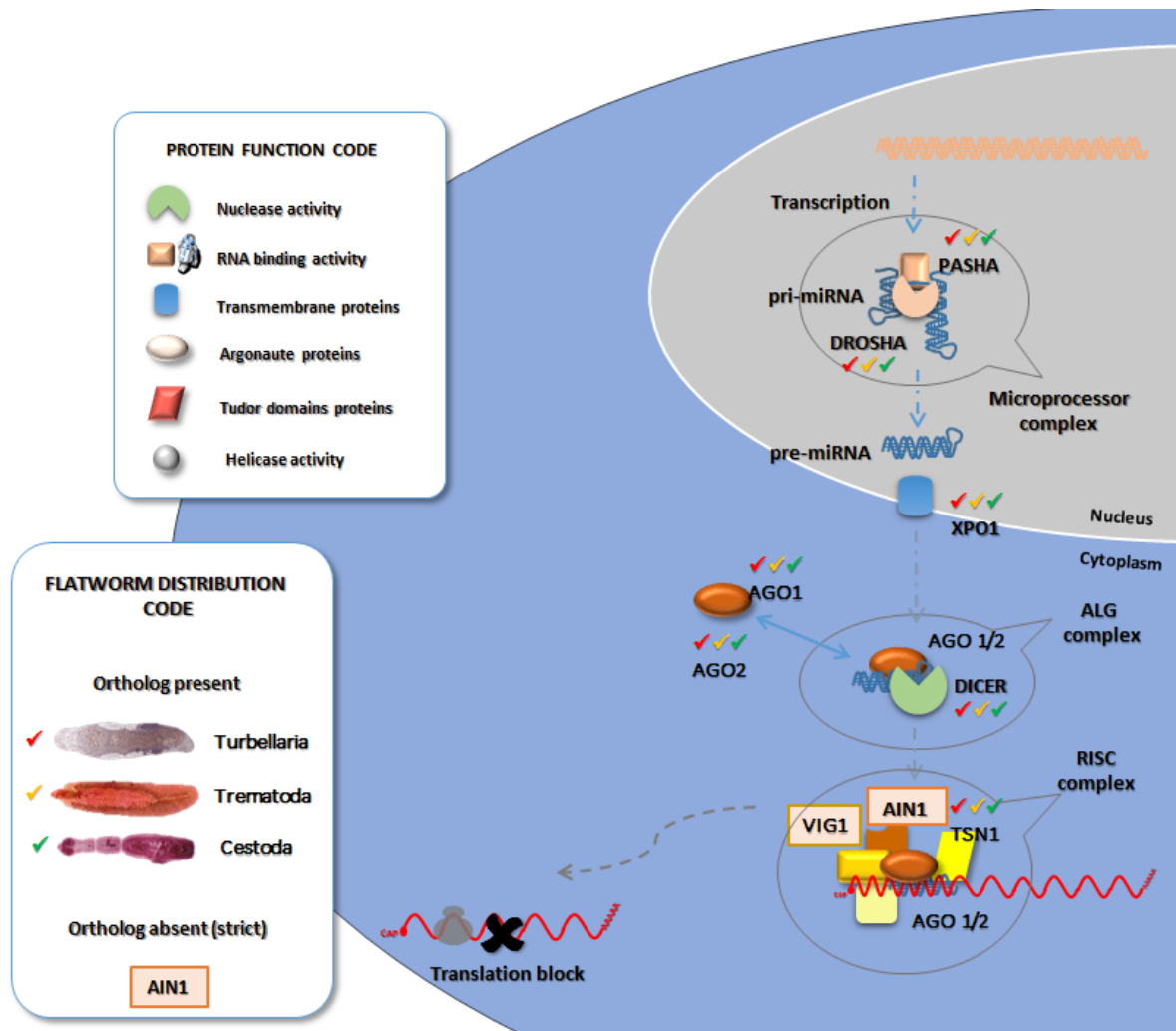


Figura 12. Esquema de la vía de miARN.

Los distintos tipos funcionales de proteínas presentes en la vía están representados con distintas figuras de acuerdo a la clave indicada a la izquierda. Se indica con tildes de color la presencia de al menos un ortólogo en las distintas clases de platelmintos. La ausencia de un ortólogo se indica con un recuadro amarillo.

También en las proteínas de tipo Argonauta hallamos además de las canónicas Ago 1/2, un segundo grupo de enzimas que parecen ser una innovación adquirida entre los platelmintos, el que es descrito en detalle más adelante (Tabla 3) (Tabla Suplementaria 1, Apéndice). Algunos otros componentes del complejo RISC presentaron homólogos conservados, como la exonucleasa TSN-1 (Tudor-Staphylococcal Nuclease), aunque no identificamos ningún gen con buena homología con los factores VIG-1 y AIN-1.

	Endoribonucleasas			Argonautas		Otros factores				
	DRSH-1	DCR-1	Dicer-like	Ago1	FL-AGO	TSN-1	PASH-1	XPO-1		
Trematodos	<i>F. hepatica</i> (PRJNA179522)	FhA 12643 +_12644 +_12645 +_12646 +_12647	FhA 11014 +_11015 +_11016 +_11017	FhA 08656 +_08657 FhA 08654 +_08655	FhA 04557 +_04558 +_04559	FhA 07567 FhA 12482 +_12483	FhA 14412 +_14413	FhA 01879 +_01880	FhA 11553 +_11554	
	<i>F. hepatica</i> (PRJEB6687)	FhB s2953B000132 +s2953B000133	FhB s1520B000104	FhB s1234B000167 FhB s1234B000168	FhB s1192B000304	FhB s5156B000132 FhB s668B000138+s 668B000139	FhB s375B000232	FhB s319B000233	FhB s2835B000195	
	<i>C. sinensis</i>	Csi 101707	Csi 111519	Csi 105018 Csi 105019	Csi 112397	Csi 100735 Csi 108572	Csi 109541	Csi 107612	Csi 107442	
	<i>O. viverrini</i>	Ovi 11305	Ovi 13440	Ovi 11263 Ovi 11264	Ovi 02735	Ovi 07377 Ovi 13773	Ovi 00614	Ovi 13119	Ovi 14337	
	<i>S. haematobium</i>	Sha 05360 +_06081	Sha 06487	Sha 03533	Sha 03584	Sha 05116	Sha 06301	Sha 00951	Sha 02166	
	<i>S. japonicum</i>	Sja 0048900	Sja 0069770	Sja 0043700	Sja 0044720	Sja 0045200 Sja 0103990	Sja 0100020	Sja 0013270	Sja 0069890	
Cestodos	<i>S. mansoni</i>	Sma 142510.1 Sma 142510.2	Sma 169750.1 Sma 169750.2	Sma 033600.1	Sma 198380.1	Sma 179320.1 Sma 102690	Sma 166110 +_081570	Sma 087220	Sma 124820.1 Sma 124820.2	
	<i>E. granulosus</i>	Egr 001182900	Egr 000085200	Egr 000181800	Egr 000346200	Egr 000739100 Egr 000911600 Egr 000911700	Egr 001073700	Egr 000576300	Egr 000331050	
	<i>E. multilocularis</i>	Emu001182900	Emu000085200	#	Emu000346200	Emu000739100 Emu000911600 Emu000911700	Emu001073700	Emu000576300	Emu000331050	
	<i>T. solium</i>	Tso 000908400	Tso 000872800	Tso 001225000 +_000756400 +_000048100	Tso 000674100	Tso 000161400 Tso 000876200 Tso 001204500	Tso 000483600	Tso 000694100	Tso 000192600	
	<i>M. corti</i>	Mco 0000725501	#	#	#	Mco 0000353801 Mco 0000379701 #	&	Mco 0000735001	Mco 0001013401	
	<i>H. microstoma</i>	Hmi 000061600	Hmi 000252400	Hmi 000200200 +_200100	Hmi 000658200	Hmi 000114300 Hmi 000950000	Hmi 000687200	Hmi 000067400	Hmi 000936400 Hmi 000935700	
	Turbelario	<i>S. mediterranea</i>	Sme.000125.07	Sme.001736.00 +_001736.01 +_001736.02 +_001736.03	Sme.004816.00	Sme.008596.00 +_008596.01	Sme.000678.05	Sme.001005.01 +_001005.00	&	Sme.000191.02
						Sme.036615.00	Sme.000678.07			

Tabla 3. Ortólogos de la vía de miARNs detectados en platelmintos.

Para facilitar la comprensión se unificó la nomenclatura renombrando los prefijos de los identificadores originales de las secuencias en wormbase-parasite por un código de tres letras que identifica la especie. Con (+) se indican las anotaciones que corresponden a un mismo transcripto y fueron editadas. Con # se indican los casos donde se detectaron las secuencias homólogas por BLAST en una región donde inicialmente no existen genes anotados. Con & se indican los casos donde el posible ortólogo en platelmintos está fragmentado en varios contigs de pequeño tamaño (Tabla Suplementaria 1, Apéndice).

VIG-1 (**V**asa **I**ntronic **G**ene-1) es una proteína de unión a ARN identificada en ensayos de purificación de proteínas por estar asociada a las proteínas Argonautas de miRISC, sin embargo distintos ensayos muestran que su función no parece ser esencial para el funcionamiento del complejo (Caudy et al. 2002; Caudy et al. 2003; Fischer 2010).

AIN-1/2 (**A**lg-1 **I**Nteracting protein-1) media la repuesta efectora del complejo RISC por la desestabilización de los ARNm blancos y ha sido señalado por ser un factor relevante en el silenciamiento traduccional (L. Zhang et al. 2007).

1.1 El complejo microprocesador

En el primer paso de procesamiento de los precursores de los miARNs se encuentra el complejo microprocesador formado por DRSH-1 (Drosha) y PASH-1 (Pasha) también llamada DGCR8 en *D. melanogaster*. Drosha es una RNasa III de Clase II (Jaronczyk, Carmichael, and Hobman 2005) y posee dos dominios Ribonucleasa III, que clivan los pri-miARNs, y un dominio DSRM de unión a ARNdc (Figura 13A). Pasha es una proteína de unión a ARN doble cadena esencial para completar este proceso (Figura 13B). Como en nematodos, insectos y vertebrados los platelmintos poseen solo un ortólogo de Drosha, y Pasha (Tabla 3).

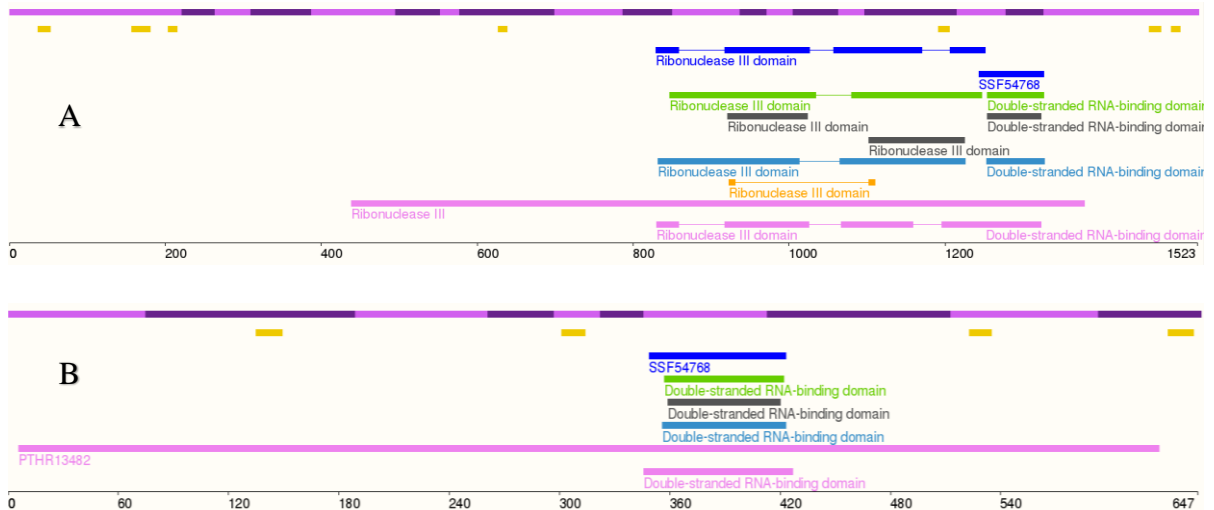


Figura 13. Dominios de las proteínas que forman el complejo Microprocesador de *C. sinensis*.

Drosha (A) presenta dos dominios Ribonucleasa III seguido de un dominio de unión a ARNdc. (B) *Pasha* presenta un dominio de unión a ARNdc.

En la figura 14 se muestra un árbol filogenético generado a partir de los genes ortólogos de *Drosha* depositados en Wormbase Parasite con la herramienta Gene Tree (<http://parasite.wormbase.org>), donde se resaltaron los ortólogos de *Drosha* en Trematodos (recuadro verde) y Cestodos (recuadro amarillo). En el diagrama de la estructura génica ubicado a la derecha se marcó la ubicación aproximada de los dominios Ribonucleasa III (RiboIII) (recuadro rosado) y DSRM (recuadro azul). Con recuadros rojos se marcan tres casos en los que se detectaron errores en las anotaciones de los modelos génicos.

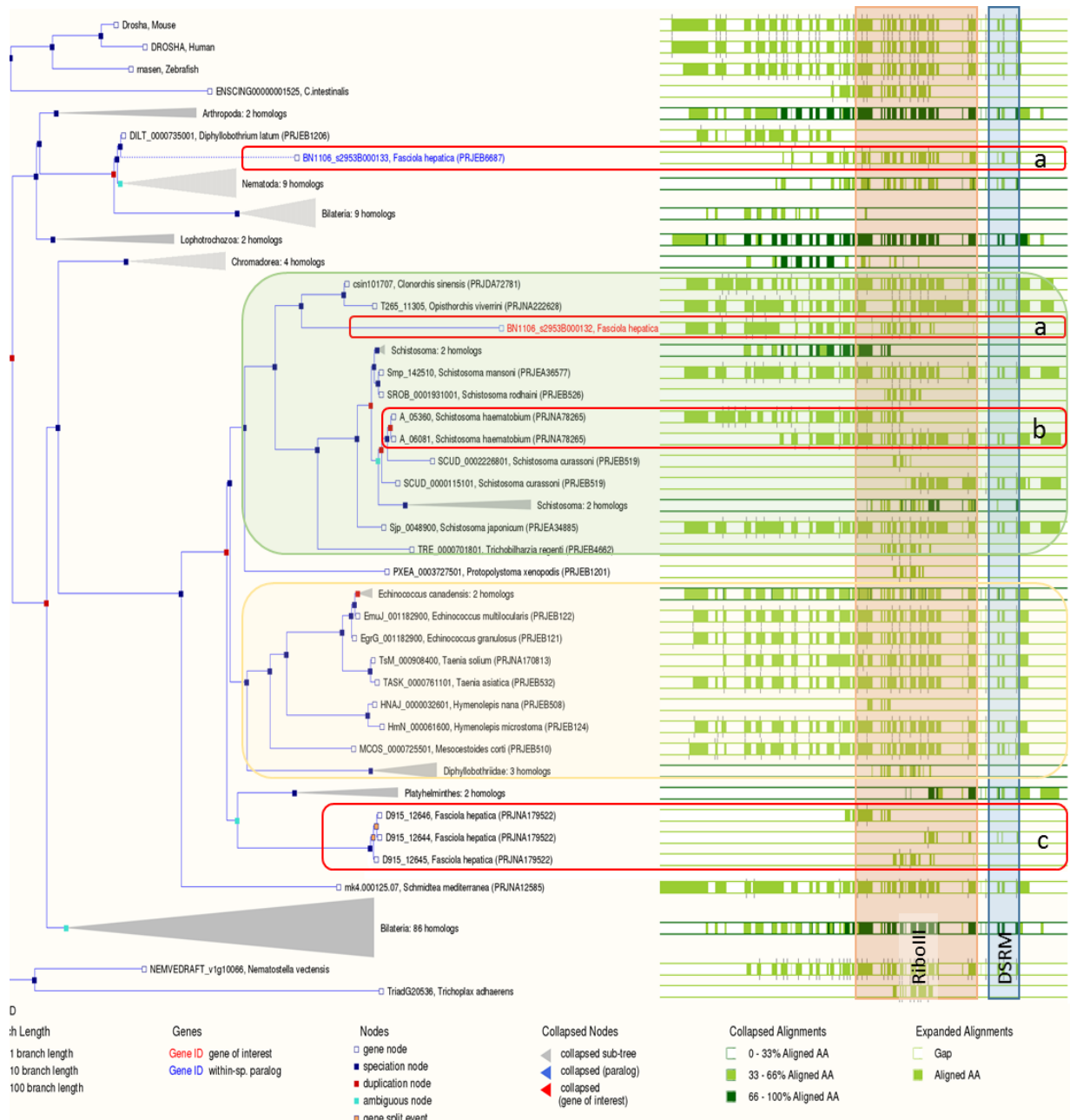


Figura 14. Árbol filogenético generado en Wormbase Parasite.

Se destacan los ortólogos de Drosha en Trematodos (recuadro verde) y Cestodos (recuadro amarillo). En rojo se muestran 3 casos problema en ortólogos de Drosha.

Para verificar si efectivamente había más de un ortólogo en estos organismos o si estos casos correspondían a errores de anotación, se mapearon proteínas Drosha estructuralmente completas en los genomas problema por tBLASTn. Se verificó que en todos los casos problema habían errores en las anotaciones genómicas (Figura 15).

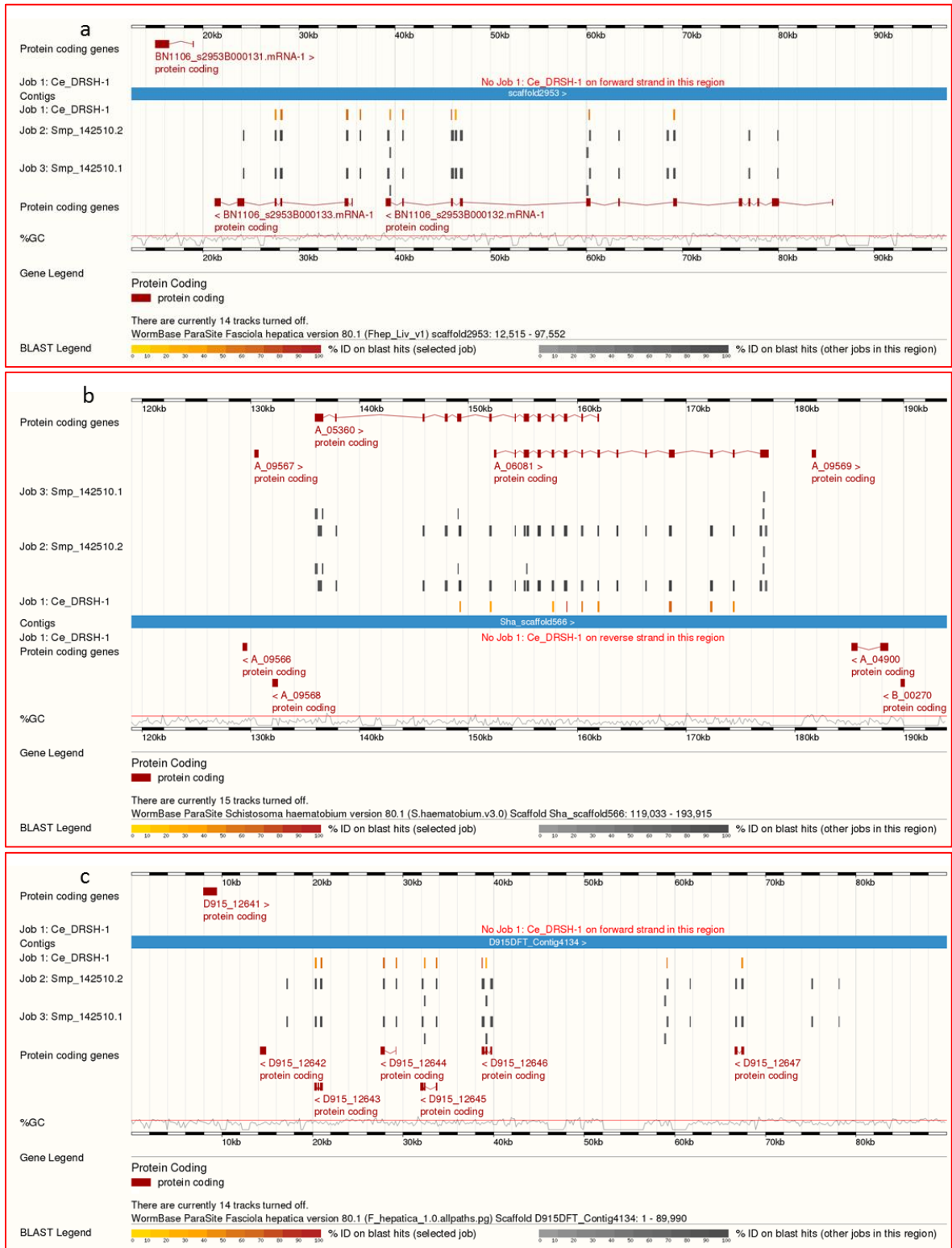


Figura 15. Salida gráfica de los resultados del BLAST realizado en Wormbase Parasite. A la izquierda de las imágenes están los identificadores de las secuencias utilizadas para la búsqueda. Las regiones donde hay homología de secuencias detectada (HSPs) se marcan en amarillo o gris en diferentes tonos según el porcentaje de identidad. En A se observa que dos genes contiguos cubren la anotación completa del gen Drosha. En B dos genes contiguos con exones en común cubren la anotación del gen Drosha. En el ensamblaje genómico representado en C, se observa que el gen está fragmentado en al menos cinco genes anotados diferentes, existiendo además otras regiones (exones) no anotados.

En la Figura 15A y B se observa que dos genes contiguos cubren la anotación completa del gen Droscha a partir de todos los resultados de BLAST. En el ensamblaje genómico representado en C, se observa que el gen está fragmentado en al menos cinco genes anotados diferentes, existiendo además otras regiones (exones) no anotados. Consecuentemente se reanotaron los tres genes en los genomas en base a los resultados obtenidos (Tabla Suplementaria 1, Apéndice). Para testear las secuencias re anotadas se alinearon localmente con los genes ortólogos de otros platelmintos identificados (Tabla 3).

1.2 Búsqueda de homólogos de Dicer

La búsqueda por homología con los genes Dicer de *C. elegans* y *D. melanogaster* dio lugar a múltiples hits en casi todos los platelmintos. Los resultados generales mostraron que al igual que en *D. melanogaster* y a diferencia de *C. elegans* los platelmintos poseen al menos dos parálogos de Dicer (Tabla 3) (Tabla Suplementaria 1, Apéndice). A partir del alineamiento de los posibles ortólogos en las diferentes especies se creó un árbol filogenético por el método de Neighbor-Joining, el que mostró que era posible organizar las secuencias en diferentes subclados.

Un ortólogo de Dicer-1 ya fue descrito en *S. mansoni* (Gomes et al. 2009; Krautz-Peterson and Skelly 2008). En el presente trabajo se detectó una segunda isoforma de este gen, siendo la isoforma Smp_169750.1 aproximadamente 200 aminoácidos más corta que la isoforma 2. Ambas secuencias junto con las de otros platelmintos formar un clado con la proteína Dicer canónica de otras especies (Figura 16). Por otro lado, se observó un segundo clado en platelmintos que no parecen tener un origen común con Dicer-2 de Insectos y que denominamos Dicer-like (Figura 16).

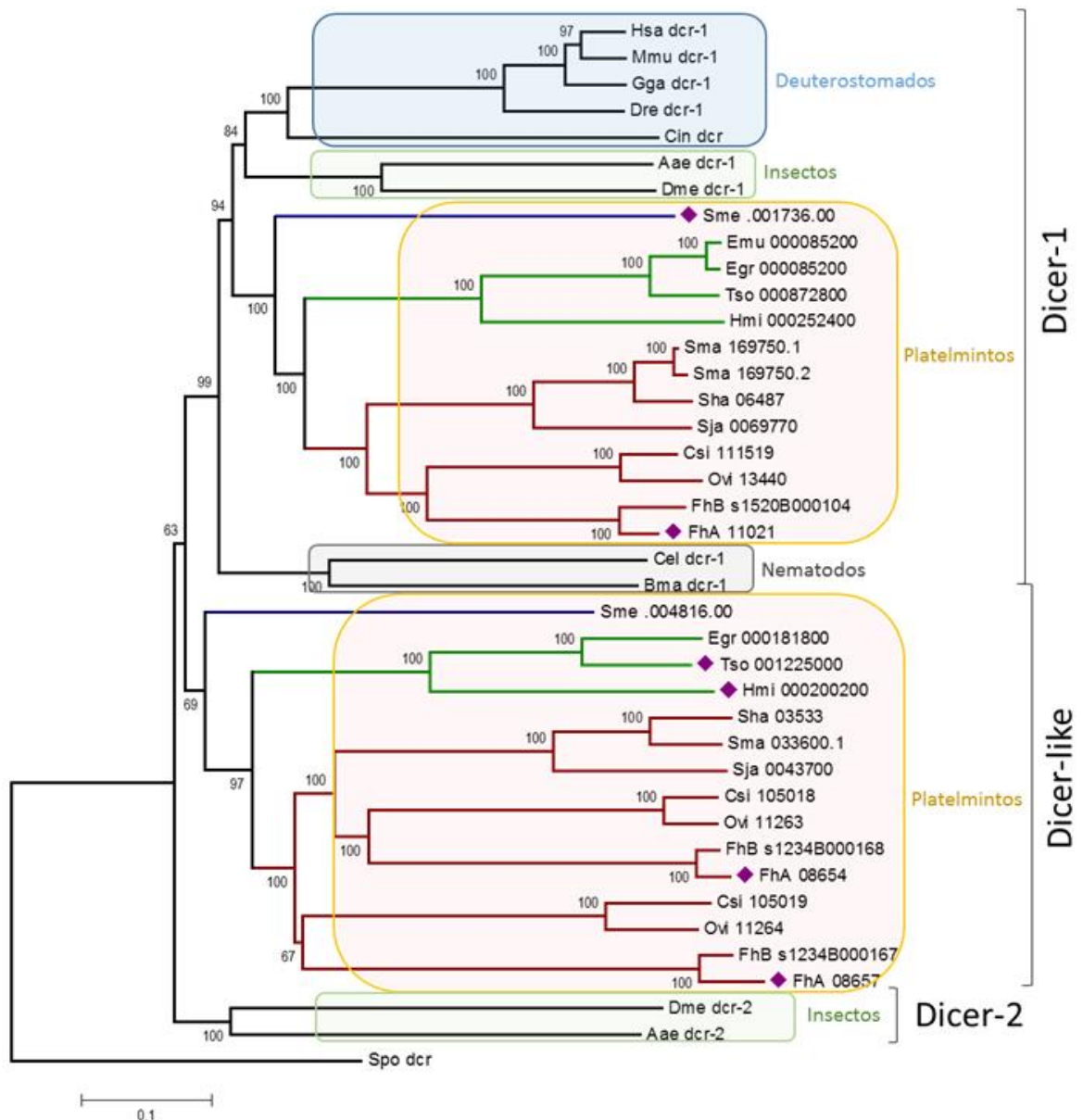


Figura 16. *Árbol filogenético generado por el método de Neighbor-joining a partir del alineamiento de ortólogos de Dicer en diferentes especies. Los porcentajes de Bootstrap se indican en cada rama.*

Se identificaron 3 subclados de los cuales Dicer-like parece ser específico de platelmintos. Recuadrados en azul están los organismos Deuterostomados, en verde los Insectos, en gris los Nematodos y en amarillo los Platelmintos. Con un ◆ se marcan las secuencias reanotadas (ver Tabla 3).

