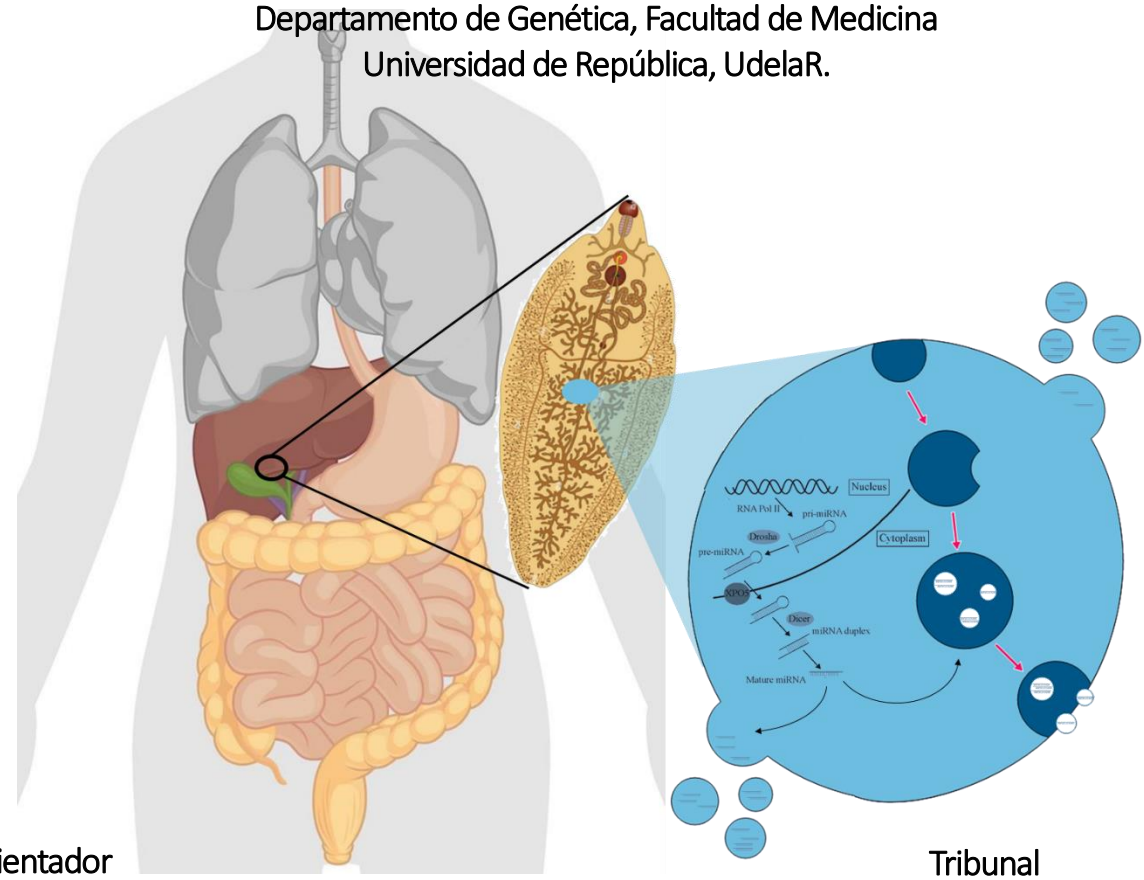
UNIVERSIDAD DE LA REPUBLICA
PROGRAMA DE DESARROLLO DE LA CIENCIAS BASICAS

AREA BIOLOGIA

Tesis presentada para acceder al Doctorado en Ciencias Biológicas

Rol de los ARN pequeños de *Fasciola hepatica* en la interacción con el huésped y la conservación de las vías de ARN pequeños en platelmintos

Msc. Santiago Fontenla Martínez
Departamento de Genética, Facultad de Medicina
Universidad de República, UdelaR.



Orientador
Dr. José F. Tort

Tribunal
Dr. Alfonso Cayota
Dra. Marcela Cucher
Dra. Luisa Berná

Agradecimientos

Al concluir esta etapa de mi vida quiero agradecer primero que nada a mi familia por su apoyo, a mis viejos, sobre todo a mi madre, a mi abuela, a mi hermana, a mis sobrinos (Bruno y Lucas) y a mi cuñado. Y no me puedo olvidar de mi familia de cuatro patas que tanto me alegra la vida.

También tengo que agradecer a Pepe Tort por haberme enseñado tanto y brindarme su apoyo a lo largo de todos estos años, siendo primero mi tutor de tesis de maestría y luego de mi tesis de doctorado. A todos los compañeros con los que he trabajado en todo este tiempo, especialmente a Fer Domínguez, por todo su trabajo y experiencia en la mesada y por guiarme en mi reciente 'vuelta' a la mesada; a Gabriel Rinaldi, que ha empujado y contribuido grandemente a mi estudio de las vías de ARN pequeño; a Alicia Costabile, con quien trabajé en varios proyectos; a Santi Radio, Pablo Smircich y al 'Pantera' Lamolle.

Al Dr. 'Gui' Correa de Oliveira y su grupo, por haberme permitido realizar una pasantía en su laboratorio durante la maestría.

También quiero agradecer a los docentes y las instituciones que me han formado a lo largo de mi carrera.

A los organismos e instituciones financiadoras: ANII, CSIC, PEDECIBA y Facultad de Medicina. A las instituciones que me becaron en cursos de corta duración: Instituto Pasteur de Montevideo, Universidad de Buenos Aires, CELFI, CABBIO, entre otras.

A mis demás compañeros del Departamento de Genética de la Facultad de Medicina, por todo el tiempo compartido en el día a día.

Abreviaturas

Ago	Argonauta
ARNdc	ARN doble cadena
ARNpq	ARN pequeños
EVs	Vesículas extracelulares
FBT	trematodos de transmisión alimentaria
IL	Interleuquina
mARN	ARN mensajero
miARN	micro ARN
ncARN	ARN no-codificante
NEJ	Juvenil recientemente desenquistado
NTD	Enfermedades tropicales desatendidas
piARN	ARN piwi
rARN	ARN ribosomal
RdRP	ARN polimerasa dependiente de ARN
RI	Respuesta inmune
RISC	RNA-induced silencing complex
siARN	ARN pequeño interferente
snARN	ARN pequeño nuclear
tARN	ARN de transferencia
tRDF	fragmentos derivados de tARNs
3´UTR	Region no-traducida 3´

Contenido

Resumen

Los platelmintos constituyen uno de los principales filos de animales invertebrados, habitan una amplia gama de ecosistemas y son uno de los filos más exitosos en adaptarse a la vida parasitaria. Los ARN pequeños no codificantes (ncARNs) han sido implicados en la regulación de complejas transiciones del desarrollo en especies parasitarias modelos. Estudios recientes han demostrado que los miARNs de los parásitos pueden liberarse al medio del huésped regulando la expresión de sus genes y, más aún, se han detectado en la circulación de animales infectados.

En este trabajo estudiamos la expresión de ncARNs contenidos en vesículas extracelulares y el patrón de expresión de pequeños ARNs a través de los estadios intra-mamífero del ciclo de vida del platelminto parásito *Fasciola hepatica*, que junto con *F. gigantica* son los agentes causantes de la Fascioliasis, una de las enfermedades helmínticas del ganado más extendida.

Generamos un pipeline para la predicción de genes diana putativos que informa solo los sitios conservados entre tres herramientas de predicción diferentes. Este pipeline se probó contra una base de datos curada de regiones 3´UTR de genes de *F. hepatica* para detectar mecanismos regulados. Hallamos que los genes blancos de los miARNs diferencialmente expresados entre estadios estaban vinculados a la regulación del desarrollo y a la invasión. Entre estos últimos, se destaca la posible regulación de la liberación de vesículas entre los estadios metacercaria y juvenil. Además, para evaluar el papel putativo de los miARNs secretados por el parásito en la interacción con el huésped, predijimos blancos en las regiones 3´ UTR de *Homo sapiens*. Hallamos que varias funciones relacionadas con la inmuno-respuesta y otras vías de señalización relevantes están enriquecidas entre los blancos de los miARNs parásitos más expresados, destacando que podrían ser actores relevantes para establecer y mantener la infección.

A su vez, utilizando datos transcriptómicos y genómicos, analizamos, contrastamos y comparamos la conservación de las vías de ARN pequeños entre varias especies de vida libre (un grupo parafilético tradicionalmente conocido como 'turbelarios') y especies parasitarias

(organizadas en el clado monofilético Neodermata), para desentrañar posibles adaptaciones en la transición al parasitismo. Nuestros hallazgos mostraron que las vías completas de miARN y ARNi están presentes en todos los gusanos planos de vida libre analizados. Sorprendentemente, mientras que todos los "turbelarios" tienen proteínas Piwi, éstas se perdieron en los neodermatas parásitos. Además, dos grupos de genes Argonauta de la clase Piwi están presentes en todos los "turbelarios". Curiosamente, identificamos una clase Argonauta-Piwi divergente en gusanos planos de vida libre exclusivamente, a la que llamamos "Fliwi". Adicionalmente, otras proteínas clave de las vías Piwi están conservadas en "turbellarios", mientras que ninguna de ellas se detectó en Neodermatas.

En conclusión, los resultados muestran evidencia de posibles adaptaciones notables en la transición al parasitismo, y arrojan luz sobre un mecanismo aún incipientemente estudiado, pero potencialmente relevante, de interacción huésped-parásito.

Introducción

1. Los parásitos helmintos

1.1. Los platelmintos: problema de salud pública

Con 30.000 especies descriptas, y posiblemente muchas más por ser descubiertas, los platelmintos comprenden uno de los mayores filos de animales invertebrados. Habitando un vasto rango de ecosistemas, desde formas microscópicas viviendo sobre y entre los granos de arena hasta gusanos multicoloridos en los arrecifes de coral, las formas de vida libre están en todos lados, pero son raramente obvias. Por otro lado, el filo platelminto es uno de los más exitosos en adaptarse al parasitismo. A diferencia de lo que sucede en nematodos, entre los platelmintos el parasitismo obligado surgió solo una vez, en el clado de los neodermata, siendo todas las especies parásitos de animales, típicamente con un huésped intermedio invertebrado y uno definitivo vertebrado (Zarowiecki & Berriman, 2015)(Figura 1). El clado Neodermata está dividido en las clases: Monogenea, Cestoda y Trematoda. Tanto las tenias (clase Cestoda) como los Trematodos producen un impacto importante para la salud humana y del ganado.

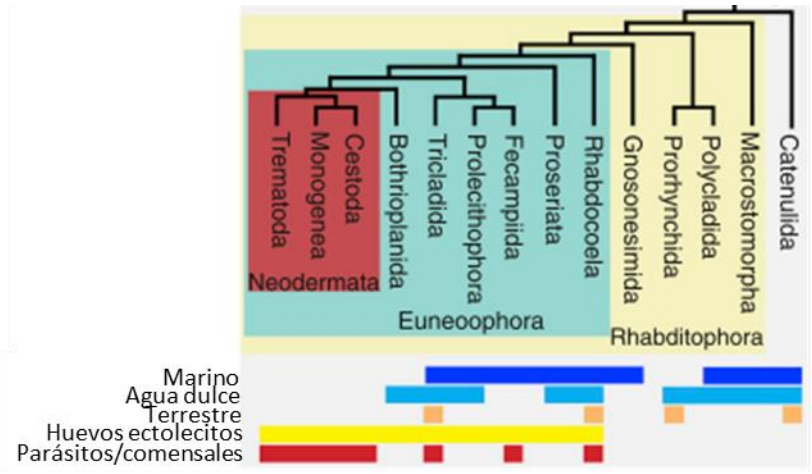


Figura 1. Árbol filogenético con la relación en el filo platelminto. *Las formas de vida libre o 'turbelaria' habitan ecosistemas diversos. El clado Neodermata está compuesto por parásitos obligados y está dividido en las clases: Monogenea, Cestoda y Trematoda. (Tomado y modificado de* (Collins, 2017)).

La schistosomiasis, causada por trematodo *Schistosoma*, es la helmintiasis humana causada por platelmintos más común, siendo endémica en 70 países, y afecta a más de 190 millones de personas (King, 2019; Robinson & Dalton, 2009). Le siguen las trematodiasis alimentarias, causadas por *Opisthorchis*, *Clonorchis*, *Fasciola* and *Paragonimus,* con 75 millones de casos, y si bien no se tienen cifras claras sobre la prevalencia de las cestodiasis, se estima que la neurocisticercosis sintomática o asintomática (causada por *Taenia solium*) afecta a 2.56-8.30 millones de personas a nivel mundial (https://www.who.int/publications/i/item/9789240032231). Las helmintiasis en general pertenecen al grupo de las enfermedades tropicales desatendidas (NTDs) y están directamente relacionadas a la pobreza. Las infecciones están ampliamente distribuidas en áreas tropicales y subtropicales, con un mayor número de infectados en los países en desarrollo del África subsahariana, las Américas y el este de Asia (Figura 2). Las NTDs causan discapacidades físicas severas, sufrimiento crónico y enormes pérdidas económicas, contribuyendo así a un ciclo arraigado de pobreza.

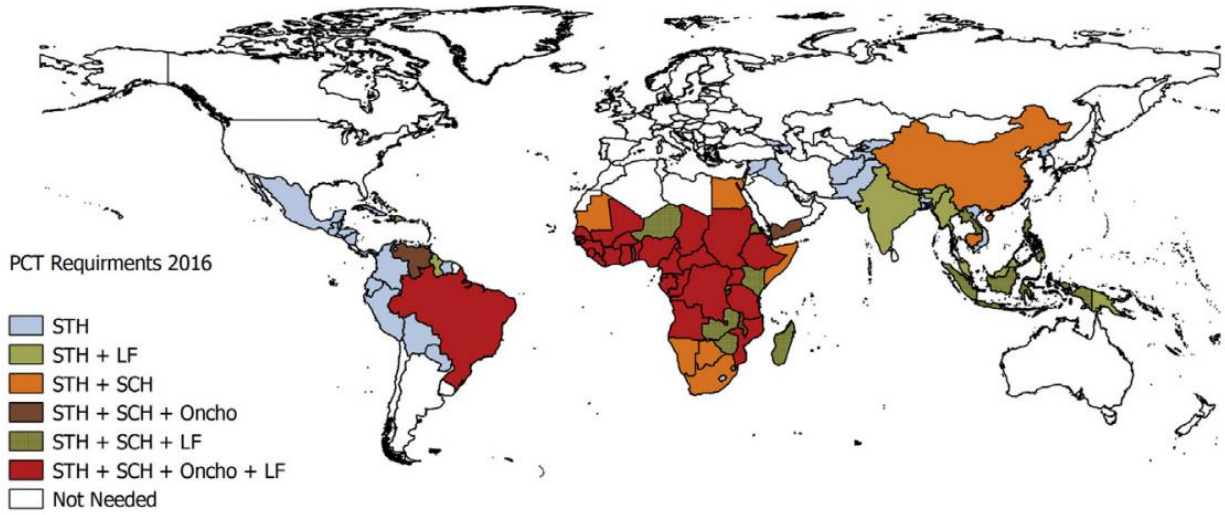


Figura 2. Distribución mundial de las helmintiasis. *Helmintiasis transmitidas por el suelo (STH), filariasis linfática (LF), oncocerquiasis (Oncho) y schistosomiasis (SCH). Tomado de* (King, 2019).

1.2. El impacto económico causado por F. hepatica

Fasciola hepatica pertenece a la clase Trematoda del phylum platelmintos y junto a *F. gigantica* producen la fasciolosis, una zoonosis de distribución mundial. Ambas especies se superponen en varias áreas de África y Asia, mientras que *F. hepatica* se distribuye predominantemente en áreas templadas y es el agente causante de la enfermedad en las Américas, Europa y Oceanía (Figura 3) (Mas-Coma et al., 2005). La fasciolosis afecta fundamentalmente al ganado productivo resultando en pérdidas estimadas en 3000 millones de USD anuales (Spithill et al., 1999). En la mayoría de los países Sudamericanos la fascioliasis es un problema productivo relevante, siendo endémica en muchas áreas dedicadas a la cría de ganado bovino y ovino, con varias otras especies oficiando de reservorio (Carmona & Tort, 2017). En Uruguay, constituye una de las parasitosis más relevantes, con presencia en todo el territorio nacional y una prevalencia promedio de 33.9% en ganado bovino (da Costa et al., 2019). Según datos de la Auditoria de Calidad de la Cadena Cárnica de 2007-08 se pierden anualmente 6 millones de dólares solo a causa del decomiso de hígado. A esto se le debe sumar las pérdidas directas (por disminución de peso, disminución de la fertilidad, calidad de la carne, lana o leche) e indirectas (por los costos de tratamiento).

A pesar de su impacto, el carácter crónico y cuasi asintomático de la enfermedad ha llevado a que casi no se desarrollen políticas de control ni se inviertan recursos en el desarrollo de nuevas terapias. A su vez, si bien la enfermedad es tratable con antihelmínticos como el triclabendazol, ya han sido descriptas cepas resistentes a esta droga en varios países como Australia, Irlanda, Reino Unido, Países Bajos, España y Uruguay (Kelley et al., 2016; Canevari et al., 2013).

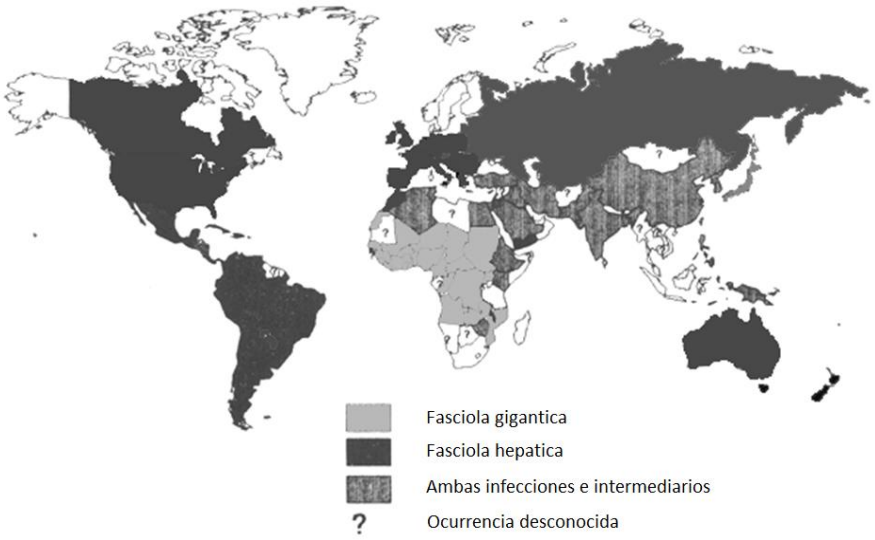


Figura 3. Distribución de Fasciola spp. *F. hepatica se distribuye predominantemente en áreas templadas mientras que F. gigantica predomina en áreas tropicales de África y Asia. Tomado y modificado de http://www.infectionlandscapes.org/2012/07/liver-flukes-part-2-fascioliasis.html.*

1.3. 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 huésped 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 huésped intermediario (Figura 4). Dentro del caracol los miracidios se transforman en esporocistos, y estos, por amplificación asexual dan lugar a redias (una forma intermedia, que permanece parásita del caracol) y cercarias. Las cercarias se liberan del caracol y se enquistan en plantas acuáticas. Estas formas quísticas de resistencia denominadas metacercarias, son las que infectan al huésped definitivo al ser ingeridas junto a la vegetación (o vegetales infectados en el caso humano) (Keiser & Utzinger, 2005).

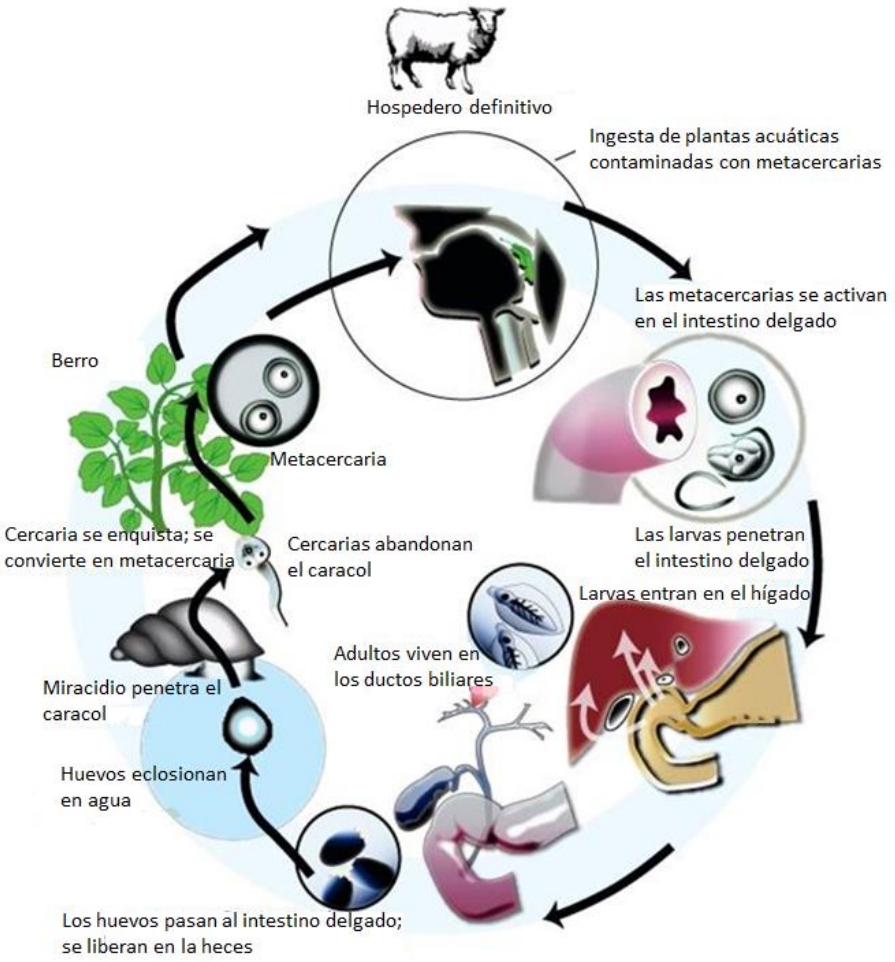


Figura 4. Ciclo de vida de F. hepatica. *Tomado y modificado de* (Alatoon et al., 2008).

Una vez en el tubo digestivo del huésped 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 ingles). 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.

Desde el punto de vista clínico, la invasión del huésped definitivo está marcada por eosinofilia, dolor abdominal y fiebre, además de cambios en las concentraciones de proteínas del plasma (albumina, globulina) e incrementos en los niveles de las enzimas hepáticas en sangre. Aparte

de esta forma aguda de la enfermedad, una etapa crónica, cuando el parásito está establecido en los ductos biliares, se caracteriza por obstrucciones intermitentes, causando síntomas que asemejan cólicos biliares y colecistitis (Garcia et al., 2007).

1.4. Variación de la expresión génica a lo largo del ciclo de vida

Durante la invasión el parásito se enfrenta a diferentes ambientes, trabajos de nuestro grupo y otros más recientes han demostrado que se producen cambios en la expresión génica ya en las primeras horas de la invasión (Cancela et al., 2008, 2010) y a lo largo de los distintos estadios posteriores (Cwiklinski et al., 2018). Un claro ejemplo es la expresión diferencial las catepsinas, que pertenecen a la familia de las cisteínproteasas similares a papaína y representan las proteasas más abundantes en la secreción de *F. hepatica*. A pesar de que dependen de solo dos clases de catepsinas [catepsina L (CL) y catepsina B (CB)], estas se han expandido y divergido formando familias multigénicas. En una publicación reciente identificamos 14 secuencias CL y 14 secuencias CB en el genoma de *F. hepatica* (Choi et al., 2020, ver ANEXO II). A su vez, algunas de estas peptidasas se expresan preferencialmente en estadios específicos. Resumidamente, CL3, CL4, CB1, CB2 y CB3 se expresan altamente en metacercaria y juvenil (Figura 5).

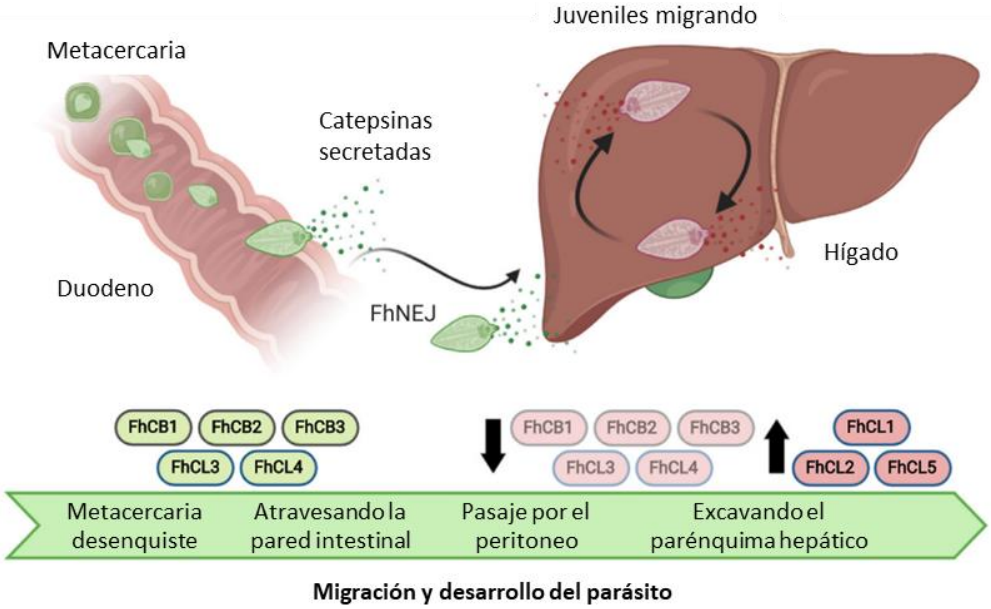


Figura 5. Regulación estadio específica de las catepsinas de F. hepatica. *(Tomado y modificado de* Gonzáles -Miguel et al., 2020*)*

Cuando el gusano inmaduro ingresa al hígado la expresión de las catepsinas mencionadas cae y la expresión de otras CL (principalmente CL1) aumenta ((McNulty et al., 2017), ver ANEXO II). Estos cambios están correlacionados con el desarrollo del parásito y con el pasaje por los tejidos del hospedero. Específicamente, la habilidad de CL3 para digerir colágeno o las propiedades bioquímicas de CB1 para resistir la inhibición por cistatinas del huésped, junto con el momento de mayor expresión, soportan el rol de estas peptidasas en la migración del parásito a través de la pared intestinal. Por lo tanto, la expresión génica a lo largo del ciclo de vida de *F. hepatica* debe estar finamente regulado para poder responder a cambios en el desarrollo y el ambiente.

2. Los ARN pequeños (ARNpq) como reguladores generales de la transcripción

En el silenciamiento génico mediado por ARN pequeños (ARNpq) los principales mediadores de la regulación de la expresión génica son ARNs de 20-30 nucleótidos de los cuales los más conocidos son los microARNs (miARNs), los ARN pequeños interferentes (siARNs) y los ARNs de unión a proteínas Piwi (piARNs).

Otros grupos de ARN pequeños con capacidad de regular la expresión génica han sido identificados recientemente, como los tARN halves derivados de tARNs (Thompson & Parker, 2009), heterochormatin small interfering RNA (hc-siARN), trans acting siRNA (ta-siARN), natural antisense short interfering RNA (nat-siARN), DNA damage-induced small RNA (qiARN) (Choudhuri, 2009) y los vault ARN (vtARN)(Frascotti et al., 2021). 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. Los miARNs pueden regular el desarrollo

Los miARNs constituyen una clase de ARN no-codificante simple hebra de 21 a 25 nucleótidos de largo. Fueron inicialmente descriptos por su papel regulador en el desarrollo. Por ejemplo, el primer miARN descripto, lin-4, es esencial en el pasaje del estadio larvario 1 (L1) a L2 del nematodo *Caenorhabditis elegans* (Lee et al., 1993). En cambio, let-7, el segundo miARN en ser

descripto, es esencial en el pasaje del estadio L4 a adulto (Reinhart et al., 2000). Posteriormente, los miARNs fueron descriptos en otros organismos modelo, como Bantam y mir-14 en *Drosophila*, cuya pérdida produce letalidad, así como defectos en el desarrollo corporal. Estos primeros miARNs fueron descubiertos en cribados para identificar genes vinculados al desarrollo, debido a que su pérdida afecta dramáticamente la progresión del estadio larvario. Estos miARNs se expresan en muchos, si no en todos, los tejidos, y afectan comportamientos esenciales de las células como la división, diferenciación o muerte. En un segundo grupo, se encuentran miARNs como lsy-6 en *C. elegans,* que es expresado en una sola neurona sensorial y es esencial para la especificación de esa sola neurona. Análogamente, mir-279 de *Drosophila* y mir-96 en el ratón también han sido implicados en contextos sensoriales muy específicos, como el censado de CO2 y la audición, respectivamente (revisado por (Alberti & Cochella, 2017)).

En el caso de los platelmintos, la primera descripción de los miARNs se hizo en el gusano de vida libre *Schimidtea mediterranea* (Palakodeti et al., 2006). La conservación de miARNs en platelmintos fue confirmada en reportes posteriores que apuntaron a estudiar el rol de estos en la regeneración por la expresión en células madre (también llamadas neoblastos) o la expresión diferencial en tejidos específicos en estos organismos (Cao et al., 2020; Friedländer et al., 2009; González-Estévez et al., 2009; Lu et al., 2009; Sasidharan et al., 2013; H. Liu et al., 2021). A su vez, los miARNs han sido descriptos en todas las clases de platelmintos parásitos (revisado por (P. Cai et al., 2016)). Durante la tesis de maestría contribuimos a este conocimiento estudiando la presencia de ARNpq reguladores en *F. hepatica* (Fontenla et al., 2015). Puesto que existían datos muy parciales de presencia de miARNs en el estadio adulto (Xu et al., 2012), nos propusimos analizar su presencia en los primeros pasos de la invasión en el huésped definitivo. Generamos datos de ARNpqs del juvenil recientemente desenquistado que nos permitió expandir a 20 familias el miRNoma de *F. hepatica* y la descripción de 5 familias nuevas específicas. A su vez, aunque ampliamos el set de miARNs descriptos nos encontramos con que varios miARNs altamente conservados en todos los animales y otros conservados en el supergrupo de los Lophotrochozoa no están presentes ni en *F. hepatica* ni en otros platelmintos parásitos (Figura 6).

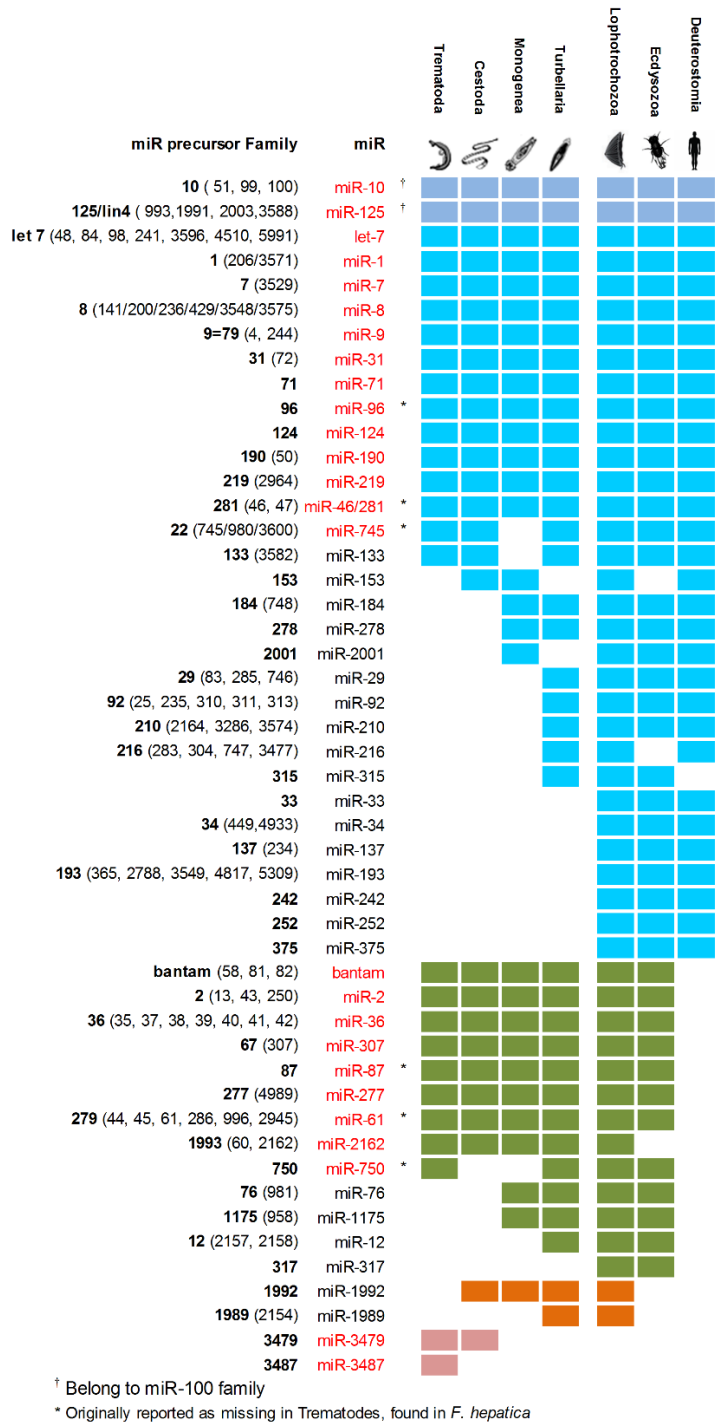


Figura 6. Conservación de familias de miARNs en platelmintos comparados con otros metazoarios. *El set de datos de F. hepatica fue comparado con aquellos presentes en otros linajes y fue incluido en el dataset 'Trematodes'. Los miARNs descriptos en el estadio juvenil se indican en letras rojas. Las perdidas de familias génicas se indican como bloques vacíos. Inferencias sobre la conservación fueron tomadas de la literatura y confirmadas por búsquedas en las bases de datos miRBase y mirOrtho. (tomado de* Fontenla et al. 2015*).*

Por otro lado, el rol regulador que juegan los miARNs en el desarrollo de *Fasciola spp.* ha sido muy escasamente estudiados hasta el momento, con un único reporte reciente en la especie hermana *F. gigantica* (R. S. Hu et al., 2021).

3. *ARN pequeños contenidos en vesículas extracelulares*

Las vesículas extracelulares (EVs) son pequeñas vesículas cubiertas por membrana secretadas por prácticamente todas las células para facilitar la transferencia de carga, principalmente lípidos, proteínas y especies de ARN. Las EVs se pueden dividir en términos generales en dos categorías principales: exosomas y microvesículas. Los exosomas tienen un origen endosómico ya que se forman por invaginación de la membrana endosómica y miden entre 30 y 100 nm de diámetro; las microvesículas se generan por la brotación y la fisión hacia afuera de la membrana plasmática y varían entre 50 y 1000 nm (van Niel et al., 2018)(Figura 7).

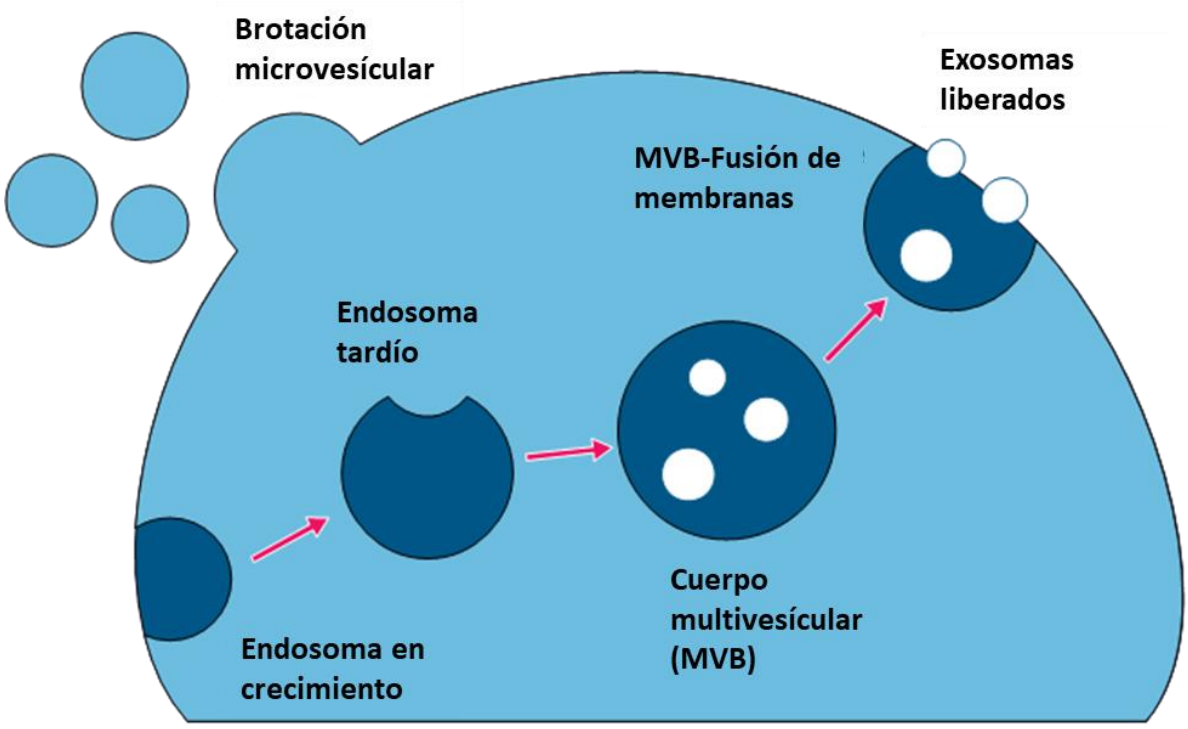


Figura 7. Biogénesis de las vesículas extracelulares. *Las EVs se pueden dividir en exosomas y microvesículas. Los exosomas tienen un origen endosómico ya que se forman por invaginación de la membrana endosómica; las microvesículas se generan por la brotación y la fisión hacia afuera de la membrana plasmática. Tomado y modificado de https://www.echelon-inc.com/exosomes-and-extracellular-vesicles-methods-and-applications/.*

Los miARNs pueden ser liberados al medio extracelular en EVs o en complejo con proteínas (Ortiz-Quintero, 2016). La secreción de EV se describió inicialmente como un medio para eliminar compuestos innecesarios de la célula (Johnstone et al., 1987). Posteriormente, trabajos pioneros reportaron la carga de mRNAs y miARNs dentro de las EVs de células de mamíferos (Mitchell et al., 2008; Valadi et al., 2007). Casi de inmediato, se identificó el potencial de estos miARN circulantes en suero como biomarcadores tempranos de cáncer (Chin & Slack, 2008). Subsecuentemente se han ido detectando miARNs contenidos en EVs en muchos organismos, con evidencia creciente sobre su papel en la regulación de la expresión génica remota (Bayraktar et al., 2017). En helmintos, los miARNs contenidos en los EV se informaron por primera vez en el trematodo *Dicrocoelium dendriticum* (Bernal et al., 2014), seguidos pronto por hallazgos similares en varios nematodos y en cestodos o trematodos, incluido *F. hepatica* (revisado por Sotillo et al., 2020).

4. Modulación de la expresión génica del huésped mamífero

Actualmente se sabe que la respuesta inmune (RI) innata tiene un rol directo frente a la infección por helmintos. Los helmintos han desarrollado estrategias para modular la RI asegurando su supervivencia. Los efectos tempranos de inmunomodulación juegan un rol esencial en el establecimiento y desarrollo del parásito y es llevado a cabo directamente por moléculas secretadas por el parásito y por el huésped que activan mecanismos de reparación (Maizels et al., 2018).

La primera interacción de las células T CD4 naive con antígenos específicos presentes en las células presentadoras de antígenos promueve su diferenciación. Dependiendo del contexto las células T naive se diferenciarán en los linajes Th1, Th2, Th17 o células T regulatorias (Treg). Cada linaje está especificado por factores de transcripción particulares y, a su vez, producen perfiles de citoquinas distintos (Figura 8).

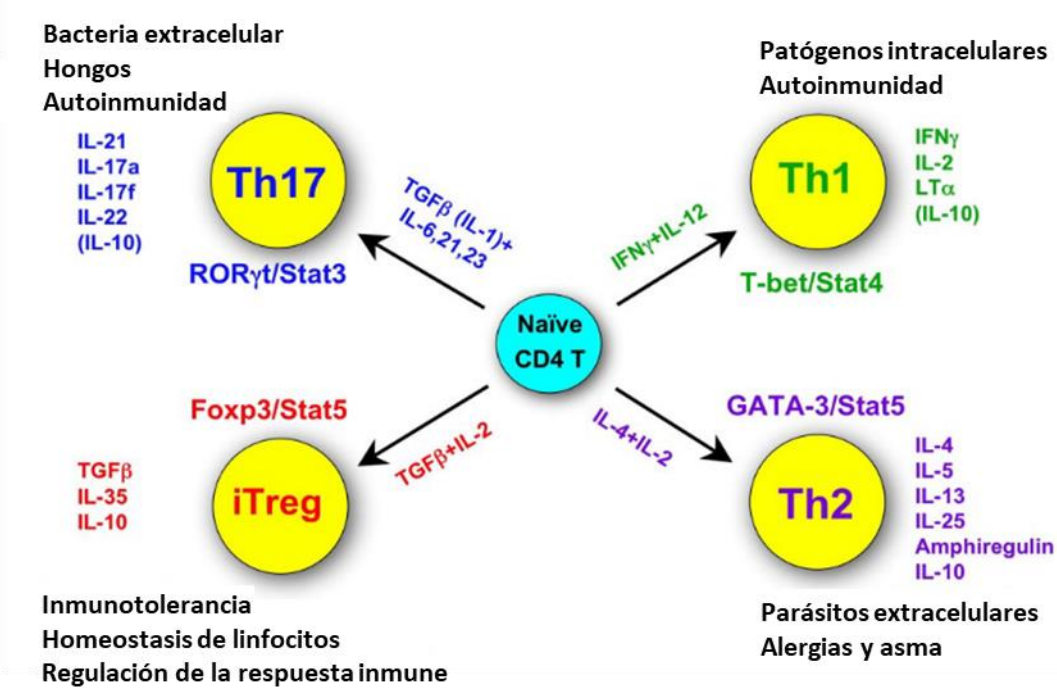


Figura 8. Distintos linajes de las células T CD4 en respuesta a distintos agentes patógenos. *Los helmintos inducen una respuesta de tipo Th2. El factor de transcripción GATA-3 induce la diferenciación en el linaje Th2 en respuesta a IL-4. Las células Th2 producen citoquinas específicas como IL-4, IL-5 e IL-13 (tomado y modificado de* (J. Zhu & Paul, 2008)*).*

Las citoquinas producidas por las células Th2 están involucradas en la respuesta a parásitos extracelulares como los helmintos. La respuesta Th2 también induce una respuesta protectora en el huésped al reducir el daño por la respuesta inmune. Esta respuesta evita, por ejemplo, que se genere un daño severo en respuesta a huevos de *Schistosoma* atrapados en la pared intestinal de animales infectados (Fairfax et al., 2012). A su vez, los helmintos también suprimen la respuesta Th1, aumentando la susceptibilidad a otros patógenos. Esta supresión de la respuesta Th1 está posiblemente asociada a la predisposición a las infecciones bacterianas en el ganado infectado por *F. hepatica* (Mabbott, 2018). En la etapa crónica, los helmintos modulan la respuesta inmune para inducir la expansión de células regulatorias (mayoritariamente Treg) y suprimir la respuesta inmune Th2. De esta manera el parásito pasa a ser tolerado por el sistema inmune como un organismo no dañino o "comensal" (Maizels et al., 2018). La inducción de tolerancia inmunológica está mediada entre otros por agonistas de los receptores de TGF-β. Esta podría, por ejemplo, ser la razón por detrás del declive en la

respuesta Th2 a la schistosomiasis luego del pico a las 9 semanas de infección (Dunne & Cooke, 2005)(Figura 9).

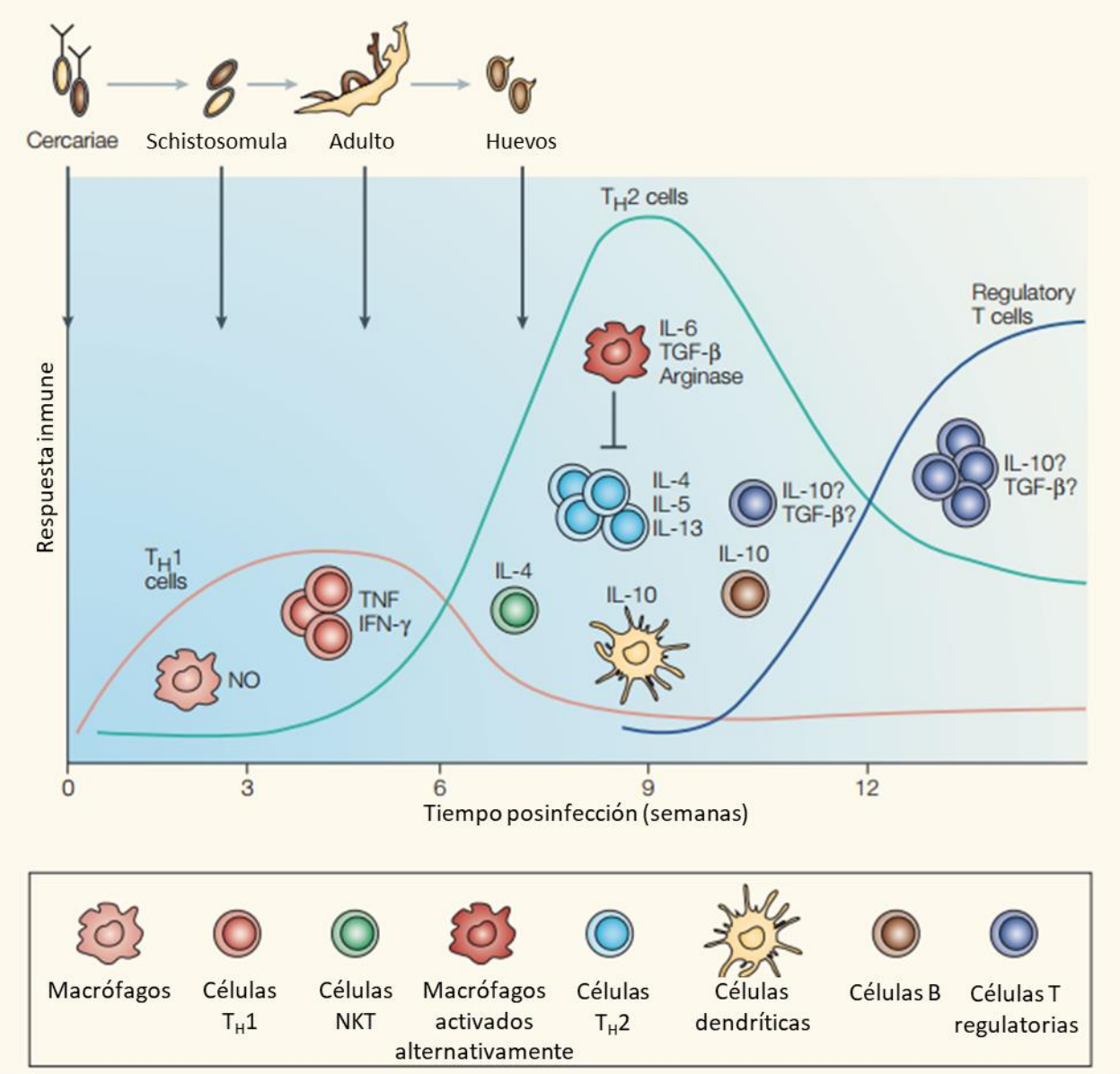


Figura 9. Respuesta inmune a lo largo de la infección por S. mansoni. *Luego de la infección la respuesta inicial desarrolla una respuesta Th1. A medida que el gusano se desarrolla y se depositan los huevos, se activan las células natural killer (NKT), las células dendríticas producen más IL-10 y se pasa a una respuesta de tipo Th2. Posteriormente, la población de macrófagos alternativamente activados y células T regulatorias aumenta llegando a la tolerancia del sistema inmune. Tomado de* (Dunne & Cooke, 2005).

El amplio éxito de los helmintos se basa en el desarrollo de estrategias biológicas complejas que involucran diferentes factores a lo largo de milenios de coevolución con sus huéspedes. Si bien

los estudios iniciales se han centrado en gran medida en las proteínas, glicoproteínas y lípidos como posibles inmunomoduladores, el descubrimiento de los miARN secretados por el parásito reveló nuevos mediadores en la interacción huésped-parásito.

4.1. Modulación mediada por miRNAs parasitarios contenidos en EVs

El papel de las EVs de origen parasitario y su rol en la interacción e inmunomodulación del huésped han sido propuestos, tanto en patógenos intracelulares, como protozoarios parásitos, hasta patógenos extracelulares, como los hongos o helmintos (Coakley et al., 2015). En helmintos, ya desde 2012 se reportó la internalización activa de EVs de *Echinostoma caproni* por células del huésped cultivadas *in vitro* (Marcilla et al., 2012) y, en 2014, se describió por primera vez el potencial inmunomodulador de los miARNs contenidos en EVs del nematodo gastrointestinal *Heligmosomoides polygyrus* (Buck et al., 2014). La supresión de los principales genes blanco de *H. polygyrus* en el huésped, *dusp1* e *Il33r*, conduce a la disminución de la respuesta inmune tipo 2, sin embargo, solo la supresión de *dusp1* parece estar mediada por miARNs. Desde entonces otros estudios han contribuido a describir el potencial regulador de lo miARNs producidos por parásitos helmintos en la comunicación e inmunomodulación del huésped, la mayoría publicados en los últimos 3 años (ver Tabla 1; la versión completa de la Tabla 1 está disponible en: https://docs.google.com/spreadsheets/d/1HuS2NYUk6IrOaCOFGSRa3yaA4iC9bkSG/edit?usp=sharing&ouid=111873225318883238450&rtpof=true&sd=true). Sin embargo, en algunos reportes se analizó el efecto modulador luego del tratamiento con EVs enteros aislados del parásito (marcados con '*' en la columna "Effect"), por lo que, es posible que otras moléculas además de los miARNs participen modulando la interacción con el huésped. Los genes que varían su expresión como resultado de la interacción directa con los miARNs parasitarios fueron resaltados en rojo en la columna "Gene effect".

Tabla 1. Función propuesta de los miARNs secretados por helmintos durante la infección

			Parásito	Estadio	miRNAs en EVs (s/n)	miRNAs más abundantes	miRNAs de interés	Efecto génico	Herramienta para predicción de blancos	Vía afectada	Efecto	Referencia
Nemathelminthes	Clades	I	*Trichinella spiralis*	Muscle larvae	S	miR-540-3p miR-8325-3p miR-1692 miR-10833 miR-2904	--	↓IL-1β ↓TNF-α ↓INF-γ ↓IL-17A ↑IL-10 ↑TGF-β ↑IL-4 ↑IL-13	miRanda TargetScan	Ts-EVs disminuyen la expresión de citoquinas proinflamatorias -Incrementan las citoquinas anti-inflamatorias	Reprime la respuesta proinflamatoria Th1/Th17.* Induce una respuesta Th2 y Treg.*	Yang et al. 2020
			Trichinella spiralis	Adulto	S	miR-51 miR-80 Tsp-miR-78	miR-31	Myf5	--	--	Represión de la miogénsis	Taylor et al. 2020
				Larvae	mayoria libre	miR-80 let-7 Tsp-miR-41						
			Trichuris muris	Adulto	S	miR-1175-3p miR-5360-5p miR-81b-3p miR-72-5p miR-44a-3p	56 miRNAs	-1791 genes blanco en ratón	miRanda	--	-Modulación del sistema inmune -Regulación de receptores -Regulación de la transcripción	Eichenberger et al. 2018
		III	*Ascaris suum*	L3 larva	S	asu-mir-100a-1	miR-1175-5p	C1QA	TargetScan PITA	-Proliferación y activación de células T -Producción de citoquinas y quimioquinas	Supresión de la respuesta inmune	Hansen et al. 2019
						asu-mir-100a-2	miR-5361-5p	CD80				
						asu-mir-5358a	miR-81					
						asu-mir-71	miR-9-3p					
						asu-mir-5359	miR-71-5p	IFNGR1				
				L4 larva		asu-mir-9	miR-57-5p	SLA-DOB				
						asu-mir-5358a asu-mir-81a	lin-4-5p	IL-13				
						asu-mir-100a-1	let-7-5p					
						asu-mir-100a-2		CD86				
				Adulto		asu-mir-71	miR-5358b-5p					
						asu-mir-81a	miR-5359-5p					
						asu-mir-5358a	miR-9-5p	TLR7				
						asu-mir-9	miR-34-5p	XCR1				
						asu-mir-100a-1	miR-87-3p	IL-33				
			Brugia malayi	L3 larva	S	lin-4 miR-71 miR-1 miR-92 miR-50	--	↑MCP-1 ↑MIP-2 ↑G-CSF ↑IL-6	--	--	Activación M1 de macrófagos*	Zamanian et al. 2015
			Brugia malayi	Microfilaria	S	--	miR-100 miR-99a miR-7	mTOR	miRDB	vía mTOR	Regulación a menos de la fosforilación por mTOR*	Ricciardi et al. 2021
							miR-34a miR-9	Ras				
							miR-31 miR-4299 miR-92a	PI3K				
							let-7	eIF-4E				
							miR-71	PDK1				
		V	*Heligmosomoides polygyrus*	Adulto	S	Bantam-b Bantam-c miR-193 miR-10 miR-200	miR-200 let-7 miR-425	↓DUSP1 ↓IL33R	--	-Señalización MAPK -Señalización Alarmina	Supresión de respuesta inmune tipo 2	Buck et al. 2014
			Nippostrongylus brasiliensis	Adulto	S	miR-81b-3p miR-56 miR-71-5p miR-36a-3p miR-2a-3p	--	↓IL-1β ↓IL-6 ↓INF-γ ↓IL-17a ↑IL-10	miRanda	Inflamación por vías de señalización de quimioquinas y citoquinas	Supresión de citoquinas pro-inflamatorias*	Eichenberger et al. 2018

Continuación de Tabla 1.

Platyhelminthes	Trematodes	Schistosomidae	*Schistosoma japonicum*	Adulto	S	miR-125b miR-61 miR-277b Bantam miR-3505	miR-125b mimics Bantam mimics	↓Pros1 ↓F11r ↓Fam212b ↓Clmp ↓Prtg	TargetScan miRanda	-Señalización TNF -Receptor Toll-like (TLR) -Interacción citoquina-receptor -Señalización RAP1	↑ Producción TNF-α ↑ monocito ↑ macrófagos ↑ Supervivencia parásito	Liu et al. 2019
			Schistosoma japonicum	--	S	--	miR-1 mimic	↓Sfrp1 ↑α-Sma ↑Col1α1 ↑Col3α1	miRDB	-Activación de Wnt/β-Catenin	Activación de HSCs produce fibrosis hepática	Wang Y. et al. 2020
			Schistosoma japonicum	--	N	--	miR-7-5p mimic	↓SKP2 ↑P27 (CDKN1B) ↓MMP9	miRDB MR-microT RNAhybrid	-Complejo SCF E3 ubiquitin-ligasa	Supresión de proliferación de células de hepatoma	Hu et al. 2019
			Schistosoma japonicum	Huevo	S	--	miR-3096 mimic	↓PIK3C2A ↓p-mTOR (Ser^{2448})	miRDB	-Señalización PIK3C2A	Supresión de proliferación de células de hepatoma	Lin et al. 2019
			Schistosoma japonicum	Huevo	S	miR-71a miR-36-3p bantam miR-71b miR-2162	miR-71a mimic	↓Sema4D ↓Plexin B1 ↓CD72 ↓α-SMA ↓Collagen I	miRanda PITA TargetScan	-TGF-β1/SMAD -IL-13/STAT6	Supresión de fibrosis hepática por ↑ de células Treg y ↓ de citoquinas Th1/Th2/Th17	Wang L. et al. 2020
			Schistosoma mansoni	Adulto	S	bantam miR-10 miR-125	miR-10	↓MAP3K7 ↓NF-kB	TargetRank	Vía de señalización NF-kB	Supresión de respuesta Th2	Meningher et al. 2020
			Schistosoma japonicum	Adulto	S	miR-10-5p miR-125b miR-61 miR-2b-5p bantam	bantam	↓Gins4 ↓Tysnd1 ↓Utp3	RNAhybrid miRanda TargetScan	--	Patogénesis hepática	Zhu et al. 2016
			Schistosoma japonicum	--	S	--	miR-61 mimics	↓PGAM1 ↓CD34	miRDB RNAhybrid MR-microT	Regulación de angiogénesis	Supresión de migración celular Anti-angiogénesis	Hu et al. 2021
			Schistosoma japonicum	Huevo	S	--	miR-2162	↓Tgfbr3 ↑α-Sma ↑Col1α1 ↑Col3α1 ↑Col4α1 ↑Timp1	miRDB	-Activación de señalización TGF-β/SMAD	Promueve fibrosis hepática	He et al. 2020
			Schistosoma mansoni	--	S	miR-10-P2a-5p miR-71-P1b-5p bantam-3p miR-10-P2c-5p miR-36-P1-3p	46 miRNAs	-321 blancos génicos compartidos entre ganado y humano. -11 blancos en vía WNT. -44 blancos en sistema inmune.	PITA TargetScan	Regulación de vías WNT canónica y no-canónica	Regulación de hematopoyesis y fisiología hepática	Ovchinnikov et al. 2020
		Fasciolidae	*Fasciola hepatica*	--		miR-277-P2-3p miR-10-P2b-5p miR-71-P1b-5p miR-10-P2a-5p miR-1993-3p				--	--	
			Fasciola hepatica	Newly excysted juvenile	S	--	miR-125b	↓Traf6	miRDB TargetScan miRanda	-Vía MAPK -Vía TNF -Vía NOD-like	Previene la deferenciación M1 de macrófagos	Tran et al. 2021
			Fasciola hepatica	--	N	--	26 miRNAs	23 genes expresados preferentemente en 6 tipos celulares inmunes	miRDB	-Regulation of innate immune cell genes	Regulate specific antiparasitic immune pathways	Ricafrente et al. 2021
			Fasciola hepatica	Adulto	S	miR-10-P2a let-7-P1 miR-279 let-7-P2 let-7-P3	21 miRNAs	-3385 blancos génicos en humanos	miRanda PITA TargetScan	-Vías Ras, MAPK, PI3K-Akt, ErgB y Wnt -Vía TNF -Activación de plaquetas, vía receptora de células B y T	-Regulación de genes relacionados a la inmunidad del húesped	Fontenla et al. 2022
			Fasciola hepatica	Newly excysted juvenile	S	miR-125b-5p miR-10-5p miR-71a-5p miR-2a-3p miR-46-3p	--	--	--	-Transcripción de genes del ciclo celular mediados por FOXO -Vía IL-4/IL-13	--	Herron et al. 2022
				Adulto	S	miR-125a-5p miR-1989-5p miR-2b-1-3p miR-2162-3p miR-10-3p	28 miRNAs	-397 blancos en ARNm de vaca	miRanda PITA RNAhybrid			
		Opistorchiidae	*Clonorchis sinensis*	Adulto	S	miR-71a-5p bantam miR-61-3p miR-10-5p let-7a-5p	let-7a-5p mimic	↓Socs1 ↓Clec7a ↑TNF-α ↑IL-6 ↑IL-1b ↑Nos2 ↓Ym1 ↓Arg1	TargetScan miRanda PicTar	-Vía de señalización NF-kB	-Activación M1 de macrófagos -Induce citoquinas proinflamatorias	Yan et al. 2021

Continuación de Tabla 1.

Platyhelminthes	Cestodes	Taeniidae	*Echinococcus multilocularis*	--	N	--	miR-71 mimic	↑Ago1 ↑Ago4 ↓Nitric oxide (NO)	--	--	Represión de la producción de NO	Zheng et al. 2016
			Echinococcus multilocularis	Protoscoleces	S	miR-71-5p let-7-5p miR-4989-5p miR-10-5p miR-4989-3p	miR-4989-3p mimic	↓NO ↑TNF-α ↑NF-kB ↑AP-1 ↑TICAM2 ↑TLR4	--	-Señalización LPS/TLR4	-Supresión en la producción de NO -Activación de la inflamación	Ding et al. 2019
			Echinococcus multilocularis	Protoscoleces	S	--	miR-4989 mimic	↓UBE2N(Ubc13)	--	-Señalización vía ubiquitinación -Vía de señalización NF-kB y p38	Supresión de la señal de ubiquitinación	Cai et al. 2021
			Taenia pisiformis	Cysticercus	S	--	let-7-5p	↓(C/EBP)-δ ↓INOS ↓IL-12 ↑IL-10	RNAhybrid miRanda TargetScan	-Inhibe la expresión de FT C/EBP-δ	Promueve la polarización M2 de macrófagos y la respuesta inmune Th2	Wang et al. 2021
			Taenia pisiformis	Cysticercus	S	novel-mir-7 miR-4989 let-7-5p novel-mir-11 novel-mir-8	--	↓INF-γ ↓iNOS ↓IL-12 ↑Arg-1 ↑IL-4 ↑IL-6 ↑IL-10 ↑IL-13	RNAhybrid miRanda TargetScan	-Tranducción de señales -Sistema inmune	Induce la secreción de citoquinas relacionadas a Th2 en macrófagos*	Wang L-Q et al. 2020
			Taenia crassiceps	Metacestode	S	let-7-5p miR-61-3p miR-190-5p miR-219-5p miR-4989-3p	--	--	miRanda	-Señalización Wnt -Señalización Cadherin -Inflamación mediada por quimioquinas y citoquinas -Señalización TGF-β -Activación células T/B	--	Ancarola et al. 2017
			Mesocestoides corti		S	let-7-5p	--					

* after treatment with whole EVs

miRNA-target interaction effect

Indirect effect

Para estudiar el papel regulador putativo de los miARNs contenidos en EVs de helmintos, en general, se ha seguido una primera estrategia de aproximación en base a predicciones *in-silico,* utilizando distintas herramientas informáticas (Tabla 1). Esto permite identificar a escala transcriptómica posibles sitios de unión de miARNs de origen parasitario en las regiones 3'UTR de los transcriptos del huésped. Estas predicciones pueden ser luego comprobadas experimentalmente utilizando métodos como medir la expresión génica luego del tratamiento con mimics de miARNs, que son moléculas diseñadas con idéntica secuencia a los miARNs parasitarios. De esta manera, se pudieron identificar vías de señalización relevantes como mTOR, MAPK, Wnt, PI3k, NF-κB, receptores tipo Toll, señalización mediada por citoquinas y factores de crecimiento entre otras (ver Tabla 1), que parecen estar siendo reguladas durante la infección por helmintos. La modulación de estas vías afecta la proliferación celular, los efectos patogénicos, la reparación del daño tisular y/o la respuesta inmune dependiendo del origen de las EVs y del tipo celular que los recibe.

En este sentido, al igual que otras moléculas moduladoras producidas y excretadas/secretadas por los helmintos durante la infección, los miARNs contenidos en EVs parecen modular la respuesta inmune de manera distinta dependiendo del estadio, generando una respuesta inflamatoria Th2 en estadios larvarios, y la supresión de la respuesta inmune en el estadio adulto (Figura 9). Por ejemplo, las EVs producidos por el estadio larvario en el músculo de *Trichinella spiralis* (Yang et al., 2020), el cisticerco de *Taenia pisiformis* (L. Wang et al., 2021; L. Q. Wang et al., 2020), y el estadio juvenil de *F. hepatica* (Tran et al., 2021) modulan citoquinas y/u otros genes que participan en vías de señalización que llevan al desarrollo de una respuesta inmune Th2 y/o a la supresión de la respuesta Th1. En el mismo sentido, los miARNs contenidos en las EVs del estadio adulto de *H. polygyrus* (Buck et al., 2014), *Nippostrongylus brasiliensis* (Eichenberger, Ryan, et al., 2018) y *S. mansoni* (Meningher et al., 2020) parecen suprimir la respuesta pro-inflamatoria o la respuesta Th2. A su vez, evidencia experimental ha mostrado que miR-71a, que es abundante en las EVs de huevos de *S. japonicum,* puede favorecer la respuesta Treg, suprimiendo la fibrosis hepática (L. Wang et al., 2020).

Por otro lado, también hay algunos reportes que describen una respuesta pro-inflamatoria frente a EVs o miARNs específicos (Tabla 1)(Yan et al., 2021; Zamanian et al., 2015). Si bien esto

puede parecer contraintuitivo, en *S. japonicum* se reportó que la inducción en la producción de TNF-α y el incremento en la población de macrófagos y monocitos asociado a los miARNs contenidos en EVs parece incrementar la supervivencia del parásito (J. Liu et al., 2019). Serán necesarios más estudios para conocer si efectos similares ocurren en otros helmintos.

A diferencia de lo que ocurre en otros helmintos como *S. japonicum*, el potencial regulador de los miARNs de *F. hepatica* sobre el huésped definitivo ha sido muy escasamente estudiado, habiendo solo dos reportes previos basados en datos obtenidos a partir de la secuenciación de EVs (Ovchinnikov et al., 2020; Tran et al., 2021), y en el caso del estudio del estadio adulto solo una muestra fue analizada (Ovchinnikov et al., 2020).

5. Las vías de generación de ARNpq

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) (Fire et al., 1991, 1998), y en el que se identificó por primera vez un miARN, lin-4 (Lee et al., 1993).

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 (Ketting et al., 1999; Tabara 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 ARNs pequeños, 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 exógeno, y la vía de ARNi endógeno aún poco conocida en ese momento. En general, las tres vías funcionan de manera similar: un primer complejo dependiente de Dicer procesa el ARN precursor, y un segundo complejo formado por una proteína de la familia

Argonauta junto con ARNpq llamado **R**NA-**I**nduced **S**ilencing **C**omplex o RISC, actúa como complejo efector o silenciador (Figura 10).

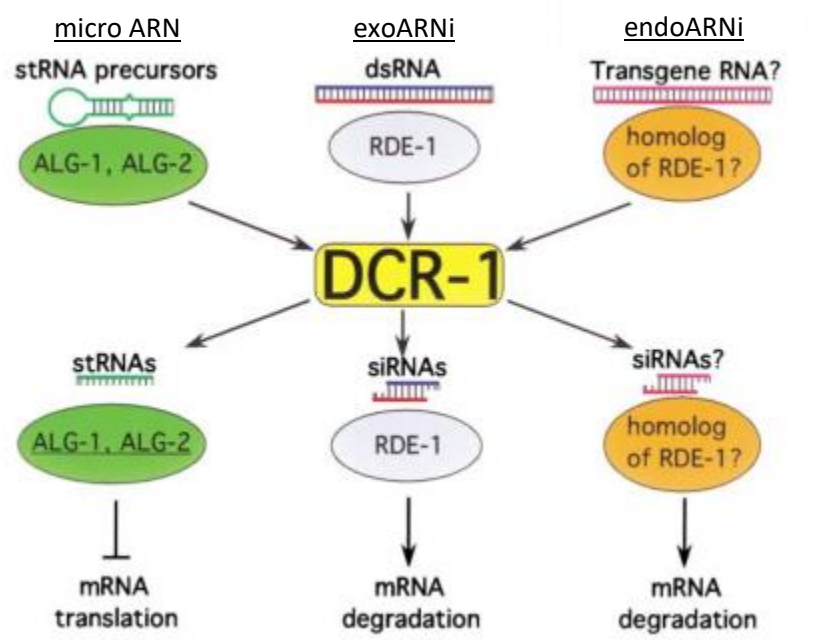


Figura 10. Esquema de las vías de ARNpq. *Se diferencian tres grandes vías de generación de ARNpq, la vía de microARNs, la vía de ARNi y la vía de ARNs endógenos. Tomado de* (Grishok et al., 2001).

5.1. 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 11).

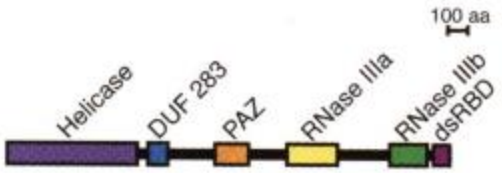


Figura 11. Estructura de Dicer. *Tomado de* (MacRae et al., 2006).

Las Dicer tienen un centro de procesamiento de ARN. Dentro de este centro, cada dominio RNasa III cliva una de las hebras del ARNdc. La dimerización intramolecular de los dominios RNasa 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 nt, con un desfasaje de dos nucleótidos en 3' (Jaronczyk et al., 2005).

5.2. La superfamilia Argonauta

Las proteínas Argonauta forman una superfamilia altamente conservada que se encuentran en diversos organismos desde archaeas a humanos (Jaronczyk et al., 2005). En algunos nematodos como *C. elegans*, la familia Argonauta está amplificada, habiéndose descripto 26 genes codificados en su genoma, y cumplen funciones altamente especializadas (Dalzell et al., 2011; Hutvagner & Simard, 2008).

Las Argonautas contienen dos dominios funcionales, un dominio PAZ y un dominio C-terminal, PIWI. El dominio PAZ se une al desfasaje 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 que 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 que silencia la traducción del ARNm (Figura 12)(Fischer, 2010; Grishok, 2005).

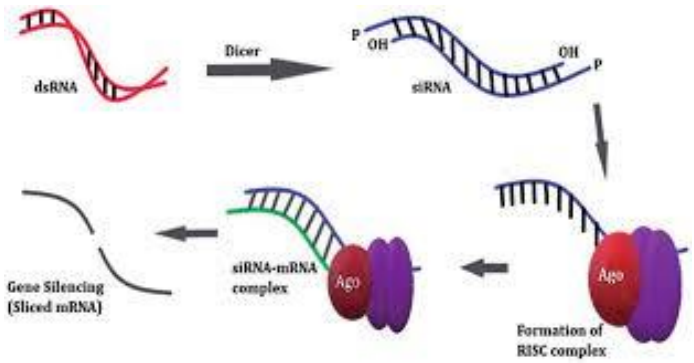


Figura 12. Formación del complejo RISC. *Los ARNdc largos son procesados por Dicer generando ARNdc de 21 nt de largo, con un desfasaje de dos nucleótidos en el extremo 3' de cada hebra. Ago se une al desfasaje de 2 nucleótidos en 3' y cliva la hebra complementaria o pasajera. El ARN pequeño maduro que se mantiene asociado a la proteína Argonauta, y junto a otros factores forman el complejo efector RISC que silencia la traducción del ARNm.*

6. Las vías de ARNpq en platelmintos

La utilidad del ARN interferente como herramienta de genómica funcional y la capacidad reguladora de los miARNs, despertó el interés por evaluar la completitud de las vías de ARNpq en distintos organismos. Este interés fue especialmente relevante en organismos no modelo, donde no existía información genómica para hacer silenciamiento por knockout. En estos organismos el knockdown por ARNi era la única herramienta para el análisis de la función génica.

Dada la escasa información que había respecto a la completitud de las vías de ARNpq en platelmintos, durante la Maestría en Bioinformática analizamos y comparamos la conservación de los genes que participan en la biogénesis de los ARNpq en los genomas de 22 especies de platelmintos parásitos y en 2 especies de platelmintos de vida libre (clase Rhabditiphora) disponibles en repositorios públicos (Tabla 2). Estos resultados se plasmaron, durante el doctorado, en una publicación en una revista arbitrada internacional ((Fontenla et al., 2017), ver ANEXO I).

Tabla 2. Especies analizadas (tomado de Fontenla et al., 2017*)*

CLASS	ORDER	FAMILY	SPECIES	GENOME SIZE	GENE COUNT	BIOPROJECT ID
RHABDITOPHORA	Macrostomida	Macrostomidae	*Macrostomum lignano* (Mli)	1040	60,534	*PRJNA284736*
	Tricladida	Dugesidae	*Schmidtea mediterranea* (Sme)	900	29,850	*PRJNA12585*
MONOGENEA	Monopisthocotylea	Gyrodactylidae	*Gyrodactylus salaris* (Gsa)	67	15,436	*PRJNA244375*
	Polyopisthocotylea	Polystomatidae	*Protopolystoma xenopodis* (Pxe)	617	37,906	*PRJEB1201*
TREMATODA	Opistorchida	Opistorchiidae	*Clonorchis sinensis* (Csi)	547	13,634	*PRJDA72781*
			Opistorchis viverrini (Ovi)	634	16,356	*PRJNA222628*
	Echinostomida	Echinostomatidae	*Echinostoma caproni* (Eca)	834	18,607	*PRJEB1207*
		Fasciolidae	*Fasciola hepatica* (FHO)	1139	15,739	*PRJNA179522*
			Fasciola hepatica (FHL)	1275	22,676	*PRJEB6687*
	Strigeidida	Schistosomatidae	*Trichobilharzia regenti* (Tre)	702	22,185	*PRJEB4662*
			Schistosoma haematobium (Sha)	385	13,073	*PRJNA78265*
			Schistosoma japonicum (Sja)	398	12,738	*PRJEA34885*
			Schistosoma mansoni (Sma)	362	10,772	*PRJEA36577*
CESTODA	Cyclophyllidea	Hymenolepididae	*Hymenolepis diminuta* (Hdi)	166	11,271	*PRJEB507*
			Hymenolepis microstoma (Hmi)	141	12,368	*PRJEB124*
			Hymenolepis nana (Hna)	163	13,777	*PRJEB508*
		Taeniidae	*Echinococcus granulosus* (Egr)	114	10,245	*PRJEB121*
			Echinococcus multilocularis (Emu)	114	10,663	*PRJEB122*
			Taenia asiatica (Tas)	136	10,331	*PRJEB532*
			Taenia solium (Tso)	122	12,481	*PRJNA170813*
			Hydatigera taeniaeformis (Hta)	104	11,649	*PRJEB534*
		Mesocestoididae	*Mesocestoides corti* (Mco)	117	10,614	*PRJEB510*
	Diphyllobotridea	Diphyllobothriidae	*Diphyllobothrium latum* (Dla)	531	19,966	*PRJEB1206*
			Schistocephalus solidus (Sso)	539	20,228	*PRJEB527*
			Spirometra erinaceieuropaei (Ser)	1259	39,557	*PRJEB1202*

Species abbreviation are indicated between brackets. Genome sizes (in Mb)

6.1. La vía de miARNs

Los pri-miARNs son transcriptos por la ARN polimerasa II, estos son procesados a pre-miARNs de 60-100 nucleótidos de largo dentro del núcleo por el complejo microprocesador formado por una proteína con actividad RNasa III llamada Drosha y una proteína de unión a ARN doble cadena (ARNdc) llamada Pasha (Figura 13). Los pre-miARNs tienen una estructura en horquilla y son exportados al citoplasma desde el núcleo por un canal transmembrana llamado Exp-5, en vertebrados y moscas, o su análogo funcional en *C. elegans*, Xpo-1. A continuación, los pre-miARNs son procesados por Dicer junto con una proteína argonauta de la clase miAgo (Ago-1 en insectos). Ago-1 separa la hebra madura de la complementaria, generando un miARN maduro. La hebra madura junto con Ago-1 y otros factores asociados forman el complejo

miRISC. Detectamos que en general la vía de generación de miARNs está conservada en todas las especies analizadas.

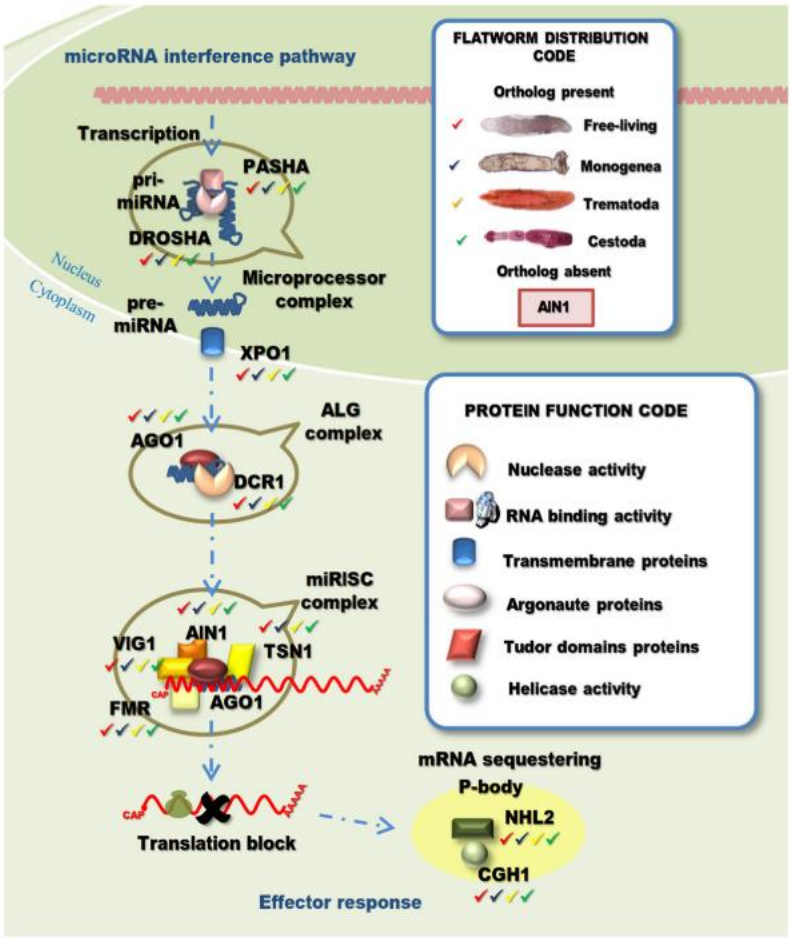


Figura 13. Conservación de genes de la vía de generación de miARNs en platelmintos. *Los pri-miARNs son transcriptos la ARN polimerasa II y procesados en pre-miARNs por Pasha y Drosha. El pre-miARN es exportado al citoplasma por una proteína transmembrana Xpo1, donde es procesado por Dicer a ARNdc cortos. Ago remueve la hebra complementaria del miARN maduro. El miARN con Ago y otros factores asociados al complejo miRISC hibridaran con el ARNm blanco (Tomado de* Fontenla et al. 2017*).*

Los miARNs hibridan con los mARNs blanco reprimiendo la traducción. En los sitios donde el apareamiento Watson-Crick es extenso, los miARNs de metazoarios pueden inducir el clivaje del ARNm mediado por Ago. Sin embargo, más comúnmente, los miARNs de metazoarios se aparean perfectamente con su sitio blanco en la región 5' del miARN centrados en los nucleótidos 2-8 llamada región 'seed' o 'semilla'. Por fuera de esta región el apareamiento es

más laxo conduciendo a la represión de la traducción, desestabilización el ARNm o ambos (Figura 14)(Bartel, 2009).

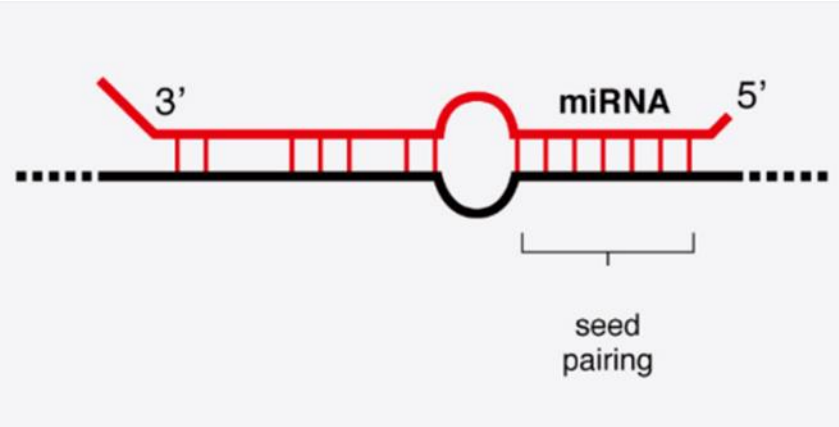


Figura 14. Apareamiento entre un miARN y su sitio blanco en un ARNm. *Más comúnmente, los miARNs de metazoarios se aparean perfectamente en la región 5' del miARN llamada región 'seed' o 'semilla'. Tomado y modificado de* Fuchs Wightman et al., 2018.

6.2. La vía de interferencia por ARNdc exógeno

La vía de ARNi está conservada a lo largo de la evolución y posiblemente surgió como un mecanismo de defensa contra ARNdc virales u otros parásitos genéticos (Tomari & Zamore, 2005). El mecanismo se desencadena cuando un ARNdc largo es internalizado por una proteína de canal transmembrana análoga a SID-1 de *C. elegans* (Figura 15). En *C. elegans,* el ARNdc es reconocido en el citoplasma por una proteína de unión a ARNdc llamada RDE-4 y presentado a Dicer para su procesamiento. RDE-4 está conservado solo en algunas especies de nematodos. Por lo tanto, también procuramos ortólogos del análogo funcional de RDE-4 en *D. melanogaster,* llamado R2D2. Sin embargo, no identificamos proteínas similares en platelmintos. De todas maneras, dado que la interferencia por ARNdc ha probado ser funcional en varias especies de platelmintos, es probable que una vía alternativa u otro mediador aun no caracterizado esté operando en estos organismos.

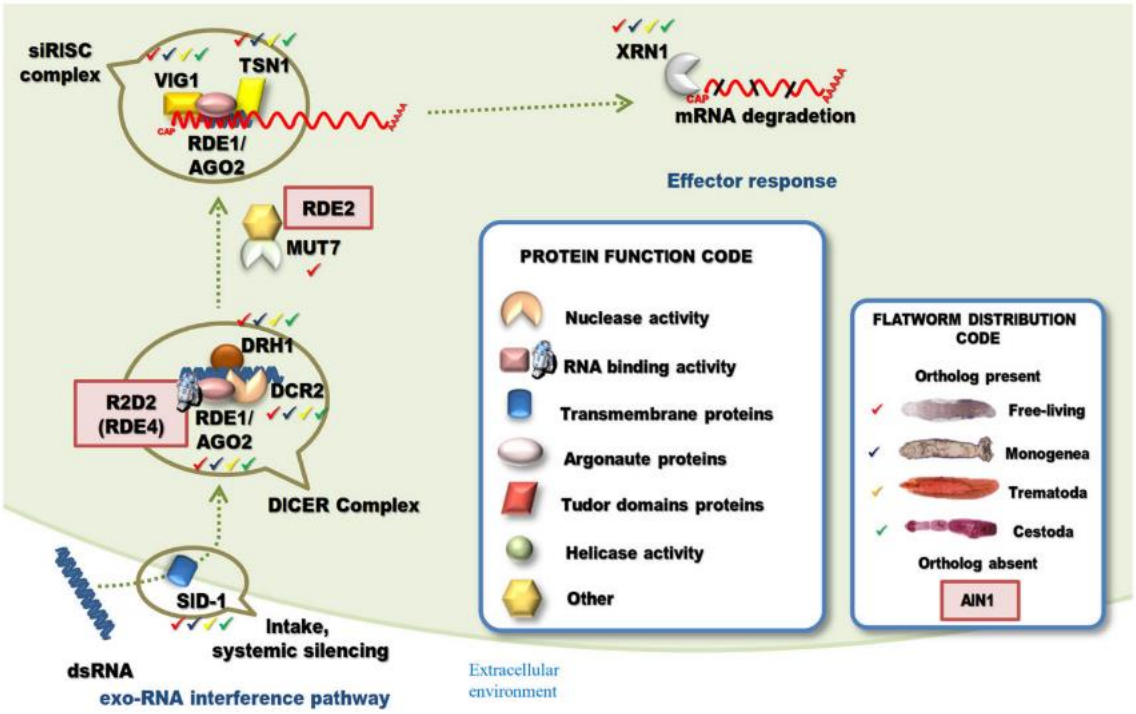


Figura 15. Conservación de los genes de la vía de interferencia por ARN exógeno en platelmintos. *Un ARNdc exógeno es internalizado por un canal transmembrana, SID-1. En el citoplasma el ARNdc es reconocido por RDE4 (un ortólogo de R2D2 de Drosophila melanogaster), esto promueve la formación del complejo DICER que procesa el ARNdc largo en ARNdc de 22 nt de largo. La proteína argonauta RDE1 (homóloga a Ago2 de D. melanogaster) procesa el ARNdc liberando el ARN pequeño interferente maduro, que junto con otros factores formaran el complejo efector siRISC. (Tomado de* Fontenla et al. 2017*)*

Una proteína con dominio DexH-box/Helicasa (DRH-1) y una proteína Argonauta de la clase siAGO (llamada Ago-2 en insectos o RDE-1 en *C. elegans*) también forman parte de este complejo. Ago-2 se une a los ARN pequeños interferentes (siARNs) estabilizándolos y cliva la hebra complementaria o pasajera. Los siARNs maduros permanecen unidos a Ago-2 formando el complejo siRISC.

XRN-1 es una exonucleasa 5'->3' asociada a Dicer que degrada el ARNm blanco del complejo RISC y está conservada en platelmintos. Por el contrario, solo detectamos ortólogos de Mut-7 en el platelminto de vida libre *Macrostomum lignano,* y RDE-2 no fue detectado en ninguna de las especies analizadas. Mut-7 y RDE-2 forman un complejo que forma parte de la ARNi en *C. elegans* cuya función es aún poco conocida.

6.3. *La vía de interferencia por ARNdc endógeno*

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 (Figura 16). En *C. elegans*, los transcriptos primarios son procesados por el complejo ERI (o **E**nhancer of **R**NA **I**nterference) formado por Dicer, DRH-3, ERI-1, ERI-3, ERI-5, ERI-6, ERI-7 y RRF-3, una ARN polimerasa dependiente de ARN (RdRP). Los factores ERI fueron nombrados de esta manera en base a resultados que indicaban que la supresión de estos genes potenciaba el efecto de la ARNi exógena, lo que sugirió que las vías de ARNi endógeno y exógeno compartían factores en común.

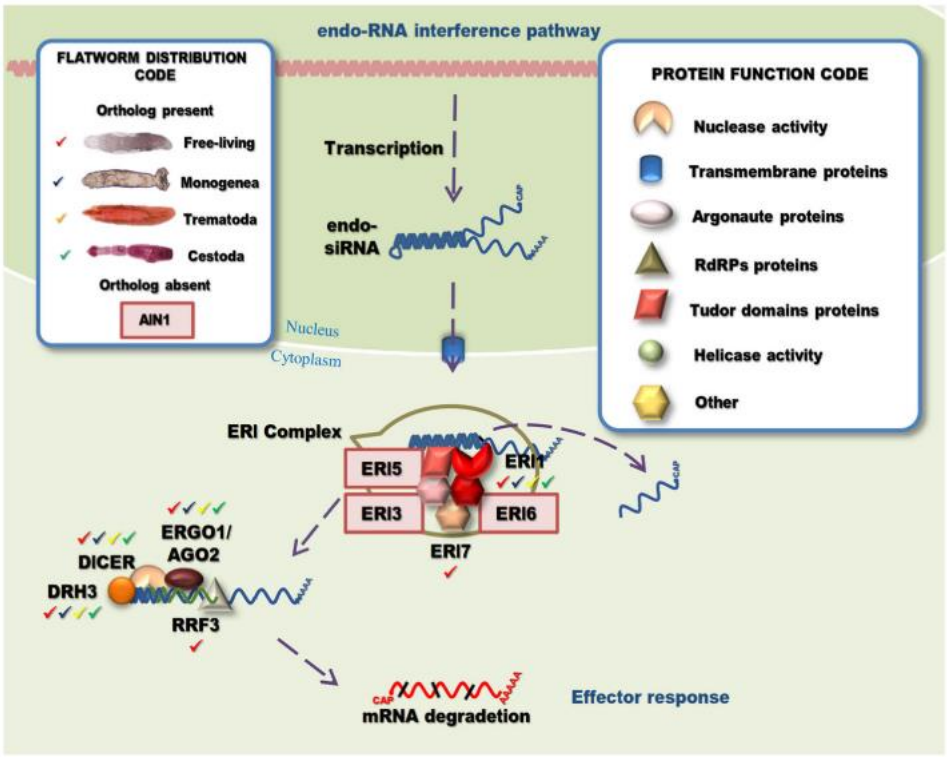


Figura 16. Conservación de los genes de la vía de interferencia por ARN endógeno en platelmintos. *El ARNnc conteniendo una región de ARNdc es transcripto en el núcleo y procesado en el citoplasma por el complejo ERI. Luego de unirse al ARNdc, ERI-1 remueve una de las hebras no apareadas creando una estructura en horquilla pasible de servir como primer para la síntesis de ARNdc largo por RRF-3. Dicer y DRH-3 luego clivan el ARNdc largo, y ERGO-1 (un homólogo de Ago2) cargado con un siARN gatilla la degradación del ARN blanco. (Tomado de* Fontenla et al. 2017*)*

En platelmintos, detectamos la presencia de homólogos de ERI-1, una enzima con actividad exonucleasa 3'->5'. Mientras que en platelmintos parásitos no pudimos detectar otros miembros del complejo ERI, si encontramos ortólogos de ERI-7 en ambas especies de vida libre analizadas, y genes con dominio RdRP solo en *Macrostomum lignano*.

6.4. Biogénesis de piARNs

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 Argonautas de la clase 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 previene la acumulación de cambios en el genoma en las siguientes generaciones.