Se observó que en cestodos Dicer-like presenta una estructura similar a la descrita en las Dicer canónicas, con 4 dominios (DEAD/Helicasa C, dimerización de Dicer y dos dominios Ribonucleasa III) (Figura 17 a) (Jaronczyk, Carmichael, and Hobman 2005).

Mientras tanto, en trematodos Dicer-like es estructuralmente distinta y solo presenta los dos dominios Ribonucleasa III (Figura 17 b). Si bien la proteína Dicer-like identificada en el turbelario *S. mediterranea* es parcial, detectamos un dominio DEAD/Helicasa C junto con los dos dominios Ribonucleasa III, por lo que la estructura proteica de la Dicer-like de Turbelarios sería similar a la de los Cestodos.

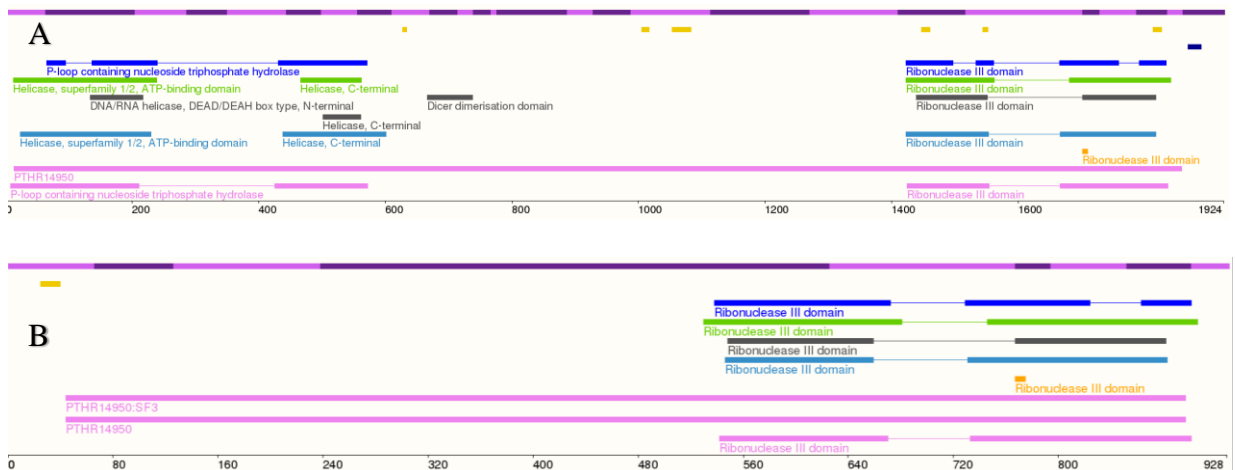


Figura 17. Dominios funcionales de Dicer-like en Cestodos (a) y Trematodos (b).

La estructura de Dicer-like en Cestodos (a) presenta un DEAD-box/Helicasa C, un dominio de dimerización de Dicer y dos dominios Ribonucleasa III, mientras que en Trematodos (b) solo tiene dos dominios Ribonucleasa III.

Se sabe que los Cestodos y los Trematodos evolucionaron como organismos adaptados al parasitismo a partir de organismos ancestrales de vida libre. Los Turbelarios, que mantuvieron su modo de vida libre son posiblemente los organismos evolutivamente más cercanos a este ancestro común de los platelmintos parásitos. En ese sentido la pérdida de dominios en las Dicer-like podría haber sucedido en un ancestro común de los Trematodos. En todo caso, una pregunta interesante que surge de este resultado es el efecto que podría tener en las vías de silenciamiento mediada por ARN pequeños la presencia de una enzima Dicer-like estructuralmente distinta entre Trematodos por un lado, y Cestodos y Turbelarios por otro. La variación estructural en la forma presente en Trematodos sugiere que esta podría cumplir otras funciones.

Hallamos a su vez, que dentro del grupo de los Trematodos existe una diferencia en el número de genes Dicer-like (ver Figura 16). Como puede observarse en la Figura 18a, en *S. mansoni* donde existe una exhaustiva anotación del genoma aparece una única Dicer-like, y también se registra una única copia en los otros Schistosomatidae. Sin embargo, en los genomas de Fasciolidae y Opisthorchidae hay una duplicación invertida de los genes Dicer-like (Figura 18 b y c).

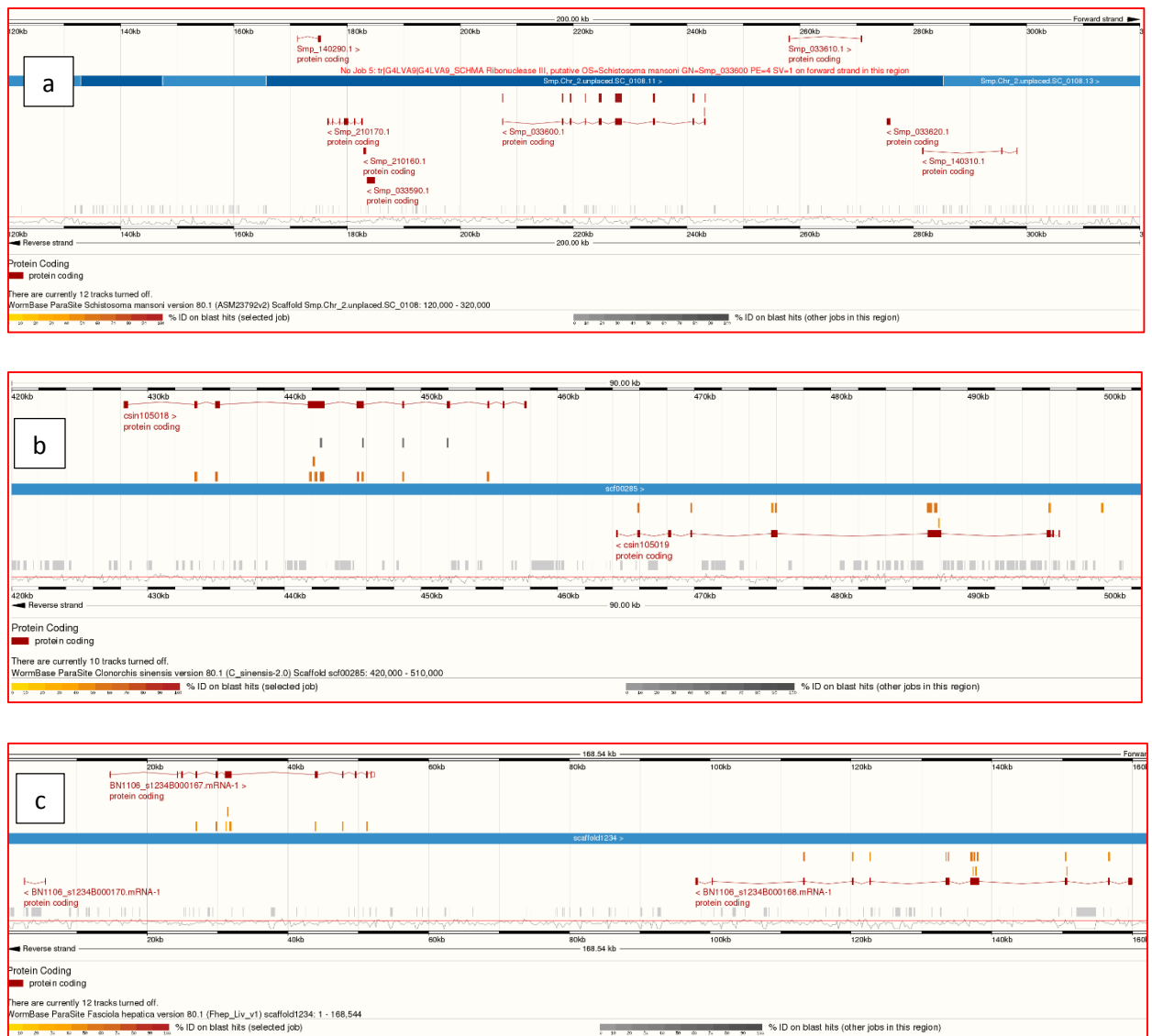


Figura 18. Duplicación invertida de los genes Dicer-like en Fasciolidae y Opisthorchidae. No se observa duplicación de Dicer-like en el genoma de *S. mansoni* (a) como puede verse a partir de una región de 200 Kb del cromosoma 2 centrada en el gen Dicer-like Smp-033600. Sin embargo, en los genomas de los Opisthorchidae (b) y Fasciolidae (c) se observa una duplicación invertida de los estos genes. En *C. sinensis* ambos parálogos están separados por menos de 10 kb (b) en *F. hepatica* (c) la región intergénica es casi de 50kb.

Los genes Dicer-like están separadas por una región intergénica variable, mientras que *C. sinensis* hay 10 kb entre ambos genes (Figura 18b) la región intergénica en el genoma de *F. hepatica* es mayor (50 kb) (Figura 18c). La duplicación de Dicer-like en los genomas de los Fasciolidae y Opisthorchidae pudo haber ocurrido por un evento de recombinación desigual en un ancestro común a ambos grupos.

Con excepción de *M. corti* todos los platelmintos presentaron al menos un homólogo anotado de DCR-1 (Tabla 3) (Tabla Suplementaria 1, Apéndice). Para descartar que la ausencia de ortólogos en *M. corti* se deba al uso de la secuencia de *C. elegans* como query se utilizaron las secuencias ortólogas detectadas en *E. granulosus* para interrogar por BLASTp el proteoma de *M. corti*, sin resultados positivos. Sin embargo cuando se repitió la búsqueda directamente sobre el genoma de *M. corti* por tBLASTn, se lograron identificar múltiples hits en regiones no anotadas de dos contigs diferentes, por lo que confirmamos su presencia. En la Figura 19 se observa que ambas regiones presentaron altos porcentajes de identidad con los genes Dicer (Egr000085200) y Dicer-like (Egr 000181800), respectivamente.

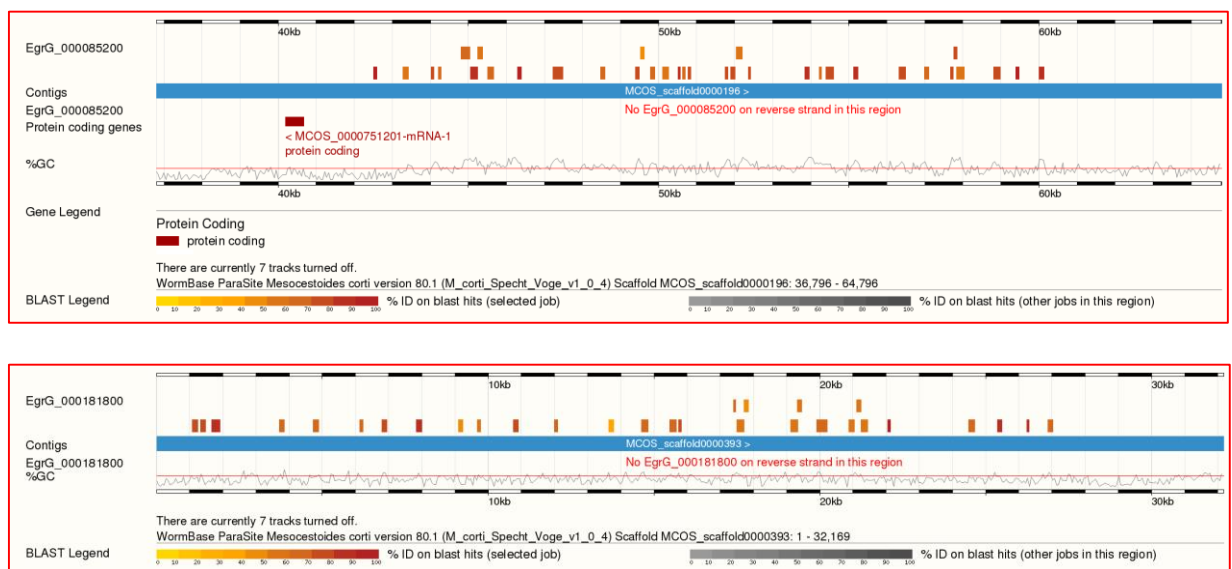


Figura 19. Alineamiento por tBLASTn de los genes de Dicer-1 y Dicer-like de *E. granulosus* en el genoma de *M. corti*.

Se identificaron regiones con altos porcentajes de identidad (marcada por la intensidad de los colores de los hits) para cada uno de los genes de *E. granulosus* utilizados como query.

1.3 Identificación de Argonautas en platelmintos

Las proteínas Argonautas forman una superfamilia proteica que está conservada en organismos tan diversos como archaea o humanos (Jaronczyk, Carmichael, and Hobman 2005). En humanos hay 8 genes Argonauta, en *Drosophila* hay 5 y en *C. elegans* hay 26 (Hutvagner and Simard 2008). En *C. elegans* algunos de los genes de la familia Argonauta tienen funciones muy especializadas (Hutvagner and Simard 2008; Fischer 2010 ; Jamalkandi and Masoudi-Nejad 2011). Las Argonautas se clasifican en tres grupos parálogos: las proteínas Argonauta que son similares a la AGO1 descritas plantas; las Piwi que forman parte de la vía de síntesis de piARNs; y el grupo 3 formado por las proteínas WAGO (worm Ago) identificadas en *C. elegans* (Hutvagner and Simard 2008).

ALG-1 y ALG-2 son las Argonautas canónicas de *C. elegans* y forman un complejo con Dicer que procesa los pre-miARN a miARN maduro, y median el silenciamiento traduccional por el complejo efector miRISC. La búsqueda por homología mostró que hay genes ortólogos a las proteínas de la clase Argonauta en todos los genomas de los Platelminos analizados. El alineamiento generado a partir de estas secuencias halladas en platelmintos se utilizó para generar un árbol filogenético por el método de Neighbor-Joining con un bootstrap de 500 réplicas. En el árbol filogenético se observan cuatro subclados claramente distintos (Figura 20).

En todos los platelmintos se hallaron genes en el subclado Ago, aunque observamos que, como en *C. elegans*, en *S. mediterranea* hay dos genes ortólogos. No se hallaron genes ortólogos en el subclado Piwi en Cestodos y Trematodos pero si en *S. mediterranea*. Los genes Piwi participan en las vías de síntesis de piARNs, los que mantienen la estabilidad genómica en las células de la línea germinal. La ausencia de proteínas de la vía de síntesis de piARNs ya ha sido documentada en platelmintos parásitos (Skinner et al. 2014). A su vez, también se ha reportado la presencia genes Piwi-like en *S. mediterranea* (Palakodeti et al. 2008).

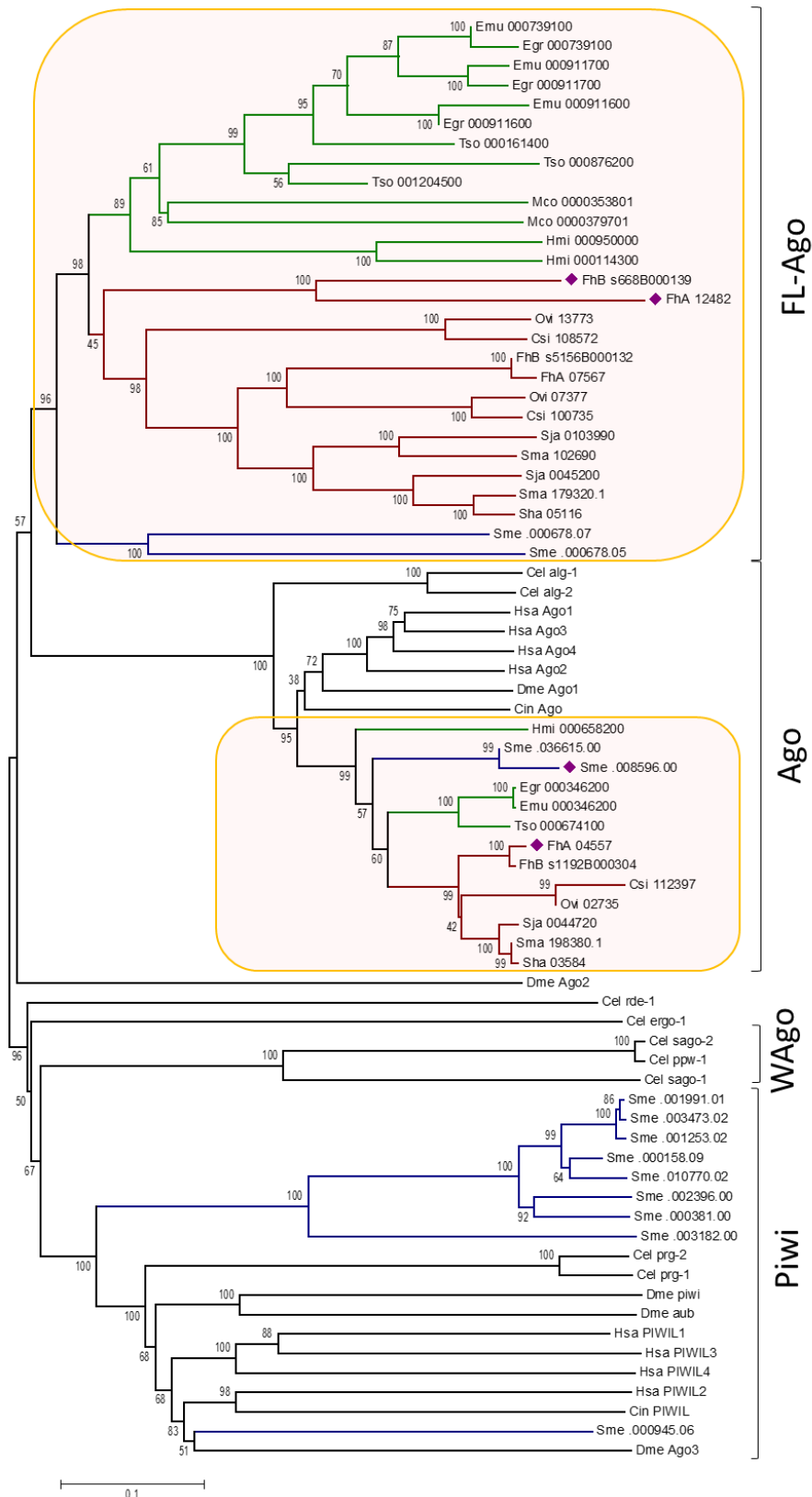


Figura 20. Árbol filogenético generado por el método de Neighbor-joining a partir del alineamiento de ortólogos de Argonautes en diferentes especies. Los porcentajes de Bootstrap se indican en cada rama.

Con corchetes se indican los diferentes subclados. Los recuadros amarillos indican las ramas conservadas en todos los platelminths. Con un ◆ se marcan las secuencias reanotadas (ver Tabla 3).

El tercer subclado, corresponde a las proteínas Wago, que son particulares de nematodos (Hutvagner and Simard 2008; Dalzell et al. 2011). Corroboramos la presencia de un cuarto grupo compuesto únicamente por genes de platelmintos que denominamos Flatworm-Agos (FL-AGO), y que parecen ser una novedad evolutiva de estos organismos (Figura 20, Tabla 3). La existencia de este grupo nuevo de proteínas Argonautas particulares de platelmintos ya fueron reportadas previamente (Zheng 2012; Skinner et al. 2014). Sin embargo, este es el primer trabajo que reporta la presencia de las Flatworm-Agos en tantas especies de platelmintos. La presencia de esta familia en todos los platelmintos que analizamos confirma un origen evolutivo temprano, probablemente previo a la especiación.

1.4 Localización genómica de las FL-Agos

Observamos que en los genomas de *Echinococcus spp.* las FL-Agos están dispuestos en tándem (Figura 21). En *E. multilocularis* los genes Emu000911600 y Emu000911700 están anotadas sobre el mismo contig, y el gen 000739100 en otro (Figura 21A). En *E. granulosus* por otro lado los genes Egr00739100 y Egr000911600 están próximos entre si y el gen 000911700 esta anotada en un contig diferente (Figura 21B). Si bien en un principio se pensó en una anotación cruzada, al analizar el árbol filogenético de las proteínas Argonautas (Figura 20) comprobamos que los genes con la misma numeración correspondían a ortólogos entre las dos especies.

Dada la cercanía evolutiva de las dos especies y a que en ambos genomas los genes FL-Agos mapean en las regiones terminales de sus respectivos contigs creemos que el mapeo cruzado en contigs distintos de estos genes en los dos genomas se deben probablemente a artefactos de los ensamblajes y no a diferencias genómicas reales. Si nuestra hipótesis es correcta, habría una disposición en tándem de los tres loci en los genomas de *Echinococcus*.

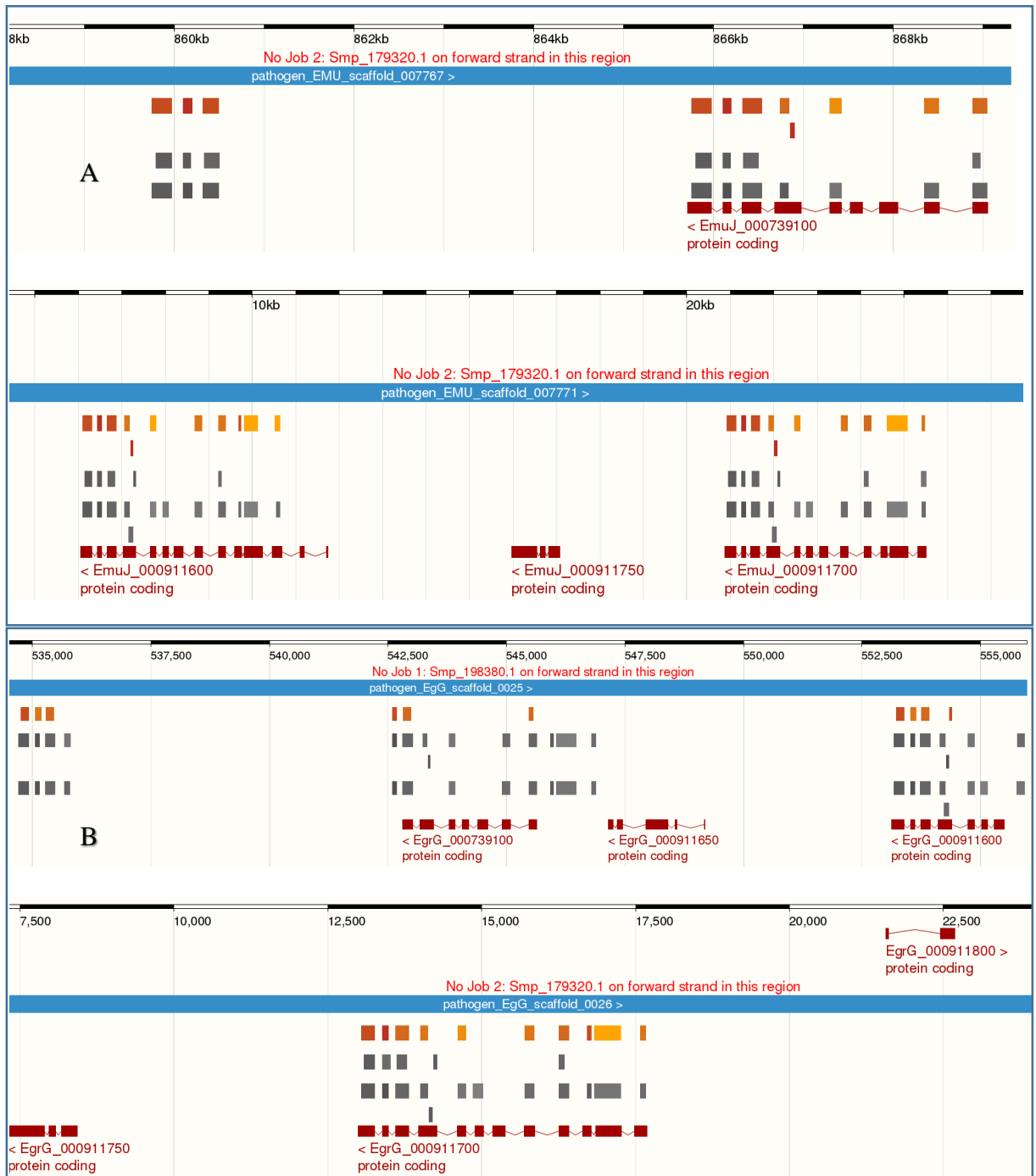


Figura 21. Disposición en tándem de las Flatworm-Agos en los genoma de *E. multilocularis* (A) y *E. granulosus* (B).

Para comprobar si esta estructura génica se mantenía en otros platelmintos observamos las coordenadas de mapeo de los genes de las FL-Agos. En ningún otro cestodo se observó la disposición en tándem descrita en *Echinococcus*. Sin embargo, no es posible descartar que exista en *T. solium* y *H. microstoma* dado el pequeño tamaño de los contigs donde están anotadas las FL-Agos.

Encontramos que las FL-Agos también están dispuestas en tándem en el genoma de *S. mansoni* pero separadas por una distancia de 45 kb aproximadamente. El tamaño reducido de los contigs en otros trematodos no permite aún determinar si esta característica es compartida.

Finalmente, al analizar el árbol filogenético de Argonautas no se observó ningún gen ortólogo de *M. corti* en el subclado de las Ago-like (Figura 20), y solo se observaron dos ortólogos en el subclado de las FL-Agos cuando en otros cestodos hay tres. Para investigar si existían Argonautas no anotadas en el genoma de *M. corti* se usaron los ortólogos de *S. mansoni* para realizar una búsqueda por tBLASTn. A partir de estas búsquedas se detectaron dos regiones putativas no anotadas con altos porcentajes de identidad que podrían corresponder a las proteínas Argonautas faltantes (Figura 22).

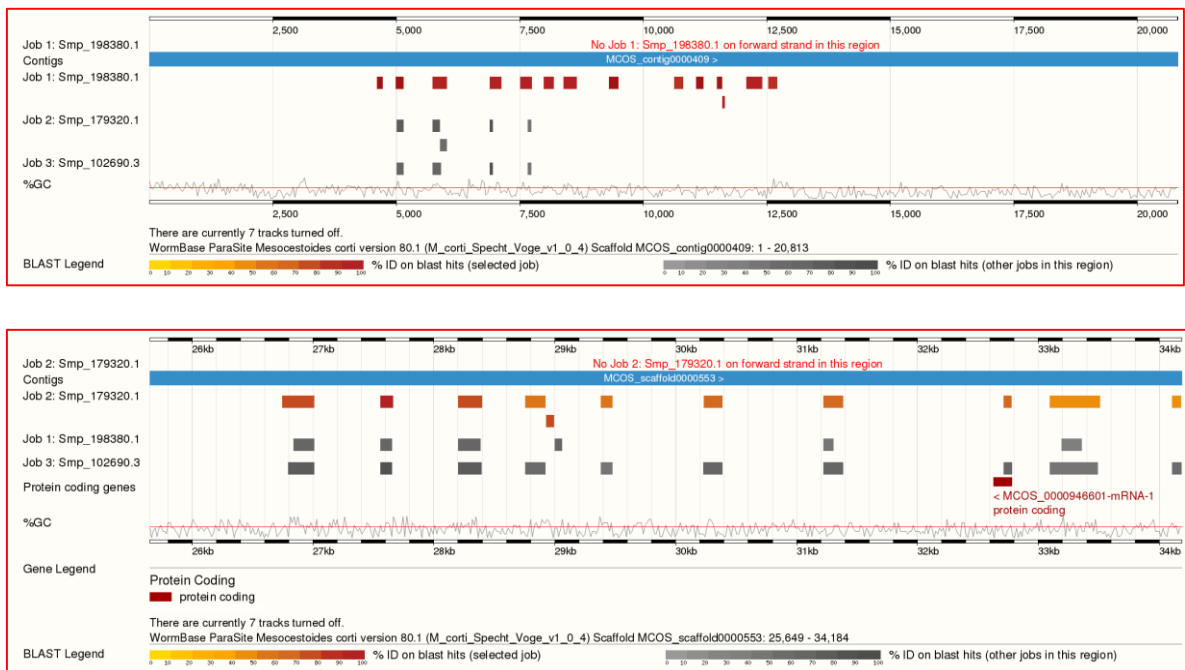


Figura 22. Detección de posibles proteínas Argonauta no anotadas en el genoma de *M. corti*.

Se observa los resultados de las búsquedas por tBLASTn con los ortólogos de proteínas Argonauta detectadas en *S. mansoni*.

La presencia de genes relevantes como Dicer y Argonautas nos lleva a concluir que la vía de regulación de la transcripción génica mediada por ARN pequeños parece estar conservada y ser funcional en platelmintos. Sin embargo, se observaron particularidades como la duplicación y pérdida de algunos genes, y genes carentes de dominios proteicos funcionales. Las diferencias halladas podrían resultar en que las vías de biogénesis por ARN pequeños en platelmintos presenten diferencias particulares con las descritas en organismos modelo como *C. elegans* y, por lo tanto, merecen un estudio más detallado.

Capítulo 2 miARNs expresados en el estadio NEJ de *F. hepatica*

Los miARNs son secuencias de ARN de cadena simple de 21 a 23 nt de longitud que regulan epigenéticamente la expresión génica por la degradación o el secuestro de sus ARNm blanco. Su presencia ha sido demostrada en varios platelmintos incluyendo el estadio adulto de *F. hepatica*, pero se desconoce si la expresión es similar en otros estadios.

Para evaluar la expresión de miARNs en el estadio invasivo se secuenció con tecnología Illumina la fracción de ARN pequeños obtenidos a partir de juveniles desenquistados *in-vitro*. Los reads obtenidos fueron filtrados por calidad y se removieron secuencias contaminantes.

Los reads filtrados se compararon con bases de datos de miARN de repositorios especializados y de otros generados localmente a partir de miARN reportados en platelmintos. Del total de reads obtenidos en la secuenciación aproximadamente 30% correspondieron a ARNs pequeños que fueron agrupados en 40 miARNs distintos. Todos los miARNs reportados en adultos fueron hallados en NEJ con altos niveles de expresión y se agregaron 19 familias al miRNoma de este trematodo (Tabla 1 del Artículo: el miRNoma del juvenil de *F. hepatica*, pág. 63). Se analizaron los niveles de expresión observándose grandes diferencias entre diversos miARNs.

Para predecir nuevos miARNs, se generó un ensamblaje local de datos genómicos disponibles públicamente, y este ensamblaje fue utilizado como referencia para mapear los reads obtenidos usando el paquete MiRDeep2. Usando criterios estrictos de filtrado se identificaron cinco candidatos con buen plegamiento y significancia estadística. Para saber si estas secuencias noveles estaban presentes en otros trematodos se interrogaron los genomas de *C. sinensis*, *S. mansoni* y *O. viverrini*, con resultados negativos. Consecuentemente estos miARNs noveles son exclusivos y podrían estar involucrados en la regulación de vías asociadas al ciclo de vida y/o al desarrollo de *F. hepatica*.

Para determinar la distribución filogenética de los miARNs en el estadio NEJ se compararon las entradas con anotaciones similares en miRBase y mirOrtho y se identificaron como presentes o ausentes en cada taxón. Se buscó confirmar o refutar estudios previos que sugerían una simplificación en el miRNoma de los platelmintos (Fromm et al. 2013; Bai et al. 2014).

Evaluamos la conservación de secuencias en las familias de miRNAs detectadas mediante alineamiento de las secuencias de precursores y miARNs maduros de platelmintos y otros organismos. Si bien los miARNs están entre las secuencias evolutivamente más conservadas (Wheeler et al. 2009), observamos una menor conservación fuera de la región semilla en platelmintos en comparación a otros linajes.

Finalmente analizamos un grupo particular de reads con un tamaño medio de 32 nt que presentaban homología con ARNts. Se mapeó la fracción de reads de 25-35 nt al genoma, se comparó con los ARNt predichos con tRNAscan-SE en el genoma ensamblado localmente.

Se encontró que correspondían a mitades de ARNt, un tipo de moléculas potencialmente reguladores recientemente descritas en otros sistemas (Thompson and Parker 2009). Si bien no sabemos el papel que podrían tener los ARNt halves en *F. hepatica*, este trabajo plantea nuevas preguntas sobre la regulación génica en *Fasciola*.

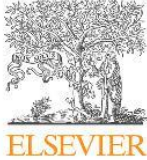
Los resultados correspondientes a este trabajo se publicaron en el artículo que se adjunta a continuación:

Fontenla S, Dell’Oca N, Smircich P, Tort JF, Siles-Lucas M. The miRNome of *Fasciola hepatica* juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms. *Int J Parasitol.* 2015; 45: 901–913.

2 Artículo: el miRNoma del juvenil de *F. hepatica*

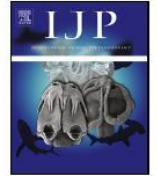
ARTICLE IN PRESS

International Journal for Parasitology xxx (2015) xxx–xxx



Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

The miRnome of *Fasciola hepatica* juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms [☆]

Santiago Fontenla ^a, Nicolás Dell'Oca ^a, Pablo Smircich ^{a,b}, José F. Tort ^{a,*}, Mar Siles-Lucas ^c^aDepartamento de Genética, Facultad de Medicina, Universidad de la República (UdelaR), Montevideo, Uruguay^bLaboratorio de Interacciones Moleculares, Facultad de Ciencias, Universidad de la República (UdelaR), Montevideo, Uruguay^cInstituto de Recursos Naturales y Agrobiología de Salamanca (IRNASA), Consejo Superior de Investigaciones Científicas (CSIC), Salamanca, Spain

ARTICLE INFO

Article history:

Received 22 May 2015

Received in revised form 27 June 2015

Accepted 30 June 2015

Available online xxxx

Keywords:

Fasciola hepatica
Newly excysted juveniles
miRNAs
tRNA halves

ABSTRACT

The liver fluke *Fasciola hepatica* is a foodborne zoonotic parasite affecting livestock worldwide, with increasing relevance in human health. The first developmental stage that the host meets after ingestion of the parasite is the newly excysted juvenile, that actively transverse the gut wall and migrates to its final location in the liver. The regulation of the early developmental events in newly excysted juveniles is still poorly understood and a relevant target for control strategies. Here we investigated the putative involvement of small regulatory RNAs in the invasion process. The small RNA population of the newly excysted juvenile fall into two classes, one represented by micro (mi)RNAs and a secondary group of larger (32–33 nucleotides) tRNA-derived sequences. We identified 40 different miRNAs, most of those belonging to ancient miRNAs conserved in protostomes and metazoans, notably with a highly predominant miR-125b variant. Remarkably, several protostomian and metazoan conserved families were not detected in consonance with previous reports of drastic miRnome reduction in parasitic flatworms. Additionally, a set of five novel miRNAs was identified, probably associated with specific gene regulation expression needs in *F. hepatica*. While sequence conservation in mature miRNA is high across the metazoan tree, we observed that flatworm miRNAs are more divergent, suggesting that mutation rates in parasitic flatworms could be high. Finally, the distinctive presence of tRNA-derived sequences, mostly 5' tRNA halves of selected tRNAs in the small RNA population of newly excysted juveniles, raises the possibility that both miRNA and tRNA fragments participate in the regulation of gene expression in this parasite.

© 2015 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Fasciolosis is one of the most widespread helminthic diseases of livestock, caused by the liver flukes *Fasciola hepatica* in temperate areas and *Fasciola gigantica* in tropical regions of Africa, the Middle East and south eastern Asia. Economic losses related to this infection are estimated in billions of US \$, mainly due to reduction in meat, wool and milk output by infected animals, with additional costs derived from liver condemnation and use of flukicide drugs (Mas-Coma et al., 2005; Charlier et al., 2014). In addition to its impact on production, fasciolosis is becoming a major concern in human health. The World Health Organization has estimated that

at least 2.4 million people are infected in more than 70 countries worldwide (http://www.who.int/foodborne_trematode_infections/fascioliasis/en/). While treatment with anthelmintic drugs is effective, it does not usually prevent re-infection and growing reports of resistance are emerging in several countries (Brennan et al., 2007). Novel drugs and/or strategies for controlling this infection are urgently needed. Understanding the molecular mechanisms behind the early invasion process might provide clues for novel control strategies.

Fasciola hepatica infection in livestock or humans occurs when water plants (e.g., watercress) contaminated with the resistant stage of the parasite (metacercariae) are ingested. Gastric and duodenal contents promote larval excystation, and the newly excysted juveniles (NEJs) actively migrate through the intestinal wall, the peritoneal cavity and the liver parenchyma into the biliary ducts in the liver where they reach maturity (Keiser and Utzinger, 2009). Clinically, parasite invasion is characterised by marked eosinophilia, abdominal pain and fever. Intermittent obstruction

[☆] Note: New sequence data reported in this paper are available in the GEO database under accession number [GSE66490](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE66490).

* Corresponding author at: Departamento de Genética, Facultad de Medicina, Universidad de la República (UdelaR), Gral. Flores 2125, CP 11800 Montevideo, Uruguay. Tel.: +598 2924 9562.

E-mail address: jtort@fmed.edu.uy (J.F. Tort).

<http://dx.doi.org/10.1016/j.ijpara.2015.06.007>

0020-7519/© 2015 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

ARTICLE IN PRESS

2

S. Fontenla et al. / International Journal for Parasitology xxx (2015) xxx–xxx

of bile ducts after worm entry causes symptoms that resemble biliary colic and cholecystitis (García et al., 2007). Chronic infection has been associated with hepatic fibrosis and cirrhosis (Marcos et al., 2008). During the invasion process the invasive NEJs are confronted with different environments in their passage from the intestine to the liver. It has been demonstrated that changes in gene expression underlie the early stages of invasion (Cancela et al., 2008, 2010) but little is known of how these processes are regulated.

MicroRNAs (miRNAs) have recently emerged as relevant regulators of metazoan gene expression, being involved in many different biological processes, and play a key role in developmental timing (Berezikov, 2011). These small regulatory single-stranded RNAs are widely conserved through metazoan evolution and once acquired they are rarely lost in descendent branches (Sempere et al., 2007; Wheeler et al., 2009; Tarver et al., 2013). They are transcribed as hairpin precursors that are cleaved by the nuclear ribonuclease III Drosha. The resulting double-stranded (ds) pre-miRNAs are exported to the cytoplasm by exportin 5, where they are further processed by the ribonuclease III Dicer, producing a 21–23 nucleotide (nt) dsRNA. One of the strands of this duplex is the mature miRNA that is loaded into the miRNA Induced Silencing Complex (miRISC). By pairing to complementary sequences most commonly located in the 3' untranslated region (3'-UTR) of diverse target mRNAs, different miRNAs can trigger mRNA degradation and/or blocking of translation (Bartel, 2009). In addition, miRNAs can silence gene expression by sequestering mRNAs and by chromatin remodelling of target genes (Grishok, 2005; Eulalio et al., 2008; Jamalkandi and Masoudi-Nejad, 2009; Fischer, 2010). MiRNAs* corresponding to the complementary strand of mature miRNAs, and first thought to be degraded after separation of mature miRNAs, may also have a regulatory function in gene expression (Childiyal et al., 2010).

Regulation of gene expression by miRNAs has been extensively demonstrated in all metazoans, including several reports in flatworms (Manzano-Román and Siles-Lucas, 2012). Early studies detected miRNAs in free living planarians (Palakodeti et al., 2006; Friedländer et al., 2009; Lu et al., 2009), and were promptly followed by studies in schistosomes (reviewed in Zhu et al., 2014), cestodes (Cucher et al., 2011; Ai et al., 2012; Bai et al., 2014) and a monogenean species (Fromm et al., 2013). Within trematodes other than schistosomes, miRNAs has only been reported in *Clonorchis sinensis* (Xu et al., 2010) and the *Fasciola* spp. *F. hepatica* and *F. gigantica* (Xu et al., 2012). This later study describes a few miRNA from the adult liver stage of both species. To gain further insight into the putative role of small regulatory RNAs in the invasion process in *F. hepatica*, we investigated the expression profile of miRNAs in the NEJ stage of the parasite, comparing it with what has been described in other flatworms, both free living and parasitic. We describe the presence of several conserved miRNA families and we discover some novel miRNAs that might be liver fluke-specific. In addition, we describe flatworm miRNA sequences less conserved in comparison with other metazoans, and discuss the relevance of these findings in view of their parasitic lifestyle.

2. Materials and methods

2.1. In vitro excystation of metacercariae

Fasciola hepatica metacercariae (field strain Shrewsbury) were purchased from Ridgeway Research Ltd. (UK). Five thousand metacercariae were excysted in vitro as described previously (Hernández-González et al., 2010). Briefly, pure CO₂ was bubbled for 30 s in 10 ml of cold distilled water. Sodium dithionite to a final concentration of 0.02 M was added to the water, and the tube

was closed and incubated at 37 °C until a fine cloudy precipitate was produced. Metacercariae were then added to the tube and further incubated at 37 °C for 1 h. After incubation, parasites were washed twice by sedimentation with warm distilled water, and 5 ml of Hank's balanced salt solution (Sigma, USA) plus 10% (v/v) rabbit bile and 30 mM HEPES (Sigma) pH 7.4 was added to the dry metacercariae. Parasites were then incubated for 5 h at 37 °C. Emerging active parasites were collected with a 20 µl pipette, washed twice with sterile PBS, and stored dry at –80 °C until use.

2.2. Small RNA isolation and sequencing

Samples were homogenised using the Tissulyser II (QIAGEN, USA) and the RNA extraction was performed with the RNeasy Mini Kit (QIAGEN, USA) following the manufacturer's instructions. The amount and integrity of total RNA was determined using a 2100 BioAnalyzer (Agilent, USA). Three micrograms of total RNA were size-fractionated on a 15% TBE urea polyacrylamide gel and a 15–30 bp fraction was excised. The RNA was retrotranscribed into single-stranded cDNA using the Superscript II reverse transcriptase (Invitrogen, USA) and the Illumina small RNA RT-Primer following the manufacturers' instructions. The cDNA was PCR amplified for 15 cycles using an Illumina small RNA primer set. The PCR products were quantified and subjected to single end Illumina sequencing for 75 cycles. Raw sequencing reads, as well as processed data, were deposited in the GEO database (www.ncbi.nlm.nih.gov/geo/) under accession number **GSE66490** and submitted to miRbase (www.mirbase.org/).

2.3. Data processing and filtering

After masking of adaptor sequences and removal of reads smaller than 18 nt, the clean reads were subjected to quality analysis with Fastx Tools (http://hannonlab.cshl.edu/fastx_toolkit/). Reads where 90% of the bases had a Phred score higher than 20 were conserved. Redundant reads were merged and counted. High quality reads were screened against transcriptomic (mRNA) data of platyhelminths available at GenBank (www.ncbi.nlm.nih.gov/nucleotide/), SchistoDB (schistodb.net/schisto/) (Zerlotini et al., 2013) and HelmDB (gasser-research.vet.unimelb.edu.au/helmdb/) (Mangiola et al., 2013), discarding all reads matching over 85%. Data were further screened with the same stringent criteria (mismatches ≤ 3) against other non-coding RNA databases such as Rfam (rfam.xfam.org/) (Burge et al., 2013) and fRNAdb (www.ncrna.org/frnadb/) (Mituyama et al., 2009) (both excluding miRNA) and those reported in *Schistosoma mansoni* and *Schistosoma japonicum* (Copeland et al., 2009) to identify and remove non-coding (nc) RNAs, rRNA, tRNA, small nuclear (sn)RNA, small nucleolar (sno) RNA and other ncrRNAs.

2.4. Data analysis and miRNA classification

The processed read dataset was used to perform homology searches against the Sanger miRBase release 20 (www.mirbase.org/) (Kozomara and Griffiths-Jones, 2014), the miRNA sections of Rfam (Burge et al., 2013), and miRNAs present in the functional RNA database (www.ncrna.org/frnadb/) (Mituyama et al., 2009). Furthermore, processed reads were compared with a local database generated with putative flatworm miRNAs reported in the literature that were not integrated in miRBase, Rfam or fRNAdb. This database included sequences from the *S. japonicum* miRNA database (http://omics.biosino.org:14000/kweb/sj_miRNA/) and those published in *S. japonicum* (Huang et al., 2009), *S. mansoni* (Simões et al., 2011), *C. sinensis* (Xu et al., 2010), *F. hepatica* and *F. gigantica* (Xu et al., 2012), and the monogenean *Gyrodactylus salaris*

Please cite this article in press as: Fontenla, S., et al. The miRnome of *Fasciola hepatica* juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms. Int. J. Parasitol. (2015), <http://dx.doi.org/10.1016/j.ijpara.2015.06.007>

ARTICLE IN PRESS

S. Fontenla et al./International Journal for Parasitology xxx (2015) xxx–xxx

3

(Fromm et al., 2013). Reads showing high similarity (>85%) to known miRNAs of other organisms (mismatches ≤ 3) were classified into the same miRNA family. All miRNAs showing the same annotation and slight differences in length or sequence were inspected and their read counts added if they had less than three mismatches. At this stage all miRNA candidates with less than 10 reads or lengths below 20 or over 23 nts were discarded. The searches were performed using miRExpress (mirexpress.mbc.nctu.edu.tw/) (Wang et al., 2009) and the miRNA module of CLC Genomics Workbench v6.5 (CLC Inc, Aarhus, Denmark) with similar outputs from both procedures. Consequently the remaining analyses were performed with the dataset generated by CLC Genomics Workbench. Read counts were normalised and expressed as reads per million.

2.5. Phylogenetic distribution and conservation of miRNA families found in *F. hepatica* NEJs

The phylogenetic distribution of the miRNA candidates was assessed based on the family classifications made by Wheeler et al. (2009) and Tarver et al. (2013), and further confirmed by comparing the NEJ putative miRNAs with entries with similar annotations in miRBase (Kozomara and Griffiths-Jones, 2014) and/or mirOrtho (cegg.unige.ch/mirortho/) (Gerlach et al., 2009). The absence and presence of members of the different families in deuterostomes, ecdysozoans, lophotrochozoans and platyhelminths were registered. Sequences of different precursors and mature miRNAs from selected species covering the metazoan diversity (see list below) were downloaded from miRBase and aligned with the corresponding miRNAs from flatworms using the multiple aligner MAFFT (Katoh and Standley, 2013). Alignments were visualised within CLC Genomics Workbench, generating the corresponding consensus logos. Conservation was quantified by counting substitutions using the most parsimonious interpretation following the procedure described by Wheeler et al. (2009). Briefly, if a substitution is shared by closely related species or sister taxa (i.e. shared in cestodes) it was counted as a single change; if more than two different residues were present in a position within a taxa they were counted as separate individual changes. Changes by position were averaged for deuterostomes, ecdysozoans and platyhelminths.

Species considered for alignments were: Deuterostomia (human, cattle, mouse, chicken, clawed frog, zebrafish, lancelet (cephalochordate), Ciona (urochordate) and sea urchin (echinodermata)); Ecdysozoa (Arthropoda: (fruit fly, mosquito, bee, silkworm, beetle), nematodes: *Caenorhabditis elegans* (free living roundworm), *Pristionchus pacificus* (free living roundworm), *Haemonchus contortus* (barber's pole worm, parasite) and *Brugia malayi* (lymphatic filariae, parasite)), Lophotrochozoa (*Capitella teleta* (annelid worm) and *Lottia gigantea* (owl limpet, mollusk)); Platyhelminthes (*Schmidtea mediterranea*, (free living planaria), *G. salaris* (monogenean fish ectoparasite), Trematoda: *S. mansoni*, *S. japonicum* (blood flukes), *C. sinensis* (bile duct fluke), *F. hepatica* (liver fluke), Cestoda: *Echinococcus granulosus* (hydatid cyst tapeworm) and *Echinococcus multilocularis* (alveolar cyst tapeworm)).

2.6. Identification of novel miRNAs

After requesting permission, *F. hepatica* genomic sequencing reads generated by the Wellcome Trust Sanger Center (England) and deposited at the European Bioinformatics Institute repository (ftp.sra.ebi.ac.uk/vol1/fastq/ERR065/ERR065038/) were downloaded. The crude pair end reads were preprocessed by trimming and quality filtering, and assembled de novo with CLC Genomic Workbench. In order to identify novel *F. hepatica*-specific miRNAs, sequences that produced no matches in the previous analysis and

with lengths between 20 and 23 nts were mapped with miRDeep2 (Friedländer et al., 2012) on this genomic assembly. Putative predicted miRNAs were considered novel if they had a miRDeep2 score equal to or higher than 10. The resulting predictions were then manually checked for possible false positives. The same approach was followed to confirm those miRNA candidates found by homology, using a less stringent miRDeep2 score cut-off of 1, since only confirmation was required. Resulting predictions were manually inspected in order to discard false positives. Furthermore, since two different assemblies of the *F. hepatica* genome were recently released (PRJEB6687 (Cwiklinski et al., 2015), PRINA179522 (Martin et al., 2015)) all the identified miRNAs were mapped to these assemblies in order to confirm their presence and location.

2.7. Identification of tRNA-derived sequences

The primary analysis of the 25–35 nt long reads showed homology to tRNA sequences. In order to confirm this, we mapped the 25–35 nt fraction to the partial genome assembly of *F. hepatica*. Reads that mapped to the genome were searched, using BLAST, against tRNAs predicted with tRNAscan-SE (Lowe and Eddy, 1997) with a cutoff of E^{-10} . A total of 99,014 reads matched tRNAs. The tRNA-derived sequences were mapped to the structures available in the tRNAscan-SE to determine their origin.

3. Results and discussion

3.1. Small RNAs identified in the juvenile liver fluke have two length classes corresponding to miRNA and tRNA halves

In order to initiate the characterisation of the miRNAs in the invasive juvenile stage of *F. hepatica*, we obtained and sequenced the small RNA population from in vitro NEJs. From an initial set of nearly 10 million raw reads, more than 8.6 million remained after adapter clipping and quality trimming, indicating the good quality of the dataset generated. These reads were collapsed into more than 2 million different unique sequences, with diverse levels of redundancy. The length distribution of this collection of high quality sequences show a prominent peak at 21 nts, compatible with miRNAs and a secondary peak at 32–33 nts (Supplementary Fig. S1). Since there was no public genome assembly for *F. hepatica* to map these sequences when we initiated this study, our analysis pipeline was based on homology searches to public databases. More than a quarter of the sequences (569,978) were discarded after comparison with all the available flatworm transcriptomes, since those might have originated from partial mRNA transcripts. Another set of sequences were also discarded due to their homology to other functional ncRNAs such as rRNA, tRNA, snRNA, snoRNA, and other ncRNAs (see Section 2.3). The remaining set consisted of 1,326,249 putative small RNAs representing more than 3 million reads, approximately 30% of the initial dataset (Fig. 1).

In order to classify this collection of small RNAs, we compared them with miRBase and other putative miRNAs from flatworms described in the literature but not included in miRBase (see Section 2.4 for details). We identified representatives of 40 miRNAs, most of those belonging to conserved families, as well as a few putative miRNAs only described to date in flatworms (Table 1). All the conserved miRNA families previously reported in the adult stage of *F. hepatica* (fhe-mir-10, fhe-let-7, fhe-mir-71, fhe-mir-124 -reported before as sja miR-12-, fhe-mir-2 and fhe-bantam) (Xu et al., 2012) were detected, being expressed at very high levels in the juvenile stage. Furthermore, we extended the miRnome of *F. hepatica* by identifying representatives of 19 additional conserved families of miRNAs. None of the novel sequences reported in adults

Please cite this article in press as: Fontenla, S., et al. The miRnome of *Fasciola hepatica* juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms. Int. J. Parasitol. (2015), <http://dx.doi.org/10.1016/j.ijpara.2015.06.007>

ARTICLE IN PRESS

4

S. Fontenla et al. / International Journal for Parasitology xxx (2015) xxx–xxx

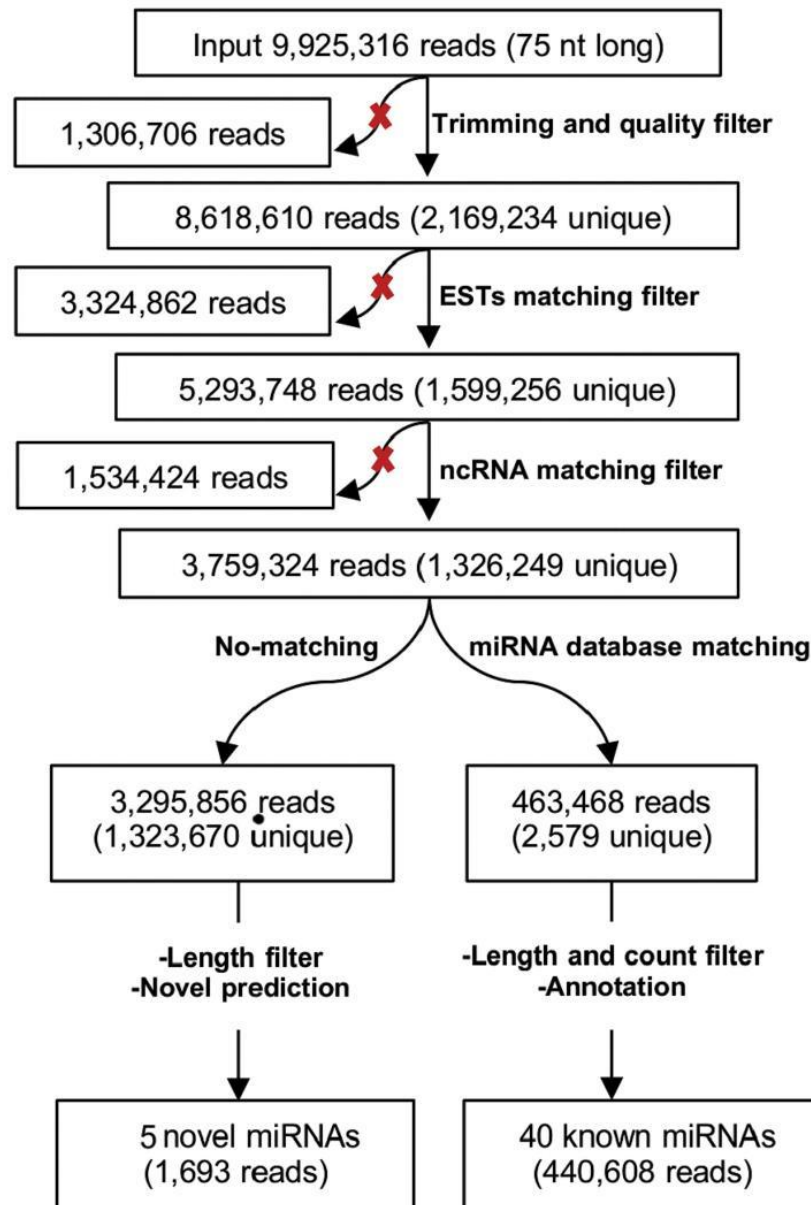


Fig. 1. Data processing pipeline of small RNA sequencing reads from *Fasciola hepatica* newly excysted juveniles. Schematic representation of the data processing pipeline followed; crude reads (in nucleotides) were first trimmed and quality filtered, then filtered against expressed sequence tags (ESTs) and non-coding (nc)RNA (rRNA, tRNA, small nuclear (sn)RNA, small nucleolar (sno)RNA). Filtered reads were screened for homology against micro (mi)RNA databases (mirBASE, Rfam and FRNAdB) and a local database of published miRNAs from flatworms. Conserved miRNAs were annotated by homology to miRBase and miROrtho, while novel miRNAs were predicted with miRDeep2 from the remaining reads based on an in house assembly of the publically available *Fasciola hepatica* genomic data.