La metilasa HEN-1, implicada en la ARNi y en la vía de piARNs asociados a Argonautas (Agos) tipo Piwi (Montgomery et al., 2012) fue detectada únicamente en *M. lignano* y *S. mediterranea,* lo que es consistente con la presencia Agos tipo Piwi en los platelmintos de vida libre estudiados y su ausencia en los parásitos. A su vez, Vasa, otro factor clave de la vía de piARNs, también parece estar ausente en platelmintos parásitos (Skinner et al., 2014), lo que soporta la hipótesis de una pérdida evolutiva de los piARNs en los Neodermata y su conservación en platelmintos de vida libre. Sin embargo, esta hipótesis está fundada en el análisis de un número muy escaso de especies de vida libre. Por lo tanto, nos preguntamos si las vías estarán conservadas de la misma manera en otros clados del grupo parafilético 'Turbelaria'. Si esto es así, nuestros hallazgos podrían indicar una relación directa entre la divergencia en las vías de los ARNpq y los mecanismos que condujeron al parasitismo de los Neodermatas.

Hipótesis

En función de los antecedentes presentados es posible que los ARNpq jueguen un papel en el desarrollo de *F. hepatica* y en el contexto de la interacción huésped-parásito al ser secretados en vesículas extracelulares. A su vez, dada la relevancia de los ARNpq en la regulación de diversas funciones, la ausencia de genes de las vías de ARNpq entre los neodermata versus otros clados de platelmintos podría estar vinculada al parasitismo. En consecuencia, nos planteamos la siguiente hipótesis: los ARN pequeños deben cumplir un papel esencial en la regulación del desarrollo y en la interacción con el huésped, y a su vez, podrían haber divergencias asociadas a la conservación de las vías de síntesis de ARN pequeños que podrían estar involucrados en la adaptación al parasitismo en platelmintos.

Objetivos

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

A) Analizar los perfiles de ARNpq en los estadios asociados al huésped mamífero y en vesículas extracelulares de *F. hepatica,* y estudiar su potencial regulador.

Objetivos específicos:

i) Identificar y comparar la expresión de miARNs entre los estadios intra-mamífero de *F. hepatica,* y entre el estadio adulto y EVs.

ii) Identificar los genes blanco y las funciones reguladas por los miARNs diferencialmente expresados entre estadios de *F. hepatica*.

iii) Identificar los genes blanco y las funciones reguladas en el huésped por los miARNs más abundantes en las EVs de *F. hepatica.*

iv) Caracterizar otras especies de ARNnc menos estudiadas como los tARNs y los vault ARNs.

B) Comparar la conservación de las vías de ARNpq entre platelmintos de vida libre y parásitos.

Objetivos específicos:

i) Identificar transcriptomas de buena calidad de distintas especies de platelmintos de vida libre disponibles en base de datos públicas.

ii) Identificar los genes involucrados en las vías de generación ARNpq en base a ortología.

iii) Inferir la relación evolutiva de familias génicas claves de las vías de ARNpq.

iv) Estudiar la conservación de secuencia a nivel de dominios e inferir si hay selección positiva entre las subfamilias génicas de Argonauta.

Materiales y métodos

Dado que los resultados presentados en el presente trabajo ya se han publicado en revistas arbitradas internacionales, los materiales y métodos están descriptos en las secciones correspondientes de los artículos presentados. En la siguiente sección se describe la estrategia seguida para abordar cada uno de los objetivos planteados. Adicionalment, en el ANEXO III se encuentran los programas y parámetros usados en la tesis.

Estrategia general

A) Analizar los perfiles de ARNpq en los estadios intramamífero y en vesículas extracelulares de F. hepatica, y estudiar su potencial regulador.

Se analizaron muestras correspondientes a diversos estadios de vida intra-mamífero. Las muestras de metacercarias dormante (MD) y activadas *in vitro* (MA) y gusanos adultos (AD1; AD2) fueron generadas por nosotros en el marco del proyecto del genoma de *F. hepatica* (estudio SRP040521). Para la obtención de estas muestras el ARN total se extrajo usando el kit mirVana Total Isolation (Thermo Fisher) separando las fracciones grande (>200 pb) y pequeña (<200 pb). Las bibliotecas se construyeron utilizando el kit NEBNext Multiplex Small RNA Library Prep.

Los datos de juveniles (NEJ) fueron generados y analizados previamente por nosotros (Fontenla et al., 2015), y reanalizados y comparados con otros estadios en este trabajo. En este caso se utilizó el kit RNeasy Mini para la extracción del ARN total. Este se separó por tamaño en gel de poliacrilamida, seleccionando posteriormente la fracción de ARN pequeño. Se uso la transcriptasa reversa Superscript II (Invitrogen) e Illumina small RNA RT-Primer según instrucciones del fabricante para la retrotranscribir el ARN y construir las bibliotecas.

Las muestras de EV se obtuvieron de gusanos adultos recuperados de ganado vacuno infectado naturalmente en un matadero local en Dungannon, Irlanda del Norte. El ARN total de estas muestras se aisló usando TRizol (Life Technologies). Los ARN pequeños fueron subsequentemente aislados usando el kit SeraMir (SystemBio). Las bibliotecas se generaron usando el kit TailorMix miRNA Sample Preparation Version 2 (SeqMatic LLC). La construcción de las bibliotecas y la secuenciación de estas muestras fueron realizadas por System Biosciences, USA.

Las lecturas de secuenciación fueron procesadas en paralelo. Se removieron los adaptadores y se filtraron las lecturas de secuenciación eliminando aquellos de baja calidad y tamaño menor a 18 nucleótidos. Las lecturas restantes se mapearon con Bowtie (Langmead et al., 2009) al

genoma de *F. hepatica*, permitiendo un máximo de 2 mismatches. Identificando sucesivamente las lecturas correspondientes a mRNA, tRNA, rRNA, snRNA y regiones repetitivas del genoma (es decir, transposones, regiones de baja complejidad o repetidos en tándem) (Figura 17).
Las lecturas no asignadas a ARN funcionales se colapsaron y remapearon al genoma de *F. hepatica*. Se utilizó el pipeline de miRDeep2 (Friedländer et al., 2012) para identificar y cuantificar los miARN nuevos y conservados, usando miARNs depositados en miRBase v22.1 y otros miARNs de *F. hepatica* identificados previamente, pero no depositados en miRBase (Ricafrente et al., 2021) como set de miARNs "conocidos" (ver ANEXO III).

Además, la presencia de secuencias derivadas del huésped se evaluó volviendo a mapear las lecturas que no mapearon en el genoma de *F. hepatica*, al genoma de *Bos taurus*.

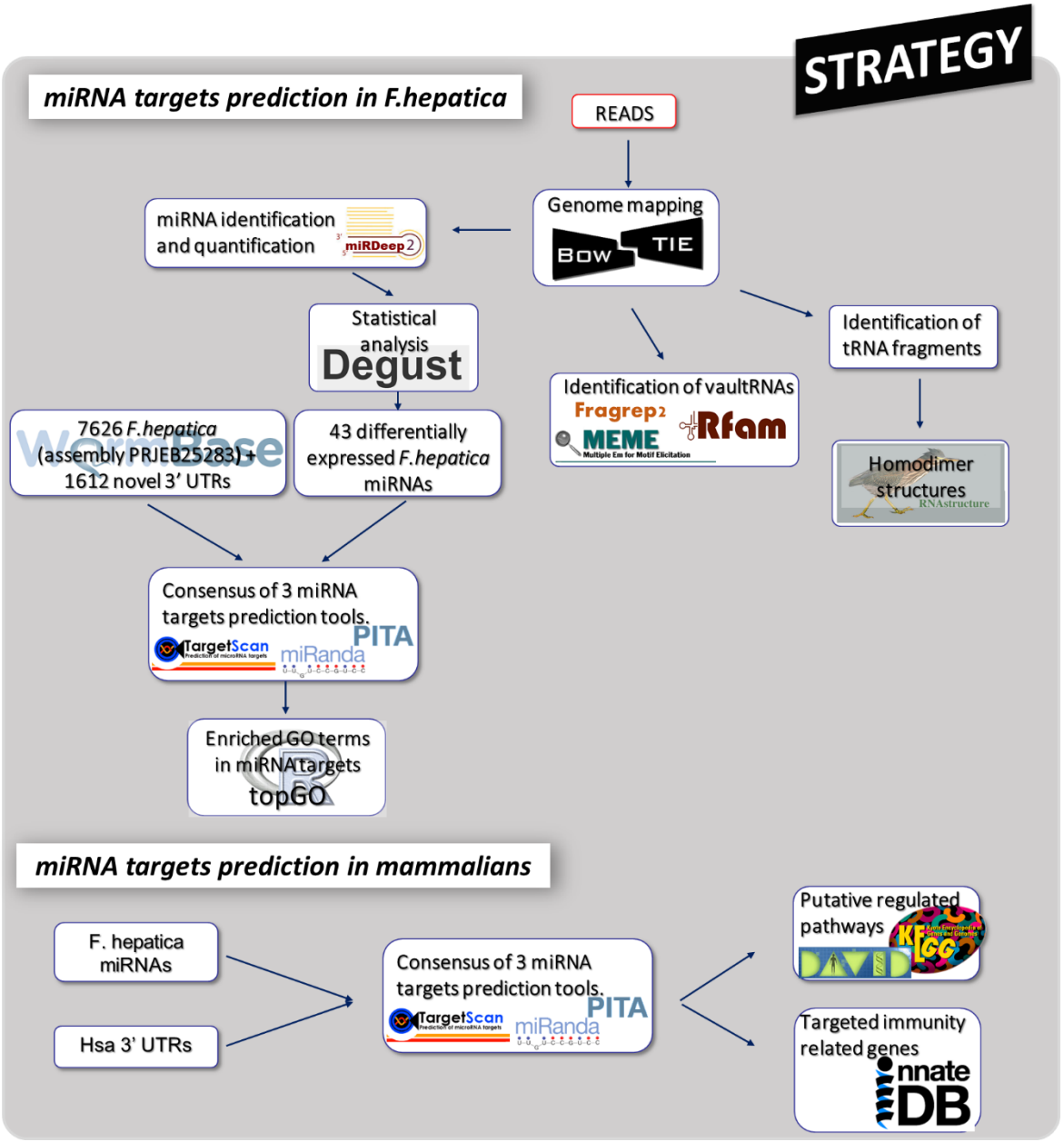


Figura 17. Esquema de la estrategia seguida en el objetivo (A).

El análisis estadístico de la expresión de miARNs se realizó con DEGUST v4.1.1 (https://degust.erc.monash.edu/). Las muestras se agruparon por estadio; y se analizaron con el método Voom/Limma, con normalización a conteo por millón (CPM). Los miARN con un log2FoldChange ≥ 2 y una tasa de descubrimientos falsos (FDR) ≤ 0.01 fueron considerados como diferencialmente expresados entre estadios.

Se utilizaron tres algoritmos: miRanda (Betel et al., 2010), PITA (Kertesz et al., 2007) y TargetScan v7.0 (Agarwal et al., 2015), con opciones por defecto para predecir los blancos de los miARNs noveles y conservados diferencialmente expresados en una base de datos conteniendo 7626 regiones 3' no traducidas (3'UTR) anotadas en el ensamblaje PRJEB25283 de *F. hepatica* y 1612 3'UTR de transcriptos nuevos obtenidas de una re-predicción génica (Langleib et al. en preparación). Se consideraron solo aquellas predicciones coincidentes entre los tres algoritmos.

Se utilizó TopGO (Alexa & Rahnenfuhrer, 2019) para identificar las ontologías génicas enriquecidas (GO) entre los blancos de los miARN expresados diferencialmente, los términos GO con P-valores inferiores o iguales a 0.01 se consideraron significativamente enriquecidos. Para predecir los blancos en el huésped mamífero se descargaron los 3'UTR de humanos de la base de datos TargetScan (http://www.targetscan.org/cgi-bin/targetscan/data_download.vert72.cgi). Como se describió anteriormente, las predicciones se ejecutaron con opciones predeterminadas. Se utilizó la base de datos DAVID (Huang et al., 2009) para identificar funciones reguladas en genes blanco, considerando solo las vías reguladas por dos o más miARN. Además, se utilizó InnateDB (Breuer et al., 2013) para identificar los genes relacionados con la inmunidad bajo regulación.

Las lecturas que se asignaron a la base de datos de ARNt se cuantificaron y clasificaron mediante scripts propios. Para cada fragmento de ARNt, la secuencia con el recuento más alto se usó para la predicción *in-silico* de plegamiento y dimerización utilizando la herramienta 'Fold RNA Biomolecular' del paquete RNAstructure6.3 (Bellaousov et al., 2013).

Para la identificación de vault RNAs (vtARNs) conservados en el genoma de *F. hepatica* se utilizó la región más conservada de los vtARNs identificados en *Lottia gigantea* y *Helobdella robusta* obtenido de la base de datos Rfam (https://rfam.xfam.org/ RF00006) para realizar una búsqueda por BLASTn en el genoma de *F. hepatica*. Los vtARNs putativos se inspeccionaron adicionalmente con MEME suite (Bailey et al., 2009) para confirmar la conservación de motivos relevantes, y RNAfold (Lorenz et al., 2011) para evaluar la presencia de la característica

estructura secundaria similar a un 'mango de sartén'. A continuación, utilizamos fragrep2 (Mosig et al., 2006) para re-inspeccionar el genoma de *F. hepatica*.

B) Comparar la conservación de las vías de ARNpq entre platelmintos de vida libre y parásitos.

Las proteínas de las vías de ARNpq de *Macrostomum lignano*, *Schmidtea mediterranea*, *Gyrodactylus salaris*, *S. mansoni* y *E. multilocularis* junto con otras especies de Neodermatas ya habían sido caracterizadas previamente por nosotros (Fontenla et al., 2017, ANEXO I).

Los transcriptomas de platelmintos turbelarios no-modelo fueron publicados por Laumer et al. (2015). El transcriptoma del monogenea basal *Eudiplozoon nipponicum* fue publicado por (Ilgová et al., 2017). También se utilizaron los datos genómicos publicados recientemente por nosotros de los trematodos *F. gigantica* y *Fasciolopsis buski* ((Choi et al., 2020), ver ANEXO II), y los genomas de cuatro especies del género *Paragonimus* publicados por (Oey et al., 2018; Rosa et al., 2020). Todos los datos fueron descargados de repositorios públicos y/o están disponibles públicamente.

Previo a analizar la conservación de las vías de ARNpq en los datos transcriptomicos, se estudió la calidad de los datos con BUSCO. En base a esto, se seleccionaron siete ensamblajes que abarcaran razonablemente la diversidad del clado Turbelaria. Se utilizó TransDecoder para definir los marcos abierto de lectura (ORF) y traducir los transcritos (Figura 18)(ver ANEXO III).

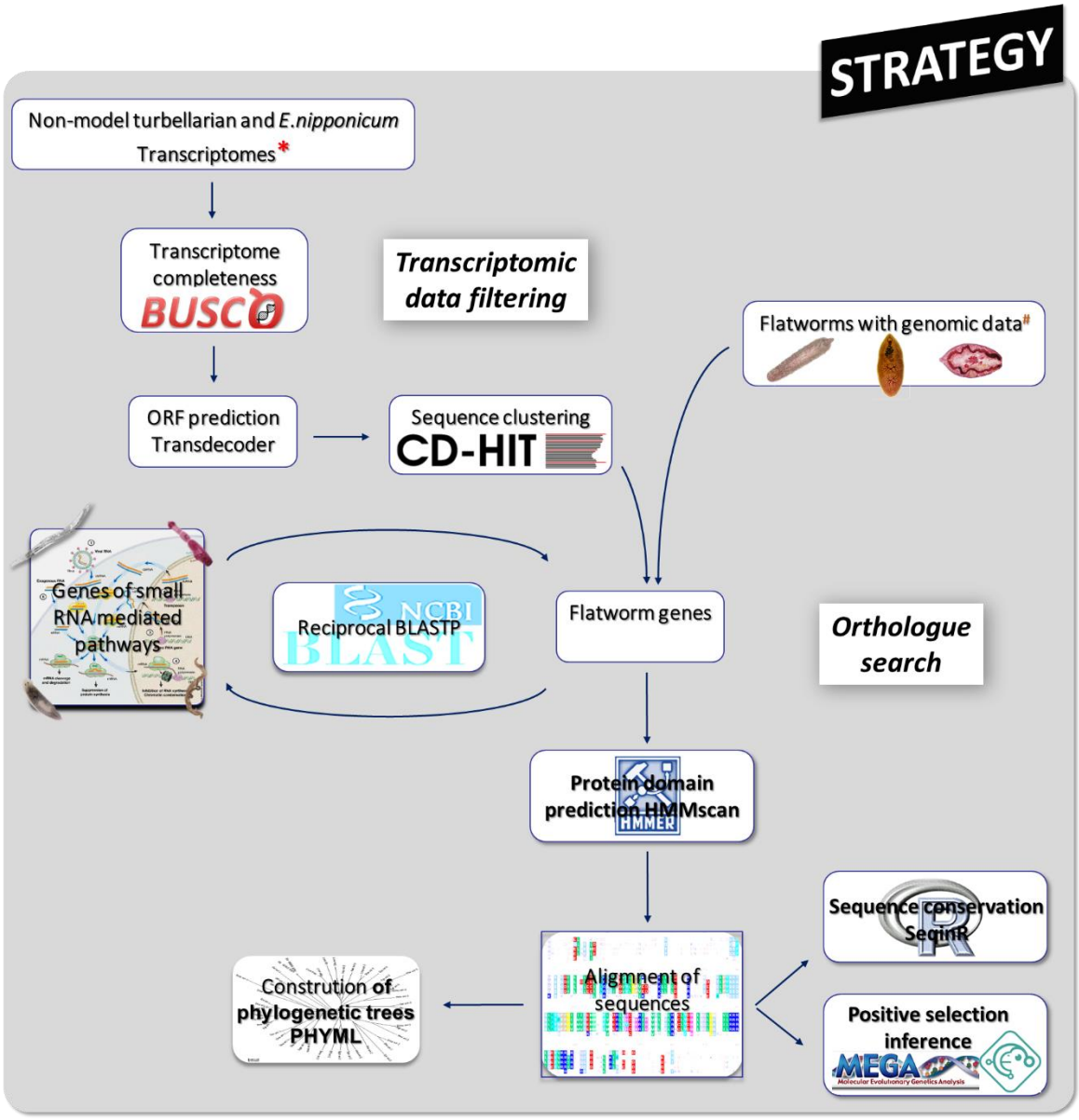


Figura 18. Estrategia seguida en el objetivo B. *(*) Datos generados por* (Laumer et al., 2015) *y por* (Ilgová et al., 2017)*, respectivamente. (#) Genomas de F. gigantica y F. buski generados por* (Choi et al., 2020)*; genomas de Paragonimus generados por* (Oey et al., 2018; Rosa et al., 2020)*.*

Siguiendo un esquema similar al utilizado previamente (Fontenla et al., 2017), se aplicó una estrategia de BLASTp recíproco para identificar los genes vinculados a las vías de ARNpq en los transcriptomas traducidos y en los genomas publicados recientemente. El BLASTp directo se realizó usando como 'carnada' los genes de las vías de ARNpq identificadas en platelmintos, y genes particulares de *C. elegans* y *Drosophila melanogaster* no detectados en nuestro análisis

previo. A su vez, se comprobó la homología funcional de las proteínas putativas detectadas mediante predicciones de dominios proteicos por HMMScan.

Para evitar sobreestimar el número de genes en los datos transcriptómicos o reportar duplicaciones genómicas enteras como duplicaciones génicas (como ocurre en el genoma tetraploide de *M. lignano*), se agruparon las secuencias con una similaridad mayor al 90%.

MAFFT con opción de alineamiento local e información estructural fue usado para generar los alineamientos de secuencias. PhyML fue usado para construir los árboles filogenéticos de familias génicas claves en las vías de biogénesis de ARNpq.

Previamente habíamos reportado la ausencia de Argonautas de la clase Piwi en Neodermatas y la aparición de un cluster de genes de la familia Argonauta específico de platelmintos, denominado 'FL-Agos' por nosotros, que presentaban amplificaciones independientes en las especies parásitas (Fontenla et al., 2017). Estos resultados obtenidos en nuestro set extendido de especies eran, a su vez, coincidentes con resultados publicados por otros grupos en sets más reducidos (Skinner et al., 2014; Zheng, 2012). Sospechando que podía estar ocurriendo un fenómeno de diversificación en la familia de las Argonautas, decidimos estudiar la conservación de secuencia a nivel de dominios proteicos e inferir si había selección positiva entre las subfamilias génicas de Argonauta de platelmintos.

Resultados

Los resultados correspondientes a las distintas secciones indicadas en Objetivos y Estrategia General, ya han sido publicados en revistas arbitradas internacionales, por lo que se incluyen estas publicaciones en las sub-secciones correspondientes.

A. Rol de los ARN pequeños en el desarrollo de *Fasciola hepatica* y en la interacción con el huésped

ARTICULO: Rol de los ARN pequeños de Fasciola hepatica en la interacción con el huésped mamífero.

Santiago Fontenla, Mauricio Langleib, Eduardo de la Torre Escudero, María Fernanda Domínguez, Mark W. Robinson, José Tort.

Frontiers Cellular and Infection Microbiology (2022), doi: 10.3389/fcimb.2021.812141

En este trabajo, presentamos los resultados que responden al objetivo (A): Analizar los perfiles de ARNpq en los estadios asociados al huésped mamífero y en vesículas extracelulares de *F. hepatica,* estudiando su potencial regulador.

Hallamos que los perfiles de expresión de miARNs de los estadios intramamífero muestran importantes variaciones, a pesar de que hay un set de miARNs que es altamente expresado en todos los estadios. En base al pipeline de predicción de interacción 3'UTR-miARNs, detectamos funciones relevantes vinculadas al desarrollo y otras que podrían ser relevantes en el proceso invasivo entre los blancos de los miARNs diferencialmente expresados entre estadios de *F. hepatica*. A su vez, entre los blancos génicos en el huésped de los miARNs más abundantes en EVs de *F. hepatica,* hallamos varias funciones relacionadas con la modulación o respuesta inmune.

Adicionalmente, detectamos fragmentos derivados del procesamiento de tARNs en todos los estadios del desarrollo analizados, y documentamos la presencia de fragmentos largos noveles de tARNs enriquecidos en vesículas. También, confirmamos la presencia de al menos 5 vtARNs putativos, que son expresados en los estadios estudiados y están enriquecidos en vesículas.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 20 January 2022
doi: 10.3389/fcimb.2021.812141

Role of *Fasciola hepatica* Small RNAs in the Interaction With the Mammalian Host

Santiago Fontenla[1*], *Mauricio Langleib*[1,2], *Eduardo de la Torre-Escudero*[3], *Maria Fernanda Domínguez*[1], *Mark W. Robinson*[3] *and José Tort*[1*]

[1] Departamento de Genética, Facultad de Medicina, Universidad de la República (UdelaR), Montevideo, Uruguay, [2] Departamento de Desarrollo Biotecnológico, Instituto de Higiene, Facultad de Medicina, Universidad de la República (UdelaR), Montevideo, Uruguay, [3] School of Biological Sciences, Queen's University Belfast, Belfast, Ireland

OPEN ACCESS

Edited by:
Pengfei Cai,
QIMR Berghofer Medical Research Institute, Australia

Reviewed by:
Juan F. Quintana,
University of Glasgow, United Kingdom
Eileen Devaney,
University of Glasgow, United Kingdom

***Correspondence:**
Santiago Fontenla
sfontenla@fmed.edu.uy
José Tort
jtort@fmed.edu.uy

Specialty section:
This article was submitted to Parasite and Host, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: 09 November 2021
Accepted: 29 December 2021
Published: 20 January 2022

Citation:
Fontenla S, Langleib M, de la Torre-Escudero E, Domínguez MF, Robinson MW and Tort J (2022) Role of Fasciola hepatica Small RNAs in the Interaction With the Mammalian Host. Front. Cell. Infect. Microbiol. 11:812141. doi: 10.3389/fcimb.2021.812141

MicroRNAs (miRNAs) are important post-transcriptional regulators of gene expression being involved in many different biological processes and play a key role in developmental timing. Additionally, recent studies have shown that miRNAs released from parasites are capable of regulating the expression of host genes. In the present work, we studied the expression patterns of ncRNAs of various intra-mammalian life-cycle stages of the liver fluke, *Fasciola hepatica*, as well as those packaged into extracellular vesicles and shed by the adult fluke. The miRNA expression profile of the intra-mammalian stages shows important variations, despite a set of predominant miRNAs that are highly expressed across all stages. No substantial variations in miRNA expression between dormant and activated metacercariae were detected, suggesting that they might not be central players in regulating fluke gene expression during this crucial step in the invasion of the definitive host. We generated a curated pipeline for the prediction of putative target genes that reports only sites conserved between three different prediction approaches. This pipeline was tested against an iso-seq curated database of the 3' UTR regions of *F. hepatica* genes to detect miRNA regulation networks within liver fluke. Several functions related to the host immune response or modulation were enriched among the targets of the most highly expressed parasite miRNAs, stressing that they might be key players during the establishment and maintenance of infection. Additionally, we detected fragments derived from the processing of tRNAs, in all developmental stages analyzed, and documented the presence of novel long tRNA fragments enriched in vesicles. We confirmed the presence of at least 5 putative vault RNAs (vtRNAs), that are expressed across different stages and enriched in vesicles. The presence of tRNA fragments and vtRNAs in vesicles raise the possibility that they could be involved in the host-parasite interaction.

Keywords: *Fasciola hepatica*, micro RNA (miRNA), development, tRNA, vault RNA (vtRNA), Extracellular vesicles (EVs), host-parasite

INTRODUCTION

Fasciolosis, or liver fluke disease, is caused by infection with two major liver fluke species: *Fasciola hepatica* in temperate regions of all continents and *Fasciola gigantica*, which is more restricted to tropical regions. Fasciolosis in ruminants is widespread and is responsible for massive economic losses to the livestock industry, estimated globally to be US$3.2 billion annually due to reduced production yields and associated treatment costs (Zerna et al., 2021). It is now also recognized as a neglected tropical disease of humans by the World Health Organization (WHO) (World Health Organization, 2007) with 17 million people being infected and another 180 million people at risk of acquiring the infection predominantly in developing countries. However, the actual numbers of infections in humans and animals are likely underestimated due to the lack of comprehensive or coordinated investigations, and limited availability of diagnostic tools in some developing countries (Harrington et al., 2017).

Liver flukes have complex life cycles. The definitive hosts, (either livestock or humans) acquire the infection by consuming water or plants (e.g., watercress) contaminated with parasite cysts (metacercariae). Gastric and duodenal contents promote excystation, and the newly excysted juveniles (NEJs) actively migrate across the intestinal wall into the peritoneal cavity, and then into the liver parenchyma. The final destination are the biliary ducts within the liver where they reach maturity, producing thousands of eggs that are released to the environment in the stool (Keiser and Utzinger, 2009).

Clinically, the acute phase of *F. hepatica* infection is characterized by marked eosinophilia, abdominal pain and fever. Beyond this acute form of the disease, a chronic stage once the parasite is established is characterized by intermittent obstruction of bile ducts, causing symptoms that resemble biliary colic and cholecystitis (Garcia et al., 2007). Like most helminths, *Fasciola* spp. have developed strategies to modulate the host immune response to promote their survival and reproduction. These play an essential role in the establishment of infection and are mediated by molecules secreted by the parasite and those of the host that activate repair mechanisms (Dalton et al., 2013).

While initial studies have focused largely on proteins as possible immunomodulators, the discovery of microRNAs that are secreted by the parasite revealed new mediators in the host-parasite interaction. MicroRNAs (miRNAs) are single-stranded non-coding 21 to 25 nucleotide long sequences, that negatively regulate the translation of messenger RNAs (mRNAs) by blockage of translation and/or mRNA destabilization (Bartel, 2009). MiRNAs were first described in the roundworm *Caenorhabditis elegans* with roles in the regulation of development (Lee et al., 1993; Reinhart et al., 2000). Since then, miRNAs have been described in all eukaryotes (Wheeler et al., 2009; Tarver et al., 2013) and perform varied functional roles including cellular differentiation, apoptosis, metabolism and silencing of transposable elements (Bushati and Cohen, 2007; Tristán-Ramos et al., 2020). MiRNAs have also been described in several platyhelminths (Palakodeti et al., 2006; Basika et al., 2016; Cai et al., 2016; Stroehlein et al., 2018), including the NEJ and adult stages of *F. hepatica* (Xu et al., 2012; Fontenla et al., 2015; Fromm et al., 2015) and more recently in eight stages of the life cycle of *F. gigantica* (Hu et al., 2021).

The secretion of EVs was initially described as a means of eliminating unneeded compounds from the cell (Johnstone et al., 1987) but later works reported the mRNA and miRNA cargo within the EVs of mammalian cells suggesting that they have roles in regulation of gene expression (Valadi et al., 2007; Mitchell et al., 2008). In helminths, miRNAs contained in EVs was first reported in the trematode *Dicrocoelium dendrticum*, soon followed by reports in the clade V nematode *Heligmosomoides polygyrus* (Bernal et al., 2014; Buck et al., 2014), and later by similar findings in several other nematode, trematode and cestode species (Fromm et al., 2015; Nowacki et al., 2015; Sotillo et al., 2020; Avni and Avni, 2021; Cucher et al., 2021).

Parasite-specific miRNAs have been detected in the blood of hosts infected with several nematodes including *Onchocerca ochengi* and *O. volvulus* (Quintana et al., 2015), *Loa* (Tritten et al., 2014b), *Dirofilaria immitis* (Tritten et al., 2014a), *Litomosoides sigmodontis* (Buck et al., 2014; Quintana et al., 2019) and trematodes such as schistosomes (Cheng et al., 2013; Hoy et al., 2014; Cai et al., 2015; Nowacki et al., 2015; Meningher et al., 2017; Mu et al., 2020) and *F. gigantica* (Guo and Guo, 2019). These findings have stimulated interest in the possible value of worm-derived miRNAs as diagnostic biomarkers for helminthiases [reviewed by Mu et al., (2021)].

Alongside this diagnostic interest, recent efforts have focused on investigating the immunomodulatory potential of secreted helminth miRNAs (Arora et al., 2017). For example, *H. polygyrus* EV miRNAs can be internalized by the mouse epithelial cell line Mode K, and inhibit host type 2 innate immunity (Buck et al., 2014). In line with this, miR-10 contained in *Schistosoma mansoni* EVs, internalized by T helper cells, inhibits NF-κB activity essential for Th2 differentiation (Meningher et al., 2020). Several miRNAs of *S. japonicum* were reported to be involved in the regulation of host macrophages by inducing TNF-α production and monocyte proliferation (Liu et al., 2019). In another case, a particular *S. japonicum* miRNA (miR-3096) showed antitumor effects both *in vivo* and *in vitro* (Lin et al., 2019).

Numerous miRNAs have been identified within the EVs secreted by *F. hepatica* (Fromm et al., 2015). Moreover, *F. hepatica* EVs can be internalized by host macrophages (de la Torre-Escudero et al., 2019) and their miRNAs have been detected in macrophages of infected animals (Tran et al., 2021), indicating that *F. hepatica* secreted miRNAs, either in vesicles or free, could also be regulating host genes. However, miRNAs are not the only small RNA class with potential regulatory roles detected in the EVs of helminths (Cucher et al., 2021). To further characterize possible liver fluke regulatory RNAs, we compiled and analyzed the repertoire and expression of diverse small non-coding RNAs (sncRNA) in the key stages in the invasion and establishment of *F. hepatica* infection in the definitive host. Additionally, we studied the

profile of sncRNAs contained in EVs released by the adult stage. Diverse miRNAs show expression differences across the life stages analyzed, and by miRNA target prediction both on the parasite and hosts genes, we attempted to gain insights in the regulation of the internal homeostasis of the parasite and its interaction with the host. Beside these, we found a restricted set of tRNA fragments (tRFs) that are abundant in all samples analyzed. Interestingly, those enriched in EVs can form homodimer structures that are, potentially, more stable and resistant to degradation by nucleases. Furthermore, we identified a new class of ncRNA previously unreported in trematodes, vault RNAs (vtRNAs), that are abundant in EVs.

METHODS

Data Generation and Sequencing

Analyzed samples corresponding to diverse life stages (including dormant (MD) and *in vitro* activated metacercariae (MA), and adult worms (AD1; AD2) were generated by us and sequenced at The Genome Center at Washington University in St. Louis, Missouri, USA (WUGSC) within the frame of *F. hepatica* genome project (study SRP040521), (McNulty et al., 2017). Briefly, total RNA was extracted using mirVana Total Isolation kit (Thermo Fisher) in large (>200 bp) and small (<200 bp) fractions. Small RNA sequencing libraries were constructed using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England Biolabs, Beverly, MA) according to the manufacturer's instructions, and sequenced on the Illumina platform. Sequencing data is publicly available under accession numbers: SRR3584125 [MD], SRR3584124 [MA], SRR3584126 [AD1], SRR3584122 [AD2]. Newly excysted juveniles (NEJ [SRS862512]) data have been generated and analyzed by us previously (Fontenla et al., 2015).

EV samples were obtained from 50 adult worms recovered from a naturally-infected cow at a local abattoir in Dungannon, Northern Ireland. Briefly the flukes were thoroughly washed with PBS (3 x 200ml) to void their gut contents and then maintained in RPMI 1640 culture medium containing 0.1% glucose, 100 U penicillin and 100 mg/ml streptomycin (Sigma), at 2 worms/ml for 5 h at 37°C. EVs were isolated from the culture media by differential centrifugation as described by Cwiklinski et al. (2015) and treated with RNAase (Qiagen) to remove extracellular RNA not packaged within the EV. Total RNA was isolated from EV pellets using TRizol (Life Technologies) according to the manufacturer's instructions. Small RNAs were subsequently isolated from the *F. hepatica* EV RNA preparations using the SeraMir kit (SystemBio). Libraries were generated with the TailorMix Micro RNA Sample Preparation version 2 protocol (SeqMatic LLC, Fremont, USA), amplified and Illumina sequenced. EV library construction and next generation sequencing was performed by System Biosciences, US. Sequencing data was deposited at the SRA repository with the accession number PRJNA782636.

All the data are freely available from SRA repository (https://www.ncbi.nlm.nih.gov/sra) with the given accession numbers. Quality and consistency of samples were evaluated by multidimensional scaling (MDS) (Mead, 1992) showing that they were comparable, despite being obtained from field isolates since lab pure lines or strains are still unavailable for *F. hepatica*.

Quality Control and Genome Mapping

Trim galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to trim adapters and to remove reads shorter than 18 bp or with a phred score lower than 20. The remining reads were mapped to PRJEB25283 *F. hepatica* genome (https://parasite.wormbase.org) with Bowtie (Langmead et al., 2009), allowing a maximum of 2 mismatches (option -v 2). The coordinates corresponding to mRNAs, tRNAs, rRNAs, snRNAs and repetitive regions of the genome (i.e. transposons, low complexity regions or tandem repeats) were parsed out from the genome annotation, and the sequences obtained with bedtools getfasta with option -s (Quinlan and Hall, 2010). Next, Bowtie was used to classify reads by subsequently mapping to each dataset. Remaining reads were used to detect novel and conserved miRNAs.

Additionally, the presence of host-derived sequences was assessed by re-mapping the reads that failed to map to the genome of *F. hepatica*, to the genome of *Bos taurus* (downloaded from https://www.ensembl.org/Bos_taurus/Info/Index) with Bowtie (option -v 2) (**Table S1**).

A pipeline with the commands and scripts used in the analysis is available at GitHub repository (https://github.com/santifo/miRNA_analysis/blob/main/Pipeline).

Since MDS did not show significant differences in small RNA species and quantities between duplicated samples, and also between dormant and activated metacercariae (**Figure S1**), samples were considered as duplicates, pooled and averaged by stage: metacercariae (MC), adult (AD) and extracellular vesicles (EV), for further analysis.

Classification and Identification of Known and Novel miRNAs

Reads not mapped to functional RNAs were collapsed and remapped to the genome with Bowtie with identical options as before. SAM output files were converted to arf format with bwa_sam_converter.pl. MiRDeep2 pipeline (Friedländer et al., 2012). was used to identify and quantify conserved and novel miRNAs, with a "known" dataset consisting of miRNAs deposited in miRBase v22.1 (Kozomara and Griffiths-Jones, 2014) and other *F. hepatica* miRNAs identified previously, but not deposited in miRBase (Ricafrente et al., 2021). Putative novel miRNAs were manually inspected, removing those with a score ≤ 5 and/or that showed drifting in the read stacking and/or had poor folding. Furthermore, BLAST and SSEARCH tools available at miRBase were used to confirm that novel miRNAs were not detected previously in other species. Additionally, MAFFT with local alignment parameters (Katoh and Standley, 2013), was used to align novel *F. hepatica* miRNAs to all Lophotrochozoan miRNAs deposited in miRBase, to search for identical 'seed' homology. Furthermore, a BLASTn (e-value 1e-4) was

implemented to inspect if the novel precursors detected in *F. hepatica* could be conserved in any of the other platyhelminthes genomes deposited in WormBase Parasite (https://parasite.wormbase.org). No homologies were found except in the sister species *F. gigantica* (**Table S2**).

Statistical Analysis of miRNA Expression

Statistical analysis of miRNA expression was performed with DEGUST v4.1.1 (https://degust.erc.monash.edu/). Samples were grouped by stage; and analyzed with Voom/Limma method, with count per million (CPM) normalization. Differentially expressed miRNAs were defined from pairwise comparisons of the $\log_2$-transformed expression estimates, establishing a minimum fold change of 2 and a false discovery rate (FDR) corrected P-values lower or equal to 0.01. MiRNAs with a total count below 10 reads were removed from the analysis. A hierarchical clustering with hclust function, was applied to differentially expressed miRNA between metacercariae, NEJ and adult.

MiRNA Target Prediction and Functional Enrichment of *F. hepatica* Gene Targets

F. hepatica conserved and novel miRNAs were used to predict targets in an iso-seq curated database of *F. hepatica* genes (FHISCDB). This database consisted of 7626 3' untranslated regions (3'UTRs) annotated in PRJEB25283 assembly, and 1612 3'UTRs of novel transcripts obtained from a prediction based on remapping all RNA-seq data available in repositories, and iso-seq sequencing reads from intra snail stages of *F. hepatica* (Langleib et al. in preparation). Transdecoder (available at https://github.com/TransDecoder/TransDecoder/) was used to predict the open reading frame (ORFs) and 3'UTR on novel transcripts.

Three prediction algorithms were used: miRanda (Betel et al., 2010), PITA (Kertesz et al., 2007) and TargetScan v7.0 (Agarwal et al., 2015), with default options. Given that each algorithm implements different mechanisms to predict target sites [reviewed by (Peterson et al., 2014)], they usually differ in a few bases since they include different lengths outside the seed region. Therefore, to provide a consistent matching position of each target, we have reported the position on the 3'UTR matching the first base (start) of the "seed region" of the miRNA in column D of **Table S4**.

Genes defined in the FHISCDB were functionally reannotated with eggNOG-mapper (Huerta-Cepas et al., 2019). TopGO (Alexa and Rahnenfuhrer, 2019) (with option: statistic = "fisher") was used to identify enriched gene ontology (GO) terms between the gene targets of differentially expressed miRNAs, GO terms with a P-values lower or equal to 0.01 were considered significantly enriched.

miRNA Target Prediction in Mammals and Identification of Regulated Functions

The conserved and novel miRNAs of *F. hepatica* were used to predict complementary target sites in host species genes. 3'UTRs from sheep, cattle and human were downloaded from TargetScan database (http://www.targetscan.org/cgi-bin/targetscan/data_download.vert72.cgi). Although the initial idea was to use a 3 by 3 strategy, retaining only the targeted positions within the 3'UTR predicted by the three algorithms (miRanda, PITA and TargetScan v7.0) conserved in the three host species, the low quality of 3' annotation in the sheep and cattle datasets resulted in many sites conserved only in human and one of the ruminant species. Consequently, we decided to retain all those targets identified by the three algorithms in the human dataset. As described previously, predictions were run with defaults options. DAVID database (Huang et al., 2009) was used to identify regulated functions in targeted genes, considering only the pathways regulated by two or more miRNAs. Furthermore, InnateDB (Breuer et al., 2013) was used to identify the genes related to immunity under regulation.

Quantification and Folding of tRNA Derived Fragments

The reads that mapped to the tRNA database were quantified and classified using in-house scripts. Reads containing a terminal CCA-3' anchor were subsequently quantified using a similar processing, except that the tRNAs of reference were modified by adding a 'CCA' trinucleotide anchor to their 3' end. The output files were analyzed to obtain the information on alignment positions from first to end base found in the tRNA precursor. These positions were used to categorize the tRNAs fragments. The fragments resulting from the cleavage at the anticodon loop were named 5' tRNA half fragment (5'tHF) and 3' tRNA half fragment (3'tHF); if the fragment was contained but shorter than the 5' or 3' halves they were named 5' fragments (5'tRF) and 3' fragments (3'tRFs). The fragments that started at the 5' or 3' ends and were longer than the 5'tHF or 3'tHF were classified as 5' tRNA long fragments (5'tLF) or 3' tRNA long fragments (3'tLF). The fragments that were not contained within these categories or that corresponded to full tRNAs were poorly represented and were not further classified. For each tRNA fragment, the sequence with the highest count was used as input for *in-silico*, folding and dimerization prediction using the 'Fold RNA Biomolecular' tool of RNAstructure6.3 package (Bellaousov et al., 2013).

Detection of Vault RNAs

Vault RNAs are short (80 to 150 nt) polymerase III transcripts, that show no sequence conservation except in two short regions, box A and box B, that correspond to internal polymerase III promoter elements. Since finding them by traditional homology methods is unpractical, we followed a modified version of the approach described by Stadler et al. (Stadler et al., 2009). We used the most conserved region, the box A, plus the 3' flanking 'TTACTTCG' of *L. gigantea* and *H. robusta* obtained from Rfam database (RF00006 (https://rfam.xfam.org/).) to perform a relaxed BLASTn search in the genome of *F. hepatica*. Typical polymerase III-terminator motifs (consisting mainly of a run of T's) were sought in the 150 pb region downstream of the blast HSPs hits.

Putative vtRNAs were further inspected with MEME suite (Bailey et al., 2009) to confirm the conservation of relevant motifs, and RNAfold (Lorenz et al., 2011) to evaluate the

presence of the characteristic panhandle-like secondary structure. Next, we used fragrep2 (Mosig et al., 2006) pipeline to align and generate a *F. hepatica* specific pattern of vtRNAs with the conserved motifs. Fragrep2 tool implements an algorithm for detecting the pattern fragments that occur in a given order but are interrupted by non-conserved sequences of highly variable length. The pattern was used to re-search the genome. Additional putative vtRNAs detected were also evaluated as described. The genomic locations of all vtRNAs were inspected to confirm that no other features were annotated and that the vtRNAs genes were being expressed.

RESULTS

Small Non-Coding RNAs Are Present in All Stages, and Also in Secreted Vesicles

Small RNAs produced by metacercariae, newly excysted juveniles (NEJ) and adults of *F. hepatica* were analyzed and compared with novel data generated from the sequencing of small RNAs present in extracellular vesicles (EV) released by the adult stage.

Sequencing data samples were analyzed in parallel as described in the methods section. Despite differences in the initial number of raw reads, roughly 77% passed the quality filtering and trimming stages (**Table S1**). More than 80% of these high-quality reads mapped to the genome of *F. hepatica* (PRJEB25283), except for NEJ where the mapping values were lower at 66% (**Table S1**). The length distribution of the reads that mapped to the genome showed a prominent peak at 21-22 nt in all samples, consistent with the presence of miRNAs, and a secondary less prominent peak at 31-33 nt that is consistent with our previous analysis (Fontenla et al., 2015) and that corresponded to sequences that mapped to tRNAs (see below) (**Figure S2**). Reads that mapped to the genome were subsequently classified by mapping to different datasets (see Methods). While the majority of the reads mapped to coding RNA and repeated regions, a small but significant fraction corresponded to putative functional non-coding RNA. On average, 13.7% of the reads corresponded to the category of functional ncRNAs that grouped: miRNAs, tRNAs, rRNAs and vtRNAs (described here for the first time in trematodes). In all samples, miRNAs were the most abundant ncRNA followed by tRNA-derived fragments (**Figure 1**).

MiRNA Profiles Vary Greatly Between Different Stages

Compiling all the data published so far on *Fasciola hepatica* miRNAs (**Table S2**), we observe that *F. hepatica*'s miRNome is still reduced with respect to other Lophotrochozoans or even free-living flatworms [flatworms are expected to share at least 46 miRNA families (Tarver et al., 2013)]. As described previously by us and others (Fromm et al., 2013; Fontenla et al., 2015), neodermatans are characterized by a reduction in the conserved miRNA families. At present the miRNome of *F. hepatica* is composed of 34 conserved families (13 of them with more than one miRNA within the family). An almost identical complement of miRNAs was recently reported in *F. gigantica* (Hu et al., 2021) and previously in *S. mansoni* (Ovchinnikov et al., 2020) (summarized in **Figure 2A**),

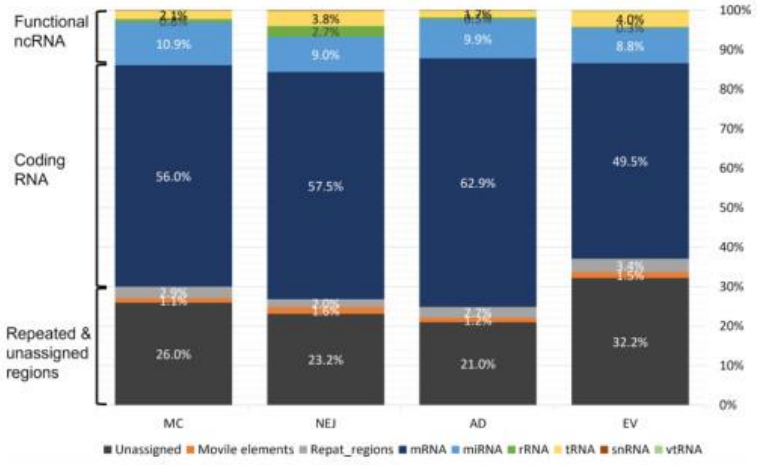


FIGURE 1 | Mapping of small RNA sequences from different stages of *Fasciola hepatica* to the genome. Samples were obtained from metacercaria (MC), newly excysted juvenile (NEJ) adult worms (AD), and adult-released extracellular vesicles (EVs). Mapping reads were classified in three main categories: functional non-coding (nc)RNA (rRNA, tRNA, small nuclear (sn)RNA, vault (vt)RNA and miRNA), coding RNAs (mRNA) and those associated to repeat and unassigned regions (mobile elements (e.g., transposons), regions with repeated sequences or regions that remain unannotated).

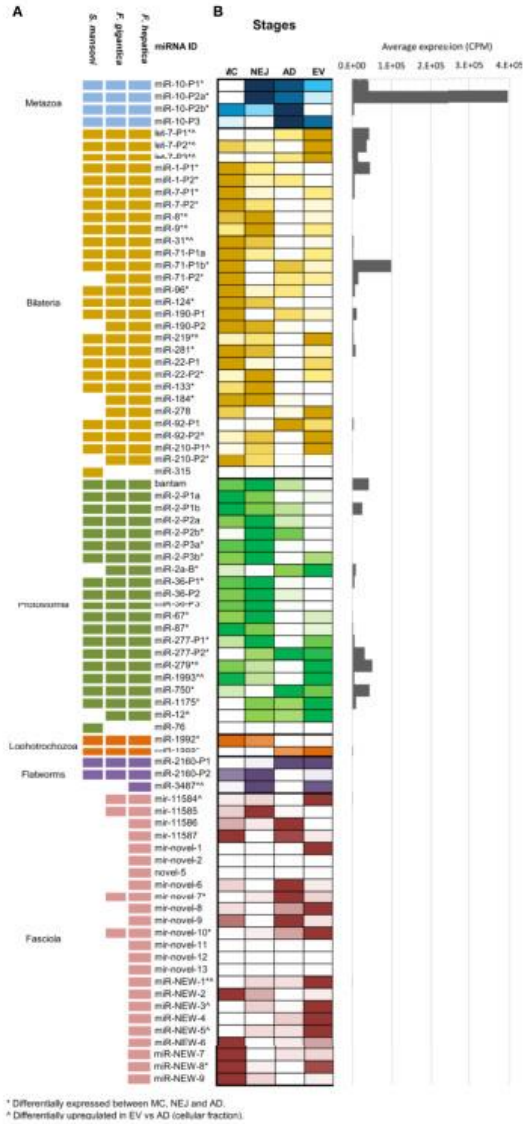


FIGURE 2 | MiRNA complement and relative abundance across species and stages. **(A)** Updated miRNA complements in *F. hepatica*, *F. gigantica* and *S. mansoni*. individual miRNAs belonging to the same miRNA family have the same id plus an individual precursor number. Shaded blocks indicate evolutive conservation across: Metazoan (light blue), Bilateria (yellow), Protostomia (green), Lophotrochozoa (orange), flatworms (purple) or Fasciola species (pink). Empty (white) blocks indicate that the miRNA has not been detected in the species **(B)** Heatmap depicting the comparative abundance of individual miRNAs between MC, NEJ, AD stages, and the adult EV fraction. Read counts were normalized to count per million (CPM) and color scale was applied. White shaded blocks indicate lower expression, darker colors indicate high abundance. (*) Indicates significant variations between cellular fractions of MC, NEJ, and AD (FDR ≤ 0.01 & Log2FC ≥ 2). (^) Indicates significantly upregulated in the EV fraction respect to the somatic fraction of AD (FDR ≤ 0.01 & Log2FC ≥ 2). A histogram with the average expression of each miRNA shows that only a reduced set are always upregulated.

indicating that these losses could have been in the origin of the adaptation to parasitism of the clade. In addition, to the conserved miRNA, 15 novel *F. hepatica* exclusive miRNAs have been described previously (Ricafrente et al., 2021), four of which were also confirmed in the miRNA complement of *F. gigantica* (Hu et al., 2021). Here, we further extended the Fasciola specific miRNome by adding nine new well supported miRNAs (fhe-miR-NEW-1 to 9 in **Figure 2** and **Figure S3**), with seven of them, also predicted to be conserved in the *F. gigantica* genome (**Table S2**).

When the normalized expression of miRNAs was compared across the intra-mammalian stages, we observed that a reduced set of evolutionary conserved families are very predominant (**Figure 2B** and **Table S3**). The metazoan conserved fhe-miR-10-P2a (named fhe-miR-125b in miRBase) was the most highly expressed miRNA across all stages and in the EV fraction of the adult stage, despite variations in the proportional representation. Other conserved miRNAs highly expressed across stages and EVs were fhe-miR-71-P1b found in all bilaterians, and bantam, a miRNA characteristic of protostomes (**Figure S4**).

Interestingly, 13 of the 28 miRNAs conserved between Bilaterians were significantly more expressed in MC respect to the other stages [(*) in **Figure 2B**], while more than a third of the protostomian specific miRNAs showed higher abundance in NEJ (**Figure 2B**). miRNAs conserved across Lophotrochozoans and Flatworms showed low expression in all samples, despite significant variations for fhe-miR-1992, fhe-miR-1989 and fhe-miR-3479. However, when we compared our results with those described in the sister species *F. gigantica* (Hu et al., 2021), we noticed that several lophotrochozoan and flatworm conserved miRNAs (e.g. miR-1992, miR-1989, miR-2160-P1 & -P2, and miR-11584), showed higher expression in egg and the intra-snail stages (**Figure S5**). Significant variations were also detected for several *Fasciola* specific miRNAs across diverse stages.

MiRNAs as Regulators of Development and Metabolism in *F. hepatica*

To analyze the role of miRNAs in the regulation of the homeostasis of *F. hepatica*, we classified the differentially expressed mature and two co-mature star miRNAs into five expression clusters (**Figure 3A**). Target sites in the transcripts of *F. hepatica* were predicted for the five clusters (**Table S4**). We found 8851 targeted positions in 3369 transcripts predicted by the three algorithms used, including 423 transcripts

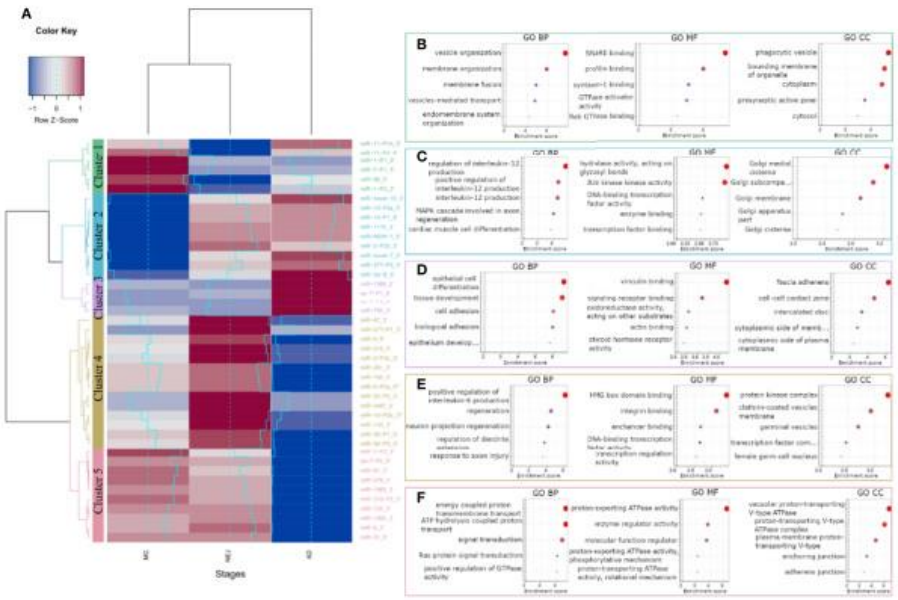


FIGURE 3 | GO terms enrichment of differentially expressed miRNAs between metacercariae, NEJ and adult stages. **(A)** Heatmap of the differences in the expression between stages. MiRNAs were grouped into 5 clusters based on their expression profiles: cluster 1 (green), cluster 2 (blue), cluster 3 (purple), cluster 4 (yellow) and cluster 5 (pink). A gradual change in color ranging from blue to red indicates the low-to-high expression, additionally light-blue line reflects the variation in the Log2FC. **(B–F)** Top 5 biological process (BP), molecular function (MF) and cellular component (CC) terms ordered by p-value, enriched in the predicted targeted genes of clusters 1-5, respectively (a color code like in A was applied). Enrichment score was calculated as -Log of p-value, terms with a score ≥ 2 (p-value = 0.01, dashed line) were considered significant.

corresponding to novel genes predicted by us (see *Methods* section, Langleib et al. in preparation). GO enrichment analysis of the targeted genes by each cluster of miRNA suggests that they might be regulating functions related to development, signaling pathways and transport, among others (**Figures 3B–F** and **Figures S6A–E**).

Cluster 1 represents a set of miRNAs highly expressed in MC while lowly expressed in NEJs, probably highlighting the regulation of relevant genes in the first interaction with the host and in the invasion process. Interestingly the genes targeted by cluster 1 miRNAs are enriched in function and processes associated with vesicle organization and transport, and membrane fusion (**Figure 3B** and **Figure S6A**).

MiRNAs strongly downregulated in MC while upregulated in NEJ and adults are grouped in cluster 2 (**Figure 3A**). Genes related to the Golgi are enriched within cluster 2 targets, suggestive of a role in vesicular cargo maturation and trafficking. Since some of the enriched targets to Cluster 2 miRNAs are genes involved in positive regulation of IL-12 signaling, the upregulation of these miRNAs in NEJ and adults might have a role dealing with the IL-12 mediated responses generated by the host macrophages and dendritic cells (**Figure 3C**). Other functions highlighted in the top 20 processes enriched like "response to inorganic substances" and "response to temperature stimuli" (**Figure S6B**) could be related to sensing stimuli relevant for excystment and development.

Cluster 3 includes miRNAs highly expressed in adults, and enrichment terms on targeted genes include: cell differentiation and tissue morphogenesis of epithelial cells and cell to cell contact. These processes are very relevant in the maturation of juvenile forms, during the formation of the syncytial tegument (**Figure 3D** and **Figure S6C**).

On the other hand, Cluster 4 contains miRNAs highly expressed in NEJ but strongly downregulated in adults. Here we found functions associated with nervous development, transcription regulation and interestingly some terms related to germinal functions, that might be associated with the production of eggs in the adult stage (**Figure 3E** and **Figure S6D**).

Cluster 5 represents miRNAs strongly downregulated in adults but expressed in MC and NEJ, and here enriched terms are related to signal transduction particularly those associated with vacuolar proton transport ATPases. These enzymes are associated with the acidification of vacuoles, a requisite for the function of several proteases relevant in feeding and the interaction with the host (**Figure 3F** and **Figure S6E**).

A Reduced Set of miRNAs Are Highly Represented in the EVs of Trematodes

The analysis of miRNAs cargo in EVs showed a sharply reduced set of miRNAs with a highly biased enrichment of particular miRNA signatures, e.g., the top 10 abundant miRNAs corresponding to 87.9% of the miRNA population. Additionally, 16 miRNAs were significantly upregulated in EVs with respect to the adult stage (**Figure 2B**), most prominently fhe-let-7-P1, -P2, -P3 and fhe-miR-279 that were among the most highly expressed miRNAs in the EV data but were significantly less dominant in the cellular fraction of the adult stage (**Figure S4**).

When we compared our *F. hepatica* EV miRNA profile with those previously reported for *F. hepatica* and *S. mansoni* (Ovchinnikov et al., 2020), we observed that the same reduced set of miRNA families were overrepresented but with differences in ranking between species and studies (**Figure 4**). For example, miR-10-P2a and miR-71-P1b were in the top 10 of the three experiments (**Figure 4A**), with the former being the most abundant both in the cellular and EV fractions in our samples of *F. hepatica*'s and in the *S. mansoni* study (Ovchinnikov et al., 2020) (**Figure 4B**). Interestingly, we observed more similarities between the profiles we obtained and those reported for *S. mansoni* than with the other *F. hepatica* experiment (**Figure 4A**). If we extend the analysis to all the adult EV miRNAs detected in trematodes (reviewed by Sotillo et al., 2020; Avni and Avni, 2021), we consistently observed that the top 10 abundant miRNAs accounted for more than 80% of the total miRNA population. These highly expressed miRNAs correspond to ancient metazoan miRNA families, including miR-10, let-7, miR-71 and the lophotrochozoan specific families: bantam, miR-2, miR-279 and miR-277, among others (**Table S5**).

Analysis of Putative Host Genes Targeted by Parasite EV-Derived miRNAs Highlights Relevant Regulatory Functions

We predicted complementary binding position in the 3' UTR of host genes for the most abundant *F. hepatica* miRNAs detected in both EVs studies (indicated in **Figure 4B**). We found 7105 targeted sites, with identical coordinates between the algorithms used, in 3385 different genes (**Table S6**). Next, we used DAVID database (https://david.ncifcrf.gov/tools.jsp) to identify enriched pathways only retaining those that were regulated by more than one miRNA. Several fundamental signaling pathways such as Ras, MAPK, PI3K-Akt, ErgB and Wnt emerged as putative enriched targets (**Figure 5**). Additionally, pathways specifically related to immune response such as TNF signaling, Platelet activation, T cell receptor signaling, B cell receptor signaling and leukocyte transendothelial migration were also found under regulation. Therefore, we consulted InnateDB (https://www.innatedb.com/), to better identify the genes related to immunity under regulation by the miRNAs contained in the EVs of *F. hepatica*. From the initial gene set identified, we only retained 77 genes that were targeted in two or more positions, as they may be under a more rigorous regulation by the miRNAs of *F. hepatica* (**Table S7**). Interestingly, the miRNAs with most targets among these genes (fhe-let-7-P1 & -P2, fhe-miR-10-P2a & -P2b, fhe-miR-1 and fhe-miR-277-P1 & -P2) were also among the most abundant miRNAs both in our data and in the previous *F. hepatica* EVs study (Ovchinnikov et al., 2020).

TRNAs-Derived Fragments Were Detected in EVs and in All Somatic Samples

Similar to what we have previously reported in NEJ only (Fontenla et al., 2015), reads mapping to tRNAs corresponded

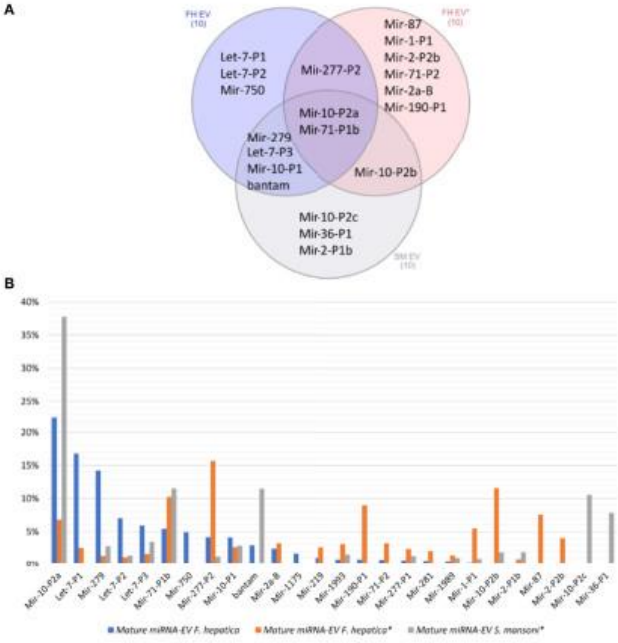


FIGURE 4 | **(A)** Venn diagram of the top 10 abundant miRNAs in the EVs of two *F. hepatica* experiments and *S. mansoni*. **(B)** Bar plot shows the percentage of the total represented by the most abundant orthologous miRNAs. (*) Data generated by (Ovchinnikov et al., 2020).

to a longer class of small RNA sequences (31-35 nucleotides in length, with a prominent peak at nucleotide 32) (**Figure S7**). These reads corresponded predominantly to diverse fragments generated from cleavage of a restricted set of tRNAs. By inspecting read alignments, we detected diverse cleavage points over the mature tRNAs resulting in distinct type of fragments (classified and described in Methods, **Figure 6A**). Overall, the most abundant class across all samples were the 5'halves of tRNAs (5'tHF) generated by cleavage in the anticodon loop, but also other fragments were detected. Shorter 5' tRF were frequently detected in metacercarial samples and were the product of cleavage before the anticodon loop or in the D loop of the precursor tRNA, while short 3'fragments were abundant in adult samples and were produced by cleavage at the T loop. Longer 5' fragments (5'tLF) generated by cleavage at the T loop were observed in EVs (**Figure 6B** and **Figure S7**). Full length tRNAs were only found in EVs, representing less than 5% of the tRNA mapping reads, although 65% of them correspond to the Selenocysteine $tRNA^{Sec\_TCA}$ (**Table S8**).

Similarly, a marked skew in representation of tRNA precursors was observed in all fragment classes. Within 5'tHFs, five precursors corresponding to $tRNA^{Gly\_GCC}$, $tRNA^{Lys\_TTT}$, $tRNA^{Cys\_GCA}$, $tRNA^{Met\_CAT}$ and $tRNA^{Asp\_GTC}$ were preferentially retained in all the samples, although in different proportions (**Figure 6B**). 5'tRFs were particularly abundant in the MC stage, with $tRNA^{Gly\_GCC}$, $tRNA^{Tyr\_GTA}$, $tRNA^{Gly\_CCC}$ representing more than half of them. These same 5'tRF species were also enriched in other stages and EVs although they were proportionally less significant. In adult samples, 3'tRF were abundant, enriched particularly in $tRNA^{Val\_CAC}$, generated by cleavage at the T-loop (in 99.4% of the cases), and to a lesser extent in the 3'tRF derived from $tRNA^{Met\_CAT}$. Within EVs, a substantial fraction (26%) of tRNA derived molecules were 5'tLFs, (corresponding to the prominent peak at nucleotide 53 in **Figure S7**), but again with a very skewed representation, with $tRNA^{Gln\_CTG}$ derived molecules representing more than 75% of their class.

Since it has been proposed that homo and heterodimerization can stabilize tRNA fragments, preventing their degradation in the extracellular medium (Tosar et al., 2018), we analyzed the preferred cleavage site of the tRNAs packed in EVs, and used the resulting fragments to predict if homodimer structures were possible (**Figure 7**). Interestingly, 5'tHF-$tRNA^{Gly\_GCC}$, 5'tLF-$tRNA^{Gln\_CTG}$ and 5'tHF-RNA^{Cys_GCA}, the most abundant in EVs, formed homodimeric structures that hid the 3' ends in a similar way to what was described for the 5' $tRNA^{Gly\_GCC}$ tHFs of humans (see **Figures 7A–D**). Despite similarities, the *F. hepatica* Gly-GCC homodimer is maintained by an uninterrupted 8 nt-long stretch of Watson and Crick or G:U

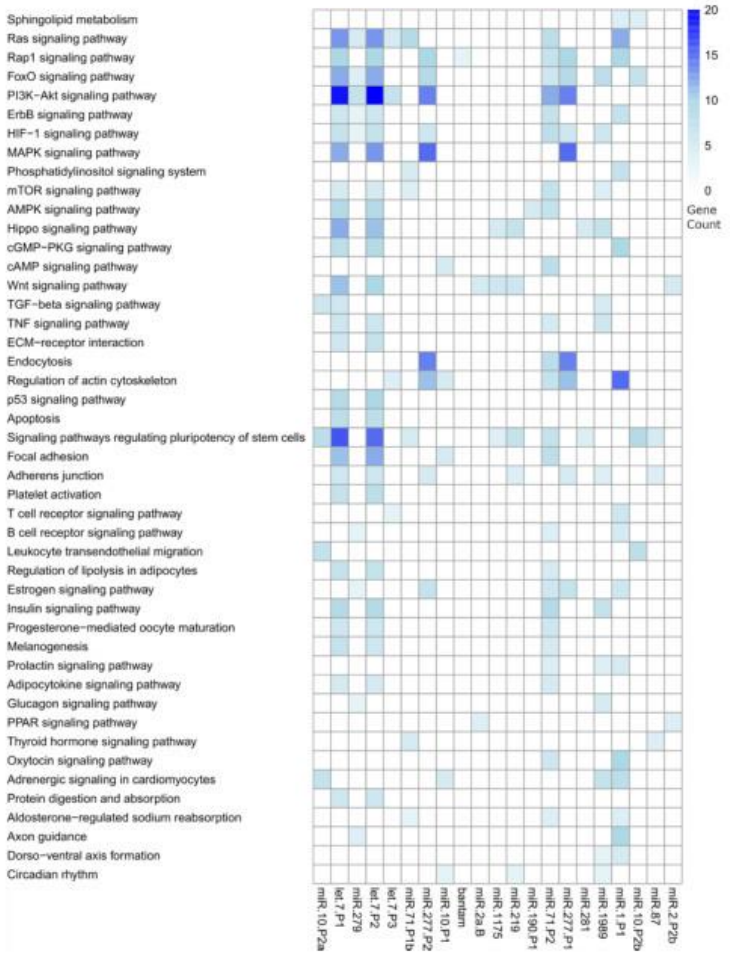


FIGURE 5 | Heatmap showing the significant pathways regulated by the miRNAs contained in the EVs of *F. hepatica*. DAVID database was used to calculate enriched pathways. Counts of genes targeted within each pathway are indicated by a color scale. The list of targeted transcripts is reported in **Table S6** and those related to immunity are presented in **Table S7**.

pairing, while there are 12 base pairings in the human counterpart (**Figures 7A, B**). Additionally, we observed the formation of a stable homodimer from a 5'tLF, generated by cleavage at the T-arm of $tRNA^{Gln\_CTG}$ (**Figure 7C**, blue arrows), around the pseudouridine, instead of the anticodon loop. This Gln_{CTG}/Gln_{CTG} homodimer tRDFs is bonded by the interaction of 10 nt, with an uninterrupted stretch of 6 nt.

While the 5'tHF-RNA^{Met_CAT} generates homodimers with hidden 3' ends, only 4 bp bonded the monomers, suggesting it might be more prone to dissociation (**Figure S8A**). In contrast, the homodimers predicted with other frequent tRNA fragments as 5'tHF-RNA^{Lys_TTT}, 5'tHF-RNA^{Asp_GTC} or 5'tLF-RNA^{Gly_TCC} generated structures with free 3'ends that are probably more susceptible to degradation by exonucleases (**Figures S8B–D**). No stable heterodimers were predicted for any of the tRNA fragments analyzed.

Vault (vt)RNAs Are Present in All Stages but Enriched in EVs

Besides the well-known presence of miRNA and tRNA fragments in EVs, the detection of the major vault proteins

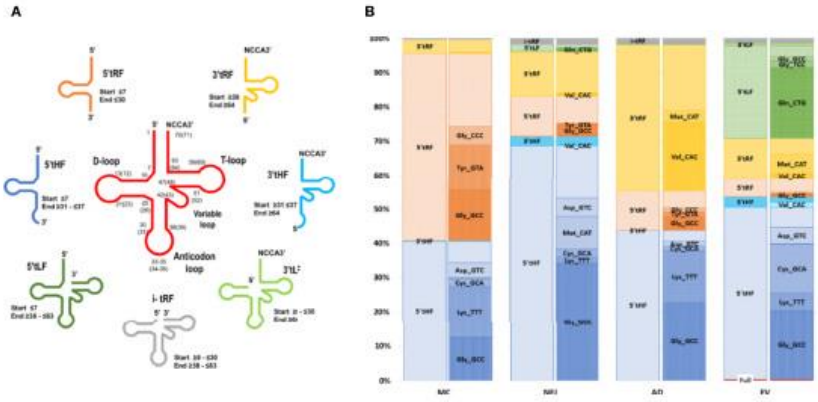


FIGURE 6 | Expression profiles across samples and folding of tRNAs-derived fragments. **(A)** Depiction of tRNA fragments detected in the study, a color code was used for each class. **(B)** Bar charts grouped by sample showing the percentage of each tRNA fragment class detected and proportion of individual fragments related to the tRNA of origin and the sequence of the anticodon.

(MVP) as part of the EVs cargo (Cwiklinski et al., 2015) raised the question if the corresponding vtRNAs were also present. Vault particles are short polymerase III transcripts with lengths varying between 80 and 150 nucleotides, with sequence conservation restricted to two short regions, box A and box B, that correspond to internal polymerase III promoter elements (**Figure 8A**).

For the detection of vtRNAs, we applied a combination of relaxed homology and pattern searches, motif detection, structural prediction and manual curation (described in detail in 'Methods'). We detected 5 putative vtRNAs that passed the filtration process. By using MEME suite, we found two common motifs between the putative *F. hepatica*'s vtRNAs and recognized animal vtRNAs sequences (motifs 1 and 3 in **Figure 8A**). While the conserved box A was included in motif 1, the box B included in motif 2 was not detected by MEME suite and was inferred from the sequence alignment. The sequence conservation of box B was very low being limited mostly to the sequence 'GTTC' or one nucleotide variations of it (**Figure 8A**). The folding prediction of the putative vtRNAs produced panhandle-like secondary structures with an extended stem-loop connecting 5' and 3' ends similar to those characterized in other animals (**Figure 8B**) (Stadler et al., 2009).

Interestingly four of the five putative vtRNAs detected were clustered in a 133.3 Kb region on the scaffold 1448, with the remaining structure (vtRNA4) in another scaffold (scaffold 1781, **Table S9**). While all of them were expressed in all stages analyzed, vtRNA3 was predominant in all samples, and particularly enriched in EVs (more than 20-fold in relation to the corresponding adult stage) (**Figure 8C**).

When we inspected the genomic regions predicted to contain vtRNA in the wormbase parasite database, we found good RNA-seq reads coverage in samples from different stages and diverse experiments, confirming their expression. Once again, vtRNA3 was the most expressed overall (**Figure S9**).

DISCUSSION

In this study we analyzed the presence of diverse small RNA populations across distinct life stages of *F. hepatica*, and those present in vesicles secreted by the adult stage. Although samples were obtained and sequenced at different times, those from similar stages tend to cluster in multidimensional analysis, suggesting they represent the population consistently found at these stages (**Figure S1**). Furthermore, read data were normalized to account for differences in sequencing depth. A complex set of different small RNAs was identified in all samples analyzed, composed by miRNAs, tRNA fragments and the longer vtRNAs.

MiRNAs Are Associated With Host Invasion and Development of *F. hepatica*

Our analysis of the miRNome of the main stages of *F. hepatica* associated with the definitive host, adds nine novel miRNAs to those already described and recompiled recently (Ricafrente et al., 2021). The comparison with a recent study in the sister species *F. gigantica* (Hu et al., 2021), confirms the presence of 34 conserved families and a growing set of miRNAs described so far only in Fasciolidae.

Additionally, we *in-silico* correlated the differentially expressed miRNAs with the putative biological processes under regulation, with results suggestive of relevant roles in the normal development of the parasite and possible roles in the invasion process. Even though, our results await further experimental validation, the regulatory roles of miRNAs are well known

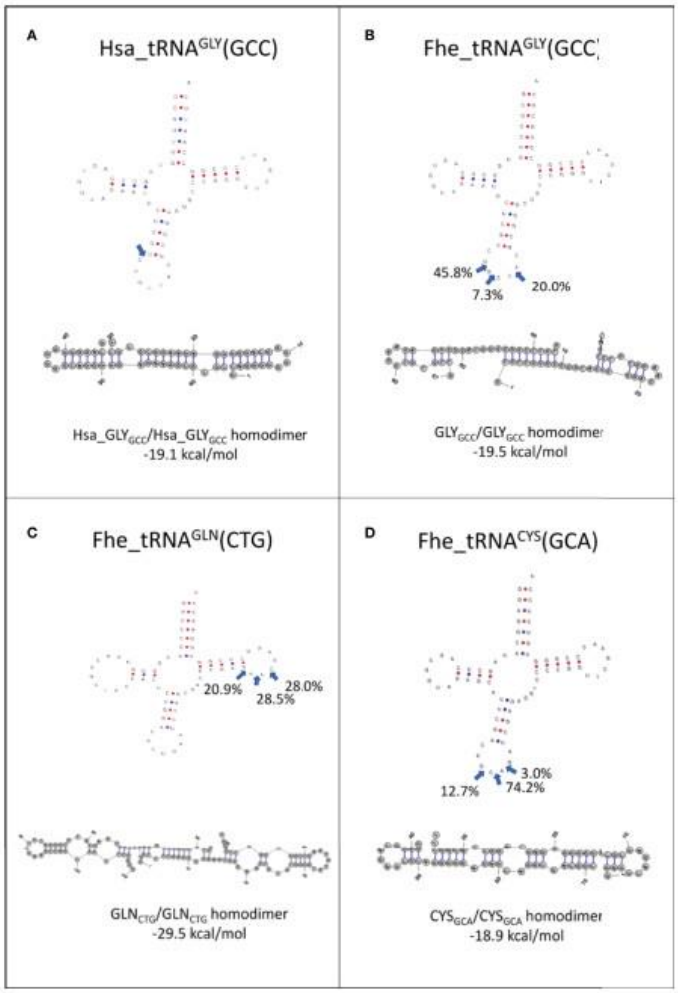


FIGURE 7 | Folding of tRNAs-derived fragments. Full length tRNA secondary structures were predicted with tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/), position and frequency of the most frequent cleavage sites in EVs are indicated (blue arrows). Homodimer structures and energy in kcal/mol for: **(A)** Hsa_Gly$_{GCC}$, **(B)** Fhe_Gly$_{GCC}$, **(C)** Fhe_Gln$_{CTG}$ and **(D)** Fhe_Cys$_{GCA}$ as predicted by RNAstructure6.3 for the most abundant fragment.

(Bushati and Cohen, 2007). Studies in the model animals *C. elegans* and *Drosophila*, show that the most conserved animal miRNAs are abundant and usually involved in core developmental process, while scarce miRNAs are usually related to specific functions, and sometimes expressed in very restricted groups of cells or tissues (Alberti and Cochella, 2017). In this sense, in planarians, particular miRNA families (also conserved in *F. hepatica)*, were found to show preferential expression in neoblasts or to be upregulated in the regenerating tissue (González-Estévez et al., 2009; Sasidharan et al., 2013; Cao et al., 2020). It is possible that similar cell type-miRNA preference occurs in *F. hepatica*, however, additional experiments are needed to corroborate this hypothesis. In parasitic neodermatans, the role of miRNAs in the regulation of cellular processes, up until now, was mainly restricted to descriptions in Schistosomes (Sun et al., 2014; Zhu et al., 2016; Protasio et al., 2017; Yu et al., 2019), *Echinococcus* spp. (Guo et al., 2017; Macchiaroli et al., 2017; Pérez et al., 2019; Bai et al., 2020;

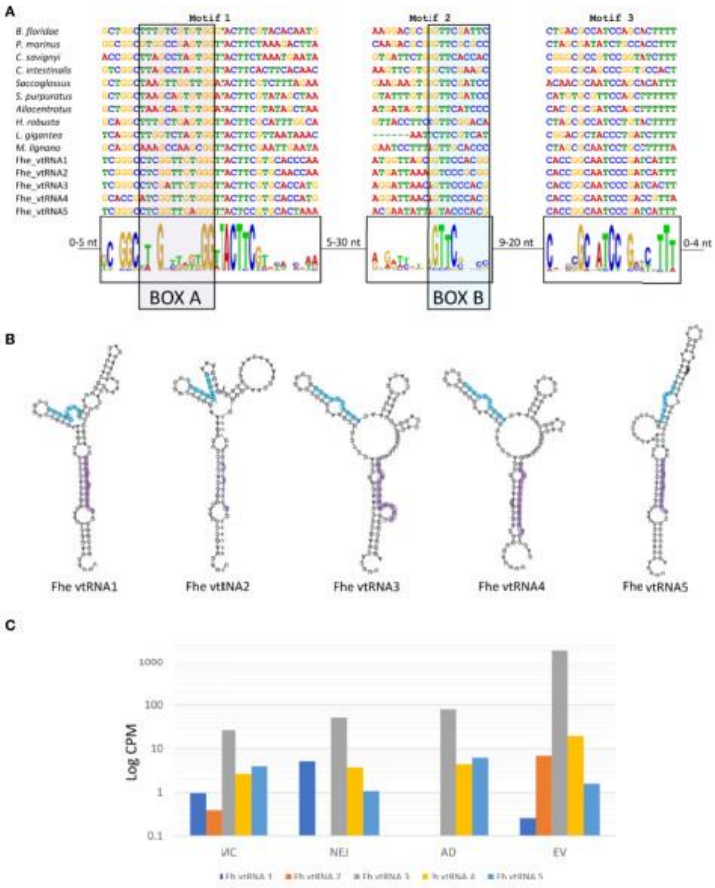


FIGURE 8 | Structure and expression of vtRNAs detected in the genome of *F. hepatica*. **(A)** *F. hepatica* vtRNAs were compared with vtRNAs detected in other species; chordata: *Petromyzon marinus*, *Branchiostoma floridae*, *Ciona intestinalis* and *C. savigni*; hemichordate: Sacoglosus; Echinodermata, *Strongilocentrotus purpuratus* and Allocentrotus; Annelida: *Helobdella robusta*; Mollusca: *Lottia gigantea*; Platyhelminthes: *Macrostomum lignano*. Motifs 1 and 3 were detected with MEME suite, motif 2 was inferred from the sequence alignment. Motif 1 and motif 2 contain the internal promoter sites of RNA polymerase III [indicated as BOX A (highlighted in purple) and BOX B (highlighted in light blue)]. The sequence range in the variable region between the motifs are indicated. **(B)** Fhe-vtRNAs panhandle-like secondary structure, Box A is highlighted. **(C)** Bar chart with the expression profile of fhe vtRNAs in the stages analyzed. Reads were averaged between samples, normalized to count per million and a logarithmic scale was applied to improve data visualization.

Macchiaroli et al., 2021) and more recently *F. gigantica* (Hu et al., 2021). Orthologous miRNAs of those described in these studies are present in *F. hepatica*, suggestive of common regulatory pathways. Interestingly, our enrichment study highlighted different clusters of miRNA expression, that in turn might regulate diverse processes. Is interesting that several of the miRNAs (fhe-miR-71-P1b, fhe-miR-71-P2, fhe-miR-1-P1, fhe-miR-1-P2, fhe-miR-96 and fhe-miR-7-P1) might be regulating the formation and release of vesicles. A recent characterization of EVs in different developmental stages of *F. hepatica* has reported the secretion of large amounts of different types of vesicles in NEJ. These vesicles were found to be pre-formed in the metacercariae (Sánchez-López et al., 2020). Furthermore, secretory vesicles with distinct morphologies were observed in the tegument of *F. gigantica* NEJs before and after penetration of the host intestine indicating that secretion by this life-cycle stage requires a fine level of control (Hanna et al., 2019). Therefore, the change in the expression of the miRNAs of cluster 1 (**Figure 3A**) could be, at

least in part, regulating the release of the preformed vesicles until appropriate signals from the host trigger their release from the apical plasma membrane.

MiRNAs Contained in *F. hepatica* EVs May Target Immune-Related Host Genes

During the establishment of the infection the strategy adopted by helminth parasites is to manipulate and modulate immunity in order to defuse immune defenses. Multiple mediators (i.e. proteins, glycans, lipids and nucleic acids) intervene in this process (Coakley et al., 2016). The presence of miRNAs and other small RNAs packed in EVs suggests that they might also be involved in this regulation.

Reasoning in a dose-effect manner, we studied the possible effects of the most abundant EV miRNAs, by predicting targeted sites in the 3' UTR of host genes. Several essential pathways like TGF-β, MAPK, PI3K/AKT, Wnt signaling and particular pathways related to the establishment of an immune response were highlighted as putative enriched targets (**Figure 5**), suggesting a possible role in modulating host's responses. These results are consistent with previous reports in helminths (reviewed by (Arora et al., 2017)) and with more recent publications that highlight several signaling pathways (Wnt, MAPK, TNF and NOD-like) as potential targets of *F. hepatica*'s miRNAs (Ovchinnikov et al., 2020; Tran et al., 2021). In this sense, fhe-miR-125b (named miR-10-P2a here) was found as the most abundant parasitic miRNA within peritoneal macrophages of infected mice (Tran et al., 2021). Similarly, parasite-derived sma-miR-10 and sma-bantam were found in cells isolated from Peyers patches and mesenteric lymph nodes of *S. mansoni* infected mice (Meningher et al., 2020). A reduction of MAP3K7 expression and reduced NF-κB activity by sma-miR-10 was observed *in vitro*, suggesting a mechanism for the downregulation of the Th2 response. Additionally, sja-bantam and sja-miR-125b were found to induce Th2 key mediator, TNF-α in macrophages *in vitro* and *in vivo*, facilitating parasite development and egg deposition (Liu et al., 2019). Furthermore, parasite-derived let-7 has been predicted to regulate the Wnt signaling pathway and T/B cell activation in mouse genes, and to induce a Th2 immune response in macrophages treated with miRNA mimics in cestodes (Ancarola et al., 2017; Wang et al., 2021). Th2 modulation is a common theme in parasite immune-evasion (Coakley et al., 2016). Interestingly, the miR-10, bantam and let-7 family members are consistently amongst the most abundant miRNA signatures in EVs secreted by adult trematodes (**Table S5**), suggesting that their enrichment is not produced by chance, therefore, constituting central players in modulating host responses to the parasite.

EVs Pack Homodimeric tRDFs That Could Resist Exonuclease Degradation

From early on, tRNA-derived fragments (tRDF) have been claimed to be involved in the regulation of cellular stress, since their emergence was upregulated in cells exposed to different types of stress stimuli. These tRDFs have been linked to diverse processes including translational regulation, proliferation, apoptosis, stress-granule formation, mRNA stabilization, transposon expression, ribosome biogenesis and the inheritance of acquired traits (Tosar and Cayota, 2020).

Our analysis highlighted that a diverse set of fragments derived from a restricted set of tRNAs were present in different stages related to the mammalian invasion, and in the extracellular vesicles of the adult worm. While 5' HF from almost every tRNA precursor can be found, their abundance is skewed with those derived from $tRNA^{Gly\_GCC}$ as the prominent in all the stages analyzed, and only four others (namely Lys_TTT, Cys_GCA, Asp_GTC, and Met_CAT) highly represented in somatic or EVs samples (**Figure 6B**). Similarly, shorter tRNA fragments detected showed a marked skew in representation, with those derived of $tRNA^{Gly\_GCC}$ again as one of the predominant classes. Interestingly, we detected the presence of a novel type of fragment generated by preferential cleavage at the T loop of the precursor, derived mainly from $tRNA^{Gln\_CTG}$ that were highly abundant in EVs. Moreover, most of the tRNA precursors highlighted here were also abundant along the life cycle of *F. gigantica* (Hu et al., 2021) or in the EVs of *S. mansoni* (Nowacki et al., 2015). A marked skew in the representativity of the tRNA derived molecules contained in EVs have also been reported in several species of nematodes, trematodes and cestodes was also evident (Nowacki et al., 2015; Quintana et al., 2019; Wang et al., 2020; Cucher et al., 2021). This could be a consequence of sub-representation due to the difficulties imposed by modified bases in the amplification/sequencing process (Potapov et al., 2018). Novel methods relying on different enzymes and adapters are being developed to overcome this technical restriction, allowing the capture and sequencing of RNAs containing modified bases (Qin et al., 2016; Xu et al., 2019; Shi et al., 2021; Tosar et al., 2021). Surely these methods would provide in the future a more complex picture of the tRNA derived fragments. But so far, we could analyze and compare those detected by the more traditional small RNA sequencing methodologies in different species. On the other hand, the repeated enrichment of certain fragments across stages and species might correspond to their increased resistance to degradation. It has been reported that human tRNA fragments form homodimers and heterodimers that hide the 3'ends and show resistance to degradation by exonucleases (Tosar et al., 2018). We show that the three abundant tRNA fragments found in *F. hepatica* could form homodimers with hidden 3' ends, similar to those described in humans. While in humans the tRDFs with dimerization capacity were found to be more abundant in the extracellular space, not associated to the EV fraction (Tosar et al., 2018), we found them in EVs, but we did not investigate so far their presence as free circulating tRDFs.

The roles of tRDFs are still poorly understood in platyhelminthes, however it was recently reported that they may be involved in the regeneration of planaria (Cao et al., 2020). Interestingly, in these organisms, Piwi proteins and Ago1, but not Dicer or Ago2, were associated with the generation and/or function of 5'tHFs and 5'tRFs, respectively (Lakshmanan et al., 2021). However, the absence of Piwi genes in all Neodermatans

implies that, at least the biogenesis of 5'tHFs could be different in trematodes and cestodes (Fontenla et al., 2017; Fontenla et al., 2021) to those described in free-living flatworms. In any case, many questions related to the mechanisms of generation and the possible roles of tRDFs remain open.

VtRNAs Are Enriched in the Secreted Fraction

Vault ribonucleoproteins are large hollow barrel-like shape particles in the cytoplasm of many eukaryotic cells. Although little is known yet on their function, they have been involved in key regulatory roles including autophagy, apoptosis and modulation of gene expression [reviewed by Frascotti et al., (2021)]. Even more, recent evidence has shown that human vtRNA1-1 could guide sequence-specific cleavage of a complementary target RNA (Frascotti et al., 2021).

Within flatworms vtRNAs have only been identified previously in *Macrostomum lignano* (https://rfam.xfam.org/), and although they were expected to be present in *S. mansoni* since a homolog of the Major Vault Protein (MVP) is coded in its genome, no clear sequences were identified (Copeland et al., 2009). Proteomic analysis of *F. hepatica* EVs detected the presence of MVPs, raising the question if the corresponding vtRNAs were also present. Our approach based on a relaxed homology search with a short but most conserved region of phylogenetically proximal lophotrochozoan vtRNAs, followed by a thorough manual curation was key to detect 5 putative vtRNAs in *F. hepatica*. The identified sequences have short, conserved regions similar to other vtRNAs and can be folded producing similar structures, suggesting they are bona-fide vtRNAs. Furthermore, a rapid analysis of available transcriptomic data provided evidence of their expression in diverse samples and stages. Taken together, these results suggest that a similar approach can now be followed to seek their presence in other trematodes. Although the functional role of vtRNAs from *F. hepatica* is unknown, their selective packaging into fluke EVs suggests that they could participate in host-parasite interactions.

CONCLUSIONS

We have analyzed the profiles of expression of the small RNAs complement in three intra-mammalian stages of the life cycle of *F. hepatica*. We consistently detected the presence of miRNAs, tRNA derived molecules and vtRNAs in all the samples. This is the first description of the still little known vtRNAs in trematodes. Interestingly the same three types of RNAs were present in EVs generated by the adult worms, stressing their putative role in crosstalk to the host.

Within miRNAs, those more abundant in EVs correspond to conserved families predicted to target several host signaling pathways. Interestingly, this seems to be a common theme in helminths, with increasing reports of uptake of parasite-derived miRNAs by host cells, and *in vitro* evidence of downregulation of host genes associated with the immune response (Arora et al., 2017). Additionally, a skewed population of tRNA fragments were detected in all the stages analyzed. We described here a new class of tRNA fragment, produced by the cleavage at the T-loop particularly abundant in the EV fraction. The most abundant tRNA fragments of the EV fraction can form stable homodimeric structures that might explain their increased stability. The roles of these tRNA fragments in regulation of *F. hepatica* and/or in the interaction with the host are speculative and await further validation.