(Xu et al., 2012) was detected here, with the exception of three sequences (Fgi-miR-novel-10, Fgi-miR-novel-15 and Fhe-miR-novel-7) that actually correspond to repetitive sequences rather than bona fide miRNAs.

Despite the substantial skew of miRBase to vertebrate sequences, most of the *F. hepatica* miRNA sequences find their best matches within miRNA sequences from other helminths. The length distribution of the identified miRNAs coincides with the 21 nt peak observed in the crude high quality reads, indicating that miRNAs are the prominent class of small RNAs found in the com-

mon liver fluke. We also analysed the secondary peak centred at 32 nts, finding that it corresponds mainly to tRNA-derived sequences, mostly tRNA halves (see Section 3.5). However, the vast majority of the dataset remained unannotated with positive hits representing less than 1% of the total small RNAs (2,579 miRNAs) and 12% of the reads (463,468).

In order to identify putative novel miRNAs within this non-matching group of reads, we decided to compare them with partial genomic sequences available at the moment. An in-house generated assembly of the available (but partial) *F. hepatica* genome

Please cite this article in press as: Fontenla, S., et al. The miRNome of *Fasciola hepatica* juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms. Int. J. Parasitol. (2015), <http://dx.doi.org/10.1016/j.ijpara.2015.06.007>

ARTICLE IN PRESS

S. Fontenla et al. / International Journal for Parasitology xxx (2015) xxx–xxx

5

Table 1
Conserved micro (mi)RNAs identified in newly excysted juveniles of *Fasciola hepatica*.

miRNome		Small RNA sequence
bantam-3p	^a	UGAGAUCGCGAUUAAAAGCUG
let-7c-5p		AGAGGUAGUGACUCAUUGACU
let-7-5p	^a	UGAGGUAGUGGUUUAUAGU
miR-1-3p		UGGAAUUGGCGAAGUAUGGU
miR-2a-3p		UCACAGCCAGAAUUGAUGAACG
miR-2b-3p	^a	UAUCACAGCCUGCUUGGACA
miR-2c-3p		UAUCACAGCCUGCUUAGGGCU
miR-2d-3p		UAUCACAGCCUGCUUAGGGU
miR-2e-3p	^a	UAUCACAGUCCAAGCUUUGG
miR-2e-5p		UACCAACUAGACUGCGUUUAU
miR-2f-3p		UCACAGCCAAUUAUGAGCCA
miR-7-5p		UGGAAAGACUGGUAUUGUUGU
miR-8-3p		UAUAUCUGUUGUUAAGAUCC
miR-9-5p		UCUUUGGUUAUCAAGCAGUAUG
miR-10-5p	^a	AACCCUGUAGACCCGAGUUGCA
miR-10-3p		AAAUJCCGAGUCUACAAGGAAC
miR-31-5p		UGGCAAGAUUAUGCGAAGCUG
miR-36a-3p		UCACCCGGUAGACAUUAUCAC
miR-36b-3p		CCACCCGGUAGACAUUAUCAC
miR-46/281-3p		UGUCAUGGAGUUGUCUCUACA
miR-61-3p		UGACUAGAAAGUGCACUCACUUC
miR-61-5p		UGUGAGUCUCUUUCUUGUCAUG
miR-71a-5p	^a	UGAAAGACGUAUGUAGAGAGU
miR-71b-5p	^a	UGAAAGACUUGAGUAGUGAG
miR-87-3p		GUGAGCAAAGUUUACGGUGUA
miR-96-5p		CUUGGCACUUUGGAAUUGUCA
miR-124-3p	^a	UAAGGCACCGCGUGAAUGUCA
miR-125a-5p		UCCUGAGACCCUAGAGUUUC
miR-125b-5p		CCCCUGAGACUGAAUUAUGCUC
miR-190-5p		AGAUUAGUUUGGGUUAUCUUGG
miR-190-3p		CCAGUGACCAACAUAUUUCUC
miR-219-5p		UGAUUUGCCAUUCGCAUUUCUUG
miR-277-3p		UAAAUGCAUUUCUGGCCCC
miR-307/67-3p		UCACAACCUACUUGAUUGAGGGG
miR-745b-3p		AAAGCUGCCAAGCGAAGGGC
miR-745-3p		UGCUCUUUAUAGAGCUGUG
miR-750-3p		CCAGAUUCAGUCUUCAGCUCU
miR-2162-3p		UAUUUAUGCAACAUAUUCACUCU
miR-3479-3p		UAUUGCACUUUCUUCGCCUUA
miR-3487-5p		UCCCGUAUUGCAACUGUUGU

^a miRNAs already found in adult *F. hepatica* (Xu et al., 2012).

sequences obtained from the Sanger Center was used to identify putative novel miRNAs using the miRNA prediction programme miRDeep2. The analysis produced five novel precursor sequences with good confidence statistics and appropriate hairpin structures (Fig. 2, Table 2 and Supplementary Fig. S2). These novel *F. hepatica*-specific miRNA sequences were used to interrogate the available genomes of *S. japonicum*, *Opisthorchis viverrini* and *C. sinensis* in order to explore whether they represent conserved trematode novelties, but were not detected in any of them. Consequently they might be specific, modulating, relevant pathways associated with the life cycle and developmental stages of the liver fluke.

Besides these putative novel miRNAs, we confirmed all those predicted by homology by mapping them to our in-house assembly with miRDeep2. Very recently, two draft assemblies of the *F. hepatica* genome were made available (Cwiklinski et al., 2015; Martin et al., 2015), providing an essential tool to further investigate the miRNome of this species. Since these novel assemblies are based on different datasets than the one we used, we confirmed all the miRNAs identified by mapping them to these reference genomes. The supporting evidence is summarised in Supplementary Table S1.

3.2. Ancient metazoan and protostome conserved miRNAs are the most abundant in juvenile liver flukes

The diverse miRNAs found in *F. hepatica* juveniles are quantitatively expressed at very different levels. Few conserved

sequences are strongly predominant in the sample, while a wide dynamic range of miRNAs that varies from single to hundreds of thousands of reads can be detected. Only one miRNA, miR-125b, accounts for more than three quarters of the total reads identified in *F. hepatica* juveniles, and just five miRNAs represent 90% of the total reads (Fig. 3). Similarly skewed expression levels have been observed in *S. japonicum* (Cai et al., 2011, 2013). Unfortunately, the adult *F. hepatica* dataset is very scant, and no comparable quantitative data is available (Xu et al., 2012), so it was not possible to check whether expression levels change from juvenile to adult stages.

The most abundantly expressed miRNA is miR-125b, one of the oldest known miRNAs well conserved in all metazoans (Sokol, 2012). MiR-125 has been related to very diverse functions, and it is highly studied in humans due to its association with both tumour promoting and tumour suppression phenotypes. Several mRNA targets for miR-125 have been identified including transcription factors, apoptotic pathway genes and genes associated with host defense, especially in response to bacterial or viral infections (Sun et al., 2013). In several metazoans, miR-125 is part of a genomic cluster that also includes the precursors for miR-100 and let-7. Interestingly, let-7 and miR-125 are clustered in *S. mansoni* and in the planarian *S. mediterranea* (Hertel et al., 2012), but they are separated in the genomes of the cestodes, while miR-100 is absent from all flatworms analysed to date (Jin et al., 2013). Let-7 is the third most highly expressed miRNA in NEJs, if we consider a novel variant of let-7 (let-7c) which is poorly conserved out of the seed region (Table 1). This variant is identical to a putative miRNA described as novel in *S. japonicum* (Sja-novel-137), that was also among those miRNAs highly expressed in adults and schistosomules (Hao et al., 2010) and has homology to the planarian sme-let-7c and sme-let-7d. The canonical let-7 was also found here but expressed at very low levels. It is noteworthy that another putative miRNA described as novel in *S. japonicum*, Sja novel-148 (Hao et al., 2010), corresponds to miR-125b. Despite the clustered organisation of let-7 and miR-125 in planarians and schistosomes, our mapping of data to all available genome assemblies suggests that they are separated in *F. hepatica* as in cestodes (see Supplementary Table S1).

One of the most ancient miRNAs that emerged in the protostome lineage, bantam, is also thoroughly represented in NEJs. In insects, bantam has been implicated in the regulation of cell proliferation and apoptosis (Brennecke et al., 2003) and more recently it has been demonstrated to have a role in systemic growth by regulating ecdysone production (Boulan et al., 2013). Interestingly, bantam expression was differentially upregulated in *S. mansoni* females from paired infections when compared with female only infections, suggesting that it plays a role in the pairing-dependent female sexual maturation (Sun et al., 2014). Contrasting roles in this process might be played by miR-1 and miR-71 which were upregulated in unpaired females (Sun et al., 2014). Both miRNAs were detected in the NEJs of *F. hepatica*, but only miR-71 is highly expressed.

MiR-71 is found only in protostomes and resides in a cluster with members of the unrelated miR-2 family. This cluster has been duplicated in the genome of schistosomes (De Souza Gomes et al., 2011), and four additional copies exist in planarians, but a single cluster is found in the genomes of cestodes (Jin et al., 2013). The reason for the expansion of the cluster in specific flatworm lineages remains unknown. We found all the described members of the miR-2 families and two isoforms of miR-71. To further investigate the genomic organisation of the loci responsible for the expression of these miRNAs, we mapped these reads to our partial genome assembly and found two clusters as in other trematodes (Fig. 4). This arrangement was also confirmed by mapping to the novel genome assemblies.

Please cite this article in press as: Fontenla, S., et al. The miRNome of *Fasciola hepatica* juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms. Int. J. Parasitol. (2015), <http://dx.doi.org/10.1016/j.ijpara.2015.06.007>

ARTICLE IN PRESS

6

S. Fontenla et al. / International Journal for Parasitology xxx (2015) xxx–xxx

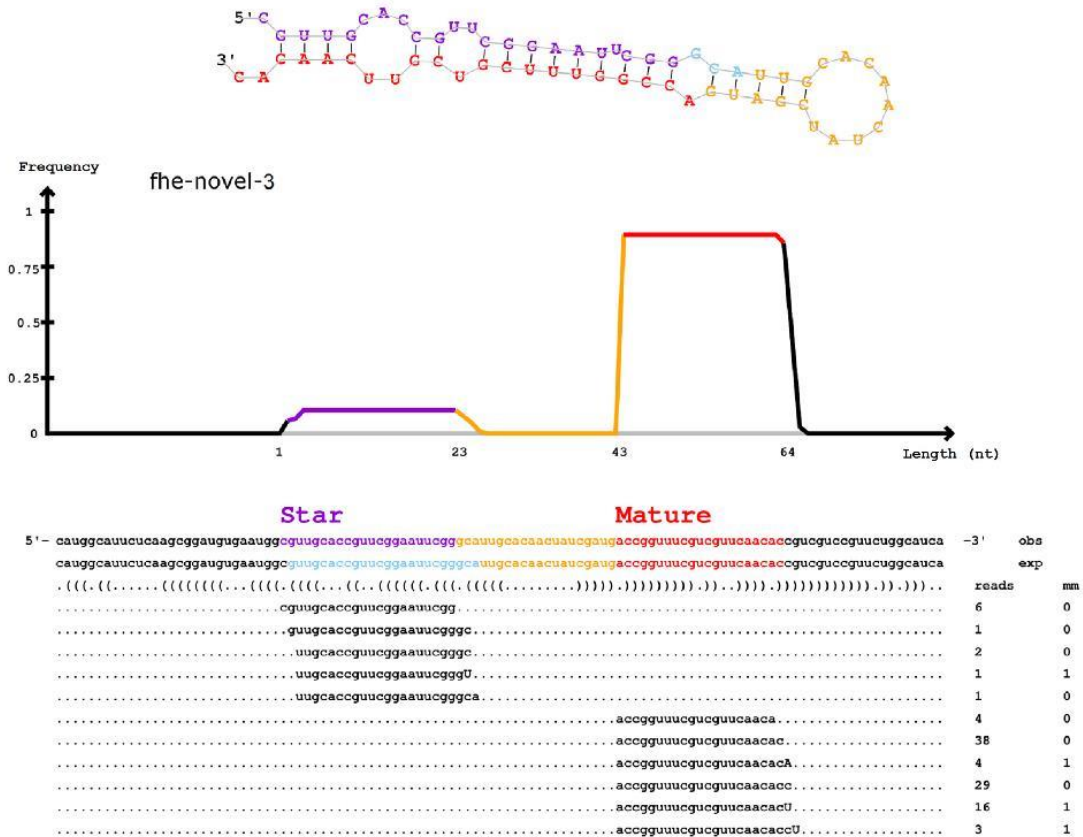


Fig. 2. Novel *Fasciola hepatica* micro (mi)RNA predictions. Hairpin structure and mapping of one of the novel miRNAs (fhe-novel-3) to a locally generated assembly of the publicly available liver fluke genomic data; counts and mismatches for each read are indicated. Images corresponding to all predicted novel miRNAs can be found in Supplementary Fig. S2.

Table 2

Annotation of novel micro RNAs of *Fasciola hepatica* predicted with miRDeep2.

Provisional ID	Mature sequence	Strand	Count	Star sequence	Strand	Count	Genome PRJNA179522 ^a	Genome PRJEB6687 ^a
fhe-novel-1	CAUUAUAUAAGAUAUAGGUCUCU	3p	278	CGCCUGAAUCUUAUAUAAG	5p	1	Cont1047:150722..150789:-	scaffold651:176763..176696:-
fhe-novel-2	GUAAGACGAUCGUAUUGACG	3p	134	UGACUGCUAUCGUAGGGUCUGU	5p	41	Cont1698:19871..19950:+	scaffold422:255752..255674:-
fhe-novel-3	ACCGGUUUCGUCGUUAACAC	3p	94	CGUUGCACCGUUCGGAAUUCGG	5p	11	Cont3792:23865..23927:+	scaffold1311:338595..338657:+
fhe-novel-4	UUCCGGCAGCUUAGUACAGCU	3p	52	UACCCAGUUGAGGAAGUUCGCG	5p	1	Cont5373:19137..19207:+	scaffold3850:1867..1934:+
fhe-novel-5	UAGAGUACCGUAGAUUUG	3p	51	UAAGAUCUACA AUUCUGAAC	5p	2	Cont7415:7171..7255:-	Scaffold7417:3262..3346:-

^a Position in genome.

Little is known of the function of mir-71, but interestingly it has been found highly expressed in schistosomules and adults of *S. japonicum* (Hao et al., 2010), and in the cestodes *Taenia saginata* (Ai et al., 2012) and *E. granulosus* (Bai et al., 2014). In the nematode *C. elegans*, miR-71 has been involved in stress responses (Zhang et al., 2011) and mediating the effects of germ cell loss on life span (Boulas and Horvitz, 2012). Interestingly, it has been shown that miR-71b expression is down-regulated in planarians upon exposure to irradiation, suggesting a role in maintenance of neoblast function (Lu et al., 2009). Whether this association stands in parasitic flatworms deserves a more detailed study.

The miR-2 family members are also expressed at relatively high levels in *F. hepatica* NEJs. The different miR-2 isoforms usually share the first 10 nucleotides of the functional 3' arm and conse-

quently they might share the same targets. In *Drosophila melanogaster*, three apoptotic genes relevant in neural development were identified as targets, and target predictions in *C. elegans* also suggest a role in neural development (Marco et al., 2012). In contrast to the high conservation within protostomes, in flatworms miR-2 family members showed seed shifts and sequence changes (see Section 3.4).

Other miRNAs found only in protostomes and unfortunately poorly characterised (miR-277, miR-750), were also detected as relatively highly expressed in NEJs. A study of target detection in insects identified several enzymes of the aliphatic amino acid catabolism as putative targets of miR-277, probably acting on starvation or stress conditions (Stark et al., 2003). Most of these enzymes detected as targets in insects are missing in the genomes of the

Please cite this article in press as: Fontenla, S., et al. The miRnome of *Fasciola hepatica* juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms. *Int. J. Parasitol.* (2015), <http://dx.doi.org/10.1016/j.ijpara.2015.06.007>

ARTICLE IN PRESS

S. Fontenla et al./International Journal for Parasitology xxx (2015) xxx–xxx

7

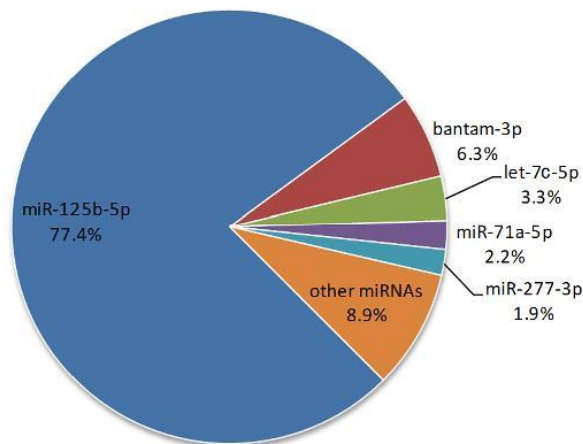


Fig. 3. Skewed expression of conserved micro (mi)RNAs in *Fasciola hepatica* newly excysted juveniles. Relative abundance of known miRNAs in juvenile liver flukes (represented as a percentage of the total conserved miRNA reads), indicate that only five miRNAs represent more than 90% of the miRNAs detected.

schistosomes (*Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium, 2009) suggesting that miR-277 might play a different role in flatworms.

3.3. The *F. hepatica* miRnome confirms flatworm-specific miRNA family losses

Animal evolutionary history shows that after a miRNA emerges in a particular lineage, it is rarely lost in the descendant lineages (Sempere et al., 2007). miRNA acquisition has been related to morphological complexity, exemplified by the gain of 128 miRNAs and

no family losses in primates (Berezikov, 2011). Given their strong conservation, it was suggested that miRNAs are under intense negative selection and consequently considered to be relevant phylogenetic markers (Wheeler et al., 2009; Tarver et al., 2013). According to their position within lophotrochozoans, flatworms are expected to share at least 46 miRNA families (Tarver et al., 2013), but recent studies have described that several miRNA families are missing, especially in the parasitic lineages (Fromm et al., 2013). In agreement with this study, our extended dataset of the *F. hepatica* miRnome shows that only half of the families expected for a lophotrochozoan are present (23 families). Very recently a similar reduction has been confirmed in cestodes (Bai et al., 2014).

To further characterise the possible diminution of the flatworm miRnome, we compared the available miRNA complements of selected species representing the three major metazoan lineages (deuterostomes, ecdysozoans and lophotrochozoans) with those found in platyhelminthes. Consistent with previously published data, we found that eight families expressed in all bilateral metazoans are missing in all flatworms (Fig. 5). The simplification of the miRnome is even more extensive in the parasitic neodermatans, since seven additional ancient miRNAs are absent, while those are still present in free-living planarians. Further reductions were found in cestodes and trematodes that share the absence of miR-278, miR-184, miR-76, miR-1175 and miR-2001. Besides these reductions, we found evidence of the presence in *F. hepatica* of miR-96, miR-281, miR-745 (member of the miR-22 seed family) originally reported as missing in trematodes (Fromm et al., 2013) and miR-750 previously reported only in planarians (Lu et al., 2009). In addition, miR-3487 originally reported as exclusive to schistosomes was also detected in *F. hepatica* and might represent an innovation which has emerged in trematodes. Mir-3479, also originally described in schistosomes and detected in *F. hepatica*, was recently reported in *E. granulosus* (Bai et al., 2014). From what is known, this is the first miRNA family exclusively shared between trematodes and cestodes. Further innovations are represented by

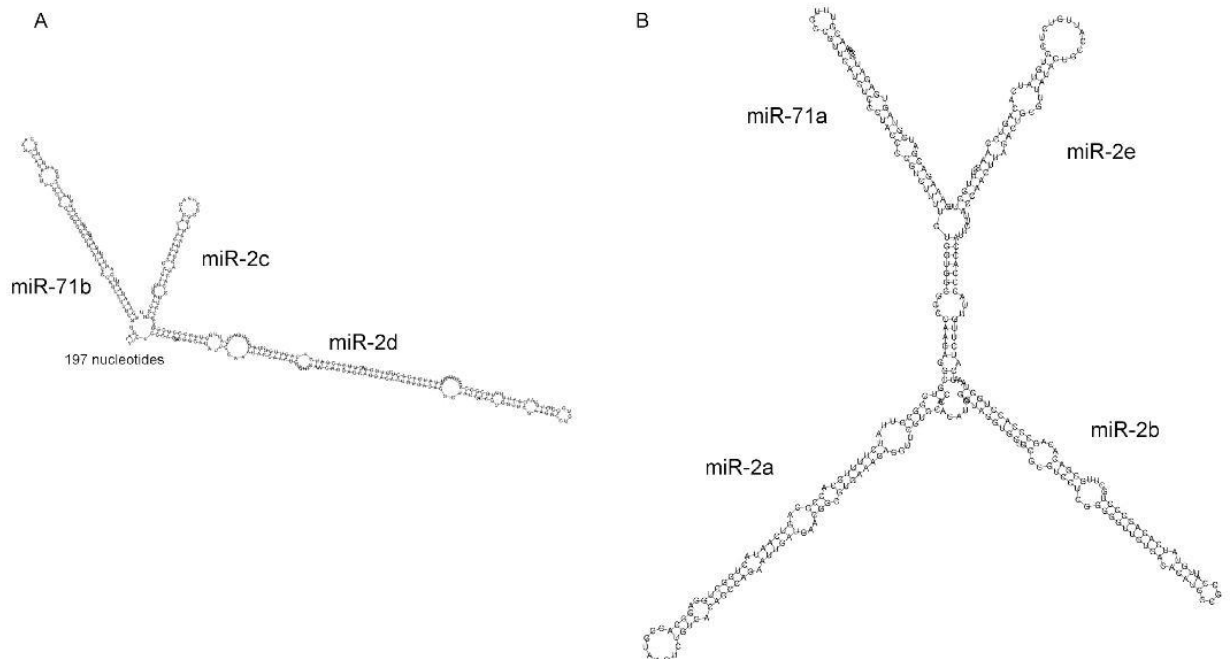


Fig. 4. Cluster structure of miR-71 and miR-2 in *Fasciola hepatica*. miR-71 and miR-2 family members were mapped to the genome assembly. Secondary structure prediction of the region was performed with RNAfold in the Vienna package (www.tbi.univie.ac.at/RNA/RNAfold.html). The two identified clusters containing (A) miR-71b, miR-2c and miR-2d and (B) miR-71a, miR-2a, miR-2b and miR-2e are depicted.

Please cite this article in press as: Fontenla, S., et al. The miRnome of *Fasciola hepatica* juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms. Int. J. Parasitol. (2015), <http://dx.doi.org/10.1016/j.ijpara.2015.06.007>

ARTICLE IN PRESS

8

S. Fontenla et al. / International Journal for Parasitology xxx (2015) xxx–xxx

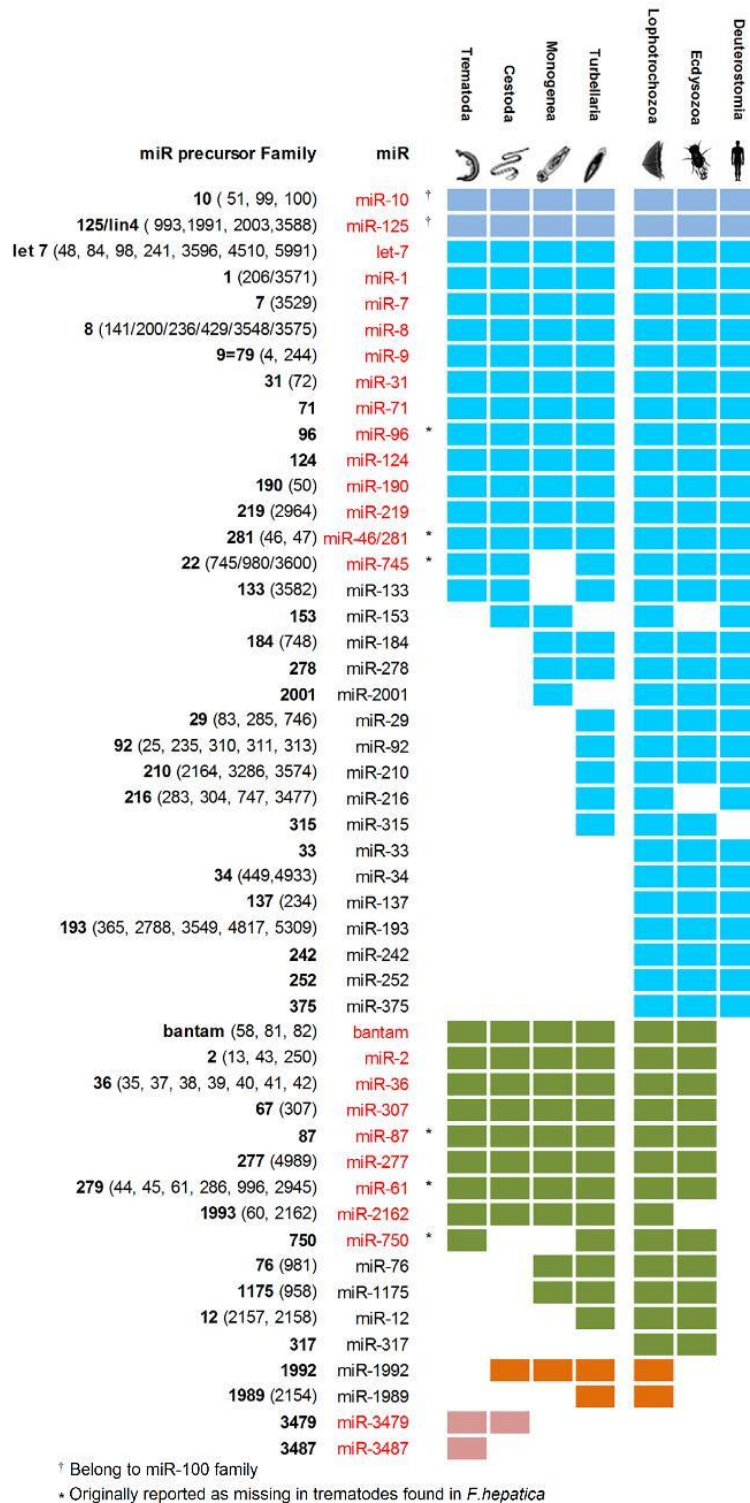


Fig. 5. Micro(mi)RNA family conservation in flatworms compared with other metazoans. The *Fasciola hepatica* dataset was compared with those present in other lineages and included in the trematode dataset. miRNAs newly excysted juveniles (this work) are indicated (red (light grey) lettering). Gains and losses of miRNA families in the different lineages are indicated. Losses are depicted as empty (white) blocks while conserved miRNA families are highlighted in shaded blocks (blue (upper, very light grey), conserved in Metazoans; light blue (light grey), conserved in Bilateria; green (mid, dark grey), conserved in Protostomia; orange (black), conserved in Lophotrochozoa; pink (lower, light grey), conserved in Trematoda and Cestoda). Conservation inferences are taken from the literature (Wheeler et al., 2009; Tarver et al., 2013) and were confirmed by database searches against miRBASE and mirOrtho. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Fontenla, S., et al. The miRnome of *Fasciola hepatica* juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms. Int. J. Parasitol. (2015), <http://dx.doi.org/10.1016/j.ijpara.2015.06.007>

ARTICLE IN PRESS

S. Fontenla et al./International Journal for Parasitology xxx (2015) xxx-xxx

9

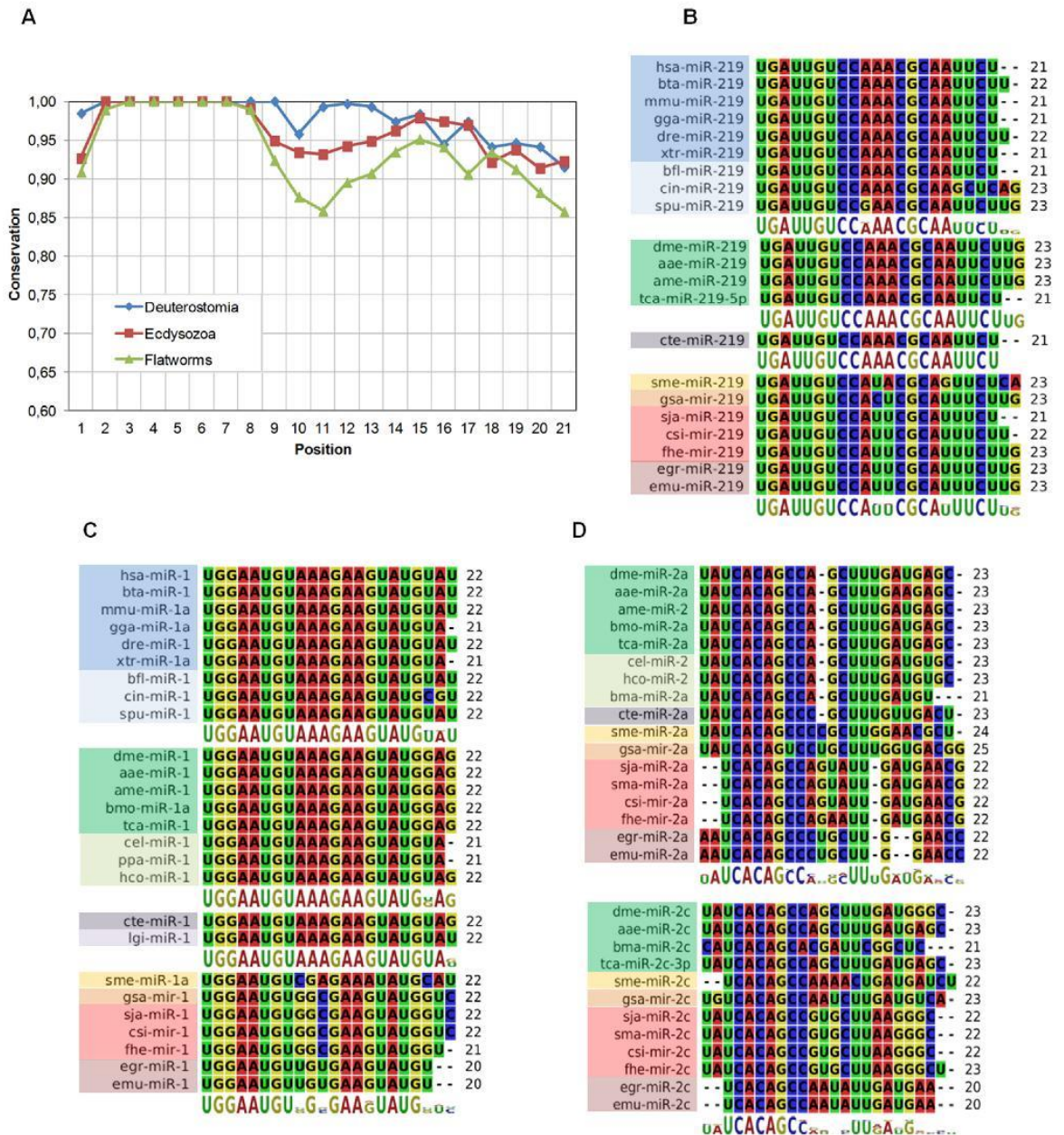


Fig. 6. Micro (mi)RNA sequence conservation through evolution. Alignments of sequences and consensus logos were generated for the different lineages comparing Deuterostomes (from sea urchin to mammals, colour coded in shades of blue), Ecdysozoans (including arthropods and nematodes, green) Lophotrochozoans (represented by the owl limpet (mollusk) and the annelid *Capitella teleta*, lavender shaded), and flatworms (planaria (yellow), monogenean (orange), trematodes (red) and cestodes (purple)). (A) Average sequence conservation by position between Deuterostomia, Ecdysozoa and Platyhelminthes was calculated based on scoring changes by position for all the identified miRNAs. Less conservation of Platyhelminthes outside the seed region (2–8 nt) is evident. (B) miR-219 represents one of the less variable miRNAs found across different taxa, with flatworms being the more divergent. Within flatworms, different substitutions allow differentiating flatworm lineages (see Section 3.4 for description). (C) miR-1a families. (D) Seed shifts observed in miR-2 subfamilies of flatworms. miR-2a 3p alignment shows a conserved two base shift of the miRNA sequences in trematodes. The seed shift in miR-2c-3p is in planarians and cestodes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

five novel *F. hepatica*-specific miRNAs detected in this study which are absent in the genomes of other trematodes. These novel miRNAs are expressed at modest levels (according to their low read counts), and they might be involved in the fine modulation of the expression of *F. hepatica*-specific genes or functions. Interestingly, slightly different repertoires are shown between *F. hepatica*,

S. mansoni and *S. japonicum*, with less conserved and more species-specific miRNAs in the blood flukes. This might reflect truly biological variations within trematodes, but could be also be due to under-sampling of the miRNA populations since not all relevant life stages of these parasites have yet been studied. Similarly, novel miRNAs are being described in each flatworm species studied to

Please cite this article in press as: Fontenla, S., et al. The miRnome of *Fasciola hepatica* juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms. Int. J. Parasitol. (2015), <http://dx.doi.org/10.1016/j.ijpara.2015.06.007>

ARTICLE IN PRESS

10

S. Fontenla et al. / International Journal for Parasitology xxx (2015) xxx–xxx

Table 3Annotation of tRNA halves found in *Fasciola hepatica* newly excysted juveniles.

tRNA Family	tRNA-anticodon	Sequence	Fragment	Total count	Counts by cleavage point												
					30	31	32	33	34	35	36–67	36–68	36–69	40–72			
Gly	tRNA ^{Gly} (GCC)	CGCCGGUGGUUACAGUGGUAGAUAUGCUCGCCUCG	5' half	66,887	244	364	52,716	12,218	1,217	128							
	tRNA ^{Gly} (UCC)	GCGAAGCUGGUGUAAUGGUUUAUCAUGGUUGCCU	5' half	944			569	375									
Met	tRNA ^{Met} (CAU)	AGCAGAGUGGCGCAGUGGAAGCGUGCGGGCCC	5' half	12,341		11,039	349	953									
Lys	tRNA ^{Lys} (UUU)	GCCCGAUAGCUCAGUCGUGAGCAUCAGACUU	5' half	3,701			1,555	2,090	56								
Cys	tRNA ^{Cys} (GCA)	GGAGGCAUAGCUCAGUGGUAAGCAUUCAGCUCG	5' half	3,594			2,523	57	1,014								
Gln	tRNA ^{Gln} (CUG)	GGCCCAUGGUGUAGCGGUUAUCAUCUGGACUC	5' half	2,556				1,996	560								
	tRNA ^{Gln} (UUG)	GGUCCCAUGGUGUAGCGGUUAUCAUCUGGACUC	5' half	1,313			279	1,034									
Tyr	tRNA ^{Tyr} (GUA)	CCGACGAUAGCUCAGUUGGUAAGCGGAGGACU	5' half	2,148			959	1,189									
Glu	tRNA ^{Glu} (CUC)	CACCCAGGCGGCCGGUUCGACUCCCGCGCGG	3' half	1,908								325	500	526	557		
	tRNA ^{Glu} (CUC)	UCCCGGUGGUCUAGCGGUUAGGAUUCUGGUCU	5' half	383			73	95	215								
Pro	tRNA ^{Pro} (AGG)	GGCCUGGUGGUCUAGCGGUUAUGAUUCGCUU	5' half	1,385				1,385									
	tRNA ^{Pro} (CGG)	GGCUCAGUGGUCUAGCGGUUAUGAUUCGCUU	5' half	302		50	252										
Ala	tRNA ^{Pro} (UGG)	GGCCUAGUGGUCUAGCGGUUAUGAUUCGCUU	5' half	165		92	73										
	tRNA ^{Ala} (CGC)	GGGGAUGUAGCUCAGUGGUAAGCGCCUUGCUUC	5' half	774			344	430									
	tRNA ^{Ala} (AGC)	GGGGUACUAGCUCAAUGGUAAGCGCCUUGCUUC	5' half	59				59									
Asp	tRNA ^{Asp} (GUC)	UCCUCGGUAGUUAUGUGGUAUAUCUCGCGCU	5' half	471			184	287									
	tRNA ^{Asp} (GUC)	CACGUGGAAGGCCCGGUUCCGAUUCGCGCGCGG	3' half	111												111	
Leu	tRNA ^{Leu} (UAA)	GGUAGGUUGCCGAGUGGUUAAGCGCGGUGAC	5' half	104			104										

date (as included here), but those might not always reflect specific new adaptations. Some of those might correspond to shared innovations that were undetected in previous studies of the related taxa. Searching for homologues in related species' genomes when available (as we did here) might end up being a necessary and/or recommended step in novel miRNA description (Thomson et al., 2014). In any case, while miRNA predictions can be made based on genomes, experimental detection of the actual miRNAs is necessary for validation.

While free-living trematodes are devoid of 10 conserved miRNA families, in monogeneans 18 families are missing, whereas 21 families have not yet been detected in cestodes and 22 in trematodes. The remarkable progressive decrease in the conserved miRNA repertoire in flatworms could be associated with the parasitic way of life. Interestingly, this decrease correlates with variations in genome size and gene counts, with the planarians displaying the largest genomes – more than 30,000 predicted genes – trematodes with smaller genomes (1000–300 Mb) and close to 15,000 genes, and cestodes and monogeneans with the smallest genomes (100–150 Mb) and 15,000 predicted genes. Furthermore, a recent study has identified a substantial reduction in the core bilaterian gene complement, including the homeodomain-containing genes, loss of the *piwi* and *vasa* genes that are considered essential for animal development (Skinner et al., 2014), loss of functional fatty acid biosynthesis pathways and the absence of peroxisomes as shared genomic features for the ecto- and endoparasitic lineages (Hahn et al., 2014). Whether these gene losses precede or are the consequence of the absence of the corresponding modulating miRNAs is still not clear, raising novel questions with possibly wide evolutionary implications. On the other hand, novel miRNAs are consistently being described in flatworms but it is still difficult to assess whether novel families shared across flatworms might have emerged to compensate for those lost, or if novelties emerged repeatedly in different flatworm lineages to drive diverse functions in the parasites' biology.

3.4. Low sequence conservation in ancient miRNA suggests a high mutational rate in parasitic flatworms

miRNAs are among the most evolutionarily conserved sequences with low rates of mutations. A 3.5% substitution rate was calculated for 93 miRNA genes across 14 nephrozoan taxa,

which makes miRNAs one of the most conserved sequences throughout evolution (Wheeler et al., 2009). Within mature miRNAs the 'seed' region that spans from positions 2–8 is the main determinant of target selectivity and consequently extremely conserved. A second highly preserved region involves residues 13–16 of the mature miRNA (Wheeler et al., 2009). When analysing conserved miRNA families present in *F. hepatica*, we observed several differences in the sequence outside the seed regions. Interestingly, this sequence divergence was not restricted to *F. hepatica* but seems to be characteristic of flatworms in general. While sequence conservation is notoriously high even between distant deuterostomes, ecdysozoans and lophotrochozoans, platyhelminthes show marked and in general lineage-specific variations, with strict conservation restricted to the seed region (Fig. 6A). For example, an almost complete conservation of miR-219 is observed from lophotrochozoans to mammals, while in flatworms different substitutions at positions 11, 12, 17, 22 and 23 allow us to differentiate planarians from monogeneans and these from cestodes and trematodes that share the miR-219 sequence (Fig. 6B). A second and more drastic example is provided by miR-1a, with almost complete conservation observed in deuterostomes from sea urchins to humans with single variations in position 20 of *Ciona intestinalis*; the same position allows differentiation of insects from nematodes within ecdysozoans, while position 22 discriminates between the annelid *C. teleta* and the mollusk *L. gigantea* (owl limpet) (Fig. 6C). Platyhelminth sequences, on the other hand, presented multiple differences that are lineage-specific. All flatworm sequences showed one A to G transition at position 10, indicating that this substitution probably first appeared in a common ancestor of these organisms. Besides this change, the planarian *S. mediterranea* presented three other changes, a transversion at position 9 and transitions at positions 15 and 20. Cestode sequences are shorter and with transversions at positions 9 and 11; monogeneans and trematodes shared three transversions at positions 11, 20 and 21, and two transitions at positions 9 and 22 (Fig. 6C). Similar examples can be found in several miRNA families (Supplementary Fig. S3).

Sequence variation was accompanied by seed shifts in the case of the miR-2 family members. While a +2 base shift is observed in miR-2c of planarians and cestodes, a similar change is detected in the miR-2a of trematodes (Fig. 6D). Interestingly, miR-2b of flatworms and ecdysozoans share the seed, but the other protostomes

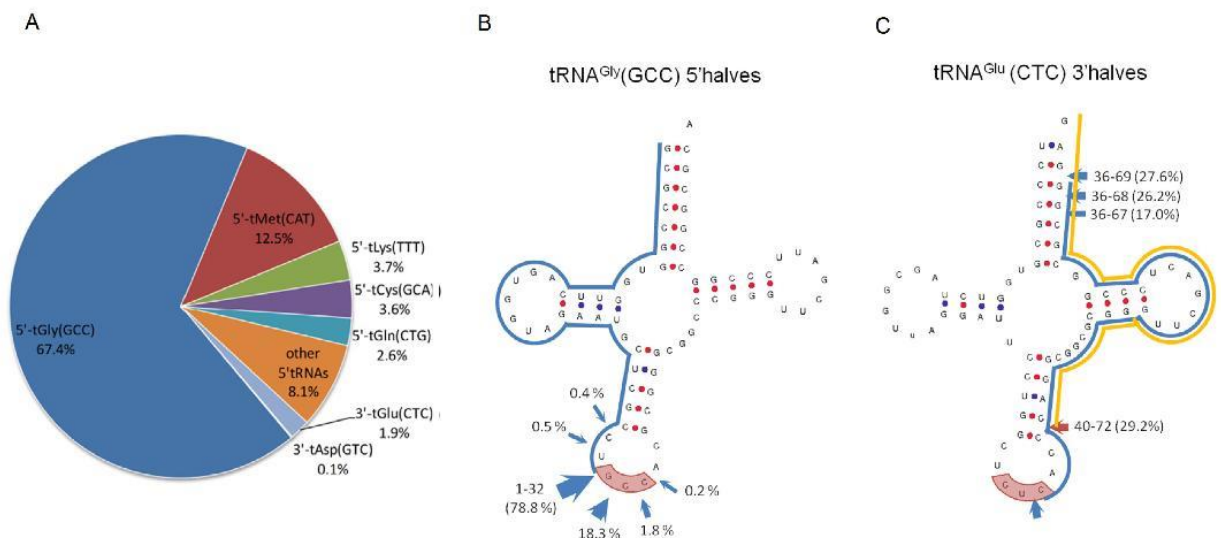


Fig. 7. Fragmented tRNAs detected in the small RNAome of *Fasciola hepatica* newly excysted juveniles. (A) The proportional distribution of the RNA halves detected is indicated. (B) Approximately 98% of the tRNA halves detected were 5' halves with tRNA^{Gly} (GCC)-derived being the most frequent. (C) Within the 3' fragments, the most abundant is tRNA^{Glu} (CTC), presenting two types of fragments, one generated by a single cleavage at the 3' base of the anticodon loop (yellow (clear)) and the remaining by two cleavages (blue (dark)). Variation at the cleavage site is indicated by arrows. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

show a +2 shift (not shown). Despite this generalised picture of miRNA sequence variation, a few families such as miR-219 or miR-124 have very low sequence variation while others are more prone to change (Supplementary Fig. S3).

It is worth highlighting that further minor variants might also exist. Within the *F. hepatica* dataset for a given miRNA, the most abundant sequence was taken as the canonical but variants at 5' and 3' with lower counts or shifts to the 3' or 5' were detected (see Section 2.4). These variations might represent diverse isomiRs that might be biologically relevant, since they eventually provide fine modulation at target recognition (Morin et al., 2008; Wang et al., 2010). A typical example of this is the let-7 family. We found a canonical let-7 that is expressed at a very low level. However, a highly expressed variant with the same seed and homologous to a putative miRNA described as novel in *S. japonicum* (Sja-novel-137) (Hao et al., 2010) was found (let-7c). Four different let-7 variants have emerged in free-living planarians, but diversification in the let-7 complement in trematodes has not yet been described. Despite this generalised picture of miRNA sequence variation, a few families such as miR-219 or miR-124 have very low sequence variation, while others are more prone to change (Supplementary Fig. S3).

The sequence variations observed to date demand a more detailed study of the residue change rate in flatworm miRNAs. The high variation rate observed in the flatworm sequences might affect the targets identified by these miRNAs. In any case, if they are coevolving, a high substitution rate seems to be characteristic of flatworms. The origins and consequences of these variations are interesting questions that emerge from this study.

3.5. The 5' halves of some tRNAs are present in the small RNA population

A second peak of abundance at 32 nts was observed in the NEJ small RNA dataset (see Supplementary Fig. S1). This size corresponds approximately to the 30 nt average size of the piwi-interacting RNAs (piRNA). piRNAs are related to post-transcriptional silencing of retrotransposons and repetitive sequences

in the germline (Weick and Miska, 2014). piRNAs are highly expressed in neoblasts of the free-living flatworm *S. mediterranea* (Palakodeti et al., 2008; Friedländer et al., 2009). While neoblasts do exist in parasitic flatworms, piRNAs have not been identified in cestodes or trematodes. Consistent with this absence, the canonical PIWI protein is missing in these organisms (Zheng, 2012; Tsai et al., 2013), posing the question of how these organisms maintain genomic integrity (Skinner et al., 2014). It has been proposed that other small RNAs may carry out these key roles in flatworms and in some nematodes in which piRNAs have not been identified (Britton et al., 2014). In agreement with this, we did not find sequences that might be consistently assigned to piRNAs in our dataset.

However some of these longer reads were flagged as being related to tRNA sequences. In order to confirm this, we mapped the 25–35 nt fraction to the *F. hepatica* partial genome assembly, and compared this dataset by stringent BLAST searches against all the predicted tRNAs. A total of 99,014 reads matched tRNAs, surprisingly corresponding almost exclusively to tRNA halves. A strong skew towards the 5' half of tRNAs was observed, with only 2.04% of the tRNA-derived sequences corresponding to 3' tRNA halves (Table 3). Notably, there is a marked preference for the tRNA^{Gly} (68.3%), with a major fraction that accounted by the transfer with the GCC anticodon (67.3%). The second most abundant was tRNA^{Met}, representing more than 12% of the identified tRNAs (Fig. 7). The two halves were found only for two tRNAs (Asp and Glu), and interestingly the tRNA^{Glu} 3' half had a higher count than the 5' half. tRNA halves are tRNA-derived fragments of 30–35 nts long generated by the cleavage of tRNAs at anticodon loops and have been claimed to be post-transcriptional regulators in response to stress (García-Silva et al., 2013). tRNA fragmentation in response to stress has been reported in a wide variety of organisms ranging from bacteria to humans (Thompson and Parker, 2009). tRNA halves have been reported to be major post-transcriptional regulators in *Trypanosoma cruzi*, an organism where canonical small RNA regulation is hampered due to the absence of a canonical RNA pathway (García-Silva et al., 2010). In general, tRNAs are processed by different RNases that are usually bound to organelle membranes and separated from cytoplasmic RNA.

ARTICLE IN PRESS

12

S. Fontenla et al. / International Journal for Parasitology xxx (2015) xxx–xxx

During oxidative stress these nucleases move into the cytoplasm to process tRNAs, and the cleaved halves could be used by the silencing machinery of the cell (Kumar et al., 2014). More recently, in *T. cruzi* the argonaute family protein TcPIWI-tryp has been shown to immunoprecipitate with tRNA halves and was claimed to be involved in the interference mechanism (García-Silva et al., 2014a). In these organisms the tRNA half containing complexes localise mostly to intracellular vesicles called reservosomes that are involved in parasite communication via the endocytic/exocytic pathway. Interestingly, these vesicles might take part in host cell susceptibility to parasitic infection, since changes in gene expression were found in HeLa cells exposed to parasite extracellular vesicles (García-Silva et al., 2014b). Our finding of tRNA-derived small RNAs in the juvenile stage opens new and interesting questions. Assuming that tRNA-derived sequences originated in response to stress as in other organisms, it is still not known if this is triggered by culture conditions or if it happens naturally as a response by the parasite to the initial contact with the host, and consequently plays a role in invasion. Further work is needed in order to confirm if they are present in other parasitic stages and if they can be modulated experimentally by culture conditions.

3.6. Conclusion

We found that a diverse set of small RNAs is produced by the invading NEJ of *F. hepatica*. These molecules essentially fall into two classes, one represented by miRNAs and a secondary group of larger (32–33 nts) tRNA-derived sequences. Some ancient miRNAs conserved in metazoans and protostomes are very highly expressed by the NEJs. Our results also confirm that no less than eight conserved miRNA families have been lost in all flatworms, while more than 20 are missing in all parasitic trematodes analysed to date. Besides this reduction, the miRnome is enriched by novel flatworm and *F. hepatica*-specific miRNAs generally expressed at lower levels. Interestingly, sequence conservation in flatworm miRNAs was lower than in other lineages, suggesting a high mutational rate in parasitic flatworms. While no piRNA was found, a distinctive set of tRNA-derived sequences represented mostly by 5' halves of selected tRNAs was detected. The singularities found in the small RNA repertoire of flatworms suggest that more effort is needed to understand their role in parasite evolution and adaptation.

Acknowledgments

We would like to thank Drs. Gabriel Rinaldi and Paul Brindley for valuable comments, discussions and suggestions. The work was partially supported by the Maestría de Bioinformática, Programa de Desarrollo de Ciencias Básicas (PEDECIBA), Uruguay, and Comisión Sectorial de Investigación Científica (CSIC-UdelaR), Uruguay. SF is recipient of a scholarship from the Sistema Nacional de Becas from the Agencia Nacional de Investigación e Innovación (SNB-ANII), Uruguay; SF and NDO are postgraduate students in PEDECIBA (Bioinformatics and Biology) and received support from this program. PS and JFT were supported by the Sistema Nacional de Investigadores from ANII (SNI-ANII). MSL was supported by the Consejo Superior de Investigación Científica (CSIC), Spain. No other specific funding was received for this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ijpara.2015.06.007>.

References

- Ai, L., Xu, M.J., Chen, M.X., Zhang, Y.N., Chen, S.H., Guo, J., Cai, Y.C., Zhou, X.N., Zhu, X. Q., Chen, J.X., 2012. Characterization of microRNAs in *Taenia saginata* of zoonotic significance by Solexa deep sequencing and bioinformatics analysis. *Parasitol. Res.* 110, 2373–2378.
- Bai, Y., Zhang, Z., Jin, L., Kang, H., Zhu, Y., Zhang, L., Li, X., Ma, F., Zhao, L., Shi, B., Li, J., McManus, D.P., Zhang, W., Wang, S., 2014. Genome-wide sequencing of small RNAs reveals a tissue-specific loss of conserved microRNA families in *Echinococcus granulosus*. *BMC Genomics* 15, 736.
- Bartel, D.P., 2009. MicroRNA target recognition and regulatory functions. *Cell* 136, 215–233.
- Berezikov, E., 2011. Evolution of microRNA diversity and regulation in animals. *Nat. Rev. Genet.* 12, 846–860.
- Boulan, L., Martín, D., Milán, M., 2013. Bantam miRNA promotes systemic growth by connecting insulin signaling and ecdysone production. *Curr. Biol.* 23, 473–478.
- Boulias, K., Horvitz, H.R., 2012. The *C. elegans* microRNA mir-71 acts in neurons to promote germline-mediated longevity through regulation of DAF-16/FOXO. *Cell Metab.* 15, 439–450.
- Brennan, G.P., Fairweather, I., Trudgett, A., Hoey, E., McCoy, McConville, M., Meaney, M., Robinson, M., McFerran, N., Ryan, L., Lanusse, C., Mottier, L., Alvarez, L., Solana, H., Virkel, G., Brophy, P.M., 2007. Understanding triclabendazole resistance. *Exp. Mol. Pathol.* 82, 104–109.
- Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., Cohen, S.M., 2003. Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell* 113, 25–36.
- Britton, C., Winter, A.D., Gillan, V., Devaney, E., 2014. MicroRNAs of parasitic helminths – identification, characterization and potential as drug targets. *Int. J. Parasitol. Drugs Drug Resist.* 4, 85–94.
- Burge, S.W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E.P., Eddy, S.R., Gardner, P.P., Bateman, A., 2013. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* 41, D226–D232.
- Cai, P., Hou, N., Piao, X., Liu, S., Liu, H., Yang, F., Wang, J., Jin, Q., Wang, H., Chen, Q., 2011. Profiles of small non-coding RNAs in *Schistosoma japonicum* during development. *PLoS Negl. Trop. Dis.* 5, e1256.
- Cai, P., Piao, X., Hao, L., Liu, S., Hou, N., Wang, H., Chen, Q., 2013. A deep analysis of the small non-coding RNA population in *Schistosoma japonicum* eggs. *PLoS One* 8, e64003.
- Cancela, M., Acosta, D., Rinaldi, G., Silva, E., Durán, R., Roche, L., Zaha, A., Carmona, C., Tort, J.F., 2008. A distinctive repertoire of cathepsins is expressed by juvenile invasive *Fasciola hepatica*. *Biochimie* 90, 1461–1475.
- Cancela, M., Ruétalo, N., Dell'Oca, N., da Silva, E., Smircich, P., Rinaldi, G., Roche, L., Carmona, C., Alvarez-Valín, F., Zaha, A., Tort, J.F., 2010. Survey of transcripts expressed by the invasive juvenile stage of the liver fluke *Fasciola hepatica*. *BMC Genomics* 11, 227.
- Charlier, J., van der Voort, M., Kenyon, F., Skuce, P., Vercurusse, J., 2014. Chasing helminths and their economic impact on farmed ruminants. *Trends Parasitol.* 30, 361–367.
- Copeland, C.S., Marz, M., Rose, D., Hertel, J., Brindley, P.J., Santana, C.B., Kehr, S., Attolini, C.S., Stadler, P.F., 2009. Homology-based annotation of non-coding RNAs in the genomes of *Schistosoma mansoni* and *Schistosoma japonicum*. *BMC Genomics* 10, 464.
- Cucher, M., Prada, L., Mourglia-Ettlin, G., Dematteis, S., Camicia, F., Asurmendi, S., Rosenzvit, M., 2011. Identification of *Echinococcus granulosus* microRNAs and their expression in different life cycle stages and parasite genotypes. *Int. J. Parasitol.* 41, 439–448.
- Cwiklinski, K., Dalton, J.P., Dufresne, P.J., La Course, J., Williams, D.J., Hodgkinson, J., Paterson, S., 2015. The *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. *Genome Biol.* 16, 71.
- De Souza Gomes, M., Munyappa, M.K., Carvalho, S.G., Guerra-Sá, R., Spillane, C., 2011. Genome-wide identification of novel microRNAs and their target genes in the human parasite *Schistosoma mansoni*. *Genomics* 98, 96–111.
- Eulalio, A., Huntzinger, E., Izaurralde, E., 2008. Getting to the root of miRNA-mediated gene silencing. *Cell* 132, 9–14.
- Fischer, S.E.J., 2010. Small RNA-mediated gene silencing pathways in *C. elegans*. *Int. J. Biochem. Cell Biol.* 42, 1306–1315.
- Friedländer, M.R., Adamidi, C., Han, T., Lebedeva, S., Isenbarger, T.A., Hirst, M., Marra, M., Nusbaum, C., Lee, W.L., Jenkin, J.C., Sánchez Alvarado, A., Kim, J.K., Rajewsky, N., 2009. High-resolution profiling and discovery of planarian small RNAs. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11546–11551.
- Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., Rajewsky, N., 2012. MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52.
- Fromm, B., Worren, M.M., Hahn, C., Hovig, E., Bachmann, L., 2013. Substantial loss of conserved and gain of novel microRNA families in flatworms. *Mol. Biol. Evol.* 30, 2619–2628.
- García, H.H., Moro, P.L., Schantz, P.M., 2007. Zoonotic helminth infections of humans: echinococcosis, cysticercosis and fascioliasis. *Curr. Opin. Infect. Dis.* 20, 489–494.
- García-Silva, M.R., Frugier, M., Tosar, J.P., Correa-Dominguez, A., Ronalte-Alves, L., Parodi-Talice, A., Rovira, C., Robello, C., Goldenberg, S., Cayota, A., 2010. A population of tRNA-derived small RNAs is actively produced in *Trypanosoma cruzi* and recruited to specific cytoplasmic granules. *Mol. Biochem. Parasitol.* 171, 64–73.