Indeed, further experimental approaches are needed to understand the roles of all these small RNA classes; their combined presence in EVs suggests a concerted action in the interaction and modulation of the host responses, that deserves to be investigated.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available in NCBI under accession number PRJNA782636.

AUTHOR CONTRIBUTIONS

Conceptualization: SF and JT. Samples preparation and sequencing: ET and MR. Analysis: SF and ML. Drafting the manuscript: SF, MR, MD, and JT. All authors contributed to the article and approved the submitted version.

FUNDING

The research was founded by CSIC grant 521_2019 and BBSRC grant BB/L019612/1. SF, MD, and JT are part of the National Researchers System (SNI). JT is part of the Basic Sciences Development Program (PEDECIBA).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.812141/full#supplementary-material

REFERENCES

Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P. (2015). Predicting Effective microRNA Target Sites in Mammalian mRNAs. *eLife* 4, 1–38. doi: 10.7554/eLife.05005

Alberti, C., and Cochella, L. (2017). A Framework for Understanding the Roles of miRNAs in Animal Development. *Development* 144, 2548–2559. doi: 10.1242/dev.146613

Alexa, A., and Rahnenfuhrer, J. (2019). topGO: Enrichment Analysis for Gene Ontology. R package version 2.36.0. doi: 10.18129/B9.bioc.topGO

Ancarola, M. E., Marcilla, A., Herz, M., Macchiaroli, N., Pérez, M., Asurmendi, S., et al. (2017). Cestode Parasites Release Extracellular Vesicles With microRNAs and Immunodiagnostic Protein Cargo. *Int. J. Parasitol.* 47, 675–686. doi: 10.1016/j.ijpara.2017.05.003

Arora, N., Tripathi, S., Singh, A. K., Mondal, P., Mishra, A., and Prasad, A. (2017). Micromanagement of Immune System: Role of miRNAs in Helminthic Infections. *Front. Microbiol.* 8:586. doi: 10.3389/fmicb.2017.00586

Avni, D., and Avni, O. (2021). Extracellular Vesicles: Schistosomal Long-Range Precise Weapon to Manipulate the Immune Response. *Front. Cell. Infect. Microbiol.* 11, 649480. doi: 10.3389/fcimb.2021.649480

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME Suite: Tools for Motif Discovery and Searching. *Nucleic Acids Res.* 37, 202–208. doi: 10.1093/nar/gkp335

Bai, Y., Zhang, Z., Jin, L., Zhu, Y., Zhao, L., Shi, B., et al. (2020). Dynamic Changes in the Global Transcriptome and MicroRNAome Reveal Complex miRNA-mRNA Regulation in Early Stages of the Bi-Directional Development of Echinococcus Granulosus Protoscoleces. *Front. Microbiol.* 11, 654. doi: 10.3389/fmicb.2020.00654

Bartel, D. P. (2009). MicroRNA Target Recognition and Regulatory Functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002.MicroRNA

Basika, T., Macchiaroli, N., Cucher, N., Espinola, S., Kamenetzky, L., Zaha, A., et al. (2016). Identification and Profiling of microRNAs in Two Developmental Stages of Mesocestoides Corti, a Model Cestode Parasite. *Mol. Biochem. Parasitol.* 210, 37–49. doi: 10.1016/j.molbiopara.2016.08.004

Bellaousov, S., Reuter, J. S., Seetin, M. G., and Mathews, D. H. (2013). RNAstructure: Web Servers for RNA Secondary Structure Prediction and Analysis. *Nucleic Acids Res.* 41, W471–W474. doi: 10.1093/nar/gkt290

Bernal, D., Trelis, M., Montaner, S., Cantalapiedra, F., Galiano, A., Hackenberg, M., et al. (2014). Surface Analysis of Dicrocoelium Dendriticum. The Molecular Characterization of Exosomes Reveals the Presence of miRNAs. *J. Proteomics* 105, 232–241. doi: 10.1016/j.jprot.2014.02.012

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive Modeling of microRNA Targets Predicts Functional Non-Conserved and Non-Canonical Sites. *Genome Biol.* 11, R90. doi: 10.1186/gb-2010-11-8-r90

Breuer, K., Foroushani, A. K., Laird, M. R., Chen, C., Sribnaia, A., Lo, R., et al. (2013). InnateDB: Systems Biology of Innate Immunity and Beyond - Recent Updates and Continuing Curation. *Nucleic Acids Res.* 41, 1228–1233. doi: 10.1093/nar/gks1147

Buck, A. H., Coakley, G., Simbari, F., McSorley, H. J., Quintana, J. F., le Bihan, T., et al. (2014). Exosomes Secreted by Nematode Parasites Transfer Small RNAs to Mammalian Cells and Modulate Innate Immunity. *Nat. Commun.* 5, 5488. doi: 10.1038/ncomms6488

Bushati, N., and Cohen, S. M. (2007). MicroRNA Functions. Annual Review of Cell and Developmental Biology. *Annu. Rev. Cell Dev. Biol.* 23, 175–205. doi: 10.1146/annurev.cellbio.23.090506.123406

Cai, P., Gobert, G. N., and McManus, D. P. (2016). MicroRNAs in Parasitic Helminthiases: Current Status and Future Perspectives. *Trends Parasitol.* 32, 71–86. doi: 10.1016/j.pt.2015.09.003

Cai, P., Gobert, G. N., You, H., Duke, M., and McManus, D. P. (2015). Circulating miRNAs: Potential Novel Biomarkers for Hepatopathology Progression and Diagnosis of Schistosomiasis Japonica in Two Murine Models. *PloS Negl. Trop. Dis.* 9, e0003965. doi: 10.1371/journal.pntd.0003965

Cao, Z., Rosenkranz, D., Wu, S., Liu, H., Pang, Q., Zhang, X., et al. (2020). Different Classes of Small RNAs Are Essential for Head Regeneration in the Planarian Dugesia Japonica. *BMC Genomics* 21, 1–11. doi: 10.1186/s12864-020-07234-1

Cheng, G., Luo, R., Hu, C., Cao, J., and Jin, Y. (2013). Deep Sequencing-Based Identification of Pathogen-Specific microRNAs in the Plasma of Rabbits Infected With Schistosoma Japonicum. *Parasitology* 140, 1751–1761. doi: 10.1017/S0031182013000917

Coakley, G., Buck, A. H., and Maizels, R. M. (2016). Host Parasite Communications–Messages From Helminths for the Immune System: Parasite Communication and Cell-Cell Interactions. *Mol. Biochem. Parasitol.* 208, 33–40. doi: 10.1016/j.molbiopara.2016.06.003

Copeland, C. S., Marz, M., Rose, D., Hertel, J., Brindley, P. J., Santana, C. B., et al. (2009). Homology-Based Annotation of Non-Coding RNAs in the Genomes of Schistosoma Mansoni and Schistosoma Japonicum. *BMC Genomics* 10, 464. doi: 10.1186/1471-2164-10-464

Cucher, M. A., Ancarola, M. E., and Kamenetzky, L. (2021). The Challenging World of Extracellular RNAs of Helminth Parasites. *Mol. Immunol.* 134, 150–160. doi: 10.1016/j.molimm.2021.03.011

Cwiklinski, K., de la Torre-Escudero, E., Trelis, M., Bernal, D., Dufresne, P. J., Brennan, G. P., et al. (2015). The Extracellular Vesicles of the Helminth Pathogen, Fasciola Hepatica : Biogenesis Pathways and Cargo Molecules Involved in Parasite Pathogenesis. *Mol. Cell. Proteomics* 14, 3258–3273. doi: 10.1074/mcp.M115.053934

Dalton, J. P., Robinson, M. W., Mulcahy, G., O'Neill, S. M., and Donnelly, S. (2013). Immunomodulatory Molecules of Fasciola Hepatica: Candidates for Both Vaccine and Immunotherapeutic Development. *Vet. Parasitol.* 195, 272–285. doi: 10.1016/j.vetpar.2013.04.008

de la Torre-Escudero, E., Gerlach, J. Q., Bennett, A. P. S., Cwiklinski, K., Jewhurst, H. L., Huson, K. M., et al. (2019). Surface Molecules of Extracellular Vesicles Secreted by the Helminth Pathogen Fasciola Hepatica Direct Their Internalisation by Host Cells. *PloS Negl. Trop. Dis.* 13, e0007087. doi: 10.1371/journal.pntd.0007087

Fontenla, S., Dell'Oca, N., Smircich, P., Tort, J. F., and Siles-Lucas, M. (2015). The Mirnome of Fasciola Hepatica Juveniles Endorses the Existence of a Reduced Set of Highly Divergent Micro RNAs in Parasitic Flatworms. *Int. J. Parasitol.* 45, 901–913. doi: 10.1016/j.ijpara.2015.06.007

Fontenla, S., Rinaldi, G., Smircich, P., and Tort, J. F. (2017). Conservation and Diversification of Small RNA Pathways Within Flatworms. *BMC Evolut. Biol.* 17, 215. doi: 10.1186/s12862-017-1061-5

Fontenla, S., Rinaldi, G., and Tort, J. F. (2021). Lost and Found: Piwi and Argonaute Pathways in Flatworms. *Front. Cell. Infect. Microbiol.* 11, 653695. doi: 10.3389/fcimb.2021.653695

Frascotti, G., Galbiati, E., Mazzucchelli, M., Pozzi, M., Salvioni, L., Vertemara, J., et al. (2021). The Vault Nanoparticle: A Gigantic Ribonucleoprotein Assembly Involved in Diverse Physiological and Pathological Phenomena and an Ideal Nanovector for Drug Delivery and Therapy. *Cancers* 13, 1–37. doi: 10.3390/cancers13040707

Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., and 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., Trelis, M., Hackenberg, M., Cantalapiedra, F., Bernal, D., and Marcilla, A. (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. *Int. J. Parasitol.* 45, 697–702. doi: 10.1016/j.ijpara.2015.06.002

Fromm, B., Worren, M. M., Hahn, C., Hovig, E., and Bachmann, L. (2013). Substantial Loss of Conserved and Gain of Novel MicroRNA Families in Flatworms. *Mol. Biol. Evol.* 30, 2619–2628. doi: 10.1093/molbev/mst155

Garcia, H. H., Moro, P. L., and Schantz, P. M. (2007). Zoonotic Helminth Infections of Humans: Echinococcosis, Cysticercosis and Fascioliasis. *Curr. Opin. Infect. Dis.* 20, 489–494. doi: 10.1097/QCO.0b013e3282a95e39

González-Estévez, C., Arseni, V., Thambyrajah, R. S., Felix, D. A., and Aboobaker, A. A. (2009). Diverse miRNA Spatial Expression Patterns Suggest Important Roles in Homeostasis and Regeneration in Planarians. *Int. J. Dev. Biol.* 53, 493–505. doi: 10.1387/ijdb.082825cg

Guo, X., and Guo, A. (2019). Profiling Circulating microRNAs in Serum of Fasciola Gigantica-Infected Buffalo. *Mol. Biochem. Parasitol.* 232, 1–5. doi: 10.1016/j.molbiopara.2019.111201

Guo, X., Zhang, X., Yang, J., Jin, X., Ding, J., Xiang, H., et al. (2017). Suppression of Nemo-Like Kinase by miR-71 in Echinococcus Multilocularis. *Exp. Parasitol.* 183, 1–5. doi: 10.1016/j.exppara.2017.10.004

Hanna, R. E. B., Moffett, D., Robinson, M. W., Jura, W. G. Z. O., Brennan, G. P., and Fairweather, I. (2019). Fasciola Gigantica: Comparison of the Tegumental Ultrastructure in Newly Excysted Metacercariae and In Vitro Penetrated Juvenile Flukes Indicates Intracellular Sources of Molecules With Vaccinal and Immunomodulatory Potential. *Vet. Parasitol.* 265, 38–47. doi: 10.1016/j.vetpar.2018.11.019

Harrington, D., Lamberton, P. H. L., and McGregor, A. (2017). Human Liver Flukes. *Lancet Gastroenterol. Hepatol.* 2, 680–689. doi: 10.1016/S2468-1253(17)30111-5

Hoy, A. M., Lundie, R. J., Ivens, A., Quintana, J. F., Nausch, N., Forster, T., et al. (2014). Parasite-Derived MicroRNAs in Host Serum As Novel Biomarkers of Helminth Infection. *PloS Negl. Trop. Dis.* 8, e2701. doi: 10.1371/journal.pntd.0002701

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). EggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Hu, R. S., Zhang, X. X., Ma, Q. N., Elsheikha, H. M., Ehsan, M., Zhao, Q., et al. (2021). Differential Expression of microRNAs and tRNA Fragments Mediate the Adaptation of the Liver Fluke Fasciola Gigantica to Its Intermediate Snail and Definitive Mammalian Hosts. *Int. J. Parasitol.* 51, 405–414. doi: 10.1016/j.ijpara.2020.10.009

Johnstone, R. M., Adam, M., Hammond, J. R., Orr, L., and Turbide, C. (1987). Vesicle Formation During Reticulocyte Maturation. Association of Plasma Membrane Activities With Released Vesicles (Exosomes). *J. Biol. Chem.* 262, 9412–9420. doi: 10.1016/s0021-9258(18)48095-7

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Keiser, J., and Utzinger, J. (2009). Food-Borne Trematodiases. *Clin. Microbiol. Rev.* 22, 466–483. doi: 10.1128/CMR.00012-09

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The Role of Site Accessibility in microRNA Target Recognition. *Nat. Genet.* 39, 1278–1284. doi: 10.1038/ng2135

Kozomara, A., and Griffiths-Jones, S. (2014). Mirbase: Annotating High Confidence microRNAs Using Deep Sequencing Data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181

Lakshmanan, V., Sujith, T. N., Bansal, D., Shivaprasad, P., Palakodeti, D., and Krishna, S. (2021). Comprehensive Annotation and Characterization of Planarian tRNA and tRNA-Derived Fragments (tRFs). *Rna* 27, 477–495. doi: 10.1261/rna.077701.120

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome. *Genome Biol.* 10, R25. doi: 10.1186/gb-2009-10-3-r25

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. Elegans Heterochronic Gene Lin-4 Encodes Small RNAs With Antisense Complementarity to Lin-14. *Cell* 75, 843–854. doi: 10.1016/0092-8674(93)90529-Y

Lin, Y., Zhu, S., Hu, C., Wang, J., Jiang, P., Zhu, L., et al. (2019). Cross-Species Suppression of Hepatoma Cell Growth and Migration by a Schistosoma Japonicum MicroRNA. *Mol. Ther. - Nucleic Acids* 18, 400–412. doi: 10.1016/j.omtn.2019.09.006

Liu, J., Zhu, L., Wang, J., Qiu, L., Chen, Y., Davis, R. E., et al. (2019). Schistosoma Japonicum Extracellular Vesicle Mirna Cargo Regulates Host Macrophage Functions Facilitating Parasitism. *PLoS Pathog.* 15, 1–27. doi: 10.1371/journal.ppat.1007817

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6, 26. doi: 10.1186/1748-7188-6-26

Macchiaroli, N., Maldonado, L. L., Zarowiecki, M., Cucher, M., Gismondi, M. I., Kamenetzky, L., et al. (2017). Genome-Wide Identification of microRNA Targets in the Neglected Disease Pathogens of the Genus Echinococcus. *Mol. Biochem. Parasitol.* 214, 91–100. doi: 10.1016/j.molbiopara.2017.04.001

Macchiaroli, N., Preza, M., Pérez, M. G., Kamenetzky, L., Cucher, M., Koziol, U., et al. (2021). Expression Profiling of Echinococcus Multilocularis Mirnas Throughout Metacestode Development *In Vitro*. *PLoS Negl. Trop. Dis.* 15, 1–32. doi: 10.1371/journal.pntd.0009297

McNulty, S. N., Tort, J. F., Rinaldi, G., Fischer, K., Rosa, B. A., Smircich, P., et al. (2017). Genomes of Fasciola Hepatica From the Americas Reveal Colonization With Neorickettsia Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers. *PLoS Genet.* 13, e1006537. doi: 10.1371/journal.pgen.1006537

Mead, A. (1992). Review of the Development of Multidimensional Scaling Methods. *Stat.* 41, 27. doi: 10.2307/2348634

Meningher, T., Barsheshet, Y., Ofir-Birin, Y., Gold, D., Brant, B., Dekel, E., et al. (2020). Schistosomal Extracellular Vesicle-Enclosed miRNAs Modulate Host T Helper Cell Differentiation. *EMBO Rep.* 21, 1–17. doi: 10.15252/embr.201947882

Meningher, T., Lerman, G., Regev-Rudzki, N., Gold, D., Ben-Dov, I. Z., Sidi, Y., et al. (2017). Schistosomal microRNAs Isolated From Extracellular Vesicles in Sera of Infected Patients: A New Tool for Diagnosis and Follow-Up of Human Schistosomiasis. *J. Infect. Dis.* 215, 378–386. doi: 10.1093/infdis/jiw539

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., et al. (2008). Circulating microRNAs as Stable Blood-Based Markers for Cancer Detection. *Proc. Natl. Acad. Sci. U. S. A.* 105, 10513–10518. doi: 10.1073/pnas.0804549105

Mosig, A., Sameith, K., and Stadler, P. (2006). Fragrep: An Efficient Search Tool for Fragmented Patterns in Genomic Sequences. *Genom. Proteomics Bioinf.* 4, 56–60. doi: 10.1016/S1672-0229(06)60017-X

Mu, Y., Cai, P., Olveda, R. M., Ross, A. G., Olveda, D. U., and McManus, D. P. (2020). Parasite-Derived Circulating microRNAs as Biomarkers for the Detection of Human Schistosoma Japonicum Infection. *Parasitology* 147, 889–896. doi: 10.1017/S0031182019001690

Mu, Y., McManus, D. P., Gordon, C. A., and Cai, P. (2021). Parasitic Helminth-Derived microRNAs and Extracellular Vesicle Cargos as Biomarkers for Helminthic Infections. *Front. Cell. Infect. Microbiol.* 11, 708952. doi: 10.3389/fcimb.2021.708952

Nowacki, F. C., Swain, M. T., Klychnikov, O. I., Niazi, U., Ivens, A., Quintana, J. F., et al. (2015). Protein and Small Non-Coding RNA-Enriched Extracellular Vesicles Are Released by the Pathogenic Blood Fluke Schistosoma Mansoni. *J. Extracell. Vesicles* 4, 28665. doi: 10.3402/jev.v4.28665

Ovchinnikov, V. Y., Kashina, E., Mordvinov, V. A., and Fromm, B. (2020). EV-Transported microRNAs of Schistosoma Mansoni and Fasciola Hepatica: Potential Targets in Definitive Hosts. *Infect. Genet. Evol.* 85, 104528. doi: 10.1016/j.meegid.2020.104528

Palakodeti, D., Smielewska, M., and Graveley, B. R. (2006). MicroRNAs From the Planarian Schmidtea Mediterranea: A Model System for Stem Cell Biology. *RNA (New York N. Y.)* 12, 1640–1649. doi: 10.1261/rna.117206

Pérez, M. G., Spiliotis, M., Rego, N., Macchiaroli, N., Kamenetzky, L., Holroyd, N., et al. (2019). Deciphering the Role of miR-71 in Echinococcus Multilocularis Early Development *In Vitro*. *PLoS Negl. Trop. Dis.* 13, 1–21. doi: 10.1371/JOURNAL.PNTD.0007932

Peterson, S. M., Thompson, J. a, Ufkin, M. L., Sathyanarayana, P., Liaw, L., and Congdon, C. B. (2014). Common Features of microRNA Target Prediction Tools. *Front. Genet.* 5, 23. doi: 10.3389/fgene.2014.00023

Potapov, V., Fu, X., Dai, N., Corrêa, I. R. J., Tanner, N. A., and Ong, J. L. (2018). Base Modifications Affecting RNA Polymerase and Reverse Transcriptase Fidelity. *Nucleic Acids Res.* 46, 5753–5763. doi: 10.1093/nar/gky341

Protasio, A., van Dongen, S., Collins, J., Quintais, L., Ribeiro, D. M., Sessler, F., et al. (2017). MiR-277/4989 Regulate Transcriptional Landscape During Juvenile to Adult Transition in the Parasitic Helminth Schistosoma Mansoni. *PLoS Negl. Trop. Dis.* 11, e0005559. doi: 10.1371/journal.pntd.0005559

Qin, Y., Yao, J., Wu, D. C., Nottingham, R. M., Mohr, S., Hunicke-Smith, S., et al. (2016). High-Throughput Sequencing of Human Plasma RNA by Using Thermostable Group II Intron Reverse Transcriptases. *RNA (New York N.Y.)* 22, 111–128. doi: 10.1261/rna.054809.115

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033

Quintana, J. F., Kumar, S., Ivens, A., Chow, F. W. N., Hoy, A. M., Fulton, A., et al. (2019). Comparative Analysis of Small RNAs Released by the Filarial Nematode Litomosoides Sigmodontis *In Vitro* and *In Vivo*. *PLoS Negl. Trop. Dis.* 13, 1–29. doi: 10.1371/journal.pntd.0007811

Quintana, J. F., Makepeace, B. L., Babayan, S. A., Ivens, A., Pfarr, K. M., Blaxter, M., et al. (2015). Extracellular Onchocerca-Derived Small RNAs in Host Nodules and Blood. *Parasites Vectors* 8, 58. doi: 10.1186/s13071-015-0656-1

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., et al. (2000). The 21-Nucleotide Let-7 RNA Regulates Developmental Timing in Caenorhabditis Elegans. *Nature* 403, 901–906. doi: 10.1038/35002607

Ricafrente, A., Nguyen, H., Tran, N., and Donnelly, S. (2021). An Evaluation of the Fasciola Hepatica Mirnome Predicts a Targeted Regulation of Mammalian Innate Immune Responses. *Front. Immunol.* 11, 608686. doi: 10.3389/fimmu.2020.608686

Sánchez-López, C. M., Trelis, M., Jara, L., Cantalapiedra, F., Marcilla, A., and Bernal, D. (2020). Diversity of Extracellular Vesicles From Different Developmental Stages of Fasciola Hepatica. *Int. J. Parasitol.* 50, 663–669. doi: 10.1016/j.ijpara.2020.03.011

Sasidharan, V., Lu, Y.-C., Bansal, D., Dasari, P., Poduval, D., Seshasayee, A., et al. (2013). Identification of Neoblast- and Regeneration-Specific miRNAs in the Planarian Schmidtea Mediterranea. *RNA (New York N. Y.)* 19, 1394–1404. doi: 10.1261/rna.038653.113

Shi, J., Zhang, Y., Tan, D., Zhang, X., Yan, M., Zhang, Y., et al. (2021). PANDORA-Seq Expands the Repertoire of Regulatory Small RNAs by Overcoming RNA Modifications. *Nat. Cell Biol.* 23, 424–436. doi: 10.1038/s41556-021-00652-7

Sotillo, J., Robinson, M. W., Kimber, M. J., Cucher, M., Ancarola, M. E., Nejsum, P., et al. (2020). The Protein and microRNA Cargo of Extracellular Vesicles From Parasitic Helminths – Current Status and Research Priorities. *Int. J. Parasitol.* 50, 635–645. doi: 10.1016/j.ijpara.2020.04.010

Stadler, P. F., Chen, J. J. L., Hackermüller, J., Hoffmann, S., Horn, F., Khaitovich, P., et al. (2009). Evolution of Vault RNAs. *Mol. Biol. Evol.* 26, 1975–1991. doi: 10.1093/molbev/msp112

Stroehlein, A. J., Young, N. D., Korhonen, P. K., Hall, R. S., Jex, A. R., Webster, B. L., et al. (2018). The Small RNA Complement of Adult Schistosoma Haematobium. *PLoS Negl. Trop. Dis.* 12, e0006535. doi: 10.1371/journal.pntd.0006535

Sun, J., Wang, S., Li, C., Ren, Y., and 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. *Parasites Vectors* 7, 177. doi: 10.1186/1756-3305-7-177

Tarver, J. E., Sperling, E., Nailor, A., Heimberg, A. M., Robinson, J. M., King, B. L., et al. (2013). miRNAs: Small Genes With Big Potential in Metazoan Phylogenetics. *Mol. Biol. Evol.* 30, 2369–2382. doi: 10.1093/molbev/mst133

Tosar, J. P., and Cayota, A. (2020). Extracellular tRNAs and tRNA-Derived Fragments. *RNA Biol.* 17, 1149–1167. doi: 10.1080/15476286.2020.1729584

Tosar, J. P., Gámbaro, F., Castellano, M., and Cayota, A. (2021). RI-SEC-Seq: Comprehensive Profiling of Nonvesicular Extracellular RNAs With Different Stabilities. *Bio-Protocol* 11, e3918. doi: 10.21769/BioProtoc.3918

Tosar, J. P., Gámbaro, F., Darré, L., Pantano, S., Westhof, E., and Cayota, A. (2018). Dimerization Confers Increased Stability to Nucleases in 5' Halves From Glycine and Glutamic Acid tRNAs. *Nucleic Acids Res.* 46, 9081–9093. doi: 10.1093/nar/gky495

Tran, N., Ricafrente, A., To, J., Lund, M., Marques, T. M., Gama-Carvalho, M., et al. (2021). Fasciola Hepatica Hijacks Host Macrophage miRNA Machinery to Modulate Early Innate Immune Responses. *Sci. Rep.* 11, 1–11. doi: 10.1038/s41598-021-86125-1

Tristán-Ramos, P., Rubio-Roldan, A., Peris, G., Sánchez, L., Amador-Cubero, S., Viollet, S., et al. (2020). The Tumor Suppressor microRNA Let-7 Inhibits Human LINE-1 Retrotransposition. *Nat. Commun.* 11, 5712. doi: 10.1038/s41467-020-19430-4

Tritten, L., Burkman, E., Moorhead, A., Satti, M., Geary, J., Mackenzie, C., et al. (2014a). Detection of Circulating Parasite-Derived MicroRNAs in Filarial Infections. *PLoS Negl. Trop. Dis.* 8, e2971. doi: 10.1371/journal.pntd.0002971

Tritten, L., O'Neill, M., Nutting, C., Wanji, S., Njouendoui, A., Fombad, F., et al. (2014b). Loa Loa and Onchocerca Ochengi miRNAs Detected in Host Circulation. *Mol. Biochem. Parasitol.* 198, 14–17. doi: 10.1016/j.molbiopara.2014.11.001

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J. J., and Lötvall, J. O. (2007). Exosome-Mediated Transfer of mRNAs and microRNAs Is a Novel Mechanism of Genetic Exchange Between Cells. *Nat. Cell Biol.* 9, 654–659. doi: 10.1038/ncb1596

Wang, L., Liu, T., Chen, G., Li, Y., Zhang, S., Mao, L., et al. (2021). Exosomal Microrna Let-7-5p From Taenia Pisiformis Cysticercus Prompted Macrophage to M2 Polarization Through Inhibiting the Expression of C/Ebp δ. *Microorganisms* 9, 1403. doi: 10.3390/microorganisms9071403

Wang, L. Q., Liu, T. L., Liang, P. H., Zhang, S. H., Li, T. S., Li, Y. P., et al. (2020). Characterization of Exosome-Like Vesicles Derived From Taenia Pisiformis Cysticercus and Their Immunoregulatory Role on Macrophages. *Parasites Vectors* 13, 1–16. doi: 10.1186/s13071-020-04186-z

Wheeler, B. M., Heimberg, A. M., Moy, V. N., Sperling, E., Holstein, T. W., Heber, S., et al. (2009). The Deep Evolution of Metazoan microRNAs. *Evol. Dev.* 11, 50–68. doi: 10.1111/j.1525-142X.2008.00302.x

World Health Organization. (2007). *Fact Sheet on Fascioliasis. Action Against Worms*. Available at: http://www.who.int/neglected_diseases/preventive_chemotherapy/Newsletter10.pdf.

Xu, M.-J., Ai, L., Fu, J.-H., Nisbet, A. J., Liu, Q.-Y., Chen, M.-X., et al. (2012). Comparative Characterization of microRNAs From the Liver Flukes Fasciola Gigantica and F. Hepatica. *PLoS One* 7, e53387. doi: 10.1371/journal.pone.0053387

Xu, H., Yao, J., Wu, D. C., and Lambowitz, A. M. (2019). Improved TGIRT-Seq Methods for Comprehensive Transcriptome Profiling With Decreased Adapter Dimer Formation and Bias Correction. *Sci. Rep.* 9, 7953. doi: 10.1038/s41598-019-44457-z

Yu, J., Yu, Y., Li, Q., Chen, M., Shen, H., Zhang, R., et al. (2019). Comprehensive Analysis of miRNA Profiles Reveals the Role of Schistosoma Japonicum miRNAs at Different Developmental Stages. *Vet. Res.* 50, 23. doi: 10.1186/s13567-019-0642-2

Zerna, G., Spithill, T. W., and Beddoe, T. (2021). Current Status for Controlling the Overlooked Caprine Fasciolosis. *Animals* 11, 1819. doi: 10.3390/ani11061819

Zhu, L., Zhao, J., Wang, J., Hu, C., Peng, J., Luo, R., et al. (2016). MicroRNAs Are Involved in the Regulation of Ovary Development in the Pathogenic Blood Fluke Schistosoma Japonicum. *PLoS Pathog.* 12, 1–23. doi: 10.1371/journal.ppat.1005423

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

B. Conservación de las vías de ARN pequeños entre platelmintos de vida libre y parásitos

ARTICULO: Perdidos y encontrados: las vías de Piwi y Argonauta en gusanos planos

Santiago Fontenla, Gabriel Rinaldi, José Tort.

Frontiers Cellular and Infection Microbiology 11 (2021), doi: 10.3389/fcimb.2021.653695

En este trabajo, presentamos los resultados que responden al objetivo (B): Comparar la conservación de las vías de ARNpq entre platelmintos de vida libre y parásitos. Para esto analizamos la conservación de las vías de ARNpq entre varias especies del grupo parafilético tradicionalmente conocido como 'turbelarios' y especies parásitas (organizadas en el clado monofilético Neodermata) para desentrañar posibles adaptaciones durante la transición al parasitismo. Nuestros hallazgos muestran que las vías de miARNs e ARNi están presentes en todas las especies de vida libre analizadas. A su vez, mientras que todos los 'turbelarios' tienen proteínas Piwi, estas se perdieron en todos los Neodermatas. Inclusive, identificamos dos clusters de genes Piwi en 'turbelarios', uno de ellos notoriamente divergente, que denominamos 'Fliwi'. Adicionalmente, otros factores clave de la vía Piwi, como Vasa o Zucchini, también están conservados en todos los 'turbelarios' pero ausentes en Neodermatas.

Además, confirmamos que las Argonautas platelminto-especifica denominada 'FL-Ago' está conservada ancestralmente en los todos platelmintos pero está expandida solo en especies parásitas.

A nivel de la secuencia proteica, hallamos menor conservación entre las FL-Agos respecto a otras subfamilias de las Argonautas, sugiriendo que podrían haber sufrido una mayor presión evolutiva.

frontiers in Cellular and Infection Microbiology

ORIGINAL RESEARCH
published: 27 May 2021
doi: 10.3389/fcimb.2021.653695

Lost and Found: Piwi and Argonaute Pathways in Flatworms

Santiago Fontenla[1]*, *Gabriel Rinaldi*[2] *and Jose F. Tort*[1]*

[1] Departamento de Genética, Facultad de Medicina, Universidad de la República (UDELAR), Montevideo, Uruguay,
[2] Wellcome Sanger Institute, Hinxton, United Kingdom

Platyhelminthes comprise one of the major phyla of invertebrate animals, inhabiting a wide range of ecosystems, and one of the most successful in adapting to parasitic life. Small non-coding RNAs have been implicated in regulating complex developmental transitions in model parasitic species. Notably, parasitic flatworms have lost Piwi RNA pathways but gained a novel Argonaute gene. Herein, we analyzed, contrasted and compared the conservation of small RNA pathways among several free-living species (a paraphyletic group traditionally known as 'turbellarians') and parasitic species (organized in the monophyletic clade Neodermata) to disentangle possible adaptations during the transition to parasitism. Our findings showed that complete miRNA and RNAi pathways are present in all analyzed free-living flatworms. Remarkably, whilst all 'turbellarians' have Piwi proteins, these were lost in parasitic Neodermantans. Moreover, two clusters of Piwi class Argonaute genes are present in all 'turbellarians'. Interestingly, we identified a divergent Piwi class Argonaute in free living flatworms exclusively, which we named 'Fliwi'. In addition, other key proteins of the Piwi pathways were conserved in 'turbellarians', while none of them were detected in Neodermatans. Besides Piwi and the canonical Argonaute proteins, a flatworm-specific class of Argonautes (FL-Ago) was identified in the analyzed species confirming its ancestrallity to all Platyhelminthes. Remarkably, this clade was expanded in parasitic Neodermatans, but not in free-living species. These phyla-specific Argonautes showed lower sequence conservation compared to other Argonaute proteins, suggesting that they might have been subjected to high evolutionary rates. However, key residues involved in the interaction with the small RNA and mRNA cleavage in the canonical Argonautes were more conserved in the FL-Agos than in the Piwi Argonautes. Whether this is related to specialized functions and adaptations to parasitism in Neodermatans remains unclear. In conclusion, differences detected in gene conservation, sequence and structure of the Argonaute family suggest tentative biological and evolutionary diversifications that are unique to Platyhelminthes. The remarkable divergencies in the small RNA pathways between free-living and parasitic flatworms indicate that they may have been involved in the adaptation to parasitism of Neodermatans.

Keywords: Piwi, Ago, Vasa, RNAi pathways, miRNA, piRNA, siRNA, flatworms

OPEN ACCESS

Edited by:
Sandra R. Maruyama, Federal University of São Carlos, Brazil

Reviewed by:
Fabrizio Lombardo, Sapienza University of Rome, Italy
Darren J. Obbard, University of Edinburgh, United Kingdom

***Correspondence:**
Santiago Fontenla
sfontenla@fmed.edu.uy
Jose F. Tort
jtort@fmed.edu.uy

Specialty section:
This article was submitted to Parasite and Host, a section of the journal Frontiers in Cellular and Infection Microbiology

Received: *15 January 2021*
Accepted: *10 May 2021*
Published: *27 May 2021*

Citation:
Fontenla S, Rinaldi G and Tort JF (2021) Lost and Found: Piwi and Argonaute Pathways in Flatworms. Front. Cell. Infect. Microbiol. 11:653695. doi: 10.3389/fcimb.2021.653695

INTRODUCTION

At the beginning of this century an unexpected and complex 'RNA world' started to be unveiled (Guil and Esteller, 2015), resulting in the discovery of novel layers of fine-tuned mechanisms for gene expression regulation, unimaginable until then. Regulatory activities were assigned to a growing range of new species of non-coding RNAs (Guil and Esteller, 2015). Single stranded non-coding RNAs of 20 to 30 nucleotides long are key mediators in small RNA pathways that underlie diverse biological processes. MicroRNAs (miRNAs) are post-transcriptional regulators involved in cell development and differentiation, metabolism, DNA methylation, neurological development, immune response, defense against viral infections and cancer (Huang and Zhang, 2014). Piwi-RNAs (piRNAs) are small non-coding RNAs specifically involved in the maintenance of genome stability by silencing transposable elements (TE) in germline cells (Weick and Miska, 2014). Finally, RNA interference (RNAi) is a pathway mediated by short-interfering RNAs (siRNAs) that might have originally emerged as response to double-strand RNA (dsRNA) generated during some virus infections (Ding and Voinnet, 2007). The presence of dsRNA molecules in the cell cytoplasm triggers a post-transcriptional degradation of complementary mRNA molecules. Consequently, this pathway has been exploited as a reverse genetic tool to silence specific genes (Han, 2018). Currently, more than two decades after it was first applied to a flatworm species (Sánchez Alvarado and Newmark, 1999), it is still the main tool to study gene function in worms (Mourão et al., 2012; Wang et al., 2020).

The regulatory pathways mediated by small RNAs have been extensively studied in the nematode *Caenorhabditis elegans*, first model species in which post-transcriptional gene-silencing mediated by dsRNA was described (Fire et al., 1998). In addition, *C. elegans* was the first organism from which a miRNA was isolated: lin-4, (Lee, 1993). On the other hand, *Schmidtea mediterranea* was the first free-living flatworm species to be silenced by RNAi (Sánchez Alvarado and Newmark, 1999). Planarians have long been models for tissue regeneration and stem cells homeostasis, and the emergence of RNAi as functional genomic tool has transformed the field (Reddien and Alvarado, 2004; Blythe et al., 2010; Sandmann et al., 2011).

Planarians are free living flatworms of the order Tricladida, phylum Platyhelminthes. Platyhelminthes are one of the major phyla of invertebrate animals, traditionally divided into four classes: the free living 'turbellarians', the ectoparasitic Monogenea, and the endoparasitic Trematoda (flukes) and Cestoda (tapeworms). All the parasitic classes are grouped in Neodermata given they all share the presence of a syncytial unciliated epidermis (the neodermis) that seem to be crucial for host immune system evasion and nutrient absorption (Caira and Littlewood, 2013). Studies based on rRNA (Larsson and Jondelius, 2008; Laumer and Giribet, 2014) and transcriptomic data (Egger et al., 2015; Laumer et al., 2015) showed that the 'turbellarians' constitute a paraphyletic group, splitting now the phylum Platyhelminthes into two clades; the ancestral Catenulida and the Rhabditophora, that contains several free-living orders and the parasitic neodermatans. More recently the Macrostomorpha was placed as the earliest diverging Rhabditophoran linage and the Tricladida as part of the later evolved 'turbellarians' (Egger et al., 2015; Laumer et al., 2015) (**Figure S1**).

The success of RNAi in planarians encouraged its use in parasitic species where genetic tools were desperately needed. RNAi has proven to be functional in other free-living and parasitic species (Orii et al., 2003; Rinaldi et al., 2008; Kuales et al., 2011; Dell'Oca et al., 2014; Moguel et al., 2015) and miRNAs have been detected in almost all flatworm lineages (Palakodeti et al., 2006; Cucher et al., 2011; Fromm et al., 2013; Fontenla et al., 2015; Cai et al., 2016; Protasio et al., 2017). Whilst piRNAs were early found in free living planarians (Palakodeti et al., 2008; Friedländer et al., 2009), remarkably, they have not been identified in parasitic species.

Our previous analysis of the small RNA pathways in parasitic flatworm genomes strongly indicated relevant gene losses within the neodermatans that may have been associated with the adaptation to parasitism (Fontenla et al., 2017). However, these observations were limited by the paucity of data from free-living species, represented only by *Macrostomum lignano* and the planaria *S. mediterranea*.

The availability of transcriptomic data from several early diverging free-living species (Laumer et al., 2015) allowed us now to expand our analysis and provide a complete picture of the phylum Platyhelminthes, including a representative set of 'turbellarian' species[1], adding also novel monogenean (Ilgová et al., 2017) and trematode genomes (Oey et al., 2018; Choi et al., 2020; Rosa et al., 2020). The emerging picture provides evidence of substantial differences in the distribution of small RNA pathways proteins suggestive of diverse regulatory possibilities in both free living and parasitic flatworms. Additionally, these findings shine a light into tentative relations between the divergency of small RNA pathways and mechanisms driving parasitism in organisms that are responsible for an enormous disease burden in both human and animals.

METHODS

Data Acquisition

Small RNA pathways proteins of *Macrostomum lignano*, *Schmidtea mediterranea*, *Gyrodactylus salaris*, *Schistosoma mansoni* and *Echinococcus multilocularis* together with other Neodermata species were characterized as described (Fontenla et al., 2017). The recently published genomes of the trematodes *Fasciola gigantica* and *Fasciolopsis buski* (Choi et al., 2020) and four species of the genus *Paragonimus* (Oey et al., 2018; Rosa et al., 2020) were also included. Transcriptomic data on several early diverging flatworms (Laumer et al., 2015) were obtained from public repository Data Dryad (doi: 10.5061/dryad.622q4). Transcriptomic data on *Eudiplozoon nipponicum* was

[1] Here and across the manuscript we used the old term 'turbellarian' as a proxy to refer to all the non-Neodermatan species, including the Catenulida and the early diverging Rhabdithopora taxons.

downloaded from GitHub repository (Ilgová et al., 2017). To study the quality of the transcriptomic data, BUSCO v4.1.4 (Seppey et al., 2019) was used with option –l metazoa to search for conserved Metazoan genes. Considering the levels of missing and fragmented transcripts we selected the seven best 'turbellarian' assemblies, comprising a reasonable overview of the 'turbellarian' clade diversity. The species analyzed were *Stenostomum leucops*, *Prostheceraeus vittatus*, *Geocentrophora applanata*, *Rhynchomesostoma rostratum*, *Monocelis fusca*, *Kronborgia cf. amphipodicola* and *Bothrioplana semperi*. TransDecoder.LongOrfs function of TransDecoder v4.1.0 software (available at https://github.com/TransDecoder/TransDecoder/) was used to predict longest open reading frames (ORFs) on transcripts. Detailed information and source of other sequences used in the construction of the gene trees can be found in **Table S10**.

Identification of Small RNA Pathways Proteins

Flatworm small RNA pathways proteins previously identified by us (Fontenla et al., 2017), and *C. elegans* factors that we failed to detect in our previous search were used as query to interrogate with BLASTp the translated transcriptomes and genomes. We also inspected the presence of *D. melanogaster*'s Zuc and Vasa (Fontenla et al., 2017) using the same approach against all the species. Matched sequences were acquired and used to perform reverse BLASTp against the proteomes of *S. mansoni*, *M. lignano*, *C. elegans* and *D. melanogaster* retaining only the best hit. HMMScan (Johnson et al., 2010) was used to predict the functional domains in putative small RNA pathways proteins, and sequences with no functional domains were discarded from the analysis. HMMScan prediction was performed in the complete transcriptomes of *S. mediterranea* and *S. mansoni* as quality control, to confirm that distant homologous genes with the conserved function were not discarded in the BLAST search. The procedure did not show a different outcome to the blast results Detected sequences are available as **Supplementary Material** (**Folder S1**).

Construction of Phylogenetic Trees

To avoid overestimating the number of genes in transcriptomic data or report entire genome duplications as gene expansions, CD-HIT (Huang et al., 2010) was used to cluster sequences with more than 90% similarity (option –c 0.9). MAFFT (Katoh and Standley, 2013) with local alignment option and structural information was used to align the selected sequences. Due to the fragmented nature of transcriptomic data, sequence alignments were manually curated using BioEdit (Hall, 1999), removing sequences that were too short (< 140 aa). Maximum Likelihood trees with statistical branch support (SH-like) were generated with PhyML (Guindon et al., 2010), with models inferred with Smart Model Selection (SMS) (Lefort et al., 2017). Trees were visualized with Evolview (He et al., 2016), and enriched by adding domain structure information. Argonaute unrooted tree (**Figure 1**) was visualized with MEGA version X (Kumar et al., 2018).

Molecular Evolutionary Analysis of Argonaute Subfamilies

For the heatmap showing the sequence conservation of the Argonaute subfamilies, the consensus function of R package seqinR (Charif and Lobry, 2007) was used to build matrices with the residues count at each position of the alignments. Next, we applied a color scale to the most abundant residue at each

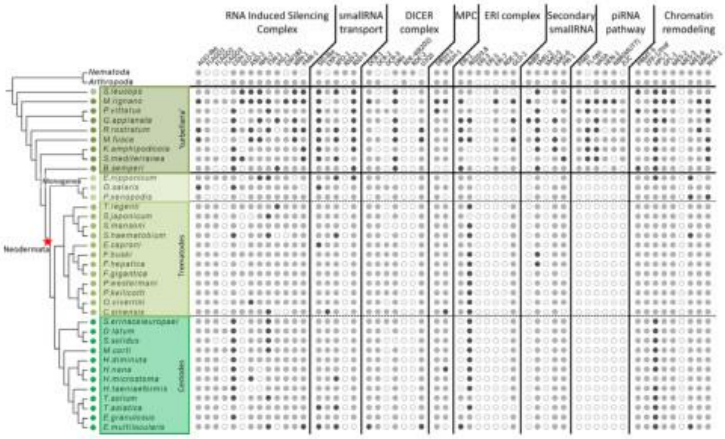


FIGURE 1 | Distribution of miRNA, piRNA and RNAi proteins across flatworms, arthropods and nematodes. White circles indicate absence, grey circles indicate one homologue detected, black circles indicate two or more homologues detected. In arthropods and nematodes only presence/absence is indicated.

position. To calculate the percentage of conserved positions by domain, the positions with conserved residues in more than 50% of the sequences were added and corrected by the domain length.

For positive selection inference, Argonaute transcripts were codon aligned in MEGA version X (Kumar et al., 2018) with Muscle aligner. The alignment was edited with BioEdit (Hall, 1999) and MEGA version X Codon-based Z-test of Selection tool used to compute synonymous and nonsynonymous substitutions: the hypothesis tested was positive selection (dN>dS) in sequence pairs with the Nei-Gojobori method (Nei and Gojobori, 1986), with a *p-value* threshold ≤ 0.05.

Additionally, we tested for evidence of positive selected sites (PSSs) using the mixed effects model of evolution (MEME) (Murrell et al., 2012) method. MEME applies a branch-site random effects phylogenetic framework that allows the distribution of dN/dS to vary from site to site as well as from branch to branch, thereby identifying residues that have undergone episodic selection (i.e. positive selection that varies temporally throughout the tree). Only likelihood ratio test (LRT) with *p-value* ≤ 0.05 were considered as statistically significant evidence of PSSs.

RESULTS

While miRNA and siRNA Pathways Are Conserved Across Platyhelminths, the Complete Piwi Pathway Is Lost in Parasitic Flatworms

The presence of small RNA pathways was investigated on available transcriptomes from 25 early diverging flatworms (Laumer et al., 2015). We first evaluated the quality of the assemblies against a set of conserved metazoans genes using BUSCO v 4.1.4. Based on the level of fragmentation and number of missing orthologues, we selected assemblies from seven species, that added to *M. lignano*, and *S. mediterranea* capture the diversity of free-living flatworms. In addition, we included a novel dataset from underrepresented monogeneans and six novel trematode genomes (Oey et al., 2018; Choi et al., 2020; Rosa et al., 2020). Taken together this dataset provides a comprehensive phylogenetic view of the platyhelminth clade diversity (**Figure S1**).

The homology search with a curated set of proteins involved in small RNA pathways not only showed the presence of most of them in all free-living species, but also outstanding absences in neodermatans. While several proteins show differential distribution among diverse classes (**Figures S2–S6** and **Tables S1–S6**), a remarkable feature is the complete absence of all the piwi pathway components in all parasitic species (**Figure 1**).

The absence of piwi proteins in parasitic trematodes and cestodes have been previously proposed (Skinner et al., 2014; Fontenla et al., 2017), raising questions on how the parasitic species control the activity of repetitive mobile elements. Our extensive search of other piwi pathway genes across flatworms clearly shows that the complete pathway is missing in neodermatans while is conserved in free-living species.

Amplifications in the Argonaute Family Show Differential Distributions Across Flatworms

Argonautes, small-RNA binding proteins, are central components of all the small RNA pathways (Niaz, 2018). Phylogenetic studies have traditionally classified the family into the Ago class (that is further subdivided in two subclasses comprising miRNA Agos and the siRNA associated proteins), the Piwi class, and the Wago clade, this latter, comprising nematode specific argonaute proteins (Wynant et al., 2017).

The comparison of Argonaute superfamily proteins from diverse metazoans, now including our extended sampling of flatworms, reveals interesting differences in their distribution, particularly between free-living and parasitic species.

All flatworms have putative orthologues to miRNA class proteins that constitute a well-defined clade.

Rather than grouping with siRNA class Agos from basal metazoans (poriferans and cnidarians), ecdysozoans (nematodes and arthropods) or other lophotrochozoa (mollusks, annelids, gastrotrichs and rotifers), all the other flatworm Ago sequences cluster together in a well-defined clade (**Figure 2**). This flatworm specific clade has been previously reported by us and others, and termed FLAgos (Zheng, 2012; Skinner et al., 2014; Fontenla et al., 2017).

Interestingly, while a single gene is found in all the free-living species (with the only exception of *S. mediterranea* that showed a gene duplication), the parasitic cestodes and trematodes have experienced gene amplifications leading to two or more genes (**Figures 2** and **3**). The two FLAgo genes from the model trematode *S. mansoni* (Smp_179320 and Smp_102690) are organized in tandem in chromosome 1, but show clear differential expression among developmental stages (**Figure S7**). While a similar tandem gene arrangement could be detected in *S. japonicum* genome, it is not possible to assess if this is a general trend in trematodes due to the still fragmentary nature of the assemblies for other species. Similarly, within cestodes three FLAgo genes are in tandem in *Echinococcus multilocularis* and a more complex array of amplified genes is evident in *Hymenolepis microstoma*. With the information available so far these amplifications appear as independent events in cestodes and trematodes (**Figure 3**). Since single genes are recovered in all free-living species, a more parsimonious hypothesis would be an initial duplication at the origin of neodermatans. However, more detailed analyses and better genome assemblies are needed, in particular considering that the gene trees suggest these are rapidly evolving Agos.

Two subgroups of piwi class genes have been previously described (Wynant et al., 2017; Jehn et al., 2018). Consistently within the free-living flatworms Piwis two sub-groups are evident. But while single genes from all free-living species cluster within the Piwi2 clade, a second separate clade is formed with all the remaining Piwi genes from free-living Rhabditophorans, here termed FLiwi (**Figures 2** and **4**). The complete absence of any piwi homologue in neodermatans is quite evident in monogeneans, trematodes and cestodes, and

this is confirmed based on the extensive sampling on available genomes.

The FLiwi cluster showed several independent gene duplications that have probably occurred after speciation, although we cannot rule out that gene duplications might be overestimated in those species in those species for which only transcriptomic data is available. We cannot rule out that transcripts from different genes were counted as one if the sequences were too similar due to very recent gene duplication events, as was the case for *S. mediterranea*'s FLiwis (Kim et al., 2020). Interestingly, *S. mediterranea* displayed the widest expansion of FLiwis (**Figure 4** dark green arch). Here, eight genes clustered in this group, some with identical sequences that can be collapsed to five quite similar genes (SmeT032534, SmeT030029, SmeT030034 and SmeT030033 present >99% identity). Whether these are functional genes or represent transcribed pseudogenes remains unknown (Kim et al., 2020).

Interestingly, single Ago (Sle67413) and Piwi (Sle58342) genes can be detected in the catenulid *S. leucops*, that generally is placed outside of the clades from other flatworms. It is not possible to assess if the absence of further genes is real or is due to partial sampling of the available transcriptome. In any case, it is quite interesting the placing outside of other flatworms since the inclusion of the group within platyhelminthes was in debate until recently (Larsson and Jondelius, 2008; Laumer and Giribet, 2014).

The phylogenetic tree shown in **Figure 2** suggests that FLAgo and FLiwi have probably experimented highly evolutionary rates, as indicated by their long branches. This may also explain the observed branching pattern inconsistent with the

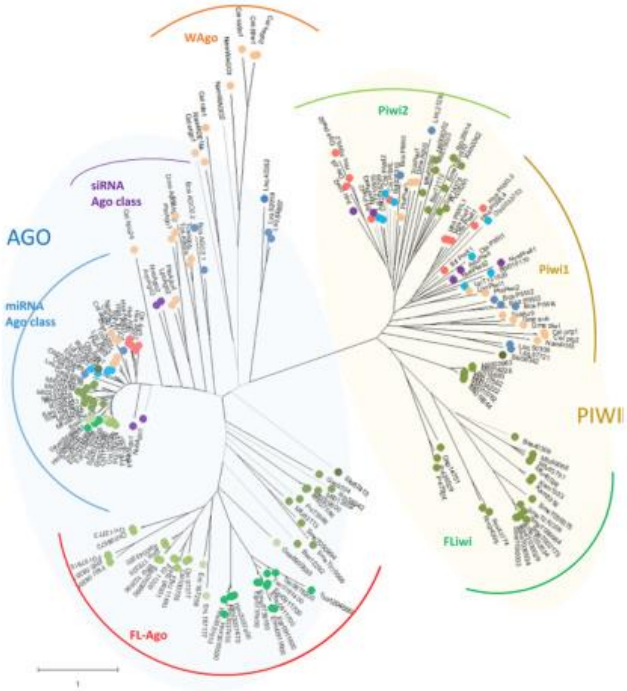


FIGURE 2 | Unrooted maximum likelihood phylogenetic tree of Argonaute proteins belonging to Deuterostomia (•: Hsa, *Homo sapiens*; Gga, *Gallus gallus*; Dre, *Danio rerio*; Bfl, *Branchiostoma floridae*; Cin, *Ciona intestinalis*), Ecdysozoa (•: Cel, *Caenorhabditis elegans*; Nam, *Necator americanus*; Dme, *Drosophila melanogaster*; Tca, *Tribolium casteneum*; Pte, *Parasteatoda tepidariorum*; Lpo, *Limulus polyphemus*), Mollusca & Annelida (•: Obi, *Octopus bimaculoides*; Bgl, *Biomphalaria glabrata*; Lgi, *Lottia gigantea*; Cte, *Capitella teleta*), Rotifera & Gastrotricha (•: Bca, *Brachionus calyciflorus*; Lsq, *Lepidermella squamata*) 'turbellaria' Catenulida (•: Sle, *Stenostomum leucops*), 'turbellaria' Rhabditophora (•: Mli, *Macrostomum lignano*; Pvi, *Prosthecereaus vittatus*; Gap, *Geocentrophora applanate*; Rro, *Rhynchomesostoma rostratum*; Mfu, *Monocelis fusca*; Kam, *Kronborgia amphipodicola*; Bse, *Bothrioplana semperi*), Monogenea (•: Eni, *Eudiplozoon nipponicum*; Gsa, *Gyrodactilus salaris*), Trematode (•: FBU, *Fasciolopsis buski*; FHE, *Fasciola hepatica*; FGI, *Fasciola gigantica*; Ovi, *Opisthorchis viverrini*; *Clonorchis sinensis*; Smp, *Schistosoma mansoni*; Sjp, *Schistosoma japonicum*), Cestode (•: Hmi, *Hymenolepis microstoma*; Tso, *Taenia solium*; Emu, *Echinococcus multiloculars*; Egr, *Echinococcus granulosus*), Cnidaria & Porifera (•: Nve, *Nematstella vectensis*; Aqu, *Amphimedon queenslandica*). Black branches indicate SH-like approximate likelihood ratios ≥ 90. Arches indicates the subgroups of Argonautes (blue shaded) and Piwi (green shaded) genes.

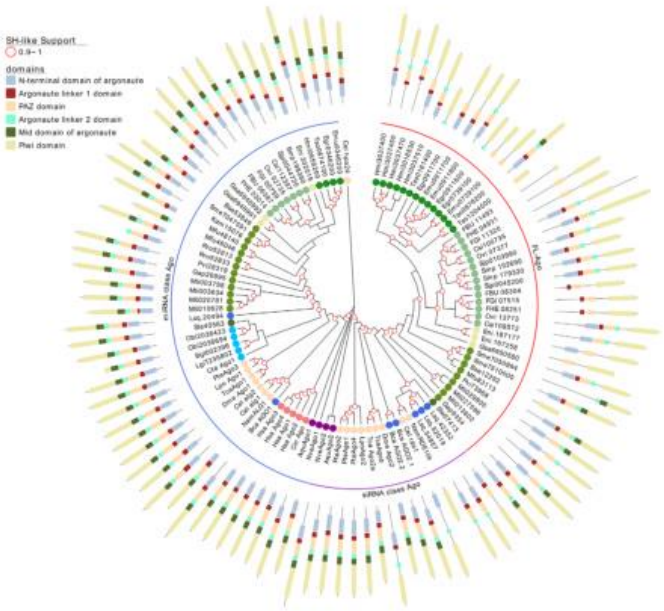


FIGURE 3 | A maximum likelihood tree of canonical Argonaute proteins and FLAgos. SH-like approximate likelihood ratios are indicated. Conserved protein domains were predicted with HMMScan. Blue, purple and red arches correspond to miRNA Ago-class, siRNA Ago-class and FL-Agos, respectively. Abbreviations are as indicated in **Figure 2**.

species tree (**Figure 2**). In this sense, a parsimonious explanation would be that FLAgo might represent rapidly evolving siRNA class genes, while Fliwi could correspond to fast evolving piwi1 type proteins. Further evidence is needed to functionally validate these proteins.; therefore, we decided to investigate other aspects of these intriguing genes.

FLAgos and FLiwis Structures Are More Variable Than Canonical Ago and Piwi Counterparts

Argonaute proteins consist of five distinct domains: the N-terminal, PAZ, Mid, PIWI and two linker regions, L1 and L2. When analyzing the domain conservation among the Argonaute proteins detected across flatworms, it was obvious that while canonical miAgo class genes are highly conserved in their structure, FL-Agos have a more variable structure with MID domain being poorly detected. Both Piwi subfamilies display PAZ and PIWI domains, and while the N-terminal and linker 1 domain are generally identified in the free-living Piwi 2 genes, they are devoid of the linker 2 domain (**Figures 3** and **4**). Conversely, although more structurally variable probably due to independent gene amplifications, most of the FLiwi proteins have linker 2 domain in addition to PAZ/PIWI domains, while the detection of the other domains is more scattered.

Next, we studied the conservation at sequence level of the flatworm Argonaute genes. The canonical miAgo family showed the highest overall conservation which reached up to 85% of conserved positions in the MID domain (**Figure 5A** and **Table S7** for complete list of conservation by domain). Piwi, FLiwi and FLAgos showed lower general sequence conservation. The most conserved domain across clades at amino acid level is the PIWI domain (**Figure 5A**). Remarkably, the less conserved domain in the canonical Ago (i.e. the N-terminal domain) showed more conservation than any of the domains in both Piwi clades and FL-Agos (**Figure 5A** and **Table S7**). This is suggestive of faster evolution rates of the FLAgo and FLiwi clades, consistent with previous reports (Wynant et al., 2017).

Since the differences in sequences conservation may be associated with selection, we measured the variation in the rate of non-synonymous/synonymous (dN/dS) substitutions of Argonaute family proteins. We found that within each subgroup of Argonautes, the FLAgos showed the highest number of statistically significant dN/dS substitutions when free-living to parasitic pairs of genes were considered (**Table S8**). The second group with dN>dS was the FLiwis followed by the miRNA class Agos and the Piwis. When considering variation between subfamilies the highest rate of dN/dS substitutions were observed between the Agos *vs* FLAgos, Ago/

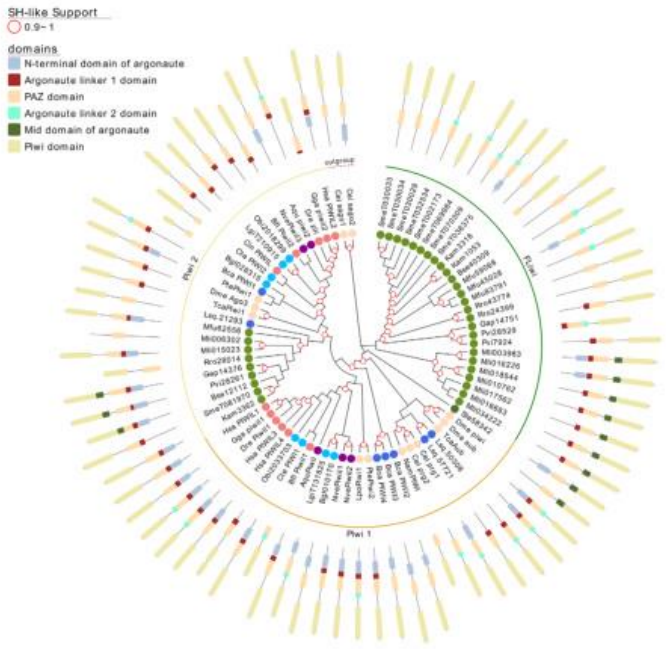


FIGURE 4 | A maximum likelihood tree of Piwi proteins. Wagos were used as outgroup. SH-like approximate likelihood ratios are indicated. Conserved protein domains were predicted with HMMScan. Light green, yellow and green arches corresponds to the cluster of the Piwi 2, Piwi 1 and FLiwis, respectively. Abbreviations are as indicated in **Figure 2**.

Piwis, followed by Piwi/FLAgos and Ago/FLiwi. Using MEME, we aimed to detect sites that were subjected to positive or diversifying selection (see *Methods*). FLAgos showed the highest count of positions under positive diversifying selection, with a significant likelihood ratio test (LRT) of non-synonymous to synonymous substitutions (p-value ≤ 0.05) (dots in **Figure 5A**). Meanwhile, the Piwis showed 3 putative sites under positive selection, and the FLiwis and canonical Agos groups only had one position each (**Figure 5A**).

Small RNA Binding and Catalytic Amino Acids Conservation in FLAgos and FLiwis

Structural studies have shined a light on the mechanisms of action of Argonautes. Human Ago2 bonded to miRNA has a bi-lobed architecture, with the guide miRNA threaded through a central cleft formed by the N-PAZ and MID-PIWI domains. Multitude interactions within this tight binding pocket were identified, involving mainly residues of the MID domain capped on the 5' side by PIWI domain residues (Elkayam et al., 2012; Kong et al., 2017). There, the target mRNA has access to mate with the miRNA and is cleaved by a RNAase H fold comprised by the Asp-Asp-His (DDH) triad in the PIWI domain (Tolia and Joshua-Tor, 2007). Consequently, we sought conservation of these active sites and functional residues in the different Ago classes (indicated by black and grey bars at **Figure 5A**).

Fourteen out of 51 functionally relevant positions were generally conserved in all the Argonaute proteins. Further fifteen positions were conserved between Ago-like and FLAgos, and a minor group of residues were shared between Piwi and FLiwi (**Figure S8**). These include four of the eight residues involved in positioning the guide RNA with respect to the active site to ensure that cleavage of targets occurs at a well-defined and predictable position. Three other positions are conserved only between canonical Agos and FLAgos (**Figure S8**). On the other hand, both Piwis and FLiwis present well conserved substitutions at positions 588 [K→ polar (Q)] and 845 [R→ hydrophobic (L, M or F)] (**Figure S8**), that could also be relevant in the binding to the small RNA.

The QSKN motif (positions 566 to 569) of the MID domain (blue bar in **Figures 5A** and **S8**) was reported to be involved in the binding to the seed region of the miRNA in the *Echinococcus canadensis* canonical Ago genes (Maldonado et al., 2017). Interestingly, while this QSKN motif is conserved in all Neodermatans, (**Figure S8**), the second position was occupied by a non-polar Alanine residue (QAKN), in 'turbellarians' as in *D. melanogaster* and *C. elegans*. In Deuterostomes, the same

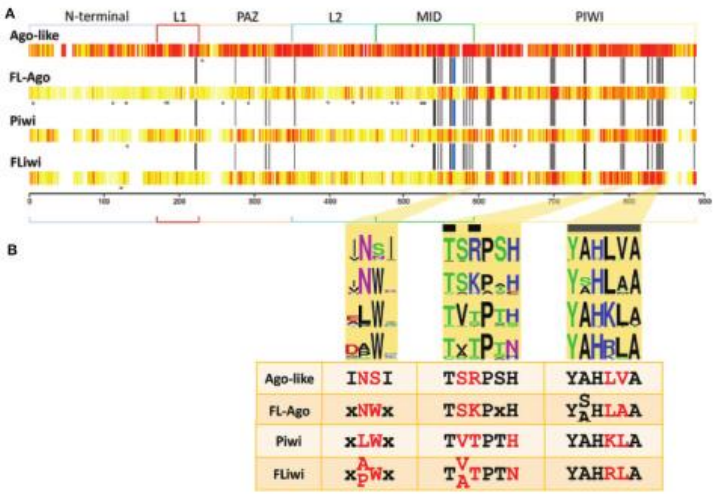


FIGURE 5 | Sequence conservation among Argonaute subfamilies. **(A)** Conservation is shown in a white (less conserved) to red (most conserved) color scale. N-terminal, linker 1 (L1), PAZ, linker 2 (L2), MID and PIWI domains are indicated. Black bars indicate the positions that have been reported to interact with the miRNA (Elkayam et al., 2012). Grey bars indicate the positions of the catalytic DDH triad. Blue bar indicate the position that has been reported to interact with the seed region of miRNAs (Maldonado et al., 2017). Dots indicate positions with statistically significant LRT (p-value ≤ 0.05) detected with MEME tool. **(B)** Highlight of three regions of the MID and PIWI domains that produce a signature for each Ago subfamily. Black bars indicate the positions that have been reported to interact with the miRNA. Grey bars indicate the positions of the catalytic DDH triad.

position was occupied by the non-polar aliphatic residue Valine (QVKN) (or Methionine in Has_Ago2, QMKN). Although the functional implications of this substitution are not clear, the restricted conservation of this motif in the Ago-like genes suggests that it is relevant in this subfamily, and importantly, may be useful as a marker of linage in the future.

The RNase H activity associated with a conserved DDH triad (grey bars in **Figures 5A** and **S8**) is well conserved in all the miRNA class Ago genes of Platyhelminthes. In the case of the Piwi class subfamily only SmeT002173 possess substitutions in the catalytic triad. The absence of the DDH triad would not imply that the catalytic activity was lost, as Dme_piwi has been shown to possess "slicer" activity even though it contains a DDK active site (Tolia and Joshua-Tor, 2007). Furthermore, it was shown for Hsa Ago3 that irrespective of RNase triad conservation, the catalytic activity changes depending on the guide RNA that is loaded (Park et al., 2017). In any event, additional *in silico* approaches as structural homology modeling, ligand docking and molecular dynamics, as well as, experimental evidence involving site-specific mutagenesis and cleavage assays are needed to define the ligand-protein interaction and characterize the catalytic activity of the diverse flatworm Argonautes.

Based on the residue conservation we selected three short motifs with different conserved residues among the diverse flatworm Argonaute subfamilies that can be useful to differentiate them, classifying and assigning novel members (**Figure 5B**). The first motif consists of a duo located in the carboxi-terminal end of the MID domain and is an Asn (N) preferentially followed by Ser (S) in canonical Agos and Asn-Trp (NW) in FLAgos. The non-polar Asn is substituted by Leu (L) and Ala (A) or Pro (P) in Piwi and FLiwis, respectively. The second and third motifs are within the PIWI domain. Motif 2 is TSRPSH in miRNA class Agos, while is TSKPxH in FLAgos, TVTPTH in Piwi and TVTPTN or TATPTN in FLiwis. The third motif contains the Histidine residue of the DDH triad and corresponds to a sextet that has the sequence YALHVA in Ago-like, YSLHAA or YALHAA in FLAgos, YAHKLA in Piwis or YAHRLA in FLiwis. We suggest that the analysis of these motifs may be useful to classify Argonaute proteins in flatworms, and also might provide a means of rapidly identify members in other metazoan species.

Trematodes Display a Shorter Dicer-2 Gene

Ribonuclease III family proteins represent central player in the small RNA pathways. Dicers belong to the ribonuclease III family with the ability to process dsRNA. Dicer (Dcr) is responsible for recognizing a hairpin (in pre-miRNA) or long dsRNA and processing them into 22-23 nt miRNA-miRNA* or siRNA duplexes (Jaronczyk et al., 2005). These small RNA duplexes are bound and processed by Ago proteins to form the RNA interference silencing complex (RISC). Like arthropods, flatworms have two Dcr genes, Dcr-1 and Dcr-2 (including a

putative *S. leucops* Dcr-2 placed in the root of Dcr-1 group) with a paralogue of Dcr-2 in some species we named Dcr-3 (**Figure 6A**) (Gao et al., 2014; Fontenla et al., 2017).

In Neodermatans, Dcr-2 is variable in number and organization (Fontenla et al., 2017). Strikingly, trematodes in contrast to cestodes have a shorter version of Dcr-2 with only the RNAse III domains and, in some cases, a PAZ domain (**Figure 6A**). Additionally, a duplication of Dcr-2 is evident in foodborne trematodes (FBT) compared to blood flukes (BF) (**Figure 6A** green arch). In FBT Dcr-2 and Dcr-3 genes are organized as inverted tandem repeats with the exception of *F. buski* and *Paragonimus spp*, where the fragmented nature of the assembled genomes does not allow to confirm or discard this gene arrangement.

The comparison of Dcr-2 genes from the cestode *E. multilocularis*, the planaria *S. mediterranea* and the trematode *S. mansoni* showed that the second exon of Sma033600 matches with exon 15 of the cestode or planarian counterparts. In contrast, exon 1 of cestode Dcr-2 matches with exon 1 in the planarian gene (**Figure 6B**). This observation suggests that the shortening of Dcr-2 is the result of a genomic reorganization that occurred ancestrally in trematodes, probably by an unequal crossover between different chromatids or an intra-chromatid recombination that resulted in the deletion of about half of the ancestral gene. Consequently, this may have led to the absence of the helicase and dsRNA binding domains. It is tempting to speculate that these proteins might only recognize ssRNA as substrates, but experimental evidence is missing.

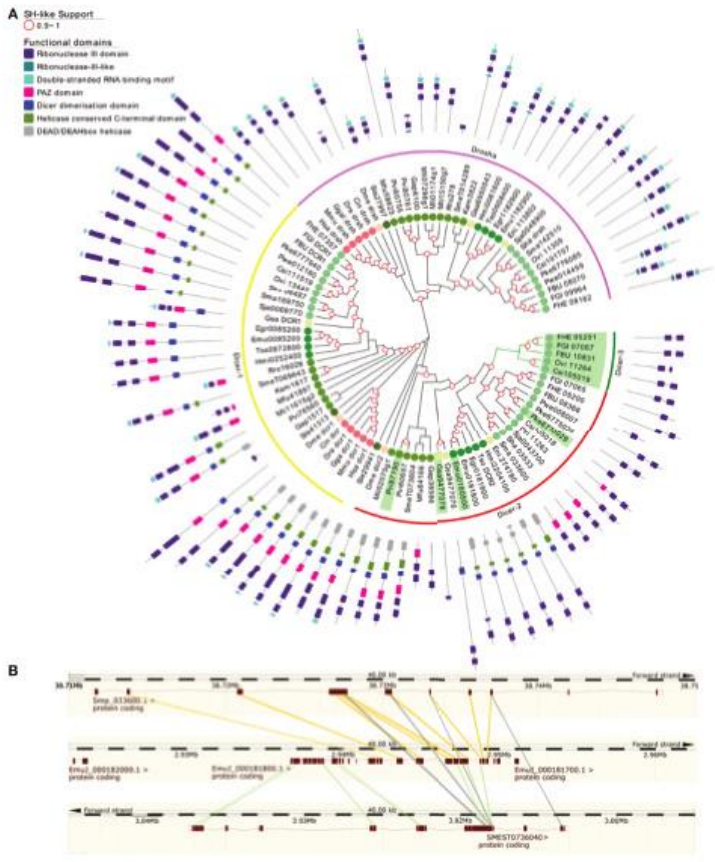


FIGURE 6 | Ribonuclease III subfamilies of flatworms **(A)** A maximum likelihood tree of Ribonuclease III subfamilies. SH-like approximate likelihood ratios are indicated. Conserved protein domains were predicted with HMMScan. Dcr -2 (red arch) duplications are highlighted in green. Dcr-2 is ancestrally duplicated in FBT (dicer-3, green arch). Dcr-2 has missing domains in trematodes compared to cestodes, similar structural heterogeneity was observed between 'turbellarians'. Abbreviations are as indicated in **Figure 2**. **(B)** High scoring pairs between Dcr-2 genes of *S. mansoni*, *E. multilocularis* and *S. mediterranea*.

Although some variation in the structure of 'turbellarians' Dicers was also detected, we cannot rule out these may be artefactual due to the fragmented nature of these transcriptomes.

Belle/PL10 Is Duplicated in Flatworms While Vasa Is Lost in Neodermatan Parasites

Since several other piRNA pathway genes, besides piwi itself, seem to be absent in neodermatans, we investigated in more detail other relevant members involved in the pathway. Vasa is a germline specific DEAD box RNA helicase and plays an essential role in regulating germ cell differentiation (Abdelhaleem, 2005). Taking advantage of our extended set of transcriptomes and genomes of free-living and parasitic flatworms, we analyzed the conservation of Vasa and its paralogous gene, Belle. We founded that while 'turbellarians' conserved homologue genes to Vasa and Belle (green and blue arches in **Figure 7**), Neodermatans have lost vasa orthologues while maintaining Belle/PL10 homologs (blue branches in **Figure 7**). The only exception is Ngvlg3 a Vasa orthologue detected in the monogenean *Neobenedenia girellae*. Interestingly, this gene was not found to be expressed in any tissue and its knockdown by RNAi produced no phenotypic effect on the worm (Ohashi et al., 2007). Thus, we speculate that given the basal position of *N. girellae* within the Neodermatans is possible that Ngvlg3 may be a non-functional pseudogene of the free-living ancestors.

Notably, Belle-like gene is duplicated in all Neodermatans and in *B. semperi*, the closest free-living ancestor of the Neodermatans (blue arch in **Figure 7**). We also found a third group of helicases that we classified as Belle related (light blue arch in **Figure 7**) given that they show a higher percentage of identity with Belle and/or laf-1 compared to other of the *D. melanogaster*/*C. elegans* helicases (**Table S9**). All genes considered within this family have a similar domain structure, with conserved DEAD/DEAH-box and C-terminal helicase domains (**Figure 7**). An amino terminal repeat of Zinc knuckle

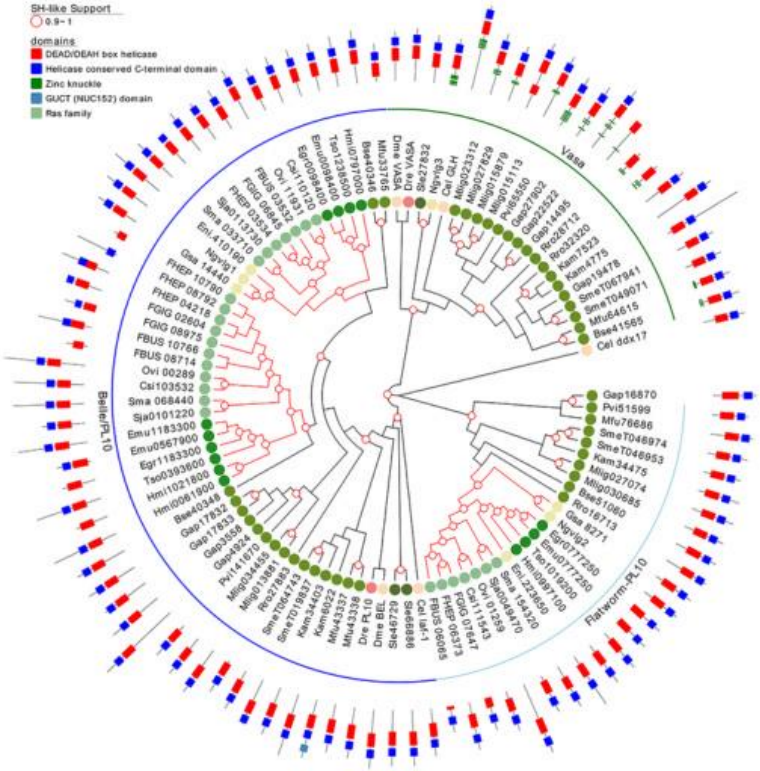


FIGURE 7 | A maximum likelihood tree of Vasa and Vasa-like proteins. SH-like approximate likelihood ratios are indicated. Conserved protein domains were predicted with HMMScan. Green and blue arches correspond to Vasa and Belle/PL10, respectively. Light blue arch corresponds to flatworm PL10. Red branches correspond to Neodermatan species. Species abbreviations are indicated as in **Figure 2**.

was detected in several Vasa genes and in the *C. elegans* ortholog (GLH gene). Remarkably, Vasa homologous genes were amplified in several 'turbellaria' including *M. lignano* where 4 genes were detected after clustering highly similar sequences.

Other PiwiRNA Pathway Proteins Are Conserved Only in Free Living Flatworms

An RNA dependent RNA polymerase (RdRP) amplifies the signal leading to the generation of secondary siRNAs in *C. elegans*. Two 'turbellarian' species (*G. applanata* and *M. fusca*) had sequences with RdRPs functional domains (**Figures 8B**, **S5** and **Table S6**) in addition to *M. lignano*, (Fontenla et al., 2017). RdRPs were not detected in *S. leucops* or *P. vittatus* the two other most ancestral species of Platyhelminthes analyzed here. Interestingly, we found RdRPs in the phylum Gastrotricha, sister phylum to all Platyhelminthes (Egger et al., 2015; Laumer et al., 2015) (**Figure S5B**), suggesting that RdRPs were conserved in the common ancestor to both phyla and were lost during the evolution of Platyhelminthes. While, it is clear that the piRNA pathway does not depend on RdRPs in *S. mediterreanea* and probably other species where RdRPs are not present, it might be possible to occur in species where RdRPs are conserved, as in *C. elegans* (**Figure 8B**).

We also found homologues to other piwi pathway genes like HEN1, Zuc or Mut7 in almost all 'turbellarian' species although they are completely absent in neodermatans. (**Figures 1**, **8A** and **Table S6**). MUT-7 is a protein that contains an exonuclease domain that, in complex with RDE-2, is required in the RNAi pathway in *C. elegans* (Tops et al., 2005)(**Figure S4**). However, in *D. melanogaster* RDE-2 is missing and the orthologue of MUT-7, Nibbler (Nbr), is involved in the maturation of piRNA 3' ends (Wang et al., 2016). As in insects, RDE-2 is missing in flatworms suggesting that the MUT-7/Nbr homolog in 'turbellarians' is involved in piRNA biogenesis but not in the RNAi pathway in free living flatworms (**Figures 1**, **8A** and **Table S6**).

DISCUSSION

Platyhelminthes comprise one of the early diverging phyla of bilateral metazoans, originated approximately 839 million years ago (Hedges et al., 2015), inhabiting a wide range of ecosystems and particularly successful in adapting to parasitic life. Since small non-coding RNAs have been implicated in regulating developmental transitions, we reasoned that they might be involved in the adaptation to parasitism way of life. Therefore, comparing the conservation of small RNA pathways among the paraphyletic group of free-living ('turbellarians') and the monophyletic parasitic Neodermatans may provide evolutionary clues to disentangle possible adaptations to parasitism.

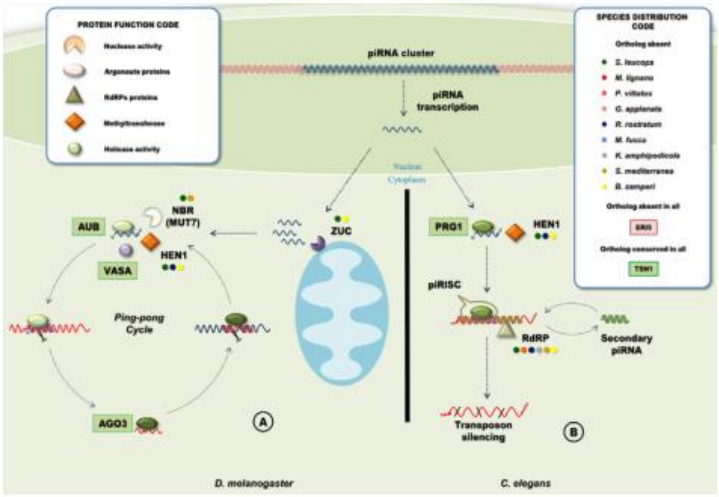


FIGURE 8 | Piwi-interacting RNA pathways. **(A)** PiRNA pathway in *D. melanogaster*. PiRNA precursors are cleaved by a protein of the mitochondrial membrane, Zucchini (Zuc) producing the primary piRNAs that are loaded into Aub, maturation of piRNA requires the 2-O methylation and cleavage of the piRNA 3' end by HEN1 and NBR (MUT7), respectively. The primary piRNA and Aub form the piRISC, secondary piRNAs are generated in a slicer-dependent amplification loop that silence cytoplasmic TE transcripts named "Ping-Pong" cycle (Tóth et al., 2016). **(B)** PiRNA pathway in *C. elegans*. In the cytoplasm, piRNAs are bounded by a PRG protein and methylated at the 3' end by HEN1. The processed piRNA with PRG form the piRISC that will cleavage target RNA, target recognition is followed by the generation of secondary siRNAs mediated by RdRPs (Weick and Miska, 2014). Circles indicate species with missing homologous genes ("Species Distribution Code" box). A 'shape' code was used to indicate predicted function of factors ("Protein Function Code" box).

The Argonaute Gene Family

We have previously shown (Fontenla et al., 2017) that key proteins involved in small RNA pathways are conserved in all major clades of Platyhelminthes; however, clear differences between clades are evident, particularly the complete absence of the Piwi pathway genes in Neodermatans.

While miRNA class Ago proteins are conserved in all flatworms, a platyhelminth-specific family that we named FLAgos (Fontenla et al., 2017), showed independent gene amplifications in Trematodes and Cestodes but not in 'turbellarians', with the only exception of a duplication in *S. mediterranea* (**Figure 3**). This group might have originated as a highly divergent siRNA class Ago, that was further amplified and diversified. On the other hand, while the parasitic Neodermatans lack Piwi-like proteins, these are amplified in free-living flatworms, including the canonical Piwi 2 class, and a new group, the FLiwis, specific of free living Rhabditophora (**Figure 4**). As in FLAgos, since Piwi1 class homologues are missing, is possible that the Fliwi group represents a fast-evolving piwi1 class that diverged within flatworms. Interestingly, neither FLAgos, nor FLiwi are shared by other lophotrochozoans lineages analyzed in this study, despite amplifications can also been detected within them.

Piwi proteins are involved in the biogenesis and activities of piRNAs, being crucial at silencing transposable elements. In *S. mediterranea*, Piwi genes were reported to be essential in the regeneration and homeostasis of neoblasts, the pluripotent stem cells of Platyhelminthes (Reddien et al., 2005; Palakodeti et al., 2008). Additionally, *S. mediterranea* piwi-1 (SmeT036375) is highly expressed in blastomeres and is critical during embryogenesis and organogenesis (Davies et al., 2017). The highest levels of piwi-1 are found in epidermal progenitors and tetraspanin-1-positive neoblasts. Differentiation of pluripotent neoblasts into fate-determined progenitors and terminally differentiated cells is accompanied by a successive reduction of piwi-1 levels (Kim et al., 2020). In *M. lignano*, piwi-1 (Mli034222) but not piwi-2 (Mli016226) was found to be involved in the piRNA pathway in both germline and somatic cells, as well as in the maintenance of stem cells (Zhou et al., 2015). It is of interest to decipher if similar functional restrictions or labor division is found in other 'turbellarians', now that we show that the key genes are conserved.

FLAgo proteins show less sequence conservation, due to a higher substitution rate, a feature generally associated with acquiring novel functions. We show evidence of site-specific diversifying selection mainly in FLAgos compared to other Argonaute groups. We hypothesize that the sequence variation observed for FLAgos might have been associated to specialization in the gene function within this cluster. Further experimental evidence is needed in to validate this hypothesis, for instance evaluating if the substitutions detected are associated with changes in functional domains or protein conformations. Even more, flatworm-specific proteins like the FLAgos (not conserved in the host) could be targets for new drugs or vaccines. In that sense, a drug that specifically target this highly diverting subgroup could be a potential innovation in the treatment of helminthiasis. Such drug has already been proposed in *in silico* modeling to target Hsa Ago-2 (Schmidt et al., 2013).

The Dicer Family

In flatworms, the Dicer family is organized in two subgroups. While the Dcr-1 group is invariable with only a single gene per species, the Dcr-2 group is heterogenous both in number of copies and structure (**Figure 6A**). The shortening of Dcr-2 is probably the consequence of a genomic reorganization that eliminated the first 14 exons of the gene in the ancestor of all trematodes (**Figure 6B**). Transposable elements (TE) are recognized as contributors to genomic innovation as well as genomic instability across a wide variety of species (Klein and O'Neill, 2018). We have previously reported the accumulation of TEs in the genomes of trematodes, especially FBT, as an extreme example, more than 50% of the genomes of *Fasciola* spp. corresponds to repetitive elements (Choi et al., 2020). It is tempting to speculate that the accumulation of TEs in the ancestor of the trematode class has driven the Dcr2 reorganization and, possibly, further accumulation of TEs in the FBTs contributed to generate a duplication of Dcr2 (Dcr3) in that linage. Whether these shorter Dcr 2/3 are functional or represent a pseudogene awaits confirmation. However, while Dcr1 and Dcr2 are express across different stages, Dcr3 seems to be limited to eggs (Fontenla et al., 2017).

The detection of Dcr-2 proteins in *G. applanate* and *M. fusca* with very similar structures to the ones in trematodes, may indicate that the genomic reorganization detected in trematodes may have occurred more than once during the evolution of flatworms. However, given the fragmentation of the 'turbellarian' transcriptomes, further genomic data are needed to verify these observations.

The Vasa and Vasa-Like Genes

Vasa genes in planarians are expressed in ovary and testis of sexual worms and in the totipotent cells (neoblasts) of asexual ones (Shibata et al., 1999). Increase in the expression of Vasa was detected in growing blastema of regenerating planarians and lost in irradiated organisms (Shibata et al., 1999). Belle (also known as PL10), on the other hand, is a Vasa-related protein, that has conserved roles in fertility and development, and co-localizes with Vasa to the germline (Johnstone et al., 2005). Vasa was proposed to be part of the biogenesis of piRNAs and to be differentially conserved between 'turbellarians' and Neodermatans (Skinner et al., 2014). Like the absence of Piwi, the absence of Vasa in Neodermatans results puzzling. Piwi and Vasa are widely known among developmental biologists as germline markers. Vasa was even found to be expressed in the germline of early branching Metazoans like the ctenophore clade, suggesting a central role in the development of all Metazoans (Mochizuki et al., 2001; Rebscher et al., 2007; Alié et al., 2011). Even more, vasa mutants in *D. melanogaster* fail to form pole cells, the precursor of the gonadal germ cell population, and show deletions of abdominal segments (Schupbach and Wieschaus, 1986). It has been proposed that Vasa genes arose by duplication of an ancestral PL-10-related gene before

appearance of sponges but after the diversion of fungi and plants (Mochizuki et al., 2001). Our data suggest that a second duplication of PL-10 took place in an early ancestor of the Platyhelminthes producing a flatworm specific family of PL-10-related genes. If that is the case, we speculate that a redundant role of Vasa, PL-10-related and flatworm PL-10 in 'turbellarians' was further simplified in Neodermatans with the loss of Vasa. Why the loss of Vasa was evolutionary favored in Neodermatans is still unknown; however, the germline expression and role in the gametogenesis of flatworm specific PL-10 has been reported in Neodermatans by RNAi assays (Ohashi et al., 2007). Moreover, vasa-like genes, i.e. PL-10 are strongly expressed in the ovary of *Schistosoma mansoni* female adult worms and showed high expression in female adults and eggs laid *in vitro* by worms in culture (Skinner et al., 2012). More recently, RNAi against vasa/PL10-like gene -1 in *S. mansoni* adult female worms resulted in smaller ovaries and a reduced number of ovarian dividing cells (Skinner et al., 2020). Similarly, the knockdown of vasa-like genes in *S. japonicum* induced changes in the morphology of the reproductive organs, especially in the female ovary, vitellarium and the male testes. In addition, a significant reduction in egg production in knocked-down parasites was evident (He et al., 2018).

PiRNA Biogenesis in 'Turbellarians'

The piwiRNAs (piRNAs) are generated either from RNA transcripts of active transposable element (TE) copies or from transcripts originated from specialized loci in the genome called piRNA clusters. In general, piRNAs generated from piRNA clusters are mostly antisense to TE mRNA sequences (Tóth et al., 2016). However, these regulatory non-coding RNAs are also originated from different biogenesis pathways depending on the species. In *D. melanogaster* and vertebrates, piRNAs are 26-30 nt in length. These are derived from single-stranded piRNA precursors and processed in the cytoplasm by Zucchini (Zuc). This is a protein with endonuclease activity for single-stranded RNA and expressed in the mitochondrial surface with a predicted phospholipase D-like domain (**Figure 8A**). The piRNAs generated this way preferentially have a 5'-end uracil. HEN1 is required for 2'-O-methylation of maturing piRNAs (Horwich et al., 2007; Montgomery et al., 2012). Mature piRNAs are bounded by the Piwi protein Aubergine (Aub), to form the piRISC that targets and degrades TE mRNAs. Ago3, on the other hand, binds to TE mRNA cleaved sequences that contains an adenosine residue at position 10 and will target piRNA sequences resulting in an amplification loop named ping-pong cycle (Weick and Miska, 2014). The ping-pong cycle includes not only Aub and Ago3, but also Vasa, that has two proposed roles in piRNA processing. First, Vasa participates in the assembly of the ping-pong complex (Xiol et al., 2014). Second, the RNA-unwinding activity of Vasa helps to release cleaved products from the piRNA-protein complex to facilitate the ping-pong cycle (Nishida et al., 2015).

C. elegans piRNAs are shorter (21 nt long), also with a 5' uracil, and require methylation by HEN1 for maturation. Mature piRNAs bound to the Piwi orthologue PRG1, form the piRISC complex that targets and silence TE mRNAs. However, instead of the ping pong mechanism, the amplification of the silencing signal relies on RNA dependent RNA polymerases (RdRPs) (**Figure 8B**) (Weick and Miska, 2014).

The conservation of Zuc, Nbr, Vasa and RdRPs in some 'turbellarian' species rises a question about the biogenesis of piRNAs in the free-living clades. In *M. lignano*, it has been reported that the knock down of vasa produce a severe reduction in the piRNA fraction (Zhou et al., 2015). Additionally, in *M. lignano*, *S. mediterranea* and *Dugesia japonica* where the small RNA population has been sequenced, piRNAs are ~32 nt in length preferentially displaying U at the 5' end (Palakodeti et al., 2008; Friedländer et al., 2009; Qin et al., 2012; Zhou et al., 2015) like the ones described in *D. melanogaster*. Besides, a preference for A at position 10 and the overlap of reads by 10 nt suggest that a ping-pong cycle occur in these species with no evidence of any other mechanism of amplification. Therefore, it is possible to speculate that the RdRPs detected in *M. lignano* (and some other 'turbellarians') might not be involved primarily in the amplification of piRNAs. In any event, additional experimental evidence, possibly involving RNAi against RdRPs genes and sequencing of the small RNA population, is needed to test this hypothesis.

Alternative Solutions to Piwi Absence in Parasites

The absence of Piwi in addition to the amplification of the FLAgos in Neodermatans raise the hypothesis that some of the FLAgos substitute the role of the Piwi proteins in this clade (Skinner et al., 2014). In this regard, Cai et al. (2012) sequenced the population of small RNAs associated to the FL-Ago SjAgo2 (Sja_0045200) and found that it was preferentially associated with siRNAs derived from LINE and LTR retrotransposons, the main targets of Piwi proteins (see below). This observation suggests that the FLAgos could be at least partially mimicking the role of the lost Piwi genes in Neodermatans. Furthermore, the silencing of SmAgo2 resulted in a moderate increased expression of transposable elements, suggesting that this protein might be involved in regulating transposons in *Schistosoma mansoni* (Protasio et al., 2020).

The absence of piwi pathway proteins in parasitic species seems to be a consistent trend. We here provide strong evidence of the complete absence of the pathway across all parasitic flatworms. Similarly, the absence of Piwi have been reported in all nematode clades, except clade V, the one containing the model species *C. elegans*, and some animal parasites as *Haemonchus contortus* and *Pristionchus pacificus* (Sarkies et al., 2015). Other proteins of the piwi pathway were also absent in non-clade V nematode, confirming the absence of a functional pathway. Furthermore, the piwi pathway was also found absent in the dust and scabies mite genomes (Mondal et al., 2018). Both in nematodes and in dust mites there is evidence that alternative siRNA related mechanisms are involved in controlling TEs. The amplification of genes associated with the main small RNA

pathways in parasitic flatworms is suggestive of a similar cooption of functions. Furthermore, as already mentioned primary evidence show that this might be the case in parasitic flatworms (Cai et al., 2012; Protasio et al., 2020). Since TEs are recognized as contributors to genomic innovation as well as genomic instability across a wide variety of species (Klein and O'Neill, 2018), is tempting to speculate that piwi loss might be associated to rapid genomic reorganization leading to adopt a parasitic way of life.

CONCLUSIONS

We provide a strong bioinformatic support for the presence and absence of key proteins involved in small RNA pathways in early diverging free-living flatworms, suggesting that miRNA regulation, piRNA mediated silencing and RNAi are ancestral regulatory mechanisms in flatworms. In addition, differences observed in later evolving parasitic species strongly suggest that small RNA mediated mechanisms might have been also relevant during the transition to parasitism.

A long and still unsettled discussion has taken place regarding the biological simplification occurred in flatworms, especially in the parasitic Neodermatans. It is unclear, yet, if the process of loss of redundancy is the product of an adaptive mechanism to parasitism (Tsai et al., 2013; McNulty et al., 2017) or is an ancestral characteristic acquired by the Neodermata clade (Hahn et al., 2014). Interestingly, in the present study we found that the 'turbellarian' parasite *K. amphipodicola* showed no major differences in the conservation of small RNA pathway factors respect to other phylogenetically related free-living 'turbellarians', like *S. mediterranea*, including the conservation of a putative functional piRNA pathway that has been lost in Neodermatans (Skinner et al., 2014). Then, it is possible to hypothesize that the overall simplification associated to the Neodermatans is not a characteristic needed for parasitism in the phylum Platyhelminthes. However, the complete absence of the piwi pathway mediators in all trematodes and cestodes is suggestive of an early loss in an ancestor of the Neodermatans. This single loss might have had dramatic evolutionary consequences, since transposable elements might have driven genome instability that led to biochemical, morphological and functional transformations, for instance the origin of the neodermis (Caira and Littlewood, 2013), and other changes that favored the adaptation to a novel lifestyle. Along these lines, it is tempting to think that the later independent amplification of FLAgos, and Belle/PL10 in trematodes and cestodes might have resulted from an adaptation to these changes, either for control of transposons, and/or generating novel regulatory mechanisms mediated by small non-coding RNAs.

Different experimental approaches can be considered to test this hypothesis. Functional genomic tools tested in flatworms like RNAi (Dell'Oca et al., 2014), transgenesis (Rinaldi et al., 2012; Suttiprapa et al., 2016), genome-editing by CRISPR-Cas9 (Lok et al., 2017; Ittiprasert et al., 2019; Sankaranarayanan et al., 2020) and immunoprecipitation assays (Free et al., 2009) could be used to define the function of the factors reported here or to detect novel ones. Additionally, experimental evidence involving chromatin immunoprecipitation of nucleosome core particles followed by high throughput sequencing has proven to be useful to detect chromatin modification triggered by dsRNA (Gu et al., 2012).

To conclude, these findings together with our previous report (Fontenla et al., 2017), describe novel features of the biology and evolution that are unique to Platyhelminthes, implying that subtle mechanisms involved in the small RNA pathways of flatworms are different to the ones described in model organisms like mammals, *C. elegans* or *D. melanogaster*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

SF performed the acquisition, bioinformatics analysis and interpretation of data and contributed in writing the manuscript. GR was involved in drafting the manuscript and critical revision of its content. JT participated in the design of the study and the interpretation of data, drafting the manuscript and critical revision of its content. All authors contributed to the article and approved the submitted version.

FUNDING

The research was founded by grant Iniciación a la Investigación 2019 (CSIC-UdelaR). SF, GR and JT are researchers from the Sistema Nacional de Investigadores (SNI-ANII). JT is also researcher from Pedeciba.

ACKNOWLEDGMENTS

We thank Dr. Andrés Iriarte for discussion and comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2021.653695/full#supplementary-material

REFERENCES

Abdelhaleem, M. (2005). RNA Helicases: Regulators of Differentiation. *Clin. Biochem.* 38, 499–503. doi: 10.1016/j.clinbiochem.2005.01.010

Alié, A., Leclère, L., Jager, M., Dayraud, C., Chang, P., Le Guyader, H., et al. (2011). Somatic Stem Cells Express Piwi and Vasa Genes in an Adult Ctenophore: Ancient Association of "Germline Genes" With Stemness. *Dev. Biol.* 350, 183–197. doi: 10.1016/j.ydbio.2010.10.019

Blythe, M. J., Kao, D., Malla, S., Rowsell, J., Wilson, R., Evans, D., et al. (2010). A Dual Platform Approach to Transcript Discovery for the Planarian Schmidtea Mediterranea to Establish RNAseq for Stem Cell and Regeneration Biology. *PLoS One* 5, e15617. doi: 10.1371/journal.pone.0015617

Cai, P., Gobert, G. N., and McManus, D. P. (2016). MicroRNAs in Parasitic Helminthiases: Current Status and Future Perspectives. *Trends Parasitol.* 32, 71–86. doi: 10.1016/j.pt.2015.09.003

Cai, P., Piao, X., Hou, N., Liu, S., Wang, H., and Chen, Q. (2012). Identification and Characterization of Argonaute Protein, Ago2 and its Associated Small RNAs in Schistosoma Japonicum. *PLoS Negl. Trop. Dis.* 6:e41337. doi: 10.1371/journal.pntd.0001745

Caira, J. N., and Littlewood, D. T. J. (2013). "Worms, Platyhelminthes," in *Encyclopedia of Biodiversity* (Waltham: Elsevier), 437–469. doi: 10.1016/B978-0-12-384719-5.00166-0

Charif, D., and Lobry, J. R. (2007). "Seqin{R} 1.0-2: A Contributed Package to the {R} Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis," in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations Biological and Medical Physics, Biomedical Engineering*. Eds. U. Bastolla, M. Porto, H. E. Roman and M. Vendruscolo (New York: Springer Verlag), 207–232.

Choi, Y.-J., Fontenla, S., Fischer, P. U., Le, T. H., Costábile, A., Blair, D., et al. (2019). Adaptive Radiation of the Flukes of the Family Fasciolidae Inferred From Genome-Wide Comparisons of Key Species. *Mol. Biol. Evol.* 37, 84–99. doi: 10.1093/molbev/msz204

Cucher, M., Prada, L., Mourglia-Ettlin, G., Dematteis, S., Camicia, F., Asurmendi, S., et al. (2011). Identification of Echinococcus Granulosus microRNAs and Their Expression in Different Life Cycle Stages and Parasite Genotypes. *Int. J. Parasitol.* 41, 439–448. doi: 10.1016/j.ijpara.2010.11.010

Davies, E. L., Lei, K., Seidel, C. W., Kroesen, A. E., McKinney, S. A., Guo, L., et al. (2017). Embryonic Origin of Adult Stem Cells Required for Tissue Homeostasis and Regeneration. *Elife* 6, 1–35. doi: 10.7554/eLife.21052

Dell'Oca, N., Basika, T., Corvo, I., Castillo, E., Brindley, P. J., Rinaldi, G., et al. (2014). RNA Interference in Fasciola Hepatica Newly Excysted Juveniles: Long dsRNA Induces More Persistent Silencing Than SiRNA. *Mol. Biochem. Parasitol.* 197, 28–35. doi: 10.1016/j.molbiopara.2014.10.001

Ding, S. W., and Voinnet, O. (2007). Antiviral Immunity Directed by Small RNAs. *Cell* 130, 413–426. doi: 10.1016/j.cell.2007.07.039

Egger, B., Lapraz, F., Tomiczek, B., Müller, S., Dessimoz, C., Girstmair, J., et al. (2015). A Transcriptomic-Phylogenomic Analysis of the Evolutionary Relationships of Flatworms. *Curr. Biol.* 25, 1347–1353. doi: 10.1016/j.cub.2015.03.034

Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., et al. (2012). The Structure of Human Argonaute-2 in Complex With Mir-20a. *Cell* 150, 100–110. doi: 10.1016/j.cell.2012.05.017

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and Specific Genetic Interference by Double-Stranded RNA in Caenorhabditis Elegans. *Nature* 391, 806–811. doi: 10.1038/35888

Fontenla, S., Dell'Oca, N., Smircich, P., Tort, J. F., and Siles-Lucas, M. (2015). The miRnome of Fasciola Hepatica Juveniles Endorses the Existence of a Reduced Set of Highly Divergent Micro RNAs in Parasitic Flatworms. *Int. J. Parasitol.* 45, 901–913. doi: 10.1016/j.ijpara.2015.06.007

Fontenla, S., Rinaldi, G., Smircich, P., and Tort, J. F. (2017). Conservation and Diversification of Small RNA Pathways Within Flatworms. *BMC Evol. Biol.* 17, 215. doi: 10.1186/s12862-017-1061-5

Free, R. B., Hazelwood, L. A., and Sibley, D. R. (2009). Identifying Novel Protein-Protein Interactions Using Co-Immunoprecipitation and Mass Spectroscopy. *Curr. Protoc. Neurosci.* 0 5, Unit–5.28. doi: 10.1002/0471142301.ns0528s46

Friedländer, M. R., Adamidi, C., Han, T., Lebedeva, S., Isenbarger, T., Hirst, M., et al. (2009). High-Resolution Profiling and Discovery of Planarian Small RNAs. *Proc. Natl. Acad. Sci. U. S. A.* 106, 11546–11551. doi: 10.1073/pnas.0905222106

Fromm, B., Worren, M. M., Hahn, C., Hovig, E., and Bachmann, L. (2013). Substantial Loss of Conserved and Gain of Novel Microrna Families in Flatworms. *Mol. Biol. Evol.* 30, 2619–2628. doi: 10.1093/molbev/mst155

Gao, Z., Wang, M., Blair, D., Zheng, Y., and Dou, Y. (2014). Phylogenetic Analysis of the Endoribonuclease Dicer Family. *PLoS One* 9, 1–7. doi: 10.1371/journal.pone.0095350

Gu, S. G., Pak, J., Guang, S., Maniar, J. M., Kennedy, S., and Fire, A. (2012). Amplification of siRNA in Caenorhabditis Elegans Generates a Transgenerational Sequence-Targeted Histone H3 Lysine 9 Methylation Footprint. *Nat. Genet.* 44, 157–164. doi: 10.1038/ng.1039

Guil, S., and Esteller, M. (2015). Rna-RNA Interactions in Gene Regulation: The Coding and Noncoding Players. *Trends Biochem. Sci.* 40, 248–256. doi: 10.1016/j.tibs.2015.03.001

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Hahn, C., Fromm, B., and Bachmann, L. (2014). Comparative Genomics of Flatworms (Platyhelminthes) Reveals Shared Genomic Features of Ecto- and Endoparasitic Neodermata. *Genome Biol. Evol.* 6, 1105–1117. doi: 10.1093/gbe/evu078

Hall, T. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/Nt. *Nucleic Acids Symp. Ser.* 41, 95–98.

Han, H. (2018). Rna Interference to Knock Down Gene Expression. *Methods Mol. Biol.* 1706, 293–302. doi: 10.1007/978-1-4939-7471-9_16

Hedges, S. B., Marin, J., Suleski, M., Paymer, M., and Kumar, S. (2015). Tree of Life Reveals Clock-Like Speciation and Diversification. *Mol. Biol. Evol.* 32 (4), 835–845. doi: 10.1093/molbev/msv037

He, Z., Zhang, H., Gao, S., Lercher, M. J., Chen, W.-H., and Hu, S. (2016). Evolview v2: An Online Visualization and Management Tool for Customized and Annotated Phylogenetic Trees. *Nucleic Acids Res.* 44, W236–41. doi: 10.1093/nar/gkw370

He, S., Zhu, L., Liu, F., Liu, Q., Shao, Y., Hua, M., et al. (2018). Functions of the Vasa Gene in Schistosoma Japonicum as Assessed by RNA Interference. *Gene* 638, 13–19. doi: 10.1016/j.gene.2017.09.054

Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., et al. (2007). The Drosophila Rna Methyltransferase, Dmhen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC. *Curr. Biol.* 17 (14), 1265–1272. doi: 10.1016/j.cub.2007.06.030

Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). Cd-Hit Suite: A Web Server for Clustering and Comparing Biological Sequences. *Bioinformatics* 26, 680–682. doi: 10.1093/bioinformatics/btq003

Huang, B., and Zhang, R. (2014). Regulatory non-Coding RNAs: Revolutionizing the RNA World. *Mol. Biol. Rep.* 41, 3915–3923. doi: 10.1007/s11033-014-3259-6

Ilgová, J., Jedličková, L., Dvořáková, H., Benovics, M., Mikeš, L., Janda, L., et al. (2017). A Novel Type I Cystatin of Parasite Origin With Atypical Legumain-Binding Domain. *Sci. Rep.* 7, 17526. doi: 10.1038/s41598-017-17598-2

Ittiprasert, W., Mann, V. H., Karinshak, S. E., Coghlan, A., Rinaldi, G., Sankaranarayanan, G., et al. (2019). Programmed Genome Editing of the Omega-1 Ribonuclease of the Blood Fluke, Schistosoma Mansoni. *Elife* 8, e41337. doi: 10.7554/eLife.41337

Jaronczyk, K., Carmichael, J. B., and Hobman, T. C. (2005). Exploring the Functions of RNA Interference Pathway Proteins: Some Functions are More RISCy Than Others? *Biochem. J.* 387, 561–571. doi: 10.1042/BJ20041822

Jehn, J., Gebert, D., Pipilescu, F., Stern, S., Kiefer, J. S. T., Hewel, C., et al. (2018). PIWI Genes and piRNAs are Ubiquitously Expressed in Mollusks and Show Patterns of Lineage-Specific Adaptation. *Commun. Biol.* 1 (1), 137. doi: 10.1038/s42003-018-0141-4

Johnson, L. S., Eddy, S. R., and Portugaly, E. (2010). Hidden Markov Model Speed Heuristic and Iterative HMM Search Procedure. *BMC Bioinf.* 11, 431. doi: 10.1186/1471-2105-11-431

Johnstone, O., Deuring, R., Bock, R., Linder, P., Fuller, M. T., and Lasko, P. (2005). Belle is a Drosophila DEAD-Box Protein Required for Viability and in the Germ Line. *Dev. Biol.* 277, 92–101. doi: 10.1016/j.ydbio.2004.09.009

Katoh, K., and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kim, I. V., Riedelbauch, S., and Kuhn, C. D. (2020). The piRNA Pathway in Planarian Flatworms: New Model, New Insights. *Biol. Chem.* 401, 1123–1141. doi: 10.1515/hsz-2019-0445

Klein, S. J., and O'Neill, R. J. (2018). Transposable Elements: Genome Innovation, Chromosome Diversity, and Centromere Conflict. *Chromosom. Res.* 26, 5–23. doi: 10.1007/s10577-017-9569-5

Kong, R., Xu, L., Piao, L., Zhang, D., Hou, T.-J., and Chang, S. (2017). Exploring the RNA-bound and RNA-free Human Argonaute-2 by Molecular Dynamics Simulation Method. *Chem. Biol. Drug Des.* 90, 753–763. doi: 10.1111/cbdd.12997

Kuales, G., De Mulder, K., Glashauser, J., Salvenmoser, W., Takashima, S., Hartenstein, V., et al. (2011). Boule-Like Genes Regulate Male and Female Gametogenesis in the Flatworm Macrostomum Lignano. *Dev. Biol.* 357, 117–132. doi: 10.1016/j.ydbio.2011.06.030

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). Mega X: Molecular Evolutionary Genetics Analysis Across Computing Platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Larsson, K., and Jondelius, U. (2008). Phylogeny of Catenulida and Support for Platyhelminthes. *Org. Divers. Evol.* 8, 378–387. doi: 10.1016/j.ode.2008.09.002

Laumer, C. E., and Giribet, G. (2014). Inclusive Taxon Sampling Suggests a Single, Stepwise Origin of Ectolecithality in Platyhelminthes. *Biol. J. Linn. Soc* 111, 570–588. doi: 10.1111/bij.12236

Laumer, C. E., Hejnol, A., and Giribet, G. (2015). Nuclear Genomic Signals of the "Microturbellarian" Roots of Platyhelminth Evolutionary Innovation. *Elife* 4, 1–31. doi: 10.7554/eLife.05503

Lee, R. C. (1993). The C . Elegans Heterochronic Gene Lin-4 Encodes Small RNAs With Antisense Complementarity to & Il-14. *Cell* 75, 843–854. doi: 10.1016/0092-8674(93)90529-Y

Lefort, V., Longueville, J. E., and Gascuel, O. (2017). Sms: Smart Model Selection in Phyml. *Mol. Biol. Evol.* 34, 2422–2424. doi: 10.1093/molbev/msx149

Lok, J. B., Shao, H., Massey, H. C., and Li, X. (2017). Transgenesis in Strongyloides and Related Parasitic Nematodes: Historical Perspectives, Current Functional Genomic Applications and Progress Towards Gene Disruption and Editing. *Parasitology* 144, 327–342. doi: 10.1017/S0031182016000391

Maldonado, L. L., Assis, J., Araújo, F. M. G., Salim, A. C. M., Macchiaroli, N., Cucher, M., et al. (2017). The *Echinococcus Canadensis* (G7) Genome: A Key Knowledge of Parasitic Platyhelminth Human Diseases. *BMC Genomics* 18, 204. doi: 10.1186/s12864-017-3574-0

McNulty, S. N., Tort, J. F., Rinaldi, G., Fischer, K., Rosa, B. A., Smircich, P., et al. (2017). Genomes of Fasciola Hepatica From the Americas Reveal Colonization With Neorickettsia Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers. *PloS Genet.* 13, e1006537. doi: 10.1371/journal.pgen.1006537

Mochizuki, K., Nishimiya-Fujisawa, C., and Fujisawa, T. (2001). Universal Occurrence of the Vasa-Related Genes Among Metazoans and Their Germline Expression in Hydra. *Dev. Genes Evol.* 211, 299–308. doi: 10.1007/s004270100156

Moguel, B., Bobes, R. J., Carrero, J. C., and Laclette, J. P. (2015). Transfection of Platyhelminthes. *BioMed. Res. Int.* 2015:206161. doi: 10.1155/2015/206161

Mondal, M., Klimov, P., and Flynt, A. S. (2018). Rewired RNAi-Mediated Genome Surveillance in House Dust Mites. *PloS Genet.* 14 (1), e1007183. doi: 10.1371/journal.pgen.1007183

Montgomery, T. A., Rim, Y.-S., Zhang, C., Dowen, R. H., Phillips, C. M., Fischer, S. E. J., et al. (2012). Piwi Associated siRNAs and PiRNAs Specifically Require the Caenorhabditis Elegans HEN1 Ortholog Henn-1. *PloS Genet.* 8 (4), e1002616. doi: 10.1371/journal.pgen.1002616

Mourão, M. M., Grunau, C., Loverde, P. T., Jones, M. K., and Oliveira, G. (2012). Recent Advances in Schistosoma Genomics. *Parasite Immunol.* 34, 151–162. doi: 10.1111/j.1365-3024.2011.01349.x

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S. L. (2012). Detecting Individual Sites Subject to Episodic Diversifying Selection. *PloS Genet.* 8, e1002764. doi: 10.1371/journal.pgen.1002764

Nei, M., and Gojobori, T. (1986). Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. *Mol. Biol. Evol.* 3, 418–426. doi: 10.1093/oxfordjournals.molbev.a040410

Niaz, S. (2018). The AGO Proteins: An Overview. *Biol. Chem.* 399, 525–547. doi: 10.1016/j.celrep.2014.12.013

Nishida, K. M., Iwasaki, Y. W., Murota, Y., Nagao, A., Mannen, T., Kato, Y., et al. (2015). Respective Functions of Two Distinct Siwi Complexes Assembled During PIWI-Interacting RNA Biogenesis in Bombyx Germ Cells. *Cell Rep.* 10, 193–203. doi: 10.1016/j.celrep.2014.12.013

Oey, H., Zakrzewski, M., Narain, K., Devi, K. R., Agatsuma, T., Nawaratna, S., et al. (2018). Whole-Genome Sequence of the Oriental Lung Fluke Paragonimus Westermani. *Gigascience* 8, 1–8. doi: 10.1093/gigascience/giy146

Ohashi, H., Umeda, N., Hirazawa, N., Ozaki, Y., Miura, C., and Miura, T. (2007). Expression of Vasa (Vas)-Related Genes in Germ Cells and Specific Interference With Gene Functions by Double-Stranded RNA in the Monogenean, Neobenedenia Girellae. *Int. J. Parasitol.* 37, 515–523. doi: 10.1016/j.ijpara.2006.11.003

Orii, H., Mochii, M., and Watanabe, K. (2003). A Simple "Soaking Method" for RNA Interference in the Planarian Dugesia Japonica. *Dev. Genes Evol.* 213, 138–141. doi: 10.1007/s00427-003-0310-3

Palakodeti, D., Smielewska, M., and Graveley, B. R. (2006). MicroRNAs From the Planarian Schmidtea Mediterranea: A Model System for Stem Cell Biology. *RNA* 12, 1640–1649. doi: 10.1261/rna.117206

Palakodeti, D., Smielewska, M., Lu, Y.-C., Yeo, G. W., and 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. doi: 10.1261/rna.1085008

Park, M. S., Phan, H. D., Busch, F., Hinckley, S. H., Brackbill, J. A., Wysocki, V. H., et al. (2017). Human Argonaute3 has Slicer Activity. *Nucleic Acids Res.* 45, 11867–11877. doi: 10.1093/nar/gkx916

Protasio, A. V., Rawlinson, K. A., Miska, E. A., Berriman, M., and Rinaldi, G. (2020). Evidence for Transposable Element Control by Argonautes in a Parasitic Flatworm Lacking the piRNA Pathway. *bioRxiv [Preprint]*. Available at: https://www.biorxiv.org/content/10.1101/670372v3.full.pdf+html (Accessed January 10, 2021).

Protasio, A. V., Van Dongen, S., Collins, J., Quintais, L., Ribeiro, M., Sessler, F., et al. (2017). MiR-277 / 4989 Regulate Transcriptional Landscape During Juvenile to Adult Transition in the Parasitic Helminth Schistosoma Mansoni. *PLoS Negl. Trop. Dis.* 11, e0005559. doi: 10.1371/journal.pntd.0005559

Qin, Y. F., Zhao, J. M., Bao, Z. X., Zhu, Z. Y., Mai, J., Huang, Y. B., et al. (2012). Identification of Small non-Coding RNAs in the Planarian Dugesia Japonica Via Deep Sequencing. *Genomics* 99, 315–321. doi: 10.1016/j.ygeno.2012.03.001

Rebscher, N., Zelada-González, F., Banisch, T. U., Raible, F., and Arendt, D. (2007). Vasa Unveils a Common Origin of Germ Cells and of Somatic Stem Cells From the Posterior Growth Zone in the Polychaete Platynereis Dumerilii. *Dev. Biol.* 306, 599–611. doi: 10.1016/j.ydbio.2007.03.521

Reddien, P. W., and Alvarado, A. S. (2004). Fundamentals of Planarian Regeneration. *Annu. Rev. Cell Dev. Biol.* 20, 725–757. doi: 10.1146/annurev.cellbio.20.010403.095114

Reddien, P. W., Oviedo, N. J., Jennings, J. R., Jenkin, J. C., and Sanchez Alvarado, A. (2005). SMEDWI-2 is a PIWI-like Protein That Regulates Planarian Stem Cells. *Science* 310, 1327–1330. doi: 10.1126/science.1116110

Rinaldi, G., Eckert, S. E., Tsai, I. J., Suttiprapa, S., Kines, K. J., Tort, J. F., et al. (2012). Germline Transgenesis and Insertional Mutagenesis in Schistosoma Mansoni Mediated by Murine Leukemia Virus. *PloS Pathog.* 8, 16. doi: 10.1371/journal.ppat.1002820

Rinaldi, G., Morales, M. E., Cancela, M., Castillo, E., Brindley, P. J., and Tort, J. F. (2008). Development of Functional Genomic Tools in Trematodes: RNA Interference and Luciferase Reporter Gene Activity in Fasciola Hepatica. *PloS Negl. Trop. Dis.* 2, e260. doi: 10.1371/journal.pntd.0000260

Rosa, B. A., Choi, Y.-J., McNulty, S. N., Jung, H., Martin, J., Agatsuma, T., et al. (2020). Comparative Genomics and Transcriptomics of 4 Paragonimus Species Provide Insights Into Lung Fluke Parasitism and Pathogenesis. *Gigascience* 9, 1–16. doi: 10.1093/gigascience/giaa073

Sánchez Alvarado, A., and Newmark, P. (1999). Double-Stranded RNA Specifically Disrupts Gene Expression During Planarian Regeneration. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5049–5054. doi: 10.1073/pnas.96.9.5049

Sandmann, T., Vogg, M. C., Owlarn, S., Boutros, M., and Bartscherer, K. (2011). The Head-Regeneration Transcriptome of the Planarian Schmidtea Mediterranea. *Genome Biol.* 12, R76. doi: 10.1186/gb-2011-12-8-r76

Sankaranarayanan, G., Coghlan, A., Driguez, P., Lotkowska, M. E., Sanders, M., Holroyd, N., et al. (2020). Large CRISPR-Cas-induced Deletions in the Oxamniquine Resistance Locus of the Human Parasite Schistosoma Mansoni. *Wellcome Open Res.* 5, 178. doi: 10.12688/wellcomeopenres.16031.1

Sarkies, P., Selkirk, M. E., Jones, J. T., Blok, V., Boothby, T., Goldstein, B., et al. (2015). Ancient and Novel Small RNA Pathways Compensate for the Loss of piRNAs in Multiple Independent Nematode Lineages. *PloS Biol.* 13, e1002061. doi: 10.1371/journal.pbio.1002061

Schmidt, M. F., Korb, O., and Abell, C. (2013). MicroRNA-Specific Argonaute 2 Protein Inhibitors. *ACS Chem. Biol.* 8, 2122–2126. doi: 10.1021/cb400246k

Schupbach, T., and Wieschaus, E. (1986). Germline Autonomy of Maternal-Effect Mutations Altering the Embryonic Body Pattern of Drosophila. *Dev. Biol.* 113, 443–448. doi: 10.1016/0012-1606(86)90179-X

Seppey, M., Manni, M., and Zdobnov, E. M. (2019). Busco: Assessing Genome Assembly and Annotation Completeness. *Methods Mol. Biol.* 1962, 227–245. doi: 10.1007/978-1-4939-9173-0_14

Shibata, N., Umesono, Y., Orii, H., Sakurai, T., Watanabe, K., and Agata, K. (1999). Expression of Vasa(Vas)-Related Genes in Germline Cells and Totipotent Somatic Stem Cells of Planarians. *Dev. Biol.* 206, 73–87. doi: 10.1006/dbio.1998.9130

Skinner, D. E., Popratiloff, A., Alrefaei, Y. N., Mann, V. H., Rinaldi, G., and Brindley, P. J. (2020). Functional Analysis of Vasa/PL10-Like Genes in the Ovary of Schistosoma Mansoni. *Mol. Biochem. Parasitol.* 236:111259. doi: 10.1016/j.molbiopara.2020.111259

Skinner, D. E., Rinaldi, G., Koziol, U., Brehm, K., and Brindley, P. J. (2014). How Might Flukes and Tapeworms Maintain Genome Integrity Without a Canonical piRNA Pathway? *Trends Parasitol.* 30, 123–129. doi: 10.1371/journal.pntd.0001686

Skinner, D. E., Rinaldi, G., Suttiprapa, S., Mann, V. H., Smircich, P., Cogswell, A. A., et al. (2012). Vasa-Like Dead-Box RNA Helicases of *Schistosoma mansoni*. *PLoS Negl. Trop. Dis.* 6, 26–31. doi: 10.1371/journal.pntd.0001686

Suttiprapa, S., Rinaldi, G., Tsai, I. J., Mann, V. H., Dubrovsky, L., Yan, H., et al. (2016). HIV-1 Integrates Widely Throughout the Genome of the Human Blood Fluke Schistosoma Mansoni. *PLoS Pathog.* 12, 1–26. doi: 10.1371/journal.ppat.1005931

Tolia, N. H., and Joshua-Tor, L. (2007). Slicer and the Argonautes. *Nat. Chem. Biol.* 3, 36–43. doi: 10.1038/nchembio848

Tops, B. B., Tabara, H., Sijen, T., Simmer, F., Mello, C. C., Plasterk, R. H., et al. (2005). RDE-2 Interacts With MUT-7 to Mediate RNA Interference in Caenorhabditis Elegans. *Nucleic Acids Res.* 33, 347–355. doi: 10.1093/nar/gki183

Tóth, K. F., Pezic, D., Stuwe, E., and Webster, A. (2016). The piRNA Pathway Guards the Germline Genome Against Transposable Elements. *Adv. Exp. Med. Biol.* 886, 51–77. doi: 10.1007/978-94-017-7417-8_4

Tsai, I. J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sánchez-Flores, A., Brooks, K. L., et al. (2013). The Genomes of Four Tapeworm Species Reveal Adaptations to Parasitism. *Nature* 496, 57–63. doi: 10.1038/nature12031

Wang, H., Ma, Z., Niu, K., Xiao, Y., Wu, X., Pan, C., et al. (2016). Antagonistic Roles of Nibbler and Hen1 in Modulating piRNA 3' Ends in Drosophila. *Development* 143, 530–539. doi: 10.1242/dev.128116

Wang, J., Paz, C., Padalino, G., Coghlan, A., Lu, Z., Gradinaru, I., et al. (2020). Large-Scale RNAi Screening Uncovers Therapeutic Targets in the Parasite Schistosoma Mansoni. *Science* 369, 1649–1653. doi: 10.1126/science.abb7699

Weick, E.-M., and Miska, E. (2014). piRNAs: From Biogenesis to Function. *Development* 141, 3458–3471. doi: 10.1242/dev.094037

Wynant, N., Santos, D., and Vanden Broeck, J. (2017). The Evolution of Animal Argonautes: Evidence for the Absence of Antiviral AGO Argonautes in Vertebrates. *Sci. Rep.* 7, 9230. doi: 10.1038/s41598-017-08043-5

Xiol, J., Spinelli, P., Laussmann, M. A., Homolka, D., Yang, Z., Cora, E., et al. (2014). RNA Clamping by Vasa Assembles a piRNA Amplifier Complex on Transposon Transcripts. *Cell* 157, 1698–1711. doi: 10.1016/j.cell.2014.05.018

Zheng, Y. (2012). Phylogenetic Analysis of the Argonaute Protein Family in Platyhelminths. *Mol. Phylogenet. Evol.* 66, 1050–1054. doi: 10.1016/j.ympev.2012.11.014

Zhou, X., Battistoni, G., El Demerdash, O., Gurtowski, J., Wunderer, J., Falciatori, I., et al. (2015). Dual Functions of Macpiwil in Transposon Silencing and Stem Cell Maintenance in the Flatworm Macrostomum Lignano. *RNA* 21, 1885–1897. doi: 10.1261/rna.052456.115

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Discusión General

Dado que los resultados ya fueron discutidos en detalle las secciones correspondientes de los artículos presentados, aquí solo se discutirán algunos aspectos generales.

A. Rol de los ARN pequeños en el desarrollo de *Fasciola hepatica* y en la interacción con el huésped

En este estudio analizamos la presencia de diversas poblaciones de ARNpq a lo largo de distintos estadios de vida de *F. hepatica*, y aquellos presentes en EVs del estadio adulto. Si bien las muestras se obtuvieron y fueron secuenciadas en tiempos diferentes, aquellas pertenecientes a estadios similares tendieron a agruparse en el análisis multidimensional (Figura S1 del manuscrito presentado en Resultados A), lo que sugiere que representan poblaciones consistentemente encontradas en esos estadios. A su vez, los conteos fueron normalizados para tener en consideración diferencias en la profundidad de la secuenciación. En nuestros datos identificamos un set complejo de ARNpq, compuesto por miARNs, fragmentos de tARNs y vtARNs.

Entre los miARNs diferencialmente expresados entre estadios, nuestras predicciones *in-silico* sugieren roles relevantes en el desarrollo y posiblemente en el proceso de invasión parasitario. Si bien es necesaria validación experimental, los roles reguladores de los miARNs son bien conocidos (Bushati & Cohen, 2007). Entre estas se destaca la regulación de la formación y liberación de vesículas durante la invasión del huésped, dado que la secreción de EVs durante la invasión parece ser un proceso finamente regulado (Hanna et al., 2019; Sánchez-López et al., 2020) (ver Figura 3B y Figura S6A del manuscrito presentado en Resultados A). Los cambios en la expresión de los miARNs durante la invasión podrían, al menos en parte, estar controlando la liberación de vesículas preformadas.

Curiosamente, otro análisis en estadios similares a los analizados por nosotros, pero con una estrategia diferente, publicado con posterioridad a la publicación de nuestros datos, reportó que los miARNs de *F. hepatica* pueden ser reguladores de vías metabólicas y de señalización, y también podrían estar regulando procesos vinculados a la invasión (Herron et al., 2022). En este caso, por la regulación de la secreción de proteasas, factores vinculados a la adquisición de nutrientes, componentes individuales de exosomas entre otros.

A su vez, los miARNs más abundantes en las EVs correspondieron a familias conservadas posiblemente dirigidas a regular vías de señalización e inmunomodulación del huésped (Figura 5, Tabla S6 y S7 del manuscrito presentado en Resultados A). En concordancia con nuestros resultados, Herron y col. (Herron et al., 2022) también reportaron la regulación de vías de señalización, como FOXO, o genes vinculados a la respuesta inmune, como IL-4 e IL-13, entre los blancos génicos en el huésped de los miARNs contenidos en EVs de *F. hepatica*. Curiosamente, esta parece ser una estrategia común en los helmintos, con un número creciente de reportes describiendo la absorción de miARN derivados de parásitos por parte de las células huésped, y evidencia *in vitro* de disminución en la expresión de genes del huésped asociados con la respuesta inmune (Arora et al., 2017)(Tabla 1). La modulación Th2 es un tema común en la evasión inmune de parásitos (Coakley et al., 2016). Resulta interesante que los miembros de la familia miR-10, bantam y let-7 vinculados a la modulación de esta respuesta se encuentran consistentemente entre los miARN más abundantes en las EVs secretados por trematodos adultos (Tabla S5 del manuscrito presentado en Resultados A), lo que sugiere que su enriquecimiento no se produce por casualidad y, por lo tanto, constituyen elementos centrales en la modulación de la respuesta del huésped.

Los fragmentos derivados de tRNA (tRDF) han sido relacionados con la regulación de diversos procesos que incluyen la traducción, la proliferación, la apoptosis, la formación de gránulos de estrés, la estabilización del ARNm, la expresión de transposones, la biogénesis de los ribosomas y la herencia de rasgos adquiridos (Tosar & Cayota, 2020). En este trabajo, detectamos una población sesgada de tRDFs en todas las etapas analizadas y describimos una nueva clase de fragmento de ARNt, producido por la escisión en el T-loop particularmente abundante en la fracción EV (Figura 6B y S7 del manuscrito presentado en Resultados A). Los fragmentos de

ARNt más abundantes de la fracción EV pueden formar estructuras homodiméricas estables que podrían explicar su mayor estabilidad.

Las funciones de los tRDFs son aún poco conocidas en los platelmintos, sin embargo, recientemente se informó que pueden estar involucrados en la regeneración de las planarias (Cao et al., 2020). De todas maneras, las funciones de estos fragmentos en la regulación de *F. hepatica* y/o en la interacción con el huésped son especulativas y requieren validación adicional.

Las ribonucleoproteínas vault son partículas grandes y huecas en forma de barril que se encuentran en el citoplasma de muchas células eucariotas. Aunque todavía se sabe poco sobre su función, parecen estar involucradas en funciones reguladoras clave que incluyen la autofagia, la apoptosis y la modulación de la expresión génica (revisado por Frascotti et al., 2021).

Nuestro enfoque basado en una búsqueda por homología laxa seguido de una curación manual exhaustiva fue clave para detectar vtARNs putativos en el genoma *F. hepatica* (Figura 8 y S9, y Tabla S9 del manuscrito presentado en Resultados A). Aunque se desconoce el papel funcional de los vtRNA de *F. hepatica*, su empaquetamiento selectivo en EVs de trematodos sugiere que podrían participar en las interacciones huésped-parásito.

B. Comparación de las vías de ARNpq entre platelmintos de vida libre y parásitos.

A principio de este siglo se develo un inesperado ‘mundo de ARN’, que ha adicionado nuevas capas de regulación a los sistemas biológicos, en muchos casos poco comprendidos aun hoy en día, especialmente en organismos no modelo. Estas actividades regulatorias son mediadas por un número creciente de clases de ARNs no codificantes. En el caso de los piARNs, estos son necesarios para silenciar elementos transponibles en las células de la línea germinal y mantener la estabilidad cromosómica. Previamente mostramos que proteínas claves involucradas en las vías de ARNpq están conservadas en los clados mayoritarios de platelmintos (Fontenla et al., 2017); sin embargo, detectamos evidentes diferencias entre clados, particularmente la completa ausencia de los genes de la vía Piwi en Neodermatas.

Si bien las proteínas de clase Ago de miARN se conservan en todos los platelmintos, una familia específica de platelmintos que llamamos FLAgos (Fontenla et al., 2017), mostró amplificaciones génicas independientes en trematodos y cestodos. Este grupo podría haberse originado a partir de una clase Ago de siARN altamente divergente, que se amplificó y diversificó rápidamente. Por otro lado, mientras que los Neodermatas parásitos carecen de proteínas tipo Piwi, estas se amplificaron en platelmintos de vida libre, incluyendo la clase canónica Piwi 2, y un nuevo grupo, los FLiwis, específicos de Rhabditophora de vida libre (Figuras 2 y 4 del manuscrito presentado en Resultados B). Al igual que en FLAgos, dado que faltan los homólogos de la clase Piwi1, es posible que el grupo Fliwi represente una clase piwi1 de rápida evolución que divergió dentro de los platelmintos.

A su vez, las proteínas FLAgo mostraron una menor conservación de secuencia, debido a una mayor tasa de sustitución, una característica generalmente asociada con la adquisición de funciones noveles. Mostramos evidencia de selección diversificadora sitio-específica principalmente en FLAgos en comparación con otros grupos de Argonauta (Figura 5 del manuscrito presentado en Resultados B). Dada el sesgo evolutivo de las FLAgos (no conservados en el huésped) creemos que podrían ser blancos para el desarrollo de nuevos medicamentos o vacunas. Fármacos anti-Ago ya se han propuesto en modelos *in silico* (Schmidt et al., 2013).

Entre los genes de la familia Dicer, reportamos el acortamiento de Dcr-2 en todos los trematodos analizados con una posterior duplicación (Dcr-3) en los FBT (Figura 6 del manuscrito presentado en Resultados B), posiblemente como consecuencia de reorganizaciones genómicas por la acumulación de TEs (Klein & O'Neill, 2018). Queda por confirmar si las Dcr 2/3 más cortos son funcionales o representan pseudogenes. La detección de proteínas Dcr-2 en *G. applanata* y *M. fusca* con estructuras muy similares a las de los trematodos, puede indicar que la reorganización genómica detectada en los trematodos pudo haber ocurrido más de una vez durante la evolución de los platelmintos. Sin embargo, dada la fragmentación de los transcriptomas de 'turbelarios', se necesitan más datos genómicos para verificar estas observaciones.

Vasa es parte de la biogénesis de los piRNA y se encuentra diferencialmente conservada entre 'turbelarios' y neodermatas (Skinner et al., 2014). Al igual que la ausencia de Piwi, la ausencia de Vasa en Neodermatas resulta desconcertante. Piwi y Vasa son ampliamente conocidos entre los biólogos del desarrollo como marcadores de la línea germinal. Incluso se encontró que Vasa se expresa en la línea germinal de metazoos de ramificación temprana como el clado ctenófora, lo que sugiere un papel central en el desarrollo de todos los metazoos (Alié et al., 2011; Mochizuki et al., 2001; Rebscher et al., 2007). Se ha propuesto que los genes Vasa surgieron por duplicación de un gen ancestral relacionado con PL-10 antes de la aparición en las esponjas, pero después de la separación de hongos y plantas (Mochizuki et al., 2001). Nuestros datos sugieren que una segunda duplicación de PL-10 tuvo lugar en un ancestro temprano de los platelmintos que produjo una familia específica de genes relacionados con PL-10 en platelmintos (Figura 7 del manuscrito presentado en Resultados B). Todavía se desconoce por qué la pérdida de Vasa fue favorecida evolutivamente en los neodermatas; sin embargo, la expresión en la línea germinal y el papel en la gametogénesis de PL-10 específico de platelmintos ha sido reportado en los neodermatas (Ohashi et al., 2007). De igual forma, se ha reportado que los genes vasa-like, es decir, PL-10, se expresan en células de ovario, testículo y huevo de *S. mansoni* y *S. japonicum* (S. He et al., 2018; Skinner et al., 2012, 2020).

En el mismo sentido, comprobamos que otros genes de la vía Piwi también están ausentes en Neodermatas pero conservados en todos los 'turbelarios'. La ausencia de proteínas de la vía piwi en especies parásitas parece ser una tendencia constante. Asimismo, se ha reportado la ausencia de Piwi en todos los clados de nematodos, excepto en el clado V, que contiene la especie modelo *C. elegans*, y algunos parásitos animales como *Haemonchus contortus* y *Pristionchus pacificus* (Sarkies et al., 2015). Otras proteínas de la vía piwi también están ausentes en nematodos que no pertenece al clado V, lo que confirma la ausencia de una vía funcional. Además, la vía piwi también se encontró ausente en los genomas de los ácaros del polvo y la sarna (Mondal et al., 2018). Tanto en los trematodos como en los nematodos y en los ácaros, existe evidencia de que mecanismos alternativos relacionados con el siARN están involucrados en el control de los TEs (P. Cai et al., 2012; Protasio et al., 2020). La amplificación de genes asociados con las principales vías de ARNpq en platelmintos parásitos sugiere una

adaptación similar de funciones. Dado que se reconoce que los TEs contribuyen a la innovación genómica, así como a la inestabilidad genómica en una amplia variedad de especies (Klein & O'Neill, 2018), es tentador especular que la pérdida de la vía Piwi podría estar asociada con una rápida reorganización genómica que ha llevado a adoptar una forma de vida parasitaria.

Conclusiones y Perspectivas

Hemos analizado los perfiles de expresión del complemento de ARNs pequeños en tres estadios intra-mamífero del ciclo de vida y en las vesículas extracelulares del estadio adulto de *F. hepatica*. Detectamos perfiles de miARNs que varían significativamente su expresión y, a casi 30 años de la primera descripción del rol de miARNs en la regulación del desarrollo de *C. elegans*, realizamos la primera descripción *in-silico* de la función putativa de los miARNs en los estadios invasivos de *F. hepatica*. A su vez, hallamos que los miARNs pueden ser relevantes no solo en la regulación de procesos vinculados al desarrollo, sino que también pueden estar regulando mecanismos directamente vinculados a la invasión. En el mismo sentido, describimos las vías centrales de señalización y las vinculadas a la respuesta inmune del huésped que podrían estar siendo reguladas por los miARNs parasitarios contenidos en EVs. Este puntapié inicial será de gran valor en ensayos experimentales que pretendan abordar el estudio de miARNs clave tanto para el desarrollo o el proceso de invasión, como en la regulación de blancos génicos en el huésped.

Por otro lado, trabajos recientes han propuesto el potencial uso de los miARNs circulantes como biomarcadores no invasivos en sangre. A su vez, se han detectado miRNAs libres circulantes en la sangre de huéspedes infectados con varias especies de nematodos y platelmintos (Mu et al., 2021). Es posible que los miARN más abundantes en EVs sean detectables en la circulación de animales infectados con *F. hepatica*. Si este es el caso, se podrían desarrollar nuevos mecanismos para el diagnóstico de la enfermedad.

Estudiamos las poblaciones de fragmentos de ARNt en todas las etapas analizadas, y describimos una nueva clase de fragmento de ARNt particularmente abundante en la fracción extravesicular. A su vez, inferimos por predicciones *in-silico* que los fragmentos de ARNt más abundantes de la fracción EV podrían estar formando estructuras homodiméricas estables, futuros ensayos posiblemente mediante tratamiento con exonucleasas podrán corroborar esta observación. Del mismo modo, será relevante corroborar si estas estructuras resistentes a la

degradación se encuentran en la fracción libre, no asociada a EVs, como ha sido descripto en humanos (Tosar et al., 2018).

Realizamos la primera descripción de vtARNs en Neodermatas, el método utilizado en nuestro trabajo podrá ser utilizada para caracterizaciones similares en otros platelmintos. Si bien las moléculas vault son aún poco conocidas pueden ser potencialmente relevante dado las múltiples funciones regulatorias en las que están involucradas en organismos modelo (Frascotti et al., 2021). A su vez dada su abundancia en EVs, será relevante estudiar el rol que pueden estar jugando en la interacción con el huésped.

Por otro lado, brindamos fuerte evidencia bioinformática sobre la presencia y ausencia de proteínas clave involucradas en las vías de ARNpq en platelmintos de vida libre de divergencia temprana, lo que sugiere que la regulación por miARNs, el silenciamiento mediado por piARN y ARNi son mecanismos regulatorios ancestrales en platelmintos. Además, las diferencias observadas en especies parasitarias que evolucionaron más recientemente sugieren fuertemente que los mecanismos mediados por ARN pequeños podrían haber sido también relevantes durante la transición al parasitismo.

La ausencia total de los mediadores de la vía piwi en todos los trematodos y cestodos sugiere una pérdida temprana en un ancestro de los neodermatas. Esta única pérdida podría haber tenido consecuencias evolutivas dramáticas, ya que los elementos transponibles podrían haber impulsado la inestabilidad del genoma que condujo a transformaciones bioquímicas, morfológicas y funcionales, por ejemplo, el origen de la neodermis (Caira & Littlewood, 2013), y cambios similares que favorecieron la adaptación a un nuevo estilo de vida. En esta línea altamente especulativa, es tentador pensar que la posterior amplificación independiente de FLAgos y Belle/PL10 en trematodos y cestodos podría haber resultado de una adaptación a estos cambios, ya sea para el control de los transposones y/o para generar nuevos mecanismos reguladores mediados por ARN pequeños no codificantes.

Se pueden considerar diferentes enfoques experimentales para probar esta hipótesis. Herramientas genómicas funcionales probadas en platelmintos como ARNi (Dell'Oca et al., 2014), transgénesis (Rinaldi et al., 2012; Suttiprapa et al., 2016), edición del genoma mediante

CRISPR-Cas9 (Ittiprasert et al., 2019; Lok et al., 2017; Sankaranarayanan et al., 2020) u ensayos de inmunoprecipitación (Free et al., 2009) podrían usarse para definir la función de los factores informados aquí y/o para detectar otros nuevos.

En todo caso, se necesitan más enfoques experimentales para comprender las funciones de todas las clases de ARN pequeños detectados; su presencia combinada en las EVs sugiere una acción concertada en la interacción y modulación de las respuestas del huésped, que merece ser investigada. A su vez, la presente descripción de las vías de ARNpq, junto con nuestro informe anterior (Fontenla et al., 2017), describen características novedosas de la biología y la evolución que son exclusivas de platelmintos, lo que implica que hay sutiles mecanismos involucrados en las vías de ARNpq de los platelmintos que son diferentes a los descritos en organismos modelo como mamíferos, *C. elegans* o *D. melanogaster*.

En conclusión, si bien queda mucho terreno por investigar nuestros resultados aportan nuevas perspectivas en campos aun insipientemente estudiados que podrían redundar en el diseño de nuevos mecanismos de control y tratamiento de enfermedades negligenciadas pero con un impacto enorme para el desarrollo humano.

Referencias

Agarwal, V., Bell, G. W., Nam, J. W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *ELife*, *4*(AUGUST2015), 1–38. https://doi.org/10.7554/eLife.05005

Alatoom, A., Cavuoti, D., Southern, P., and Gander, R. (2008). Fasciola hepatica Infection in the United States. *Lab. Med.* 39, 425–428. doi:10.1309/MLFFXA77UBBFH4GP.

Alberti, C., & Cochella, L. (2017). A framework for understanding the roles of miRNAs in animal development. *Development*, *144*(14), 2548–2559. https://doi.org/10.1242/dev.146613

Alexa, A., & Rahnenfuhrer, J. (2019). *topGO: Enrichment Analysis for Gene Ontology*.

Alié, A., Leclère, L., Jager, M., Dayraud, C., Chang, P., le Guyader, H., Quéinnec, E., & Manuel, M. (2011). Somatic stem cells express Piwi and Vasa genes in an adult ctenophore: Ancient association of "germline genes" with stemness. *Developmental Biology*, *350*(1), 183–197. https://doi.org/10.1016/j.ydbio.2010.10.019

Ancarola, M. E., Marcilla, A., Herz, M., Macchiaroli, N., Pérez, M., Asurmendi, S., Brehm, K., Poncini, C., Rosenzvit, M., & Cucher, M. (2017). Cestode parasites release extracellular vesicles with microRNAs and immunodiagnostic protein cargo. *International Journal for Parasitology*, *47*(10–11), 675–686. https://doi.org/10.1016/j.ijpara.2017.05.003

Arora, N., Tripathi, S., Singh, A. K., Mondal, P., Mishra, A., & Prasad, A. (2017). Micromanagement of immune system: Role of miRNAs in helminthic infections. *Frontiers in Microbiology*, *8*(APR), 1–13. https://doi.org/10.3389/fmicb.2017.00586

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, *37*(SUPPL. 2), 202–208. https://doi.org/10.1093/nar/gkp335

Bayraktar, R., van Roosbroeck, K., & Calin, G. A. (2017). Cell-to-cell communication: microRNAs as hormones. *Molecular Oncology*, *11*(12), 1673–1686. https://doi.org/10.1002/1878-0261.12144

Bartel, D. P. (2009). MicroRNA Target Recognition and Regulatory Functions. *Cell* 136, 215–233. doi:10.1016/j.cell.2009.01.002.MicroRNA.

Bellaousov, S., Reuter, J. S., Seetin, M. G., & Mathews, D. H. (2013). RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Research*, *41*(Web Server issue), W471-4. https://doi.org/10.1093/nar/gkt290

Bernal, D., Trelis, M., Montaner, S., Cantalapiedra, F., Galiano, A., Hackenberg, M., & Marcilla, A. (2014). Surface analysis of dicrocoelium dendriticum. The molecular characterization of exosomes reveals the presence of miRNAs. *Journal of Proteomics*, *105*, 232–241. https://doi.org/10.1016/j.jprot.2014.02.012

Bernstein, E., Caudy, A. A., Hammond, S. M., & Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, *409*. https://doi.org/10.1038/35053110

Betel, D., Koppal, A., Agius, P., Sander, C., & Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biology*, *11*(8), R90. https://doi.org/10.1186/gb-2010-11-8-r90

Breuer, K., Foroushani, A. K., Laird, M. R., Chen, C., Sribnaia, A., Lo, R., Winsor, G. L., Hancock, R. E. W., Brinkman, F. S. L., & Lynn, D. J. (2013). InnateDB: Systems biology of innate immunity and beyond - Recent updates and continuing curation. *Nucleic Acids Research*, *41*(D1), 1228–1233. https://doi.org/10.1093/nar/gks1147

Buck, A. H., Coakley, G., Simbari, F., McSorley, H. J., Quintana, J. F., le Bihan, T., Kumar, S., Abreu-Goodger, C., Lear, M., Harcus, Y., Ceroni, A., Babayan, S. A., Blaxter, M., Ivens, A., & Maizels, R. M. (2014). Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nature Communications*, *5*, 5488. https://doi.org/10.1038/ncomms6488

Bushati, N., & Cohen, S. M. (2007). microRNA functions. *Annual Review of Cell and Developmental Biology*, *23*, 175–205. https://doi.org/10.1146/annurev.cellbio.23.090506.123406

Cai, M., Ding, J., Li, Y., He, G., Yang, J., Liu, T., Guo, X., Yang, X., Wang, X., Cho, W. C., Fasihi Harandi, M., & Zheng, Y. (2021). Echinococcus multilocularis infection induces UBE2N suppression via exosomal emu-miR-4989. *Acta Tropica*, *223*(March), 106087. https://doi.org/10.1016/j.actatropica.2021.106087

Cai, P., Gobert, G. N., & McManus, D. P. (2016). MicroRNAs in Parasitic Helminthiases: Current Status and Future Perspectives. *Trends in Parasitology*, *32*(1), 71–86. https://doi.org/10.1016/j.pt.2015.09.003

Cai, P., Piao, X., Hou, N., Liu, S., Wang, H., & Chen, Q. (2012). Identification and Characterization of Argonaute Protein, Ago2 and Its Associated Small RNAs in Schistosoma japonicum. *PLoS Neglected Tropical Diseases*, *6*(7), e1745. https://doi.org/10.1371/journal.pntd.0001745

Caira, J. N., & Littlewood, D. T. J. (2013). Worms, Platyhelminthes. In S. A. B. T.-E. of B. (Second E. Levin (Ed.), *Encyclopedia of Biodiversity* (pp. 437–469). Elsevier. https://doi.org/10.1016/B978-0-12-384719-5.00166-0

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*(10), 1461–1475. https://doi.org/10.1016/j.biochi.2008.04.020

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. https://doi.org/10.1186/1471-2164-11-227

Canevari, J., Ceballos, L., Sanabria, R., Romero, J., Olaechea, F., Ortiz, P., et al. (2013). Testing albendazole resistance in Fasciola hepatica: validation of an egg hatch test with isolates from South America and the United Kingdom. 1–7. doi:10.1017/S0022149X13000163.

Cao, Z., Rosenkranz, D., Wu, S., Liu, H., Pang, Q., Zhang, X., Liu, B., & Zhao, B. (2020). Different classes of small RNAs are essential for head regeneration in the planarian Dugesia japonica. *BMC Genomics*, *21*(1), 1–11. https://doi.org/10.1186/s12864-020-07234-1

Carmona, C., & Tort, J. F. (2017). Fasciolosis in South America: epidemiology and control challenges. *Journal of Helminthology*, *91*(2), 99–109. https://doi.org/10.1017/S0022149X16000560

Chin, L. J., & Slack, F. J. (2008). A truth serum for cancer microRNAs have major potential as cancer biomarkers. *Cell Research*, *18*(10), 983–984. https://doi.org/10.1038/cr.2008.290

Choi, Y. J., Fontenla, S., Fischer, P. U., Le, T. H., Costábile, A., Blair, D., Brindley, P. J., Tort, J. F., Cabada, M. M., & Mitreva, M. (2020). Adaptive Radiation of the Flukes of the Family Fasciolidae Inferred from Genome-Wide Comparisons of Key Species. *Molecular Biology and Evolution*, *37*(1), 84–99. https://doi.org/10.1093/molbev/msz204

Choudhuri, S. (2009). Lesser known relatives of miRNA. *Biochemical and Biophysical Research Communications*, *388*(2), 177–180. https://doi.org/10.1016/j.bbrc.2009.08.039

Coakley, G., Buck, A. H., & Maizels, R. M. (2016). Host parasite communications--Messages from helminths for the immune system: Parasite communication and cell-cell

interactions. *Molecular and Biochemical Parasitology*, *208*(1), 33–40. https://doi.org/10.1016/j.molbiopara.2016.06.003

Coakley, G., Maizels, R. M., & Buck, A. H. (2015). Exosomes and Other Extracellular Vesicles: The New Communicators in Parasite Infections. *Trends in Parasitology*, *31*(10), 477–489. https://doi.org/10.1016/j.pt.2015.06.009

Collins, J. J. (2017). Platyhelminthes. In *Current Biology*. https://doi.org/10.1016/j.cub.2017.02.016

Cwiklinski, K., Jewhurst, H., McVeigh, P., Barbour, T., Maule, A. G., Tort, J., O'Neill, S. M., Robinson, M. W., Donnelly, S., & Dalton, J. P. (2018). Infection by the Helminth Parasite Fasciola hepatica Requires Rapid Regulation of Metabolic, Virulence, and Invasive Factors to Adjust to Its Mammalian Host. *Molecular & Cellular Proteomics : MCP*, *17*(4), 792–809. https://doi.org/10.1074/mcp.RA117.000445

da Costa, R. A., Corbellini, L. G., Castro-Janer, E., & Riet-Correa, F. (2019). Evaluation of losses in carcasses of cattle naturally infected with Fasciola hepatica: effects on weight by age range and on carcass quality parameters. *International Journal for Parasitology*, *49*(11), 867–872. https://doi.org/10.1016/j.ijpara.2019.06.005

Dalzell, J. J., McVeigh, P., Warnock, N. D., Mitreva, M., Bird, D. M., Abad, P., Fleming, C. C., Day, T. a, Mousley, A., Marks, N. J., & Maule, A. G. (2011). RNAi effector diversity in nematodes. *PLoS Neglected Tropical Diseases*, *5*(6), e1176. https://doi.org/10.1371/journal.pntd.0001176

Dell'Oca, N., Basika, T., Corvo, I., Castillo, E., Brindley, P. J., Rinaldi, G., & Tort, J. F. (2014). RNA interference in Fasciola hepatica newly excysted juveniles: Long dsRNA induces more persistent silencing than siRNA. *Molecular and Biochemical Parasitology*, *197*(1–2), 28–35. https://doi.org/10.1016/j.molbiopara.2014.10.001

Ding, J., He, G., Wu, J., Yang, J., Guo, X., Yang, X., Wang, Y., Kandil, O. M., Kutyrev, I., Ayaz, M., & Zheng, Y. (2019). miRNA-seq of Echinococcus multilocularis Extracellular Vesicles and Immunomodulatory Effects of miR-4989. *Frontiers in Microbiology*, *10*(November), 1–9. https://doi.org/10.3389/fmicb.2019.02707

Dunne, D. W., & Cooke, A. (2005). A worm's eye view of the immune system: Consequences for evolution of human autoimmune disease. *Nature Reviews Immunology*, *5*(5), 420–426. https://doi.org/10.1038/nri1601

Eichenberger, R. M., Ryan, S., Jones, L., Buitrago, G., Polster, R., Montes de Oca, M., Zuvelek, J., Giacomin, P. R., Dent, L. A., Engwerda, C. R., Field, M. A., Sotillo, J., & Loukas, A. (2018). Hookworm Secreted Extracellular Vesicles Interact With Host Cells and Prevent Inducible Colitis in Mice. *Frontiers in Immunology*, *9*(APR), 1–14. https://doi.org/10.3389/fimmu.2018.00850

Eichenberger, R. M., Talukder, M. H., Field, M. A., Wangchuk, P., Giacomin, P., Loukas, A., & Sotillo, J. (2018). Characterization of Trichuris muris secreted proteins and extracellular vesicles provides new insights into host–parasite communication. *Journal of Extracellular Vesicles*, *7*(1), 1428004. https://doi.org/10.1080/20013078.2018.1428004

Fairfax, K., Nascimento, M., Huang, S. C. C., Everts, B., & Pearce, E. J. (2012). Th2 responses in schistosomiasis. *Seminars in Immunopathology*, *34*(6), 863–871. https://doi.org/10.1007/s00281-012-0354-4

Fire, A., Albertson, D., Harrison, S. W., & Moerman, D. G. (1991). Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. *Development (Cambridge, England)*, *113*(2), 503–514.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, *391*(6669), 806–811. https://doi.org/10.1038/35888

Fischer, S. E. J. (2010). Small RNA-mediated gene silencing pathways in C. elegans. *The International Journal of Biochemistry & Cell Biology*, *42*(8), 1306–1315. https://doi.org/10.1016/j.biocel.2010.03.006

Fontenla, S., Dell'Oca, N., Smircich, P., Tort, J. F., & Siles-Lucas, M. (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), 901–913. https://doi.org/10.1016/j.ijpara.2015.06.007

Fontenla, S., Rinaldi, G., Smircich, P., & Tort, J. F. (2017). Conservation and diversification of small RNA pathways within flatworms. *BMC Evolutionary Biology*, *17*(1), 215. https://doi.org/10.1186/s12862-017-1061-5

Frascotti, G., Galbiati, E., Mazzucchelli, M., Pozzi, M., Salvioni, L., Vertemara, J., & Tortora, P. (2021). The vault nanoparticle: A gigantic ribonucleoprotein assembly involved in diverse physiological and pathological phenomena and an ideal nanovector for drug delivery and therapy. *Cancers*, *13*(4), 1–37. https://doi.org/10.3390/cancers13040707

Free, R. B., Hazelwood, L. A., & Sibley, D. R. (2009). Identifying Novel Protein-Protein Interactions Using Co-Immunoprecipitation and Mass Spectroscopy. *Current Protocols in Neuroscience / Editorial Board, Jacqueline N. Crawley ... [et Al.]*, *0 5*, Unit-5.28. https://doi.org/10.1002/0471142301.ns0528s46

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. *Proceedings*

of the National Academy of Sciences of the United States of America, *106*(28), 11546–11551. https://doi.org/10.1073/pnas.0905222106

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 Research*, *40*(1), 37–52. https://doi.org/10.1093/nar/gkr688

Fuchs Wightman, F., Giono, L. E., Fededa, J. P., and de la Mata, M. (2018). Target RNAs Strike Back on MicroRNAs. *Front. Genet.* 9. doi:10.3389/fgene.2018.00435.

Garcia, H. H., Moro, P. L., and Schantz, P. M. (2007). Zoonotic helminth infections of humans: echinococcosis, cysticercosis and fascioliasis. *Curr. Opin. Infect. Dis*. 20, 489–94. doi:10.1097/QCO.0b013e3282a95e39.

González-Estévez, C., Arseni, V., Thambyrajah, R. S., Felix, D. A., & Aboobaker, A. A. (2009). Diverse miRNA spatial expression patterns suggest important roles in homeostasis and regeneration in planarians. *International Journal of Developmental Biology*, *53*(4), 493–505. https://doi.org/10.1387/ijdb.082825cg

González-Miguel, J., Becerro-Recio, D., and Siles-Lucas, M. (2021). Insights into Fasciola hepatica Juveniles: Crossing the Fasciolosis Rubicon. *Trends Parasitol*. 37, 35–47. doi:10.1016/j.pt.2020.09.007.

Grishok, A. (2005). RNAi mechanisms in Caenorhabditis elegans. *FEBS Letters*, *579*(26), 5932–5939. https://doi.org/10.1016/j.febslet.2005.08.001

Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., & Mello, C. C. (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. https://doi.org/10.1016/S0092-8674(01)00431-7

Hanna, R. E. B., Moffett, D., Robinson, M. W., Jura, W. G. Z. O., Brennan, G. P., & Fairweather, I. (2019). Fasciola gigantica: Comparison of the tegumental ultrastructure in newly excysted metacercariae and in vitro penetrated juvenile flukes indicates intracellular sources of molecules with vaccinal and immunomodulatory potential. *Veterinary Parasitology*, *265*(October 2018), 38–47. https://doi.org/10.1016/j.vetpar.2018.11.019

Hansen, E. P., Fromm, B., Andersen, S. D., Marcilla, A., Andersen, K. L., Borup, A., Williams, A. R., Jex, A. R., Gasser, R. B., Young, N. D., Hall, R. S., Stensballe, A., Ovchinnikov, V., Yan, Y., Fredholm, M., Thamsborg, S. M., & Nejsum, P. (2019). Exploration of extracellular vesicles from Ascaris suum provides evidence of parasite–host cross talk. *Journal of Extracellular Vesicles*, *8*(1), 1578116. https://doi.org/10.1080/20013078.2019.1578116

He, S., Zhu, L., Liu, F., Liu, Q., Shao, Y., Hua, M., Ding, H., Shao, W., Du, Y., Hou, X., Ren, C., Liu, M., & Shen, J. (2018). Functions of the Vasa gene in Schistosoma japonicum as assessed by RNA interference. *Gene*, *638*, 13–19. https://doi.org/10.1016/j.gene.2017.09.054

He, X., Wang, Y., Fan, X., Lei, N., Tian, Y., Zhang, D., & Pan, W. (2020). A schistosome miRNA promotes host hepatic fibrosis by targeting transforming growth factor beta receptor III. *Journal of Hepatology*, *72*(3), 519–527. https://doi.org/10.1016/j.jhep.2019.10.029

Herron, C. M., O'Connor, A., Robb, E., McCammick, E., Hill, C., Marks, N. J., Robinson, M. W., Maule, A. G., & McVeigh, P. (2022). Developmental Regulation and Functional Prediction of microRNAs in an Expanded Fasciola hepatica miRNome. *Frontiers in Cellular and Infection Microbiology*, *12*(February), 1–13. https://doi.org/10.3389/fcimb.2022.811123

Hu, C., Li, Y., Pan, D., Wang, J., Zhu, L., Lin, Y., Zhu, S., & Pan, W. (2021). A Schistosoma japonicum MicroRNA Exerts Antitumor Effects Through Inhibition of Both Cell Migration and Angiogenesis by Targeting PGAM1. *Frontiers in Oncology*, *11*(June), 1–11. https://doi.org/10.3389/fonc.2021.652395

Hu, C., Zhu, S., Wang, J., Lin, Y., Ma, L., Zhu, L., Jiang, P., Li, Z., & Pan, W. (2019). Schistosoma japonicum MiRNA-7-5p Inhibits the Growth and Migration of Hepatoma Cells via Cross-Species Regulation of S-Phase Kinase-Associated Protein 2. *Frontiers in Oncology*, *9*(MAR), 1–12. https://doi.org/10.3389/fonc.2019.00175

Hu, R. S., Zhang, X. X., Ma, Q. N., Elsheikha, H. M., Ehsan, M., Zhao, Q., Fromm, B., & Zhu, X. Q. (2021). Differential expression of microRNAs and tRNA fragments mediate the adaptation of the liver fluke Fasciola gigantica to its intermediate snail and definitive mammalian hosts. *International Journal for Parasitology*, *51*(5), 405–414. https://doi.org/10.1016/j.ijpara.2020.10.009

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. *Nature Reviews. Molecular Cell Biology*, *9*(1), 22–32. https://doi.org/10.1038/nrm2321

Ilgová, J., Jedličková, L., Dvořáková, H., Benovics, M., Mikeš, L., Janda, L., Vorel, J., Roudnický, P., Potěšil, D., Zdráhal, Z., Gelnar, M., & Kašný, M. (2017). A novel type I cystatin of parasite origin with atypical legumain-binding domain. *Scientific Reports*, *7*(1), 17526. https://doi.org/10.1038/s41598-017-17598-2

Ittiprasert, W., Mann, V. H., Karinshak, S. E., Coghlan, A., Rinaldi, G., Sankaranarayanan, G., Chaidee, A., Tanno, T., Kumkhaek, C., Prangtaworn, P., Mentink-Kane, M. M., Cochran, C. J., Driguez, P., Holroyd, N., Tracey, A., Rodpai, R., Everts, B., Hokke, C. H., Hoffmann, K. F., … Brindley, P. J. (2019). Programmed genome editing of the omega-1 ribonuclease of the blood fluke, Schistosoma mansoni. *ELife*, *8*. https://doi.org/10.7554/eLife.41337

Jaronczyk, K., Carmichael, J. B., & Hobman, T. C. (2005). Exploring the functions of RNA interference pathway proteins: some functions are more RISCy than others? *Biochem J*, *387*(Pt 3), 561–571. https://doi.org/10.1042/BJ20041822

Johnstone, R. M., Adam, M., Hammond, J. R., Orr, L., & Turbide, C. (1987). Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). *The Journal of Biological Chemistry*, *262*(19), 9412–9420. http://www.ncbi.nlm.nih.gov/pubmed/3597417

Keiser, J., & Utzinger, J. (2005). Emerging foodborne trematodiasis. *Emerging Infectious Diseases*, *11*(10), 1507–1514. https://doi.org/10.3201/eid1110.050614

Kelley, J. M., Elliott, T. P., Beddoe, T., Anderson, G., Skuce, P., and Spithill, T. W. (2016). Current Threat of Triclabendazole Resistance in Fasciola hepatica. *Trends Parasitol.* 32, 458–469. doi:10.1016/j.pt.2016.03.002.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., & Segal, E. (2007). The role of site accessibility in microRNA target recognition. *Nature Genetics*, *39*(10), 1278–1284. https://doi.org/10.1038/ng2135

Ketting, R. F., Haverkamp, T. H. A., van Luenen, H. G. A. M., & Plasterk, R. H. A. (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), 133–141. https://doi.org/10.1016/S0092-8674(00)81645-1

King, C. H. (2019). Helminthiasis Epidemiology and Control: Scoring Successes and Meeting the Remaining Challenges. In *Advances in Parasitology* (Vol. 103). Elsevier Ltd. https://doi.org/10.1016/bs.apar.2018.08.001

Klein, S. J., & O'Neill, R. J. (2018). Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Research*, *26*(1–2), 5–23. https://doi.org/10.1007/s10577-017-9569-5

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Laumer, C. E., Hejnol, A., & Giribet, G. (2015). Nuclear genomic signals of the "microturbellarian" roots of platyhelminth evolutionary innovation. *ELife*, *4*, 1–31. https://doi.org/10.7554/eLife.05503

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, *75*(5), 843–854. https://doi.org/10.1016/0092-8674(93)90529-Y

Lin, Y., Zhu, S., Hu, C., Wang, J., Jiang, P., Zhu, L., Li, Z., Wang, S., Zhang, Y., Xu, X., & Pan, W. (2019). Cross-Species Suppression of Hepatoma Cell Growth and Migration by a Schistosoma japonicum MicroRNA. *Molecular Therapy - Nucleic Acids*, *18*(December), 400–412. https://doi.org/10.1016/j.omtn.2019.09.006

Liu, H., Song, Q., Zhen, H., Deng, H., Zhao, B., & Cao, Z. (2021). miR-8b is involved in brain and eye regeneration of Dugesia japonica in head regeneration. *Biology Open*, *10*(6), 8–13. https://doi.org/10.1242/bio.058538

Liu, J., Zhu, L., Wang, J., Qiu, L., Chen, Y., Davis, R. E., & Cheng, G. (2019). Schistosoma japonicum extracellular vesicle mirna cargo regulates host macrophage functions facilitating parasitism. *PLoS Pathogens*, *15*(6), 1–27. https://doi.org/10.1371/journal.ppat.1007817

LOK, J. B., SHAO, H., MASSEY, H. C., & LI, X. (2017). Transgenesis in Strongyloides and related parasitic nematodes: historical perspectives, current functional genomic applications and progress towards gene disruption and editing. *Parasitology*, *144*(3), 327–342. https://doi.org/DOI: 10.1017/S0031182016000391

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, *6*(1), 26. https://doi.org/10.1186/1748-7188-6-26

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*. 1483–1491. https://doi.org/10.1261/rna.1702009.The

Mabbott, N. A. (2018). The Influence of Parasite Infections on Host Immunity to Co-infection With Other Pathogens. *Frontiers in Immunology*, *9*(November), 2579. https://doi.org/10.3389/fimmu.2018.02579

MacRae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z., Adams, P. D., & Doudna, J. A. (2006). Structural Basis for Double-Stranded RNA Processing by Dicer. *Science*, *311*(5758), 195–198. http://www.jstor.org.proxy.timbo.org.uy:443/stable/3843253

Maizels, R. M., Smits, H. H., & McSorley, H. J. (2018). Modulation of Host Immunity by Helminths: The Expanding Repertoire of Parasite Effector Molecules. *Immunity*, *49*(5), 801–818. https://doi.org/10.1016/j.immuni.2018.10.016

Marcilla, A., Trelis, M., Cortés, A., Sotillo, J., Cantalapiedra, F., Minguez, M. T., Valero, M. L., Sánchez del Pino, M. M., Muñoz-Antoli, C., Toledo, R., & Bernal, D. (2012). Extracellular Vesicles from Parasitic Helminths Contain Specific Excretory/Secretory Proteins and Are Internalized in Intestinal Host Cells. *PLoS ONE*, *7*(9). https://doi.org/10.1371/journal.pone.0045974

Mas-Coma, S., Bargues, M. D., & Valero, M. a. (2005). Fascioliasis and other plant-borne trematode zoonoses. *International Journal for Parasitology*, *35*(11–12), 1255–1278. https://doi.org/10.1016/j.ijpara.2005.07.010

McNulty, S. N., Tort, J. F., Rinaldi, G., Fischer, K., Rosa, B. A., Smircich, P., Fontenla, S., Choi, Y.-J., Tyagi, R., Hallsworth-Pepin, K., Mann, V. H., Kammili, L., Latham, P. S., Dell'Oca, N., Dominguez, F., Carmona, C., Fischer, P. U., Brindley, P. J., & Mitreva, M. (2017). Genomes of Fasciola hepatica from the Americas Reveal Colonization with Neorickettsia Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers. *PLoS Genetics*, *13*(1), e1006537. https://doi.org/10.1371/journal.pgen.1006537

Meningher, T., Barsheshet, Y., Ofir-Birin, Y., Gold, D., Brant, B., Dekel, E., Sidi, Y., Schwartz, E., Regev-Rudzki, N., Avni, O., & Avni, D. (2020). Schistosomal extracellular vesicle-enclosed miRNAs modulate host T helper cell differentiation. *EMBO Reports*, *21*(1), 1–17. https://doi.org/10.15252/embr.201947882

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., Peterson, A., Noteboom, J., O'Briant, K. C., Allen, A., Lin, D. W., Urban, N., Drescher, C. W., Knudsen, B. S., Stirewalt, D. L., Gentleman, R., Vessella, R. L., Nelson, P. S., Martin, D. B., & Tewari, M. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(30), 10513–10518. https://doi.org/10.1073/pnas.0804549105

Mochizuki, K., Nishimiya-Fujisawa, C., & Fujisawa, T. (2001). Universal occurrence of the vasa-related genes among metazoans and their germline expression in Hydra. *Development Genes and Evolution*, *211*(6), 299–308. https://doi.org/10.1007/s004270100156

Mondal, M., Klimov, P., & Flynt, A. S. (2018). Rewired RNAi-mediated genome surveillance in house dust mites. *PLOS Genetics*, *14*(1), e1007183. https://doi.org/10.1371/journal.pgen.1007183

Montgomery, T. A., Rim, Y.-S., Zhang, C., Dowen, R. H., Phillips, C. M., Fischer, S. E. J., & Ruvkun, G. (2012). PIWI Associated siRNAs and piRNAs Specifically Require the Caenorhabditis elegans HEN1 Ortholog henn-1. *PLoS Genetics*, *8*(4), e1002616. https://doi.org/10.1371/journal.pgen.1002616

Mosig, A., Sameith, K., & Stadler, P. (2006). Fragrep: An efficient search tool for fragmented patterns in genomic sequences. *Genomics, Proteomics and Bioinformatics*, *4*(1), 56–60. https://doi.org/10.1016/S1672-0229(06)60017-X

Mu, Y., McManus, D. P., Gordon, C. A., & Cai, P. (2021). Parasitic Helminth-Derived microRNAs and Extracellular Vesicle Cargos as Biomarkers for Helminthic Infections. *Frontiers in Cellular and Infection Microbiology*, *11*(June), 1–8. https://doi.org/10.3389/fcimb.2021.708952

Oey, H., Zakrzewski, M., Narain, K., Devi, K. R., Agatsuma, T., Nawaratna, S., Gobert, G. N., Jones, M. K., Ragan, M. A., McManus, D. P., & Krause, L. (2018). Whole-genome sequence of the oriental lung fluke Paragonimus westermani. *GigaScience*, *8*(1), 1–8. https://doi.org/10.1093/gigascience/giy146

Ohashi, H., Umeda, N., Hirazawa, N., Ozaki, Y., Miura, C., & Miura, T. (2007). Expression of vasa (vas)-related genes in germ cells and specific interference with gene functions by double-stranded RNA in the monogenean, Neobenedenia girellae. *International Journal for Parasitology*, *37*(5), 515–523. https://doi.org/10.1016/j.ijpara.2006.11.003

Ortiz-Quintero, B. (2016). Cell-free microRNAs in blood and other body fluids, as cancer biomarkers. *Cell Proliferation*, *49*(3), 281–303. https://doi.org/10.1111/cpr.12262

Ovchinnikov, V. Y., Kashina, E. v., Mordvinov, V. A., & Fromm, B. (2020). EV-transported microRNAs of Schistosoma mansoni and Fasciola hepatica: Potential targets in definitive hosts. *Infection, Genetics and Evolution*, *85*(June), 104528. https://doi.org/10.1016/j.meegid.2020.104528

Palakodeti, D., Smielewska, M., & Graveley, B. R. (2006). MicroRNAs from the Planarian Schmidtea mediterranea: a model system for stem cell biology. *RNA (New York, N.Y.)*, *12*(9), 1640–1649. https://doi.org/10.1261/rna.117206

Protasio, A. v, Rawlinson, K. A., Miska, E. A., Berriman, M., & Rinaldi, G. (2020). Evidence for transposable element control by Argonautes in a parasitic flatworm lacking the piRNA pathway. *BioRxiv*, 670372. https://doi.org/10.1101/670372

Rebscher, N., Zelada-González, F., Banisch, T. U., Raible, F., & Arendt, D. (2007). Vasa unveils a common origin of germ cells and of somatic stem cells from the posterior growth zone in the polychaete Platynereis dumerilii. *Developmental Biology*, *306*(2), 599–611. https://doi.org/10.1016/j.ydbio.2007.03.521

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., & Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature*, *403*. https://doi.org/10.1038/35002607

Ricafrente, A., Nguyen, H., Tran, N., & Donnelly, S. (2021). An Evaluation of the Fasciola hepatica miRnome Predicts a Targeted Regulation of Mammalian Innate Immune Responses. *Frontiers in Immunology*, *11*(January). https://doi.org/10.3389/fimmu.2020.608686

Ricciardi, A., Bennuru, S., Tariq, S., Kaur, S., Wu, W., Elkahloun, A. G., Arakelyan, A., Shaik, J., Dorward, D. W., Nutman, T. B., & Semnani, R. T. (2021). Extracellular vesicles released from the filarial parasite Brugia Malayi downregulate the host mTOR pathway. *PLoS Neglected Tropical Diseases*, *15*(1), 1–18. https://doi.org/10.1371/journal.pntd.0008884

Rinaldi, G., Eckert, S. E., Tsai, I. J., Suttiprapa, S., Kines, K. J., Tort, J. F., Mann, V. H., Turner, D. J., Berriman, M., & Brindley, P. J. (2012). Germline transgenesis and insertional mutagenesis in Schistosoma mansoni mediated by murine leukemia virus. *PLoS Pathogens*, *8*(7), 16. https://doi.org/10.1371/journal.ppat.1002820

Robinson, M. W., & Dalton, J. P. (2009). Zoonotic helminth infections with particular emphasis on fasciolosis and other trematodiases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1530), 2763–2776. https://doi.org/10.1098/rstb.2009.0089

Rosa, B. A., Choi, Y.-J., McNulty, S. N., Jung, H., Martin, J., Agatsuma, T., Sugiyama, H., Le, T. H., Doanh, P. N., Maleewong, W., Blair, D., Brindley, P. J., Fischer, P. U., & Mitreva, M. (2020). Comparative genomics and transcriptomics of 4 Paragonimus species provide insights into lung fluke parasitism and pathogenesis. *GigaScience*, *9*(7), 1–16. https://academic.oup.com/gigascience/article/doi/10.1093/gigascience/giaa073/5873693

Sánchez-López, C. M., Trelis, M., Jara, L., Cantalapiedra, F., Marcilla, A., & Bernal, D. (2020). Diversity of extracellular vesicles from different developmental stages of Fasciola hepatica. *International Journal for Parasitology*, *50*(9), 663–669. https://doi.org/10.1016/j.ijpara.2020.03.011

Sankaranarayanan, G., Coghlan, A., Driguez, P., Lotkowska, M. E., Sanders, M., Holroyd, N., Tracey, A., Berriman, M., & Rinaldi, G. (2020). Large CRISPR-Cas-induced deletions in the oxamniquine resistance locus of the human parasite Schistosoma mansoni. *Wellcome Open Research*, *5*, 178. https://doi.org/10.12688/wellcomeopenres.16031.1

Sarkies, P., Selkirk, M. E., Jones, J. T., Blok, V., Boothby, T., Goldstein, B., Hanelt, B., Ardila-Garcia, A., Fast, N. M., Schiffer, P. M., Kraus, C., Taylor, M. J., Koutsovoulos, G., Blaxter, M. L., & Miska, E. A. (2015). Ancient and Novel Small RNA Pathways Compensate for the Loss of piRNAs in Multiple Independent Nematode Lineages. *PLOS Biology*, *13*(2), e1002061. https://doi.org/10.1371/journal.pbio.1002061

Sasidharan, V., Lu, Y.-C., Bansal, D., Dasari, P., Poduval, D., Seshasayee, A., Resch, A. M., Graveley, B. R., & Palakodeti, D. (2013). Identification of neoblast- and regeneration-specific miRNAs in the planarian Schmidtea mediterranea. *RNA (New York, N.Y.)*, *19*(10), 1394–1404. https://doi.org/10.1261/rna.038653.113

Schmidt, M. F., Korb, O., & Abell, C. (2013). MicroRNA-specific argonaute 2 protein inhibitors. *ACS Chemical Biology*, *8*(10), 2122–2126. https://doi.org/10.1021/cb400246k

Skinner, D. E., Popratiloff, A., Alrefaei, Y. N., Mann, V. H., Rinaldi, G., & Brindley, P. J. (2020). Functional analysis of vasa/PL10-like genes in the ovary of Schistosoma mansoni. *Molecular and Biochemical Parasitology*, *236*, 111259. https://doi.org/10.1016/j.molbiopara.2020.111259

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 in Parasitology*, *30*(3), 123–129. https://doi.org/10.1016/j.pt.2014.01.001

Skinner, D. E., Rinaldi, G., Suttiprapa, S., Mann, V. H., Smircich, P., Cogswell, A. A., Williams, D. L., & Brindley, P. J. (2012). Vasa-like dead-box rna helicases of schistosoma mansoni. *PLoS Neglected Tropical Diseases*, *6*(6), 26–31. https://doi.org/10.1371/journal.pntd.0001686

Sotillo, J., Robinson, M. W., Kimber, M. J., Cucher, M., Ancarola, M. E., Nejsum, P., Marcilla, A., Eichenberger, R. M., & Tritten, L. (2020). The protein and microRNA cargo of extracellular vesicles from parasitic helminths – current status and research priorities. *International Journal for Parasitology*, *50*(9), 635–645. https://doi.org/10.1016/j.ijpara.2020.04.010

Spithill, T. W., Smooker, P. M., & Copeman, D. B. (1999). *"Fasciola gigantica": Epidemiology, control, immunology and molecular biology* (pp. 465-525 BT-Fasciolosis). CABI .

Suttiprapa, S., Rinaldi, G., Tsai, I. J., Mann, V. H., Dubrovsky, L., Yan, H. bin, Holroyd, N., Huckvale, T., Durrant, C., Protasio, A. v., Pushkarsky, T., Iordanskiy, S., Berriman, M., Bukrinsky, M. I., & Brindley, P. J. (2016). HIV-1 Integrates Widely throughout the Genome of the Human Blood Fluke Schistosoma mansoni. *PLoS Pathogens*, *12*(10), 1–26. https://doi.org/10.1371/journal.ppat.1005931

Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., & Mello, C. C. (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell*, *99*(2), 123–132. https://doi.org/10.1016/S0092-8674(00)81644-X

Taylor, P. J., Hagen, J., Faruqu, F. N., Al-Jamal, K. T., Quigley, B., Beeby, M., Selkirk, M. E., & Sarkies, P. (2020). Trichinella spiralis secretes abundant unencapsulated small RNAs

with potential effects on host gene expression. *International Journal for Parasitology*, *50*(9), 697–705. https://doi.org/10.1016/j.ijpara.2020.05.008

Thompson, D. M., & Parker, R. (2009). Stressing out over tRNA cleavage. *Cell*, *138*(2), 215–219. https://doi.org/10.1016/j.cell.2009.07.001

Tomari, Y., & Zamore, P. D. (2005). Perspective: Machines for RNAi. *Genes and Development*, *19*(5), 517–529. https://doi.org/10.1101/gad.1284105

Tosar, J. P., & Cayota, A. (2020). Extracellular tRNAs and tRNA-derived fragments. *RNA Biology*, *17*(8), 1149–1167. https://doi.org/10.1080/15476286.2020.1729584

Tosar, J. P., Gámbaro, F., Darré, L., Pantano, S., Westhof, E., & Cayota, A. (2018). Dimerization confers increased stability to nucleases in 5' halves from glycine and glutamic acid tRNAs. *Nucleic Acids Research*, *46*(17), 9081–9093. https://doi.org/10.1093/nar/gky495

Tran, N., Ricafrente, A., To, J., Lund, M., Marques, T. M., Gama-Carvalho, M., Cwiklinski, K., Dalton, J. P., & Donnelly, S. (2021). Fasciola hepatica hijacks host macrophage miRNA machinery to modulate early innate immune responses. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-86125-1

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J. J., & Lötvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature Cell Biology*, *9*(6), 654–659. 10.1038/ncb1596

van Niel, G., D'Angelo, G., & Raposo, G. (2018). Shedding light on the cell biology of extracellular vesicles. *Nature Reviews. Molecular Cell Biology*, *19*(4), 213–228. https://doi.org/10.1038/nrm.2017.125

Wang, L., Liao, Y., Yang, R., Yu, Z., Zhang, L., Zhu, Z., Wu, X., Shen, J., Liu, J., Xu, L., Wu, Z., & Sun, X. (2020). Sja-miR-71a in Schistosome egg-derived extracellular vesicles suppresses liver fibrosis caused by schistosomiasis via targeting semaphorin 4D. *Journal of Extracellular Vesicles*, *9*(1). https://doi.org/10.1080/20013078.2020.1785738

Wang, L., Liu, T., Chen, G., Li, Y., Zhang, S., Mao, L., Liang, P., Fasihi Harandi, M., Li, T., & Luo, X. (2021). Exosomal microRNA let-7-5p from Taenia pisiformis Cysticercus Prompted Macrophage to M2 Polarization through Inhibiting the Expression of C/EBP δ. *Microorganisms*, *9*(7), 1403. https://doi.org/10.3390/microorganisms9071403

Wang, L. Q., Liu, T. L., Liang, P. H., Zhang, S. H., Li, T. S., Li, Y. P., Liu, G. X., Mao, L., & Luo, X. N. (2020). Characterization of exosome-like vesicles derived from Taenia pisiformis cysticercus and their immunoregulatory role on macrophages. *Parasites and Vectors*, *13*(1), 1–16. https://doi.org/10.1186/s13071-020-04186-z

Wang, Y., Fan, X., Lei, N., He, X., Wang, X., Luo, X., Zhang, D., & Pan, W. (2020). A MicroRNA Derived From Schistosoma japonicum Promotes Schistosomiasis Hepatic Fibrosis by Targeting Host Secreted Frizzled-Related Protein 1. *Frontiers in Cellular and Infection Microbiology*, *10*(March), 101. https://doi.org/10.3389/fcimb.2020.00101

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*(12), e53387. https://doi.org/10.1371/journal.pone.0053387

Yan, C., Zhou, Q.-Y., Wu, J., Xu, N., Du, Y., Li, J., Liu, J.-X., Koda, S., Zhang, B.-B., Yu, Q., Yang, H.-M., Li, X.-Y., Zhang, B., Xu, Y.-H., Chen, J.-X., Wu, Z., Zhu, X.-Q., Tang, R.-X., & Zheng, K.-Y. (2021). Csi-let-7a-5p delivered by extracellular vesicles from a liver fluke activates M1-like macrophages and exacerbates biliary injuries. *Proceedings of the National Academy of Sciences*, *118*(46), e2102206118. https://doi.org/10.1073/pnas.2102206118

Yang, Y., Liu, L., Liu, X., Zhang, Y. Y., Shi, H., Jia, W., Zhu, H. F., Jia, H., Liu, M., & Bai, X. (2020). Extracellular Vesicles Derived From Trichinella spiralis Muscle Larvae Ameliorate TNBS-Induced Colitis in Mice. *Frontiers in Immunology*, *11*(June), 1–16. https://doi.org/10.3389/fimmu.2020.01174

Zamanian, M., Fraser, L. M., Agbedanu, P. N., Harischandra, H., Moorhead, A. R., Day, T. A., Bartholomay, L. C., & Kimber, M. J. (2015). Release of Small RNA-containing Exosome-like Vesicles from the Human Filarial Parasite Brugia malayi. *PLOS Neglected Tropical Diseases*, *9*(9), e0004069. https://doi.org/10.1371/journal.pntd.0004069

Zarowiecki, M., & Berriman, M. (2015). What helminth genomes have taught us about parasite evolution. *Parasitology*, *142 Suppl*(S1), S85-97. https://doi.org/10.1017/S0031182014001449

Zheng, Y. (2012). Phylogenetic analysis of the Argonaute protein family in platyhelminths. *Molecular Phylogenetics and Evolution*, *66*(3), 1050–1054. https://doi.org/10.1016/j.ympev.2012.11.014

Zheng, Y., Guo, X., He, W., Shao, Z., Zhang, X., Yang, J., Shen, Y., Luo, X., & Cao, J. (2016). Effects of Echinococcus multilocularis miR-71 mimics on murine macrophage RAW264.7 cells. *International Immunopharmacology*, *34*, 259–262. https://doi.org/10.1016/j.intimp.2016.03.015

Zhu, J., & Paul, W. E. (2008). CD4 T cells: fates, functions, and faults. *Blood*, *112*(5), 1557–1569. https://doi.org/10.1182/blood-2008-05-078154

Zhu, L., Liu, J., Dao, J., Lu, K., Li, H., Gu, H., Liu, J., Feng, X., & Cheng, G. (2016). Molecular characterization of S. japonicum exosome-like vesicles reveals their regulatory roles in

parasite-host interactions. *Scientific Reports*, *6*(April), 25885. https://doi.org/10.1038/srep25885

ANEXO I

Parte de los resultados obtenidos durante la maestría en bioinformática fueron publicados en el periodo del doctorado, y se incluyen aquí como anexo.