Please cite this article in press as: Fontenla, S., et al. The miRnome of *Fasciola hepatica* juveniles endorses the existence of a reduced set of highly divergent micro RNAs in parasitic flatworms. *Int. J. Parasitol.* (2015), <http://dx.doi.org/10.1016/j.ijpara.2015.06.007>

ARTICLE IN PRESS

S. Fontenla et al. / International Journal for Parasitology xxx (2015) xxx–xxx

13

- García-Silva, M.R., Cabrera-Cabrera, F., Güida, M.C., Cayota, A., 2013. Novel aspects of tRNA-derived small RNAs with potential impact in infectious diseases. *ABB* 4, 17–25.
- García-Silva, M.R., Cabrera-Cabrera, F., das Neves, R.F.C., Souto-Padrón, T., de Souza, W., Cayota, A., 2014. Gene expression changes induced by *Trypanosoma cruzi* shed microvesicles in mammalian host cells: relevance of tRNA-derived halves. *Biomed. Res. Int.* 2014, 305239.
- García-Silva, M.R., Sanguinetti, J., Cabrera-Cabrera, F., Franzén, O., Cayota, A., 2014. A particular set of small non-coding RNAs is bound to the distinctive Argonaute protein of *Trypanosoma cruzi*: insights from RNA-interference deficient organisms. *Gene* 538, 379–384.
- Gerlach, D., Kriventseva, E.V., Rahman, N., Vejnar, C.E., Zdobnov, E.M., 2009. MiROrtho: computational survey of microRNA genes. *Nucleic Acids Res.* 37, D111–D117.
- Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., Zamore, P.D., 2010. Sorting of *Drosophila* small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA* 16, 43–56.
- Grishok, A., 2005. RNAi mechanisms in *Caenorhabditis elegans*. *FEBS Lett.* 579, 5932–5939.
- Hahn, C., Fromm, B., Bachmann, L., 2014. Comparative genomics of flatworms (Platyhelminthes) reveals shared genomic features of ecto- and endoparasitic Neodermata. *Genome Biol. Evol.* 6, 1105–1117.
- Hao, L., Cai, P., Jiang, N., Wang, H., Chen, Q., 2010. Identification and characterization of microRNAs and endogenous siRNAs in *Schistosoma japonicum*. *BMC Genomics* 11, 55.
- Hernández-González, A., Valero, M.L., del Pino, M.S., Oleaga, A., Siles-Lucas, M., 2010. Proteomic analysis of *in vitro* newly excysted juveniles from *Fasciola hepatica*. *Mol. Biochem. Parasitol.* 172, 121–128.
- Hertel, J., Bartschat, S., Wintsche, A., Otto, C., Stadler, P.F., 2012. Evolution of the let-7 microRNA family. *RNA Biol.* 9, 231–241.
- Huang, J., Hao, P., Chen, H., Hu, W., Yan, Q., Liu, F., Han, Z.G., 2009. Genome-wide identification of *Schistosoma japonicum* microRNAs using a deep-sequencing approach. *PLoS One* 4, e8206.
- Jamalkandi, S.A., Masoudi-Nejad, A., 2009. Reconstruction of *Arabidopsis thaliana* fully integrated small RNA pathway. *Funct. Integr. Genomics* 9, 419–432.
- Jin, X., Lu, L., Su, H., Lou, Z., Wang, F., Zheng, Y., Xu, G.T., 2013. Comparative analysis of known miRNAs across platyhelminths. *FEBS J.* 280, 3944–3951.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
- Keiser, J., Utzinger, J., 2009. Food-borne trematodiasis. *Clin. Microbiol. Rev.* 22, 466–483.
- Kozomara, A., Griffiths-Jones, S., 2014. MiRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73.
- Kumar, P., Anaya, J., Mudunuri, S.B., Dutta, A., 2014. Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *BMC Biol.* 12, 78.
- Lowe, T.M., Eddy, S.R., 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964.
- Lu, Y., Smielewska, M., Palakodeti, D., Lovci, M.T., Aigner, S., Yeo, G.W., Graveley, B. R., 2009. Deep sequencing identifies new and regulated microRNAs in *Schmidtea mediterranea*. *RNA* 15, 1483–1491.
- Mangiola, S., Young, N.D., Korhonen, P., Mondal, A., Scheerlinck, J.P., Sternberg, P.W., Cantacessi, C., Hall, R.S., Jex, A.R., Gasser, R.B., 2013. Getting the most out of parasitic helminth transcriptomes using HelmiDB: implications for biology and biotechnology. *Biotechnol. Adv.* 31, 1109–1119.
- Manzano-Román, R., Siles-Lucas, M., 2012. MicroRNAs in parasitic diseases: potential for diagnosis and targeting. *Mol. Biochem. Parasitol.* 186, 81–86.
- Marco, A., Hooks, K.B., Griffiths-Jones, S., 2012. Evolution and function of the extended miR-2 microRNA family. *RNA Biol.* 9, 242–248.
- Marcos, L.A., Terashima, A., Gotuzzo, E., 2008. Update on hepatobiliary flukes: fascioliasis, opisthorchiasis and clonorchiasis. *Curr. Opin. Infect. Dis.* 21, 523–530.
- Martin, J., Rosa, B.A., Ozersky, P., Hallsworth-Pepin, K., Zhang, X., Bhonagiri-Palsikar, V., Tyagi, R., Wang, Q., Choi, Y., Gao, X., McNulty, S., Brindley, P.J., Mitreva, M., 2015. Helminth.net: expansions to Nematode.net and an introduction to Trematode.net. *Nucleic Acids Res.* 43, 698–706.
- Mas-Coma, S., Bargues, M.D., Valero, M.A., 2005. Fascioliasis and other plant-borne trematode zoonoses. *Int. J. Parasitol.* 35, 1255–1278.
- Mituyama, T., Yamada, K., Hattori, E., Okida, H., Ono, Y., Terai, G., Yoshizawa, A., Komori, T., Asai, K., 2009. The Functional RNA Database 3.0: databases to support mining and annotation of functional RNAs. *Nucleic Acids Res.* 37, D89–D92.
- Morin, R.D., Connor, M.D.O., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Eaves, C.J., Marra, M.A., 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* 18, 610–621.
- Palakodeti, D., Smielewska, M., Graveley, B.R., 2006. MicroRNAs from the Planarian *Schmidtea mediterranea*: a model system for stem cell biology. *RNA* 12, 1640–1649.
- Palakodeti, D., Smielewska, M., Lu, Y.C., Yeo, G.W., Graveley, B.R., 2008. The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *RNA* 14, 1174–1186.
- Schistosoma japonicum* Genome Sequencing and Functional Analysis Consortium, 2009. The *Schistosoma japonicum* genome reveals features of host–parasite interplay. *Nature* 460, 345–351.
- Sempere, L.F., Martínez, P., Cole, C., Bagaña, J., Peterson, K.J., 2007. Phylogenetic distribution of microRNAs supports the basal position of acol flatworms and the polyphyly of Platyhelminthes. *Evol. Dev.* 9, 409–415.
- Simões, M.C., Lee, J., Djikeng, A., Cerqueira, G.C., Zerlotini, A., da Silva-Pereira, R.A., Dalby, A.R., LoVerde, P., El-Sayed, N.M., Oliveira, G., 2011. Identification of *Schistosoma mansoni* microRNAs. *BMC Genomics* 12, 47.
- Skinner, D.E., Rinaldi, G., Koziol, U., Brehm, K., Brindley, P.J., 2014. How might flukes and tapeworms maintain genome integrity without a canonical piRNA pathway? *Trends Parasitol.* 30, 123–129.
- Sokol, N.S., 2012. Small temporal RNAs in animal development. *Curr. Opin. Genet. Dev.* 22, 368–373.
- Stark, A., Brennecke, J., Russell, R.B., Cohen, S.M., 2003. Identification of *Drosophila* microRNA targets. *PLoS Biol.* 1, E60.
- Sun, Y.M., Lin, K.Y., Chen, Y.Q., 2013. Diverse functions of miR-125 family in different cell contexts. *J. Hematol. Oncol.* 6, 6.
- Sun, J., Wang, S., Li, C., Ren, Y., Wang, J., 2014. Novel expression profiles of microRNAs suggest that specific miRNAs regulate gene expression for the sexual maturation of female *Schistosoma japonicum* after pairing. *Parasit. Vectors* 7, 177.
- Tarver, J.E., Sperling, E.A., Nailor, A., Heimberg, A.M., Robinson, J.M., King, B.L., Pisani, D., Donoghue, P.C., Peterson, K.J., 2013. MiRNAs: small genes with big potential in metazoan phylogenetics. *Mol. Biol. Evol.* 30, 2369–2382.
- Thompson, D.M., Parker, R., 2009. Stressing out over tRNA cleavage. *Cell* 138, 215–219.
- Thomson, R.C., Plachetzki, D.C., Mahler, D.L., Moore, B.R., 2014. A critical appraisal of the use of microRNA data in phylogenetics. *Proc. Natl. Acad. Sci. U.S.A.* 111, E3659–E3668.
- Tsai, I.J., Zarowiecki, M., Holroyd, N., Garciaarubio, A., Sanchez-Flores, A., Brooks, K. L., Tracey, A., Bobes, R.J., Frago, G., Sciutto, E., Aslett, M., Beasley, H., Bennett, H.M., Cai, J., Camicia, F., Clark, R., Cucher, M., De Silva, N., Day, T.A., Deplazes, P., Estrada, K., Fernández, C., Holland, P.W., Hou, J., Hu, S., Huckvale, T., Hung, S., Kamenetzky, L., Keane, J.A., Kiss, F., Koziol, U., Lambert, O., Liu, K., Luo, X., Luo, Y., Macchiaroli, N., Nichol, S., Paps, J., Parkinson, J., Pouchkina-Stantcheva, N., Riddiford, N., Rosenzvit, M., Salinas, G., Wasmuth, J.D., Zamanian, M., Zheng, Y., Taenia solium Genome Consortium, Cai, X., Soberón, X., Olson, P.D., Laclette, J.P., Brehm, K., Berriman, M., 2013. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* 496, 57–63.
- Wang, W.C., Lin, F.M., Chang, W.C., Lin, K.Y., Huang, H.D., Lin, N.S., 2009. MiRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinformatics* 10, 328.
- Wang, Z., Xue, X., Sun, J., Luo, R., Xu, X., Jiang, Y., Zhang, Q., Pan, W., 2010. An “in-depth” description of the small non-coding RNA population of *Schistosoma japonicum* schistosomulum. *PLoS Negl. Trop. Dis.* 4, e596.
- Weick, E.M., Miska, E.A., 2014. PiRNAs: from biogenesis to function. *Development* 141, 3458–3471.
- Wheeler, B.M., Heimberg, A.M., Moy, V.N., Sperling, E.A., Holstein, T.W., Heber, S., Peterson, K.J., 2009. The deep evolution of metazoan microRNAs. *Evol. Dev.* 11, 50–68.
- Xu, M.J., Liu, Q., Nisbet, A.J., Cai, X.Q., Yan, C., Lin, R.Q., Yuan, Z.G., Song, H.Q., He, X.H., Zhu, X.Q., 2010. Identification and characterization of microRNAs in *Clonorchis sinensis* of human health significance. *BMC Genomics* 11, 521.
- Xu, M.J., Ai, L., Fu, J.H., Nisbet, A.J., Liu, Q.Y., Chen, M.X., Zhou, D.H., Zhu, X.Q., 2012. Comparative characterization of microRNAs from the liver flukes *Fasciola gigantica* and *F. hepatica*. *PLoS One* 7, e53387.
- Zerlotini, A., Aguiar, E.R.G.R., Yu, F., Xu, H., Li, Y., Young, N.D., Gasser, R.B., Protasio, A. V., Berriman, M., Roos, D.S., Kissinger, J.C., Oliveira, G., 2013. SchistoDB: an updated genome resource for the three key schistosomes of humans. *Nucleic Acids Res.* 41, D728–D731.
- Zhang, X., Zabinsky, R., Teng, Y., Cui, M., Han, M., 2011. MicroRNAs play critical roles in the survival and recovery of *Caenorhabditis elegans* from starvation-induced L1 diapause. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17997–18002.
- Zheng, Y., 2012. Phylogenetic analysis of the Argonaute protein family in platyhelminths. *Mol. Phylogenet. Evol.* 66, 1050–1054.
- Zhu, L., Liu, J., Cheng, G., 2014. Role of microRNAs in schistosomes and schistosomiasis. *Front. Cell Infect. Microbiol.* 4, 165.

DISCUSIÓN

Las parasitosis ocasionadas por helmintos en general tienen un fuerte impacto en la salud pública por la gran cantidad de personas afectadas por estas enfermedades. Las trematodiasis juegan un gran papel dentro de estas enfermedades afectando a más de 200 millones de personas principalmente en países en vías de desarrollo. El esquema actual de tratamiento no es sustentable y es imperioso el desarrollo de nuevos métodos para combatir estas enfermedades. En la búsqueda de nuevas estrategias de control muchos investigadores han dirigido sus esfuerzos a la identificación de nuevos genes y nuevas proteínas a través de la secuenciación de los genomas y transcriptomas de estos parásitos. La minería en datos públicos con diferentes herramientas bioinformáticas es una forma rápida y económica para identificar *in-silico* posibles blancos terapéuticos para el desarrollo de nuevas drogas o vacunas.

Identificación de proteínas de la vía de miARNs en platelmintos

Dentro de este espectro, un aspecto novedoso está dado por el papel de los ARN pequeños reguladores. La secuenciación de ARN pequeños ha aportado al conocimiento de un nuevo mecanismo de regulación génica aun poco conocido en platelmintos pero potencialmente esencial en la regulación de su ciclo biológico.

Las vías de generación de ARN pequeños son esenciales para la regulación de la expresión génica, diferenciación celular, desarrollo e integridad del genoma. Si bien se sabe que los platelmintos poseen vías de generación de ARN pequeños funcionales se conoce muy poco sobre los genes participantes con unos pocos reportes en *Schistosoma* (Krautz-peterson and Skelly 2008; Krautz-Peterson et al. 2010; Gomes et al. 2009; Luo et al. 2010), y algunos datos aportados en la descripción de los genomas de cestodos (Tsai et al. 2013).

Teniendo en cuenta estos problemas en el presente trabajo realizamos una búsqueda por homología para identificar las proteínas de la vía de ARN pequeños en Trematodos, Cestodos y Turbelarios, tomando como referencia el nematodo modelo *C. elegans*.

Si bien en este trabajo se reportan los genes de la vía de miARN identificados en 12 platelmintos, durante la Maestría se generaron datos, aún preliminares, sobre los genes participantes en las otras vías de síntesis ARN pequeños que continuaremos estudiando durante el doctorado.

Para ello se identificaron los genes participantes en el nematodo modelo, y se generó localmente una base de datos con toda la información transcriptómica y genómica disponible de platelmintos, sobre la que realizar las búsquedas. Las secuencias identificadas fueron analizadas a través de alineamientos y generación de árboles filogenéticos, y determinación de dominios proteicos presentes. Adicionalmente una vez que se disponibilizaron públicamente los genomas de varios platelmintos en el sitio Wormbase-Parasite, se realizaron búsquedas en ese sitio, utilizando los reportes y mapas allí generados.

Se hallaron ortólogos para la mayoría de los genes de la vía excepto dos genes (VIG-1 y AIN-1/2) vinculados al complejo RISC. Aunque no se conoce como interactúan VIG-1 y AIN-1/2 dentro del complejo miRISC, la ausencia de ambos genes resulta llamativa y es posible suponer que podrían actuar coordinadamente. En todo caso, la ausencia de estos genes nos hace pensar que podría existir un proceso de simplificación del complejo RISC en platelmintos. Tampoco es posible descartar que los genes ortólogos de VIG-1 y AIN-1/2 de platelmintos presenten una divergencia mayor a la de otros grupos de ortólogos identificados en este trabajo y que por esa razón no hayan sido detectados con los métodos utilizados en este estudio, o que existan secuencias análogas que cumplan las funciones de las proteínas no detectadas.

Observamos que en términos generales la vía de miARNs está completa en todos los platelmintos estudiados. Sin embargo, detectamos algunas particularidades interesantes en estos organismos. Las proteínas centrales Dicer y Argonautas están presente en todos los platelmintos, aunque con particularidades llamativas.

Los platelmintos presentan más de una proteína de tipo Dicer

Dicer pertenece a la clase III de la familia de las enzimas Ribonucleasa III se caracterizan por tener un dominio Helicasa C en la región N-terminal, un dominio PAZ, dos dominios Ribonucleasa III y un dominio de unión a ARNdc (Jaronzcyk, Carmichael, and Hobman 2005). Sin embargo, proteínas Dicer funcionales que no poseen uno o varios de los dominios proteicos funcionales descritos en la estructura canónica de Dicer han sido reportados en varios organismos (Jaronzcyk, Carmichael, and Hobman 2005; MacRae et al. 2006; Shi, Tschudi, and Ullu 2006; Pompey, Foda, and Singh 2015). En los platelmintos encontramos una Dicer canónica y se encontró además un segundo grupo de proteínas Dicer que se agrupan en un subclado independiente al que denominamos Dicer-like.

Mientras que en Turbelarios y Cestodos las proteínas Dicer-like tienen dominios funcionales similares a los de las proteínas Dicer canónicas, en trematodos solo parecen estar presentes los dos dominios Ribonucleasa III. No podemos descartar que la secuencia de otros dominios funcionales sean muy divergentes y por lo tanto no pudieran ser detectadas. Aunque proteínas Dicer con solo dos dominios Ribonucleasa III han probado ser funcionales en *Trypanosoma brucei* (Shi, Tschudi, and Ullu 2006), desconocemos si las proteínas Dicer-like de Trematodos son funcionales. Los datos de expresión génica muestran que Dicer-like se expresa en *S. mansoni* (<http://www.genedb.org/>). Esto sumado al hecho que en dos familias de trematodos, Opisthorchidae y Fasciolidae, los genes Dicer-like estén duplicados en tándem y que las secuencias se conservan a nivel aminoacídico sugieren que estos genes pueden estar cumpliendo alguna función aun no detectada.

En platelmintos hay dos subgrupos de Argonautas

Las Proteínas Argonauta se caracterizan por poseer un dominio de unión a ARN llamado PAZ y un dominio catalítico (PIWI). Las Argonautas se clasifican tradicionalmente en tres subfamilias las Ago-like, Piwi-like y la clase 3 que corresponde a las proteínas Argonauta secundarias presentes en nematodos (Wago) (Hutvagner and Simard 2008).

Ya ha sido reportado que en Platelminotos existen varios genes Argonauta (Zheng 2012; Luo et al. 2010; Gomes et al. 2009). Sin embargo este es el primer trabajo que compara los genes Argonauta en múltiples especies de platelmintos lo que permitió encontrar claras diferencias entre Trematodos, Cestodos y Turbelarios.

Al analizar el árbol filogenético a partir del alineamiento de proteínas de la familia Argonauta de varias especies de la escala evolutiva observamos que las proteínas Argonautas de platelmintos se agrupan principalmente en los clados Ago y en un cuarto clado que denominamos Flatworm-Agos (FL-Ago). A su vez, si bien el grupo de las FL-Agos parece tener un origen ancestral en todos los platelmintos, observamos que el número de copias es variable. *S. mansoni* y *S. japonicum* tienen dos genes FL-Ago que parecen haber surgido en un ancestro común de estas dos especies. Dada la divergencia evolutiva cercana de los *Schistosomas* no resulta extraño que las secuencias de estas especies se agrupen en una rama dentro del árbol de las FL-Agos. Lo mismo sucede con las 3 copias detectadas en las especies de *Echinococcus*. En *T. solium*, *H. microstoma*, *M. corti* y *S. mediterranea* las duplicaciones parecen haber ocurrido dentro de cada especie por separado. En todo caso, parece existir una presión evolutiva para la ganancia de genes FL-Ago de forma independiente entre los distintos linajes platelmintos. No podemos descartar, a su vez, que los genes FL-Agos hayan surgido por conversión génica.

No se detectaron ortólogos de proteínas Piwi-like en trematodos y cestodos, pero si en turbelarios. Estos resultados concuerdan con reportes previos (Zheng 2012; Skinner et al.

2014) y con la ausencia de piARNs en los datos la secuenciación de ARN pequeños (Fontenla et al. 2015; Cucher et al. 2015). En el clado Ago correspondiente a los argonautas canónicos, solo la planaria *S. mediterranea* posee dos genes mientras que en trematodos y cestodos hay un gen único. Sin embargo en el clúster de las Flatworm-Agos la cantidad de parálogos en cada especie es heterogéneo, entre dos y tres. Solo en *S. mansoni* y *Echinococcus* se observó que los genes de las FL-Agos están dispuestos en tándem aunque no es posible descartar que en otros platelmintos exista la misma estructura genómica. Esto se debe a que los ensamblajes de los genomas de las otras especies son todavía muy primarios, representados por muchos contigs cortos, por lo que es posible que esta estructura genómica exista pero pase aun desapercibida.

Esta baja calidad de algunos de los borradores genómicos nos llevaron a la necesidad de reanotar algunos modelos génicos, pues detectamos errores de anotación, y aun más, regiones donde parece estar presente el gen en cuestión, pero que no fue predicho en la anotación inicial.

En ningún platelminto se hallaron ortólogos de las Wago (2ary-Agos) que participan principalmente en *C. elegans* como Argonautas de la vía de amplificación mediada RNA polimerasas dependientes de ARN. Este resultado, junto con otros que obtuvimos en la búsqueda de ortólogos de genes participantes en la vía de amplificación (no reportados aquí), nos permiten concluir que esta vía presente en algunos nematodos, no están presente en platelmintos.

Los resultados obtenidos evidencian claras diferencias entre los mecanismos de regulación génica entre platelmintos y nematodos, y resaltan la existencia de particularidades en platelmintos que requieren un estudio más detallado.

*Los microARN del estadio NEJ de *F. hepatica**

Cuando comenzamos este trabajo se conocían datos de algunos miARNs del estadio adulto de *F. hepatica* (Xu et al. 2012) pero se desconocía si los perfiles de expresión variaban respecto a otros estadios. Para iniciar la caracterización de los miARNs en el estadio invasivo NEJ de *F. hepatica* obtuvimos y secuenciamos los ARN pequeños de una población de juveniles obtenidos in-vitro por la activación de metacercarias.

Identificamos representantes de 40 miARNs, la mayoría de los cuales pertenecen a familias conservadas así como algunas pocas descritas solo en platelmintos. Todas las familias de miARNs reportadas previamente en el estadio adulto de *F. hepatica* (Xu et al. 2012) fueron detectadas, con altos niveles de expresión en el estadio NEJ.

Al momento de realizado este trabajo no se contaba con un ensamblaje del genoma de *F. hepatica*, por lo que fue necesario realizar un ensamblaje local para identificar miARNs noveles putativos, y logramos identificar 5 miARN noveles con buen folding y significancia estadística expresados en el estadio NEJ. Estos miARN noveles no estaban conservados en los genomas de otros trematodos por lo que podrían actuar modulando vías metabólicas y etapas del desarrollo específicas de *F. hepatica*. Previo a que este trabajo fuera publicado (Fontenla et al. 2015) se hicieron públicos dos ensamblajes del genoma de *F. hepatica* (Cwiklinski et al. 2015; Martin et al. 2015), los que permitieron confirmar todos los miARNs identificados a partir de nuestro ensamblaje local.

Los platelmintos presentan una reducción de su miRNoma

La historia de la evolución de los animales muestra que una vez que un miARN emerge en un linaje particular raramente se pierde en los linajes descendientes (Sempere et al. 2007).

Dada esta fuerte conservación, se ha sugerido que los miARNs están bajo una intensa selección negativa y consecuentemente son considerados como un marcador filogenético

relevante (Wheeler et al. 2009; Tarver et al. 2013). Dada su posición entre los Lophotorchozoarios, se espera que los platelmintos compartan al menos 46 familias de miARNs (Tarver et al. 2013), sin embargo estudios recientes han reportado la ausencia de varias familias de miARNs, especialmente en los linaje parasitarios (Fromm et al. 2013; Bai et al. 2014). En concordancia con estos estudios, nuestro dataset muestra que el miRnoma de *F. hepatica* solo presenta la mitad de las familias esperadas para un Lophotorchozoario. A pesar de esta reducción, en nuestro set de miARNs detectamos familias que habían sido reportadas como ausentes en Trematodos (Fromm et al. 2013; Lu et al. 2009). Adicionalmente, se hallaron familias de miARNs reportados en otros platelmintos endoparasitos (Bai et al. 2014), por lo que podrían representar una innovación adquirida solo en estos organismos con función aún desconocida. Pero que podrían estar regulando genes específicos de vías de metabólicas surgidas en este linaje de los platelmintos por la adaptación al parasitismo.

La menor conservación de la secuencia de los miARNs de platelmintos sugiere una alta tasa mutacional

Los miARNs están entre las secuencias evolutivamente más conservadas con una baja tasa de mutaciones (Wheeler et al. 2009). Dentro de la secuencia de los miARNs maduros la región ‘semilla’, comprendida entre las posiciones 2-8, es la principal determinante de la selección del sitio blanco y esta extremadamente conservada.

Cuando analizamos las familias de miARNs conservadas presentes en *F. hepatica* observamos varias diferencias en la secuencia fuera de la región semilla. Interesantemente, esta divergencia no solo se restringió a *F. hepatica* y parece ser característica de los platelmintos en general.

Mientras que la conservación es notoriamente alta incluso entre deuterostomados, ecdysozoas y lophotrochozoas evolutivamente distantes, los platelmintos muestran variaciones marcadas (en general linaje específicas) con conservación estricta restringida a

la región semilla. La alta tasa de variaciones observada en las secuencias los miARNs de platelmintos podría afectar los blancos identificados por estos. En todo caso, si están co-evolucionando, una alta tasa de sustituciones parece ser característica de los platelmintos. El origen de estas variaciones en secuencias tan conservadas a lo largo de la evolución es aun desconocido y las consecuencias de estas variaciones son preguntas interesantes que emergen de este estudio.