- Conservation and Diversification of Small RNA Pathways within Flatworms.

  Fontenla, Santiago, Gabriel Rinaldi, Pablo Smircich, and Jose F. Tort.

  BMC Evolutionary Biology 17 (2017), 215.

Dado el escaso conocimiento que había sobre los factores involucrados en las vías de ARNpq en platelmintos al momento de redactar este artículo, nos propusimos analizar su conservación en 24 especies de platelmintos con ensamblaje genómico disponible, la mayoría de ellas especies parásitas. Para ello empleamos factores de las vías de ARNpq que han sido bien caracterizados en la especie modelo *C. elegans,* para realizar búsquedas bioinformáticas en las bases de datos de platelmintos.

Entre nuestros resultados hallamos una familia de Argonauta especifica de platelmintos (que denominamos FL-Agos), que presentó amplificaciones independientes en los diferentes linajes y la pérdida de la familia Ago Piwi en todos los neodermata analizados. Detectamos una duplicación de Dicer en todos los platelmintos, con diferencias estructurales de Dicer-2 entre trematodos, cestodos y platelmintos de vida libre. A su vez, describimos por primera vez, la presencia de homólogos de un factor clave del complejo miRISC, GW182 en platelmintos.

Creemos que estos hallazgos contribuyeron a reflejar innovaciones específicas de la evolución de los platelmintos presumiblemente asociadas a mecanismos noveles de regulación de la expresión génica mediada por vías de ARNpq, que difieren con lo que ha sido descripto clásicamente en organismos modelo.

Fontenla et al. BMC Evolutionary Biology (2017) 17:215
DOI 10.1186/s12862-017-1061-5

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Conservation and diversification of small RNA pathways within flatworms

Santiago Fontenla[1], Gabriel Rinaldi[2], Pablo Smircich[1,3] and Jose F. Tort[1*]

Abstract

Background: Small non-coding RNAs, including miRNAs, and gene silencing mediated by RNA interference have been described in free-living and parasitic lineages of flatworms, but only few key factors of the small RNA pathways have been exhaustively investigated in a limited number of species. The availability of flatworm draft genomes and predicted proteomes allowed us to perform an extended survey of the genes involved in small non-coding RNA pathways in this phylum.

Results: Overall, findings show that the small non-coding RNA pathways are conserved in all the analyzed flatworm linages; however notable peculiarities were identified. While Piwi genes are amplified in free-living worms they are completely absent in all parasitic species. Remarkably all flatworms share a specific Argonaute family (FL-Ago) that has been independently amplified in different lineages. Other key factors such as Dicer are also duplicated, with Dicer-2 showing structural differences between trematodes, cestodes and free-living flatworms. Similarly, a very divergent GW182 Argonaute interacting protein was identified in all flatworm linages. Contrasting to this, genes involved in the amplification of the RNAi interfering signal were detected only in the ancestral free living species *Macrostomum lignano*. We here described all the putative small RNA pathways present in both free living and parasitic flatworm lineages.

Conclusion: These findings highlight innovations specifically evolved in platyhelminths presumably associated with novel mechanisms of gene expression regulation mediated by small RNA pathways that differ to what has been classically described in model organisms. Understanding these phylum-specific innovations and the differences between free living and parasitic species might provide clues to adaptations to parasitism, and would be relevant for gene-silencing technology development for parasitic flatworms that infect hundreds of million people worldwide.

Keywords: Flatworms, Small RNA pathways, miRNA, RNAi, Dicer, Argonaute

Background

RNA interference (RNAi) is a reverse genetic tool that triggers post-transcriptional knock-down of a target mRNA by the introduction of complementary double-stranded RNA (dsRNA) molecules. It is almost the only available tool to study a gene function in helminths [1]. In platyhelminthes, while this technique has been routinely established in the study of gene function in planarians [2], is still poorly developed in parasitic species. In parasitic flatworms, RNAi techniques have been optimized in schistosome species [3], and it has been proven to be functional in other few trematode species like *Fasciola hepatica* [4–7], *Opisthorchis viverrini* [8], *Clonorchis sinensis* [9], *O. felineus* [10] and in the Cestodes *Moniezia expansa* [11], *Hymenolepis microstoma* [12], *Echinococcus granulosus* [13], *E. multilocularis* [14] and *Taenia crassiceps* [15]. While this suggest that the mechanism is fully functional, several caveats exist even in the more studied species, since inefficient or inconsistent transcript knockdown have been reported, highlighted by variable levels of gene silencing between parasite developmental stages and the unsuccessful silencing of target genes in *S. mansoni* [16–19] and *F. hepatica* [6]. Is still not clear if these variations could be related to difference in the accessibility of interfering molecules among developmental stages and/or variations in the RNAi silencing pathway trough the development. Therefore, measuring the expression levels of factors involved in RNAi in different stages

* Correspondence: jtort@fmed.edu.uy
[1]Departamento de Genética, Facultad de Medicina, Universidad de la República (UDELAR), Gral. Flores 2125, CP11800 Montevideo, MVD, Uruguay
Full list of author information is available at the end of the article

would be relevant for the optimization of this functional approach. However, the lack of information regarding the genes involved in the RNAi pathway has been a limiting obstacle for this goal.

While the information on RNAi pathways in flatworms is fragmentary, several reports have detected the presence of microRNAs (miRNAs) in almost all the flatworm lineages [20–33]. MicroRNAs are relevant post-transcriptional regulators of gene expression conserved throughout all metazoa and plants [34]. Regulation mediated by miRNAs has been described in diverse biological processes including metabolism, cell development and differentiation, DNA methylation, chromatin modifications, neurological development, immune response, defense against viral infections, and cancer [35]. However, recent reports have shown a reduced miRNA complement in flatworms with significant sequence divergence in conserved families [23, 26], highlighting peculiarities of flatworm small RNA pathways.

The piRNA pathway is a small RNA mediated mechanism involved in the silencing of transposable elements in germline cells and hence, maintaining the genome stability [36]. While canonical 30–32 nt small RNAs typical of Piwi associated were detected in early studies in planarias [37–39], and later in *M. lignano* [22], they haven't been found in the major parasitic flatworms. Consistent with this absence early studies seeking the PIWI-like protein in the model trematodes failed, leading to the question of how these organisms maintain their genome stability throughout generations [40].

Despite increasing evidence supporting the conservation of functional small RNA pathways in flatworms, they are still poorly described in most of the platyhelminth species. Reports comparing small RNA pathways among flatworms generally focus on the two main proteins, Dicer and Argonaute [40–42], while more comprehensive surveys of factors are limited to the analyses of one species at a time [22, 31, 43]. Given the scarce knowledge of the factors involved in small RNA pathways in platyhelminthes, we employed well-characterized biogenesis pathways of small non-coding RNAs in the model organism *Caenorhabditis elegans* to conduct a bioinformatic search in flatworms with available transcriptomic/genomic data.

Results and discussion

The availability of 25 flatworm genome sequences, ranging from free living to parasitic monogenan, digenean and cestode lineages (Table 1) offer a good opportunity to compare and complete the still fragmentary knowledge of the small non-coding RNA pathways in platyhelminthes. Since the best described small non-coding RNA pathways are from *C. elegans*, we used a curated complete set of these factors as queries for comparisons (complete list in Additional file 1: Table S1).

We observed that while the core of the pathways are generally conserved, several factors are absent in flatworms and/or have diverged to a degree that are unrecognizable on the primary sequence level, as described in the following sections. Is worth to mention here that genome annotation quality varies between different flatworm species, and for that reason we did not rely only on predicted proteomes for the study, since this might lead to misreporting unannotated or fragmented models as gene losses or gains. To avoid this issue we also performed full genomic searches and corrected gene annotations based on these results, extending partial genes or fully annotating putative novel ones (see Methods).

Core ribonuclease III factors are differentially distributed in flatworms

The cleavage of long double stranded RNAs (dsRNA) by ribonuclease III enzymes is a key step in the biogenesis of small non-coding RNAs. Dicer (Dcr) belongs to this group of ribonucleases and process dsRNA into 22-23 nt RNAs with two-nucleotide 3′ overhangs that are recognized by Argonaute proteins [44]. Similarly, the nuclear protein Drosha (Drsh-1), involved in the generation of miRNA precursors, belongs to this group of ribonucleases. A complete conservation of *drsh-1* gene in all analyzed flatworms was observed, consistent with its central role in miRNA biogenesis (Fig. 1). While a single *drsh-1* gene has been described, two subfamilies of Dicer has been identified; a canonical Dcr-1 conserved in all metazoans, and a second subfamily, Dcr-2 present only in some species of invertebrates [41]. In *C. elegans*, there is an orthologue of Dcr-1, while in *D. melanogaster* there are orthologues of both subfamilies [41]. Expectedly, we found Dcr-1 conserved in all the flatworms analyzed, but also detected orthologues of Dcr-2 in all flatworms (Fig. 1). Interestingly, two copies of *dcr-2* genes were identified in all trematodes with the exception of the blood flukes (Additional file 1: Table S2). Dcr-2 and the second copy named Dcr-3, probably originated by an inverted duplication in a common ancestor of Fasciolidae and Opistorchidae (Additional file 2: Figure S1). To gain further insight into the organization of this family we analyzed the presence of functional domains. Flatworm Drsh-1 proteins showed two putative Ribonuclease III domains, a dsRNA-binding domain and N-termini rich in proline, serine and arginine residues (Fig. 1). On the other hand, variability in the organization of functional domains of Dcr-1 and Dcr-2 subfamilies has been reported [41]. The canonical Dcr-1 protein is characterized by an N-terminal DEAD/H-box helicase, a PAZ domain, two Ribonuclease III domains, and dsRNA-binding domain at the Carboxyl-end [44]. The predicted flatworm Dcr-1 proteins generally lack most amino terminal helicase domains conserving the remaining structure. While

Table 1 Analyzed species

CLASS	ORDER	FAMILY	SPECIES	GENOME SIZE	GENE COUNT	BIOPROJECT ID
RHABDITOPHORA	Macrostomida	Macrostomidae	*Macrostomum lignano* (Mli)	1040	60,534	*PRJNA284736*
	Tricladida	Dugesidae	*Schmidtea mediterranea* (Sme)	900	29,850	*PRJNA12585*
MONOGENEA	Monopisthocotylea	Gyrodactylidae	*Gyrodactylus salaris* (Gsa)	67	15,436	*PRJNA244375*
	Polyopisthocotylea	Polystomatidae	*Protopolystoma xenopodis* (Pxe)	617	37,906	*PRJEB1201*
TREMATODA	Opistorchida	Opistorchiidae	*Clonorchis sinensis* (Csi)	547	13,634	*PRJDA72781*
			Opistorchis viverrini (Ovi)	634	16,356	*PRJNA222628*
	Echinostomida	Echinostomatidae	*Echinostoma caproni* (Eca)	834	18,607	*PRJEB1207*
		Fasciolidae	*Fasciola hepatica* (FHO)	1139	15,739	*PRJNA179522*
			Fasciola hepatica (FHL)	1275	22,676	*PRJEB6687*
	Strigeidida	Schistosomatidae	*Trichobilharzia regenti* (Tre)	702	22,185	*PRJEB4662*
			Schistosoma haematobium (Sha)	385	13,073	*PRJNA78265*
			Schistosoma japonicum (Sja)	398	12,738	*PRJEA34885*
			Schistosoma mansoni (Sma)	362	10,772	*PRJEA36577*
CESTODA	Cyclophyllidea	Hymenolepididae	*Hymenolepis diminuta* (Hdi)	166	11,271	*PRJEB507*
			Hymenolepis microstoma (Hmi)	141	12,368	*PRJEB124*
			Hymenolepis nana (Hna)	163	13,777	*PRJEB508*
		Taeniidae	*Echinococcus granulosus* (Egr)	114	10,245	*PRJEB121*
			Echinococcus multilocularis (Emu)	114	10,663	*PRJEB122*
			Taenia asiatica (Tas)	136	10,331	*PRJEB532*
			Taenia solium (Tso)	122	12,481	*PRJNA170813*
			Hydatigera taeniaeformis (Hta)	104	11,649	*PRJEB534*
		Mesocestoididae	*Mesocestoides corti* (Mco)	117	10,614	*PRJEB510*
	Diphyllobotridea	Diphyllobothriidae	*Diphyllobothrium latum* (Dla)	531	19,966	*PRJEB1206*
			Schistocephalus solidus (Sso)	539	20,228	*PRJEB527*
			Spirometra erinaceieuropaei (Ser)	1259	39,557	*PRJEB1202*

Species abbreviation are indicated between brackets. Genome sizes (in Mb)

Dcr-2 proteins of insects present the conserved domain, this is drastically reduced in flatworms. In cestodes both the PAZ domain and the C-terminal dsRNA binding domain are absent. An even shorter Dcr-2 is present in trematodes, with only two RNAse III domains. A PAZ domain was also identified in schistosomes and in *C. sinensis*, suggesting the presence of highly divergent domains in the remaining trematodes. The presence of divergent Dcr-2 in flatworms, and Dcr-3 in Fasciolidae and Opistorchidae may indicate functional redundancy or different roles for both proteins. Available transcriptomic data from *S. mansoni* and *F. hepatica* shows that all these variants are expressed among different developmental stages; particularly, *F. hepatica dcr-3* is predominantly expressed in eggs (Additional file 2: Figure S1).

In *D. melanogaster* Dcr-1 and Dcr-2 proteins have evolved different roles, Dcr-1 processes pre-miRNA, i.e. miRNA precursor molecules, while Dcr-2 processes dsRNA in the RNAi pathway [45]. Dcr-1 would preferentially bind and process non-complementary stretches of dsRNA, characteristic of miRNAs, while Dcr-2 requires perfect complementarity between the dsRNA and target mRNA [46]. Due to the conservation of both subfamilies is tempting to hypothesize that a similar phenomenon may be occurring in flatworms, however this hypothesis awaits to be demonstrated.

Parasitic flatworms have two groups of Argonaute, while planarians have three

Argonaute proteins are essential for gene regulatory mechanisms mediated by small RNAs. The Argonaute (Ago) genes have been classified into four paralogous groups: (1) Argonaute-like (Ago-like) proteins, similar to *Arabidopsis thaliana* Ago1; (2) Piwi-like proteins, closely related to *D. melanogaster* PIWI; (3) a specific group identified in *C. elegans* called group 3 [47, 48], reported to be conserved in nematodes [49]; and (4) a cluster of proteins flatworm-specific recently identified and named cluster 1 by Zheng [42] and cluster 4 by Skinner [40].

In agreement with previous reports, our exhaustive homology search among diverse flatworms showed multiple hits in all species ranging from 3 in cestodes to

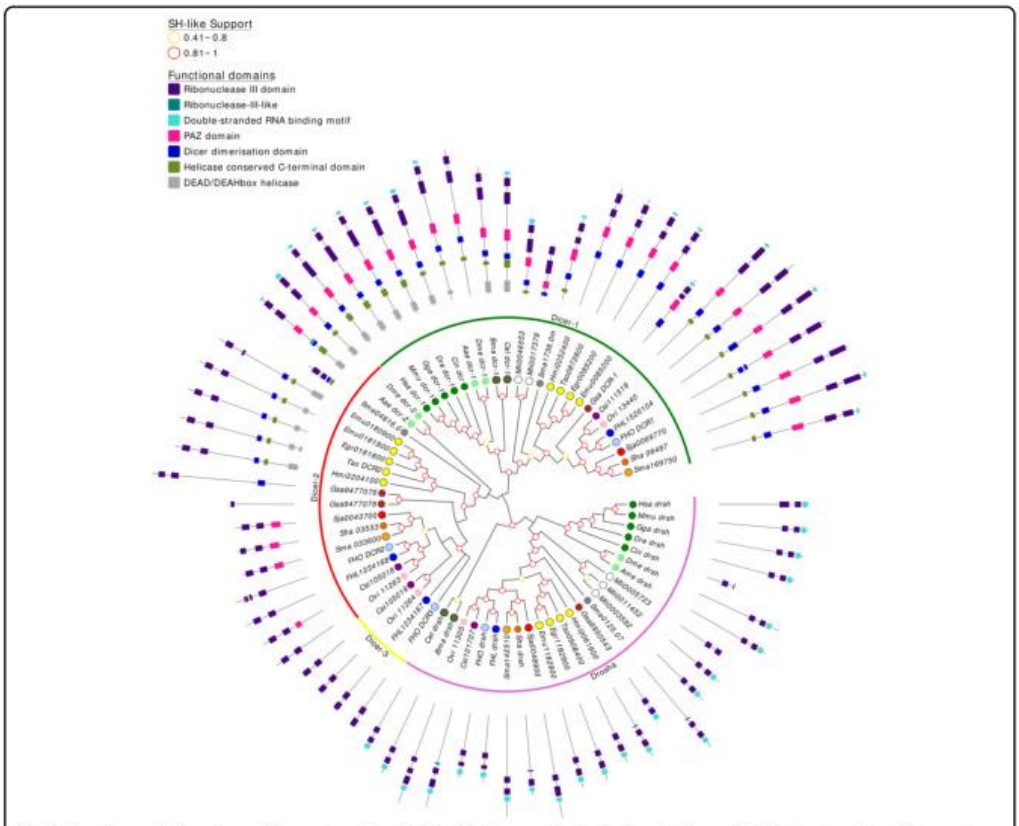


Fig. 1 A maximum likelihood tree of Ribonuclease III subfamilies. SH-like approximate likelihood ratios are indicated, values below 0.4 are not shown. Conserved protein domains were predicted with HMMScan. Dcr-2 (red arch) has missing domains in trematodes compared to cestodes, monogeneans and planarians. Dicer-2 are duplicated in Fasciolidae and Opisthorchiidae but not in blood flukes, the genes product of duplication were named Dicer-3 (yellow arch). Species abbreviation are indicated in Table 1

more than a dozen in free-living species. A more detailed phylogenetic analysis aided by structure determination allowed us to further classify these hits.

We found single orthologues of the canonical Ago subfamiliy in most parasitic flatworms, and an expansion in free living species. While *S. mediterranea* showed tree paralogues, in *M. lignano* thirteen paralogous sequences were detected (Fig. 2, Additional file 1: Table S3); this massive amplification seem consistent with the suspected tetraploidy of *M. lignano* [50].

Besides this, the monogenan *G. salaris* showed two inverted paralogous genes adjacent on the same contig suggest a recent duplication. Also in *F. hepatica* two paralogues were detected by genomic scanning, one of them unannotated in both genome assemblies available (Fig. 2, Additional file 1: Table S3). However, this novel unannotated gene has an unusual gene structure and is poorly expressed, suggesting that it might correspond to a pseudogene.

Interestingly, the domain architecture of these Ago-like proteins is very conserved from vertebrate to flatworms, i.e. an N-terminal domain followed by a linker 1 domain, the PAZ domain, which is important for the small RNA association along with the PIWI domain [48], a linker 2 domain, followed by the mid domain and PIWI, the catalytic domain (Fig. 2).

The PIWI-like proteins are core factors of the piRNA pathway involved in transposable elements (TE) silencing, and hence, maintaining the genome stability in germline cells [36]. This group was first reported in *S. mediterranea* [37, 38] and *Dugesia japonica* [39], and more recently in *M. lignano* [22]. According to these

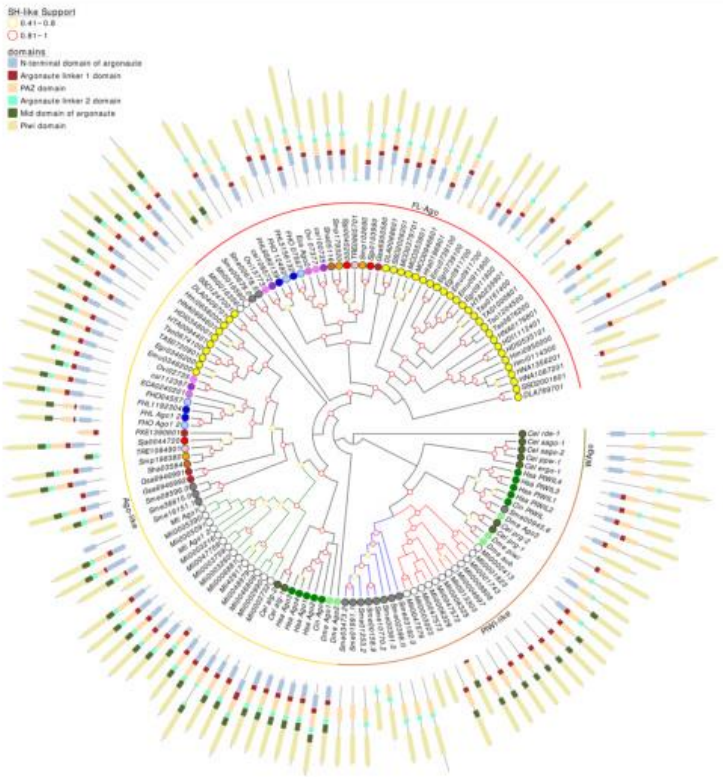


Fig. 2 A maximum likelihood tree of Argonautes. SH-like approximate likelihood ratios are indicated, values below 0.4 are not shown. Conserved protein domains were predicted with HMMScan. Green branches indicate the expansion of canonical Ago-like (yellow arch) in *M. lignano*. Blue branches and red branches indicate of PIWI-like (brown arch) in *S. mediterranea* and *M. lignano*, respectively. FL-Agos are indicated with the red arch. Species abbreviation are indicated in Table 1

early observations, we detected Piwi-like genes only in turbellarians, with a notable amplification in the genomes of both free-living species (Fig. 2, Additional file 1: Table S3). Again, the distribution of copies in *M. lignano* show amplification rounds consistent with tetraploidy. Surprisingly, structure of PIWI-like proteins from free-living flatworms showed some variations, while most *S. mediterranea* PIWIs have the linker 2 domain besides the PAZ and PIWI domains, *M. lignano*'s have the linker 1 and mid domains instead (Fig. 2). The apparently loss of PIWI-like genes during the evolution of parasitic platyhelminthes led to the question of how these organisms maintain their genome stability throughout generations [40].

While the presence of flatworm-specific Agos have been previously noticed in model species [40, 42], we now found that this group of flatworm-specific Agos (named FL-Ago by us), is actually amplified, being represented by two or three paralogous genes identified in each of the parasitic species (Fig. 2, Additional file 1: Table S3). The phylogenetic analysis suggests that independent duplications within the FL-Ago have occurred in both trematodes and cestodes. We used HMMScan and MEME to inspect FL-Agos domains finding that in most of them the mid domain was missing. The mid domain intervenes in the anchor of the 5′ phosphate of siRNA or miRNA and contains a cap-binding region that is required for efficient regulation of translation [48]. Additionally, the transcriptomic data of *S. mansoni*, *F. hepatica* and *E. granulosus* showed that FL-Agos were being expressed in several stages (Additional file 3: Figure S2). Even though we can speculate that the function of FL-Agos might be relevant it remains unknown. However, Cai et al. [51] sequenced the population of small RNA associated to Ago2 of *S. japonicum*,

detecting 19-22 nt endo-siRNA derived from TE. These results suggested that Sj.Ago2 might be a functional homologue of the *D. melanogaster* Ago2 in the short-interfering RNA (siRNA) pathway suppressing active TE in both the somatic and germline cells [40]. Considering that in all flatworm lineages different Ago proteins are amplified is tempting to speculate that they might represent diverse adaptations to cope with similar phenomena.

Most proteins of the miRNA pathway are conserved in flatworms

Pri-microRNAs are transcribed by RNApol II and processed into 60–100 nucleotides long pre-miRNA by the RNase III Drosha and its partner Pasha in a complex named microprocessor [52]. We confirmed that Drosha and Pasha are conserved in all the flatworm analyzed (Fig. 3, Additional file 1: Table S4). In addition, we identified orthologous genes of the transmembrane channel Xpo-1 that transports pre-miRNAs into the cytoplasm in *C. elegans* [53]. In vertebrates and flies Exp-5, that connects the nucleoplasm with the cytoplasm, is involved in the miRNA pathway; however, in nematodes no Exp-5 orthologue has been identified and it has been hypothesized that Xpo-1 replaces its function [53, 54]. An orthologue of Exp-5 has already been described in *S. mansoni* [43] and we detected orthologues in other flatworms by using SmExp-5 as query in tBLASTn search (Additional file 1: Table S4). In *C. elegans*, Dcr-1 further process pre-miRNAs into a ~ 22 nt RNA duplex [55]. In *D. melanogaster*, however, Dcr-1 requires Loquacious (Loqs), a dsRNA-binding protein, for pre-miRNA processing [45]. Loqs orthologues were identified in most of flatworm species analyzed (Additional file 1: Table S4),

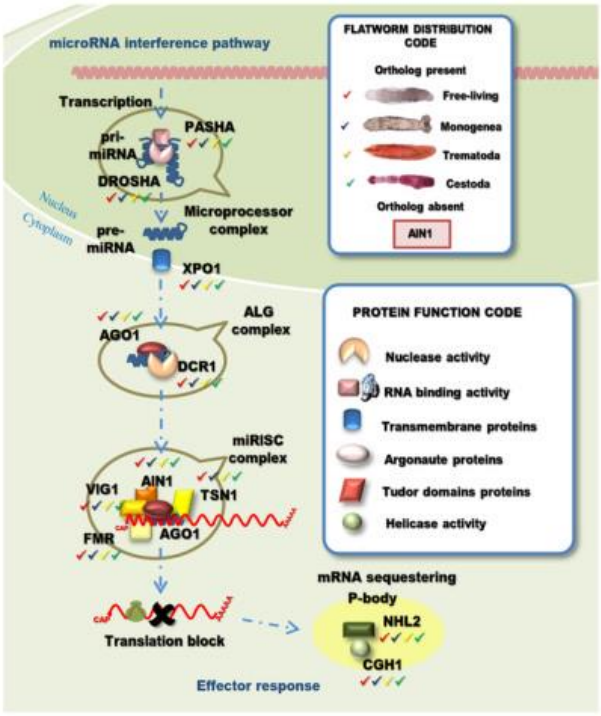


Fig. 3 MicroRNA pathway genes detected in flatworms. Pri-microRNA is transcribed in the nucleus by RNA polymerase II and processed into pre-microRNA by Pasha and Drosha (the microprocessor complex). Pre-miRNA is exported to the cytoplasm by the transmembrane protein Xpo1 where is processed by Dicer into a 21 ~ nucleotides dsRNA. Ago1/2 removes the passenger strand from the mature miRNA. The miRNA with Ago1 and other factors associated to the miRISC complex will anneal the target mRNA with 100% complementarity to the seed region (nucleotides 2–8 of miRNA) blocking the translation and sequestering the mRNA to the P-body. Factors involved in the microRNA pathway with homologous genes detected in flatworms are indicated ("Flatworm Distribution Code" box). A 'shape' code was used to indicate predicted function of factors ("Protein Function Code" box)

suggesting that this protein may be relevant for RNAi processing in flatworms.

Argonaute gene products (ALG-1 or ALG-2 in *C.elegans*, Ago1 in *D. melanogaster*) recognize short RNA duplexes, cleaving the passenger strand from the mature miRNA. The Ago protein in tandem with the miRNA and other associated factors form the miRNA Induced Silencing Complex (miRISC) that lead to the degradation of target mRNAs [55].

GW182 Is conserved in free living *M. lignano* and divergent in parasitic parasites

GW182 is one of the most critical miRISC factors. In *D. melanogaster*, Gawky (GW182 orthologue) contributes to translation repression and 'label' the target mRNA for decay via deadenylation and decapping [56]. Moreover, knock-down of GW182 resulted in the suppression of mRNA silencing mediated by miRNA with no effect on the expression levels of corresponding miRNAs or Argonaute protein [57]. GW182 family members share a common domain characterized by a central ubiquitin associated-like domain (UBA) and C-terminal RNA recognition motif (RRM). These domains are embedded in regions predicted to be unstructured that include three blocks of glycine-tryptophan repeats (the N-, middle- and C-terminal GW-repeats), and a glutamine-rich (Q rich) region located between the UBA and the RRM domains, that are difficult to retrieve by simple homology searches (Fig. 4) (reviewed at [57]). On the other hand, two proteins (AIN-1 and AIN-2) with a low number of GW-repeats but no Q-rich region or UBA and RRM domains has been described in *C. elegans*. This lack of common domain architecture suggests that AIN-1 and AIN-2 may not be members of the GW182 protein family, but rather represent functional analogs [57].

AIN-1 and AIN-2 from *C. elegans*, Gawky from *D. melanogaster*, and a GW182 family member recently reported in *Crassostrea gigas* [58], were employed as queries to interrogate the flatworm genomes in Wormbase parasite. Hits were only identified in the genome of *M. lignano* using the *C. gigas* sequence, probably due to closer evolutionary distance between these species. The *M. lignano* sequence was predicted to contain UBA and RRM domains like the GW182 family member. Despite the unstructured nature of GW182 family members, outside UBA and RRM domains there are two short but well conserved motifs in the Ago-binding domain and the silencing domain (Fig. 4) [57]. Using MEME suite, motifs were found to be partially conserved in *M. lignano* sequence (Fig. 4). In addition, we found a Q-rich region with 22.7% of glutamine residues between the UBA domain and the conserved motif of silencing domain as previously described for the Gawky sequence (Fig. 4). This suggests that the free living *M. lignano* has at least one GW182 family member. Next, we used this *M. lignano* putative GW182 as bait to search by tBLASTn in the genomes of other flatworms, finding several hits with different confidence, mainly in proteins with unknown function (Additional file 1: Table S4). MEME search on these putative GW182 orthologues identified the presence of dispersed sequence conservation (Additional file 4: Figure S3). Experimental validation of the role of the putative GW182 is now needed to confirm its essentiality in miRNA-driven gene silencing and the interaction with the miRISC complex in flatworms.

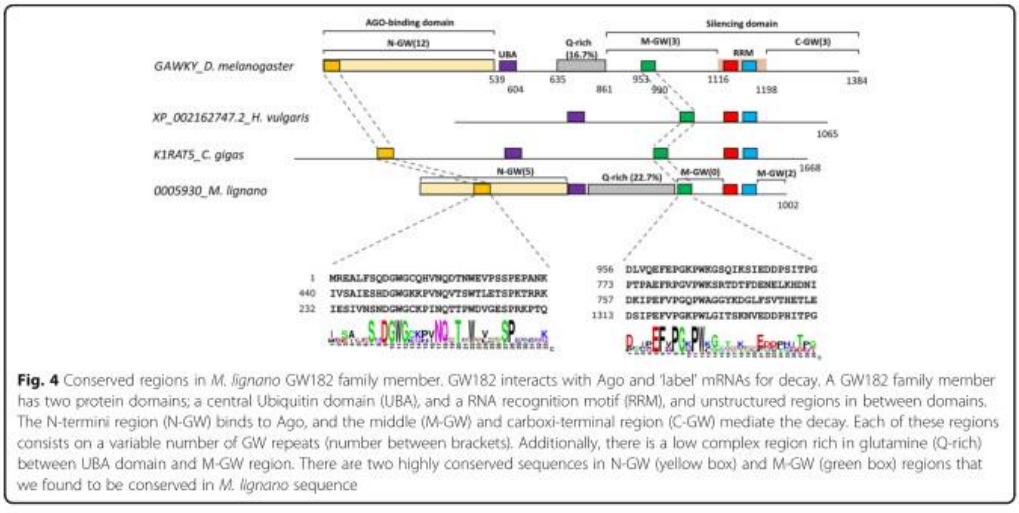


Fig. 4 Conserved regions in *M. lignano* GW182 family member. GW182 interacts with Ago and 'label' mRNAs for decay. A GW182 family member has two protein domains; a central Ubiquitin domain (UBA), and a RNA recognition motif (RRM), and unstructured regions in between domains. The N-termini region (N-GW) binds to Ago, and the middle (M-GW) and carboxi-terminal region (C-GW) mediate the decay. Each of these regions consists on a variable number of GW repeats (number between brackets). Additionally, there is a low complex region rich in glutamine (Q-rich) between UBA domain and M-GW region. There are two highly conserved sequences in N-GW (yellow box) and M-GW (green box) regions that we found to be conserved in *M. lignano* sequence

Other miRISC factors are moderately conserved in flatworms

Besides the factors analysed above, other proteins have been described in the miRISC, mainly RNA binding proteins [55, 59, 60]. Homologues genes of TSN-1 and VIG-1 were identified in almost all flatworms (Additional file 1: Table S4). We confirmed the presence of an orthologue to Fragile X Retardation factor (FXR) involved in miRNA silencing in mammals previously reported in *S. mansoni* [43]. On the other hand, no RNA binding protein GLD-1 orthologues were identified in our survey; however, we detected orthologue proteins for ASD-2, a paralog of GLD-1 in *C. elegans*, which might be a functional homolog in flatworms.

The exoRNA interference is functional in flatworms

The RNAi pathway discovered at the dawn of the century in *C. elegans*, is now known as the exo-RNAi pathway (Fig. 5). The mechanism is triggered when a long dsRNA molecule is internalized by the cell by a transmembrane channel protein called SID-1 [61]. Other transmembrane coding genes paralogous to SID-1 in *C. elegans* have been described; CHUP-1, Y37H2C1 and C08A9.3. However, none of them seem to be involved in RNAi, CHUP-1 in particular, is expressed mainly in the intestine and the terminal bulb of the pharynx where is involved in the internalization of cholesterol [62]. Interestingly, in a pairwise distance matrix of aligned sequences we observed that platyhelminthes, insects and vertebrates SID-1 homologues (named Sid-like by [63]) are closely related to CHUP-1 than to SID-1. Thus, platyhelminthes, insects and vertebrates SID-like transporters might instead be orthologous of *C. elegans* CHUP-1 [63]. It has been recently reported that like *C. elegans* SID-1, vertebrates SID-like binds to dsRNA through its extracellular domain [64] suggesting that the role of SID-1 might be exerted by SID-like proteins in other organisms.

Other key transmembrane protein involved in dsRNA uptake in *C. elegans* is SID-2. Interestingly, SID-2 is required for RNAi delivered by dsRNA soaking or feeding (environmental RNAi), but not for systemic RNAi. The protein is mainly expressed in the apical intestinal membrane where in coordination with SID-1 internalizes the interfering molecules [65]. Even though RNAi by soaking has been proven to be functional in platyhelminthes [5, 9, 16], we did not detected orthologues of SID-2 in flatworms. Therefore, it is possible that the role of SID-2 has been substituted by other proteins that enable the preliminary uptake of dsRNA like SID-1 [65].

In *C. elegans*, the dsRNA is first recognized by the dsRNA-binding protein RDE-4 and presented to DCR for processing in the cytoplasm [66]. RDE-4 is conserved only in some nematode species [16, 49], hence, orthologues of R2D2, the functional homologous of RDE-4 reported in *D. melanogaster* [67], were searched. No proteins similar to RDE-4 or R2D2 were identified in

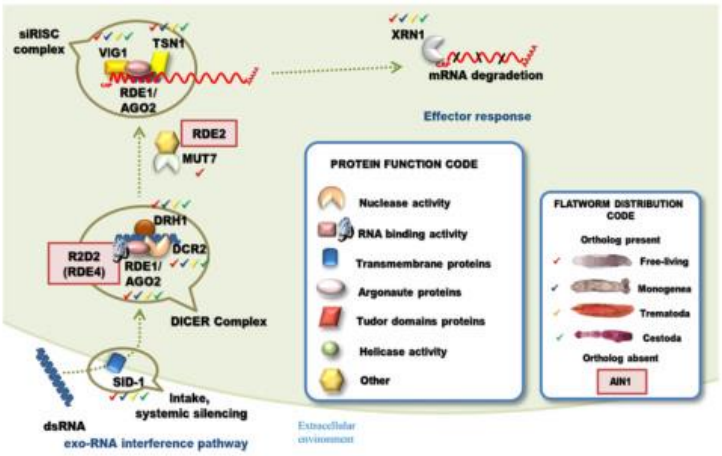


Fig. 5 Exogenous RNA interference pathway genes detected in flatworms. An exogenous double-stranded RNA is internalized by transmembrane channels SID-1. In the cytoplasm the dsRNA is recognized by RDE4 (an orthologous of R2D2 described in *D. melanogaster*), this promotes the formation of the DICER complex that process the long dsRNA into 22 nucleotides long dsRNA. The Argonaute protein RDE1 (homologous to Ago2 in *D. melanogaster*) further process the dsRNA releasing matures siRNA that together with other factors form the effector complex siRISC. mRNAs targeted by siRISC are degraded, XRN1 has 5'-3' exonuclease activity and intervenes in this process. Mut-7 and RDE-2 act downstream of RDE-1 and RDE-4 and are required for siRNA accumulation. Exogenous RNAi pathway factors with homologous genes detected in flatworms are indicated ("Flatworm Distribution Code" box). A shape code was used to indicate predicted function of factors ("Protein Function Code" box)

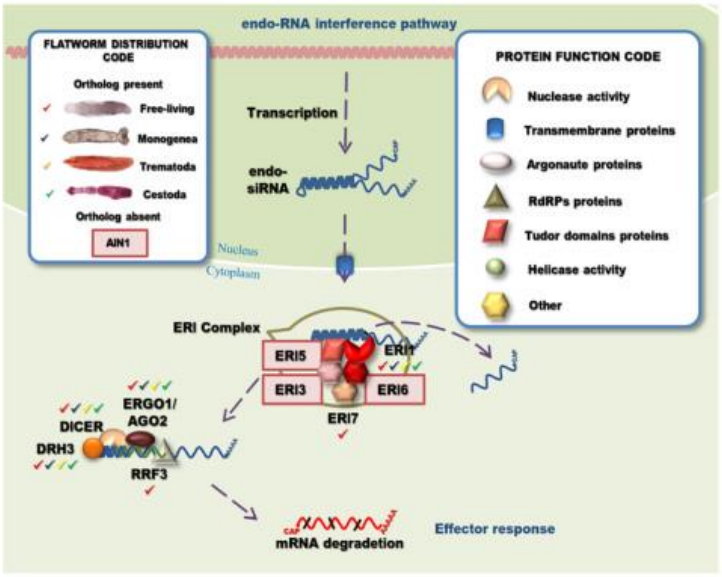


Fig. 6 Endogenous RNA interference pathway genes detected in flatworms. The non-coding RNA (ncRNA) containing a dsRNA region is transcribed in the nucleus and processed in the cytoplasm by ERI complex. After binding, the ncRNA ERI-1 removes one of unpaired strands creating a hairpin structure suitable for priming the synthesis of long dsRNA by RRF-3. Dicer and DRH-3 then, cleave the long dsRNA, and ERGO-1 (homologous to Ago2 in *D. melanogaster*) loaded with siRNA triggers the target RNA degradation. Endogenous RNAi pathway factors with homologous genes detected in flatworms are indicated ("Flatworm Distribution Code" box). A 'shape' code was used to indicate predicted function of factors ("Protein Function Code" box)

flatworms, an intriguing observation considering its critical role in the generation of siRNAs [68]. However, since RNAi has proven to be functional in several flatworms, an alternative pathway or other yet uncharacterized mediators might be operating in these organisms.

In addition to Dcr and Ago genes, we identified putative homologs of DRH-1, a DExH-box helicase protein associated to Dcr and XRN-1, a 5′-to-3′ exonuclase that degrades the mRNA targeted by the RISC complex (Fig. 5) (Additional file 1: Table S5).

No homologs of Mut-7 and RDE-2 in flatworms were detected; these factors are involved in a protein complex in *C. elegans*, required for downstream events in the RNAi not well understood [69]. In addition, Mut-7 orthologues were described in *D. melanogaster* and *H. sapiens* [70].Therefore, we used Nbr (the orthologue of Mut-7 in *D. melanogaster*) as query to search by tBLASTn on the flatworm genomes. Only *M. lignano* showed three genomic regions with high similarity (43.3% on average), two of them with annotated genes. These genes were used as bait to test the other flatworm genomes with no hits identified. Therefore, it is possible that Mut-7 homologs has been lost in the evolution of flatworms or a high evolution rate within the flatworm phylogeny made them undetectable by our method. RDE-2, on the other hand, has only been detected in *C. elegans* and other *Caenorhabditis* species [49, 71]. We can speculate that either the role of this complex is not crucial in flatworms or that there are other non-evolutionary related factors playing a similar role.

The endo-RNAi pathway is reduced in parasitic flatworms

In *C. elegans*, in addition to exo-siRNA, a endo-RNAi pathway triggered by endogenous dsRNAs derived from overlapping genes, long hairpins, and TE in the germline, has been reported [72, 73]. The primary transcripts are processed by the Enhancer of RNA Interference (ERI) complex formed by Dcr-1, DRH-3, ERI-1, ERI-3, ERI-5, ERI-6, ERI-7 and RRF-3, a RNA-dependent RNA polymerase (RdRP) [73–75]. The suppression of the ERI factors increased the RNAi effect triggered by exogenous dsRNA [76, 77], suggesting that the exogenous and endogenous RNAi pathways shared common factors [74, 78].

In flatworms, we only detected the presence of ERI-1, a ribonuclease with an 3′-5′ exonuclease domain (Fig. 6). Interestingly, ERI-1 putative orthologues were lost in the teanid and hymenolepid cyclophilidean cestodes, while still present in *M. corti*. In addition, in the

Diphyllobothriidae order only *D. latum* seem to have lost this gene (Additional file 1: Table S6). On the other hand, we detected homologue sequences of R02D3.8, a paralogous of ERI-1, in almost all platyhelminthes, including the species lacking an ERI-1 orthologue (Additional file 1: Table S6). It is possible that in cestode species lacking ERI-1 its function has been substituted by homologs of R02D3.8. In any case, experimental evidences are needed to prove this hypothesis. While we were unable to detect any other member of the ERI group in parasitic flatworms, putative orthologues of NTP hydrolase ERI-7, are found in the free living species *M. lignano* and the planarian *S. mediterranea*. As happens with other factors already described, *M. lignano*, has three predicted genes with high homology to RRF-3, i.e. the RdRP of the endo-RNAi, and like the ERI factors, it has been reported that the silencing of RRF-3 increased the effect of exo-RNAi [79].

The RNAi amplification pathway is functional in *M. lignano*

In addition to the primary silencing, a pathway that amplifies the interfering signal has been reported in *C. elegans* [55]. RNA dependent RNA Polymerases (RdRP) that polymerase dsRNA using the target mRNA bonded to the siRNA as template are protagonists in this pathway. The amplification pathway is not conserved in metazoans, being absent in insects and mammals [70].

Among the thirteen *C.elegans* factors involved in the amplification pathway, we identified PIR-1, SMG-2 and SMG-6 conserved in flatworms, but no RDE-3 orthologues were detected. However, we did find homologous sequences to its paralogous gene GLD-2 (Fig. 7a, Additional file 1: Table S7).

In addition, four RdRPs were identified in *M. lignano*, one of them with high homology to EGO-1, which plays a critical role in the adult germline exo-RNAi pathway [80]. To detect low conserved homologs of RdRPs in other flatworms, we used *M. lignano* genes as queries in tBLASTn searches; however, no evident homologs were detected.

RdRPs are conserved in early branches of the tree of life, i.e. plants, fungi, yeast and nematodes [49, 81] Therefore, the possibility that *M. lignano* RdRPs derived from contaminants such as virus, fungi or yeast cannot be ruled out. However, a phylogenetic analysis that included virus, fungi, yeast, plant, *C. elegans* and *M. lignano* RdRPs strongly suggested that the *M. lignano* genes are coded by the nuclear genome (Additional file 5: Figure S4).

M. lignano has a basal place in the evolution of platyhelminthes [82]. The absence of RdRP genes in other flatworms may be due to the acquisition of other amplification mechanism or functional RdRP homologs. In humans, for example, the telomerase reverse transcriptase (TERT) show RdRP activity by using a snoRNA as RNA template, additionally, the dsRNAs produced by this TERT are processed by Dicer, and drive the silencing of the target RNA [83]. However, RdRPs family might display a high divergence at the protein sequence in platyhelminthes, not allowing their detection by blast search.

Additionally, we detected genes involved in the persistence of RNAi. This persistence depends on SMG genes 2-, 5- and 6-, possibly by facilitating the amplification of the RNAi signal [84] (Fig. 7a, Additional file 1: Table S7).

Spreading of RNAi

In *C. elegans*, the cell-to-cell spread of the interfering signal has been reported, allowing a systemic silencing via endo/exo-RNAi pathways [55]. This phenomenon depends on SID-1 and the products of the RSD (from RNA spreading deficient) genes −2, −3 and −6. Mutants for RSD-2, RSD-3 and RSD-6 still retain the systemic RNAi in somatic cells, but are unable to further spread dsRNAs to the germline [85]. In *C.elegans* RSD-2, a protein with no predicted domains interacts with RSD-6, a Tudor domain containing protein. RSD-3 encodes an epsin N-terminal homology (ENTH) domain, suggesting that RSD-3 may play a role in the traffic of vesicles during systemic RNAi [85].

In flatworms, among all the factors of this pathway only homologous genes of RSD-3 were identified (Fig. 7b, Additional file 1: Table S8), suggesting that the RSD-2/RSD-6 system is either not conserved, or is too divergent to be detected by our method. Interestingly, similar results have been reported in insects [63], consistent with the observation that systemic RNAi is not possible in *D. melanogaster*. The demonstration of RNAi silencing by feeding in planaria [86] supports the existence of a functional spreading mechanism at least in this free living flatworm. No reports of similar feeding approaches have been reported so far in parasitic flatworms, so is not clear if spreading is limited to planaria or extended to all flatworms. In any case, efforts to unravel the players in RNAi spreading in planaria are warranted.

Chromatin remodeling mediated by siRNA is possible

In addition to the cytoplasmic RNAi, in *C. elegans* mRNAs are silenced within the nucleus. NRDE-3, an Ago protein, binds to secondary siRNAs in the cytoplasm, transports them into the nucleus, and interacts with nascent transcripts in an siRNA-dependent manner [87]. We did not detected specific orthologues of NRDE-3 in flatworms; however, it is possible that the role of NRDE-3 is played by any of the other Agos detected in flatworms, including the flatworm-specific Agos. On the other hand, the absence of a canonical amplification pathway in all flatworms with the only exception of *M. lignano*, suggests that secondary siRNA might not be

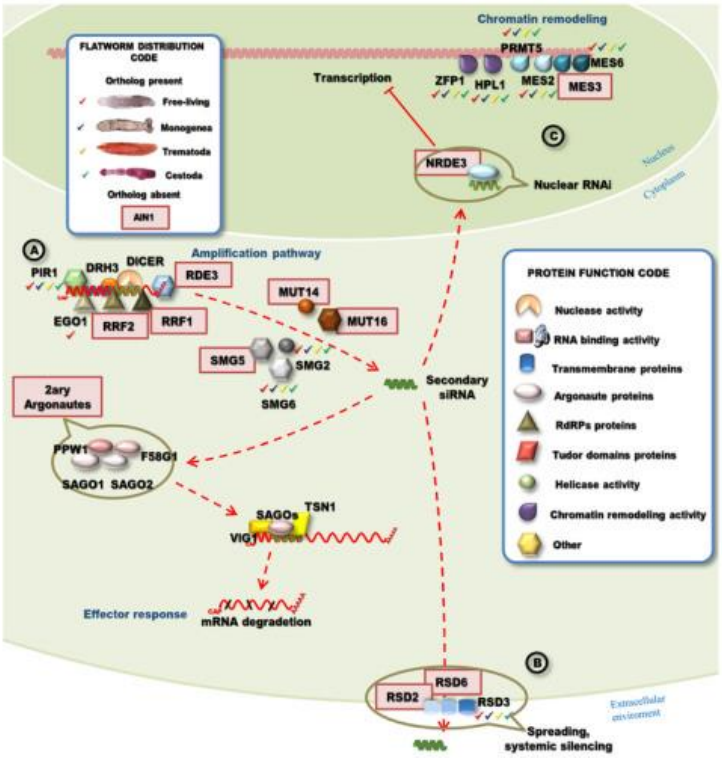


Fig. 7 RNAi amplification, spreading and transcriptional gene silencing pathways. **a** Amplification pathway. RNA dependent RNA Polymerases (RdRP) employs the 3' of the siRNA as primer to synthesis a new long dsRNA using the mRNA as template. PIR-1 and DRH-3 are required for Dicer to process long dsRNA, while RDE-3 intervenes in an intermediate step, stabilizing the product of the initial round of cleavage, allowing the amplification mediated by RdRPs. SMG genes, Mut-14 and Mut-16 act downstream of secondary siRNA generation. RNAi persists in the animals exposed to dsRNA for several days [5]. This persistence depends on SMG genes. The secondary siRNA with the Secondary Argonautes (SAGOs) produce a secondary siRISC complex. **b** Systemic spreading. The secondary siRNAs spread throughout other tissues inducing the systemic silencing of the target gene. **c** Co-transcriptional silencing and chromatin remodeling. A specialized Argonaute, NRDE-3, transports the secondary siRNA to the nucleus where produce co-transcriptional gene silencing and interacts with other factors to remodel the chromatin. Factors with homologous genes detected in flatworms are indicated ("Flatworm Distribution Code" box). A shape code was used to indicate predicted function of factors ("Protein Function Code" box)

produced and consequently co-transcriptional silencing might not be operational in parasitic species.

It has been shown that RNAi also induces transcriptional gene silencing by modifying the chromatin structure [71]. If the targeted genes are expressed in the germline the silencing signal can be transmitted epigenetically to the next generations [71]. Interestingly here as well, a differential distribution was observed. Seven genes were conserved across flatworms, including GFL-1, the ATP-dependent DEAD/H box RNA helicase RHA-1, the metyl transferase PRMT-5, the heterochromatic protein HPL-1, the chromodomain containing protein MRG-1 and the polycomb complex proteins MES-2 and MES-6 (Fig. 7c, Additional file 1: Table S9). Only MES-3, a member of the polycomb complex, was not detected in flatworms, but it has been reported that MES-3 is a novel acquisition of nematodes and is not conserved in *D. melanogaster*, which only have functional homologs of MES-2 and MES-6 [88]. Remarkably, the zinc finger protein ZPF-1 was detected only in free living species and trematodes, being absent in cestodes. However, three paralogous genes of ZPF1 exist (PHF-14,

PHF-15 and LIN-49 in Additional file 1: Table S9), being all of them present in flatworms, suggesting that the function might be conserved.

Absence of a canonical PIWI pathway in parasitic flatworms

The methylase HEN-1, implicated in the siRNAs and piRNA associated to PIWI-like Argonautes [89] is only detected in *M. lignano* and *S. mediterranea* consistent with the exclusive presence of PIWI-like Agos in free living flatworms (Additional file 1: Table S9).

The absence of Vasa, another key factor of piRNA pathway [40] and the lack of piRNA in small RNA sequencing data of parasitic platyhelminthes (Table 1) support the hypothesis of an evolutionary loss of piRNAs in parasitic flatworms. However, it has been reported that FL-Ago-2 and PL10 (a homolog of Vasa) transcripts are enriched in germline cells of *S. mansoni* [90], and as we mentioned above, FL-Ago-2 may be involved in TE silencing, suggesting that parasitic flatworms might have an alternative pathway to the PIWI pathway. In any case, other approaches as functional genomics analyses employing RNAi and transgenesis to define pathways and enzymes that participate in silencing of TEs and immunoprecipitations to characterize binding partners are needed to further decipher this enigma.

Conclusions

Our data indicate that in all analyzed species most of the small RNA pathways are conserved, providing a strong bioinformatic evidence that gene-silencing mediated by RNAi and endogenous gene expression regulation mediated by small RNAs is possible in all the members of the phylum. While in parasitic flatworms there is a marked reduction in the repertoire of small RNA pathway factors, in *S. mediterranea* reduction is less pronounced and even lesser in *M. lignano* where a putative amplification pathway of siRNAs was identified (Fig. 8).

A general phenomenon of simplification has occurred in flatworms, especially in parasitic species [91–93]. We found that the losses of metabolic pathways are differential and related to the environment in the invaded tissue. In addition, a reduction in the set of flatworm miRNAs, greater in parasitic species compared to free-living organisms [23, 26], might suggest that the absence of miRNA families could be related to the loss of target mRNAs. This process of loss of redundancy in the biological system appears very early in the evolution of parasitic flatworms and is still unclear if it represents an adaptation to parasitism or is an ancestral characteristic shared between the Neodermata (as proposed by Hahn et al. [93]). However, it is tempting to hypothesize that, in the case of small RNAs, factors with similar function have been removed of the genomes of parasitic flatworms, while remaining factors could display multi-role in different pathways. This reduction in the small RNA machinery could affect the biogenesis of some small RNA, and could be related to the reported pronounced loss of miRNAs in the Neodermata species [23, 26].

On the other hand, we detected flatworm specific amplifications and variants. We found that Dcr-2 is structurally different in trematodes and that is duplicated in Opistorchidae and Fasciolidae. In addition, a flatworm specific group of argonaute proteins (FL-Agos) exist, being expanded independently in the diverse linages (Fig. 8). The relevance of these novelties are currently unknown, studies that combine both in silico and experimental approaches are needed to unravel their function.

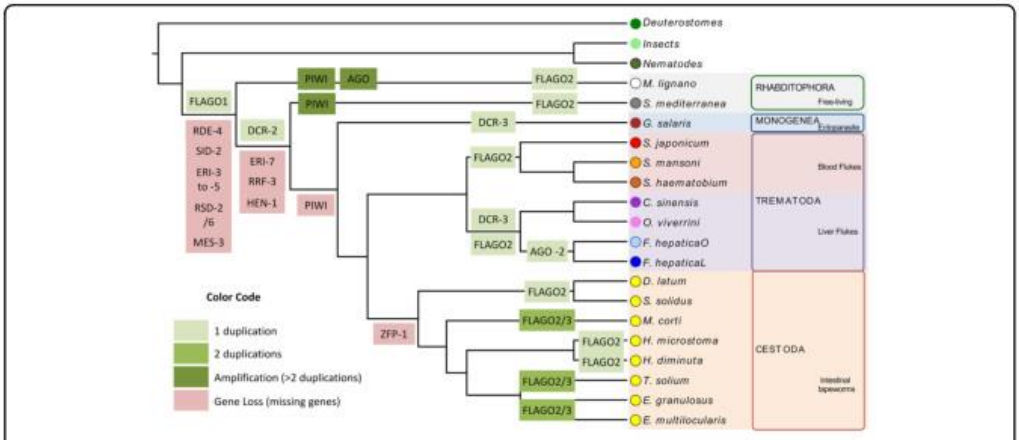


Fig. 8 Summary of small RNA pathways gene gains and losses in the evolution of flatworms. A schematic representation of the phylogeny of flatworms indicating major gene duplications or gene losses

Here we used a strict homology search method against genomes that are in most cases in a draft stage. We did not rely only in current annotation (since this might be partial or incorrect for many species) adding a whole genome homology search to confirm the putative absences. We are quite confident that this combined approach provide a good picture of the main aspects of small RNA pathways in flatworms. Other in silico approaches like PSI-BLAST search [94], a method for low-homology structural search [95] or transitive homology approaches [96] can be used in the future to search for more divergent sequences.

Additionally, experimental approaches are needed to confirm some of the results reported here and to detect novel factors that could be developed uniquely in flatworms. Functional genomics utilizing RNAi [5], transgenesis [97, 98], and more recently genome-editing by CRISPR-Cas9 [99] would help to define the function of some of the factors reported here, while immunoprecipitation assays could be used in the identification of novel factors [100].

In any case, we believe that the results reported here are relevant not only to shine a light on the basic biology and gene regulation mechanisms in these organisms, but also to optimize the RNAi as functional genomic tool for parasitic worms, agents of neglected tropical diseases affecting hundreds of million people worldwide.

Methods

Identification of small RNA pathways proteins and generation of local repository

Sixty-one proteins involved in small non-coding RNA pathways in *C. elegans* were identified from the literature (Additional file 1: Table S1), and the sequences were downloaded from wormbase (www.wormbase.org, release WS244). A local repository with the genomes and proteomes of 8 species of Trematodes, 12 species of Cestodes, 2 Monogeneans and 2 Turbelaria was generated (Table 1). The data was downloaded from Wormbase Parasite (http://parasite.wormbase.org) and The *Gyrodactylus salaris* Genome Project (http://invitro.titan.uio.no/gyrodactylus/). Results were later corroborated with the new genomic annotations published in Wormbase Parasite release 8.

Homologous search with BLAST

The protein sequences of all *C. elegans* small RNA pathways factors were used to interrogate the proteomes of platyhelminthes using BLASTp [101]. All hits in the platyhelminthes proteomes with an expected value ≤1 E-05 were retained and reciprocally BLASTed (using the same e-value cut-off) to the *C. elegans* proteome. Only the best reciprocal hits between flatworms and *C. elegans* were retained for further inspection.

When no homologous protein was identified or the retrieved gene was incomplete or fragmented, a genomic search using tBLASTn was conducted. A careful inspection of the resulting High-scoring Segment Pairs (HSP) aligned to the genomes were useful to extend gene annotation or to fully annotate putative novel genes. These were confirmed by reverse tBLASTn to *C. elegans*. Alternatively, if a positive hit was retrieved in some flatworm species but not in others, these were used as queries for new genomic and proteomic reciprocal blast searches. This strategy helped to overcome the disparity on quality and completeness of genomic assemblies and/or gene annotation among the genomes deposited in Wormbase Parasite (http://parasite.wormbase.org/).

Protein domains prediction

To confirm that the identified sequences from platyhelminthes were functional homologs of the *C. elegans* proteins, domains were predicted with HMMScan [102], using an expected value cut-off of 1 E-03. Only those sequences with a similar functional profile to the *C. elegans* protein were considered as orthologs. However, we consider flatworm genes to be homologs when reverse BLAST produced a hit to a paralogous of the gene of *C. elegans* involved in the small non-coding RNA pathways, to avoid confusion these cases are explicitly indicated in the results section. In cases where protein domains were not detected with HMMScan, MEME suite [103] was used to identify conserved amino acidic positions.

Sequence alignment and phylogenetic tree building

Local alignment with MAFFT [104] was used to align the detected sequences in platyhelminthes with the *C. elegans* reference proteins and orthologous detected in other species. Alignments were visualized and edited when needed with BioEdit version 7.2.5 [105]. The alignments were used to build Maximum Likelihood (ML) phylogenetic trees with PhyML [106], with statistical branch support (SH-like). The model used to build the trees was inferred with Modelgenerator [107]. The trees were visualized with Evolview [108].

Additional files

Additional file 1: Table S1. *C. elegans* small non-coding RNA pathway genes used for homology search. **Table S2.** Ribonuclease III genes identified in flatworms. **Table S3.** Argonaute genes identified in flatworms. **Table S4.** microRNA pathway genes detected in flatworms. **Table S5.** Exogenous RNA interference pathway genes detected in flatworms. **Table S6.** Endogenous RNA interference pathway genes detected in flatworms. **Table S7.** Amplification pathway genes detected in flatworms. **Table S8.** RNA interference spreading signal genes detected in flatworms. **Table S9.** Chromatin remodeling genes associated to RNA interference detected in flatworms. (XLSX 47 kb)

Additional file 2: Figure S1. Dcr-2 and Dcr-3 genomic location and expression in Fasciolidae and Opisthorchiidae. In *C. sinensis* and *O. viverrini* genomes both paralogues are separated by less than 10 kb, while, in *F. hepatica* the intergenic region is almost 50 kb. Transcriptomic data of *F.*

hepatica show that both genes are transcribed in several developmental stages. (TIFF 1513 kb)

Additional file 3: Figure S2. FL-Agos genomic location and expression in *S. mansoni*, *F. hepatica* and *E. granulosus*. Transcriptomic data show that all FL-Agos are expressed in several developmental stages. Genes size was found to be variable ranging from around 4 kb in *E. granulosus* to 6.5 kb in *S. mansoni* and almost 40 kb in one of the *F. hepatica* genes. (TIFF 5708 kb)

Additional file 4: Figure S3. Motifs detected in putative GW182 sequences of flatworms. Motifs detected in UBA and RRM domains are indicated in purple and lightblue boxes, respectively. The GW182 family conserved motif of AGO binding domain was also found in all flatworms (yellow box). Two additional motifs conserved only among flatworms were detected. The motif at the AGO binding domain (white box) is conserved in all flatworms, while, the other (black box) is rich in glutamine residues (Q) and is only conserved in trematodes and cestodes. Sequences of common motifs to all flatworm linages were aligned and residue conservation is indicated. Additionally, the number of GW repeats for each sequence are indicated. Species with parcial or no predicted gene model are not shown (see Additional file 1: Table S3). (TIFF 3464 kb)

Additional file 5: Figure S4. A maximum likelihood tree of RNA dependent RNA Polymerases. One hundred iterations bootstrap was calculated. Values below 0.4 are not shown. (TIFF 1023 kb)

Abbreviations

AGO: Argonaute family proteins; DCR: Dicer protein; dsRNA: double stranded RNA; endoRNAi: endogenous RNAi; ERI: Enhanced RNAi proteins; exoRNAi: exogenous RNAi; FL-AGO: Flatworm Specific Argonaute Proteins; miRISC: miRNA processing RISC complex; miRNA: microRNA; piRNA: Piwi protein interacting RNA; RdRP: RNA dependent RNA Polymerase; RISC: RNA Induced Silencing Complex; RNAi: RNA interference; siRISC: siRNA processing RISC complex; siRNA: small interfering RNA

Acknowledgments

Not applicable.

Funding

The research was funded by grant CABBIO2012-UY04. SF received a scholarship from ANII (award number POS_NAC_2015_1_110044). PS and JT are researchers from the Sistema Nacional de Investigadores (SNI-ANII), and Pedeciba.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary files or are redirected to databases where they are publically available.

Authors' contributions

SF performed the acquisition, bioinformatics analysis and interpretation of data and contributed in writing the manuscript. GR was involved in drafting the manuscript and critical revision of its content. PS participated in the design of the study and the interpretation of data. JT participated in the design of the study and the interpretation of data, drafting the manuscript and critical revision of its content. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Departamento de Genética, Facultad de Medicina, Universidad de la República (UDELAR), Gral. Flores 2125, CP11800 Montevideo, MVD, Uruguay. [2]Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [3]Laboratorio de Interacciones Moleculares, Facultad de Ciencias, Universidad de la República (UdelaR), Montevideo, Uruguay.

Received: 7 June 2017 Accepted: 5 September 2017
Published online: 11 September 2017

References

1. Mourão MM, Grunau C, Loverde PT, Jones MK, Oliveira G. Recent advances in *Schistosoma* genomics. Parasite Immunol. 2012;34(2–3):151–62.
2. Newmark PA. Opening a new can of worms: a large-scale RNAi screen in planarians. Dev Cell. 2005;8(5):623–24.
3. Moguel B, Bobes RJ, Carrero JC, Laclette JP. Transfection of Platyhelminthes. Biomed Res Int. 2015;2015:206161.
4. Rinaldi G, Morales ME, Cancela M, Castillo E, Brindley PJ, Tort JF. Development of functional genomic tools in trematodes: RNA interference and luciferase reporter gene activity in *Fasciola hepatica*. PLoS Negl Trop Dis. 2008;2(7):e260.
5. Dell'Oca N, Basika T, Corvo I, Castillo E, Brindley PJ, Rinaldi G, et al. RNA interference in *Fasciola hepatica* newly excysted juveniles: long dsRNAinduces more persistent silencing than siRNA. Mol Biochem Parasitol. 2014;197(1-2):28–35.
6. McGonigle L, Mousley A, Marks NJ, Brennan GP, Dalton JP, Spithill TW, et al. The silencing of cysteine proteases in *Fasciola hepatica* newly excysted juveniles using RNA interference reduces gut penetration. Int J Parasitol. 2008;38:149–55.
7. McVeigh P, McCammick EM, McCusker P, Morphew RM, Mousley A, Abidi A, et al. RNAi dynamics in juvenile *Fasciola spp.* liver flukes reveals the persistence of gene silencing in vitro. PLoS Negl Trop Dis. 2014;8(9):e3185.
8. Sripa J, Pinlaor P, Brindley PJ, Sripa B, Kaewkes S, Robinson MW, et al. RNA interference targeting *Opisthorchis viverrini*. Parasitol Int. 2011;60:283–88.
9. Wang X, Chen W, Tian Y, Huang Y, Li X, Yu X. RNAi-mediated silencing of enolase confirms its biological importance in *Clonorchis sinensis*. Parasitol Res. 2014;113:1451–458.
10. Pakharukova MY, Vavilin VA, Sripa B, Laha T, Brindley PJ, Mordvinov VA. Functional analysis of the unique cytochrome P450 of the liver fluke *Opisthorchis felineus*. PLoS Negl Trop Dis. 2015;9(12):e0004258.
11. Pierson L, Mousley A, Devine L, Marks NJ, Day TA, Maule AG. RNA interference in a cestode reveals specific silencing of selected highly expressed gene transcripts. Int. J. Parasitol. 2010;40:605–15.
12. Pouchkina-Stantcheva NN, Cunningham LJ, Hrčkova G, Olson PD. RNA-mediatedgene suppression and in vitro culture in *Hymenolepis microstoma*. Int J Parasitol. 2013;43:641–46.
13. Hu D, Song X, Xie Y, Zhong X, Wang N, Zheng Y, et al. Molecular insights into a tetraspanin in the hydatid tapeworm *Echinococcus granulosus*. Parasit Vectors. 2015;8:311.
14. Spiliotis M, Mizukami C, Oku Y, Kiss F, Brehm K, Gottstein B. *Echinococcus multilocularis* primary cells: Improved isolation, small-scale cultivation and RNA interference. Mol. Biochem. 2010;174:83–87.
15. Moguel B, Moreno-Mendoza N, Bobes RJ, Carrero JC, Chimal-Monroy J, Díaz-Hernández ME, et al. Transient transgenesis of the tapeworm *Taenia crassiceps*. Springerplus. 2015;4:496.
16. Geldhof P, Visser a, Clark D, Saunders G, Britton C, Gilleard J, et al. RNA interference in parasitic helminths: current situation, potential pitfalls and future prospects. Parasitology. 2007;134:609–19.
17. Tran MH, Freitas TC, Cooper L, Gaze S, Gatton ML, Jones MK, et al. Suppression of mRNAs encoding tegument Tetraspanins from *Schistosoma mansoni* results in impaired tegument turnover. PLoS Pathog. 2010;6:e1000840.
18. Krautz-Peterson G, Radwanska M, Ndegwa D, Shoemaker CB, Skelly PJ. Optimizing gene suppression in schistosomes using RNA interference. Mol Biochem Parasitol. 2007;153:194–02.
19. Krautz-Peterson G, Bhardwaj R, Faghiri Z, Tararam CA, Skelly PJ. RNA interference in schistosomes: machinery and methodology. Parasitology. 2010;137:485–95.
20. Palakodeti D, Smielewska M, Graveley BR. MicroRNAs from the planarian *Schmidtea mediterranea*: a model system for stem cell biology. RNA. 2006; 12:1640–649.

21. Qin YF, Zhao JM, Bao ZX, Zhu ZY, Mai J, Huang YB, et al. Identification ofsmall non-coding RNAs in the planarian *Dugesia japonica* via deep sequencing. Genomics. 2012;99:315–21.
22. Zhou X, Battistoni G, El Demerdash O, Gurtowski J, Wunderer J, Falciatori I, et al. Dual functions of Macpiwi1 in transposon silencing and stem cell maintenance in the flatworm *Macrostomum lignano*. RNA. 2015;21:1885–897.
23. Fromm B, Worren MM, Hahn C, Hovig E, Bachmann L. Substantial loss of conserved and gain of novel microRNA families in flatworms. Mol Biol Evol. 2013;30:2619–628.
24. Zhu L, Liu J, Cheng G. Role of microRNAs in schistosomes and schistosomiasis. Front Cell Infect Microbiol. 2014;4:165.
25. Wang CR, Xu MJ, Fu JH, Nisbet AJ, Chang QC, Zhou DH, et al. Characterization of MicroRNAs from *Orientobilharzia turkestanicum*, a neglected blood fluke of human and animal health significance. PLoS One. 2012;7(10):e47001.
26. 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;45:901–13.
27. Xu M-J, Wang C-R, Huang S-Y, Fu J-H, Zhou D-H, Chang Q-C, et al. Identification and characterization of microRNAs in the pancreatic fluke *Eurytrema pancreaticum*. Parasit Vectors. 2013;6:25.
28. Chen M-X, Hu W, Li J, He J-J, Ai L, Chen J-X. Identification and characterization of microRNAs in the zoonotic fluke *Fasciolopsis buski*. Parasitol Res 2016;115: 2433–438.
29. Ovchinnikov VY, Afonnikov DA, Vasiliev GV, Kashina EV, Sripa B, Mordvinov VA, et al. Identification of microRNA genes in three Opisthorchiids. PLoS Negl. Trop. Dis. 2015;9(4):e0003680.
30. Bernal D, Trelis M, Montaner S, Cantalapiedra F, Galiano A, Hackenberg M, et al. Surface analysis of *Dicrocoelium dendriticum*. The molecular characterization of exosomes reveals the presence of miRNAs. J. Proteomics. 2014;105:232–41.
31. Cucher M, Prada L, Mourglia-Ettlin G, Dematteis S, Camicia F, Asurmendi S, et al. Identification of *Echinococcus granulosus* microRNAs and their expression in different life cycle stages and parasite genotypes. Int. J. Parasitol. 2011;41:439–48.
32. Ai L, Xu MJ, Chen MX, Zhang YN, Chen SH, Guo J, et al. Characterization of microRNAs in *Taenia saginata* of zoonotic significance by Solexa deep sequencing and bioinformatics analysis. Parasitol Res. 2012;110:2373–378.
33. Basika T, Macchiaroli N, Cucher N, Espinola S, Kamenetzky L, Zaha A, et al. Identification and profiling of microRNAs in two developmental stages of *Mesocestoides corti*, a model cestode parasite. Mol. Biochem. Parasitol. 2016; 210(1-2):37-49.
34. Huang B, Zhang R. Regulatory non-coding RNAs: revolutionizing the RNA world. Mol Biol Rep. 2014;41:3915–923.
35. Huang Y, Shen XJ, Zou Q, Wang SP, Tang SM, Zhang GZ. Biological functions of microRNAs: a review. J Physiol Biochem. 2011;67:129–39.
36. Weick E-M, Miska E a. piRNAs: from biogenesis to function. Development. 2014;141:3458–71.
37. Reddien PW, Oviedo NJ, Jennings JR, Jenkin JC, Sanchez AA. SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. Science 2005;310:1327–30.
38. Palakodeti D, Smielewska M, Lu Y-C, Yeo GW, Graveley BR. The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. RNA. 2008;14:1174–186.
39. Rossi L, Salvetti A, Lena A, Batistoni R, Deri P, Pugliesi C, et al. DjPiwi-1, a member of the PAZ-Piwi gene family, defines a subpopulation of planarian stem cells. Dev Genes Evol. 2006;216:335–46.
40. Skinner DE, Rinaldi G, Koziol U, Brehm K, Brindley PJ. How might flukes and tapeworms maintain genome integrity without a canonical piRNA pathway? Trends Parasitol. 2014;30:123–29.
41. Gao Z, Wang M, Blair D, Zheng Y, Dou Y. Phylogenetic analysis of the endoribonuclease Dicer family. PLoS One. 2014;9(4):e95350.
42. Zheng Y. Phylogenetic analysis of the Argonaute protein family in platyhelminths. Mol. Phylogenet. Evol. 2012;66:1050–054.
43. Gomes MS, Cabral FJ, Jannotti-Passos LK, Carvalho O, Rodrigues V, Baba EH, et al. Preliminary analysis of miRNA pathway in *Schistosoma mansoni*. Parasitol. Int. 2009;58:61–8.
44. Jaronczyk K, Carmichael JB, Hobman TC. Exploring the functions of RNA interference pathway proteins: some functions are more RISCy than others? Biochem J. 2005;387:S61–71.
45. Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol. 2009;10:126–39.
46. Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, et al. Distinct roles for *Drosophila* Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell. 2004;117:69–1.
47. Tolia NH, Joshua-Tor L. Slicer and the argonautes. Nat Chem Biol. 2007;3:36–43.
48. Hutvagner G, Simard MJ. Argonaute proteins: key players in RNA silencing. Nat Rev Mol Cell Biol. 2008;9:22–32.
49. Dalzell JJ, McVeigh P, Warnock ND, Mitreva M, Bird DM, Abad P, et al. RNAi effector diversity in nematodes. PLoS Negl Trop Dis. 2011;5:e1176.
50. Wasik K, Gurtowski J, Zhou X, Ramos OM, Delás MJ, Battistoni G, et al. Genome and transcriptome of the regeneration-competent flatworm, *Macrostomum lignano* Proc Natl Acad Sci. 2015;112(40):12462–7.
51. Cai P, Piao X, Hou N, Liu S, Wang H, Chen Q. Identification and characterization of Argonaute protein, Ago2 and its associated small RNAs in *Schistosoma japonicum*. PLoS Negl Trop Dis. 2012;6(7):e1745.
52. Newman MA, Hammond SM. Emerging paradigms of regulated microRNA processing. Genes Dev. 2010;24(11):1086–92.
53. Büssing I, Yang J-S, Lai EC, Grosshans H. The nuclear export receptor XPO-1 supports primary miRNA processing in *C. elegans* and *Drosophila*. EMBO J. 2010;29:1830–839.
54. Murphy D, Dancis B, Brown JR. The evolution of core proteins involved in microRNA biogenesis. BMC Evol Biol. 2008;8:92.
55. Fischer SEJ. Small RNA-mediated gene silencing pathways in *C. elegans*. Int. J. Biochem. Cell Biol. 2010;42:1306–15.
56. Eulalio A, Huntzinger E, Izaurralde E. Getting to the root of miRNA-mediated gene silencing. Cell. 2008;132:9–14.
57. Eulalio A, Tritschler F, Izaurralde E. The GW182 protein family in animal cells: new insights into domains required for miRNA-mediated gene silencing. RNA. 2009;15:1433–442.
58. Zielezinski A, Karlowski WM. Early origin and adaptive evolution of the GW182protein family, the key component of RNA silencing in animals. RNA Biol. 2015;12:761–70.
59. Caudy A a, Myers M, Hannon GJ, Hammond SM. Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev. 2002;16:2491–496.
60. Caudy A a, Ketting RF, Hammond SM, Denli AM, AMP B, BBJ T, et al. A micrococcal nuclease homologue in RNAi effector complexes. Nature. 2003;425:411–14.
61. van Roessel P, Brand AH. Spreading silence with Sid. Genome Biol. 2004;5:208.
62. Valdes VJ, Athie A, Salinas LS, Navarro RE, Vaca L. CUP-1 is a novel protein involved in dietary cholesterol uptake in *Caenorhabditis elegans*. PLoS One. 2012;7:e33962.
63. Tomoyasu Y, Miller SC, Tomita S, Schoppmeier M, Grossmann D, Bucher G. Exploring systemic RNA interference in insects: a genome-wide survey for RNAi genes in *Tribolium*. Genome Biol. 2008;9:R10.
64. Li W, Koutmou KS, Leahy DJ, Li M. Systemic RNA Interference Deficiency-1(SID-1) extracellular domain selectively binds long double-stranded RNA and is required for RNA transport by SID-1. J. Biol. Chem. 2015;90(31):18904-8913.
65. McEwan DL, Weisman AS, Hunter CP. Uptake of Extracellular Double-Stranded RNA by SID-2. Mol. Cell. 2012;47:746–54.
66. Tabara H, Yigit E, Siomi H, Mello CC. The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in *C. elegans*. Cell. 2002;109:861–71.
67. Liu Q, Rand TA, Kalidas S, Du F, Kim H-E, Smith DP, et al. R2D2, A bridge between the initiation and effector steps of the *Drosophila* RNAi pathway. Science. 2003;301:1921–925.
68. Parrish S, Fire A. Distinct roles for RDE-1 and RDE-4 during RNA interference in *Caenorhabditis elegans*. RNA. 2001;7:1397–402.
69. Tops BB, Tabara H, Sijen T, Simmer F, Mello CC, Plasterk RH, et al. RDE-2interacts with MUT-7 to mediate RNA interference in *Caenorhabditis elegans*. Nucleic Acids Res. 2005;33:347–55.
70. Tabach Y, Billi AC, Hayes GD, Newman M. A, Zuk O, Gabel H, et al. Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. Nature. 2013;493:694–98.
71. Grishok A. RNAi mechanisms in *Caenorhabditis elegans*. FEBS Lett. 2005;579:5932–39.
72. Azimzadeh Jamalkandi S, Masoudi-Nejad A. RNAi pathway integration in *Caenorhabditis elegans* development. Funct Integr Genomics. 2011;11:389–05.
73. Pavelec DM, Lachowiec J, Duchaine TF, Smith HE, Kennedy S. Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in *Caenorhabditis elegans*. Genetics. 2009;183:1283–295.

74. Duchaine TF, Wohlschlegel J a, Kennedy S, Bei Y, Conte D, Pang K, et al. Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. Cell. 2006;124:343–54.
75. Fischer SEJ, Bulter MD, Pan Q, Ruvkun G. Trans-splicing in *C. elegans* generates the negative RNAi regulator ERI-6/7. Nature. 2008;455:491–96.
76. Kennedy S, Wang D, Ruvkun G. A conserved siRNA-degrading RNase negatively regulates RNA interference in *C elegans*. Nature. 2004;427:645–49.
77. Timmons L. Endogenous inhibitors of RNA interference in *Caenorhabditis elegans*. BioEssays. 2004;26:715–18.
78. Zhuang JJ, Hunter CP. The influence of competition among *C. elegans* small RNA pathways on development. Genes (Basel). 2012;3:671–85.
79. Simmer F, Tijsterman M, Parrish S, Koushika SP, Nonet ML, Fire A, et al. Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. Curr. Biol. 2002;12:1317–319.
80. Maniar JM, Fire AZ. EGO-1, a *C. elegans* RdRP, modulates gene expression via production of mRNA-templated short antisense RNAs. Curr. Biol. 2011; 21:449–59.
81. Maida Y, Masutomi K. RNA-dependent RNA polymerases in RNA silencing. Biol Chem. 2011;392:299–304.
82. Egger B, Lapraz F, Tomiczek B, Müller S, Dessimoz C, Girstmair J, et al. A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. Curr Biol. 2015;25:1347–353.
83. Maida Y, Yasukawa M, Furuuchi M, Lassmann T, Possemato R, Okamoto N, et al. An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. Nature Nature Publishing Group. 2009;461(7261):230–35.
84. Mango SE. Stop making nonSense: the *C. elegans* smg genes. Trends Genet. 2001;17:646–53.
85. Tijsterman M, May RC, Simmer F, Okihara KL, Plasterk RHA. Genes required for systemic RNA interference in *Caenorhabditis elegans*. Curr Biol. 2004;14:111–16.
86. Newmark P a, Reddien PW, Cebrià F, Sánchez Alvarado A. Ingestion of bacterially expressed double-stranded RNA inhibits gene expression in planarians. Proc. Natl. Acad. Sci. U. S. A. 2003;100(Suppl:1):1861–865.
87. Guang S, Bochner AF, Pavelec DM, Burkhart KB, Lachowiec J, Kennedy S. An Argonaute transports siRNAs from the cytoplasm to the nucleus. Science, 2008;321(5888):537–41.
88. Ketel CS, Andersen EF, Vargas ML, Suh J, Strome S, Simon JA. Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. Mol Cell Biol. 2005;25:6857–868.
89. Montgomery TA, Rim Y-S, Zhang C, Dowen RH, Phillips CM, Fischer SEJ, et al. PIWI associated siRNAs and piRNAs specifically require the *Caenorhabditis elegans* HEN1 Ortholog henn-1. PLoS Genet. 2012;8:e1002616.
90. Wang B, Collins JJ, Newmark PA. Functional genomic characterization of neoblast-like stem cells in larval *Schistosoma mansoni*. Elife. 2013;2:e00768.
91. Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. Nature. 2013;496:57–63.
92. McNulty SN, Tort JF, Rinaldi G, Fischer K, Rosa BA, Smircich P, et al. Genomes of *Fasciola hepatica* from the Americas Reveal Colonization with Neorickettsia Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers. PLoS Genet. 2017;13:e1006537.
93. Hahn C, Fromm B, Bachmann L. Comparative genomics of flatworms(platyhelminthes) reveals shared genomic features of ecto- and endoparasitic neodermata. Genome Biol. Evolution. 2014;6:1105–117.
94. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.
95. Peng J, Xu J. Low-homology protein threading. Bioinformatics. 2010;26(12): i294–i300.
96. Martin-Duran JM, Ryan JF, Vellutini B, Pang K, Hejnol A. Increased taxon sampling reveals thousands of hidden orthologs in flatworms. Genome Res. 2017;27(7):1263–272.
97. Rinaldi G, Eckert SE, Tsai IJ, Suttiprapa S, Kines KJ, Tort JF, et al. Germline transgenesis and insertional mutagenesis in Schistosoma mansoni mediated by murine leukemia virus. PLoS Pathog. 2012;8(7):e1002820.
98. Suttiprapa S, Rinaldi G, Tsai IJ, Mann VH, Dubrovsky L, Yan H bin, et al. HIV-1 integrates widely throughout the genome of the human blood fluke *Schistosoma mansoni*. PLoS Pathog. 2016;12(10):e1005931.
99. LOK JB, SHAO H, MASSEY HC, LI X. Transgenesis in *Strongyloides* and related parasitic nematodes: historical perspectives, current functional genomic applications and progress towards gene disruption and editing. Parasitology. 2017;144(3):327–42.
100. Free RB, Hazelwood LA, Sibley DR. Identifying Novel Protein-Protein Interactions Using Co-Immunoprecipitation and Mass Spectroscopy. Curr. Protoc. Neurosci. 2009;0 5(Unit-5):28.
101. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol England. 1990;215:403–10.
102. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics. 2010;11:431.
103. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME suite: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:202–08.
104. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
105. Hall T. BioEdit: a user-friendly biological sequence alignment editor andanalysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 1999:95–98.
106. 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;59:307–21.
107. Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol. 2006;6:29.
108. He Z, Zhang H, Gao S, Lercher MJ, Chen W-H, Hu S. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. Nucleic Acids Res. 2016;44(W1):W236–41.

ANEXO II

Durante el transcurso de la tesis colabore activamente con otros proyectos en curso del grupo, los que resultaron en la publicación de 4 artículos que se incluyen como anexos.

- Genomes of Fasciola hepatica from the Americas Reveal Colonization with Neorickettsia Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers.

  McNulty, Samantha N., Jose F. Tort, Gabriel Rinaldi, Kerstin Fischer, Bruce A. Rosa, Pablo Smircich, **Santiago Fontenla**, Young-Jun Choi, Rahul Tyagi, Kymberlie Hallsworth-Pepin, Victoria H. Mann, Lakshmi Kammili, Patricia S. Latham, Nicolas Dell'Oca, Fernanda Dominguez, Carlos Carmona, Peter U. Fischer, Paul J. Brindley, and Makedonka Mitreva.

  PLoS Genetics 13 (2017), e1006537.

  Este trabajo se realizó como parte de un vínculo de colaboración que mantiene nuestro grupo con el grupo de la Dra. Mitreva en la Universidad de Washington (Missouri, USA), para secuenciar y analizar genomas de FBT. Desde Uruguay aportamos muestras y participamos en el análisis de uno de los genomas de referencia de *F. hepatica.* Entre los hallazgos principales, describimos diversas adaptaciones como la expansión de familias génicas relevantes en el proceso de invasión y la reducción o pérdida de vías metabólicas que podrían estar vinculadas a la adquisición del modo de vida parasitario. Sin embargo, nuestro análisis de genómica comparada indico que *F. hepatica* y otros FBT parecen ser menos restringidos metabólicamente que schistosomas y cestodos.

  A su vez, detectamos una endobacteria asociada a los aislados de *F. hepatica* de Uruguay y Oregon, USA. La endobacteria detectada está cercanamente emparentada al género *Neorickettsia,* agente etiológico de la fiebre Sennetsu en humanos y de Potomac en caballos. Esta fue la primera descripción de interacción trans-reinos asociada a *F. hepatica,* y podría tener implicancias biológicas, evolutivas y etiopatogénicas en ambos organismos.

- Adaptive Radiation of the Flukes of the Family Fasciolidae Inferred from Genome-Wide Comparisons of Key Species

  Choi, Young-Jun, **Santiago Fontenla**, Peter U. Fischer, Thanh Hoa Le, Alicia Costábile, David Blair, Paul J. Brindley, Jose F. Tort, Miguel M. Cabada, and Makedonka Mitreva.

  Molecular Biology and Evolution 37 (2020), 84–99.

  Continuando con el vínculo de colaboración con el grupo de la Dra. Mitreva, publicamos los ensamblajes de los genomas de *F. gigantica* y *Fasciolopsis buski*, y realizamos un análisis de genómica comparada entre gusanos de la familia *Fasciolidae*: *F. hepatica, F. gigantica* y *F. buski*. Entre los resultados más relevantes reportamos la expansión del tamaño de los genomas en *Fasciola* respecto a otros trematodos. Estas expansiones genómicas no mostraron estar asociadas a un aumento en el número de genes, pero si a un aumento en el tamaño de los mismos y ocurrieron, sobre todo, por la acumulación de elementos transponibles y secuencias repetidas. Esta acumulación mostro ser linaje-especifica con la expansión preferencial de familias de TEs particulares entre los *Fasciolidae*. Detectamos expansiones diferenciales de familias génicas como Cistein-Catepsin proteasas, proteínas de unión a ácidos grasos, proteínas disulfide-isomerasas y chaperonas moleculares, que resaltan la importancia de las proteínas excretadas-secretadas para los gusanos hepáticos, y que parecen estar vinculadas a los mecanismos de invasión y/o al medio ambiente del órgano parasitado en el huésped definitivo por cada especie.

  Analizamos los tiempos evolutivos de ramificación de los distintos linajes e inferimos que podrían asociados a rápidos eventos de cambio ecológico o climático. Este estudio ha proporcionado nuevos conocimientos sobre la evolución del genoma de estos importantes patógenos, y ha generado recursos genómicos que permitirán el desarrollo de mejores herramientas diagnóstico y nuevos fármacos.

- Pleiotropic Alterations in Gene Expression in Latin American Fasciola Hepatica Isolates with Different Susceptibility to Drugs.

  Radio, Santiago, **Santiago Fontenla**, Victoria Solana, Anna C. Matos Salim, Flávio Marcos Gomes Araújo, Pedro Ortiz, Cristian Hoban, Estefan Miranda, Valeria Gayo, Fabiano Sviatopolk-Mirsky Pais, Hugo Solana, Guilherme Oliveira, Pablo Smircich, and José F. Tort.

  Parasites & Vectors 11 (2018), 56.

  Este trabajo fue realizado en el marco de un proyecto CABBIO, e involucro a grupos de Argentina, Brasil, Perú, México y Uruguay. En este estudio apuntamos a la caracterización del transcriptoma de dos aislados Sudamericanos resistentes a drogas para conocer los mecanismos moleculares putativos subyacentes a la resistencia. Se estudiaron tres aislados, la cepa Cajamarca aislada de Perú, resistente a triclabendazol y albendazol, y el aislado Rubino de Uruguay, resistente a albendazol pero sensible a triclabendazol, que fueron comparadas con una cepa completamente sensible a ambas drogas (aislado de Cenapa, Mexico). Entre los resultados más destacados está la identificación de variaciones en la expresión de proteínas del citoesqueleto, la adenilato ciclasa, entre otros lo cual es consistente con los blancos putativos reportados del triclabendazol y con posibles métodos de detoxificación.

  Estos resultados pusieron énfasis en el valor de las aproximaciones transcriptómicas para proveer de nuevas visiones apuntadas a mejorar el entendimiento de los modos de acción y los mecanismos de resistencia a las drogas, y serán de valor para el desarrollo de nuevas estrategias de tratamiento.

- Compositional Analysis of Flatworm Genomes Shows Strong Codon Usage Biases Across All Classes.

  Lamolle, Guillermo, **Santiago Fontenla**, Gastón Rijo, Jose F. Tort, and Pablo Smircich.

  Frontiers in Genetics 10 (2019), 771.

  Este trabajo fue parte de un esfuerzo realizado por bioinformáticos nacionales. Hicimos un análisis comparativo en base a los genomas de 22 especies representativas de los principales clados y estilos de vida del filo Platelmintos. Seleccionamos un conjunto de 700 genes ortólogos conservados en todas las especies, midiendo los cambios en el contenido de GC, el uso de codones y aminoácidos en posiciones ortólogas. Los valores de GC en la 3er posición del codón abarcó un amplio rango, lo que permitió discriminar dos grupos diferentes entre los turbelarios, cestodos y trematodos, respectivamente. A su vez, un agrupamiento jerárquico por uso de codones difirió notablemente del árbol filogenético. Además, detectamos un sesgo en el uso de codones sinónimos que fue más dramático en genomas extremadamente pobres o ricos en GC, es decir, los esquistosomas pobres en GC prefirieron usar codones sinónimos terminados en AT, mientras que *M. lignano* rico en GC mostró la conducta contraria. Curiosamente, estos sesgos afectaron el uso de aminoácidos, con el uso preferencial de aminoácidos codificados por codones siguiendo la tendencia del contenido de GC. Estos están asociados con sustituciones no sinónimas en posiciones ortólogas. El análisis detallado de los cambios sinónimos y no sinónimos proporcionó evidencia de un mecanismo de dos golpes en el que tanto las fuerzas de mutación como las de selección parecen impulsar las diversas estrategias de codificación de los platelmintos.

PLOS | GENETICS

RESEARCH ARTICLE

Genomes of *Fasciola hepatica* from the Americas Reveal Colonization with *Neorickettsia* Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers

Samantha N. McNulty[1], Jose F. Tort[2], Gabriel Rinaldi[3¤], Kerstin Fischer[4], Bruce A. Rosa[1], Pablo Smircich[2], Santiago Fontenla[2], Young-Jun Choi[1], Rahul Tyagi[1], Kymberlie Hallsworth-Pepin[1], Victoria H. Mann[3], Lakshmi Kammili[5], Patricia S. Latham[5], Nicolas Dell'Oca[2], Fernanda Dominguez[2], Carlos Carmona[6], Peter U. Fischer[4], Paul J. Brindley[3], Makedonka Mitreva[1,4]*

1 McDonnell Genome Institute at Washington University, St. Louis, Missouri, United States of America, **2** Departamento de Genética, Facultad de Medicina, Universidad de la República (UDELAR), Montevideo, Uruguay, **3** Department of Microbiology, Immunology and Tropical Medicine, and Research Center for Neglected Diseases of Poverty, School of Medicine & Health Sciences, George Washington University, Washington, DC, United States of America, **4** Division of Infectious Diseases, Department of Medicine, Washington University School of Medicine, St. Louis, Missouri, United States of America, **5** Department of Pathology, School of Medicine & Health Sciences, George Washington University, Washington, DC, United States of America, **6** Unidad de Biología Parasitaria, Instituto de Biología, Facultad de Ciencias, Instituto de Higiene, Montevideo, Uruguay

¤ Current address: Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, United Kingdom
* mmitreva@wustl.edu

OPEN ACCESS

Citation: McNulty SN, Tort JF, Rinaldi G, Fischer K, Rosa BA, Smircich P, et al. (2017) Genomes of *Fasciola hepatica* from the Americas Reveal Colonization with *Neorickettsia* Endobacteria Related to the Agents of Potomac Horse and Human Sennetsu Fevers. PLoS Genet 13(1): e1006537. doi:10.1371/journal.pgen.1006537

Editor: Eva H. Stukenbrock, Max-Planck-Institut fur Evolutionsbiologie, GERMANY

Received: June 9, 2016

Accepted: December 15, 2016

Published: January 6, 2017

Data Availability Statement: All sequence data from this project is available at NCBI. Raw reads from F. hepatica have been deposited in the GenBank sequence read archive (SRA), as BioProject PRJNA179522. Raw RNAseq reads from *F. hepatica* were submitted to the GenBank sequence read archive (SRA) under the same BioProject id, with the following accession numbers: SRX1037419, SRX1037422, SRX1037423, SRX1037421, SRX1037418. The *F. hepatica Neorickettsia* genome and predicted

Abstract

Food borne trematodes (FBTs) are an assemblage of platyhelminth parasites transmitted through the food chain, four of which are recognized as neglected tropical diseases (NTDs). Fascioliasis stands out among the other NTDs due to its broad and significant impact on both human and animal health, as *Fasciola sp.*, are also considered major pathogens of domesticated ruminants. Here we present a reference genome sequence of the common liver fluke, *Fasciola hepatica* isolated from sheep, complementing previously reported isolate from cattle. A total of 14,642 genes were predicted from the 1.14 GB genome of the liver fluke. Comparative genomics indicated that *F. hepatica* Oregon and related food-borne trematodes are metabolically less constrained than schistosomes and cestodes, taking advantage of the richer millieux offered by the hepatobiliary organs. Protease families differentially expanded between diverse trematodes may facilitate migration and survival within the heterogeneous environments and niches within the mammalian host. Surprisingly, the sequencing of Oregon and Uruguay *F. hepatica* isolates led to the first discovery of an endobacteria in this species. Two contigs from the *F. hepatica* Oregon assembly were joined to complete the 859,205 bp genome of a novel *Neorickettsia* endobacterium (*nFh*) closely related to the etiological agents of human Sennetsu and Potomac horse fevers. Immunohistochemical studies targeting a *Neorickettsia* surface protein found *nFh* in specific organs

features are available under accession RefSeq NZ_AGCN00000000.1. All other relevant data are within the paper and its Supporting Information files.

Funding: Sequencing of the genomes of Fasciola/Neorickettsia was supported by the 'Sequencing the etiological agents of the Food-Borne Trematodiases' project (NIH-NHGRI award number U54HG003079). The Barnes Jewish Hospital Foundation supported the immunolocalization studies. The Comisión Sectorial de Investigación Científica (CSIC-UDELAR) financially supported the Uruguayan co-authors. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

and tissues of the adult trematode including the female reproductive tract, eggs, the Mehlis' gland, seminal vesicle, and oral suckers, suggesting putative routes for fluke-to-fluke and fluke-to-host transmission. The genomes of *F. hepatica* and *nFh* will serve as a resource for further exploration of the biology of *F. hepatica*, and specifically its newly discovered trans-kingdom interaction with *nFh* and the impact of both species on disease in ruminants and humans.

Author Summary

This report presents novel findings revealing (a) the genome sequence of the food-borne trematode *Fasciola hepatica* (the liver fluke) isolated from sheep, which stands out among neglected tropical diseases due to its zoonotic impact on both human and animal health and (b) the first instance (and the genome) of the rickettsial endobacterium of the genus *Neorickettsia* in *F. hepatica*. Using stage-specific gene expression data, we identified liver fluke proteins likely involved in host-parasite interactions, and using immunolocalization, we confirmed *Neorickettsia* in organs and tissues of the adult trematode. The presence of the bacteria in fluke reproductive tissues and eggs suggests a possible mechanism for vertical transmission, and the presence of bacteria in the oral sucker used to anchor flukes to the lining of the biliary tract suggests a potential mechanism for horizontal transmission to the mammalian host. This is of interest because related *Neorickettsia* cause severe, even deadly, illness in a variety of species, including humans. This is the first report to localize *Neorickettsia* endobacteria within the tissues of adult *F. hepatica*. The discoveries in our manuscript have wide impact for the fields of both the pathophysiology and evolution of *Fasciola* and related FBTs, and the transmission strategies of *Neorickettsia*.

Introduction

Food borne trematodes (FBTs) are an assemblage of platyhelminth parasites that are transmitted through the food chain [1]. Among the four major groups of FBT infections recognized as neglected tropical diseases (NTDs) by the World Health Organization [2], fascioliasis stands out due to its zoonotic impact on both human and animal health [3]. *Fasciola* species are major pathogens of domesticated ruminants, but they infect numerous other species of mammals, including people [4]. Due to the significant burden to livestock globally, with annual losses exceeding US $3.2 billion [5] and public health with ~50 million infected people [4], these parasites are among the most-extensively studied FBTs.

Like other digenetic trematodes, *Fasciola hepatica* has a complex developmental cycle [1]. The hermaphroditic adult stage resides in the host bile ducts and reproduces sexually, releasing thousands of eggs each day that pass with the bile into the intestines and exit in the fecal stream. Eggs that reach fresh water embryonate over a couple of weeks, hatching a free-swimming miracidium that seeks out and infects a snail of the family Lymnaeidae. Within the snail, the parasite progresses through sporocyst, redia, and daughter redia stages by asexual replication and development, resulting in the release thousands of the cercariae [6]. The free-living, aquatic cercaria encysts as the metacercarial stage on solid substrates, including vegetation at the margins of the watercourse. When infected vegetation (for example, uncooked watercress) are ingested by a suitable host, the metacercaria excysts in the duodenum, transverses the wall of the small intestine, migrates through the peritoneal cavity, and penetrates the Glisson's

capsule of the liver [7]. The migration of the juvenile fluke though the liver parenchyma into the biliary ducts damages the liver and provokes reactions associated with the acute phase of the infection. This phase is accompanied by systemic disease including fever, nausea and abdominal pain. Once the adult is established in the bile ducts, anemia, inflammation, fibrosis, cholangitis and biliary stasis may ensue. In this chronic phase adult worms can survive several years in the absence of intervention [8, 9]. Despite its potent and broad action against other human parasitic flatworms the anthelmintic drug praziquantel has no effect on *F. hepatica* [10]. Triclabendazole (TCBZ) is the drug of choice since its effective against juveniles and adult liver flukes, but resistance to this benzimidazole has emerged in livestock in different countries [11]. There have been recent reports of human fascioliasis refractive to TCBZ treatment in Peru and Chile [12, 13], highlighting a need for alternative drugs and treatments.

In addition to being important pathogens themselves, some digeneans serve as vectors of bacterial pathogens. *Neorickettsia* (family Anaplasmataceae) belongs to a poorly characterized assemblage of obligate, intracellular α-Proteobacteria associated with serious, even fatal disease in mammals [14]. These bacteria can be horizontally transmitted from the fluke to host tissue invading and multiplying within mammalian cells such as macrophages, monocytes and other cells types, e.g. intestinal epithelium, eventually leading to severe disease. *Neorickettsia* can be detected by PCR in trematodes spanning the major lineages of the Digenea [15, 16], but it has never been reported from a trematode that is itself a prevalent human and livestock pathogen. Furthermore, the fact that *Neorickettsia* is not found among all fluke species (or all members of infected species) suggests that these endobacteria are not essential to fluke survival. Indeed, the exact nature of their relationship is remains unclear.

Here we describe the second reported reference genome of the common liver fluke, *F. hepatica* and the first discovery and genome sequences of the *Neorickettsia* endobacteria of *F. hepatica*. In contrast to the previously sequenced isolate from cattle from the UK, the presently described strain, taken from a sheep in Oregon, US, was infected with a *Neorickettsia* species closely related to the etiological agents of Potomac horse and Sennetsu fevers. Histological, PCR, and gene sequence analyses revealed its presence in tissues of liver fluke isolates from Oregon, and in one of several liver fluke isolates from Uruguay that were screened. Taken together, these genomes represent a benchmark resource for studies of trematode and *Neorickettsia* biology, pathogenesis and evolution.

Results and Discussion

The general features of the nuclear genome of *Fasciola hepatica*

The nuclear genome of *F. hepatica* Oregon was sequenced and assembled with a total length of 1.14 Gb, N50 number of 2,036 and N50 length of 161 kb (S1 Table). Completeness was estimated at 90.6% using the CEGMA method [17]. GC content was similar to other Food Borne Trematodes (FBTs) including *Clonorchis sinensis* [18] and *Opisthorchis viverrini* [19], but differed from blood flukes. Intriguingly, the genome of *F. hepatica* Oregon had a markedly higher repeat content (55.29%, S1 Table) than other FBTs, including the recently published genome of *F. hepatica* United Kingdom (32.0%) isolated from cattle. We detected > 92 Mb corresponding to LTR elements, 268 Mb corresponding to LINEs, and 235 Mb of unclassified repetitive sequences, values all higher than other trematodes. Functional RNAs including rRNA, tRNA and miRNA (S2 Table) were identified, representing 0.002% of the coding genome, most supported by RNAseq data.

Consistent with other FBTs [19, 20], a very small percentage of the genome assembly was predicted to encode proteins (1.08%, considering only exonic regions). A total of 14,642 protein-coding genes were identified using a combination of *de novo* and evidence-based

methods. Predicted genes had an average of 3.3 exons and 2.3 introns, average footprint of 3,078 bp, and average coding length of 837 bp (S1 Table). Comparisons of protein-coding genes between *F. hepatica* Oregon and *F. hepatica* UK (S1 Fig) revealed that most functional elements (e.g. KEGG orthologous groups) were shared, despite the fact that the gene models showed relatively poor overlap. In general, both genome annotations showed a predominance of short genes compared with other trematodes. This could be an indication of incomplete gene models in both assemblies, as the size, complexity, and incompleteness of both hindered gene prediction. Accordingly, long reads from third generation sequencing and additional RNAseq data will be needed to improve gene predictions, as demonstrated for *S. mansoni* [21] and *C. sinensis* [20].

Comprehensive functional annotation of the deduced proteins of *F. hepatica* Oregon is provided in S3 Table, including (a) 3,907 unique InterPro protein domains predicted from 8,609 proteins, associated with 1,147 unique gene ontology (GO) terms, (b) 3,175 proteins associated with 2,685 KEGG orthologous groups, (c) 339 proteins classified as putative proteases, (d) 65 proteins classified as protease inhibitors, and (e) 855 of proteins predicted to be secreted. Majority of the genes (94% of the predicted 13,740/14,642) were supported by RNAseq data from the developmental stages sampled for this study (eggs, metacercariae, and adult flukes; S2 Fig). Of the >6,000 genes expressed in these stages, ~2,500 showed no differential expression, with GO terms related to core cellular functions such as translation, RNA processing, and vesicular transport (S3 Table).

Among the differentially expressed gene sets resulting from the DESeq analysis, stage-specific overexpressed gene sets were identified (e.g., for metacercariae, genes significantly overexpressed in the metacercarial stage relative to the adult and to the egg, but not differentially expressed between adult and egg). Using these criteria, four of the top five most significantly metacercaria-overexpressed genes (2,076 total) were cysteine proteases, including four papain-like family proteases and one C13-family protease ($P < 10^{-15}$ for all comparisons), while the most significantly adult-overexpressed was also a papain-family cysteine protease ($P < 10^{-38}$). Among the other 1,169 adult-overexpressed genes, four of the top 11 were tubulin genes ($P < 10^{-20}$). Fewer genes (259) were egg-overexpressed since adult females contain eggs expressing transcripts, but the most significantly differentially expressed gene in this set ($P < 10^{-13}$) was a glucose-6-phosphate dehydrogenase.

Comparative analysis reveals phylogenetic conservation and diversification among trematodes

Approximately 88.5% of the 14,642 inferred proteins from the *F. hepatica* Oregon isolate found at least one BLAST hit ($E < 1e^{-05}$) to non-*Fasciola* proteins in the non-redundant database (NR), with most matching sequences from other FBTs (particularly the liver flukes *C. sinensis* and *O. viverrini*). 11.5% of genes are *Fasciola*-specific with respect to NR (S3 Table), 4.9% of which were assigned additional functional annotations (Interpro domains, GO terms or KEGG orthologous groups; compared to 67.7% for genes with non-*Fasciola* NR matches). The putative *Fasciola*-specific genes were enriched for GO terms related to cysteine-type endopeptidase inhibitor activity and neurotransmitter secretion (S4 Table).

Protein conservation among flatworm parasites and their hosts was analyzed by clustering predicted proteins from 10 genomes into orthologous protein families (OPFs; Fig 1A), and 7,624 *F. hepatica* proteins were included in 5,721 unique OPFs with proteins from other flatworms including the free-living planarian *Schmidtea mediterranea*, trematodes (including schistosomes, the liver flukes *C. sinensis* and *O. viverrini*), and mammalian hosts (human, cow and sheep). Some 2,875 *F. hepatica* proteins (1,451 OPFs) were conserved across the 10 species,

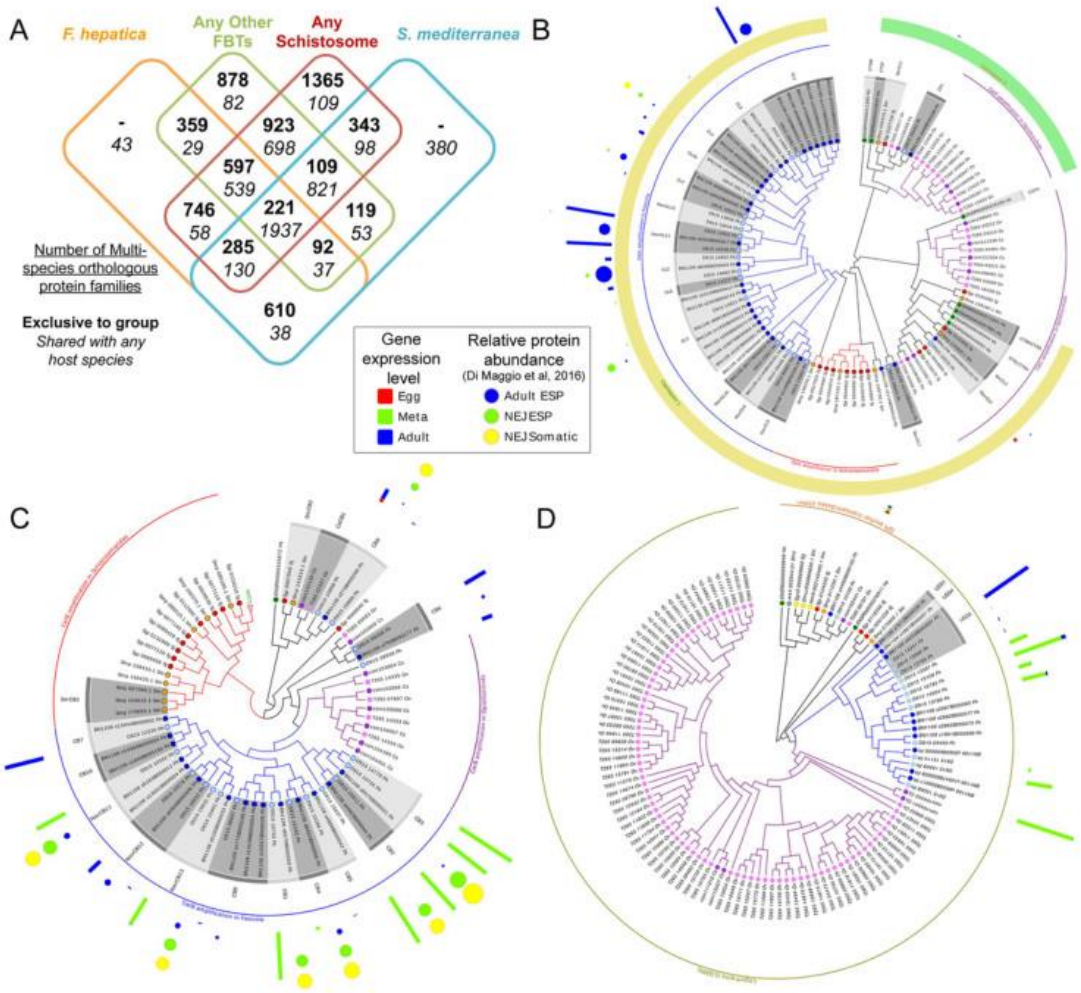


Fig 1. Protein families in trematodes. (A) A Venn diagram demonstrating the phylogenetic distribution of orthologous protein families among the trematode species analyzed (B) Differential amplification of cathepsins L and F in trematodes. A maximum likelihood tree of the genes annotated as members of the C1A protease family from trematodes shows that while a single set of cathepsins F are detected in Schistosomes and *F. hepatica* strains, an expansion of cathepsins F is observed in *O. viverrini* and *C. sinensis* (green arch). Similarly, the cathepsins L-like of the trematodes (yellow arch) show a basal node more related to the mammalian enzymes and several independent amplifications in schistosomes, opistorchiids and *Fasciola*. Most known cathepsins variants are supported by expression data in different stages (red, green and blue bars), and proteomic data (yellow, green and blue dots) from a recent report [34]. Several putative novel variants are indicated, most of them not expressed at the stages analyzed. (C) Cathepsin B subfamily of the C1A protease family. A basal cathepsin B node and differential expansions in schistosomatids (red arch), fasciolids (blue arch) and opisthorchiids (purple arch) is observed. As in cathepsins L novel isoforms are identified, some of them supported by expression data. (D) Legumain is differentially amplified in food borne trematodes. Maximum likelihood tree of the genes annotated as members of the C13 protease family. While a single gene corresponding to the glycosyl-phosphatidyl-inositol anchor transamidase exists in all trematodes, amplification of the legumains was evident in the food borne trematodes *F. hepatica*, *C. sinensis and O. viverrini*, but not in the blood flukes *S. mansoni* and *S. japonicum*. These amplification events are independent in the different lineages. Species included in the tree are color coded (human, emerald dots, planaria (grey) cestodes (yellow) *S. japonicum* (dark red) *S. mansoni* (orange) *C. sinensis* (purple) *O. viverrini* (pink), *F. hepatica* Oregon strain (sky blue) *F. hepatica* Liverpool strain (navy).

doi:10.1371/journal.pgen.1006537.g001

and these were more likely than other genes to be differentially expressed across developmental stages of *F. hepatica* ($P = 7 \times 10^{-5}$, binomial distribution test; S3 Table). In contrast, 393 *F. hepatica* genes (359 OPFs) were conserved between *F. hepatica* and at least one other FBT (*C. sinensis* and/or *O. viverrini*); these were significantly enriched for GO terms related to microtubule-based processes, cysteine endopeptidase activity and pH regulation (S4 Table). An additional 29 genes (29 OPFs) were conserved with at least one FBT and at least one host species; these were significantly enriched for the GO terms related to phagocytosis and L-ascorbic acid binding.

OPF analysis also enabled the identification of gene sets specific to schistosomes; 1,365 OPFs conserved among at least two of the three species of *Schistosoma* were analyzed. Like the FBTs, these were significantly enriched for GO terms related to microtubule activity and cysteine endopeptidases (see below). This suggests that certain functions are conserved across the platyhelminth clades despite clear divergence at the sequences level, a possible indication of rapid evolution [22].

Expanded families of secreted/excreted proteases suggest a role in host-parasite interaction

Excreted and secreted proteins (ESPs) play a crucial role in parasitism. We identified 855 (5.8%) proteins with computationally predicted signal peptides but no transmembrane domains (S3 Table), indicating they may be secreted/ excreted. These proteins were significantly enriched for GO terms related to proteolysis (S4 Table). Again, the most significantly enriched molecular process GO term was "cysteine-type endopeptidase activity" (GO:0004197). Secreted cysteine proteases have a well-defined role in the biology of *F. hepatica* and liver fluke disease [23]. Cathepsin L's are predominant in adult ESPs, where they participate in feeding, immune evasion and immune modulation. Distinct suites of cathepsin L's and cathepsin B's are abundant in the juvenile fluke, participating in excystment, migration through gut wall and liver capsule, and immune evasion [24, 25]. Although it was known that liver fluke cathepsins constitute a multigene family [26], the complexity and diversity within the family is now apparent. In addition to the six known cathepsin L's, other isoforms were detected consistently in both the Oregon and the UK isolates, raising the total count to 14; most of these overexpressed in the adult stage (Fig 1B). Independent amplifications of cathepsin L's occurred in schistosomes and the opisthorchiids, but the resulting gene copy number is less than in *F. hepatica*. In contrast, Cathepsin F's showed a divergent pattern in these lineages, with single enzymes in *Fasciola* and schistosomes, and an amplified family in the carcinogenic fish-borne liver flukes (Fig 1B). A similar pattern of independent amplifications among trematodes was observed for cathepsin B's (Fig 1C), again with a distinct expansion in *F. hepatica*. In contrast to cathepsin Ls, cathepsin Bs were overexpressed in metacercariae (MC), confirming biochemical, genetic and proteomic evidence of differential expression along the life cycle [27]. Interestingly, within both the cathepsins L and B, a clade comprising a single enzyme from each trematode species and vertebrates was identified, which might be basal to all the lineage-specific expansions. *F. hepatica* enzymes of this clade have not been described yet, and, notably, they are expressed in eggs (Fig 1B and 1C).

The remarkable amplification and diversity of secreted cysteine proteases in trematode lineages suggests key roles during parasite adaptation. Diverse trematodes express different (and amplified) subfamilies of cathepsins, reflecting their host parasite relationships, including host niche, organ sites, and transmission strategies. For example, cathepsins B and an L3 (CL3) participate in transit of the juvenile liver fluke through the gut wall with collagenolysis [28, 29], whereas juvenile of the fish-borne liver flukes ascend into the biliary tree through the ampulla

of Vater [30]. In turn, cathepsins F and the aspartic protease cathepsin D are characteristically overrepresented in these carcinogenic liver flukes [19, 31]. The blood flukes, on the other hand, invade the skin of their hosts, with conserved serine proteases essential to this process, with the cysteine protease cathepsin B providing critical activities in some species [32].

Asparaginyl endopeptidases, Class C13 (also known as legumain) [33] were expanded in *F. hepatica* with ≥10 members, and were also differentially expanded among the trematodes; 3 copies in *S. mansoni*, 5 in *S. japonicum*, 4 in *C. sinensis* and ~ 100 in *O. viverrini*. (Fig 1D). These proteases might participate in the activation of cathepsins and the digestion of infected host tissues, liberating essential amino acids.

A recent proteomic study of ESPs from juvenile and adult *F. hepatica* provide further support for the differential expression of these gene families [34], confirming our transcriptomic data (S3 Table). For example, within the cathepsin B family, members of 10 out of 13 clusters are detected by LC-MS/MS in ESPs; while 3 isoforms are exclusively expressed by adults, 2 are characteristic of juveniles (also detected by RNAseq in metacercariae), and 5 are expressed in both stages but clearly predominant in juveniles (Fig 1C). Similarly, a predominance of expression of cathepsin Ls variants in adults is observed at proteomic level consistent with our transcriptomic data (Fig 1B). Within these some of the novel clades here described were detected as being expressed.

FBTs are metabolically less constrained than blood flukes and cestodes

Metabolic pathways predicted in the *F. hepatica* Oregon strain were compared to those of other sequenced flatworms. All parasitic flatworms showed a significant reduction in metabolic capabilities compared to free living platyhelminth species, including planaria (Fig 2). As shown previously [35], parasitic flatworms depend on the hosts for provision of fatty acids. Unlike the blood flukes, however, *F. hepatica* and the other liver flukes possess enzymatic pathways for fatty acid elongation by reversal of beta-oxidation (S3A Fig) and fatty acid catabolism (S3B Fig), allowing them to take advantage of the fatty acid rich environment of bile.

Additional differences between blood and liver flukes were evident in amino acid metabolism. Inabilities to synthesize several amino acids were generally observed in neodermatan flatworms, including *Fasciola*. However, the liver flukes operate a complete catabolic pathway of aliphatic amino acids, and enzymes of these pathways (e.g. branched chain amino acid aminotransferase [BCAT, EC. 2.6.1.42]) are missing in schistosomes (S3C Fig). Aliphatic amino acids are more abundant in the bile than in blood, which may have been exploited since it facilitates access to protein synthesis precursors and alternative energy sources. While expanded families of secreted proteases are involved in protein digestion, conserved oligopeptide transporters that mediate the uptake of di- and tri-peptides in metazoans were not identified [36], suggesting the lack of these specific transporters in trematodes at large. Consequently, protein digestion up to individual amino acids may occur extracellularly, explaining the rare presence of usually cytoplasmic enzymes as leucine aminopeptidase in the secreted products and vesicles released by the parasite [37].

Further metabolic differences may have evolved in liver flukes compared to blood flukes in relation to an environment characterized by low oxygen tension. Flukes switch from aerobic to anaerobic metabolism in the low oxygen environment of the bile duct, but instead of fermenting carbohydrates to lactate, the parasite exploits the more energy-efficient malate dismutation pathway [38], where phosphoenolpyruvate from glycolysis is converted to oxaloacetate via the phosphoenolpyruvate kinase (PEPCK), and further reduced to malate. After entering the mitochondria, some malate is oxidized to acetate, and some is reduced to succinate and transformed to propionate, in a series of reactions that reverts the Krebs cycle (Fig 3A),

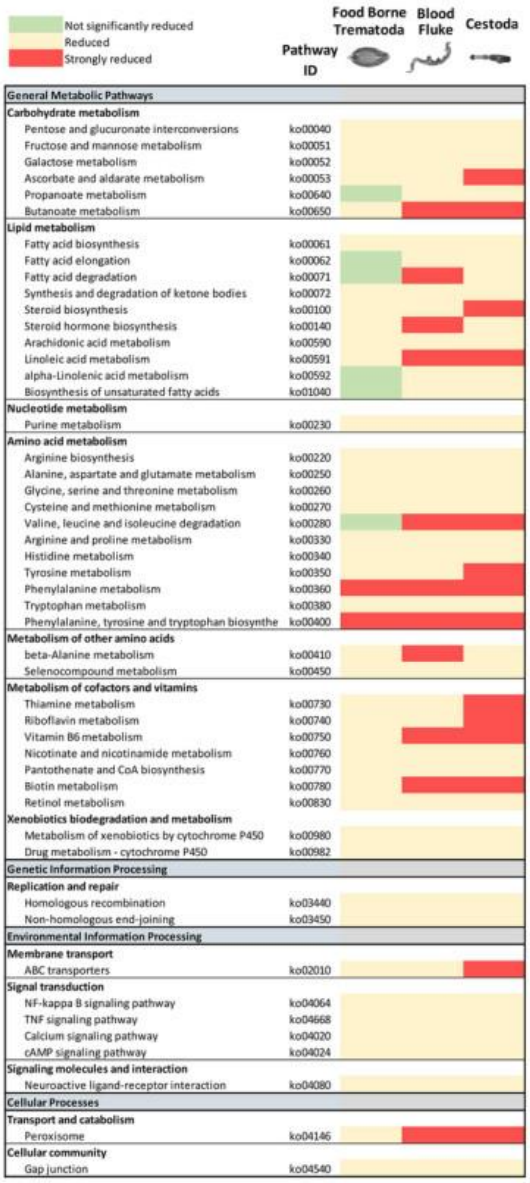


Fig 2. Metabolic pathway reduction in parasitic flatworms. The global number of proteins assigned to different metabolic pathway was compared between different parasitic flatworms, the free-living planaria and cattle predicted proteomes. Those pathways showing a significant reduction in the parasitic species in relation to those present in planaria are indicated (strongly reduced: less than 30% conservation; reduced: conservation between 30 and 80%; not significantly reduced: more than 80% conservation). While in general

several pathways are reduced in all parasitic species, some pathways are differential between food borne trematodes (*F. hepatica*, *O. viverrini*, *C. sinensis*), blood flukes (*S. mansoni* and *S. japonicum*) and cestodes (*E. multilocularis*, *H. microstoma* and *T. solium*). Most notably some lipid metabolism and amino acid pathways (i.e. aliphatic amino acid degradation) are not reduced in FBT (green) while they are reduced in the other groups. In general, FBT seem to be less constrained than blood flukes, with cestodes being the most restricted metabolically.

doi:10.1371/journal.pgen.1006537.g002

providing a source of electrons for the respiratory chain finally yielding five ATP molecules per glucose molecule. While the whole pathway was precisely described biochemically, we now for the first time identify the cognate enzymes (Fig 3A).

To provide further insight into the metabolism of *F. hepatica*, we compared the metabolic pathway modules present and complete in *F. hepatica*, the carcinogenic liver flukes *C. sinensis* and *O. viverrini*, the blood fluke *S. mansoni* and mammalian hosts of *F. hepatica* (sheep, cow and human) were incorporated in the analysis (S4A Fig). Twenty-five KEGG pathway modules were complete in our assembly (i.e., contain the complement of KO's necessary to convert the initial substrate to the final product based on strict completion [39]). These values are similar in other trematodes, but far lower than in mammals. Use of a lenient completion criterion of $\leq$2 missing steps in a module extended these values, with similar trends (S4B Fig). The analysis also identified modules that differ between the liver flukes and *S. mansoni*. Module M00020 (serine biosynthesis) also showed differences consistent with those observed in amino acid metabolism. Further differences in inositol phosphate metabolism were detected with two steps missing in *S. mansoni* (R03427 and R04372, INPP1 and INPP4 phosphatases). Module M00087 (beta-oxidation) occurs in liver flukes but is absent from schistosomes. Notably, this module revealed differences with the host, since in step 2 (R04738), *F. hepatica* shares two KOs with mammals, but there is an additional, putative platyhelminth-specific ortholog (K01692, EC 4.2.1.17) corresponding to enoyl coA-hydratase, which warrants investigation in flatworms. Differences between liver fluke and mammal were evident in module M00009m, corresponding to the tricarboxylic acid cycle (Fig 3B). The fumarate forming reaction (R01082) is dependent on fumarate hydratase class II enzyme (EC 4.2.1.2B, K01679) by the hosts, while a second fumarate hydratase, class I enzyme for this step (EC 4.2.1.2A, K01676), was detected in trematodes, an enzyme that might participate in the reverse step of the malate dismutation pathway (above).

A complete *Neorickettsia* genome identified in *F. hepatica* Oregon and Uruguay

The most striking feature of the *F. hepatica* Oregon isolate was an apparent infection with *Neorickettsia* endobacteria (*nFh*). Alpha-protobacterial sequences were first identified among "contaminating" sequences in the *F. hepatica* genome and the presence of *Neorickettsia* was confirmed and validated by both PCR and 16S rRNA sequencing (S5 Fig). The genome of *nFh* was assembled from 241,957 2x100bp read pairs that were identified during the sequencing of *F. hepatica* Oregon. A single 859,205 bp scaffold with average 56.3x sequence coverage was constructed from two contigs joined by 189 bp of inferred gaps (Fig 4). This novel *Neorickettsia* genome was similar in size and GC content to those of previously sequenced *Neorickettsia* species [40, 41]. Full genome alignments indicated that, with the exception of a small inversion, it shared nearly complete synteny with the genomes of *Neorickettsia risticii* and *N. sennetsu* (Fig 5A). Synteny among *Neorickettsia* species may reflect the lethality of large genome rearrangements due to a reduced set of DNA repair genes, but this may have increased the genetic variation in a stable intra-trematode environment by accumulation of mutations in non-essential genes [41].

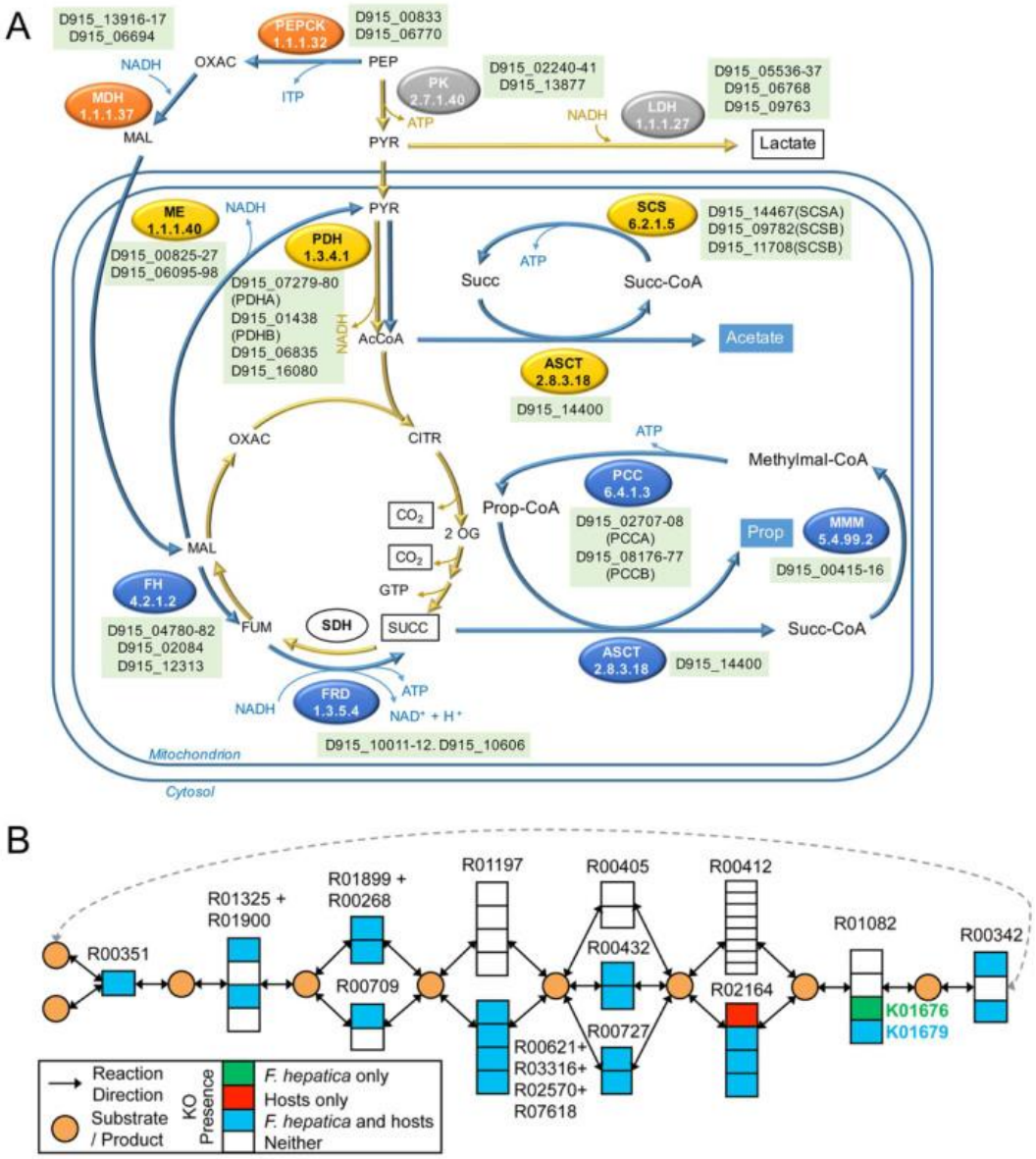


Fig 3. Metabolic pathways in *F. hepatica*. (A) Energy metabolism in anaerobic mitochondria of *F. hepatica* by malate dismutase. While the classical anaerobic fermentation to lactate is present (enzymes indicated in gray) when oxygen tension is low, the malate dismutation pathway is preferred (blue arrows). Phosphenol pyruvate reduction to malate occurs in the cytoplasm (orange). Within the mitochondria, part of the malate is oxidized to acetate (yellow) while other fraction reduced to succinate and further transformed to propionate (blue). Genes predicted for key enzymes involved in anaerobic respiration are indicated. Abbreviations: PEP, phosphenolpyruvate; OXAC, oxaloacetate; MAL, malate; FUM, fumarate; SUCC, succinate; PYR,

pyruvate; AcCoA, acetyl-CoA; CITR, citrate. Enzymes indicated are: PK, pyruvate kinase; LDH, lactate dehydrogenase, PEPCK, phosphenolpyruvate carboxykinase (ATP dependent); MDH, malate dehydrogenase; ME, malic enzyme; PDH, pyruvate dehydrogenase; ASCT, acetate:succinate CoA-transferase; SCS, succinyl-CoA synthetase; FH, fumarate hydratase; FRD, fumarate reductase; SDH, succinate dehydrogenase, MMM methylmalonyl-CoA mutase; PCC, propionyl-CoA carboxylase. (B) Parasite specific enzyme usage in TCA cycle. The KEGG module for TCA cycle (M00009) is shown with groups of orthologous enzymes indicated using KEGG orthology (KO) IDs. An interesting example of alternate enzyme usage is shown for fumarate hydratase (reaction R01082), catalyzed by a class II enzyme (EC 4.2.1.2B; K01679) in both the host and parasite. However, *F. hepatica* also has a Platyhelminthes-specific class I fumarate hydratase (EC 4.2.1.2A; K01676), not annotated in any of the host proteomes. Such knowledge can be leveraged to design worm specific therapies with potentially low (or no) impact on the host health.

doi:10.1371/journal.pgen.1006537.g003

Table 1 outlines the inferred features of *nFh*. Similar to related species, *nFh* encodes 33 tRNA genes and one copy each of 5S, 16S, and 23S rRNA genes. A total of 744 protein-coding genes were predicted, slightly fewer than *N. risticii* and *N. sennetsu* (Table 1). Gene conservation analysis among representative bacterial species of the Anaplasmataceae identified three orthologous protein families (OPFs) that were conserved in all analyzed species except *nFh* (S5

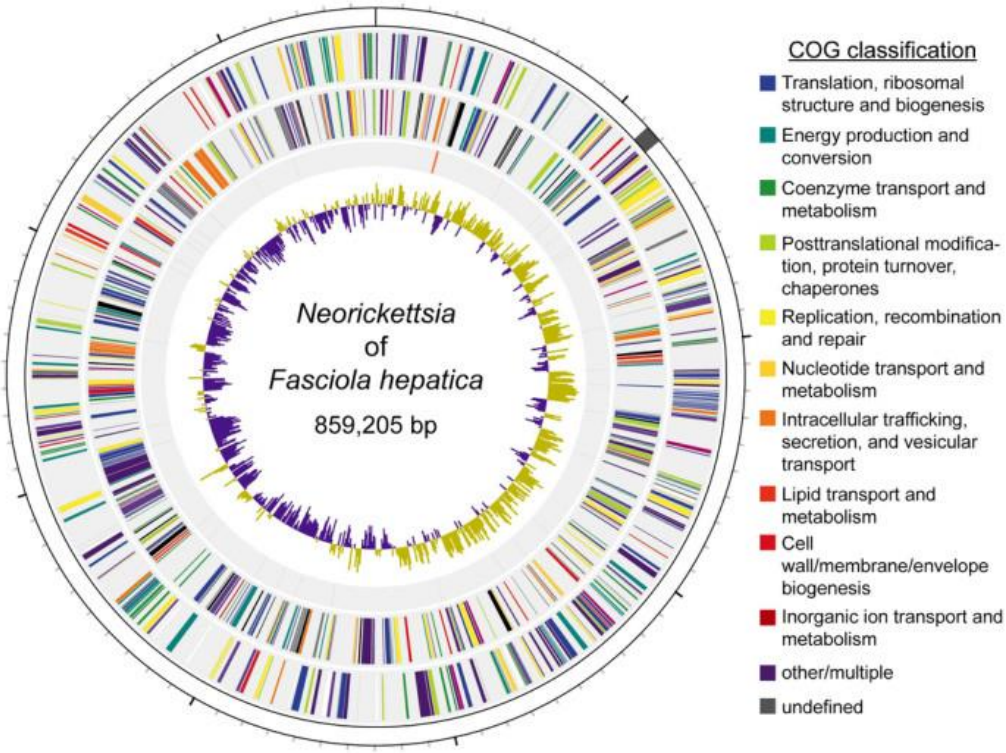


Fig 4. The genome of the *Neorickettsia* endobacterium of *Fasciola hepatica*. The first track (from outside to inside) represents the 859,205bp genome of the *Neorickettsia* endobacterium of *Fasciola hepatica* (100 kb major ticks, 10 kb minor ticks). The genome shows nearly complete synteny with *Neorickettsia risticii* and *Neorickettsia sennetsu* with the exception of an inversion (shaded in grey). The second and third tracks represent the 744 inferred protein coding genes on the plus and minus strands, respectively. Genes are coded based on their NCBI Clusters of Orthologous Groups of proteins database classification. The fourth track represents RNA coding genes, including 3 ribosomal (red), 33 transfer (black), and 1 short, noncoding (blue). The fifth track depicts the G-C skew [(G-C)/(G+C)] calculated over 500bp windows.

doi:10.1371/journal.pgen.1006537.g004

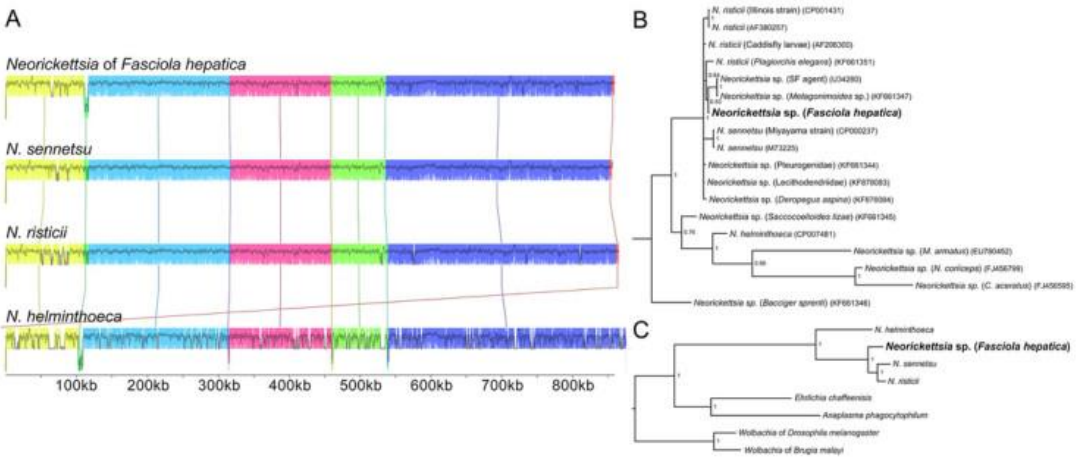


Fig 5. Phylogenetic affinities of the *Neorickettsia* symbiont of *Fasciola hepatica*. (A) The genomes of the four fully sequenced *Neorickettsia* species were aligned. Syntenic blocks are colored. The genome of the *Neorickettsia* of *Fasciola hepatica* (PRJNA295290) shares almost complete synteny with the genomes of *Neorickettsia risticii* (PRJNA19099) and *Neorickettsia sennetsu* (PRJNA357), with the exception of a small inversion that is also present in *Neorickettsia helminthoeca* (PRJNA187358). (B) Bayesian inference phylogenetic analysis based on available 16S ribosomal RNA sequences of *Neorickettsia* species, retrieved from [15]. While the resolution of the sub-clade consisting of *Neorickettsia risticii* and *Neorickettsia sennetsu* is sub-optimal, the tree indicates that the *Neorickettsia* of *Fasciola hepatica* (*nFh*) may be most closely related to a strain that occurs in *Metagonimoides* species and an agent of Sennetsu fever. (C) A Bayesian inference phylogenetic tree based on the protein sequences of 473 single-copy, orthologous protein families conserved in the represented species clearly indicates that *nFh* is more closely related to *N. risticii* and *N. sennetsu* than *N. helminthoeca* or other species of the family Anaplasmataceae. NCBI GenBank accession numbers are indicated. Trematode hosts or other defining features are indicated for uncharacterized species.

doi:10.1371/journal.pgen.1006537.g005

Table 1. The sequenced genomes of *Neorickettsia* species.

	Neorickettsia* of *F. hepatica	***Neorickettsia sennetsu***	***Neorickettsia risticii***	***Neorickettsia helminthoeca***
Trematode host	*F. hepatica*	Unknown	*Acanthatrium orgonense*	*Nanophyetus salmincola*
Vertebrate host	Unknown	Human	Horse, bat	Canids
Disease	Unknown	Sennetsu fever	Potomac horse fever	Salmon poisoning of dogs
RefSeq Accession	NZ_AGCN 00000000.1	NC_007798.1	NC_013009.1	NZ_CP007481.1
Assembly size	859,205 bp	859,006 bp	879,977 bp	884,232 bp
Length of inferred gaps	189 bp	—	—	—
GC content	41.4%	41.1%	41.3%	41.7%
rRNA	3	3	3	3
tRNA	33	33	33	33
ncRNA	1	3	1	1
Pseudo genes	8	2	11	17
Protein coding genes	744	753	759	772
Average CDS length	968 bp	966 bp	962	970
Minimum CDS length	4,761 bp	5,766 bp	4,761 bp	4,776 bp
Maximum CDS length	156 bp	156 bp	147 bp	135 bp
% coding	83.8%	84.7%	82.9%	84.6%

doi:10.1371/journal.pgen.1006537.t001

Table). Closer inspection revealed that these genes may be present and intact, although absent from the gene calls. Two additional OPFs were conserved in all sequenced *Neorickettsia* except *nFh* (S5 Table), but in both of these cases the corresponding sequences were identified but stop codons appeared to disrupt them.

S5 Table presents a functional annotation of the 744 predicted proteins of *nFh*, including (a) 1,453 unique InterPro protein domains predicted from 620 proteins and associated with 596 unique gene ontology (GO) terms, (b) 720 proteins associated with 509 KEGG orthologous groups, further binned into 120 enzymatic pathways and 101 pathway modules, (c) 25 proteins classified as putative proteases, (d) two protease inhibitors, and (e) 25 proteins with secretion signals (which were enriched for biological process GO terms related to proteolysis and protein transport; S4 Table). Protein transporters such as porins, identified in previous proteomic studies might transport nutrients from the host cytoplasm [42]. Whether *Neorickettsia* enzymes interact with those of the fluke is a fascinating but unresolved question. Revealingly, however, *N. risticii* synthesizes nucleotides, vitamins, and cofactors that the fluke cannot, raising the possibility that they may be harvested by the trematode for their mutual advantage [41].

While sequencing reads from the previously reported UK strains (SRA Project ID: ERP006249) did not map to our *nFh* genome (Fig 6), suggesting that no *Neorickettsia* DNA was present in the samples, one of our Uruguay isolate (out of five that were screened) tested positive for the presence of *Neorickettsia* by 16S rDNA PCR (S5 Fig). Whole genome sequencing of this sample recovered the genome of *nFh* (99.9% breadth of genome coverage; S6 Table), allowing a comparative analysis of sequence variation in both the *Neorickettsia* and the fluke genomes. In total, 15 single nucleotide variants were identified between the two *nFh* genomes, 11 of which occurred within the coding regions (7 non-synonymous and 4 synonymous variants; S7 Table). Notably, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC 2.7.7.60; AS219_00645), a key enzyme in the MEP pathway of isoprenoid biosynthesis [43], was found among the genes that harbored non-synonymous SNPs. It has been hypothesized that the *Wolbachia* endosymbionts of *Brugia malayi* and *Dirofilaria immitis* rely on their helminth host for the completion of the MEP pathway [44, 45]. Additionally, in many pathogenic and opportunistic bacteria, the MEP pathway intermediate (HMB-PP) has the capacity to modulate vertebrate host immune response [46], suggesting an interesting possibility of the involvement of the isoprenoid biosynthesis pathway in the host-parasite-endosymbiont interactions. Overall, the genetic distance (1-ibs, identity by state) between the Oregon (US) and Uruguay (UY) *nFh* genomes (1.75×10^{-5}) was three orders of magnitude lower than that estimated between the respective nuclear genomes of *F. hepatica* (1.08×10^{-2}; S8 Table). The observed level of genetic divergence between the US and UY isolates indicated that these flukes are not substantially more closely related to each other than either is to the five published UK isolates (S8 Table) although their *Neorickettsia* endosymbionts are genetically close to each other.

Phylogenetic affinities of *Neorickettsia* endobacteria of *F. hepatica*

A phylogenetic analysis was undertaken using the 16S rRNA sequences from *nFh* and 16 other species and isolates; the findings were similar to those previous reports [15]. Based on the 16S rRNA locus, *nFh* is closely related to the agent of Sennetsu fever (a strain of *N. sennetsu*) and a *Neorickettsia* isolate isolated from species of *Metagonimoides* (Heterophyidae) (Fig 5B). A complementary phylogenetic analysis was undertaken with conserved, single-copy homologues from sequenced species of *Neorickettsia* and representatives of the Anaplasmataceae (Fig 5C); in regard to this collection of 473 gene families, *nFh* appeared to be closer to *N. risticii* and *N. sennetsu* than to *N. helminthoeca*, the agent of salmon poisoning in dogs, consistent with synteny-based observations.

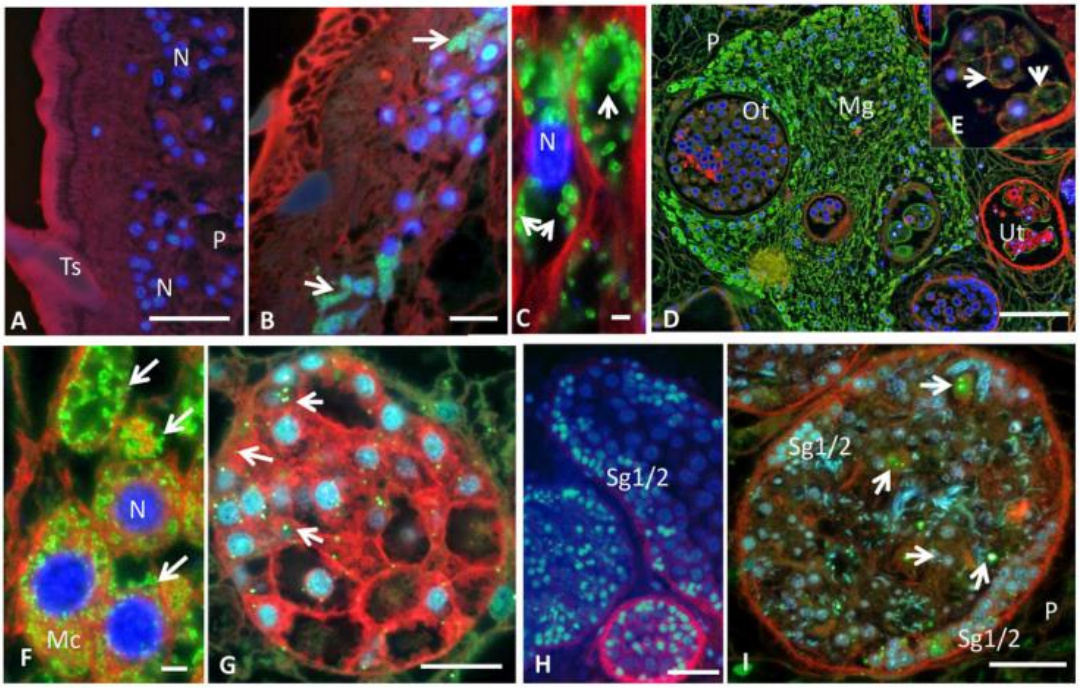


Fig 6. Immunofluorescence detection of *Neorickettsia* in adult *Fasciola hepatica* using polyclonal anti-serum raised against a recombinant surface protein of *Neorickettsia* of *P. elegans* (PeNsp-3, green labeling). DAPI (blue) and wheat hemagglutinin (red) were used to detect double stranded DNA and plasma membranes, respectively. (A) No green labeling was seen in the tegument of *F. hepatica* from Uruguay that were known to be devoid of *Neorickettsia*. (B) Clusters of *Neorickettsia* (arrows) in the tegument close to tegumental nuclei in *F. hepatica* from Oregon. (C) Numerous 'donut'-shaped endobacteria (arrows) in the parenchyma in *F. hepatica* from Oregon. (D) Labeling of large numbers of *Neorickettsia* in the Mehlis' gland and labeling of single endobacterium in the ootype or intrauterine eggs of *F. hepatica* from Oregon. (E) Magnification of a region proximal to (D) showing granular staining of single endobacterium in intrauterine eggs. (F) Clusters of *Neorickettsia* rods (arrows) in the cytoplasm of Mehlis' cells of *F. hepatica* Oregon. (G) Individual *Neorickettsia* endobacteria (arrows) in a vitelline follicle with different stages of vitelline cells of *F. hepatica* Oregon. (H) No green staining indicative of *Neorickettsia* were found in the testis of *F. hepatica* from Uruguay tested negative for *Neorickettsia* by PCR. (I) *Neorickettsia* endobacteria (arrows) in the testis of *F. hepatica* Oregon with spermatogonia in the periphery and developing spermatozoa in the center. N, nucleus; Ts, tegument spine; P, parenchyma; Mg, Mehlis' gland; Ot, ootype; Ut, uterus; Mc, Mehlis' cell; Sg1/2, primary and secondary spermatogonia; Bar corresponds in A-B, D-I to 100 μm and in C to 1 μm.

doi:10.1371/journal.pgen.1006537.g006

Approximately 97% of the predicted proteins of *nFh* showed BLASTP matches in NR (e-value ≤ 1e-05, S5 Table). The top hits were to *N. risticii* or other *Neorickettsia* species. Indeed, most *nFh* genes (721 of 744) belonged to 719 OPFs shared with other *Neorickettsia* species (S5 Table). Of the 22 genes that were excluded from OPFs, 17 failed to find a match in NR and most lacked other functional annotations; the other five matched to hypothetical proteins from *N. risticii* and *N. sennetsu*. A single OPF was identified with members from *nFh*, *Wolbachia* species, *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*. The *nFh* gene assigned to this OPF was annotated as a replicative DNA helicase. Further assessment will be needed to validate these genes and explore their roles.

A set of 625 OPFs contained members common to all four sequenced *Neorickettsia* genomes; 83 of these OPFs were specific to *Neorickettsia*. The *nFh* proteins included in the *Neorickettsia*-specific OPFs were enriched for cellular component GO terms related to the

outer membrane and biological process GO terms related to transport (S4 Table). The association with the cell surface suggested a role in endobacterial-digenean host interactions.

Neorickettsia-like endobacteria localized within tissues of *F. hepatica* Oregon

Although reports of *Neorickettsia*-infected trematodes have emerged, they relied on detection by PCR; localization within the trematode was poorly established. Gibson *et al.* [47] employed Ig from the serum of a horse infected with *N. risticii* to detect *Neorickettsia* in eggs from the bat-infecting trematode *Acanthatrium oregonense* to support the hypothesis of vertical transmission. The same serum was used to localize *Neorickettsia* in discrete developmental stages of *Plagiorchis elegans*, a trematode of rodents and birds [14]. We attempted to use the horse serum to localize *nFh* in adult *F. hepatica* Oregon, but background staining interfered with interpretation of the signals. However, polyclonal antibodies raised against recombinant surface protein-3 of *P. elegans Neorickettsia* (*PeN*sp-3) provided useful to support localization studies (Fig 6). The *PeN*sp-3 protein (Genbank KX082665) and of *nFh* (AS219_03540; S5 Table) share 98% identity. Minimal background signal occurred with the *PeN*sp-3 antisera, and *nFh* were sensitively detected as a 'donut'-shaped structure surrounding the blue DAPI-stained nucleus, consistent with the staining pattern expected for surface proteins (Fig 6C) [48]. Whereas staining of *Neorickettsia* surface protein was not observed in *Neorickettsia*-negative (as confirmed by PCR; Fig 6A and 6H) *F. hepatica* from Uruguay endobacteria were detected in six of six individual adult *F. hepatica* worms from Oregon.

Because of the size of *F. hepatica* (~2 cm; F) and because *Neorickettsia* may be transmitted vertically, analysis focused on intra-uterine eggs and reproductive tissues. Endobacteria were frequently detected in varying numbers in the ovary, ootype, Mehlis' gland, vitelline glands and in intrauterine eggs (Fig 6D–6G) as well as mature eggs isolated from liver tissue (Fig 7). The presence of *nFh* in female reproductive tissue is highly suggestive of vertical transmission. Furthermore, we analyzed by PCR adult flukes obtained after an experimental infection with Oregon metacercariae, detecting a few individual worms positive for the presence of nFh (S7 Fig). More interestingly, eggs collected from this assay were both PCR positive and presented the characteristic images of *Neorickettsia* supports the notion of vertical transmission.

Surprisingly, we also observed *nFh* in the testis and other parts of the male reproductive organs (Fig 6I and S8 Fig). Although *F. hepatica* is a hermaphrodite, cross-fertilization is assumed to be the usual reproductive strategy [49]. The presence of *Neorickettsia* in spermatozoa and seminal fluid could provide an alternative route for fluke-to-fluke transmission, as it was described for tick-borne pathogens [50], though further studies will be needed to explore this possibility.

The somatic tissues of *F. hepatica* Oregon were mostly *Neorickettsia*-free. Clusters of *nFh* were occasionally seen in the tegument adjacent to some syncytial nuclei (Fig 6B) and in intestinal tissue, particularly near the oral suckers. Liver flukes use the oral suckers to penetrate the host tissues and anchor themselves to the bile ducts, thus providing a potential mechanism for fluke-to-host transmission of *nFh*. Several infectious diseases described in the medical and veterinary literature are attributable to *Neorickettsia* carried by digenean parasites; among the more relevant are the 'Salmon Poisoning Disease' (SPD) of dogs and the Elokomin fluke fever (EFF) of fish-eating mammals in the west coast of North America, Sennetsu fever described in humans mainly in Japan and southeast Asia, and Potomac Horse Fever (PHF) in the east coast of North America [15, 16]. Notably, PHF, also known as 'churrido equino', has been described in horses in the Lake Merin region of Uruguay and Brazil (reviewed in [51]). Several species of *Neorickettsia* based on pathology, serology, antigen profile and/or genomic sequence, are

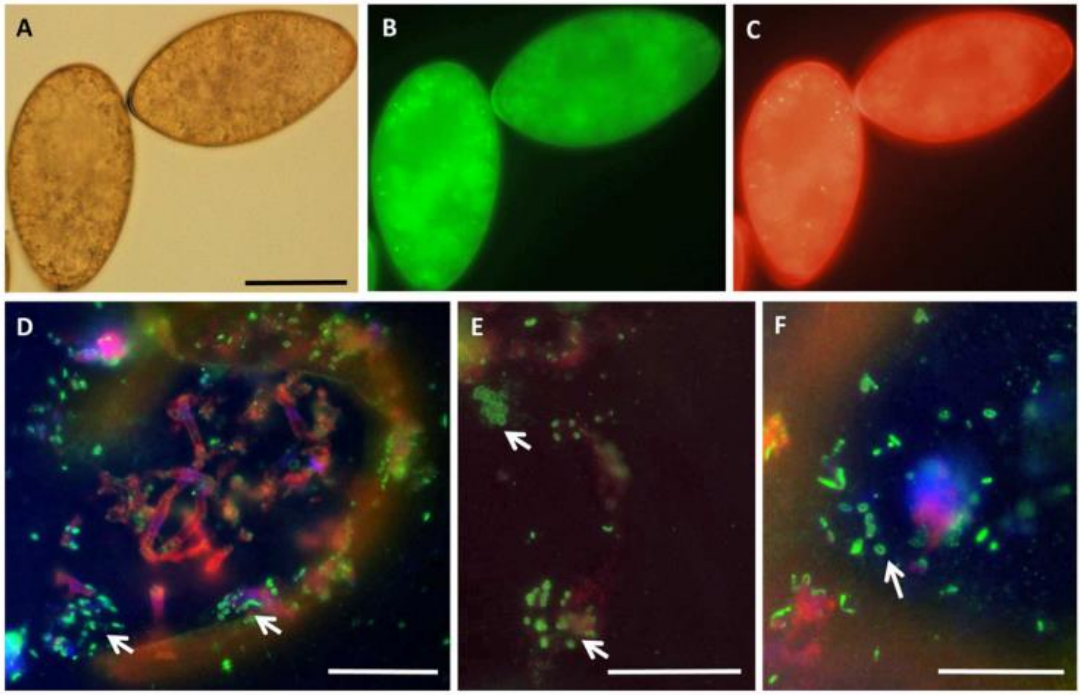


Fig 7. Immunofluorescence detection of *Neorickettsia* in eggs of *F. hepatica* from Oregon using polyclonal anti-serum raised against a recombinant surface protein of Neorickettsia of *P. elegans* (PeNsp-3, green labeling, D-F). (A) Unstained eggs recovered from the liver by regular light microcopy. (B) and (C) Unstained eggs recovered from the liver by immunofluorescence microscopy using different filters demonstrating auto-fluorescence. (D-F) Cross-sections of eggs showing various amounts of *Neorickettsia* (arrows). Bar corresponds to 50 μm.

doi:10.1371/journal.pgen.1006537.g007

considered the causative agents for these diseases; in particular, *Neorickettsia* (*Ehrlichia*) *sennetsu* causes acute, debilitating, mononucleosis-like disease [52], and has been implicated as a significant cause of human fevers of unknown etiology in southeastern Asia [53, 54].

Whereas the disease potential of *Neorickettsia* found in *F. hepatica* remains to be established, the *F. hepatica-nFh* association should be explored as a cryptic rickettsial pathogen of humans and ruminants in regions endemic for fasciolosis [55]. Additionally, more thorough studies of both the vertical transmission of *nFh* among the developmental stages of the liver fluke, and the potential horizontal transmission to the mammalian host might shine a light on the mechanisms behind the pathology induced by *Neorickettsia* endosymbionts of digenean parasites.

Methods

Liver flukes

Two isolates of *Fasciola hepatica* were analyzed: adult worms collected from livers of naturally infected sheep from a commercial slaughterhouse in Oregon (provided by Baldwin Aquatics Inc., Monmouth, Oregon), i.e. Oregon isolate; and worms isolated from livers of naturally

infected sheep obtained from a commercial slaughterhouse in Montevideo, Uruguay, i.e. Uruguay isolate. For transcriptomic analysis, total RNAs were obtained from the egg, metacercarial and adult developmental stages (in duplicate). Eggs were collected from gall bladder of naturally infected sheep. Metacercariae were purchased from Baldwin Aquatics Inc. (Monmouth, Oregon). Tissue sections for the histological analysis were prepared from adult worms of the Oregon and Uruguay isolates.

DNA & RNA isolation, sequencing, assembly, annotation of the genome of *F. hepatica* Oregon isolate and transcriptome analysis

Fresh or ethanol-preserved adult worms were fragmented using a scalpel blade, and genomic DNA (gDNA) was extracted and purified using the kit E.Z.N.A. SQ Tissue DNA Kit (Omega Bio-tek), and the yield and purified assessed by Bio-Analyzer. Whole genome shotgun fragment and paired-end sequencing libraries (3 kb and 8 kb) were constructed from the gDNAs, as described [39, 56], and sequenced on the Illumina HiSeq2000 platform.

Linker and adapter sequences were trimmed, and cleaned reads were assembled using ALLPATHS-LG [57]. Pygap, an in-house assembly improvement tool, was used to join and extend contigs using unassembled reads when possible. Annotation of different features present in the assembly was done as previously described [58] and outlined in S1 Text.

Total RNA was extracted from eggs and adults from the gall bladder of naturally infected sheep and metacercariae (Baldwin Aquatics Inc., Monmouth, Oregon) using TRIzol reagent (Invitrogen/Life Technologies, Carlsbad, CA) according to the manufacturer's instructions, and treated with Ambion Turbo DNase (Ambion/Applied Biosystems, Austin, TX). As previously described [59], RNA quality and yield were assessed, the purified RNA was poly(A) selected, reverse transcribed, paired-end cDNA libraries were generated, sequenced on the Illumina HiSeq 2000 platform and reads were analytically processed. Remaining, high-quality RNAseq reads (from one egg, two metacercariae and two adult biological replicates) were aligned to the genome assembly and constitutively expressed and differentially expressed genes were identified using standard protocols outlined in S1 Text.

Identification, assembly and annotation of the genome of *Neorickettsia* from *F. hepatica* Oregon, isolate *nFh*

A total of 126 contigs were identified as being from bacterial origin in the *F. hepatica* genome assembly, and BLAST analyses indicated significant homology to *Neorickettsia* species. The total complement of raw reads were re-mapped to the 126 *Neorickettsia* contigs using BWA--MEM version 7.10 with default parameters [60], and matching reads were assembled and assembly improved using standard protocols (*see* S1 Text). The genome assembly was annotated via the NCBI prokaryotic genome annotation pipeline [61].

Functional annotation of deduced proteins of *F. hepatica* and *nFh*, and MultiParanoid analyses

Deduced protein sequences were subjected to BLASTP against informative databases, including NCBI NR, InterPro, gene ontology (GO), KEGG, MEROPS using default cutoffs and release versions as specified in the S1 Text. Module completion was assessed as described [39], and transmembrane domains and classical secretion peptides were predicted using standard protocols (*see* S1 Text).

Inferred protein sequences of *F. hepatica* were compared to proteins from other trematodes & cognate mammalian hosts and from *Neorickettsia* were compared to proteins from

representative species from the Anaphasmataceae, included all four fully sequenced *Neorickettsia* (accession numbers are provided in S1 Text). Orthologous protein families (OPFs) were constructed from pairwise InParanoid comparisons using MultiParanoid [62]. More detailed phylogenetic analyses at a level of rRNA sequences or single copy genes were performed as outlined in the S1 Text. *F. hepatica* Oregon and *F. hepatica* UK gene sets were compared using orthologs identified by Orthofinder v. 0.7.1 [63].

Genetic variation in *Neorickettsia* and *F. hepatica* nuclear genomes

We sequenced the genomic DNA of an *F. hepatica* isolate from Uruguay that was PCR-positive for *Neorickettsia* using the Illumina platform (2 × 100bp paired-end sequencing), as previously described [56]. We included the published genomes of the United Kingdom isolates (SRA Project ID: ERP006249) in the variant analysis to help contextualize our data. Genomic reads were mapped against the combined Oregon reference assembly of *Neorickettsia and F. hepatica* using bwa v0.7.15 [60], followed by removal of PCR and optical duplicates using picard tools v2.6.0 [64]. Single-nucleotide variants were called via local de-novo assembly of haplotypes using the GATK pipeline v3.6 [65]. The following set of quality filters were applied to obtain high-confidence SNP calls: DP (maximum depth) > median depth+(median absolute deviation×1.4826)×2; QD (variant confidence divided by the unfiltered depth of non-reference samples) < 2.0; FS (Phred-scaled p-value using Fisher's Exact Test to detect strand bias in the reads) > 60.0; MQ (Root Mean Square of the mapping quality of the reads across all samples) < 40.0; MQRankSum (Mann-Whitney Rank Sum Test for mapping qualities) < -12.5; ReadPosRankSum (Mann-Whitney Rank Sum Test for the distance from the end of the read for reads with the alternate allele) < -8.0. Using SnpEff [66], variants were annotated based on their genomic locations and predicted coding effects. The genetic distance between isolates (1-ibs, identity by state) were computed using PLINK v1.90 after excluding loci with missing genotypes in any of the isolates.

Detection of *Neorickettsia* by PCR

To investigate vertical transmission of *nFh*, genomic DNA was extracted from individual worms obtained after two experimental infections in bovines (*Bos taurus*) performed at the Experimental Farm of the Institute of Hygiene, Montevideo, Uruguay, following international standards for care of research animals, and approved by the National Committee of Experimental Animal Health (CHEA). Polled Hereford calves negative to *F. hepatica* by fecal egg count and ELISA (Piacenza et al., 1999) and treated orally with ivemectin 1% (Mexiver, Laboratorios Santa Elena), were used in immunization studies that included challenge infection by mouth with 400 metacercariae (MCs). The MCs were obtained from Baldwin Aquatics, Oregon (assay 1) or DILAVE, Uruguay (assay 2). The cattle were euthanized at a commercial abattoir on week 20, and adult flukes were recovered from the liver of each of the calves, and flukes from each calf stored separately in >70% ethanol. Liver fluke eggs from the gall bladder of the calves also were recovered, and stored in the immunization group. DNA was extracted from individual worms of each assay, and from the pooled eggs, as described above. The presence of *Neorickettsia* within the *F. hepatica* adult flukes was investigated by nested PCR directed to the 16s rRNA gene, as described [67].

Histological examination of *F. hepatica* and its *Neorickettsia* endobacterium

Oregon strain flukes from sheep (and *Neorickettsia*-negative flukes from Uruguay) were fixed first in 70% ethanol and then in 10% buffered formalin overnight, tissue processed

(Shandon 1000 Tissue Processor, Thermo Scientific, Waltham, MA, USA), embedded in paraffin, sectioned at 5 μm. Serial sections were used for immunohistochemical studies and hematoxylin & eosin staining (S6 Fig). The sections stained with hematoxylin & eosin (according to standard technique) were used to assess morphology and determine the anatomical structures to be expected on adjacent slides used for immunohistochemical localization of *nFh*. Unstained tissue sections were rehydrated and blocked with 5% bovine serum albumin (Sigma, St. Louis MO, USA) for 30 min to prevent non-specific antibody binding. Polyclonal mouse antisera raised against a recombinant *Neorickettsia* surface protein from *Plagiorchis elegans* (Genbank Accession KX082665, *PeN*sp-3) diluted 1:250 in phosphate buffered saline containing 0.1% Triton-X and 1% bovine serum albumin was used as the primary antibody. Anti-mouse IgG Alexa Fluor 488 (Invitrogen) was used as a secondary antibody for fluorescence microscopy. Wheat germ agglutinin 633 (200 μg/ml, Invitrogen, Carlsbad, CA, USA) and DAPI (Prolong Antifade with DAPI, Molecular Probes by Life Technologies, Carlsbad, CA, USA) were used to label membranes and double-stranded DNA, respectively. Sections were examined using a wide field fluorescence microscope (WFFM, Zeiss Axios Imager Upright Fluorescence Microscope) with plan-apochromat 100X oil, 63X or 40X objectives. Fluorescence microscopy was performed at the Washington University Molecular Microbiology Imaging Facility (http://micro.imaging.wustl.edu/).

Supporting Information

S1 Fig. A comparison of the protein coding genes of *Fasciola hepatica* Oregon and *Fasciola hepatica* UK. (A) Histogram indicates the size distribution of the predicted protein coding sequences of both *Fasciola* genomes. *F. hepatica* UK contains an abundance of very short genes (as small as 3bp) but more large (>600bp) genes. While the proteins predicted from the two genomes do not correspond well with one another (B), functional elements appear to be shared (C). KEGG Orthologous groups (KO) shared among the *F. hepatica* genomes.
(PDF)

S2 Fig. Gene expression information for *Fasciola hepatica*. Gene expression was profiled in eggs, metacercariae (MC), and sexually mature adults (hermaphrodites). (A) Clustering of samples based on gene expression (fragments per kilobase per million reads mapped, FPKM) indicated that eggs were more closely related to adults (which, themselves, contain eggs) than to metacercariae. (B) Differential expression of *F. hepatica* genes between the diverse egg, metacercariae and adult life cycle stages. Differentially expressed genes were significantly more likely than other genes to be phylogenetically conserved across all species test ($P = 0.006$) and more likely to contain transmembrane domains ($P = 0.015$). Genes with higher expression in metacercariae were enriched for several GO terms related to signal transduction and organismal development (S4 Table), and were less likely to be conserved with other FBTs ($P = 1 \times 10^{-7}$), which is not surprising given that *F. hepatica* metacercariae encyst on plants rather than within fish or crustaceans. In contrast, the 3,811 genes upregulated in adult flukes were enriched for microtubule based movement, redox regulation, and metabolic processes (S4 Table), as previously found in expression studies of *F. hepatica* from the UK [68]. The genes overexpressed in adults compared to metacercariae were more likely to be FBT conserved and specific ($P = 2 \times 10^{-7}$) and to have orthologs in mammals but not the free-living platyhelminth *S. mediterranea* ($P = 6 \times 10^{-10}$), suggesting potential roles in host interaction. (C) Summary of Illumina RNAseq reads, and SRA accessions.
(TIF)

S3 Fig. Metabolic and catabolic liver fluke pathway details. (A) Fatty acid elongation by reversal of beta-oxidation (B) Fatty acid degradation (C) Aliphatic amino acid catabolism. Enzymes present in *F. hepatica* are identified in yellow, and those present in *S. mansoni* in blue. Image generated with the KEGGscan_pathway at trematode.net (http://trematode.net).
(TIF)

S4 Fig. Clustering of species and modules. Clustering based on complete (A) and lenient (B) completion of KEGG metabolic pathways modules in each species (light green is "incomplete with < 3 reaction steps. Modules with 2 steps have been manually filled in after combining "strict" and "lenient" results). Oa = *Ovis aries*, Bt = *Bos taurus*, Hs = *Homo sapiens*, FhOREGON = *F. hepatica*, Oregon strain, FhUK = *F. hepatica*, UK strain, Sm = *Schistosoma mansoni*, Cs = *Clonorchis sinensis*, Ov = *Opisthorchis viverrini*.
(TIF)

S5 Fig. PCR analysis for *Neorickettsia* from diverse isolates of *Fasciola hepatica*. Nested PCR for the bacterial 16s RNA gene in five different *F. hepatica* flukes from Uruguay (lanes 3–7) and the reference Oregon strain (8). nF*h*-positive signals were observed in one sample from Uruguay and the Oregon isolate.
(TIF)

S6 Fig. Examples of standard H&E stain of sections of adult *F. hepatica* to help identify organs and tissues examined for localization of *Neorickettsia*. (A) Cross-section of the proximal part of *F. hepatica*. (B) Cross-section of the distal part of *F. hepatica*. (C) Longitudinal section of *F. hepatica*. Bar corresponds to 1 mm.
(TIF)

S7 Fig. Detection of nFh 16s rRNA gene by PCR. Primary (top panel) and secondary (nested) PCR for the bacterial 16s RNA gene following protocol previously described [65]. DNA from *nFh*-negative sample from Uruguay (lane 11) and *nFh*-positive sample from Oregon (lane 13). To further test if the bacteria might have been transmitted through the parasite, we tested by PCR individual flukes isolated after two different experimental infections performed with metacercariae from Uruguay (lanes 3–10) and Oregon (lanes 16–23) respectively, and eggs collected from these experimental infections (lanes 1–2, and 14–15). Since nested PCR was performed using dilution of primary amplicons without band purification, carry over of first round primers occurred, visible as a doublet band below the primary amplification, corresponding to the expected nested product (lower band) and byproducts between the external and internal primers. The identity of the *nFh* 16s rRNA gene was confirmed by nucleotide sequencing.
(TIF)

S8 Fig. Localization of *Neorickettsia* in male reproductive tissue of *F. hepatica* Oregon. (A) Cross-section of the vas deferens with overlay of the individual stains for plasma membranes (wheat germ agglutinin, WGA), double stranded DNA (DAPI) and *Neorickettsia* (Nsp). (B) Individual stain for plasma membranes (red). (C) Individual stain for DNA. Note the strong blue stain of spermatozoa and the lighter bluish stain of low DNA content *Neorickettsia* (arrows). D. Individual green stain for *Neorickettsia* (arrows). Bar corresponds to 100 μm.
(TIF)

S1 Table. The nuclear genome assemblies of *F. hepatica* Oregon and other food borne trematodes.
(DOCX)

S2 Table. Repetitive elements, tRNA and rRNA in the genome of *Fasciola hepatica* Oregon.
(DOCX)

S3 Table. *F. hepatica* genome summary table.
(XLSX)

S4 Table. Functional enrichment among *F. hepatica* gene sets of interest.
(XLSX)

S5 Table. *Neorickettsia F. hepatica* genome summary table.
(XLSX)

S6 Table. Coverage statistics for *Fasciola hepatica* and *Neorickettsia*.
(DOCX)

S7 Table. Non-synonymous SNPs between the *Neorickettsia* genomes of *Fasciola hepatica* Oregon (US) and Uruguay (UY) isolates.
(DOCX)

S8 Table. Genetic distance (1-ibs, identity by state) between *Fasciola hepatica* isolates based on the nuclear SNPs.
(DOCX)

S1 Text. Supporting materials and methods.
(DOCX)

Acknowledgments

We thank Mr. Norman Baldwin, Baldwin Aquatics Inc. for developmental stages *F. hepatica*, Dr. Vasyl Tkach (University of North Dakota) for providing *Neorickettsia*-infected *P. elegans* material, and Dr. Yasuko Rikihisa (Ohio State University) for serum from a horse infected with *Neorickettsia*.

Author Contributions

Conceptualization: MM PJB.

Formal analysis: SNM YJC JFT SF PS FD CC KHP.

Funding acquisition: PJB JFT MM.

Methodology: JFT GR KF PUF PSL LK VHM NDO LK VHM MM.

Resources: JFT GR VHM CC PHL KF.

Visualization: SNM RT BAR PS KF SF.

Writing – original draft: SNM JFT RT BAR PUF MM.

Writing – review & editing: MM PJB PUF JFT GR.

References

1. Keiser J, Utzinger J. Food-borne trematodiases. Clin Microbiol Rev. 2009; 22(3):466–83. doi: 10.1128/CMR.00012-09 PMID: 19597009
2. Hotez PJ, Alvarado M, Basanez MG, Bolliger I, Bourne R, Boussinesq M, et al. The global burden of disease study 2010: interpretation and implications for the neglected tropical diseases. PLoS Negl Trop Dis. 2014; 8(7):e2865. doi: 10.1371/journal.pntd.0002865 PMID: 25058013

3. Torgerson PR, Devleesschauwer B, Praet N, Speybroeck N, Willingham AL, Kasuga F, et al. World Health Organization Estimates of the Global and Regional Disease Burden of 11 Foodborne Parasitic Diseases, 2010: A Data Synthesis. PLoS Med. 2015; 12(12):e1001920. doi: 10.1371/journal.pmed.1001920 PMID: 26633705
4. Nyindo M, Lukambagire AH. Fascioliasis: An Ongoing Zoonotic Trematode Infection. Biomed Res Int. 2015; 2015:786195. doi: 10.1155/2015/786195 PMID: 26417603
5. Charlier J, Vercruysse J, Morgan E, van Dijk J, Williams DJ. Recent advances in the diagnosis, impact on production and prediction of Fasciola hepatica in cattle. Parasitology. 2014; 141(3):326–35. Epub 2013/11/16. doi: 10.1017/S0031182013001662 PMID: 24229764
6. Rondelaud D, Belfaiza M, Vignoles P, Moncef M, Dreyfuss G. Redial generations of Fasciola hepatica: a review. J Helminthol. 2009; 83(3):245–54. doi: 10.1017/S0022149X09222528 PMID: 19203397
7. Burden DJ, Bland AP, Hammet NC, Hughes DL. *Fasciola hepatica*: migration of newly excysted juveniles in resistant rats. Experimental parasitology. 1983; 56(2):277–88. Epub 1983/10/01. PMID: 6617806
8. Lopez M, White AC Jr., Cabada MM. Burden of *Fasciola hepatica* Infection among children from Paucartambo in Cusco, Peru. The American journal of tropical medicine and hygiene. 2012; 86(3):481–5. Epub 2012/03/10. doi: 10.4269/ajtmh.2012.11-0448 PMID: 22403322
9. Marcos L, Terashima A, Leguia G, Canales M, Espinoza J, Gotuzzo E. La infección por *Fasciola Hepática* en el Perú: una enfermedad emergente. Revista de Gastroenterología del Perú. 2007; 27:389–96. PMID: 18183272
10. Chai JY. Praziquantel treatment in trematode and cestode infections: an update. Infection & chemotherapy. 2013; 45(1):32–43. Epub 2013/11/23.
11. Kelley JM, Elliott TP, Beddoe T, Anderson G, Skuce P, Spithill TW. Current Threat of Triclabendazole Resistance in *Fasciola hepatica*. Trends Parasitol. 2016.
12. Cabada MM, Lopez M, Cruz M, Delgado JR, Hill V, White AC Jr. Treatment Failure after Multiple Courses of Triclabendazole among Patients with Fascioliasis in Cusco, Peru: A Case Series. PLoS Negl Trop Dis. 2016; 10(1):e0004361. doi: 10.1371/journal.pntd.0004361 PMID: 26808543
13. Gil LC, Diaz A, Rueda C, Martinez C, Castillo D, Apt W. Resistant human fasciolasis: report of four patients. Rev Med Chil. 2014; 142(10):1330–3. doi: 10.4067/S0034-98872014001000014 PMID: 25601119
14. Greiman SE, Rikihisa Y, Cain J, Vaughan JA, Tkach VV. Germs within worms: localization of *Neorickettsia sp.* within life cycle stages of digenean *plagiorchis elegans*. Applied and environmental microbiology. 2016. Epub 2016/02/14.
15. Greiman SE, Tkach VV, Pulis E, Fayton TJ, Curran SS. Large scale screening of digeneans for *Neorickettsia* endosymbionts using real-time PCR reveals new *Neorickettsia* genotypes, host associations and geographic records. PLoS One. 2014; 9(6):e98453. doi: 10.1371/journal.pone.0098453 PMID: 24911315
16. Vaughan JA, Tkach VV, Greiman SE. Neorickettsial endosymbionts of the digenea: diversity, transmission and distribution. Adv Parasitol. 2012; 79:253–97. doi: 10.1016/B978-0-12-398457-9.00003-2 PMID: 22726644
17. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007; 23(9):1061–7. doi: 10.1093/bioinformatics/btm071 PMID: 17332020
18. Wang X, Chen W, Huang Y, Sun J, Men J, Liu H, et al. The draft genome of the carcinogenic human liver fluke *Clonorchis sinensis*. Genome Biol. 2011; 12(10):R107. Epub 2011/10/26. doi: 10.1186/gb-2011-12-10-r107 PMID: 22023798
19. Young ND, Nagarajan N, Lin SJ, Korhonen PK, Jex AR, Hall RS, et al. The *Opisthorchis viverrini* genome provides insights into life in the bile duct. Nature communications. 2014; 5.
20. Huang Y, Chen W, Wang X, Liu H, Chen Y, Guo L, et al. The carcinogenic liver fluke, *Clonorchis sinensis*: new assembly, reannotation and analysis of the genome and characterization of tissue transcriptomes. PLoS One. 2013; 8(1):e54732. Epub 2013/02/06. doi: 10.1371/journal.pone.0054732 PMID: 23382950
21. Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, Aslett MA, et al. A Systematically Improved High Quality Genome and Transcriptome of the Human Blood Fluke *Schistosoma mansoni*. PLoS Negl Trop Dis. 2012; 6(1):e1455. doi: 10.1371/journal.pntd.0001455 PMID: 22253936
22. Irving JA, Spithill TW, Pike RN, Whisstock JC, Smooker PM. The evolution of enzyme specificity in *Fasciola spp.* J Mol Evol. 2003; 57(1):1–15. doi: 10.1007/s00239-002-2434-x PMID: 12962301
23. Stack C, Dalton JP, Robinson MW. The phylogeny, structure and function of trematode cysteine proteases, with particular emphasis on the *Fasciola hepatica* cathepsin L family. Adv Exp Med Biol. 2011; 712:116–35. doi: 10.1007/978-1-4419-8414-2_8 PMID: 21660662

24. Cancela M, Ruetalo N, Dell'Oca N, da Silva E, Smircich P, Rinaldi G, et al. Survey of transcripts expressed by the invasive juvenile stage of the liver fluke *Fasciola hepatica*. BMC Genomics. 2010; 11:227. doi: 10.1186/1471-2164-11-227 PMID: 20374642

25. Cancela M, Acosta D, Rinaldi G, Silva E, Duran R, Roche L, et al. A distinctive repertoire of cathepsins is expressed by juvenile invasive *Fasciola hepatica*. Biochimie. 2008; 90(10):1461–75. Epub 2008/06/25. doi: 10.1016/j.biochi.2008.04.020 PMID: 18573308

26. Robinson MW, Tort JF, Lowther J, Donnelly SM, Wong E, Xu W, et al. Proteomics and phylogenetic analysis of the cathepsin L protease family of the helminth pathogen *Fasciola hepatica*: expansion of a repertoire of virulence-associated factors. Mol Cell Proteomics. 2008; 7.

27. Lowther J, Robinson MW, Donnelly SM, Xu W, Stack CM, Matthews JM, et al. The Importance of pH in Regulating the Function of the *Fasciola hepatica* Cathepsin L1 Cysteine Protease. PLoS Negl Trop Dis. 2009; 3(1):e369. doi: 10.1371/journal.pntd.0000369 PMID: 19172172

28. Robinson MW, Corvo I, Jones PM, George AM, Padula MP, To J, et al. Collagenolytic activities of the major secreted cathepsin L peptidases involved in the virulence of the helminth pathogen, *Fasciola hepatica*. PLoS Negl Trop Dis. 2011; 5(4):e1012. doi: 10.1371/journal.pntd.0001012 PMID: 21483711

29. Corvo I, Cancela M, Cappetta M, Pi-Denis N, Tort JF, Roche L. The major cathepsin L secreted by the invasive juvenile *Fasciola hepatica* prefers proline in the S2 subsite and can cleave collagen. Mol Biochem Parasitol. 2009; 167.