Secuencias derivadas de ARNt integran la población de ARN pequeños

Los piARNs son secuencias de ~30 nt que están relacionados al silenciamiento post-transcripcional de transposones y secuencias repetidas en la línea germinal (Weick and Miska 2014). A su vez han detectado altos niveles de expresión de piARNs en los neoblastos *S. mediterranea* (Palakodeti et al. 2008; Friedländer et al. 2009).

Mientras que los neoblastos existen en los platelmintos parásitos, como ya hemos mencionado, no se han detectado piARNs en Cestodos y Trematodos, lo cual plantea la interrogante de como se mantiene la integridad genómica en estos organismos (Skinner et al. 2014). Se ha propuesto que otros ARN pequeños podrían asumir el rol de los piARNs en platelmintos y en algunos otros nematodos en los que no se ha identificado piARNs (Britton et al. 2014).

En nuestro set de datos hallamos una población de ARN pequeños de 32 nt que correspondieron a mitades ARNt. A su vez, dentro de esta población observamos clara sobrerrepresentación de secuencias que correspondían a mitades 5'.

Los ARNt halves son secuencias de entre 30-35 nt derivadas de la fragmentación de ARNt en el bucle anticodon y se las ha reportado como reguladores post-transcripcionales en respuesta al estrés en una gran variedad de organismos (Garcia-Silva et al. 2013; Thompson and Parker 2009).

A su vez se ha reportado en *T. cruzi* que los ARNt halves son secretados y podrían participar en la interacción parásito-hospedero durante el proceso invasivo (García-Silva et al. 2014). Nuestros hallazgos de ARN pequeños derivados de ARNts en el estadio NEJ plantea nuevas e interesantes preguntas sobre su papel como reguladores de la expresión génica en *F. hepatica* y/o si intervienen en la invasión, o si por otro lado son producidos por el parásito en respuesta al estrés del cultivo.

CONCLUSIÓN

Los resultados obtenidos en este trabajo aportan al conocimiento de un mecanismo de regulación aún poco estudiado en platelmintos en general y en *F. hepatica* en particular. Notablemente detectamos variaciones relevantes tanto en las vías de generación de ARN pequeños como en las poblaciones de ARN pequeños generados. Estas singularidades requieren un estudio más detallado, y pueden aportar al conocimiento de los mecanismos de regulación en general pero sobre todo resultan promisorias para el desarrollo de nuevas estrategias de control de estas parasitosis.

BIBLIOGRAFÍA

- Acosta, D. 1991. "Epidemiología y control de *Fasciola hepatica* en Uruguay." Pp. 233-236 en *Enfermedades parasitarias de importancia económica en bovinos: Bases epidemiológicas para su prevención y control en Argentina y Uruguay*, editado por A Nari y C Fiel, Editorial Agropecuaria Hemisferio Sur.
- Ai, L, M J Xu, M X Chen, Y N Zhang, S H Chen, J Guo, Y C Cai, X N Zhou, X Q Zhu, and J X Chen. 2012. "Characterization of microRNAs in *Taenia Saginata* of Zoonotic Significance by Solexa Deep Sequencing and Bioinformatics Analysis." *Parasitology Research* 110 (6): 2373–78. doi:10.1007/s00436-011-2773-x.
- Bai, Yun, Zhuangzhi Zhang, Lei Jin, Hui Kang, Yongqiang Zhu, Lu Zhang, Xia Li, et al. 2014. "Genome-Wide Sequencing of Small RNAs Reveals a Tissue-Specific Loss of Conserved microRNA Families in *Echinococcus Granulosus*." *BMC Genomics* 15 (1): 736. doi:10.1186/1471-2164-15-736.
- Bedford, Mark T., and Stéphane Richard. 2005. "Arginine Methylation: An Emerging Regulator of Protein Function." *Molecular Cell* 18 (3): 263–72. doi:10.1016/j.molcel.2005.04.003.
- Bernstein, E, Amy a. Caudy, S M Hammond, and G J Hannon. 2001. "Role for a Bidentate Ribonuclease in the Initiation Step of RNA Interference." *Nature* 409 (6818): 363–66. doi:10.1038/35053110.
- Britton, Collette, Alan D Winter, Victoria Gillan, and Eileen Devaney. 2014. "microRNAs of Parasitic Helminths - Identification, Characterization and Potential as Drug Targets." *International Journal for Parasitology. Drugs and Drug Resistance* 4 (2). Australian Society for Parasitology: 85–94. doi:10.1016/j.ijpddr.2014.03.001.
- Burge, Sarah W, Jennifer Daub, Ruth Eberhardt, John Tate, Lars Barquist, Eric P Nawrocki, Sean R Eddy, Paul P Gardner, and Alex Bateman. 2013. "Rfam 11.0: 10 Years of RNA Families." *Nucleic Acids Research* 41 (Database issue): D226–32. doi:10.1093/nar/gks1005.
- Büssing, Ingo, Jr-Shiuan Yang, Eric C Lai, and Helge Grosshans. 2010. "The Nuclear Export Receptor XPO-1 Supports Primary miRNA Processing in *C. Elegans* and *Drosophila*." *The EMBO Journal* 29 (11): 1830–39. doi:10.1038/emboj.2010.82.
- Cai, Pengfei, Xianyu Piao, Lili Hao, Shuai Liu, Nan Hou, Heng Wang, and Qijun Chen. 2013. "A Deep Analysis of the Small Non-Coding RNA Population in *Schistosoma Japonicum* Eggs." *PloS One* 8 (5): e64003. doi:10.1371/journal.pone.0064003.
- Cancela, Martín, Daniel Acosta, Gabriel Rinaldi, Edileusa Silva, Rosario Durán, Leda

- Roche, Arnaldo Zaha, Carlos Carmona, and Jose F Tort. 2008. "A Distinctive Repertoire of Cathepsins Is Expressed by Juvenile Invasive *Fasciola Hepatica*." *Biochimie* 90 (10): 1461–75. doi:10.1016/j.biochi.2008.04.020.
- Cancela, Martín, Natalia Ruétalo, Nicolás Dell Oca, Edileuza Silva, Pablo Smircich, Gabriel Rinaldi, Leda Roche, et al. 2010. "Survey of Transcripts Expressed by the Invasive Juvenile Stage of the Liver Fluke *Fasciola Hepatica*." *BMC Genomics* 11:227.
- Caudy, Amy a, René F Ketting, Scott M Hammond, Ahmet M Denli, Anja M P Bathoorn, Bastiaan B J Tops, Jose M Silva, Mike M Myers, Gregory J Hannon, and Ronald H a Plasterk. 2003. "A Micrococcal Nuclease Homologue in RNAi Effector Complexes." *Nature* 425 (6956): 411–14. doi:10.1038/nature01956.
- Caudy, Amy a., Mike Myers, Gregory J. Hannon, and Scott M. Hammond. 2002. "Fragile X-Related Protein and VIG Associate with the RNA Interference Machinery." *Genes and Development* 16 (19): 2491–96. doi:10.1101/gad.1025202.
- Chan, Shih-Peng, Gopalakrishna Ramaswamy, Eun-Young Choi, and Frank J Slack. 2008. "Identification of Specific Let-7 microRNA Binding Complexes in *Caenorhabditis Elegans*." *RNA (New York, N.Y.)* 14 (10): 2104–14. doi:10.1261/rna.551208.
- Choudhuri, Supratim. 2009. "Lesser Known Relatives of miRNA." *Biochemical and Biophysical Research Communications* 388 (2). Elsevier Inc.: 177–80. doi:10.1016/j.bbrc.2009.08.039.
- Claycomb, Julie M., Pedro J. Batista, Ka Ming Pang, Weifeng Gu, Jessica J. Vasale, Josien C. van Wolfswinkel, Daniel A. Chaves, et al. 2009. "The Argonaute CSR-1 and Its 22G-RNA Cofactors Are Required for Holocentric Chromosome Segregation." *Cell* 139 (1). Elsevier Ltd: 123–34. doi:10.1016/j.cell.2009.09.014.
- Copeland, Claudia S, Manja Marz, Dominic Rose, Jana Hertel, Paul J Brindley, Clara Bermudez Santana, Stephanie Kehr, Camille Stephan-Otto Attolini, and Peter F Stadler. 2009. "Homology-Based Annotation of Non-Coding RNAs in the Genomes of *Schistosoma Mansoni* and *Schistosoma Japonicum*." *BMC Genomics* 10 (January): 464. doi:10.1186/1471-2164-10-464.
- Corvo, Ileana, Martín Cancela, Mónica Cappetta, Natalia Pi-Denis, José F Tort, and Leda Roche. 2009. "The Major Cathepsin L Secreted by the Invasive Juvenile *Fasciola Hepatica* Prefers Proline in the S2 Subsite and Can Cleave Collagen." *Molecular and Biochemical Parasitology* 167 (1): 41–47. doi:10.1016/j.molbiopara.2009.04.005.
- Cucher, M, L Prada, G Mourglia-Ettlin, S Dematteis, F Camicia, S Asurmendi, and M Rosenzvit. 2011. "Identification of *Echinococcus Granulosus* microRNAs and Their Expression in Different Life Cycle Stages and Parasite Genotypes." *International Journal for Parasitology* 41 (3-4). Australian Society for Parasitology Inc.: 439–48. doi:10.1016/j.ijpara.2010.11.010.
- Cucher, Marcela, Natalia Macchiaroli, Laura Kamenetzky, Lucas Maldonado, Klaus

- Brehm, and Mara Rosenzvit. 2015. "High-Throughput Characterization of *Echinococcus* Spp. Metacestode miRNomes." *International Journal for Parasitology* 45 (4). Australian Society for Parasitology Inc.: 253–67. doi:10.1016/j.ijpara.2014.12.003.
- Cwiklinski, Krystyna, John Pius Dalton, Philippe J Dufresne, James La Course, Diana JI Williams, Jane Hodgkinson, and Steve Paterson. 2015. "The *Fasciola Hepatica* Genome: Gene Duplication and Polymorphism Reveals Adaptation to the Host Environment and the Capacity for Rapid Evolution." *Genome Biology* 16 (1). ??? 71. doi:10.1186/s13059-015-0632-2.
- Dalzell, Johnathan J, Paul McVeigh, Neil D Warnock, Makedonka Mitreva, David McK Bird, Pierre Abad, Colin C Fleming, et al. 2011. "RNAi Effector Diversity in Nematodes." *PLoS Neglected Tropical Diseases* 5 (6): e1176. doi:10.1371/journal.pntd.0001176.
- Das, Partha P, Marloes P Bagijn, Leonard D Goldstein, Julie R Woolford, Nicolas J Lehrbach, Alexandra Sapetschnig, Heeran R Buhecha, et al. 2008. "Piwi and piRNAs Act Upstream of an Endogenous siRNA Pathway to Suppress Tc3 Transposon Mobility in the *Caenorhabditis Elegans* Germline." *Molecular Cell* 31 (1): 79–90. doi:10.1016/j.molcel.2008.06.003.
- de Souza Gomes, Matheus, Mohan Kumar Muniyappa, Sávio Gonçalves Carvalho, Renata Guerra-Sá, and Charles Spillane. 2011. "Genome-Wide Identification of Novel microRNAs and Their Target Genes in the Human Parasite *Schistosoma Mansoni*." *Genomics* 98 (2). Elsevier B.V.: 96–111. doi:10.1016/j.ygeno.2011.05.007.
- Dell'Oca, Nicolás, Tatiana Basika, Ileana Corvo, Estela Castillo, Paul J Brindley, Gabriel Rinaldi, and Jose F Tort. 2014. "Molecular & Biochemical Parasitology RNA Interference in *Fasciola Hepatica* Newly Excysted Juveniles : Long dsRNA Induces More Persistent Silencing than siRNA" 197: 28–35.
- Duchaine, Thomas F, James a Wohlschlegel, Scott Kennedy, Yanxia Bei, Darryl Conte, Kaming Pang, Daniel R Brownell, et al. 2006. "Functional Proteomics Reveals the Biochemical Niche of *C. Elegans* DCR-1 in Multiple Small-RNA-Mediated Pathways." *Cell* 124 (2): 343–54. doi:10.1016/j.cell.2005.11.036.
- Fire, A, S Xu, M K Montgomery, S A Kostas, S E Driver, and C C Mello. 1998. "Potent and Specific Genetic Interference by Double-Stranded RNA in *Caenorhabditis Elegans*." *Nature* 391 (6669): 806–11. doi:10.1038/35888 [doi].
- Fire, Andrew, Donna Albertson, Susan White Harrison, and D G Moerman. 1991. "Production of Antisense RNA Leads to Effective and Specific Inhibition of Gene Expression in *C. Elegans* Muscle." *Development (Cambridge, England)* 113 (2): 503–14.
- Fischer, Steve, Cristina Aurrecoechea, Brian P Brunk, Xin Gao, Omar S Harb, Eileen T Kraemer, Cary Pennington, et al. 2011. "The Strategies WDK: A Graphical Search

- Interface and Web Development Kit for Functional Genomics Databases.” *Database : The Journal of Biological Databases and Curation* 2011 (January): bar027. doi:10.1093/database/bar027.
- Fischer, Sylvia E J. 2010. “Small RNA-Mediated Gene Silencing Pathways in *C. Elegans*.” *The International Journal of Biochemistry & Cell Biology* 42 (8): 1306–15. doi:10.1016/j.biocel.2010.03.006.
- Fischer, Sylvia E.J., Maurice D. Bulter, Qi Pan, and Gary Ruvkun. 2008. “Trans-Splicing in *C. Elegans* Generates the Negative RNAi Regulator ERI-6/7.” *Nature* 455 (7212): 491–96. doi:10.1038/nature07274.Trans.
- Fontenla, Santiago, Nicolás Dell’Oca, Pablo Smircich, José F. Tort, and Mar Siles-Lucas. 2015. “The miRnome of *Fasciola Hepatica* Juveniles Endorses the Existence of a Reduced Set of Highly Divergent Micro RNAs in Parasitic Flatworms.” *International Journal for Parasitology* 45 (October). Australian Society for Parasitology Inc.: 901–13. doi:10.1016/j.ijpara.2015.06.007.
- Friedländer, Marc R, Catherine Adamidi, Ting Han, Svetlana Lebedeva, Thomas a Isenbarger, Martin Hirst, Marco Marra, et al. 2009. “High-Resolution Profiling and Discovery of Planarian Small RNAs.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (28): 11546–51. doi:10.1073/pnas.0905222106.
- Friedländer MR, Mackowiak SD, Li N, Chen W, Rajewsky N. 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40: 37–52. doi:10.1093/nar/gkr688
- Fromm, B., M. Trelis, M. Hackenberg, F. Cantalapietra, D. Bernal, and a. Marcilla. 2015. “The Revised microRNA Complement of *Fasciola Hepatica* Reveals a Plethora of Overlooked microRNAs and Evidence for Enrichment of Immuno-Regulatory microRNAs in Extracellular Vesicles.” *International Journal for Parasitology*. Australian Society for Parasitology Inc. doi:10.1016/j.ijpara.2015.06.002.
- Fromm, Bastian, Merete Molton Warren, Christoph Hahn, Eivind Hovig, and Lutz Bachmann. 2013. “Substantial Loss of Conserved and Gain of Novel microRNA Families in Flatworms.” *Molecular Biology and Evolution* 30 (12): 2619–28. doi:10.1093/molbev/mst155.
- Gao, Zeqian, Miao Wang, David Blair, Yadong Zheng, and Yongxi Dou. 2014. “Phylogenetic Analysis of the Endoribonuclease Dicer Family.” *PLoS ONE* 9 (4): 1–7. doi:10.1371/journal.pone.0095350.
- Garcia-Silva, Maria R, Florencia Cabrera-Cabrera, Roberta Ferreira Cura das Neves, Thaís Souto-Padrón, Wanderley de Souza, and Alfonso Cayota. 2014. “Gene Expression Changes Induced by *Trypanosoma Cruzi* Shed Microvesicles in Mammalian Host Cells: Relevance of tRNA-Derived Halves.” *BioMed Research International* 2014 (January): 305239. doi:10.1155/2014/305239.

- Garcia-Silva, Maria Rosa, Florencia Cabrera-Cabrera, Maria Catalina Güida, and Alfonso Cayota. 2013. "Novel Aspects of tRNA-Derived Small RNAs with Potential Impact in Infectious Diseases." *Advances in Bioscience and Biotechnology* 04 (05): 17–25. doi:10.4236/abb.2013.45A002.
- Garcia-Silva, Maria Rosa, Magali Frugier, Juan Pablo Tosar, Alejandro Correa-Dominguez, Lysangela Ronalte-Alves, Adriana Parodi-Talice, Carlos Rovira, Carlos Robello, Samuel Goldenberg, and Alfonso Cayota. 2010. "A Population of tRNA-Derived Small RNAs Is Actively Produced in Trypanosoma Cruzi and Recruited to Specific Cytoplasmic Granules." *Molecular and Biochemical Parasitology* 171 (2). Elsevier B.V.: 64–73. doi:10.1016/j.molbiopara.2010.02.003.
- Gaydos, Laura J., Andreas Rechtsteiner, Thea a. Egelhofer, Coleen R. Carroll, and Susan Strome. 2012. "Antagonism between MES-4 and Polycomb Repressive Complex 2 Promotes Appropriate Gene Expression in C. Elegans Germ Cells." *Cell Reports* 2 (5). The Authors: 1169–77. doi:10.1016/j.celrep.2012.09.019.
- Gerlach, Daniel, Evgenia V Kriventseva, Nazim Rahman, Charles E Vejnar, and Evgeny M Zdobnov. 2009. "miROrtho: Computational Survey of microRNA Genes." *Nucleic Acids Research* 37 (Database issue): D111–17. doi:10.1093/nar/gkn707.
- Gomes, Matheus S, Fernanda J Cabral, Liana K Jannotti-Passos, Omar Carvalho, Vanderlei Rodrigues, Elio H Baba, and Renata G Sá. 2009. "Preliminary Analysis of miRNA Pathway in Schistosoma Mansoni." *Parasitology International* 58 (1). Elsevier B.V.: 61–68. doi:10.1016/j.parint.2008.10.002.
- Grishok, Alla. 2005. "RNAi Mechanisms in Caenorhabditis Elegans." *FEBS Letters* 579 (26): 5932–39. doi:10.1016/j.febslet.2005.08.001.
- Grishok, Alla, Amy E. Pasquinelli, Darryl Conte, Na Li, Susan Parrish, Ilho Ha, David L. Baillie, Andrew Fire, Gary Ruvkun, and Craig C. Mello. 2001. "Genes and Mechanisms Related to RNA Interference Regulate Expression of the Small Temporal RNAs That Control C. Elegans Developmental Timing." *Cell* 106 (1): 23–34. doi:10.1016/S0092-8674(01)00431-7.
- Gu, Weifeng, Masaki Shirayama, Darryl Conte, Jessica Vasale, Pedro J Batista, Julie M Claycomb, James J Moresco, et al. 2009. "Distinct Argonaute-Mediated 22G-RNA Pathways Direct Genome Surveillance in the C. Elegans Germline." *Molecular Cell* 36 (2). Elsevier Ltd: 231–44. doi:10.1016/j.molcel.2009.09.020.
- Guang, Shouhong, Aaron F Bochner, Derek M Pavelec, Kirk B Burkhart, Jennifer Lachowiec, and Scott Kennedy. 2008. "An Argonaute Transports siRNAs from the Cytoplasm to the Nucleus" 321 (5888): 537–41. doi:10.1126/science.1157647.An.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol.* 2010;59: 307–321. doi:10.1093/sysbio/syq010
- Hall, Ta. 1999. "BioEdit: A User-Friendly Biological Sequence Alignment Editor and

- Analysis Program for Windows 95/98/NT.” *Nucleic Acids Symposium Series*.
- Hammell, Christopher M, Isabella Lubin, Peter R Boag, T Keith Blackwell, and Victor Ambros. 2009. “Nhl-2 Modulates microRNA Activity in *Caenorhabditis Elegans*.” *Cell* 136 (5). Elsevier Ltd: 926–38. doi:10.1016/j.cell.2009.01.053.
- Huang, Jian, Pei Hao, Hui Chen, Wei Hu, Qing Yan, Feng Liu, and Ze-Guang Han. 2009. “Genome-Wide Identification of *Schistosoma Japonicum* microRNAs Using a Deep-Sequencing Approach.” *PloS One* 4 (12): e8206. doi:10.1371/journal.pone.0008206.
- Hutvagner, Gyorgy, and Martin J Simard. 2008. “Argonaute Proteins: Key Players in RNA Silencing.” *Nature Reviews. Molecular Cell Biology* 9 (1): 22–32. doi:10.1038/nrm2321.
- Jamalkandi, Azimzadeh and Masoudi-Nejad, Ali. 2011. “RNAi Pathway Integration in *Caenorhabditis Elegans* Development.” *Functional & Integrative Genomics* 11 (3): 389–405. doi:10.1007/s10142-011-0236-1.
- Jaronczyk, K, J B Carmichael, and T C Hobman. 2005. “Exploring the Functions of RNA Interference Pathway Proteins: Some Functions Are More RISCy than Others?” *Biochem J* 387 (Pt 3): 561–71. doi:10.1042/BJ20041822.
- Jin, Xiaoliang, Lixia Lu, Hailong Su, Zhongzi Lou, Fang Wang, Yadong Zheng, and Guo-Tong Xu. 2013. “Comparative Analysis of Known miRNAs across Platyhelminths.” *The FEBS Journal* 280 (16): 3944–51. doi:10.1111/febs.12395.
- Johnson, L Steven, Sean R Eddy, and Elon Portugaly. 2010. “Hidden Markov Model Speed Heuristic and Iterative HMM Search Procedure.” *BMC Bioinformatics* 11: 431. doi:10.1186/1471-2105-11-431.
- Katoh, Kazutaka, and Daron M Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” *Molecular Biology and Evolution* 30 (4): 772–80. doi:10.1093/molbev/mst010.
- Keiser, Jennifer, and Jürg Utzinger. 2005. “Emerging Foodborne Trematodiasis.” *Emerging Infectious Diseases* 11 (10): 1507–14. doi:10.3201/eid1110.050614.
- Keiser, Jennifer, and Jürg Utzinger. 2009. “Food-Borne Trematodiasis.” *Clinical Microbiology Reviews* 22 (3): 466–83. doi:10.1128/CMR.00012-09.
- Ketting, René F, Thomas H.A Haverkamp, Henri G.A.M van Luenen, and Ronald H.A Plasterk. 1999. “Mut-7 of *C. Elegans*, Required for Transposon Silencing and RNA Interference, Is a Homolog of Werner Syndrome Helicase and RNaseD.” *Cell* 99 (2): 133–41. doi:10.1016/S0092-8674(00)81645-1.
- Kim, John K, Harrison W Gabel, Ravi S Kamath, Muneesh Tewari, Amy Pasquinelli, Jean-François Rual, Scott Kennedy, et al. 2005. “Functional Genomic Analysis of RNA Interference in *C. Elegans*.” *Science* 308 (5725). American Association for the Advancement of Science: 1164–67. <http://www.jstor.org.proxy.timbo.org.uy:443/stable/3842102>.

- Kozomara, Ana, and Sam Griffiths-Jones. 2014. "miRBase: Annotating High Confidence microRNAs Using Deep Sequencing Data." *Nucleic Acids Research* 42 (Database issue): D68–73. doi:10.1093/nar/gkt1181.
- Krautz-Peterson, Greice, Rita Bhardwaj, Zahra Faghiri, Cibele A Tararam, and Patrick J Skelly. 2010. "RNA Interference in Schistosomes: Machinery and Methodology." *Parasitology* 137 (3): 485–95. doi:10.1017/S0031182009991168.
- Krautz-peterson, Greice, and Patrick J Skelly. 2008. "Schistosoma Mansoni: The Dicer Gene and Its Expression." *Experimental Parasitology* 118 (1): 122–28.
- Kuzuoglu-Öztürk, Duygu, Eric Huntzinger, Steffen Schmidt, and Elisa Izaurralde. 2012. "The Caenorhabditis Elegans GW182 Protein AIN-1 Interacts with PAB-1 and Subunits of the PAN2-PAN3 and CCR4-NOT Deadenylation Complexes." *Nucleic Acids Research* 40 (12): 5651–65. doi:10.1093/nar/gks218.
- Lagesen, Karin, Peter Hallin, Einar Andreas Rødland, Hans Henrik Stærfeldt, Torbjørn Rognes, and David W. Ussery. 2007. "RNAmmer: Consistent and Rapid Annotation of Ribosomal RNA Genes." *Nucleic Acids Research* 35 (9): 3100–3108. doi:10.1093/nar/gkm160.
- Lee, Rosalind C. 1993. "The C. Elegans Heterochronic Gene Lin-4 Encodes Small RNAs with Antisense Complementarity to *lin-14*." *Genes & Development* 7 (5): 843–54. doi:10.1016/0092-8674(93)90529-Y.
- Lowe, T M, and S R Eddy. 1997. "tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence." *Nucleic Acids Research* 25 (5): 955–64. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=146525&tool=pmcentrez&rendertype=abstract>.
- Lu, Yi-chien, Magda Smielewska, Dasaradhi Palakodeti, Michael T Lovci, Stefan Aigner, Gene W Yeo, and Brenton R Graveley. 2009. "Deep Sequencing Identifies New and Regulated microRNAs in Schmidtea Mediterranea." *PLoS ONE* 4 (10): e7143. doi:10.1371/journal.pone.0071433.
- Luo, Rong, Xiangyang Xue, Zhangxun Wang, Jun Sun, Ying Zou, and Weiqing Pan. 2010. "Analysis and Characterization of the Genes Encoding the Dicer and Argonaute Proteins of Schistosoma Japonicum." *Parasites & Vectors* 3 (1): 90. doi:10.1186/1756-3305-3-90.
- MacRae, Ian J, Kaihong Zhou, Fei Li, Adrian Repic, Angela N Brooks, W Zacheus Cande, Paul D Adams, and Jennifer A Doudna. 2006. "Structural Basis for Double-Stranded RNA Processing by Dicer." *Science* 311 (5758): 195–98. American Association for the Advancement of Science. doi:10.1126/science.1126001. <http://www.jstor.org.proxy.timbo.org.uy:443/stable/3843253>.
- Maida, Yoshiko, and Kenkichi Masutomi. 2011. "RNA-Dependent RNA Polymerases in RNA Silencing." *Biological Chemistry* 392 (4): 299–304. doi:10.1515/BC.2011.035.