30. Kim TI, Yoo WG, Kwak BK, Seok JW, Hong SJ. Tracing of the Bile-chemotactic migration of juvenile *Clonorchis sinensis* in rabbits by PET-CT. PLoS Negl Trop Dis. 2011; 5(12):e1414. doi: 10.1371/journal.pntd.0001414 PMID: 22180795

31. Kang JM, Bahk YY, Cho PY, Hong SJ, Kim TS, Sohn WM, et al. A family of cathepsin F cysteine proteases of *Clonorchis sinensis* is the major secreted proteins that are expressed in the intestine of the parasite. Mol Biochem Parasitol. 2010; 170(1):7–16. doi: 10.1016/j.molbiopara.2009.11.006 PMID: 19932715

32. Dvořák J, Mashiyama ST, Braschi S, Sajid M, Knudsen GM, Hansell E, et al. Differential use of protease families for invasion by schistosome cercariae. Biochimie. 2008; 90(2):345–58. http://dx.doi.org/10.1016/j.biochi.2007.08.013. PMID: 17936488

33. Rawlings ND, Barrett AJ, Finn R. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic acids research. 2016; 44(D1):D343–50. Epub 2015/11/04. doi: 10.1093/nar/gkv1118 PMID: 26527717

34. Di Maggio LS, Tirloni L, Pinto AF, Diedrich JK, Yates Iii JR, Benavides U, et al. Across intra-mammalian stages of the liver fluke *Fasciola hepatica*: a proteomic study. Scientific reports. 2016; 6:32796. doi: 10.1038/srep32796 PMID: 27600774

35. Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. Nature. 2013; 496(7443):57–63. http://dx.doi.org/10.1038/nature12031 PMID: 23485966

36. Daniel H, Spanier B, Kottra G, Weitz D. From bacteria to man: archaic proton-dependent peptide transporters at work. Physiology (Bethesda). 2006; 21:93–102.

37. Marcilla A, De la Rubia JE, Sotillo J, Bernal D, Carmona C, Villavicencio Z, et al. Leucine Aminopeptidase Is an Immunodominant Antigen of *Fasciola hepatica* Excretory and Secretory Products in Human Infections. Clinical and Vaccine Immunology: CVI. 2008; 15(1):95–100. doi: 10.1128/CVI.00338-07 PMID: 18003812

38. Tielens AG, van Grinsven KW, Henze K, van Hellemond JJ, Martin W. Acetate formation in the energy metabolism of parasitic helminths and protists. Int J Parasitol. 2010; 40(4):387–97. doi: 10.1016/j.ijpara.2009.12.006 PMID: 20085767

39. Tyagi R, Joachim A, Ruttkowski B, Rosa BA, Martin JC, Hallsworth-Pepin K, et al. Cracking the nodule worm code advances knowledge of parasite biology and biotechnology to tackle major diseases of livestock. Biotechnology advances. 2015; 33(6 Pt 1):980–91. Epub 2015/06/01. doi: 10.1016/j.biotechadv.2015.05.004 PMID: 26026709

40. Dunning Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli SV, Eisen JA, et al. Comparative genomics of emerging human ehrlichiosis agents. PLoS Genet. 2006; 2(2):e21. doi: 10.1371/journal.pgen.0020021 PMID: 16482227

41. Lin M, Zhang C, Gibson K, Rikihisa Y. Analysis of complete genome sequence of *Neorickettsia risticii*: causative agent of Potomac horse fever. Nucleic acids research. 2009; 37(18):6076–91. doi: 10.1093/nar/gkp642 PMID: 19661282

42. Gibson K, Kumagai Y, Rikihisa Y. Proteomic analysis of *Neorickettsia sennetsu* surface-exposed proteins and porin activity of the major surface protein P51. J Bacteriol. 2010; 192(22):5898–905. doi: 10.1128/JB.00632-10 PMID: 20833807

43. Tsang A, Seidle H, Jawaid S, Zhou W, Smith C, Couch RD. *Francisella tularensis* 2-C-Methyl-D-Erythritol 4-Phosphate Cytidylyltransferase: Kinetic Characterization and Phosphoregulation. PLoS One. 2011; 6(6):e20884. doi: 10.1371/journal.pone.0020884 PMID: 21694781

44. Luck AN, Evans CC, Riggs MD, Foster JM, Moorhead AR, Slatko BE, et al. Concurrent transcriptional profiling of *Dirofilaria immitis* and its *Wolbachia* endosymbiont throughout the nematode life cycle reveals coordinated gene expression. BMC Genomics. 2014; 15(1):1–18.

45. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, et al. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 2005; 3(4):e121. doi: 10.1371/journal.pbio.0030121 PMID: 15780005

46. Eberl M, Hintz M, Reichenberg A, Kollas AK, Wiesner J, Jomaa H. Microbial isoprenoid biosynthesis and human gammadelta T cell activation. FEBS letters. 2003; 544(1–3):4–10. Epub 2003/06/05. PMID: 12782281

47. Gibson KE, Rikihisa Y, Zhang C, Martin C. *Neorickettsia risticii* is vertically transmitted in the trematode Acanthatrium oregonense and horizontally transmitted to bats. Environmental microbiology. 2005; 7 (2):203–12. Epub 2005/01/22. doi: 10.1111/j.1462-2920.2004.00683.x PMID: 15658987

48. Fischer K, Beatty WL, Jiang D, Weil GJ, Fischer PU. Tissue and stage-specific distribution of *Wolbachia* in *Brugia malayi*. PLoS Negl Trop Dis. 2011; 5(5):e1174. doi: 10.1371/journal.pntd.0001174 PMID: 21629728

49. Peoples RC, Fried B. Form and function in the digenea. Adv Exp Med Biol. 2014; 766:3–20. Epub 2014/06/07. doi: 10.1007/978-1-4939-0915-5_1 PMID: 24903361

50. Gaber MS, Khalil GM, Hoogstraal H. *Borrelia crocidurae*: venereal transfer in Egyptian *Ornithodoros erraticus* ticks. Exp Parasitol. 1982; 54(2):182–4. Epub 1982/10/01. PMID: 6889976

51. Dutra F, Schuch LF, Delucchi E, Curcio BR, Coimbra H, Raffi MB, et al. Equine monocytic Ehrlichiosis (Potomac horse fever) in horses in Uruguay and southern Brazil. J Vet Diagn Invest. 2001; 13(5):433–7. Epub 2001/10/03. PMID: 11580069

52. Rikihisa Y. The tribe *Ehrlichieae* and ehrlichial diseases. Clin Microbiol Rev. 1991; 4(3):286–308. PMID: 1889044

53. Newton PN, Rolain JM, Rasachak B, Mayxay M, Vathanatham K, Seng P, et al. *Sennetsu neorickettsiosis*: a probable fish-borne cause of fever rediscovered in Laos. Am J Trop Med Hyg. 2009; 81(2):190–4. PMID: 19635868

54. Walker DH. Sennetsu neorickettsiosis: a potentially prevalent, treatable, acutely incapacitating tropical infectious disease. Am J Trop Med Hyg. 2009; 81(2):187–8. PMID: 19635866

55. Walker DH, Ismail N. Emerging and re-emerging rickettsioses: endothelial cell infection and early disease events. Nat Rev Microbiol. 2008; 6(5):375–86. doi: 10.1038/nrmicro1866 PMID: 18414502

56. McNulty SN, Strube C, Rosa BA, Martin JC, Tyagi R, Choi YJ, et al. *Dictyocaulus viviparus* genome, variome and transcriptome elucidate lungworm biology and support future intervention. Scientific reports. 2016; 6:20316. Epub 2016/02/10. doi: 10.1038/srep20316 PMID: 26856411

57. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011; 108(4):1513–8. doi: 10.1073/pnas.1017351108 PMID: 21187386

58. Tang YT, Gao X, Rosa BA, Abubucker S, Hallsworth-Pepin K, Martin J, et al. Genome of the human hookworm Necator americanus. Nat Genet. 2014. Epub 2014/01/21.

59. McNulty SN, Fischer PU, Townsend RR, Curtis KC, Weil GJ, Mitreva M. Systems biology studies of adult *paragonimus* lung flukes facilitate the identification of immunodominant parasite antigens. PLoS Negl Trop Dis. 2014; 8(10):e3242. Epub 2014/10/21. doi: 10.1371/journal.pntd.0003242 PMID: 25329661

60. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. doi: 10.1093/bioinformatics/btp324 PMID: 19451168

61. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, et al. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. OMICS. 2008; 12(2):137–41. doi: 10.1089/omi.2008.0017 PMID: 18416670

62. Alexeyenko A, Tamas I, Liu G, Sonnhammer EL. Automatic clustering of orthologs and inparalogs shared by multiple proteomes. Bioinformatics. 2006; 22(14):e9–15. doi: 10.1093/bioinformatics/btl213 PMID: 16873526

63. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015; 16:157. Epub 2015/08/06. doi: 10.1186/s13059-015-0721-2 PMID: 26243257

64. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43(5):491–8. doi: 10.1038/ng.806 PMID: 21478889

65. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–303. Epub 2010/07/21. doi: 10.1101/gr.107524.110 PMID: 20644199

66. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin). 2012; 6(2):80–92. Epub 2012/06/26.

67. Greiman SE, Tkach VV, Vaughan JA. Transmission rates of the bacterial endosymbiont, *Neorickettsia risticii*, during the asexual reproduction phase of its digenean host, Plagiorchis elegans, within naturally infected lymnaeid snails. Parasit Vectors. 2013; 6:303. doi: 10.1186/1756-3305-6-303 PMID: 24383453

68. Cwiklinski K, Dalton JP, Dufresne PJ, La Course J, Williams DJ, Hodgkinson J, et al. The *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. Genome Biol. 2015; 16:71. doi: 10.1186/s13059-015-0632-2 PMID: 25887684

Adaptive Radiation of the Flukes of the Family Fasciolidae Inferred from Genome-Wide Comparisons of Key Species

Young-Jun Choi,[1] Santiago Fontenla,[2] Peter U. Fischer,[3] Thanh Hoa Le,[4] Alicia Costábile,[2] David Blair,[5] Paul J. Brindley,[6] Jose F. Tort,[2] Miguel M. Cabada,[7] and Makedonka Mitreva *[,1,3]

[1]McDonnell Genome Institute at Washington University in St. Louis, St. Louis, MO

[2]Departamento de Genética, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay

[3]Division of Infectious Diseases, Department of Medicine, Washington University School of Medicine, St. Louis, MO

[4]Immunology Department, Institute of Biotechnology, Vietnam Academy of Science and Technology, Hanoi, Vietnam

[5]College of Science and Engineering, James Cook University, Townsville, QLD, Australia

[6]Department of Microbiology, Immunology and Tropical Medicine, and Research Center for Neglected Diseases of Poverty, School of Medicine & Health Sciences, George Washington University, Washington, DC

[7]Division of Infectious Diseases, Department of Medicine, School of Medicine, University of Texas Medical Branch, Galveston, TX

*Corresponding author: E-mail: mmitreva@wustl.edu.

Associate editor: Keith Crandall

Abstract

Liver and intestinal flukes of the family Fasciolidae cause zoonotic food–borne infections that impact both agriculture and human health throughout the world. Their evolutionary history and the genetic basis underlying their phenotypic and ecological diversity are not well understood. To close that knowledge gap, we compared the whole genomes of *Fasciola hepatica*, *Fasciola gigantica*, and *Fasciolopsis buski* and determined that the split between *Fasciolopsis* and *Fasciola* took place ~90 Ma in the late Cretaceous period, and that between 65 and 50 Ma an intermediate host switch and a shift from intestinal to hepatic habitats occurred in the *Fasciola* lineage. The rapid climatic and ecological changes occurring during this period may have contributed to the adaptive radiation of these flukes. Expansion of cathepsins, fatty-acid-binding proteins, protein disulfide-isomerases, and molecular chaperones in the genus *Fasciola* highlights the significance of excretory–secretory proteins in these liver-dwelling flukes. *Fasciola hepatica* and *Fasciola gigantica* diverged ~5 Ma near the Miocene–Pliocene boundary that coincides with reduced faunal exchange between Africa and Eurasia. Severe decrease in the effective population size ~10 ka in *Fasciola* is consistent with a founder effect associated with its recent global spread through ruminant domestication. G-protein-coupled receptors may have key roles in adaptation of physiology and behavior to new ecological niches. This study has provided novel insights about the genome evolution of these important pathogens, has generated genomic resources to enable development of improved interventions and diagnosis, and has laid a solid foundation for genomic epidemiology to trace drug resistance and to aid surveillance.

Key words: **food–borne flukes, *Fasciola hepatica*, *Fasciola gigantica*, *Fasciolopsis buski*, genome evolution, adaptive radiation.**

Introduction

Digenetic trematodes (flukes) are a major group of helminth parasites of humans and animals. Among them, *Fasciolopsis buski* (*Fb. buski*), *Fasciola gigantica* (*Fa. gigantica*), and *Fasciola hepatica* (*Fa. hepatica*), the intestinal and liver flukes of the family Fasciolidae, cause zoonotic food–borne infections that have a substantial impact on both agriculture (3 billion USD per year) and human health (~90,000 disability-adjusted life years) throughout the world (Torgerson et al. 2015; Cwiklinski et al. 2016). *Fasciolopsis buski* (subfamily Fasciolopsinae) is a large fluke (up to 7.5 cm long, 2.5 cm wide) that infects the small intestine of humans and pigs in East and Southeast Asia, causing diarrhea, abdominal pain, fever, ascites, and bowel obstruction. *Fasciola hepatica* and *Fa. gigantica* (subfamily Fasciolinae) cause liver disease in ruminants and humans in Europe, the Americas and Australasia (where only *Fa. hepatica* is transmitted) and in Africa and Asia (where the two species overlap). When present, clinical symptoms include fever, malaise, abdominal pain, eosinophilia, and hepatomegaly during the acute phase, whereas biliary tree obstruction symptoms predominate in chronic disease. Fasciolid flukes have a heteroxenous life cycle, which involves a definitive vertebrate host (where the adult worms live, mate, and produce eggs), an intermediate molluscan host (where the larval stages develop and multiply), and a carrier (suitable aquatic plants). A previous phylogenetic study of the family Fasciolidae (Lotfy et al. 2008) indicates that *Fp. buski* (subfamily Fasciolopsinae) is descended from a relatively

early-diverging lineage, whereas *Fa. gigantica* and *Fa. hepatica* are derived sister species. Somewhere along the line leading to *Fasciola* species and other members of the Fasciolinae, two events of great importance occurred: a host switch from planorbid snails to lymnaeid snails, and a habitat switch by adults from intestinal sites to the liver. Here, we present the genomes of *Fp. buski* and *Fa. gigantica*, making the genomes of all three human-infecting fasciolid flukes available and via comparisons provide a better understanding of their evolutionary history and diversification, and the genetic bases underlying their phenotypic and ecological divergence and adaptation to different host species and habitats.

Results and Discussion

Genome Features of the Intestinal and Liver Flukes in the Family Fasciolidae

The nuclear and mitochondrial genomes of *Fb. buski* and *Fa. gigantica* were sequenced, assembled, and annotated (table 1 and supplementary table 1, Supplementary Material online). To facilitate more robust interspecies comparisons, our previously published *Fa. hepatica* genome (GenBank accession number: GCA_002763495) (McNulty et al. 2017) was reannotated using an improved methodology and updated RNA-seq and protein homology databases. The total assembly lengths of *Fp. buski* and *Fa. gigantica* draft genomes were 748 Mb and 1.13 Gb, respectively. Although the former is comparable to outgroup species in the family Opisthorchiidae, the latter with its expanded genome size is similar to *Fa. hepatica* (1.14 Gb), suggesting that the increased genome size is a derived trait that emerged in the lineage leading to *Fasciola*. Despite the genome size differences, the total numbers of protein-coding genes annotated in *Fp. buski* and the two *Fasciola* species were similar, ranging from 11,218 to 12,647 and representing 91.5% to 93.0% BUSCO completeness. These numbers were also comparable to other distantly related digenean taxa (table 1), suggesting that the relatively larger *Fasciola* genomes did not evolve through whole-genome duplications. Interestingly, the patterns of variation in transposable element (TE) contents of these and related genomes indicate that lineage-specific differential accumulation of TE families may have played a central role in genome size evolution in Fasciolidae (fig. 1A). The nonrepeat genome sizes are similar in *Fp. buski* (400 Mb), *Fa. hepatica* (372 Mb), *Fa. gigantica* (409 Mb), and the Opisthorchiidae, while smaller in the Schistosomatidae (200 Mb). However, the genomic regions containing interspersed repetitive elements are more than twice as long in *Fasciola* spp. (658–707 Mb) than in *Fp. buski* (318 Mb), and longer in fasciolids than in other trematodes. Most of the enrichment in repetitive elements is due to intergenic elements, although intronic elements are twice as long in *Fasciola* spp. than in *Fp. buski*. Notably, *Fasciola* genomes carry about 3–4 times more DNA transposons (e.g., Tc1/mariner) and about seven times more long-terminal-repeat retrotransposons (e.g., Gypsy, Pao, Copia) as compared with the *Fp. buski* genome (fig. 1B and supplementary table 2, Supplementary Material online). These are strongly enriched

Table 1. Assembly and Annotation Statistics of Analyzed Genomes.

	Fasciolopsis buski	*Fasciola gigantica*	*Fasciola hepatica* after Reannotation	*Fasciola hepatica* Original Annotation	*Clonorchis sinensis*	*Opisthorchis viverrini*	*Schistosoma mansoni*
GenBank assembly accession	LUCM00000000	SUNJ00000000	JXXN00000000	GCA_002763495	GCA_000236345	GCA_000715545	GCA_000237925
Total genome length (Mb)	748.5	1,128	1,138.3	1,138.3	547.3	620.5	364.5
N50 scaffold length (kb)	190.8	181.8	161.1	161.1	417.5	1,324	32,115.3
L50 scaffold count	1,104	1,805	2,036	2,036	408	138	4
Protein-coding genes	11,747	12,647	11,218	14,642	13,634	16,356	11,940
%BUSCO (complete)	73.9	77.2	78.6	43.9	77.5	79.5	89.4
%BUSCO (fragmented)	19.1	14.9	12.9	35.3	13.5	11.9	7.6
Mean CDS length	1,420	1,376	1,633	837	1,591	1,301	1,433
Mean intron length	3,708	3,982	4,170	2,902	2,757	3,550	2,409
Mean exons per mRNA	5.2	5.9	7.5	3.2	6.9	5.7	5.9
% of genome covered by CDS	2.2	1.5	1.6	1.1	4	3.4	4.6

Adaptive Radiation of the Flukes of Fasciolidae · doi:10.1093/molbev/msz204 MBE

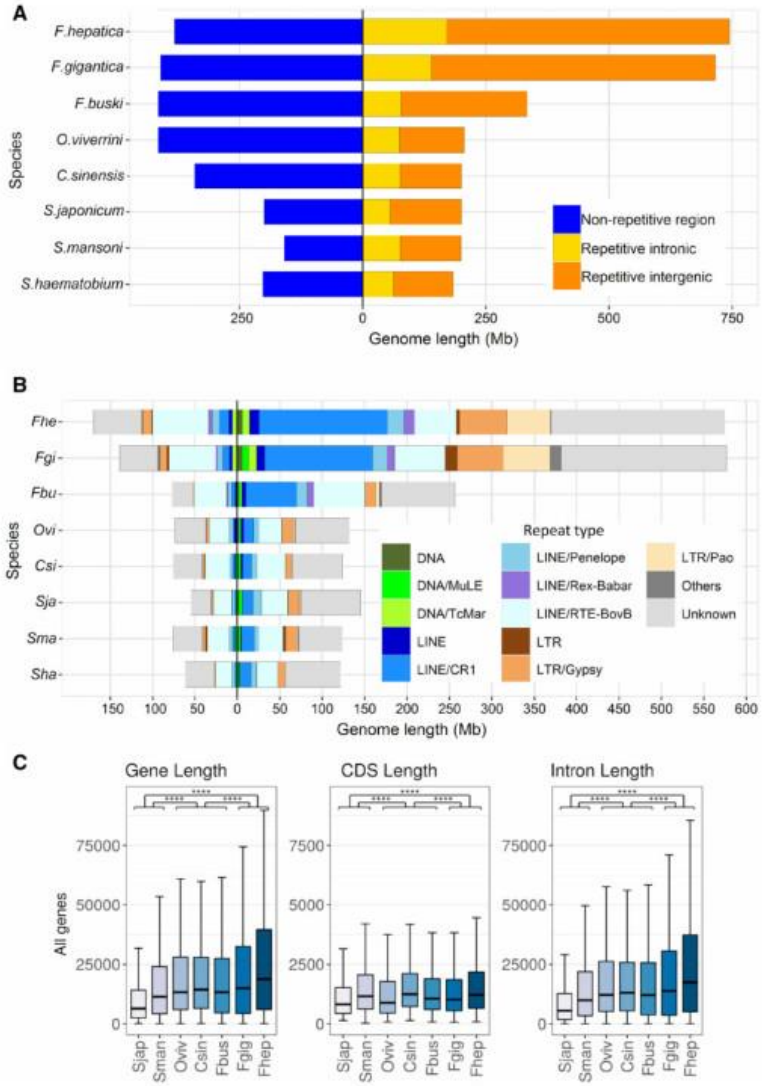


FIG. 1. Trematode genome nonrepetitive and repetitive contents. (A) Relative genome sizes of different trematodes, with nonrepetitive fraction (blue) represented to the left of the central axis, and repetitive fractions to the right indicating intronic (yellow) and intergenic (orange) repeats, respectively. (B) Classification of repeated elements present in intronic locations (to the left of the central axis) and intergenic repeats (to the right of the axis). Repeats classes are color coded. (C) Variation in gene, CDS, and intron lengths for the different trematode species. Statistically significant differences are indicated.

in intergenic regions, but other abundant long interspersed nuclear elements (LINEs) such as RTE-BovB are equally distributed between intronic and intergenic locations (fig. 1B). Gene length has also increased in Fasciolinae, due to longer introns, consistent with the increased presence of TEs within them (fig. 1C). As repeat elements degenerate through time, the sequence similarity measured as per-copy distances to consensus provides reasonable evidence that TE activity is currently low in the fasciolid genomes (supplementary fig. 1, Supplementary Material online).

The influence of TEs on animal genome size variation is widely accepted, and it is increasingly recognized that changes in TE activity may have a major effect on adaptation of populations and species facing novel habitats and large environmental perturbations (Chenais et al. 2012). TEs are potent sources of mutation that can rapidly create genetic variance,

especially following genetic bottlenecks and severe environmental changes, providing bursts of allelic and phenotypic diversity upon which selection can act (Stapley et al. 2015; Schrader and Schmitz 2018). Thus, our data lead to the hypothesis that TE-mediated genomic changes likely have contributed to the increased adaptive capacity of *Fasciola* spp. to new habitats and host species after their divergence from *Fasciolopsis*. Because TEs are highly mutagenic, either directly (e.g., insertions in coding or regulatory regions) or indirectly (e.g., chromosomal rearrangements), molecular countermeasures such as chromatin modifications suppress their activity and TE-derived transcripts are targeted for cleavage by npc silencing and *piwi*-interacting RNAs (Slotkin and Martienssen 2007). Interestingly the main silencing mechanism, the Piwi pathway, is incomplete in all parasitic flatworms including the Fasciolidae here analyzed and alternative silencing mechanisms based on conserved duplicated flatworm-specific Argonaute proteins (FLAgos) have been suggested (Skinner et al. 2014; Fontenla et al. 2017) (supplementary fig. 2, Supplementary Material online). Environmentally induced physiological or genomic stress can modulate TE activity by activating transposition or by inhibiting genomic silencing mechanisms (Rey et al. 2016), thus facilitating adaptive responses in species experiencing changed or diverse environments, as faced by invasive, pathogenic or parasitic species during the developmental cycle (Schrader and Schmitz 2018).

Lineage Diversification, Trait Evolution, and Historical Demography

Despite their public health and veterinary significance, the evolutionary history of fasciolid flukes remains understudied. The Fasciolidae may have originated in African proboscideans and later radiated in Eurasian herbivores (Lotfy et al. 2008). As the family diversified, host shifts occurred in both molluscan and mammalian hosts. There was also a switch in habitat within the definitive host from the small intestine to the liver (fig. 2). Morphological, ecological, and molecular phylogenetic data support relatively basal divergence for the lineage leading to *Fp. buski* and a derived position for the species of the genus *Fasciola* (Lotfy et al. 2008). Although the former genus has a planorbid snail as intermediate host, the latter genus exploits the Lymnaeidae, indicating a host-switch at some point in the lineage leading to *Fasciola*. The intestinal fluke *Fp. buski* has a large ventral sucker and simple bifurcated digestive ceca, whereas branched digestive caeca and a reduced ventral sucker are characteristics of the liver flukes. Accordingly, the comparative analysis of their genomes offers the opportunity to gain insights into the evolution of key processes of parasitism such as host selection, tissue tropism, and morphological adaptations. To investigate prospective correlations between biogeographical events and lineage diversification in the Fasciolidae, we constructed a dated phylogeny using a Bayesian multilocus coalescent method with 30 nuclear protein-coding genes (supplementary table 3, Supplementary Material online) and a node height prior taken from the age of Protostomia estimated in a previous fossil-calibrated eukaryote phylogeny (Parfrey et al. 2011). The molecular dating revealed that the split between the genera *Fasciolopsis* and *Fasciola* took place around 88.1 Ma (73.0–102.9, 95% highest posterior density [HPD]) in the late Cretaceous period, and the divergence between *Fa. hepatica* and *Fa. gigantica* occurred around 5.3 Ma (3.4–7.2, 95% HPD) near the Miocene–Pliocene boundary (supplementary fig. 3, Supplementary Material online). The estimated date for the divergence of *Fasciolopsis* and *Fasciola* seems rather ancient, given what we know about the evolution of mammals. Monotreme mammals were already established at 112–121 Ma (Rowe et al. 2008). Meredith *et al.* (Meredith et al. 2011) have placentals as arising around 100 Ma, mammals as a group arising much earlier than this, and ungulates a bit later. It is possible that fasciolids originated early in another group of mammals and switched into their current host groups later. Lotfy et al. (2008) suggested that fasciolids emerged in proboscideans, a hypothesis supported by the fact that the extant more basal fasciolid (*Protofasciola robusta*) lives in the small intestine of African elephants. The proboscideans evolved in Africa and radiated to Eurasia, and the host transition to ungulates might have occurred during this dispersion. To narrow down the time-window during which the most distinctive apomorphic traits of the Fasciolinae (genera *Fasciola* and *Fascioloides*) (i.e., lymnaeid snail hosts, hepatic habitats, branched intestinal caeca, dendritic testes and ovaries) originated, the ages of the stem node (the last common ancestor of Fasciolinae and Fasciolopsinae) and the crown node (the last common ancestor of all living members of Fasciolinae) were estimated using a published phylogeny of Fasciolidae that included *Parafasciolopsis fasciolaemorpha* (Fasciolopsinae) (Lotfy et al. 2008) and a whole-genome mitochondrial phylogeny including *Fascioloides magna* (Fasciolinae) (fig. 2 and supplementary fig. 4, Supplementary Material online). The data suggested that the intermediate host switch and shift from intestinal to hepatic habitats occurred between 65 Ma (stem node; 43.2–85.7, 95% HPD) and 55.9 Ma (crown node; 42.0–70.8, HPD) in the lineage leading to Fasciolinae. The profound climatic and ecological changes that occurred during this period (e.g., Cretaceous–Paleogene mass extinction and Paleocene–Eocene Thermal Maximum) may have contributed to the adaptive radiation of these flukes to new niches. The divergence time estimate of 5.3 Ma between *Fa. hepatica* and *Fa. gigantica* is substantially more recent than the previously suggested date of 19 Ma based on cathepsin L-like cysteine proteases (Irving et al. 2003). Most notably, our speciation time estimate coincides with the Miocene–Pliocene boundary that was characterized by a reduced faunal exchange between Africa and Eurasia (Bibi 2011), which may have contributed to the speciation process through an increased and sustained disruption of gene flow, resulting in two locally adapted sister taxa, that is, *Fa. gigantica* in Africa and *Fa. hepatica* in Eurasia. Based on comparative data on infectivity, life span, egg shedding and immunity among modern hosts, it was proposed that *Fa. hepatica* emerged in Eurasian ovicaprines, whereas *Fa. gigantica* originated in an African ruminant phylogenetically close to present-day bovines (Mas-Coma et al. 2009). These sister taxa still hybridize producing an intermediate form in regions where they currently occur sympatrically in Africa and Asia

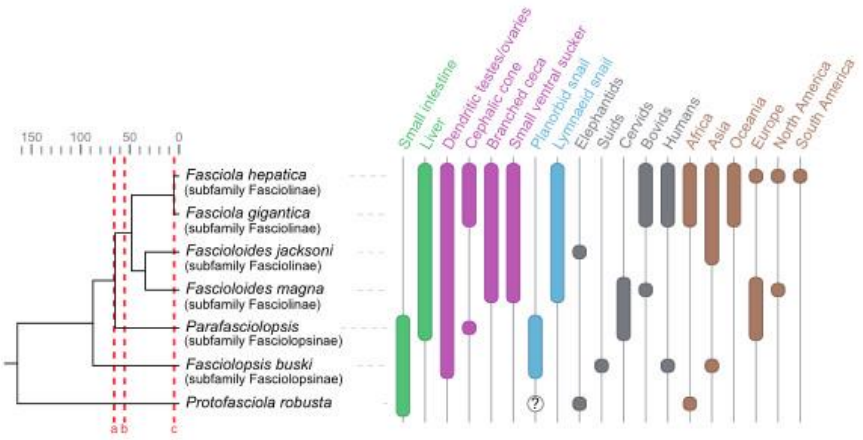


FIG. 2. Dated phylogeny and trait distribution in the Fasciolidae. Cladogram and parasite traits are based on Lotfy et al. (2008). Node ages in Ma are based on nuclear and mitochondrial molecular dating analyses: primary host body habitat (green), morphology (purple), snail host (blue), vertebrate host (gray), and geographic distribution (brown). Cretaceous–Paleogene mass extinction (a), Paleocene–Eocene Thermal Maximum (b), reduced faunal exchange between Africa and Eurasia (c).

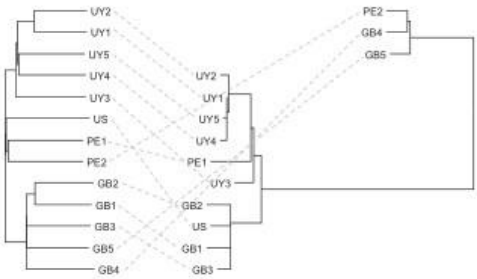


FIG. 3. Nuclear (left) and mitochondrial (right) phylogeny of *Fasciola hepatica* based on genome-wide SNPs showing mito-nuclear discordance. US, USA; UY, Uruguay; PE, Peru; GB, United Kingdom.

(Saijuntha et al. 2018). This has important implications for epidemiology such as the potential for crossover of anthelmintic resistance between the two species or emergence of more pathogenic variants.

Using genome-wide variation data of 13 samples from the United States, Uruguay, Peru, and the United Kingdom (supplementary table 1, Supplementary Material online), genetic structuring within *Fa. hepatica* was assessed. A moderate to low-level, geographic population structure, reflecting the country of origin, was observed in the nuclear genome (fig. 3) (e.g., F_{ST} between Uruguay and the United Kingdom: 0.094). Mitochondrial genome variation, however, revealed a striking pattern of discordance in which two deeply diverged clades were apparent. These haplogroups correspond to the two previously reported mitochondrial lineages in *Fa. hepatica* (defined based on a ~1.4-kb region that overlaps with cytochrome oxidase subunit III gene, tRNA-His gene, and cytochrome b gene) in geographically diverse European and Australian populations (Teofanova et al. 2011; Walker et al. 2011, 2012) (supplementary fig. 5, Supplementary Material online). Interestingly, both mitochondrial lineages were observed in Peru, suggesting that both haplogroups have been introduced to the New World, although their precise frequency and distribution in the Americas will need to be determined. Our molecular dating analysis indicated that these haplogroups originated around 1.1 Ma (0.07–2.7, 95% HPD) in the Pleistocene period (supplementary fig. 4, Supplementary Material online). Domesticated sheep (*Ovis aries*) fall into five mitochondrial haplogroups, whose radiation has been dated to be 0.92 ± 0.19 Ma, substantially predating the domestication event (~8–11 ka) (Meadows et al. 2011). It is thus tempting to hypothesize that a process of host–parasite codiversification has played a central role in the development of these haplogroups in *Fa. hepatica* where haplogroup formation in the host led to genetic structuring in the parasite. The observed mito-nuclear discordance is consistent also with high levels of nuclear gene flow and mixing of alleles within each metapopulation (Beesley et al. 2017) in contrast to the patterns of (nonrecombining) mitochondrial allelic diversity where ancestral haplotypes can persist in a population alongside derived forms.

We reconstructed the historical demography of *Fasciola* and found evidence of a rapid decline in its effective population size (i.e., a signature of founder effect) ~10–11 ka, which is consistent with a recent global spread associated with the ruminant domestication (fig. 4). An evolutionarily very recent spread of *Fasciola* spp. from their origin in the Eurasian Near East area (*Fa. hepatica*) and East Africa (*Fa. gigantica*) has been proposed based on the ribosomal DNA sequence diversity, and the spread of both species in postdomestication times likely has led to their present overlap in Africa and Asia where lymnaeid snails are suitable for the development of both

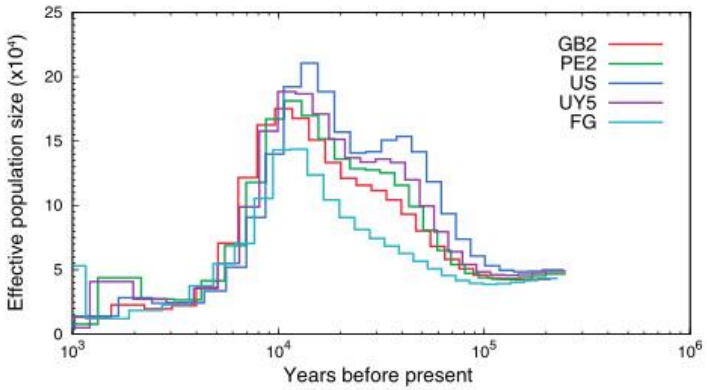


FIG. 4. *Fasciola* historical demography. The PSMC model was used to characterize historical demography by examining heterozygosity densities across the genome. US, USA; UY, Uruguay; PE, Peru; GB, United Kingdom; FG, *Fasciola gigantica*.

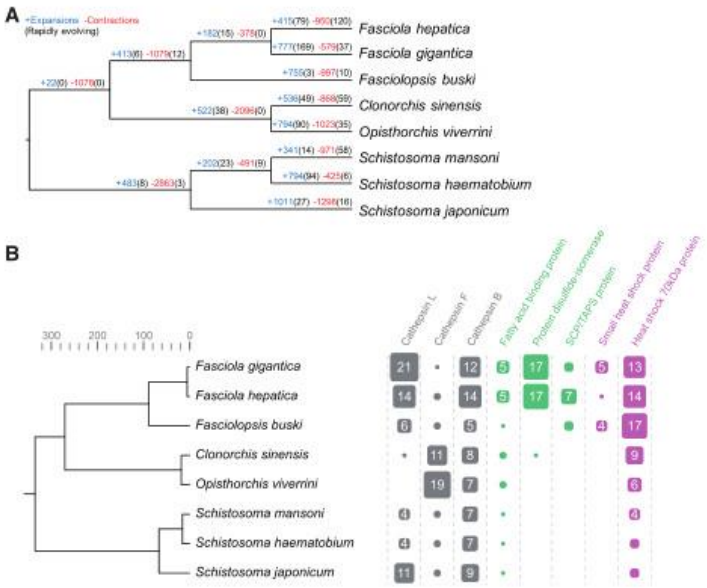


FIG. 5. Gene family dynamics in trematodes of medical importance. (*A*) Gene family gains/losses were modeled by estimating birth–death (λ) parameters while accounting for the species' phylogenetic history: when $P < 0.01$ they were considered rapidly evolving. (*B*) Gene families of interest that showed differential expansion/contraction in trematodes of medical importance. Gene family size and count are indicated by scaled boxes.

species (Mas-Coma et al. 2009). Abnormal ploidy and aspermic parthenogenesis in hybrids between the two species also suggest their near-complete genetic isolation and separate evolution in predomestication times (Mas-Coma et al. 2009).

Gene Family Dynamics in Medically Important Trematodes

To understand the genomic changes underlying phenotypic variation within fasciolid species and between families of digenetic trematodes of medical importance (Schistosomatidae, Opisthorchiidae, and Fasciolidae), we investigated large-scale differences in gene complements among lineages. Using orthologous groups (OGs) of genes identified across eight digenean species with sequenced genomes (supplementary table 4, Supplementary Material online), we modeled gene gain and loss while accounting for the species' phylogenetic history (Han et al. 2013) (fig. 5A). Based on the birth–death (λ) parameter estimate

(0.00575), statistical significance of the observed family size differences among taxa was assessed. Gene families of interest that displayed most pronounced differential expansions or contractions included the cathepsin cysteine proteases (L, F, and B families), fatty-acid-binding proteins, protein disulfide-isomerases and molecular chaperones, highlighting the significance of excretory–secretory proteins in lineage-specific adaptation (fig. 5B). Although several of these protein families were highlighted as relevant in adaptation by *Fa. hepatica* (McNulty et al. 2017), the present study provides evidence that their amplification occurs at or after the split between Fasciolinae and Fasciolopsinae.

The cathepsin superfamily encompasses several cysteine protease genes present in diverse flatworms with differential expression according to the parasite stage and multiple overlapping functions (Caffrey et al. 2018). Cathepsins constitute a substantial fraction of the excretory–secretory products and endow *Fasciola* with the ability to migrate through tissue and digest matrix, to break down proteins including hemoglobin for nutrition, and to modulate the immune response through digestion of immunoglobulins (Cancela et al. 2008; McGonigle et al. 2008; Robinson et al. 2008). In *Fa. hepatica*, a particular cathepsin L (FhCL3) with an unusual collagenolytic activity and several cathepsin Bs have been implicated in the early stages of invasion through the intestinal wall (Corvo et al. 2009; Cancela et al. 2010; Robinson et al. 2011; Meemon and Sobhon 2015). Five cathepsin B genes are annotated in *Fp. buski*, whereas 16 genes are annotated in *Fa. hepatica* and *Fa. gigantica*. Comparative analysis within cathepsins Bs (OG0000035) shows four conserved enzymes present in the three species as well as other trematodes, namely CatB9, CatB6, and a tandem duplication of CatB8 (fig. 6A and B). The remaining single gene in *Fp. buski* is at the base of an expansion of cathepsin Bs in the *Fasciola* spp. resulting in more than ten discrete genes. Given that almost all of these novel genes are shared between *Fa. hepatica* and *Fa. gigantica*, the genomic event that conferred this gene gain predated the separation of these species. A subclade of the novel cathepsin B genes occurring in *Fasciola* spp. corresponds to those that are differentially expressed during the intestine invasive stage in *Fa. hepatica* (fig. 6A and B). Similarly, within the cathepsin Ls (OG0000050), although some members are conserved (particularly CatL0, expressed in eggs in *Fa. hepatica*), an expansion and diversification process has taken place in the *Fasciolinae* lineage after it diverged from the *Fasciolopsinae* (fig. 6C). A tandem array of at least three cathepsin L genes with repeated exons that might produce diverse transcripts by alternative splicing is present in the *Fp. buski* genome (fig. 6D). Similar complex structures occur repeatedly in diverse contigs within the genomes of *Fa. gigantica* and *Fa. hepatica*, giving rise to more than a dozen cathepsin Ls (fig. 6C). It is noteworthy that the clade including the collagenolytic juvenile-specific cathepsin L3 with a suggested role in early invasion and the clade containing the mature cathepsin L1 members involved in immune evasion both seem to be related to the same *Fp. buski* cluster (fig. 6C). Although the still fragmentary status of the three assemblies does not allow to trace precisely the possible duplication events, it is plausible to consider that a region similar to the *Fp. buski* CatL cluster might have been the origin of the amplifications in the Fasciolinae lineage. In addition to amplifications of cathepsins B and L in genomes of the Fasciolinae, other independent amplifications of these gene families are observed in the Opisthorchiidae and Schistosomatidae. By contrast, cathepsin F genes (OG0000076) are amplified only in the Opisthorchiidae (Kang et al. 2010; Sripa et al. 2010) (fig. 6C). Notably, the exopeptidase cathepsin C (OG0007199), which is implicated in terminal processing of peptides in schistosomes (Hola-Jamriska et al. 1998; Caffrey et al. 2018) and *Clonorchis* (Liang et al. 2014), was absent from the three species of the Fasciolidae studied here. Cathepsin C is conserved in the Schistosomatidae and Opisthorchiidae (fig. 6A).

Asparaginyl endopeptidases, better known as legumains have been implicated in the maturation of cathepsin proenzymes (Robinson et al. 2009). We observed that *Fa. gigantica* shares the amplification of legumains (OG0000019) already described for *Fa. hepatica* (McNulty et al. 2017). By contrast, a restricted set of only three legumain genes is present in *Fp. buski* (supplementary fig. 6A, Supplementary Material online). The coincident amplification of cathepsins and legumains in the genus *Fasciola* might reflect the likely diversification of regulatory legumains involved in the maturation of the diverse cathepsin proenzymes. Other regulatory proteins of cysteine protease activity, such as the cystatin family of cysteine-protease inhibitors occurs broadly and similarly across the Fasciolidae (supplementary fig. 6B, Supplementary Material online).

Analysis of gene gain and loss showed that an orthogroup corresponding to CAP domain-containing proteins (OG0001149) was extensively amplified in the Fasciolidae compared with other trematodes. Proteins with this domain, which also is known as the SCP/TAPS domain, appear to be involved in helminth parasite–mammalian host interactions. These proteins are excreted and secreted, and are differentially expressed in the parasitic stages of hookworms (Datu et al. 2008). In platyhelminths, most studies have been undertaken in *Schistosoma mansoni*, in which different superfamily members are expressed in different developmental stages. Some CAP domain-containing proteins are specifically expressed in the intrasnail stages or in the intramammal stages (Chalmers et al. 2008; Rofatto et al. 2012). Several specific duplications have been reported for this superfamily among diverse helminth taxa (Tang et al. 2014; Hunt et al. 2016; Costabile et al. 2018; International Helminth Genomes Consortium 2019). In particular, these duplications are lineage specific, implying that each lineage has duplicated and maintained particular superfamily members, influenced by the biological differences in hosts and/or life cycle stages. To explore this further, all the orthogroups with genes annotated as CAP domain-containing proteins from trematodes were retrieved and analyzed. The phylogenetic analysis (fig. 7) reveals that although several orthogroups are lineage specific, most of them are phylogenetically related to other lineage-specific orthogroups. The Fasciolidae-specific OG0001149 is phylogenetically related to five Opisthorchiidae-specific orthogroups that diverged after the Fasciolidae/Opisthorchiidae split.

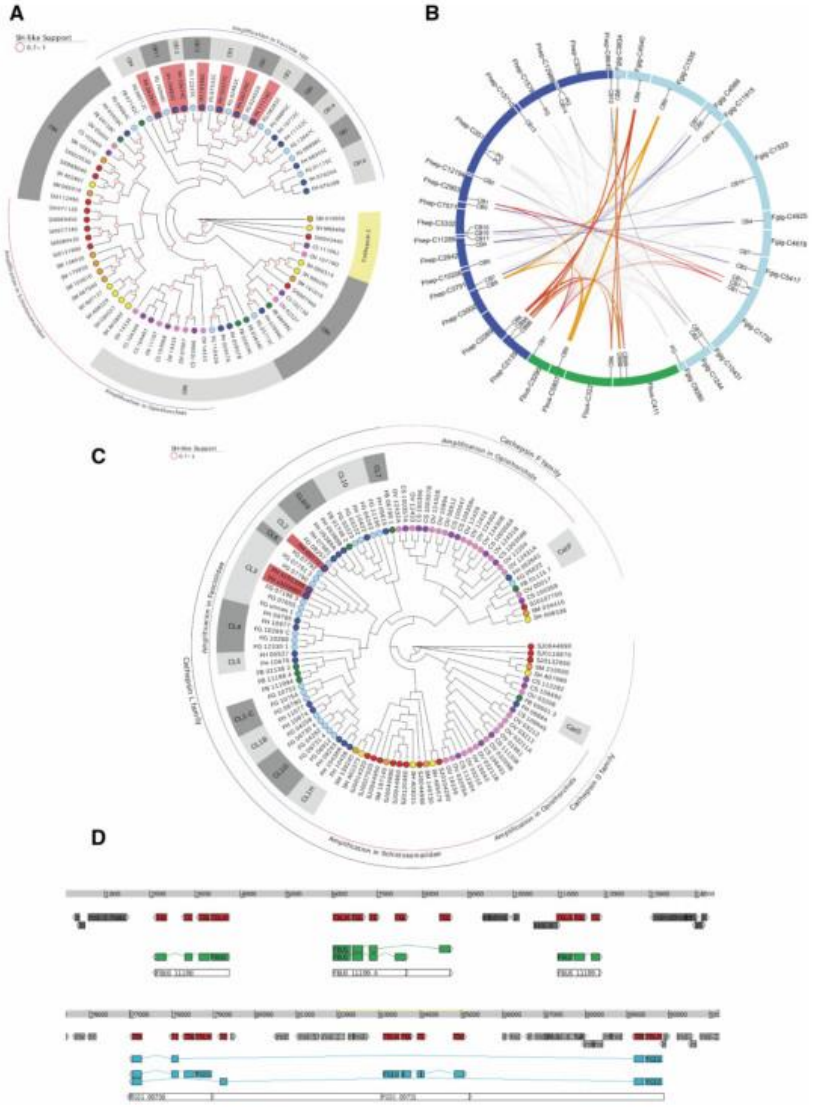


FIG. 6. Cathepsins in trematodes of medical importance. (*A*) Phylogenetic tree of cathepsins B and C from the family Fasciolidae and other trematodes. (*B*) Circos representation of the cathepsin B-containing contigs in *Fasciolopsis buski* (green), *Fasciola hepatica* (blue), *Fasciola gigantica* (light blue). The relative position of the diverse cathepsin B genes is indicated. Conservation between *Fp. buski* and the Fasciolinae cathepsin Bs is highlighted in black and those between the Fasciolinae cathepsin Bs in red. (*C*) Phylogenetic tree of cathepsins L and F from the family Fasciolidae and other trematodes. (*D*) Visualization of the gene structure of the cathepsins L cluster present in *Fp. buski* contig 4155 (top) and *Fa. gigantica* contig 4972 (bottom). Conserved regions (as detected by BLAST) are indicated in red, gene CDS are indicated in green (*Fp. buski*) and sky blue (*Fa. gigantica*). Repetitive sequences are indicated in gray. Note the presence of repeated exons that allow alternative transcripts from the same gene.

Several of their members are preferentially expressed in the invasive stage. Four other small OGs occur exclusively in the Fasciolidae, expressed across diverse life stages such as those presented in the phylogenetically conserved OG0000070. As reported for other helminths, different lineages have duplicated specific groups of genes, and this may be related to the differences in their life cycles and host immune-evasion mechanisms.

Other excretory–secretory product components like the fatty-acid-binding proteins have potent immunomodulatory

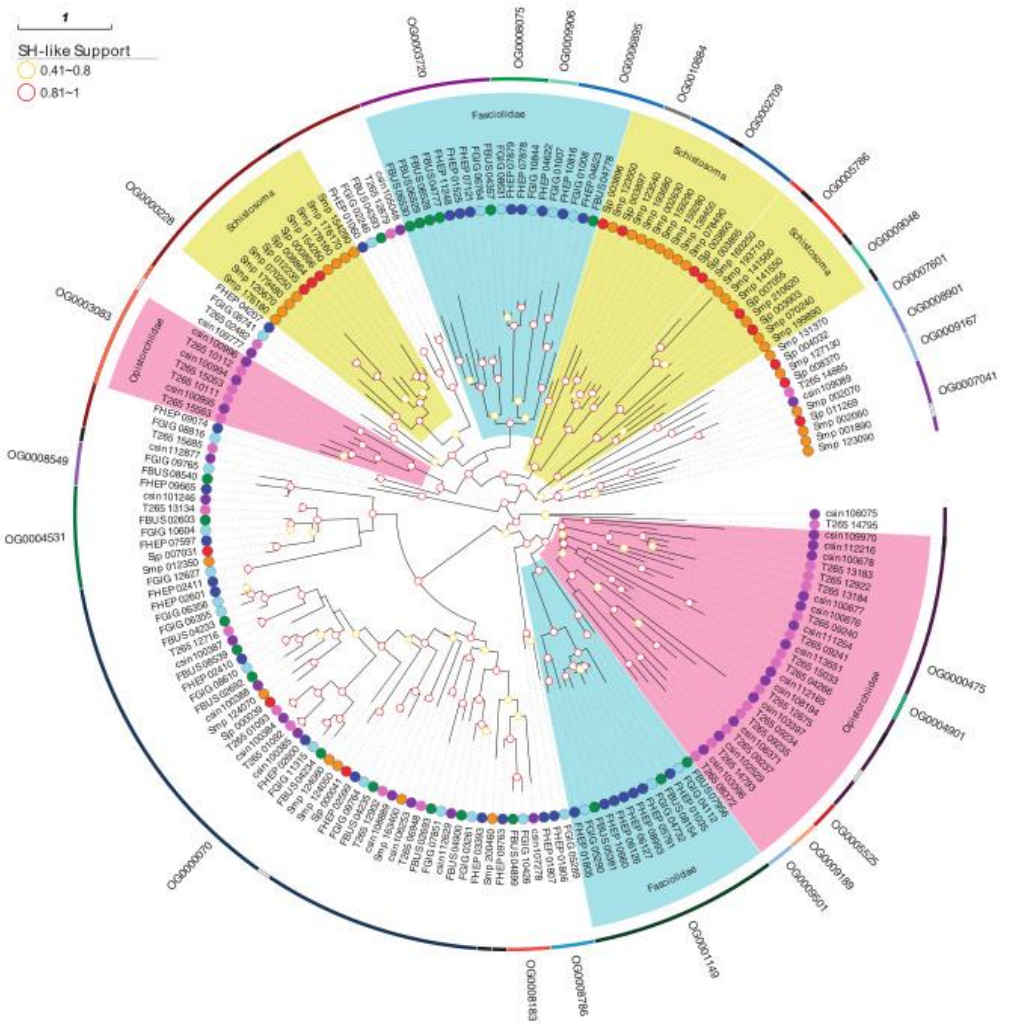


FIG. 7. CAP domain-containing proteins of the Fasciolidae and other trematodes. Phylogeny of CAP domain-containing proteins from trematodes is depicted. The outer ring indicates the diverse orthogroups containing CAP domain proteins. Lineage-specific amplifications are highlighted (pink: Opisthorchiidae, light blue: Fasciolidae, yellow: Schistosomatidae).

effects inhibiting activation of TLR4 and macrophages, allowing the parasite to establish a long-lasting infection (Figueroa-Santiago and Espino 2014; Martin et al. 2015) (supplementary fig. 7A, Supplementary Material online). Lipid transporters, such as the NPC2 protein, were also amplified in the Fasciolinae, confirming observations on *Fa. hepatica* (McNulty et al. 2017). This may represent a metabolic adaptation to the rich lipid conditions characteristic of the biliary tract, the predilection site for the adult stage of *Fa. gigantica* and *Fa. hepatica* (supplementary fig. 7B, Supplementary Material online). Heat-shock proteins expanded in fasciolid flukes (supplementary fig. 8, Supplementary Material online). Flatworm heat-shock proteins perform crucial roles in protein homeostasis and protection from stress-induced damage. They also participate in development and modulation of immune responses through their immunogenic and immunomodulative properties (Yang et al. 2012; Chung et al. 2017). The protein disulfide isomerases described in *Fasciola* may contribute to survival of the parasite in an environment like the bile ducts, with high oxidative stress and protein damage (Salazar-Calderon et al. 2003).

Genome-Wide Signatures of Adaptive Evolution in *Fasciola*

We examined genome-wide signatures of selection from patterns of genetic polymorphism (within *Fa. hepatica*) and divergence (between *Fa. hepatica* and *Fa. gigantica*) using SnIPRE, a Bayesian implementation of the McDonald and Kreitman (MK) test developed for genome-wide analysis (Eilertson et al. 2012). The Kolmogorov–Smirnov test was performed to identify enriched gene ontology (GO) terms among genes with high selection coefficient γ (indicating positive selection) (table 2). A marked enrichment was observed in GO terms related to G-protein-coupled receptors (GPCRs), indicating that GPCRs are more likely to be under positive and/or relaxed purifying selection than other genes, suggesting their involvement in adaptive evolution of these flukes. Using *S. mansoni* GPCRs ($n = 115$) as the reference (Hahnel et al. 2018), putative *Fa. hepatica* (117), *Fa. gigantica* (126), and *Fp. buski* (142) GPCRs were assigned into classes A, B, C, and F (fig. 8A and 8B). Class A GPCRs were further classified into aminergic receptors (including orphan amines, biogenic amines, and opsins), peptidergic receptors (including neuropeptide Y, neuropeptide F, and neuropeptide FF, and FMRFamide-like peptide), and the platyhelminth-specific rhodopsin-like orphan-family. Phylogenetic analysis indicated Fasciolidae-specific expansions among biogenic and orphan amine receptors (fig. 8B). The selection coefficient (γ) of GPCRs, on average, was higher than those of all other genes (Kruskal–Wallis test, $P = 1.7 \times 10^{-10}$), and this pattern was observed across all classes of GPCRs except for opsins (fig. 8C). GPCRs translate sensory inputs into cellular responses and are thus crucial for tuning physiology and behavior in response to the environment. Expansion of GPCR odorant receptors (ORs), for example, increases the repertoire of odorant signals that species detect, allowing them to occupy new ecological niches (e.g., terrestrial vs. aquatic vertebrates) (Kishida 2008). As a comparison, genomes of great apes contain about 1,000 OR genes, of which one-third appear to be functional, as acquisition of trichromatic color vision in primates caused the parallel pseudogenization of OR genes (the "vision-priority" hypothesis) (Gilad et al. 2004). Hence, it is reasonable to hypothesize that differential gene family expansion and

Table 2. Enriched GO Terms among *Fasciola* Genes with Signatures of Adaptive Evolution (i.e., high gamma values indicating positive selection).

GO Term		Kolmogorov–Smirnov Test *P*-Value
Biological process	GPCR signaling pathway	9.3E-09
	Potassium ion transport	3.0E-04
	Transmembrane transport	3.6E-03
Molecular function	GPCR activity	1.7E-11
	Ionotropic glutamate receptor activity	3.4E-04
	Extracellular-glutamate-gated ion channel activity	3.8E-04
	Calcium ion binding	7.3E-04
	Potassium channel activity	8.9E-04
	G-protein-coupled peptide receptor activity	9.4E-03
Cellular component	Integral component of membrane	2.6E-11
	Membrane	4.9E-08

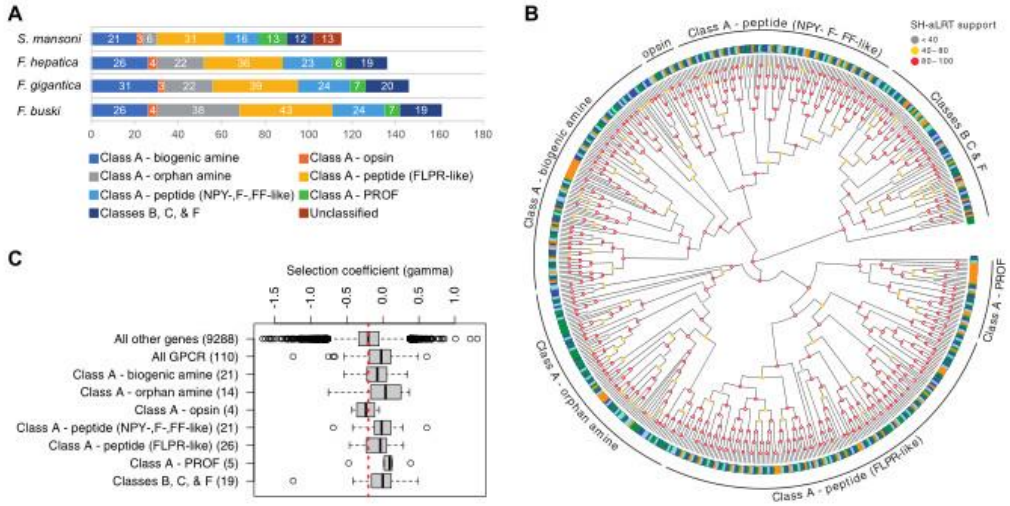


FIG. 8. G-protein-coupled receptor (GPCR). (*A*) total GPCR gene counts in each genome. (*B*) phylogeny of GPCRs and classification into different classes using *Schistosoma mansoni* GPCRs as the reference. *Fasciolopsis buski* (green), *Fasciola hepatica* (blue), *Fasciola gigantica* (light blue), *S. mansoni* (yellow) (*C*) selection coefficient of GPCRs showing signatures of positive and/or relaxed purifying selection. Positive gamma values indicate positive selection and negative gamma values indicate purifying selection.

positive selection of GPCRs, along with parallel pseudogenization of genes under relaxed purifying selection, represent one mechanism by which species of *Fasciola* adapted to a new ecological niche.

Conclusions

This comparative analysis provides novel insight into the biology and evolution of Fasciolidae and other fluke families of medical importance. Rapid climatic and ecological changes may have contributed to the adaptive radiation of fasciolids, which were accompanied by lineage-specific gene family expansions and differential rates of molecular evolution among different gene families. The genomic resources that these studies have provided should enhance development of novel interventions and diagnosis, and underpin genomic epidemiologic investigation of new disease outbreaks, virulence and drug resistance.

Materials and Methods

Parasite Specimens

Genomes of *Fa. gigantica* and *Fp. buski* were assembled de novo using specimens from Uganda (cattle liver) and Vietnam (pig intestine), respectively. Adult *Fa. gigantica* were isolated from the livers of Ankole cattle at the abattoir of Fort Portal in the Western Region of Uganda in 1993. The samples of *Fp. buski* belong to the HT strain and were collected in Ha Tay town, near Hanoi, Vietnam. Parasites were stored in 80% ethanol at −20 °C until isolation of nucleic acids. Whole-genome resequencing data for *Fa. hepatica* were generated using specimens collected from Uruguay (Montevideo) and Peru (Cusco), which were analyzed together with published samples from the United States (McNulty et al. 2017) and the United Kingdom (Cwiklinski et al. 2015).

DNA/RNA Isolation, Genome Sequencing, and Assembly

Nucleic acids were extracted using QIAGEN DNeasy (DNA) and RNeasy (RNA) mini kits and cleaned up by ethanol precipitation for *Fa. gigantica*. For *Fp. buski*, ethanol-preserved adult worms were chopped up using a scalpel blade, and genomic DNA (gDNA) was extracted and purified using the kit E.Z.N.A and SQ Tissue DNA Kit (Omega Bio-tek), and the yield and purity were assessed by Bio-Analyzer as described (McNulty et al. 2017). DNA small-insert (fragment) and mate-pair (jump) libraries were constructed using gDNA extracted from an individual adult worm and sequenced on Illumina HiSeq platform (2X100bp) as described (McNulty et al. 2017). Pacific Biosciences sequencing (PacBio RS II P5-C3/P6-C4, 20-kb library) was performed to complement the Illumina data and improve scaffolding and gap-filling. ALLPATHS-LG (release 44837) (Gnerre et al. 2011) was used to assemble the Illumina reads after adapter sequences were removed with Trimmomatic v0.36 (Bolger et al. 2014). To scaffold the assembled contigs, SSPACE-standard v3.0 (Boetzer et al. 2011) and SSPACE-longread v1.1 (Boetzer and Pirovano 2014) were sequentially run using the Illumina and PacBio reads, respectively. Gapfiller v1.10 (Boetzer and Pirovano 2012) and PBJelly v15.8.24 (English et al. 2012) were used to close gaps, and the resulting assembly was error-corrected using Pilon v1.20 (Walker et al. 2014) and screened for contaminants using blobtools v0.9.19. Mitochondrial genomes (mtDNA) were assembled with NOVOPlasty v2.6.3 using the Illumina fragment reads (Dierckxsens et al. 2017).

Genome Annotation

The nuclear genomes of *Fa. gigantica* and *Fp. buski* were annotated using the MAKER pipeline v2.31.8 (Holt and Yandell 2011). The published *Fa. hepatica* genome assembly (GenBank accession number: GCA_002763495.1) was reannotated using the same pipeline. Repetitive elements were soft-masked with RepeatMasker v4.0.6 using a species-specific repeat library created by RepeatModeler v1.0.8, RepBase repeat libraries (Bao et al. 2015), and a list of known TEs provided by MAKER (Holt and Yandell 2011). From the NCBI Sequence Read Archive, *Fa. hepatica* (SRR2038730, SRR2038734, SRR2038743, SRR2039050, SRR2039051, ERS524681-3, ERS524685-91, ERS524693, ERS524696), *Fa. gigantica* (SRR094761), and *Fp. buski* (SRR941773, SRR5929441) RNA-seq data were obtained. Additional RNA-seq was performed for *Fa. gigantica* in biological duplicates (adult stage) on the Illumina HiSeq platform (2X100bp TruSeq Stranded mRNA library) to support the genome annotation. After adapter trimming using Trimmomatic v0.36 (Bolger et al. 2014), RNA-seq reads were aligned to their respective genome assemblies using HISAT2 v2.0.5 (Kim et al. 2015) with the –dta option and subsequently assembled using StringTie v1.2.4 (Pertea et al. 2015). The resulting alignment and transcript assembly were used by BRAKER (Hoff et al. 2016) and MAKER pipelines, respectively, as extrinsic evidence data. In addition, mRNA and EST sequences for each species were retrieved from NCBI and passed to MAKER as transcript evidence. Protein sequences from UniRef100 (UniProt Consortium 2017) (Trematoda-specific, $n = 205{,}161$) and WormBase ParaSite WBPS7 (Howe et al. 2017) (*Clonorchis sinensis* PRJDA72781, *Opisthorchis viverrini* PRJNA222628, *S. mansoni* PRJEA36577) were provided to MAKER as protein homology evidence. Ab initio gene predictions from BRAKER v1.9 (Hoff et al. 2016) and AUGUSTUS v3.2.2 (trained by BRAKER and run within MAKER) were refined using the transcript and protein evidence. Previously unpredicted exons and untranslated regions were added, and split models were merged. The best-supported gene models were chosen based on Annotation Edit Distance (Eilbeck et al. 2009). To reduce false positives, gene predictions without supporting evidence were excluded during building the final annotation, with the exception of those encoding Pfam domains, as detected by InterProScan v5.19 (Jones et al. 2014). These Pfam domain sequences were rescued to improve the overall annotation accuracy by balancing sensitivity and specificity (Holt and Yandell 2011; Campbell et al. 2014). Unfiltered set of gene models are available upon request. PANNZER2 (Koskinen et al. 2015) and sma3s v2 (Casimiro-Soriguer et al. 2017) were employed to name gene products. The completeness of annotated gene sets was assessed using

BUSCO v3.0 (Waterhouse et al. 2017). GO and KEGG annotations were performed using InterProScan v5.19 (Jones et al. 2014) and BlastKOALA (Kanehisa et al. 2016), respectively. rRNA and tRNA were annotated using RNAmmer v1.2.1 (Lagesen et al. 2007) and tRNAscan-SE v1.23 (Lowe and Eddy 1997), respectively. Mitochondrial genomes were annotated using MITOS2 (Bernt et al. 2013).

Repeat Analysis

RepeatModeler v1.0.8 (with WU-BLAST as its search engine) was used to build, refine, and classify consensus models of putative interspersed repeats for each species. With the resulting repeat libraries, genomic sequences were screened using RepeatMasker v4.0.6 in "slow search" mode to generate a detailed annotation of the interspersed and simple repeats. Per-copy distances to consensus were calculated (Kimura two-parameter model, excluding CpG sites) and were plotted as repeat landscapes where divergence distribution reflected the activity of TEs on a relative time scale per genome using the calcDivergenceFromAlign.pl and createRepeatLandscape.pl scripts included in the RepeatMasker package. Intergenic and intragenic repeats were identified by comparing the genic and repeat annotation coordinates. The distribution of gene lengths, coding, and intronic sequences for different species were calculated, and the statistical significance of the observed size differences among taxa was assessed.

Molecular Divergence Dating Analysis

Diversification timeframe for Fasciolidae was estimated using StarBEAST2, a multiindividual, multilocus coalescent method implemented in *BEAST v2.4.7 (Ogilvie et al. 2017). To infer times to the most-recent common ancestor for *Fasciolopsis–Fasciola* and *Fa. hepatica–Fa. gigantica*, 30 single-copy protein-coding nuclear loci were selected randomly from OGs of genes identified across 11 protostome taxa using OrthoFinder v1.1.4 (Emms and Kelly 2015). In *S. mansoni*, which is currently the only trematode for which a chromosome-level assembly is available, all 30 loci are located on autosomes and at least 500 kb apart from each other, suggesting that these genes are unlinked and evolve independently. For each orthologous gene group, PRANK (Loytynoja and Goldman 2005) was used within the framework of GUIDANCE2 (Sela et al. 2015) to generate codon-based multiple sequence alignments with removal of unreliable columns (below the default cutoff of 0.93). A relaxed molecular clock analysis was run with the Calibrated Yule model (Heled and Drummond 2012) and bModelTest (Bouckaert and Drummond 2017) as the tree prior and the site model, respectively. A most-recent common ancestor prior was set on the root height for the species tree, taken from the age of Protostomia estimated in a previous fossil-calibrated eukaryote phylogeny (Parfrey et al. 2011)—a normal prior with mean = 632 Ma, SD = 29.3 Ma. Twenty independent Markov chain Monte Carlo chains were run, each for 2×10^9 generations, sampling every 10^5 states. The topology was held constant when estimating other parameters, including divergence times. Convergence, mixing, and ESS values for each parameter were assessed using Tracer v1.6 (Rambaut et al. 2018). The last 2×10^4 trees sampled from the stationary posterior distribution of each of the 20 Markov chain Monte Carlo runs were combined using LogCombiner v1.8.4, and a maximum clade credibility tree was generated using TreeAnnotator v2.4.7. An earlier fasciolid phylogeny (based on 28S, ITS1, ITS2, NAD1) (Lotfy et al. 2008) and a whole-genome mitochondrial phylogeny were dated, using the same methods, but adjusting gene ploidy for mitochondrial loci, and using the resulting divergence time estimates (*Fasciolopsis–Fasciola*: mean = 88.1 Ma, SD = 7.7 Ma; *Fa. hepatica–Fa. gigantica*: mean = 5.29 Ma, SD = 0.9 Ma) as node age priors.

Gene Family Evolution

OGs of genes were inferred with OrthoFinder v1.1.4 (Emms and Kelly 2015) using the longest isoform for each gene. The CAFE method (Han et al. 2013) was employed to model gene gain and loss while accounting for the species' phylogenetic history based on an ultrametric species tree, generated as described above for molecular dating, and the number of gene copies found in each species for each gene family. Birth–death (λ) parameters were estimated, and the statistical significance of the observed family size differences among taxa was assessed. Gene trees from selected Fasciolidae-enriched families were generated by aligning with MAFFT (Katoh and Standley 2014) and PHYML tree building (Guindon et al. 2010), with models predicted by Model Generator (Keane et al. 2006). Trees were visualized with Evolview (He et al. 2016). Cryptic or partial copies of gene family members were captured by tBLASTn on the genomes with coding sequence (CDS) of OG members as queries, and visualized and inspected manually with Artemis (Carver et al. 2012).

Genome Variation Analysis

Individual worm gDNA reads from Illumina small-insert libraries were aligned to the corresponding genome assemblies using BWA-MEM v0.7.15 (Li and Durbin 2009). Polymerase chain reaction and optical duplicates were removed using picard tools v2.8.3 (http://broadinstitute.github.io/picard; last accessed February 9, 2017). Reads that aligned on the edges of indels were realigned to achieve the most consistent placement. Single-nucleotide variants were called via local de novo assembly of haplotypes using GATK v3.7 (McKenna et al. 2010) and quality-filtered as previously described (Van der Auwera et al. 2013; McNulty et al. 2017). For mitochondrial loci, sample ploidy was set to 1. Full mitochondrial haplotype sequences were reconstructed based on single nucleotide polymorphisms (SNPs) for individual *Fa. hepatica* samples and were used to build maximum-likelihood phylogenetic trees with RAxML v8.2.9 under GTRCAT model and with autoMRE bootstrapping (Stamatakis 2014). The pairwise genetic distance between samples (1-ibs, identity by state) was computed using PLINK v1.90 after excluding loci with missing genotypes in any of the samples for both the nuclear and mitochondrial genomes. The computed distances were subsequently used to generate a tanglegram based on neighbor-joining trees in Dendroscope v3.5.9. To correlate

our mitochondrial genome sequences with the previously published ~1.4-kb partial mitochondrial sequences overlapping with cytochrome oxidase subunit III (cox3) gene, tRNA-His gene, and cytochrome b (cytb) gene, the corresponding regions were extracted from the genomes and subjected to phylogenetic analysis using MAFFT and RAxML (Katoh and Standley 2014; Stamatakis 2014). Variants were annotated according to their genomic locations and predicted coding effects using SnpEff (Cingolani et al. 2012).

Historical Demography

The Pairwise Sequentially Markovian Coalescent (PSMC) model (Li and Durbin 2011) was used to characterize historical demography by examining heterozygosity densities in 100-bp sliding windows across the genome. Consensus genomic sequence data (contig length > 50 kb) were generated for each diploid individual worm using SAMtools/BCFtools (mapping quality > 20; base quality > 20; median × 0.33 < depth of coverage < median × 2) (Li et al. 2009) based on the deduplicated and indel-realigned alignments. Because PSMC is sensitive to variation in coverage depth, it was run twice for each individual using parameters -N25 -t15 -r5 -p "4+25*2+4 + 6." First, it was run utilizing all mapped sequence data and then utilizing data down-sampled to 10× coverage using the DownsampleSam tool (picard). Results were scaled by a mutation rate estimated based on genome size (1.6×10^{-8} per base pair per generation for *Fasciola* spp.) (Crellen et al. 2016) and a generation time of 0.25 years (Phalee et al. 2015), resulting in distributions of N_e through time. Subsequently, 100 PSMC bootstrap replicates were performed for both full-coverage and down-sampled data to confirm consistent distributional patterns.

Genome-Wide Analysis of Signatures of Adaptive Evolution

To identify genome-wide signatures of selection from patterns of genetic polymorphism (within *Fa. hepatica*) and divergence (between *Fa. hepatica* and *Fa. gigantica*), we performed the MK test within a Bayesian framework using SnIPRE (Eilertson et al. 2012). The MK table of fixed or polymorphic replacement and silent substitutions was prepared using PopGenome (Pfeifer et al. 2014) based on exonic SNPs (identified using the GATK pipeline as described above) in 13 *Fa. hepatica* and 1 *Fa. gigantica* samples. Gene loci with >8× sequencing coverage over >70% CDS length in all samples were included in the analysis ($n = 9{,}398$). The Kolmogorov–Smirnov test was performed to identify enriched GO terms among gene with high gamma (positive selection).

GPCR Annotation and Analyses

Based on available high-confidence GPCR sequences in *S. mansoni* (Hahnel et al. 2018) and *Fa. hepatica* (McVeigh et al. 2018), orthologous GPCR sequences were identified in our fasciolid genomes by Inparanoid v4.1 (Sonnhammer and Ostlund 2015) and Reciprocal Best Hits methods. These sequences were supplemented with those annotated with the GO term GO:0004930 (GPCR activity) by InterProScan v5.19. False-positive GPCR sequences were removed through manual curation involving an iterative process of inspecting multiple sequence alignments, building phylogenetic trees, and identifying anomalous phylogenetic placements using the *S. mansoni* phylogeny (Hahnel et al. 2018) as the reference. Multiple sequence alignments were generated using TM-Aligner (Bhat et al. 2017), and phylogenetic trees were inferred by maximum likelihood using IQ-TREE (Nguyen et al. 2015) with the best-fit model automatically selected (Kalyaanamoorthy et al. 2017) and the SH-aLRT test (Guindon et al. 2010) performed with 10,000 replicates.

Command lines used to perform the analyses are made available in supplementary text 1, Supplementary Material online.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Sequencing of the genomes was supported by the "Sequencing the etiological agents of the Food–Borne Trematodiases" project (National Institutes of Health—National Human Genome Research Institute award number U54HG003079). Comparative genome analysis was funded by grants National Institutes of Health—National Institute of Allergy and Infectious Diseases AI081803 and National Institutes of Health—National Institute of General Medical Sciences GM097435 to M.M.

Author Contributions

Conceptualization: M.M., P.J.B.; formal analysis: Y.-J.C., J.F.T., S.F., A.C.; funding acquisition: P.J.B., J.F.T., M.M.; methodology: J.F.T., M.M.C., P.U.F., T.H.L., D.B., M.M.; resources: M.M.C., T.H.L., P.U.F., M.M.; visualization: Y.-J.C., J.F.T., S.F., A.C.; writing—original draft: Y.-J.C., J.F.T., M.M.; writing—review & editing: M.M., D.B., P.J.B., P.U.F., J.F.T., M.M.C.

References

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 6:11.

Beesley NJ, Williams DJ, Paterson S, Hodgkinson J. 2017. *Fasciola hepatica* demonstrates high levels of genetic diversity, a lack of population structure and high gene flow: possible implications for drug resistance. *Int J Parasitol*. 47(1):11–20.

Bernt M, Donath A, Juhling F, Externbrink F, Florentz C, Fritzsch G, Putz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 69(2):313–319.

Bhat B, Ganai NA, Andrabi SM, Shah RA, Singh A. 2017. TM-Aligner: multiple sequence alignment tool for transmembrane proteins with reduced time and improved accuracy. *Sci Rep*. 7(1):12543.

Bibi F. 2011. Mio-Pliocene faunal exchanges and African biogeography: the record of fossil bovids. *PLoS One* 6(2):e16688.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4):578–579.

Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol*. 13(6):R56.

Boetzer M, Pirovano W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* 15:211.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
Bouckaert RR, Drummond AJ. 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol Biol*. 17(1):42.
Caffrey CR, Goupil L, Rebello KM, Dalton JP, Smith D. 2018. Cysteine proteases as digestive enzymes in parasitic helminths. *PLoS Negl Trop Dis*. 12(8):e0005840.
Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, et al. 2014. MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol*. 164(2):513–524.
Cancela M, Acosta D, Rinaldi G, Silva E, Duran R, Roche L, Zaha A, Carmona C, Tort JF. 2008. A distinctive repertoire of cathepsins is expressed by juvenile invasive *Fasciola hepatica*. *Biochimie* 90(10):1461–1475.
Cancela M, Ruetalo N, Dell'Oca N, da Silva E, Smircich P, Rinaldi G, Roche L, Carmona C, Alvarez-Valin F, Zaha A, et al. 2010. Survey of transcripts expressed by the invasive juvenile stage of the liver fluke *Fasciola hepatica*. *BMC Genomics* 11(1):227.
Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28(4):464–469.
Casimiro-Soriguer CS, Munoz-Merida A, Perez-Pulido AJ. 2017. Sma3s: a universal tool for easy functional annotation of proteomes and transcriptomes. *Proteomics* 17(12):1700071.
Chalmers IW, McArdle AJ, Coulson RM, Wagner MA, Schmid R, Hirai H, Hoffmann KF. 2008. Developmentally regulated expression, alternative splicing and distinct sub-groupings in members of the *Schistosoma mansoni* venom allergen-like (SmVAL) gene family. *BMC Genomics* 9:89.
Chenais B, Caruso A, Hiard S, Casse N. 2012. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene* 509(1):7–15.
Chung EJ, Jeong YI, Lee MR, Kim YJ, Lee SE, Cho SH, Lee WJ, Park MY, Ju JW. 2017. Heat shock proteins 70 and 90 from *Clonorchis sinensis* induce Th1 response and stimulate antibody production. *Parasit Vectors*. 10(1):118.
Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6(2):80–92.
Corvo I, Cancela M, Cappetta M, Pi-Denis N, Tort JF, Roche L. 2009. The major cathepsin L secreted by the invasive juvenile *Fasciola hepatica* prefers proline in the S2 subsite and can cleave collagen. *Mol Biochem Parasitol*. 167(1):41–47.
Costabile A, Koziol U, Tort JF, Iriarte A, Castillo E. 2018. Expansion of cap superfamily proteins in the genome of *Mesocestoides corti*: an extreme case of a general bilaterian trend. *Gene Rep*. 110–120.
Crellen T, Allan F, David S, Durrant C, Huckvale T, Holroyd N, Emery AM, Rollinson D, Aanensen DM, Berriman M, et al. 2016. Whole genome resequencing of the human parasite *Schistosoma mansoni* reveals population history and effects of selection. *Sci Rep*. 6:20954.
Cwiklinski K, Dalton JP, Dufresne PJ, La Course J, Williams DJ, 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.
Cwiklinski K, O'Neill SM, Donnelly S, Dalton JP. 2016. A prospective view of animal and human Fasciolosis. *Parasite Immunol*. 38(9):558–568.
Datu BJ, Gasser RB, Nagaraj SH, Ong EK, O'Donoghue P, McInnes R, Ranganathan S, Loukas A. 2008. Transcriptional changes in the hookworm, *Ancylostoma caninum*, during the transition from a free-living to a parasitic larva. *PLoS Negl Trop Dis*. 2(1):e130.
Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 45(4):e18.
Eilbeck K, Moore B, Holt C, Yandell M. 2009. Quantitative measures for the management and comparison of annotated genomes. *BMC Bioinformatics* 10:67.
Eilertson KE, Booth JG, Bustamante CD. 2012. SnIPRE: selection inference using a Poisson random effects model. *PLoS Comput Biol*. 8(12):e1002806.
Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol*. 16:157.
English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, et al. 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* 7(11):e47768.
Figueroa-Santiago O, Espino AM. 2014. *Fasciola hepatica* fatty acid binding protein induces the alternative activation of human macrophages. *Infect Immun*. 82(12):5005–5012.
Fontenla S, Rinaldi G, Smircich P, Tort JF. 2017. Conservation and diversification of small RNA pathways within flatworms. *BMC Evol Biol*. 17(1):215.
Gilad Y, Wiebe V, Przeworski M, Lancet D, Pääbo S. 2004. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol*. 2(1):E5.
Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A*. 108(4):1513–1518.
Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59(3):307–321.
Hahnel S, Wheeler N, Lu Z, Wangwiwatsin A, McVeigh P, Maule A, Berriman M, Day T, Ribeiro P, Grevelding CG. 2018. Tissue-specific transcriptome analyses provide new insights into GPCR signalling in adult *Schistosoma mansoni*. *PLoS Pathog*. 14(1):e1006718.
Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol*. 30(8):1987–1997.
He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S. 2016. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res*. 44(W1):W236–241.
Heled J, Drummond AJ. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Syst Biol*. 61(1):138–149.
Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32(5):767–769.
Hola-Jamriska L, Tort JF, Dalton JP, Day SR, Fan J, Aaskov J, Brindley PJ. 1998. Cathepsin C from *Schistosoma japonicum*–cDNA encoding the preproenzyme and its phylogenetic relationships. *Eur J Biochem*. 255(3):527–534.
Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.
Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. 2017. WormBase ParaSite—a comprehensive resource for helminth genomics. *Mol Biochem Parasitol*. 215:2–10.
Hunt VL, Tsai IJ, Coghlan A, Reid AJ, Holroyd N, Foth BJ, Tracey A, Cotton JA, Stanley EJ, Beasley H, et al. 2016. The genomic basis of parasitism in the *Strongyloides* clade of nematodes. *Nat Genet*. 48(3):299–307.
International Helminth Genomes Consortium. 2019. Comparative genomics of the major parasitic worms. *Nat Genet*. 51:163–174.
Irving JA, Spithill TW, Pike RN, Whisstock JC, Smooker PM. 2003. The evolution of enzyme specificity in *Fasciola* spp. *J Mol Evol*. 57(1):1–15.
Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14(6):587–589.
Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol*. 428(4):726–731.
Kang JM, Bahk YY, Cho PY, Hong SJ, Kim TS, Sohn WM, Na BK. 2010. A family of cathepsin F cysteine proteases of *Clonorchis sinensis* is the major secreted proteins that are expressed in the intestine of the parasite. *Mol Biochem Parasitol*. 170(1):7–16.
Katoh K, Standley DM. 2014. MAFFT: iterative refinement and additional methods. *Methods Mol Biol*. 1079:131–146.
Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol*. 6:29.
Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 12(4):357–360.
Kishida T. 2008. Pattern of the divergence of olfactory receptor genes during tetrapod evolution. *PLoS One* 3(6):e2385.
Koskinen P, Toronen P, Nokso-Koivisto J, Holm L. 2015. PANNZER: high-throughput functional annotation of uncharacterized proteins in an error-prone environment. *Bioinformatics* 31(10):1544–1552.
Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res*. 35(9):3100–3108.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Group. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
Liang P, He L, Xu Y, Chen X, Huang Y, Ren M, Liang C, Li X, Xu J, Lu G, et al. 2014. Identification, immunolocalization, and characterization analyses of an exopeptidase of papain superfamily, (cathepsin C) from *Clonorchis sinensis*. *Parasitol Res*. 113(10):3621–3629.
Lotfy WM, Brant SV, DeJong RJ, Le TH, Demiaszkiewicz A, Rajapakse RP, Perera VB, Laursen JR, Loker ES. 2008. Evolutionary origins, diversification, and biogeography of liver flukes (Digenea, Fasciolidae). *Am J Trop Med Hyg*. 79(2):248–255.
Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 25(5):955–964.
Loytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A*. 102(30):10557–10562.
Martin I, Caban-Hernandez K, Figueroa-Santiago O, Espino AM. 2015. *Fasciola hepatica* fatty acid binding protein inhibits TLR4 activation and suppresses the inflammatory cytokines induced by lipopolysaccharide in vitro and in vivo. *J Immunol*. 194(8):3924–3936.
Mas-Coma S, Valero MA, Bargues MD. 2009. Chapter 2. *Fasciola*, lymnaeids and human fascioliasis, with a global overview on disease transmission, epidemiology, evolutionary genetics, molecular epidemiology and control. *Adv Parasitol*. 69:41–146.
McGonigle L, Mousley A, Marks NJ, Brennan GP, Dalton JP, Spithill TW, Day TA, Maule AG. 2008. The silencing of cysteine proteases in *Fasciola hepatica* newly excysted juveniles using RNA interference reduces gut penetration. *Int J Parasitol*. 38(2):149–155.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20(9):1297–1303.
McNulty SN, Tort JF, Rinaldi G, Fischer K, Rosa BA, Smircich P, Fontenla S, Choi YJ, Tyagi R, Hallsworth-Pepin K, et al. 2017. Genomes of *Fasciola hepatica* from the Americas reveal colonization with *Neorickettsia* endobacteria related to the agents of Potomac horse and human sennetsu fevers. *PLoS Genet*. 13(1):e1006537.
McVeigh P, McCammick E, McCusker P, Wells D, Hodgkinson J, Paterson S, Mousley A, Marks NJ, Maule AG. 2018. Profiling G protein-coupled receptors of *Fasciola hepatica* identifies orphan rhodopsins unique to phylum *Platyhelminthes*. *Int J Parasitol Drugs Drug Resist*. 8(1):87–103.
Meadows JR, Hiendleder S, Kijas JW. 2011. Haplogroup relationships between domestic and wild sheep resolved using a mitogenome panel. *Heredity (Edinb)*. 106(4):700–706.
Meemon K, Sobhon P. 2015. Juvenile-specific cathepsin proteases in *Fasciola* spp.: their characteristics and vaccine efficacies. *Parasitol Res*. 114(8):2807–2813.
Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simao TL, Stadler T, et al. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* 334(6055):521–524.
Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.
Ogilvie HA, Bouckaert RR, Drummond AJ. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Mol Biol Evol*. 34(8):2101–2114.
Parfrey LW, Lahr DJ, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci U S A*. 108(33):13624–13629.
Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 33(3):290–295.
Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol*. 31(7):1929–1936.
Phalee A, Wongsawad C, Rojanapaibul A, Chai JY. 2015. Experimental life history and biological characteristics of *Fasciola gigantica* (Digenea: Fasciolidae). *Korean J Parasitol*. 53(1):59–64.
Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst Biol*. 67(5):901–904.
Rey O, Danchin E, Mirouze M, Loot C, Blanchet S. 2016. Adaptation to global change: a transposable element-epigenetics perspective. *Trends Ecol Evol (Amst)*. 31(7):514–526.
Robinson MW, Corvo I, Jones PM, George AM, Padula MP, To J, Cancela M, Rinaldi G, Tort JF, Roche L, et al. 2011. Collagenolytic activities of the major secreted cathepsin L peptidases involved in the virulence of the helminth pathogen, *Fasciola hepatica*. *PLoS Negl Trop Dis*. 5(4):e1012.
Robinson MW, Menon R, Donnelly SM, Dalton JP, Ranganathan S. 2009. An integrated transcriptomics and proteomics analysis of the secretome of the helminth pathogen *Fasciola hepatica*: proteins associated with invasion and infection of the mammalian host. *Mol Cell Proteomics*. 8(8):1891–1907.
Robinson MW, Tort JF, Lowther J, Donnelly SM, Wong E, Xu W, Stack CM, Padula M, Herbert B, Dalton JP. 2008. Proteomics and phylogenetic analysis of the cathepsin L protease family of the helminth pathogen *Fasciola hepatica*: expansion of a repertoire of virulence-associated factors. *Mol Cell Proteomics*. 7(6):1111–1123.
Rofatto HK, Parker-Manuel SJ, Barbosa TC, Tararam CA, Alan Wilson R, Leite LC, Farias LP. 2012. Tissue expression patterns of *Schistosoma mansoni* Venom Allergen-Like proteins 6 and 7. *Int J Parasitol*. 42(7):613–620.
Rowe T, Rich TH, Vickers-Rich P, Springer M, Woodburne MO. 2008. The oldest platypus and its bearing on divergence timing of the platypus and echidna clades. *Proc Natl Acad Sci U S A*. 105(4):1238–1242.
Saijuntha W, Tantrawatpan C, Agatsuma T, Wang C, Intapan PM, Maleewong W, Petney TN. 2018. Revealing genetic hybridization and DNA recombination of *Fasciola hepatica* and *Fasciola gigantica* in nuclear introns of the hybrid Fasciola flukes. *Mol Biochem Parasitol*. 223:31–36.
Salazar-Calderon M, Martin-Alonso JM, Castro AM, Parra F. 2003. Cloning, heterologous expression in *Escherichia coli* and

characterization of a protein disulfide isomerase from *Fasciola hepatica*. *Mol Biochem Parasitol*. 126(1):15–23.
Schrader L, Schmitz J. 2018. The impact of transposable elements in adaptive evolution. *Mol Ecol*. 28(6):1537–1549.
Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res*. 43(W1):W7–W14.
Skinner DE, Rinaldi G, Koziol U, Brehm K, Brindley PJ. 2014. How might flukes and tapeworms maintain genome integrity without a canonical piRNA pathway? *Trends Parasitol*. 30(3):123–129.
Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet*. 8(4):272–285.
Sonnhammer EL, Ostlund G. 2015. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res*. 43(Database issue):D234–239.
Sripa J, Laha T, To J, Brindley PJ, Sripa B, Kaewkes S, Dalton JP, Robinson MW. 2010. Secreted cysteine proteases of the carcinogenic liver fluke, *Opisthorchis viverrini*: regulation of cathepsin F activation by autocatalysis and trans-processing by cathepsin B. *Cell Microbiol*. 12(6):781–795.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
Stapley J, Santure AW, Dennis SR. 2015. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. *Mol Ecol*. 24(9):2241–2252.
Tang YT, Gao X, Rosa BA, Abubucker S, Hallsworth-Pepin K, Martin J, Tyagi R, Heizer E, Zhang X, Bhonagiri-Palsikar V, et al. 2014. Genome of the human hookworm *Necator americanus*. *Nat Genet*. 46(3):261–269.
Teofanova D, Kantzoura V, Walker S, Radoslavov G, Hristov P, Theodoropoulos G, Bankov I, Trudgett A. 2011. Genetic diversity of liver flukes (*Fasciola hepatica*) from Eastern Europe. *Infect Genet Evol*. 11(1):109–115.
Torgerson PR, Devleesschauwer B, Praet N, Speybroeck N, Willingham AL, Kasuga F, Rokni MB, Zhou XN, Fevre EM, Sripa B, et al. 2015. World Health Organization estimates of the global and regional disease burden of 11 foodborne parasitic diseases, 2010: a data synthesis. *PLoS Med*. 12(12):e1001920.
UniProt Consortium. 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 45:D158–D169.
Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 43:11.10.1-33.
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
Walker SM, Johnston C, Hoey EM, Fairweather I, Borgsteede F, Gaasenbeek C, Prodohl PA, Trudgett A. 2011. Population dynamics of the liver fluke, *Fasciola hepatica*: the effect of time and spatial separation on the genetic diversity of fluke populations in the Netherlands. *Parasitology* 138(2):215–223.
Walker SM, Prodohl PA, Hoey EM, Fairweather I, Hanna RE, Brennan G, Trudgett A. 2012. Substantial genetic divergence between morphologically indistinguishable populations of *Fasciola* suggests the possibility of cryptic speciation. *Int J Parasitol*. 42(13–14): 1193–1199.
Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*.
Yang J, Yang L, Lv Z, Wang J, Zhang Q, Zheng H, Wu Z. 2012. Molecular cloning and characterization of a HSP70 gene from *Schistosoma japonicum*. *Parasitol Res*. 110(5):1785–1793.