- Mangiola, Stefano, Neil D Young, Pasi Korhonen, Alinda Mondal, Jean-Pierre Scheerlinck, Paul W Sternberg, Cinzia Cantacessi, Ross S Hall, Aaron R Jex, and Robin B Gasser. 2013. "Getting the Most out of Parasitic Helminth Transcriptomes Using HelmDB: Implications for Biology and Biotechnology." *Biotechnology Advances* 31 (8). Elsevier Inc.: 1109–19. doi:10.1016/j.biotechadv.2012.12.004.
- Mango, S E. 2001. "Stop Making nonSense: The C. Elegans Smg Genes." *Trends in Genetics : TIG* 17 (11): 646–53. doi:10.1016/S0168-9525(01)02479-9.
- Marcos, Luis a, Angelica Terashima, and Eduardo Gotuzzo. 2008. "Update on Hepatobiliary Flukes: Fascioliasis, Opisthorchiasis and Clonorchiasis." *Current Opinion in Infectious Diseases* 21 (5): 523–30. doi:10.1097/QCO.0b013e32830f9818.
- Martin, J., B. a. Rosa, P. Ozersky, K. Hallsworth-Pepin, X. Zhang, V. Bhonagiri-Palsikar, R. Tyagi, et al. 2015. "Helminth.net: Expansions to Nematode.net and an Introduction to Trematode.net." *Nucleic Acids Research* 43 (D1): D698–706. doi:10.1093/nar/gku1128.
- Mas-Coma, S., M.D. Bargues, and M.a. Valero. 2005. "Fascioliasis and Other Plant-Borne Trematode Zoonoses." *International Journal for Parasitology* 35 (11-12): 1255–78. doi:10.1016/j.ijpara.2005.07.010.
- Mituyama, Toutai, Kouichirou Yamada, Emi Hattori, Hiroaki Okida, Yukiteru Ono, Goro Terai, Aya Yoshizawa, Takashi Komori, and Kiyoshi Asai. 2009. "The Functional RNA Database 3.0: Databases to Support Mining and Annotation of Functional RNAs." *Nucleic Acids Research* 37 (Database issue): D89–92. doi:10.1093/nar/gkn805.
- Montgomery, Taiowa A., Young-Soo Rim, Chi Zhang, Robert H. Downen, Carolyn M. Phillips, Sylvia E. J. Fischer, and Gary Ruvkun. 2012. "PIWI Associated siRNAs and piRNAs Specifically Require the Caenorhabditis Elegans HEN1 Ortholog Henn-1." *PLoS Genetics* 8 (4): e1002616. doi:10.1371/journal.pgen.1002616.
- Newman, Martin A, and Scott M Hammond. 2010. "Emerging Paradigms of Regulated microRNA Processing." *Genes and Development* 24 (11): 1086–92. doi:10.1101/gad.1919710.
- Newmark, Phillip a, Peter W Reddien, Francesc Cebrià, and Alejandro Sánchez Alvarado. 2003. "Ingestion of Bacterially Expressed Double-Stranded RNA Inhibits Gene Expression in Planarians." *Proceedings of the National Academy of Sciences of the United States of America* 100 Suppl : 11861–65. doi:10.1073/pnas.1834205100.
- Orban, Tamas I, and Elisa Izaurralde. 2005. "Decay of mRNAs Targeted by RISC Requires XRN1, the Ski Complex, and the Exosome." *RNA (New York, N.Y.)* 11 (4): 459–69. doi:10.1261/rna.7231505.
- Palakodeti, Dasaradhi, Magda Smielewska, and Brenton R Graveley. 2006. "MicroRNAs from the Planarian Schmidtea Mediterranea: A Model System for Stem Cell Biology." *RNA (New York, N.Y.)* 12 (9): 1640–49. doi:10.1261/rna.117206.

- Palakodeti, Dasaradhi, Magda Smielewska, Yi-Chien Lu, Gene W Yeo, and Brenton R Graveley. 2008. "The PIWI Proteins SMEDWI-2 and SMEDWI-3 Are Required for Stem Cell Function and piRNA Expression in Planarians." *RNA (New York, N.Y.)* 14 (6): 1174–86. doi:10.1261/rna.1085008.
- Parkinson, M, S M O'Neill, and J P Dalton. 2007. "Endemic Human Fasciolosis in the Bolivian Altiplano." *Epidemiology and Infection* 135 (4): 669–74. doi:10.1017/S095026880600728X.
- Pavelec, Derek M, Jennifer Lachowicz, Thomas F Duchaine, Harold E Smith, and Scott Kennedy. 2009. "Requirement for the ERI/DICER Complex in Endogenous RNA Interference and Sperm Development in *Caenorhabditis Elegans*." *Genetics* 183 (4): 1283–95. doi:10.1534/genetics.109.108134.
- Pompey, Justine M., Bardees Foda, and Upinder Singh. 2015. "A Single RNaseIII Domain Protein from *Entamoeba Histolytica* Has dsRNA Cleavage Activity and Can Help Mediate RNAi Gene Silencing in a Heterologous System." *Plos One* 10 (7): e0133740. doi:10.1371/journal.pone.0133740.
- Reinhart, B J, F J Slack, M Basson, A E Pasquinelli, J C Bettinger, A E Rougvie, H R Horvitz, and G Ruvkun. 2000. "The 21-Nucleotide Let-7 RNA Regulates Developmental Timing in *Caenorhabditis Elegans*." *Nature* 403 (6772). ENGLAND: 901–6. doi:10.1038/35002607.
- Rinaldi, Gabriel, Maria E Morales, Martín Cancela, Estela Castillo, Paul J Brindley, and José F Tort. 2008. "Development of Functional Genomic Tools in Trematodes: RNA Interference and Luciferase Reporter Gene Activity in *Fasciola Hepatica*." *PLoS Neglected Tropical Diseases* 2 (7): e260. doi:10.1371/journal.pntd.0000260.
- Robinson, M. W., and J. P. Dalton. 2009. "Zoonotic Helminth Infections with Particular Emphasis on Fasciolosis and Other Trematodiasis." *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1530): 2763–76. doi:10.1098/rstb.2009.0089.
- Sánchez Alvarado, a, and P a Newmark. 1999. "Double-Stranded RNA Specifically Disrupts Gene Expression during Planarian Regeneration." *Proceedings of the National Academy of Sciences of the United States of America* 96 (9): 5049–54. doi:10.1073/pnas.96.9.5049.
- Sanchis, J, GV Hillyer, LM Madeira de Carvalho, MI Macchi, C Gomes, G Maldini, G Stilwell, et al. 2015. "Riesgo de Exposición a *Fasciola Hepatica* En Ganado Vacuno En Extensivo de Uruguay Y Portugal Determinado Mediante ELISA Y Un Antígeno Recombinante" 208: 201–8.
- Sanchis, J, J Suarez, Gv Hillyer, Ja Hernandez, Ma Solari, C Cazapal-Monteiro, Am Duque de Araujo, et al. 2015. "Determination of Exposure to *Fasciola Hepatica* in Horses from Uruguay Using a Recombinant-Based ELISA." *Veterinárni Medicina* 60 (09): 483–88. doi:10.17221/8439-VETMED.

- Sariözkan, S, and C YalÇin. 2011. "Estimating the Total Cost of Bovine Fasciolosis in Turkey." *Annals of Tropical Medicine and Parasitology* 105 (6): 439–44. doi:10.1179/1364859411Y.0000000031.
- Sempere, Lorenzo F, Pedro Martinez, Charles Cole, Jaume Baguña, and Kevin J Peterson. 2007. "Phylogenetic Distribution of microRNAs Supports the Basal Position of Acoel Flatworms and the Polyphyly of Platyhelminthes." *Evolution & Development* 9 (5): 409–15. doi:10.1111/j.1525-142X.2007.00180.x.
- Shi, Huafang, Christian Tschudi, and Elisabetta Ullu. 2006. "An Unusual Dicer-like 1 Protein Fuels the RNA Interference Pathway in Trypanosoma Brucei." *RNA (New York, N.Y.)* 12 (12): 2063–72. doi:10.1261/rna.246906.
- Simões, Mariana C, Jonathan Lee, Appolinaire Djikeng, Gustavo C Cerqueira, Adhemar Zerlotini, Rosiane a da Silva-Pereira, Andrew R Dalby, Philip LoVerde, Najib M El-Sayed, and Guilherme Oliveira. 2011. "Identification of Schistosoma Mansoni microRNAs." *BMC Genomics* 12 (1). BioMed Central Ltd: 47. doi:10.1186/1471-2164-12-47.
- Skinner, Danielle E, Gabriel Rinaldi, Uriel Koziol, Klaus Brehm, and Paul J Brindley. 2014. "How Might Flukes and Tapeworms Maintain Genome Integrity without a Canonical piRNA Pathway?" *Trends in Parasitology* 30 (3). Elsevier Ltd: 123–29. doi:10.1016/j.pt.2014.01.001.
- Sripa, Jittiyawadee, Porntip Pinlaor, Paul J Brindley, Banchob Sripa, Sasithorn Kaewkes, Mark W Robinson, Neil D Young, Robin B Gasser, Alex Loukas, and Thewarach Laha. 2011. "RNA Interference Targeting Cathepsin B of the Carcinogenic Liver Fluke, Opisthorchis Viverrini." *Parasitology International* 60 (3): 283–88. doi:http://dx.doi.org/10.1016/j.parint.2011.04.003.
- Stanke, Mario, and Burkhard Morgenstern. 2005. "AUGUSTUS: A Web Server for Gene Prediction in Eukaryotes That Allows User-Defined Constraints." *Nucleic Acids Research* 33 (suppl 2): W465–67. doi:10.1093/nar/gki458.
- Sun, Jun, Suwen Wang, Chen Li, Yijiu Ren, and Jinqiang Wang. 2014. "Novel Expression Profiles of microRNAs Suggest That Specific miRNAs Regulate Gene Expression for the Sexual Maturation of Female Schistosoma Japonicum after Pairing." *Parasites & Vectors* 7 (1). Parasites & Vectors: 177. doi:10.1186/1756-3305-7-177.
- Tabara, Hiroaki, Madathia Sarkissian, William G. Kelly, Jamie Fleenor, Alla Grishok, Lisa Timmons, Andrew Fire, and Craig C. Mello. 1999. "The Rde-1 Gene, RNA Interference, and Transposon Silencing in *C. Elegans*." *Cell* 99 (2): 123–32. doi:10.1016/S0092-8674(00)81644-X.
- Tabara, Hiroaki, Erbay Yigit, Haruhiko Siomi, and Craig C Mello. 2002. "The dsRNA Binding Protein RDE-4 Interacts with RDE-1, DCR-1, and a DExH-Box Helicase to Direct RNAi in *C. Elegans*." *Cell* 109 (7): 861–71. http://www.ncbi.nlm.nih.gov/pubmed/12110183.

- Tamura, K., G. Stecher, D. Peterson, A. Filipiński, and S. Kumar. 2013. "MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0." *Molecular Biology and Evolution* 30 (12): 2725–29. doi:10.1093/molbev/mst197.
- Tarver, James E, Erik a Sperling, Audrey Nailor, Alysha M Heimberg, Jeffrey M Robinson, Benjamin L King, Davide Pisani, Philip C J Donoghue, and Kevin J Peterson. 2013. "miRNAs: Small Genes with Big Potential in Metazoan Phylogenetics." *Molecular Biology and Evolution* 30 (11): 2369–82. doi:10.1093/molbev/mst133.
- Thompson, Debrah M, and Roy Parker. 2009. "Stressing out over tRNA Cleavage." *Cell* 138 (2): 215–19. doi:10.1016/j.cell.2009.07.001.
- Tops, B B, H Tabara, T Sijen, F Simmer, C C Mello, R H Plasterk, and R F Ketting. 2005. "RDE-2 Interacts with MUT-7 to Mediate RNA Interference in *Caenorhabditis Elegans*." *Nucleic Acids Res* 33 (1): 347–55. doi:10.1093/nar/gki183.
- Tsai, Isheng J., Magdalena Zarowiecki, Nancy Holroyd, Alejandro Garcíarrubio, Alejandro Sanchez-Flores, Karen L. Brooks, Alan Tracey, et al. 2013. "The Genomes of Four Tapeworm Species Reveal Adaptations to Parasitism." *Nature* 496 (March): 57–63. doi:10.1038/nature12031.
- van Roessel, Peter, and Andrea H Brand. 2004. "Spreading Silence with Sid." *Genome Biology* 5 (2): 208. doi:10.1186/gb-2004-5-2-208.
- Wang, Wei-Chi, Feng-Mao Lin, Wen-Chi Chang, Kuan-Yu Lin, Hsien-Da Huang, and Na-Sheng Lin. 2009. "miRExpress: Analyzing High-Throughput Sequencing Data for Profiling microRNA Expression." *BMC Bioinformatics* 10 (January): 328. doi:10.1186/1471-2105-10-328.
- Weick, Eva-Maria, and Eric a Miska. 2014. "piRNAs: From Biogenesis to Function." *Development (Cambridge, England)* 141 (18): 3458–71. doi:10.1242/dev.094037.
- Wheeler, Benjamin M, Alysha M Heimberg, Vanessa N Moy, Erik a Sperling, Thomas W Holstein, Steffen Heber, and Kevin J Peterson. 2009. "The Deep Evolution of Metazoan microRNAs." *Evolution & Development* 11 (1): 50–68. doi:10.1111/j.1525-142X.2008.00302.x.
- Xu, Min-Jun, Lin Ai, Jing-Hua Fu, Alasdair J Nisbet, Qing-You Liu, Mu-Xin Chen, Dong-Hui Zhou, and Xing-Quan Zhu. 2012. "Comparative Characterization of microRNAs from the Liver Flukes *Fasciola Gigantica* and *F. Hepatica*." *PloS One* 7 (12): e53387. doi:10.1371/journal.pone.0053387.
- Xu, Min-Jun, Quan Liu, Alasdair J Nisbet, Xian-Quan Cai, Chao Yan, Rui-Qing Lin, Zi-Guo Yuan, Hui-Qun Song, Xian-Hui He, and Xing-Quan Zhu. 2010. "Identification and Characterization of microRNAs in *Clonorchis Sinensis* of Human Health Significance." *BMC Genomics* 11 (January): 521. doi:10.1186/1471-2164-11-521.
- Xu, Min-Jun, Chun-Ren Wang, Si-Yang Huang, Jing-Hua Fu, Dong-Hui Zhou, Qiao-

- Cheng Chang, Xu Zheng, and Xing-Quan Zhu. 2013. "Identification and Characterization of microRNAs in the Pancreatic Fluke *Eurytrema Pancreaticum*." *Parasites & Vectors* 6 (1). Parasites & Vectors: 25. doi:10.1186/1756-3305-6-25.
- Zamore, Phillip D, and Benjamin Haley. 2005. "Ribo-Gnome: The Big World of Small RNAs." *Science (New York, N.Y.)* 309 (5740). United States: 1519–24. doi:10.1126/science.1111444.
- Zerlotini, Adhemar, Eric R G R Aguiar, Fudong Yu, Huayong Xu, Yixue Li, Neil D Young, Robin B Gasser, et al. 2013. "SchistoDB: An Updated Genome Resource for the Three Key Schistosomes of Humans." *Nucleic Acids Research* 41 (Database issue): D728–31. doi:10.1093/nar/gks1087.
- Zhang, Chi, Taiowa a Montgomery, Harrison W Gabel, Sylvia E J Fischer, Carolyn M Phillips, Noah Fahlgren, Christopher M Sullivan, James C Carrington, and Gary Ruvkun. 2011. "Mut-16 and Other Mutator Class Genes Modulate 22G and 26G siRNA Pathways in *Caenorhabditis Elegans*." *Proceedings of the National Academy of Sciences of the United States of America* 108 (4): 1201–8. doi:10.1073/pnas.1018695108.
- Zhang, Liang, Lei Ding, Tom H Cheung, Meng Qiu Dong, Jun Chen, Aileen K. Sewell, Xuedong Liu, John R. Yates, and Min Han. 2007. "Systematic Identification of *C. Elegans* miRISC Proteins, miRNAs, and mRNA Targets by Their Interactions with GW182 Proteins AIN-1 and AIN-2." *Molecular Cell* 28 (4): 598–613. doi:10.1016/j.molcel.2007.09.014.
- Zheng, Yadong. 2012. "Phylogenetic Analysis of the Argonaute Protein Family in Platyhelminths." *Molecular Phylogenetics and Evolution* 66 (3). Elsevier Inc.: 1050–54. doi:10.1016/j.ympev.2012.11.014.

APENDICE

Scripts usados para la búsqueda ortólogos de la vía de regulación mediada por de

ARNi.

Script para descarga de datos
#!/bin/sh/

```
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/hymenolepis_nana/PRJEB508/hymenolepis_nana.PRJEB508.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/hymenolepis_nana/PRJEB508/hymenolepis_nana.PRJEB508.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/mesocestoides_corti/PRJEB510/mesocestoides_corti.PRJEB510.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/mesocestoides_corti/PRJEB510/mesocestoides_corti.PRJEB510.WBPS1.proteins.fa.gz"
```

```

echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistocephalus_solidus/PRJEB527/schistocephalus_solidu
s.PRJEB527.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistocephalus_solidus/PRJEB527/schistocephalus_solidu
s.PRJEB527.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/spirometra_erinaceiuropeaei/PRJEB1202/spirometra_erina
ceiuropeaei.PRJEB1202.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/spirometra_erinaceiuropeaei/PRJEB1202/spirometra_erina
ceiuropeaei.PRJEB1202.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/taenia_asiatica/PRJEB532/taenia_asiatica.PRJEB532.WBP
S1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/taenia_asiatica/PRJEB532/taenia_asiatica.PRJEB532.WBP
S1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/taenia_solium/PRJNA170813/taenia_solium.PRJNA17081
3.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/taenia_solium/PRJNA170813/taenia_solium.PRJNA17081
3.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/echinostoma_caproni/PRJEB1207/echinostoma_caproni.P
RJEB1207.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/echinostoma_caproni/PRJEB1207/echinostoma_caproni.P
RJEB1207.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/fasciola_hepatica/PRJNA179522/fasciola_hepatica.PRJNA
179522.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/fasciola_hepatica/PRJNA179522/fasciola_hepatica.PRJNA
179522.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_curassoni/PRJEB519/schistosoma_curassoni.
PRJEB519.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_curassoni/PRJEB519/schistosoma_curassoni.
PRJEB519.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_haematobium/PRJNA78265/schistosoma_hae
matobium.PRJNA78265.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_haematobium/PRJNA78265/schistosoma_hae
matobium.PRJNA78265.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_japonicum/PRJEA34885/schistosoma_japonic
um.PRJEA34885.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_japonicum/PRJEA34885/schistosoma_japonic
um.PRJEA34885.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_mansoni/PRJEA36577/schistosoma_mansoni.
PRJEA36577.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_mansoni/PRJEA36577/schistosoma_mansoni.
PRJEA36577.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_margrebowiei/PRJEB522/schistosoma_margr
ebowiei.PRJEB522.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_margrebowiei/PRJEB522/schistosoma_margr
ebowiei.PRJEB522.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_mattheei/PRJEB523/schistosoma_mattheei.P
RJEB523.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_mattheei/PRJEB523/schistosoma_mattheei.P
RJEB523.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_rodhaini/PRJEB526/schistosoma_rodhaini.PR
JEB526.WBPS1.proteins.fa.gz"

```



```

time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schistosoma_rodhaini/PRJEB526/schistosoma_rodhaini.PR
JEB526.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/trichobilharzia_regenti/PRJEB4662/trichobilharzia_regenti.
PRJEB4662.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/trichobilharzia_regenti/PRJEB4662/trichobilharzia_regenti.
PRJEB4662.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/protopolystoma_xenopodis/PRJEB1201/protopolystoma_x
enopodis.PRJEB1201.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/protopolystoma_xenopodis/PRJEB1201/protopolystoma_x
enopodis.PRJEB1201.WBPS1.proteins.fa.gz
echo "wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schmidtea_mediterranea/PRJNA12585/schmidtea_mediterr
anea.PRJNA12585.WBPS1.proteins.fa.gz"
time wget
ftp://ftp.ebi.ac.uk/pub/databases/wormbase/parasite/releases/WBPS1/species/schmidtea_mediterranea/PRJNA12585/schmidtea_mediterr
anea.PRJNA12585.WBPS1.proteins.fa.gz

**Descompresión de archivos**
#!/bin/sh/
for x in $(cat lista)
do
gunzip $x
done
**Formateo de bases de datos**
#!/bin/sh/
for x in $(ls -l *.fa | awk '{print $9}')
do
formatdb -p T -i $x
done
**Comando para Blast Directo**
#!/bin/sh/
for x in $(ls -l *.fa | awk '{print $9}' | awk -F '!' '{print $1}')
do
blastall -p blastp -e 0.00001 -i query_CeIP.fas -d $x.fa -m 0 > ./blast_out/$x.al
blastall -p blastp -e 0.00001 -i query_CeIP.fas -d $x.fa -m 8 > ./blast_out/$x.tab
done
**Extracción de los hits obtenidos en platelmintos**
#!/bin/sh/
for x in $(ls -l *.fa | awk '{print $9}' | awk -F '!' '{print $1}')
do
grep -f $x.hits -A1 $x.fa | sed 's/--//' | sed '/^$/d' > ./hits_plathy/pl_$x
done
**Comando para Blast reverso**
#!/bin/sh/
for x in $(ls -l pl_* | awk '{print $9}' | awk -F '!' '{print $2}')
do
blastall -p blastp -b 1 -e 0.00001 -i pl_$x -d c_elegans.PRJNA13758.WS244.protein.fa -m 8 > rev_$x
done
**Extracción de las secuencias ortólogas de la vía de los proteomas de platelmintos**
#!/bin/sh/
for x in $(ls -l pl_* | awk '{print $9}' | awk -F '!' '{print $2}')
do
grep -f id_via rev_$x | awk '{print $1, $2}' | sort | uniq > r_list$x
done

for i in $(cat id_via)
do
for j in $(ls -l pl_* | awk '{print $9}' | awk -F '!' '{print $2}')
do
grep $i r_list$j >> $i.ort
done
done

for i in $(ls -l *.lista | awk '{print $9}' | awk -F '!' '{print $1}')
do
for j in $(cat $i.lista)

```

```
do
  grep '^$j\b' -A1 all_hits.fa | sed 's/--/' | sed '/^$/d' >> ./via_ort/$i.ort
done
done
```

Tabla Suplementaria 1. Posiciones genómicas de ortólogos de la vía de miARNs detectados en platelmintos

		Endoribonucleasas			Argonautas		Otros factores		
		DRSH-1	DCR-1	Dicer-like	Ago1	FL-AGO	TSN-1	PASH-1	XPO-1
Trematodos	<i>F. hepatica</i> (PRJNA17952 2)	Contig4134: 17154..78180: hebra-	Contig3056: 20589..97823: hebra-	Contig2023: 124743..148533: hebra- Contig2023: 24951..72310: hebra+	Contig812: 26102..39696: hebra+	Contig1642: 146..4638: hebra+ Contig3984: 10024..70497: hebra+	Contig6548: 10806..20454: hebra+	Contig283: 323337..373789: hebra-	Contig3369: 55388..89264: hebra-
	<i>F. hepatica</i> (PRJEB6687)	scaffold2953: 24160..79999: hebra-	scaffold1520: 3442..97224: hebra+	scaffold1234: 14666..52229: hebra+ scaffold1234: 97922..159914: hebra-	scaffold1192: 301232..348190: hebra+	scaffold5156: 149475..156723: hebra- scaffold668: 149409..189205: hebra-	scaffold375: 122739..134928: hebra+	scaffold319: 349707..405356: hebra-	scaffold2835: 114916..153841: hebra+
	<i>C. sinensis</i>	scf00068: 175047..217371: hebra-	scf01320: 690..51244: hebra+	scf00285: 428221..457732: hebra+ scf00285: 464265..496749: hebra-	scf01692: 50257..65388: hebra+	scf00024: 1048420..1070954: hebra- scf00684: 92276..170262: hebra-	scf00841: 104576..119495: hebra+	scf00546: 3956..23582: hebra+	scf00526: 4821..23172: hebra+
	<i>O. viverrini</i>	473: 148606..200673: hebra-	67: 1601303..1671562: hebra+	471: 76183..114479: hebra+ 471: 116762..147559: hebra-	35: 996642..1021277: hebra+	167: 138081..156907: hebra- 103: 810852..888417: hebra-	6: 1568195..158723: hebra-	40: 2047068..2056309: hebra-	178: 635265..647640: hebra-
	<i>S. haematobium</i>	scaffold566: 135922..177412: hebra+	scaffold62: 160262..205275: hebra+	scaffold1401: 28367..55973: hebra-	scaffold545: 323851..349867: hebra-	scaffold326: 102396..108410: hebra+ S000385: 179502..186929: hebra- S004191: 40055..55053: hebra-	scaffold385: 105665..115019: hebra-	scaffold1357: 114452..118381: hebra-	scaffold456: 173092..188317: hebra-
	<i>S. japonicum</i>	S000447: 16031..61961: hebra+	S000963: 106421..156556: hebra-	S000359: 34078..58210: hebra-	S000377: 166939..198021: hebra+	S003714: 11014..21005: hebra-	S000058: 185628..207163: hebra-	S000964: 67799..80663: hebra-	
<i>S. mansoni</i>	Chr_5: 1170439..1217340: hebra-	SC_0235: 69833..122456: hebra+	Chr_2.unplaced.SC_0108: 207465..243278: hebra-	Chr_7: 5784421..5808166: hebra-	SC_0250: 11056..17722: hebra+ SC_0250: 57859..64549: hebra+	Chr_3.unplaced.SC_220: 28895..38248: hebra+	Chr_4: 19342536..19366177: hebra-	Chr_1: 32483304..32495136: hebra-	
Cestodos	<i>E. granulosus</i>	scaffold_0012: 1983871..1991406: hebra+	scaffold_0015: 533365..545551: hebra+	scaffold_0008: 3294248..3310032: hebra-	scaffold_0011: 915440..925993: hebra+	scaffold_0025: 542807..545642: hebra- scaffold_0025: 553128..555507: hebra- scaffold_0026: 12999..17684: hebra-	scaffold_0002: 1298349..12932665: hebra-	scaffold_0017: 951529..957520: hebra-	scaffold_0031: 241664..248809: hebra-
	<i>E. multilocularis</i>	scaffold_01: 18893710..18901164: hebra+	scaffold_06: 6958478..6970628: hebra-	contig_62302: 3191820..3271153: hebra-	scaffold_09: 1957199..1967699: hebra+	scaffold_0077: 67: 865757..869041: hebra- scaffold_0077: 71: 6094..10633: hebra- scaffold_0077: 71: 20915..25481: hebra-	scaffold_02: 14077657..14081931: hebra-	scaffold_01: 16788123..16794086: hebra-	scaffold_04: 1613897..1620995: hebra-
	<i>T. solium</i>	contig_00130: 115212..122349: hebra-	contig_00416: 39951..53703: hebra+	contig_02859: 34..3444: hebra+ contig_01713: 6090..12876: hebra-	contig_00634: 735..10298: hebra-	contig_03097: 1406..4541: hebra- contig_02914: 3..4019: hebra- contig_03169: 3270..4306: hebra-	contig_00156: 56985..61279: hebra-	contig_00287: 85648..87221: hebra-	contig_00007: 322731..336585: hebra+
	<i>M. corti</i>	scaffold0000165: 36390..45851: hebra-	scaffold0000196: 44035..60090: hebra+	scaffold000003: 93: 14633..25483: hebra+	contig0000409: 4610..12701: hebra-	contig0000839: 285..12599: hebra+ contig0001006: 389..10264: hebra+ scaffold000005: 53: 26746..34179: hebra-	scaffold00010: 39: 6783..7274: hebra+ contig0002517: 16..1502: hebra+ scaffold000003: 56: 61009..60524: hebra-	scaffold00001: 76: 84209..89422: hebra+	scaffold00008: 40: 2516..7549: hebra+
	<i>H. microstoma</i>	scaffold_0001: 700214..710526: hebra+	scaffold_0006: 7609372..7629164: hebra-	scaffold_0006: 3562558..3583827: hebra-	scaffold_0008: 3633705..3646940: hebra+	scaffold_10: 196178..201057: hebra+ scaffold_354: 43753..52540: hebra+	scaffold_0001: 18509024..18512363: hebra+	scaffold_0001: 1493063..1508146: hebra-	scaffold_0004: 1242611..1245967: hebra+
Turbelario	<i>S. mediterranea</i>	v31.000125: 116128..129555: hebra-	v31.001736: 37414..61348: hebra+	v31.004816: 15251..20262: hebra-	v31.008596: 4..8604: hebra+ v31.036615: 1405..4660: hebra+	v31.000678: 44657..63988: hebra- v31.000678: 114641..132079: hebra-	v31.001005: 50039..53178: hebra+	v31.008104: 1428..5271: hebra- v31.013804: 2114..2489: hebra+	v31.000191: 46815..54482: hebra+