Radio et al. Parasites & Vectors (2018) 11:56
DOI 10.1186/s13071-017-2553-2

Parasites & Vectors

RESEARCH Open Access

Pleiotropic alterations in gene expression in Latin American *Fasciola hepatica* isolates with different susceptibility to drugs

Santiago Radio[1,2], Santiago Fontenla[1], Victoria Solana[3], Anna C. Matos Salim[4], Flávio Marcos Gomes Araújo[4], Pedro Ortiz[5], Cristian Hoban[5], Estefan Miranda[6], Valeria Gayo[7], Fabiano Sviatopolk-Mirsky Pais[4], Hugo Solana[3], Guilherme Oliveira[4,8], Pablo Smircich[1,2,9*] and José F. Tort[1*]

Abstract

Background: *Fasciola hepatica* is the main agent of fasciolosis, a zoonotic disease affecting livestock worldwide, and an emerging food-borne disease in humans. Even when effective treatments are available, drugs are costly and can result in tolerance, liver damage and normally they do not prevent reinfection. Drug-resistant strains in livestock have been reported in various countries and, more worryingly, drug resistance in human cases has emerged in South America. The present study aims to characterize the transcriptome of two South American resistant isolates, the Cajamarca isolate from Peru, resistant to both triclabendazole and albendazole (TCBZR/ABZR) and the Rubino isolate from Uruguay, resistant to ABZ (TCBZS/ABZR), and compare them to a sensitive strain (Cenapa, Mexico, TCBZS/ABZS) to reveal putative molecular mechanisms leading to drug resistance.

Results: We observed a major reduction in transcription in the Cajamarca TCBZR/ABZR isolate in comparison to the other isolates. While most of the differentially expressed genes are still unannotated, several trends could be detected. Specific reduction in the expression levels of cytoskeleton proteins was consistent with a role of tubulins as putative targets of triclabendazole (TCBZ). A marked reduction of adenylate cyclase might be underlying pleiotropic effects on diverse metabolic pathways of the parasite. Upregulation of GST mu isoforms suggests this detoxifying mechanism as one of the strategies associated with resistance.

Conclusions: Our results stress the value of transcriptomic approaches as a means of providing novel insights to advance the understanding of drug mode of action and drug resistance. The results provide evidence for pleiotropic variations in drug-resistant isolates consistent with early observations of TCBZ and ABZ effects and recent proteomic findings.

Keywords: *Fascola hepatica*, Drug resistance, American isolates, Triclabendazole, Albendazole, Transcriptomics

Background

Fasciolosis is indisputably one of the most widely distributed zoonotic diseases, affecting no less than 300 million cattle and 250 million sheep worldwide. The economical cost of the disease has been valued at 3 billion dollars annually [1, 2]. This huge economic impact from direct losses might be an underestimate considering indirect costs of treatment, or loss of animal workforce in less industrialized countries. Furthermore, fasciolosis is emerging as a relevant issue in human health, affecting roughly 2.6 million people worldwide. For this reason it has been considered as a re-emerging neglected disease by the WHO [3]. In the Americas the disease is widespread in livestock and is an important human food-borne infection in the Altiplano region of Bolivia and Peru [4].

Although effective treatments are available, drugs are costly and usually do not block reinfection. Liver fluke drug resistance is a preoccupying productive problem in Latin America and a concern in medicine since human cases have emerged [4]. Triclabendazole (TCBZ) treatment failure has been reported in Brazil [5], Argentina [6] and Peru [7–9]. Moreover, resistance to albendazole

* Correspondence: psmircich@fmed.edu.uy; jtort@fmed.edu.uy
[1]Departamento de Genética, Facultad de Medicina, Universidad de la Republica, UDELAR, Montevideo, Uruguay
Full list of author information is available at the end of the article

(ABZ) has been reported in Argentina [10], Uruguay [11], Chile and Bolivia [5]. More worryingly, ABZ resistance accompanied by reduced effectivity of TCBZ has been registered in Bolivia [12] and Peru [13], and resistance to both drugs has been described in an isolate from the Cajamarca valley of Peru [9, 11]. This phenomenon of double resistance compromises the idea of combined drug treatments. More recently, TCBZ resistance in humans has been reported in Chile and Peru [14, 15]. There is a pressing need to understand epidemiological and mechanistic aspects of drug resistance emergence.

Despite being in use for more than 30 years, the exact mechanism of action of TCBZ is still not completely elucidated [16]. One of the earliest and more complete biochemical studies provides early evidence of pleiotropic effects [17]. The study documented that the drug was absorbed through the tegument affecting the motility of the parasite in a dose-dependent manner. This effect was associated with changes in the worm's resting tegument membrane potentials. In addition, an effect on tubulin binding was also observed. Furthermore, an anaerobic metabolic stimulation was suggested by the increase in propionate and acetate production without affecting ATP levels. A general reduction of secreted proteases was also described [17], an effect that might be associated with a reported general inhibition of protein synthesis [18]. Several studies have focused on tegument disruption, one of the major effects of drug treatment [19–24]. This was generally associated with a putative role of tubulin as target, based on similarities to findings on other benzimidazole effects against nematodes [16, 25]. The ability to compete with colchicine, a known inhibitor of tubulin polymerization, was an early test for tubulin binding [17, 26]. The role of microtubules in TCBZ pathogenesis was further suggested by several lines of evidence, like the detection of cell division inhibition in vitelline and reproductive cells [27–29], the reduction in the transport of tegumental secretory bodies [22], and the inhibition of tubulin immunostaining [30, 31]. Surprisingly, TCBZ was recently reported to inhibit adenylate cyclase activity in yeast, activating the stress response [32]. This effect has not been observed in the liver fluke so far, but considering the role of cAMP as second messenger, it might provide an explanation for the pleiotropic effects of the drug.

Interestingly, TCBZ is quite specific for fasciolids, being ineffective against nematodes. On the other hand, out of the benzimidazolic drugs used for gastrointestinal roundworms, only ABZ is effective against the adult stage of *Fasciola* spp. [33]. These differences and the fact that ABZ is usually effective against TCBZ-resistant isolates suggest that different mechanisms might be underlying the effect of each benzimidazolic drug. ABZ also induces tegument damage, disruption of tegumental vesicle traffic, alterations in reproductive tissues and vitelline cells, and reduction in egg production [33, 34]. However, despite these similarities, differences in the metabolism of worms treated with these drugs are suggestive of diverse targets or mechanisms.

It has been shown that the metabolism of TCBZ to triclabendazole sulphoxide (TCBZ.SO) and TCBZ.SO to triclabendazole sulphone (TCBZ.SO2) is greater in TCBZ-R than in TCBZ-S isolates [35–37]. Interestingly, the uptake of TCBZ and TCBZ.SO by TCBZ-R fluke isolates is significantly lower than in TCBZ-S flukes, while the uptake of ABZ is similar in both strains [35, 38]. The effect can be reversed by incubating the TCBZ-R flukes in the presence of ivermectin, a substrate of P-glycoprotein (PGP) drug efflux pump. While disruption of the tegument is markedly reduced in TCBZ-R flukes, the co-incubation with R(+)-verapamil, another PGP inhibitor, gives rise to severe tegumental lesions [39, 40], highlighting PGP as one of the possible detoxifying mechanisms. A similar effect of increased tegument disruption is seen when TCBZ-R flukes are incubated with methimazole, an inhibitor of the flavin mono-oxigenases (FMO) [41]. Ketoconazole, an inhibitor of CYP450 [42, 43] also produce as similar phenotype, suggesting that these pathways might be associated with drug resistance as well. It was hypothesized that some of these enzymes might be upregulated in the resistant strains [35, 36]. Consistent with this, an increased enzyme activity of detoxifying enzymes gluthatione S-transferase (GST), carboxyl esterase and carbonyl reductase, was observed in TCBZ-treated worms [44], and a higher GST response was observed in the TCBZ-R Sligo strain in comparison to the Cullompton sensitive isolate [45]. A comparative proteomic study of the Sligo and Cullompton isolates showed variation in energy metabolic enzymes, detoxifying enzymes and structural proteins, confirming the pleiotropic nature of TCBZ effects, and reinforced suggestions of differential expression of several proteins [46].

It is therefore clear that anthelmintic drugs produce complex effects on the liver fluke, affecting the metabolism, physiology and morphology of the parasite. Drug-resistant isolates seem to utilize diverse detoxifying mechanisms to cope with these drugs [47]. We sought to shed new light into drug resistance by using a genomics approach, which has been demonstrated to be very powerful in schistosomes [48–52]. Similar efforts have been initiated in *F. hepatica* [53], starting from well characterized European TCBZ-R and TCBZ-S isolates and their genome sequences [54, 55]. In this work we add a transcriptomic perspective to the current knowledge via an in depth analysis of the basal transcriptomic state of three isolates from the Americas with different susceptibilities to TCBZ and ABZ. This report would set the baseline for upcoming comparative studies on variations of gene expression upon exposure to the different anthelminthic drugs.

Methods

Strains

Three *Fasciola hepatica* isolates with different susceptibilities to triclabendazole (TCBZ) and albendazole (ABZ) were analyzed. The "Cajamarca" isolate was originally obtained by Dr Pedro Ortiz from infected cattle in Cajamarca, Peru. It has been maintained for five years in sheep and characterized in their laboratory as resistant to both ABZ and TCBZ [9, 11]. The "Rubino" isolate (originally obtained from cattle in Salto, Uruguay) is resistant to ABZ but sensitive to TCBZ [11]. It has been maintained in sheep for eight years by Dr Valeria Gayo in the DILAVE "Miguel C. Rubino". The isolate is routinely used to test formulations of Closantel and TCBZ, since it is sensitive to both. Similarly, "Cenapa" is an isolate sensitive to both drugs that has been maintained for more than a decade in sheep, and is routinely used by the veterinary health authorities of the Mexican government to evaluate the efficacy of anthelminthics. The Cenapa isolate was kindly provided by Dr Estefan Miranda. The three laboratories followed protocols approved by the respective local Committees of Animal Experimentation, in accordance to the recommendations of Guide for the Care and Use of Laboratory Animals [56]. All isolates are maintained in sheep without selective drug pressure.

RNA-sequencing and pre-processing of the reads

Adult flukes were obtained from infected sheep livers and stored immediately in RNAlater. No macroscopic morphological differences were observed between flukes. PolyA+ RNA was purified from single adult worms of the different isolates in duplicates and used to generate paired end (PE) libraries using the TrueSeq LT kit (Illumina, San Diego, USA). Samples were sequenced in the Illumina Platform at the CPqRR sequencing facilities at FIOCRUZ, Belo Horizonte, to obtain 117 million 76 bp pair end reads. The resulting sequences were quality trimmed and mapped to the *F. hepatica* reference genome (WormBase Parasite Acc: PRJEB6687) [54] using CLC Genomics Workbench v7 (Qiagen, Aarhus C, Denmark) with default parameters. A good coverage of predicted genes (over 80%) was observed. A summary of the number of reads obtained in each step is shown in Additional file 1: Table S1. Raw sequencing data were submitted to SRA under accession PRJNA339158.

Differential expression (DE) determination

Differential expression was analyzed using different tools of the Bioconductor suite of bioinformatics packages [57, 58]. To obtain expression estimates, mapped reads were counted for each gene using the summarizeOverlaps function from the GenomicAlingments package [59] and $\log_2$-transformed. To account for sequencing depth in differential expression analysis, raw read counts were normalized with DESeq2 [60]. Replicate consistency was established by computing pairwise Pearson's correlation coefficients in R. Differentially expressed genes were defined using DESeq2 (using the Wald Test implemented in the package with 4 degrees of freedom) from pairwise comparisons of the $\log_2$-transformed normalized expression estimates, establishing a minimum fold change of 2 and a false discovery rate (FDR) (controlled using the Benjamini & Hochberg's method [61]) corrected *P*-value lower than 0.05.

Functional enrichment analysis

Lists of DE genes were analyzed with the TopGO Biocondutor package in R [62] to assess enriched gene ontology categories (using Fisher's exact test implemented in the package with 2 degrees of freedom; a P-value < 0.01 was considered significant). The parameters orderBy = "classicFisher" and ranksOf = "classicFisher" were set for visualization. GO annotation was retrieved from Wormbase Parasite (parasite.wormbase.org) [63].

SNP calling

SNPs were called from RNAseq mappings to the reference genome using mpileup (-uf parameters) and bcftools (-mv parameters) from samtools [64]. Putative phenotypic effects of the variants called were assessed with the variant effect predictor script (part of the Ensembl tools [65]. Synonymous codon variants and other low impact mutations were not included in the analysis.

KEGG orthology annotation

The GhostKOALA tool [66] was used to annotate the predicted proteome. Genes belonging to common housekeeping functions were identified using the KEGG Brite reconstruction list [67]. Selected categories, gene IDs and genome annotation are shown in Additional files as indicated in the text.

Results and discussion

A lower transcription level is observed for a significant amount of Cajamarca genes

To obtain a global picture of the differences of the transcriptomic profiles of the different isolates, samples were clustered and pairwise correlations were computed. Figure 1a shows that while duplicates were very consistent, the Cajamarca isolate displayed the lowest level of correlation to the other samples. This difference is mostly explained by comparatively low mRNA steady state levels of many transcripts in this strain (Fig. 1b and Additional file 2: Figure S1a). On the other hand, differences in the gene expression were subtler between the Rubino and Cenapa samples (Additional file 2: Figure S1b).

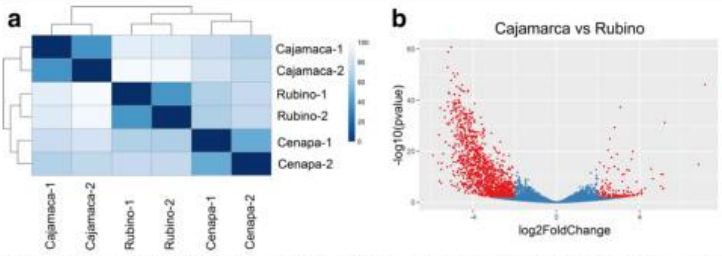


Fig. 1 Overview of sample distances and differential expression. **a** Heatmap of the Euclidean sample to sample distances shown in a white to blue scale (blue represents higher correlation). Both replicates for each sample are shown, with the corresponding clustering dendrograms on the sides. **b** Volcano plot showing the differential expression between the Cajamarca and Rubino transcripts. Red dots represent differentially expressed genes (log2 fold change > 2, *P*-value < 0.01)

Housekeeping gene expression is not biased between isolates

Since we saw a reduction of transcription in the TCBZR/ABZR Cajamarca isolate in relation to the other two isolates, we wanted to check if the asymmetrical distribution of differentially expressed genes (DEG) among the strains was due to a general lower transcript level in the Cajamarca isolate that normalization was not able to account for. Interestingly, we detected that housekeeping genes (such as aminoacyl tRNA synthetases, DNA polymerase subunits, proteasome subunits and spliceosome subunits) do not show differences of expression among any of the strains (Fig. 2 and Additional file 3: Figure S2). These results indicate that Cajamarca samples are not globally skewed toward less expression, but rather different gene families might be differentially affected. These results are puzzling since no drug selective pressure was applied to these worms, and it might reflect a basal status of the isolate.

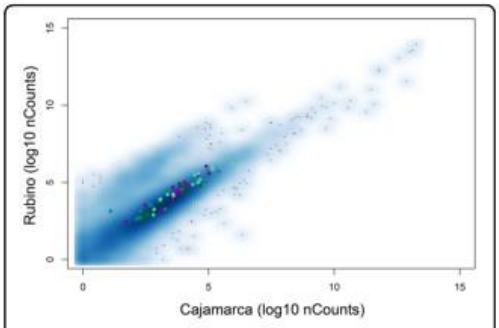


Fig. 2 Correlation of housekeeping gene expression. Scatterplot showing the correlation of Rubino normalized counts per gene *versus* their Cajamarca counterparts. Filled circles highlight the expression of genes belonging to housekeeping functions: dark cyan, aminoacyl tRNA synthetases; dark green, core DNA polymerase subunits; dark magenta, core proteasome subunits; light green, core spliceosome subunits

Notably, early studies of the effect of TCBZ reported a marked drop in protein synthesis [18], consistent with morphological observations of changes in heterochromatin, the disappearance of the nucleolus and the subsequent reduction of ribosomes, reduction in Golgi complexes, and secretory bodies in tegumental cells [22]. Therefore, it is tempting to consider that the study of differentially expressed (DE) genes might help to pinpoint some possible candidates involved in the resistance phenotype.

The top differentially expressed genes are mostly not annotated

Pairwise comparisons were performed between the isolates using DESeq2 package. Almost half of the downregulated or upregulated genes in each pairwise list are currently devoid of any annotation. We focused on the top 20 upregulated or downregulated DEG in each pairwise comparison (Additional file 4: Table S2). Even within this selected set, 41 out of 58 of the downregulated DEG are of unknown function, and diverse functions are present in the remaining annotated genes. A similar scenario was observed for the upregulated DEG where 37 of 47 were unannotated, highlighting the still incomplete nature of the available genome annotation. However, 3 out of 10 of the annotated DEGs upregulated in the Cajamarca and Rubino resistant strains, contain putative CUB domains (IPR000859). This domain (for Complement C1r/C1s, Uegf and Bmp1) comprises more than 100 amino acids and is usually found in the extracellular- or plasma membrane-associated proteins with diverse functions. Several mammalian CUB containing proteins are proteases with calcium binding EGF domains, taking part in pleiotropic functions like complement activation, developmental patterning, neurotransmission and cell signaling [68]. CUB domain containing proteins found in the *F. hepatica* genome are generally short with no other associated domains, but this might well be a consequence of the still fragmented nature of the assembly. Interestingly, upregulation of CUB domain

containing proteins has been observed in *C. elegans* in response to albendazole administration [69]. Although little is known of the relevance or function of these proteins in helminths, the upregulation in the resistant isolates even without exposure to the drug is noteworthy.

Cytoskeleton related genes are less expressed in the Cajamarca strain

The entire lists of DEG in each pairwise comparison were subjected to GO functional category enrichment analysis. Remarkably, several terms related to cytoskeleton structure and function showed a significant enrichment in the list of downregulated genes in the resistant strains, especially in the Cajamarca isolate (Fig. 3 and Additional file 5: Table S3). This is an interesting observation taking into account the putative role of tubulins as targets of ABZ, TCBZ and other benzimidazole-based drugs [70, 71].

These results prompted us to further characterize the expression of cytoskeleton-related gene families. Figure 4 shows the expression profile of the α and β tubulin gene families (Fig. 4a, b) and motor protein families (kinesins and dyneins Fig. 4c, d). In all cases the Cajamarca isolate showed lower levels of gene expression for these families when compared to the other strains, which were not significantly different among them (Student's *t*-test $P < 0.05$, see legend of Fig. 4 for exact *P*-values of each comparison). In particular, α-tubulin and β-tubulin mRNAs showed differential expression between the strains being down-represented in the Cajamarca isolate (Fig. 4a, b and Additional file 6: Table S4). Finding a skewed expression

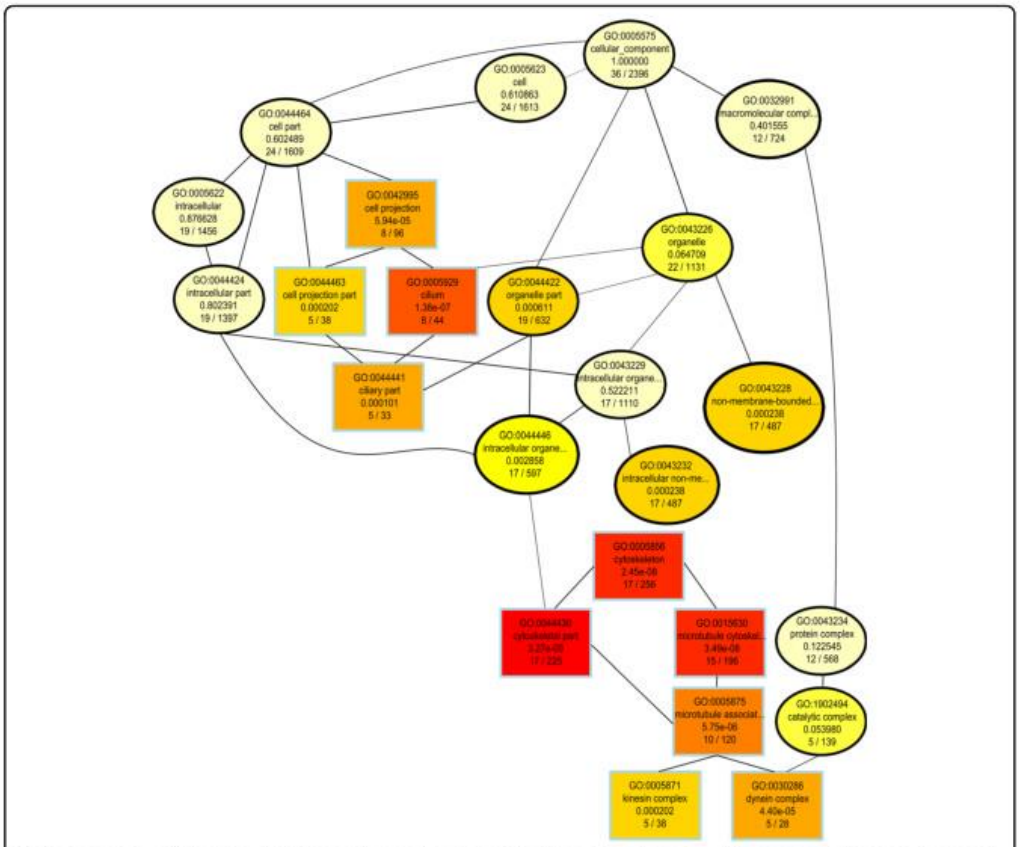


Fig. 3 Gene ontology (GO) enrichment of Cajamarca downregulated genes. Directed acyclic graph showing cellular component GO categories present in the downregulated genes when comparing the Cajamarca with the Cenapa samples. Colors indicate category overrepresentation, with red being most significant. Text lines indicate GO number, category description, Fisher's *P*-value and the number of genes in the set over the total number of genes in the genome for the GO category

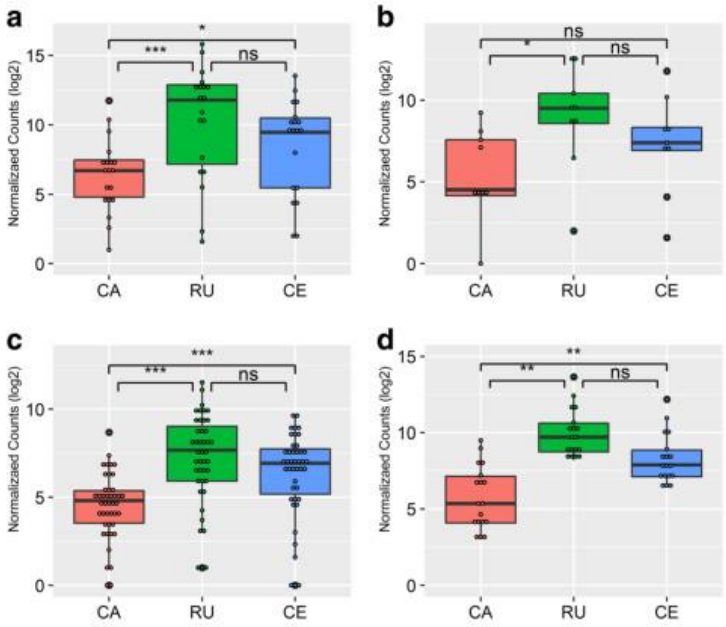


Fig. 4 Expression levels of cytoskeleton related gene families. Boxplot showing the expression levels for mRNAs coding for α-tubulins (**a**), β-tubulins (**b**), kinesins (**c**) and dyneins (**d**) in each sample. Samples were compared using a two-sided Student's *t*-test in R. For α-tubulins CA *vs* RU: $t_{(30)} = -3.586$, $P < 0.001$; CA *vs* CE: $t_{(33)} = -2.148$, $P = 0.039$; RU *vs* CE: $t_{(35)} = 1.515$, $P = 0.139$. For β-tubulins CA *vs* RU: $t_{(15)} = -2.529$, $P = 0.023$; CA *vs* CE: $t_{(16)} = -1.372$, $P = 0.189$; RU *vs* CE: $t_{(16)} = 1.162$, $P = 0.262$. For kinesins CA *vs* RU: $t_{(64)} = -5.092$, $P < 0.001$; CA *vs* CE: $t_{(67)} = -3.522$, $P < 0.001$; RU *vs* CE: $t_{(73)} = 1.608$, $P = 0.112$. For dyneins CA *vs* RU: $t_{(347)} = -2.473$, $P = 0.014$; CA *vs* CE: $t_{(343)} = -0.207$, $P = 0.836$; RU *vs* CE: $t_{(337)} = 2.407$, $P = 0.017$. Significant comparisons are marked with asterisks (*$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$). *Abbreviations*: CA, Cajamarca; RU, Rubino; CE, Cenapa; ns, not significant

of tubulins and other cytoskeleton genes in a double resistant isolate reinforces the notion of tubulin as a putative TCBZ target.

SNPs in the β-tubulin genes have been reported as molecular mechanisms of resistance development in nematodes, and in particular three amino acid substitutions in the *Haemonchus contortus* β-tubulin have been associated with resistance [72, 73]. Initial studies attempting to find the same variations in *F. hepatica* failed to find association between these changes and resistant status [30, 74]. Despite this, a strongly reduced tegument damage and little disruption of tubulin immunostaining were observed in TCBZ resistant flukes in comparison to a sensitive isolate [30]. Similarly, disruption of tegument and reduction of tubulin immunostaining were also observed upon exposure to albendazole sulphoxide (ABZ-SO) [34] highlighting tubulin as one of the putative targets in *F. hepatica*. Our results support the association of these proteins with the resistance phenomenon, but in our case, mRNA level variations and not SNPs were revealed as the putative molecular mechanism. Notably, a similar reduction in tubulin expression has been observed in *H. contortus* drug-resistant strains [75, 76]. However, a previous study failed to detect differences in transcription levels of diverse β-tubulins between the TCBZ resistant Oberon and the Leon TCBZ sensitive isolates [77]. While these contradictory observations might be pointing to different mechanisms of resistance in different strains, the observation that other motor proteins mRNA levels are reduced (Fig. 4c, d) clearly relates to the previous finding and further associates the resistance phenotype with microtubule cytoskeleton function.

Interestingly, collagen coding mRNAs were under-expressed in the Rubino isolate compared to the other strains. The difference is particularly significant with the Cajamarca isolate where an average 5-fold transcription level was detected for collagen (Additional file 7: Table S5).

Only a few genes associate with detoxifying pathways are differentially expressed

Since several putative candidates genes involved in drug resistance have been advanced [47, 78], we verified their differential expression in our three isolates. One of the

first mechanisms proposed in detoxification is the oxidation of the incoming drug by proteins involved in redox activity (defined by annotation of functional domains) (Additional file 8: Table S6). We detected downregulation of several redox proteins in the Cajamarca strain when compared with the other two isolates, and we observed extreme variations.

Another proposed mechanism for drug resistance is the conjugation of the primary metabolite to proteins, such as the glutathione S-transferase family (GST), and other detoxification enzymes. The expression pattern of these candidates was analyzed, and no significant differences were observed for the complete set when a FC > 2 cut-off was applied. However, while there is no global trend when all enzymes are considered, it is interesting to point out that some GST genes, particularly of the mu type, do have statistically significant variation and a modest upregulation in the resistant strains (FC > 2, see Additional file 9: Table S7). This is consistent with previous biochemical studies that have shown increased GST activity in the Sligo resistant strain when compared to the Cullampton sensitive strain [44, 45]. Also an increase in GST mu was observed in the only comparative proteomic study available so far [46].

Among the detoxification proteins, some mRNAs were significantly different between the strains (Additional file 9: Table S7). It has been hypothesized that ABC transporters might be upregulated in drug-resistant strains [79]. Indeed, an ABC transporter-like protein (BN1106_s3396B000087) was upregulated in the TCBZ resistant isolate (Additional file 9: Table S7). Whether this is further upregulated upon drug exposure still needs to be addressed.

Adenylate cyclase is reduced in TCBZR isolate

Recently it was reported that TCBZ can inhibit adenylate cyclase (AC) activity in yeast [32], but despite being widely studied in *F. hepatica* [80], there are still no in vitro studies of AC variation in response to drug administration. To gain insight into this possible mechanism, we investigated if AC expression was skewed in our samples. Surprisingly we observed a significant variation on the expression levels of several AC isoforms in both the resistant isolates in relation to the Cenapa sensitive strain. Notably, all isoforms tend to have consistently low transcription levels in the Cajamarca isolate with some of them being significant, while variations in both directions were found in the ABZR Rubino isolate (Table 1).

The described inhibition of AC upon exposure to TCBZ in yeast would not necessarily reduce their transcription levels. Considering the central role of AC in metabolism, its downregulation in the Cajamarca isolate

Table 1 Pairwise comparisons for all annotated adenylate cyclase genes

Gene	Cajamarca *vs* Cenapa		Rubino *vs* Cenapa		Cajamarca *vs* Rubino		Genome annotation	Kegg annotation
	Fold change (log2)	FDR	Fold change (log2)	FDR	Fold change (log2)	FDR		
BN1106_s820B000180	**-2.54**	1.00e-06	0.78	1.49e-01	**-3.32**	1.42e-11	Adenylyl cyclase class-3/4/ guanylyl cyclase (Nucleotide cyclase)	K08049 ADCY9
BN1106_s5842B000024	**-2.42**	4.10e-03	*1.78*	3.80e-03	**-4.20**	7.78e-09	Adenylyl cyclase class-3/4/ guanylyl cyclase (Nucleotide cyclase)	K08049 ADCY9
BN1106_s1582B000143	**-1.81**	1.26e-04	*1.24*	5.50e-03	**-3.05**	9.42e-13	Adenylyl cyclase class-3/4/ guanylyl cyclase (Nucleotide cyclase)	–
BN1106_s3088B000130	**-1.74**	3.43e-02	*1.36*	5.27e-02	**-3.10**	6.20e-06	Adenylyl cyclase class-3/4/ guanylyl cyclase (Nucleotide cyclase)	–
BN1106_s2758B000091	**-1.61**	1.20e-02	*1.67*	2.09e-03	**-3.28**	8.90e-10	Adenylyl cyclase class-3/4/ guanylyl cyclase (Nucleotide cyclase)	K08041 ADCY1
BN1106_s795B000308	-1.05	7.29e-02	**-1.91**	2.65e-04	0.85	2.40e-01	Adenylyl cyclase class-3/4/ guanylyl cyclase (Nucleotide cyclase)	K08049 ADCY9
BN1106_s451B000364	-0.58	6.54e-01	**-1.98**	2.50e-02	1.40	2.07e-01	Adenylyl cyclase class-3/4/ guanylyl cyclase (Nucleotide cyclase)	K08049 ADCY9
BN1106_s4307B000027	-0.48	na	-0.61	na	0.00	na	na	K08049 ADCY9
BN1106_s1588B000215	-0.12	na	-1.02	na	0.91	6.46e-01	Adenylyl cyclase class-3/4/ guanylyl cyclase (Nucleotide cyclase)	K08049 ADCY9
BN1106_s1588B000216	0.13	9.35e-01	-0.78	4.20e-01	0.91	4.60e-01	Adenylyl cyclase class-3/4/ guanylyl cyclase (Nucleotide cyclase)	K08049 ADCY9

IDs of DEG genes [FC > 2 or < -2 (log2 FC >1 or < -1) and an FDR < 0.05] in any comparison are highlighted in bold. The log2FC value is highlighted in italics and underlined for the upregulated genes and in bold and double underlined for the downregulated ones

may account for the pleiotropic reduction in gene expression observed. Since cyclic AMP is a relevant second messenger, we investigated the expression levels of the main mediators Protein Kinase A (PKA) and the RAP guanine nucleotide exchange factors. Both genes showed no alterations in their transcription profile. Nevertheless, the reduction in AC might result in a concomitant AMPc drop, which in turn would activate diverse stress responses through all these effectors without altering their transcription levels. This might provide an explanation for the pleiotropic effects on motility, cytoskeleton, carbohydrate metabolism, and activation of stress and detoxifying enzymes. In any case, carefully controlled in vitro experiments of TCBZ inhibition of AC activity in flukes are needed to confirm these hypotheses.

Interestingly, our results are generally consistent with the comparative proteomics results that found variations in several metabolic enzymes, as well as in stress response proteins and structural proteins [46]. Several of the proteins and genes found to be differentially expressed at transcript level in this study (Additional file 10: Table S8) were also reported as differential in the previous proteomics work, such as the detoxifying enzymes already mentioned (redox proteins and GSTs).

Conclusions

In the first transcriptomic analysis of *F. hepatica* isolates with different levels of drug susceptibility, we were able to highlight diverse protein functions and families that show differential gene expression. Notably, several of the affected genes and pathways correspond to those that are being proposed as normally altered upon drug exposure. The presence of variation in expression levels in these pathways in resistant isolates is suggestive, but we cannot conclude with the available evidence that it is related to the resistant phenotype. Further experiments assessing expression levels of these mRNAs in the different isolates upon controlled exposure are necessary to either confirm or reject that RNA levels actually vary upon drug exposure. While those experiments are on the way, we can highlight that the differentially expressed genes in resistant isolates are diverse, and correlate quite well with initial biochemical and structural observations of the pleiotropic nature of drug effects, particularly in the case of TCBZ [17–24]. Moreover they also are coincident with more recent proteomic characterization of drug sensitive and resistant isolates from European origins [46]. Interestingly a recent transcriptomic study of drug-resistant isolates of *Trypanosoma cruzi* also showed altered expression of genes associated with putative drug action mechanisms [81]. The fact that the parasites can survive drug exposure strongly suggests that resistance is the result of additive subtle changes in the expression, and consequently protein metabolic activity. A corollary of this observation is that resistance in different isolates might rely on diverse mechanisms or targets. This highlights the need for studying diverse isolates in order to gain a better understanding of drug action and parasite resistance mechanisms. The results provided by this work are a step in this direction that we hope will impact future methods for parasite control.

Additional files

Additional file 1: Table S1. General data processing overview. (XLSX 10 kb)

Additional file 2: Figure S1. Differential expression between isolate pairs. Volcano plot showing the differential expression of transcripts between Cajamarca and Cenapa (**a**) and between Rubino and Cenapa (**b**). Red dots represent differentially expressed genes (log2 fold change > 2, *P*-value < 0.01). (TIFF 1086 kb)

Additional file 3: Figure S2. a Correlation of normalized counts and housekeeping genes. Scatterplot showing the correlation of Cenapa normalized counts per gene *versus* their Cajamarca counterparts. **b** Scatterplot showing the correlation of Cenapa normalized counts per gene *versus* their Rubino counterparts. Full circles highlight the expression of genes belonging to housekeeping functions. Colors are as in Fig. 2. (TIFF 1042 kb)

Additional file 4: Table S2. Top 20 upregulated (**a**) and downregulated (**b**) genes for each pairwise comparison. (XLSX 16 kb)

Additional file 5: Table S3. GO category overrepresentation analysis for each two-way comparison. (XLSX 23 kb)

Additional file 6: Table S4. Differential expression of cytoskeleton related gene families. (XLSX 23 kb)

Additional file 7: Table S5. Differential expression of collagen genes. (XLSX 51 kb)

Additional file 8: Table S6. Differential expression of redox genes. (XLSX 40 kb)

Additional file 9: Table S7. Differential expression of detoxifying enzymes. (XLSX 3441 kb)

Additional file 10: Table S8. Differentially expression of genes identified in proteomic study. (XLSX 19 kb)

Abbreviations

ABZ: Albendazole; ABZR: Albendazole-resistant; ABZS: Albendazole-sensitive; ABZ-SO: Albendazole sulphoxide; AC: Adenylate cyclase; DEG: Differentially expressed genes; FDR: False discovery rate; FMO: Flavin mono-oxigenases; GST: Gluthatione S-transferase; PGP: P-glycoprotein; PKA: Protein kinase A; TCBZ: Triclabendazole; TCBZ.SO: Triclabendazole sulphoxide; TCBZ.SO2: Triclabendazole sulphone; TCBZR: Triclabendazole-resistant; TCBZS: Triclabendazole-sensitive

Acknowledgments

We are thankful for the support of agencies CABBIO, CAPES, ANII, PEDECIBA, and CSIC. The authors would like to thank Dr Martín Ciganda for proofreading the manuscript's text and critical review of the results.

Funding

The research was funded by grant CABBIO2012-UY04 awarded to JFT.

Availability of data and materials

The datasets generated and analyzed during the current study are available in the SRA repository, under accession PRJNA339158.

Authors' contributions

SR performed most of the bioinformatic analysis and interpretation of data, and contributed in writing the manuscript. SF contributed to the bioinformatic analysis and the revision of manuscript content. VS, AMS, FMGA, PO, CH, EM, VG, FSMP, HS and GO participated in the data acquisition, contributed to conception of the study and were involved in the critical revision of the manuscript content. PS participated in the design of the study, contributed to bioinformatic analysis and interpretation of data, and was a major contributor in

writing the manuscript. JT participated in the design of the study and the interpretation of data, drafting the manuscript and critical revision of its content. All authors read and approved the final manuscript.

Ethics approval and consent to participate
The three laboratories that collected samples for this study followed protocols approved by the respective local Committees of Animal Experimentation.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Departamento de Genética, Facultad de Medicina, Universidad de la Republica, UDELAR, Montevideo, Uruguay. [2]Present address: Instituto de Investigaciones Biológicas Clemente 28 Estable. MEC, Montevideo 29, Uruguay. [3]Laboratorio de Biología Celular y Molecular, Facultad de Ciencias Veterinarias, Universidad Nacional del Centro de la Provincia de Buenos Aires, Tandil, Argentina. [4]Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil. [5]Laboratorio de Inmunología, Facultad de Ciencias Veterinarias, Universidad Nacional de Cajamarca, Cajamarca, Peru. [6]Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación, Morelos, Mexico. [7]Departamento de Parasitología, División de Laboratorios Veterinarios (DILAVE), "Miguel C. Rubino", Ministerio de Ganadería, Agricultura y Pesca (MGAP), Montevideo, Uruguay. [8]Present address: Instituto Tecnológico Vale, Belém, Brazil. [9]Laboratorio de Interacciones Moleculares, Facultad de Ciencias, Universidad de la Republica, UDELAR, Montevideo, Uruguay.

Received: 2 August 2017 Accepted: 26 November 2017
Published online: 24 January 2018

References

1. Keiser J, Utzinger J. Food-borne trematodiases. Clin Microbiol Rev. 2009; 22:466–83.
2. Charlier J, Vercruysse J, Morgan E, van Dijk J, Williams DJL. Recent advances in the diagnosis, impact on production and prediction of *Fasciola hepatica* in cattle. Parasitology. 2014;141:326–35.
3. Fürst T, Duthaler U, Sripa B, Utzinger J, Keiser J. Trematode infections: liver and lung flukes. Infect Dis Clin N Am. 2012;26:399–419.
4. Carmona C, Tort JF. Fasciolosis in South America: epidemiology and control challenges. J Helminthol. 2017;91:99–109.
5. Oliveira DR, Ferreira DM, Stival CC, Romero F, Cavagnolli F, Kloss A, et al. Triclabendazole resistance involving *Fasciola hepatica* in sheep and goats during an outbreak in Almirante Tamandare, Paraná, Brazil. Rev Bras Parasitol Vet. 2008;17(Suppl 1):149–53.
6. Olaechea F, Lovera V, Larroza M, Raffo F, Cabrera R. Resistance of *Fasciola hepatica* against triclabendazole in cattle in Patagonia (Argentina). Vet Parasitol. 2011;178:364–6.
7. Espinoza JR, Terashima A, Herrera-Velit P, Marcos LA. Human and animal fascioliasis in Peru: impact in the economy of endemic zones. Rev Peru Med Exp Salud Publica. 2010;27:604–12.
8. de Dios Rojas J. Resistance of *Fasciola hepatica* to triclabendazole in cattle of the Cajamarca countryside. Rev Vet Argentina. 2012;7:1–6.
9. Ortiz P, Scarcella S, Cerna C, Rosales C, Cabrera M, Guzmán M, et al. Resistance of *Fasciola hepatica* against Triclabendazole in cattle in Cajamarca (Peru): a clinical trial and an in vivo efficacy test in sheep. Vet Parasitol. 2013;195(1-2):118-21.
10. Sanabria R, Ceballos L, Moreno L, Romero J, Lanusse C, Alvarez L. Identification of a field isolate of *Fasciola hepatica* resistant to albendazole and susceptible to triclabendazole. Vet Parasitol. 2013;193:105–10.
11. Canevari J, Ceballos L, Sanabria R, Romero J, Olaechea F, Ortiz P, et al. Testing albendazole resistance in *Fasciola hepatica*: validation of an egg hatch test with isolates from South America and the United Kingdom. J Helminthol. 2014;88:286–92.
12. Mamani W, Condori R. Anthelminthic resistance (*Fasciola hepatica*) in sheep against albendazole and triclabendazole, La Paz - Bolivia. Rev Inv Vet Peru. 2009;20:254–62.
13. Chavez A, Sanchez L, Arana C, Suarez F. Resistance to anthelmintics and prevalence of bovine fasciolosis in dairy farms in Jauja. Peru Rev Inv Vet Peru. 2012;23:90–7.
14. Gil LC, Díaz A, Rueda C, Martínez C, Castillo D, Apt W. Resistant human fasciolasis: report of four patients. Rev Med Chil. 2014;142:1330–3.
15. Cabada MM, Castellanos-Gonzalez A, Lopez M, Caravedo MA, Arque E, White AC. *Fasciola hepatica* infection in an indigenous community of the Peruvian jungle. Am J Trop Med Hyg. 2016;94:1309–12.
16. Brennan GP, Fairweather I, Trudgett A, Hoey E, McCoy, McConville M, et al. Understanding triclabendazole resistance. Exp Mol Pathol. 2007;82:104–9.
17. Bennett JL, Kohler P. *Fasciola hepatica*: action *in vitro* of triclabendazole on immature and adult stages. Exp Parasitol. 1987;63:49–57.
18. Stitt AW, Fairweather I, Mackender RO. The effect of triclabendazole ("Fasinex") on protein synthesis by the liver fluke, *Fasciola hepatica*. Int J Parasitol. 1995;25:421–9.
19. Halferty L, Brennan GP, Hanna REB, Edgar HW, Meaney MM, McConville M, et al. Tegumental surface changes in juvenile *Fasciola hepatica* in response to treatment in vivo with triclabendazole. Vet Parasitol. 2008;155:49–58.
20. Toner E, Brennan GP, Hanna REB, Edgar HW, Fairweather I. Tegumental surface changes in adult *Fasciola hepatica* in response to treatment in vivo with triclabendazole in the sheep host. Vet Parasitol. 2010;172:238–48.
21. Shareef PAA, Brennan GP, McVeigh P, Khan MAH, Morphew RM, Mousley A, et al. Time-dependent tegumental surface changes in juvenile *Fasciola gigantica* in response to triclabendazole treatment in goat. Acta Trop. 2014; 136:108–17.
22. Stitt AW, Fairweather I. The effect of the sulphoxide metabolite of triclabendazole ('Fasinex') on the tegument of mature and immature stages of the liver fluke, *Fasciola hepatica*. Parasitology. 1994;108:555–67.
23. Stitt AW, Fairweather I. *Fasciola hepatica*: tegumental surface changes in adult and juvenile flukes following treatment in vitro with the sulphoxide metabolite of triclabendazole (Fasinex). Parasitol Res. 1993;79:529–36.
24. Toner E, Brennan GP, Hanna REB, Edgar HW, Fairweather I. Time-dependent changes to the tegumental system and gastrodermis of adult *Fasciola hepatica* following treatment *in vivo* with triclabendazole in the sheep host. Vet Parasitol. 2010;174:218–27.
25. Hanna R. *Fasciola hepatica*: histology of the reproductive organs and differential effects of triclabendazole on drug-sensitive and drug-resistant fluke isolates and on flukes from selected field cases. Pathogens. 2015;4:431–56.
26. Fetterer RH. The effect of albendazole and triclabendazole on colchicine binding in the liver fluke *Fasciola hepatica*. J Vet Pharmacol Ther. 1986;9:49–54.
27. Hanna REB, Edgar HWJ, McConnell S, Toner E, McConville M, Brennan GP, et al. *Fasciola hepatica*: histological changes in the reproductive structures of triclabendazole (TCBZ)-sensitive and TCBZ-resistant flukes after treatment *in vivo* with TCBZ and the related benzimidazole derivative, compound alpha. Vet Parasitol. 2010;168:240–54.
28. Stitt AW, Fairweather I. *Fasciola hepatica*: disruption of the vitelline cells in vitro by the sulphoxide metabolite of triclabendazole. Parasitol Res. 1996;82:333–9.
29. Stitt AW, Fairweather I. Spermatogenesis in *Fasciola hepatica*: an ultrastructural comparison of the anthelmintic, triclabendazole ("Fasinex") and the microtubule inhibitor, tubulozole. Invertebr Reprod Dev. 1992;22:139–50.
30. Robinson M, Trudgett A, Hoey EM, Fairweather I. Triclabendazole-resistant *Fasciola hepatica*: β-tubulin and response to in vitro treatment with triclabendazole. Parasitology. 2002;124:325–38.
31. McConville M, Brennan GP, McCoy M, Castillo R, Hernandez-Campos A, Ibarra F, et al. Immature triclabendazole-resistant *Fasciola hepatica*: Tegumental responses to in vitro treatment with the sulphoxide metabolite of the experimental fasciolicide compound alpha. Parasitol Res. 2007;100:365–77.
32. Lee YJ, Shi R, Witt SN. The small molecule triclabendazole decreases the intracellular level of cyclic AMP and increases resistance to stress in *Saccharomyces cerevisiae*. PLoS One. 2013;8:e64337.
33. Fairweather I, Boray JC. Fasciolicides: efficacy, actions, resistance and its management. Vet J. 1999;158:81–112.
34. Buchanan JF, Fairweather I, Brenna GP, Trudgett A, Hoey EM. *Fasciola hepatica*: surface and internal tegumental changes induced by treatment *in vitro* with the sulphoxide metabolite of albendazole ('Valbazen'). Parasitology. 2003;126:141–53.

35. Alvarez LI, Solana HD, Mottier ML, Virkel GL, Fairweather I, Lanusse CE. Altered drug influx/efflux and enhanced metabolic activity in triclabendazole-resistant liver flukes. Parasitology. 2005;131:501–10.
36. Robinson MW, Lawson J, Trudgett A, Hoey EM, Fairweather I. The comparative metabolism of triclabendazole sulphoxide by triclabendazole-susceptible and triclabendazole-resistant *Fasciola hepatica*. Parasitol Res. 2004;92:205–10.
37. Solana H, Scarcella S, Virkel G, Ceriani C, Rodríguez J, Lanusse C. Albendazole enantiomeric metabolism and binding to cytosolic proteins in the liver fluke *Fasciola hepatica*. Vet Res Commun. 2009;33:163–73.
38. Mottier L, Alvarez L, Fairweather I, Lanusse C. Resistance-induced changes in triclabendazole transport in *Fasciola hepatica*: ivermectin reversal effect. J Parasitol. 2006;92:1355–60.
39. Savage J, Meaney M, Brennan GP, Hoey E, Trudgett A, Fairweather I. Increased action of triclabendazole (TCBZ) in vitro against a TCBZ-resistant isolate of *Fasciola hepatica* following its co-incubation with the P-glycoprotein inhibitor, R(+)-verapamil. Exp Parasitol. 2013;135:642–53.
40. Savage J, Meaney M, Brennan GP, Hoey E, Trudgett A, Fairweather I. Effect of the P-glycoprotein inhibitor, R(+)-verapamil on the drug susceptibility of a triclabendazole-resistant isolate of *Fasciola hepatica*. Vet Parasitol. 2013; 195:72–86.
41. Devine C, Brennan GP, Lanusse CE, Alvarez LI, Trudgett A, Hoey E, et al. Effect of the metabolic inhibitor, methimazole on the drug susceptibility of a triclabendazole-resistant isolate of *Fasciola hepatica*. Parasitology. 2009; 136:183–92.
42. Devine C, Brennan GP, Lanusse CE, Alvarez LI, Trudgett A, Hoey E, et al. Enhancement of triclabendazole action *in vivo* against a triclabendazole-resistant isolate of *Fasciola hepatica* by co-treatment with ketoconazole. Vet Parasitol. 2011;177:305–15.
43. Devine C, Brennan GP, Lanusse CE, Alvarez LI, Trudgett A, Hoey E, et al. Potentiation of triclabendazole action in vivo against a triclabendazole-resistant isolate of *Fasciola hepatica* following its co-administration with the metabolic inhibitor, ketoconazole. Vet Parasitol. 2012;184:37–47.
44. Scarcella S, Miranda-Miranda E, Cossio-Bayúgar R, Ceballos L, Fernandez V, Solana H. Increase of carboxylesterase activity in *Fasciola hepatica* recovered from triclabendazole treated sheep. Mol Biochem Parasitol. 2012;185:151–3.
45. Scarcella S, Lamenza P, Virkel G, Solana H. Expression differential of microsomal and cytosolic glutathione-S-transferases in *Fasciola hepatica* resistant at triclabendazole. Mol Biochem Parasitol. 2012;181:37–9.
46. Chemale G, Perally S, LaCourse EJ, Prescott MC, Jones LM, Ward D, et al. Comparative proteomic analysis of triclabendazole response in the liver fluke *Fasciola hepatica*. J Proteome Res. 2010;9:4940–51.
47. Matoušková P, Vokřál I, Lamka J, Skálová L. The role of xenobiotic-metabolizing enzymes in anthelmintic deactivation and resistance in helminths. Trends Parasitol. 2016;32:481–91.
48. Valentim CLL, Cioli D, Chevalier FD, Cao X, Taylor AB, Holloway SP, et al. Genetic and molecular basis of drug resistance and species-specific drug action in schistosome parasites. Science. 2013;342:1385–9.
49. Chevalier FD, Valentim CL, LoVerde PT, Anderson TJ. Efficient linkage mapping using exome capture and extreme QTL in schistosome parasites. BMC Genomics. 2014;15:617–29.
50. You H, McManus DP, Hu W, Smout MJ, Brindley PJ, Gobert GN. Transcriptional responses of in vivo praziquantel exposure in schistosomes identifies a functional role for calcium signaling pathway member CamKII. PLoS Pathog. 2013;9:e1003254.
51. Kasinathan RS, Morgan WM, Greenberg RM. *Schistosoma mansoni* express higher levels of multidrug resistance-associated protein 1 (SmMRP1) in juvenile worms and in response to praziquantel. Mol Biochem Parasitol. 2010;173:25–31.
52. Hines-Kay J, Cupit PM, Sanchez MC, Rosenberg GH, Hanelt B, Cunningham C. Transcriptional analysis of *Schistosoma mansoni* treated with praziquantel in vitro. Mol Biochem Parasitol. 2012;186:87–94.
53. Hodgkinson J, Cwiklinski K, Beesley NJ, Paterson S, Williams DJL. Identification of putative markers of triclabendazole resistance by a genome-wide analysis of genetically recombinant *Fasciola hepatica*. Parasitology. 2013;140:1523–33.
54. Cwiklinski K, Dalton JP, Dufresne PJ, La Course J, Williams DJ, Hodgkinson J, et al. The *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. Genome Biol. 2015;16:71–84.
55. McNulty SN, Tort JF, Rinaldi G, Fischer K, Rosa BA, Smircich P, et al. Genomes of *Fasciola hepatica* from the Americas reveal colonization with *Neorickettsia* endobacteria related to the agents of potomac horse and human sennetsu fevers. PLoS Genet. 2017;13:e1006537.
56. National Research Council (USA). Committee for the Update of the guide for the care and use of laboratory animals. Guide for the care and use of laboratory animals. 8th ed. Washington (DC): National Academies Press (USA); 2011.
57. R Development Core Team. R: A language and environment for statistical computing: the R Foundation for Statistical Computing, Vienna, Austria; 2016. URL: https://www.R-project.org/
58. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with bioconductor. Nat Methods. 2015;12:115–21.
59. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9:1–11.
60. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550–71.
61. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57:289–300.
62. Alexa A, Rahnenführer J. topGO: Enrichment analysis for Gene Ontology. R Package version 2.16.0. 2010.
63. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite - a comprehensive resource for helminth genomics. Mol Biochem Parasitol. 2017;215:2–10.
64. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
65. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17:122–36.
66. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428:726–31.
67. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45:D353–61.
68. Gaboriaud C, Gregory-Pauron L, Teillet F, Thielens NM, Bally I, Arlaud GJ. Structure and properties of the Ca2+–binding CUB domain, a widespread ligand-recognition unit involved in major biological functions. Biochem J. 2011;439:185–93.
69. Laing ST, Ivens A, Laing R, Ravikumar S, Butler V, Woods DJ, et al. Characterization of the xenobiotic response of *Caenorhabditis elegans* to the anthelmintic drug albendazole and the identification of novel drug glucoside metabolites. Biochem J. 2010;432:505–14.
70. Lacey E. The role of the cytoskeletal protein, tubulin, in the mode of action and mechanism of drug resistance to benzimidazoles. Int J Parasitol. 1988; 18:885–936.
71. Chambers E, Ryan LA, Hoey EM, Trudgett A, McFerran NV, Fairweather I, et al. Liver fluke B-tubulin isotype 2 binds albendazole and is thus a probable target of this drug. Parasitol Res. 2010;107:1257–64.
72. Von Samson-Himmelstjerna G, Blackhall WJ, McCarthy JS, Skuce PJ. Single nucleotide polymorphism (SNP) markers for benzimidazole resistance in veterinary nematodes. Parasitology. 2007;134:1077–86.
73. Demeler J, Krüger N, Krücken J, von der Heyden VC, Ramünke S, Küttler U, et al. Phylogenetic characterization of β-tubulins and development of pyrosequencing assays for benzimidazole resistance in cattle nematodes. PLoS One. 2013;8:e70212.
74. Ryan LA, Hoey E, Trudgett A, Fairweather I, Fuchs M, Robinson MW, et al. *Fasciola hepatica* expresses multiple alpha- and beta-tubulin isotypes. Mol Biochem Parasitol. 2008;159:73–8.
75. Kwa MS, Kooyman FN, Boersema JH, Roos MH. Effect of selection for benzimidazole resistance in *Haemonchus contortus* on beta-tubulin isotype 1 and isotype 2 genes. Biochem Biophys Res Commun. 1993;191:413–9.
76. Kwa MSG, Veenstra JG, Roos MH. Molecular characterization of β-tubulin genes present in benzimidazole-resistant populations of *Haemonchus contortus*. Mol Biochem Parasitol. 1993;60:133–43.
77. Fuchs M-A, Ryan LA, Chambers EL, Moore CM, Fairweather I, Trudgett A, et al. Differential expression of liver fluke β-tubulin isotypes at selected life cycle stages. Int J Parasitol. 2013;43:1133–9.
78. Kotze AC, Hunt PW, Skuce P, von Samson-Himmelstjerna G, Martin RJ, Sager H, et al. Recent advances in candidate-gene and whole-genome

approaches to the discovery of anthelmintic resistance markers and the description of drug/receptor interactions. Int J Parasitol Drugs Drug Resist. 2014;4:164–84.
79. Messerli SM, Kasinathan RS, Morgan W, Spranger S, Greenberg RM. *Schistosoma mansoni* P-glycoprotein levels increase in response to praziquantel exposure and correlate with reduced praziquantel susceptibility. Mol Biochem Parasitol. 2009;167:54–9.
80. Mansour TE. Chemotherapy of parasitic worms: new biochemical strategies. Science. 1979;205:462–9.
81. García-Huertas P, Mejía-Jaramillo AM, González L, Triana-Chávez O. Transcriptome and functional genomics reveal the participation of adenine phosphoribosyltransferase in *Trypanosoma cruzi* resistance to benznidazole. J Cell Biochem. 2017;118:1936–45.

frontiers in Genetics

ORIGINAL RESEARCH
published: 05 September 2019
doi: 10.3389/fgene.2019.00771

Compositional Analysis of Flatworm Genomes Shows Strong Codon Usage Biases Across All Classes

Guillermo Lamolle[1], Santiago Fontenla[1], Gastón Rijo[1], Jose F. Tort[1] and Pablo Smircich[2,3]**

[1] *Departamento de Genética, Facultad de Medicina, Universidad de la Republica, UDELAR, Montevideo, Uruguay,* [2] *Departamento de Genómica, Instituto de Investigaciones Biológicas Clemente Estable, IIBCE, MEC, Montevideo, Uruguay,* [3] *Laboratorio de Interacciones Moleculares, Facultad de Ciencias, Universidad de la Republica, UDELAR, Montevideo, Uruguay*

In the present work, we performed a comparative genome-wide analysis of 22 species representative of the main clades and lifestyles of the phylum Platyhelminthes. We selected a set of 700 orthologous genes conserved in all species, measuring changes in GC content, codon, and amino acid usage in orthologous positions. Values of 3rd codon position GC spanned over a wide range, allowing to discriminate two distinctive clusters within freshwater turbellarians, Cestodes and Trematodes respectively. Furthermore, a hierarchical clustering of codon usage data differs remarkably from the phylogenetic tree. Additionally, we detected a synonymous codon usage bias that was more dramatic in extreme GC-poor or GC-rich genomes, i.e., GC-poor Schistosomes preferred to use AT-rich terminated synonymous codons, while GC-rich *M. lignano* showed the opposite behavior. Interestingly, these biases impacted the amino acidic usage, with preferred amino acids encoded by codons following the GC content trend. These are associated with non-synonymous substitutions at orthologous positions. The detailed analysis of the synonymous and non-synonymous changes provides evidence for a two-hit mechanism where both mutation and selection forces drive the diverse coding strategies of flatworms.

Keywords: flatworms, GC content, synonymous codons, codon usage, non-synonymous substitutions, amino acid usage, mutation, selection

OPEN ACCESS

Edited by:
Daniel Yero,
Autonomous University of Barcelona, Spain

Reviewed by:
Deng-Feng Xie,
Sichuan University, China
Arif Uddin,
Assam University, India

****Correspondence:***
Jose F. Tort
jtort@fmed.edu.uy
Pablo Smircich
psmircich@fcien.edu.uy

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Genetics

Received: *24 May 2019*
Accepted: *22 July 2019*
Published: *05 September 2019*

Citation:
Lamolle G, Fontenla S, Rijo G, Tort JF and Smircich P (2019) Compositional Analysis of Flatworm Genomes Shows Strong Codon Usage Biases Across All Classes. Front. Genet. 10:771. doi: 10.3389/gene.2019.00771

INTRODUCTION

The phylum Platyhelminthes with more than 30,000 species is one of the major phyla of invertebrate animals containing an enormous diversity of life forms that had colonized very diverse niches (Caira and Littlewood, 2013). Almost three quarters of the flatworms are parasitic and belong to the Neodermata, a monophyletic clade characterized by a syncytial tegument and the presence of diverse specialized organs to attach to hosts like suckers and hooks. The Neodermata comprise three classes: the Monogenea (primarily external parasites of cold-blooded aquatic vertebrates), the Cestoda (obligate endoparasites of vertebrates), and the Trematoda (endoparasites of vertebrates as adults, with intermediate stages endoparasitic in other invertebrates, mainly mollusks) (Caira and Littlewood, 2013). Besides the parasitic Neodermatans, an enormous diversity of species occurs in seas, rivers, and lakes and on all continental land masses comprising one of the successful phyla of invertebrates (Collins, 2017). A few species exist as either commensals or occasional parasites of

invertebrates, but most of them are free-living predator species forming a single paraphyletic group collectively referred to as "turbellarians" (Caira and Littlewood, 2013). Studies based on rRNA (Larsson and Jondelius, 2008; Laumer and Giribet, 2014) and transcriptomic data (Egger et al., 2015; Laumer et al., 2015) showed that the phylum Platyhelminthes split early into two clades: the ancestral Catenulida and the Rhabditophora, which includes several free-living orders and the Neodermatans (**Table 1**). Taxonomically, the Macrostomorpha was placed as the earliest diverging Rhabditophoran linage and the order Tricladida, which contains the model organism *Schmidtea mediterranea*, as part of the later-evolved "turbellarians."

The huge diversity of flatworm's life forms seems to be paralleled at genomic level. The recent publication of several genomic assemblies of the phylum (most of them corresponding to parasitic Neodermatans) has revealed a wide genomic diversity. For example, genome sizes range from 67 or 104 Mbases in the monogenean *Gyrodactilus salaris* or the cestode *Hydatigera taeniaeformis*, respectively, to 1,200 Mbases in the trematode *Fasciola hepatica* (Coghlan et al., 2019). Interestingly, this variation has little correlation with gene set completeness among genomes and is mostly due to non-coding elements, including repetitive and non-repetitive elements, with repeat content ranging from less than 4% in the smallest genomes of cestodes to 68% in *Fasciola hepatica*. Additionally, guanine and cytosine (GC) contents are very diverse from 28% in the planaria *S. mediterranea* and 33% in Monogenea *Gyrodactylus salaris*, to more than 45% in *M. lignano* and the food-borne trematodes (FBT) *F. hepatica*, *C. sinensis*, and *O. volvulus* (Coghlan et al., 2019).

We wondered if these large variations in genomic composition and structure could be correlated with the morphological and ecological diversity. It is well known that genomic GC content determines codon usage across species (Bernardi and Bernardi, 1985; Plotkin and Kudla, 2011) and the use of alternative synonymous codons is a non-random process (Sharp et al., 2010; Plotkin and Kudla, 2011). Due to the degeneracy of the genetic code, most amino acids, with the exceptions of methionine and tryptophan, are encoded by more than one codon. Codon usage bias (CUB) is a phenomenon where synonymous codons are not used with equal frequencies in coding DNA. It has been suggested that codon usage bias is the result of an equilibrium between mutational bias and natural selection and that natural selection could be acting in presumably highly expressed genes (Sharp et al., 2010; Plotkin and Kudla, 2011). Besides the effect at synonymous codon usage, it has been shown that strong GC bias could lead to changes in amino acid frequencies (Behura and Severson, 2013; Li et al., 2015). While this has not been explored widely in flatworms, several advances have been made in nematodes (Cutter et al., 2006; Mitreva et al., 2006; Mazumder et al., 2017a; Mazumder et al., 2017b). It is not clear yet how genomic GC differences could be influencing the codon usage and amino acid composition of proteins in Platyhelminthes and if these variations correlate with the ecological and physiological diversity in the phylum.

First reports of flatworm codon usage predated the genomic era and were based on a low representative number of sequences in Schistosomes and *Echinococcus*. Heterogeneity was evidenced since Schistosomes preferred A+T-rich codons, while *Echinococcus* favored GC3-rich codons (Meadows and Simpson, 1989; Alvarez et al., 1993; Kalinna and McManus, 1994; Milho and Tracy, 1995). Further analysis in larger sets of genes showed that codon bias was not uniformly distributed between genes introducing the possibility of isochores (regions that differ in GC content) in the genomes of flatworms (Ellis and Morrison, 1995; Ellis et al., 1995). In agreement, a more recent compositional analysis of

TABLE 1 | List of species analyzed and their GC content.FL, Free-living; PR, parasitic; G.GC, genomic GC percentage; T.GC, transcript (CDS) GC percentage.

Class	Subclass/order	Species	Abbreviation	Habitat	G.GC	T.GC
CATENULIDA		*Stenostomum leucops*	Sleu	FL		54.0
RHABDITOPHORA	Macrostomorpha	*Macrostomum lignano*	Mlig	FL	45.9	58.7
RHABDITOPHORA	Lecithoepitheliata	*Geocentrophora applanata*	Gapp	FL		37.8
RHABDITOPHORA	Polycladida	*Prosthecereaeus vittatus*	Pvit	FL		46.4
RHABDITOPHORA	Neodalyellida/Rhabdocoela	*Rhynchomesostoma rostratum*	Rros	FL		40.1
RHABDITOPHORA	Dalyellioida/Fecamplida	*Kronborgia amphipodicola*	Kamp	PR		40.3
RHABDITOPHORA	Seriata/Proseriata	*Monocelis fusca*	Mfus	FL		40.8
RHABDITOPHORA	Seriata/Bothrioplanida	*Bothrioplana semperi*	Bsem	FL		53.0
RHABDITOPHORA	Seriata/Tricladida	*Schmidtea mediterranea*	Smed	FL	29.9	36.6
MONOGENEA	Monopisthocotylea/Gyrodactylidea	*Gyrodactilus salaris*	Gsal	PR	33.9	43.3
MONOGENEA	Polyopisthocotylea/Polystomatidea	*Protopolystoma xenopodis*	Pxen	PR	37.7	50.2
TREMATODA	Digenea/Strigeidida	*Schistosoma mansoni*	Sman	PR	35.5	36.0
TREMATODA	Digenea/Strigeidida	*Schistosoma japonicum*	Sjap	PR	34.1	36.0
TREMATODA	Digenea/Strigeidida	*Trichobilharzia regenti*	Treg	PR	37.4	37.2
TREMATODA	Digenea/Plagiorchiida	*Fasciola hepatica*	Fhep	PR	44.1	47.8
TREMATODA	Digenea/Opisthorchiida	*Clonorchis sinensis*	Csin	PR	44.0	48.4
TREMATODA	Digenea/Opisthorchiida	*Opistorchis viverrini*	Oviv	PR	43.8	48.5
CESTODA	Eucestoda/Diphyllobothriidea	*Schistocephalus solidus*	Ssol	PR	43.0	51.9
CESTODA	Eucestoda/Cyclophyllidea	*Mesocestoides corti*	Mcor	PR	36.7	51.4
CESTODA	Eucestoda/Cyclophyllidea	*Hymenolepis diminuta*	Hdim	PR	35.2	44.2
CESTODA	Eucestoda/Cyclophyllidea	*Echinococcus granulosus*	Egra	PR	41.9	50.0
CESTODA	Eucestoda/Cyclophyllidea	*Echinococcus multiloculoris*	Emul	PR	42.2	49.9

FL, free-living; PR, parasitic; G.GC, genomic GC percentage; T.GC, transcript (CDS) GC percentage.

the *S. mansoni* genome reported an isochore-like organization (Lamolle et al., 2016). Early studies analyzing the forces behind codon bias found evidences of both mutational pressure (Musto et al., 1998) in *S. mansoni* and selection (Fernandez et al., 2001) in *Echinococcus* spp. as preponderant forces. More recently, studies on *S. haematobium* and *S. japonicum* confirmed a major role of natural selection in shaping the codon usage bias in these species (Mazumder et al., 2017a; Mazumder et al., 2017b). Several studies analyzed codon usage on the available genomes and transcriptomes of *Taenidae* species showing weak codon bias and a higher GC3 in highly expressed genes explained by combined mutational and selection forces (Chen et al., 2013; Yang et al., 2014; Yang et al., 2015; Huang et al., 2017). A more preponderant contribution of selection shaping codon usage was identified in a comparative analysis in *Echinococcus* species (Maldonado et al., 2018). While these studies highlight that platyhelminthes are compositionally varied, they are focused just in the schistosomes and tapeworms. We took advantage of the wide array of transcriptomes and genomes now available to extend the study to a phylum-wide analysis of codon usage patterns, as a proxy of the molecular organization of flatworm genomes. We performed a comparative analysis at the genomic level of 22 species representative of the main clades and lifestyles of the phylum Platyhelminthes. Within these species, we picked a set of 700 orthologous gene groups conserved across the 22 species and measured changes in GC content, codon, and amino acid usage in orthologous positions. We found a class independent-wide diversity in codon and amino acid usages. Based on the study of orthologous positions in selected pairs of species with diverse GC content, we provide evidence of a combined contribution of mutational forces and selection that enforced synonymous codon usage bias and differential amino acid usage.

METHODS

Data Acquisition

Genomic and coding sequences of 22 flatworm species were used in this work. To ease data visualization, a four-letter code was used to name the species (**Table 1**). Genomic and transcriptomic data of Mlig, Smed, Gsal, Pxen, Csin, Oviv, Fhep, Treg, Sjap, Sman, Mcor, Hdim, Egra, Emul, and Ssol were obtained from the public repository Wormbase parasite (Howe et al., 2017) (https://parasite.wormbase.org/). Transcriptomic data on Sleu, Gapp, Pvit, Rros, Mfus, Kamp, and Bsem were generated by Laumer et al. (2015) and downloaded from the public repository Data Dryad (doi:10.5061/dryad.622q4).

Orthologues Determination

In-house Perl and Bash scripts that implemented a BLASTp best reciprocal hit strategy were used to identify a core of orthologous genes. An e-value cutoff of 1e-5 was used to define significant hits. The restrictive method produced one orthologue gene per species. A total of 700 orthologous groups were detected in all 22 species, and these sequences were used for the analysis, adding to more than 8 million codons analyzed (8.242.428).

Expression data for *S. mansoni* in reads per kilobase per million (RPKM) were taken from the study of Protasio et al. (2012). Expression data were available for 696 of the 700 *S. mansoni* orthologs. Expression data for the adult stages of *F. hepatica*, *E. granulosus*, *H. diminuta*, *S. mediterranea*, and *M. lignano* were downloaded from WormBase Parasite (Howe et al., 2017).

Gene Alignment and Phylogenetic Tree

Each group of 22 orthologous sequences were translated with an in-house Perl script and aligned individually with Mafft (Katoh and Standley, 2013). Individual alignments were concatenated into a unique alignment. This alignment was used to build a phylogenetic tree with PhyML (Guindon et al., 2010). PhyML was run with the following options: -b -4, to calculate statistical branch support, -s BEST, for tree topology estimation, -m LG, to indicate the model substitution matrix, and -o tl, for tree topology and branch length optimization.

For the hierarchical clustering based on GC content of synonymous codons (RSCU), several clusters were built using hclust from R Stats package with different option settings (R Core Team, 2019). A final consensus cluster was made with the Ape package (Paradis et al., 2004), which retained the most frequent groupings.

Codon Usage Analysis

Codon usage and compositional analyzes were done in R with the package seqinR (Charif and Lobry, 2007). Correlations between frequencies of each codon and GC3 were represented as heatmap with the R "Corrplot" package (Wei and Simko, 2017). In-house R scripts were used to evaluate significance of changes in frequencies between high- and low-expressed gene sets and defined preferred codons. A codon was considered "preferred" if its frequency (RSCU) significantly increases in a set of high-expression genes, compared with a low-expression set, regardless of whether it becomes the main codon for that amino acid or not. Correspondence analysis (COA) was performed in R.

Neutrality and Effective Number of Codons Plots

Neutrality plots (Sueoka, 1988) (GC3 vs GC12) of the 22 species were used to evaluate the relationship among the three codon positions. Additionally, a unique plot showing general GC3–GC12 for all species was calculated by using a concatenated super gene for each species.

The effective number of codons (ENC) is used to quantify the variation in codon usage, ranging from 20 (when only one codon per amino acid is used) to 61 (when all possible codons are used). GC3 vs ENC charts are useful to estimate selection contribution to CUB. Expected values of ENC based on mutation pressure generate a bell curve, so in these charts, the points that fall directly on the curve represent genes with neutral evolution, while the points under the curve suggest action of natural selection (Wright, 1990).

Amino Acid and Codon Substitutions Matrices

From the amino acid alignments of each COG, the sites that had gaps in one or more sequences were eliminated. Degapped COGs with less than 35 amino acids were eliminated. The resulting sequences were then concatenated, generating a "super-peptide" (without gaps) for each species. Then, the amino acid changes between each pair of species were counted (with a homemade R script), creating 20 by 20 substitution matrices. Each value of the matrix ($A_{Z,X}$) represents how many times amino acid Z is present in one species, while amino acid X is present in the corresponding orthologous position in the other species, being, therefore, an asymmetric matrix. The diagonal of the matrix represents the unchanged sites, while the sum of the remaining values in each column or row represents the total substitutions for each amino acid. To test for deviations in the amino acid usage between species, the total count for each amino acid in the species of a pair was calculated, and the average was considered as expected value to perform chi-square tests (**Figure 5**). For each reciprocal changes in the matrix ($A_{Z,X}$, $B_{X,Z}$), a chi-square test was performed considering the average of the counts as expected value. For simplicity of analysis, we focused in three comparisons between species with different global GC: cestodes (Hdim and Egra), trematodes (Sman and Fhep), and free-living species (Smed and Mlig). The last comparison involved the two more divergent species in GC content.

Based on the back-translation of the alignments, we generated a 61 × 61 (stop codons deleted) codon substitution matrix for the six selected species following similar procedures as the ones described in the previous section.

RESULTS

Global GC Composition Varies Across Diverse Flatworm Taxa

As a first approach to analyze if there is a GC compositional difference in the phylum Platyhelminthes, we inspected the difference in the global genomic and transcriptomic G+C content. At first glance, it was clear that there is no correlation between genomic and transcriptomic GC, so it was not possible to use transcript GC to infer genomic GC. In most of the species, transcripts were GC richer than global genomic GC with the only exception of *T. regenti* (**Table 1**). However, while Schistosomatidae species show almost no difference in GC content between the overall genome and the coding region, Cestodes transcripts, for example, were on average 9.7% GC richer than all the genome considered together.

GC Composition Varies Across Diverse Flatworm Taxa

To further analyze the GC composition in the coding region, we searched for a set of orthologous conserved genes in the available genomes and transcriptomes. Based on a best reciprocal hit BLAST search, we selected 700 orthologous genes present in the 22 species. A maximum likelihood tree confirmed that the orthologous groups strongly represented the accepted phylogeny of the species analyzed (organisms form the same group cluster together, with the only exception of the two Monogenea, which represent two distinct subclasses) (**Figure 1A**). Since GC varies across the different species, we calculated the relative synonymous codon usage (RSCU) and the GC values by codon position in this set of conserved genes. The clustering of the species based on the relative synonymous codon usage (RSCU) data showed an important reorganization respect to the phylogenetic tree (**Figure 1B**). Three main clusters were clearly appreciable: the first with high GC3 (with free-living species), the second with low GC3 values (including other free-living species and the schistosomatids), and a third with intermediate GC3 values. GC2 was the less variable between groups (0.42, 0.38, and 0.41 on average in groups 1, 2, and 3, respectively). Additionally, we noticed three subgroups within group 3: one that had lower GC1–2 than the rest but had high GC3 composed only by the monogean *G. salaris*; the cluster of *P. xenopodis, P. vittatus, M. fusca*, and the cestode *H. diminuta* that had lower GC3; and the subgroup composed by trematodes (*F. hepatica, O. viverrini, C. sinensis*) and cestodes (*S. solidus, M. corti, E. multilocularis,* and *E. granulosus*) that had higher G1-3 compared with other species of the group. This shows that global synonymous codon usage varies widely across the phylum.

The relation of GC values in 1st and 2nd position versus those presented in 3rd codon position (neutrality plot) is usually used to evaluate if the variations in codon usage are driven by mutation or selection. Neutrality plots for the 22 organisms based on the 700 orthologue genes were analyzed. In all cases, low slopes were found for the regression curve (maximum value of 0.2) (**Supplementary Figure 1**). Careful inspection of the plots indicates that this can be explained by a low variability of GC1/2 among the genes (ranges between 0.4 and 0.6). These results suggest a contribution of selection in shaping codon usage for these organisms. To visualize all species together, we plotted the GC1–2 versus GC3 of the concatenated orthologue groups (COGs) for each species (see Methods). Expectedly, while GC1–2 presented little variation across species, the best discriminator was variation at GC3 (**Figure 2**). For example, between the most GC biased genomes, the GC-poor *S. mediterranea* and the GC-rich *M. lignano*, there was only 10% variation in GC1–2 axis but 40% variation in GC3. Interestingly, based on GC3 variability, we found two distinctive clusters within the freshwater "turbellarians," trematodes and cestodes. The trematodes species clearly differentiated the blood-dwelling flukes grouped on the lowest side of the GC3 spectrum to the food-borne liver flukes allocated in the middle upper GC3 range. Similarly, cestodes tend to cluster on the upper side of the GC3 range with the exception of the Hymenolepidae that fall on the lower-middle of the GC range. Freshwater "turbellarians" showed the largest variability in both the GC12 and GC3 range grouping into very distant clusters. However, we found no clear evolutionary-GC content correlation as species belonging to different lineages were mixed in both groups.

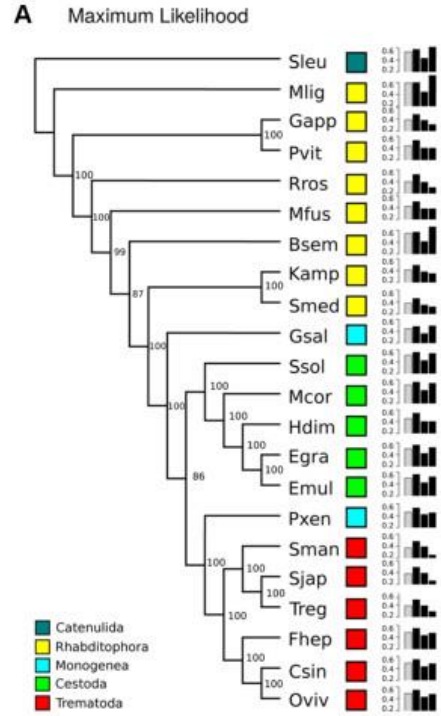


FIGURE 1 | GC composition variation across Platyhelminthes. **(A)** Maximum likelihood phylogenetic tree built with a set of 700 orthologue genes. Light blue, green, and red boxes correspond to the Neodermata species; blue and yellow boxes correspond to "turbellaria" species (see abbreviations in **Table 1**). SH-like support values are indicated for each node. **(B)** Hierarchical clustering based on GC content of relative synonymous codons (RSCU). Histograms reflect overall GC content in all orthologues (gray bar) and by codon position (black bars). GC3 is the most variable between species.

GC Bias in Coding Sequences Affect Codon Usage

Codon usage bias is a general feature of genomes that has been widely associated with GC content (mutational bias) and natural selection (Bernardi and Bernardi, 1985; Sharp et al., 2010; Plotkin and Kudla, 2011). In this context, we decided to study its extent and its relationship with the genomic GC frequency discussed in the previous section. To this end, heatmaps were plotted to visualize the correlations between GC3 and codon usage (Palidwor et al., 2010). While codon usage bias is observed for all species, the more compositionally skewed organisms (the three plots on the right) show more intense correlations, indicating that the phenomena are stronger in these organisms as might be expected. Also, in most cases, the correlation values were positive for GC-ended codons and negative for the AT-ended ones (**Figure 3**). To further characterize this relationship, the distribution of the frequency of synonymous codons was analyzed for all organisms. A dramatic split of GC- vs AT-ended codons is observed in the species with the more biased GC genomes as the AT-rich model trematode *S. mansoni*. Notably, the split is seen in opposite directions in the free-living flatworms *S. mediterranea* and *M. lignano* that are at the extremes of the GC distribution (**Figure 4**). A less marked but significant difference is seen within the cestodes consistent with a more balanced GC content, a feature confirmed in the analysis of the 22 species across flatworm diversity (**Supplementary Figure 2**). These observations suggest that genome-wide mutational bias is a major contributor to the observed codon frequency profiles for each organism.

The GC3 vs ENC charts for the analyzed species (**Supplementary Figure 3**) show a combined contribution of selection and mutation for most of the species supporting the trends observed previously, while heavily biased genomes of Schistosomes fall on the curve, suggesting a strong effect of mutational bias.

Differential Codon Usage Is Associated With Expression Levels

While mutation bias influences codon usage in a genome-wide fashion, selection may also act on coding sequences to select for specific codons. This theory predicts that more frequent codons are actually more efficient and/or more accurate during translation of

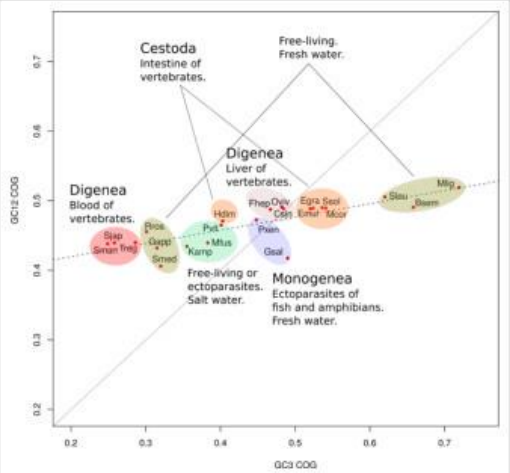


FIGURE 2 | GC variation and lifestyles correlations. Mean of GC12 versus GC3 for orthologue groups. GC3 is a better discriminator of species than GC12. GC content was linked to the ecological niche of the adult form.

the mRNA (Sharp et al., 2010; Plotkin and Kudla, 2011). To test if this phenomenon is observed in flatworms, steady-state mRNA levels for the available species were collected to differentiate high- and low-expression genes. As shown in **Figure 5**, where the two main components of a PCA of codon usage for *S. mansoni* are plotted, high- and low-expression genes do present a distinct usage profile. Interestingly, when comparing the 10% higher- and lower-expressed genes in the adult stage, a preference for using GC-rich codons is observed narrowing the distribution in the highly expressed genes and extending it in the lowly expressed (**Supplementary Figure 4**). Even though these results may be explained by biased repair mechanisms acting on highly transcribed sequences, this result is also compatible with translational selection acting on these genes to drive the observed bias.

Amino Acid Usage Is Also Biased in Diverse Flatworm Lineages

The strong bias observed in codon usage is expected to be associated with the 3rd codon position allowing synonymous changes. However, variations might also exist at the amino acid level (Li et al., 2015). To investigate this, we analyzed the amino acid usage within the set of 700 orthologue genes in pairs of species. Subtle but significant differences in the amino acid frequencies can be detected in cestodes and trematodes mainly involving the amino acids encoded by AT-rich [Ile (AUR), Asn (AAY), Lys (AAR)] or GC-rich [Arg (CGN), Ala (GCN)] codons (**Figure 6** and **Supplementary Table 1**). The variations are more pronounced in the comparison of the free-living species, and in all the cases, the variation follows the GC trend of the species. Since these results are based on a set of orthologue genes, the variations in amino acid frequencies indicate that not only synonymous changes account for the variability observed but also non-synonymous substitutions are taking place.

Synonymous and Non-Synonymous Substitutions in Conserved Orthologous Genes

We decided to investigate if particular directional changes could be detected when analyzing orthologous positions in the three

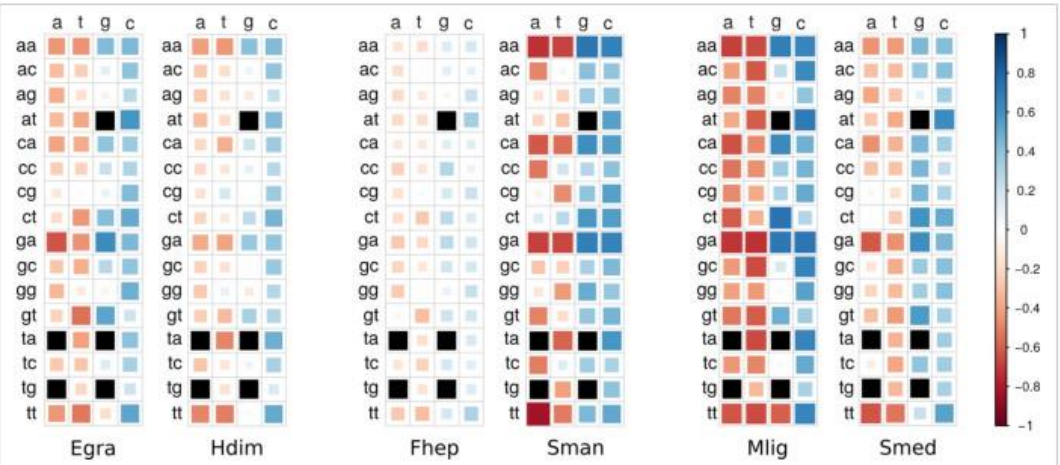


FIGURE 3 | Heatmaps of correlations between codon usage and GC3. Correlation values are coded according to the color scale depicted in the side bar (blue = positive, red = negative). Color intensity and the size of the rectangle are proportional to the correlation coefficients. Black squares: ATG (Met), TGG (Trp), and STOP codons. Organism name abbreviations are as in **Table 1**.

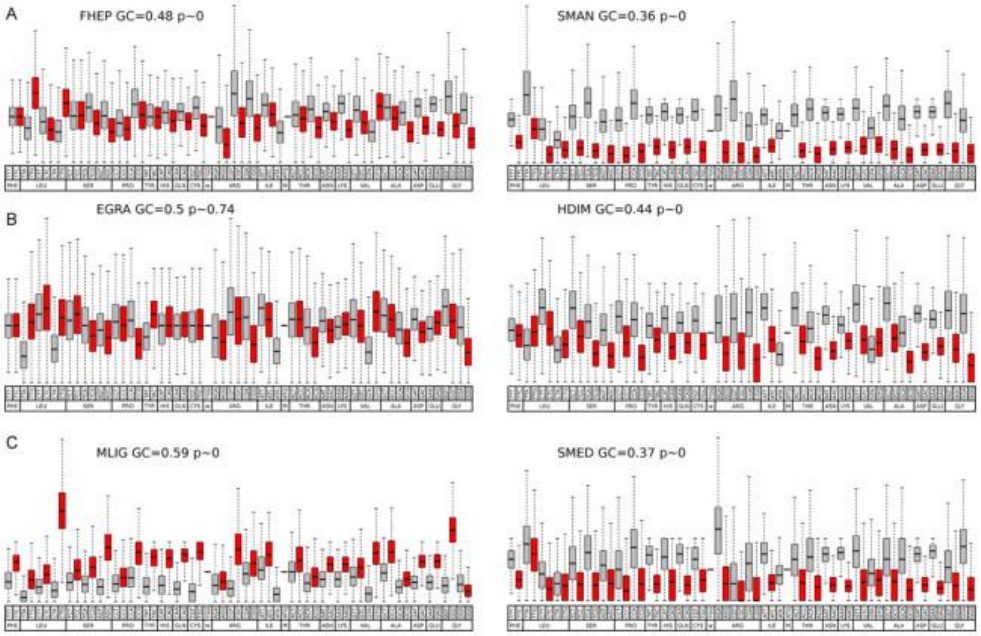


FIGURE 4 | Codon usage in representative species of flatworms. Boxplot of codon usage in six representative species of the diverse lineages and GC content across flatworms. **(A)** comparison of the trematodes *F. hepatica* and *S. mansoni*; **(B)** comparison of the cestodes *E. granulosus* and *H. diminuta*; **(C)** comparison of the free-living *M. lignano* and *S. mediterranea*. GC- and AT-ended codons are grey and red coded for ease of visualization. Mean GC values and p values are indicated.

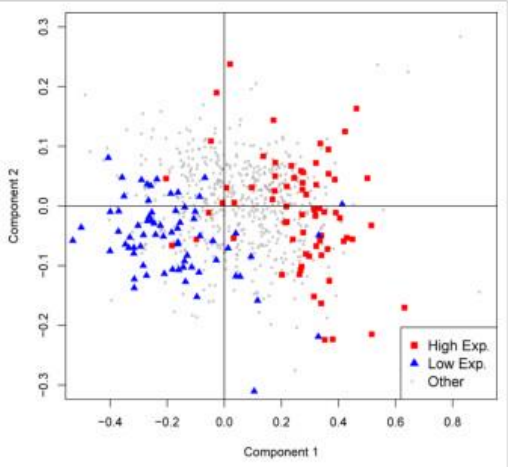


FIGURE 5 | GC variation and expression level in *S. mansoni*. The first two axes of a principal component analysis on codon frequencies are plotted. Genes with expression levels above the 90th percentile or below the 10th percentile are colored red and blue, respectively.

paired species. For this, we selected the ungapped regions of each pairwise alignment of orthologues and generated substitution matrixes based on the aligned orthologous positions. Amino acid conservation was generally high, with tryptophan (Trp) and glycine (Gly) as the more conserved residues in all the species, confirming Ile, Ser, Ala, Asn, and Arg as the more variable (**Supplementary Table 2**). Expectedly, the most frequent changes involved amino acids with similar properties (**Figure 7A**), particularly those involving aliphatic and hydrophilic residues. However, several reciprocal changes showed significant differences in the counts (**Figure 7B**). Similar effects can be observed in the other pairwise comparisons (**Supplementary Figure 5**).

To gain further insights into these phenomena, we evaluated the substitutions at the codon level generating substitutions matrices for the three pair of species selected (**Supplementary Table 3**).

The conservation at codon level as expected was much lower with a strong component of synonymous changes (**Supplementary Table 3**). The lower frequency of GC-rich codons in *S. mansoni* (depicted in **Figure 4**) is explained by a marked increase of synonymous substitutions toward AT-rich codons (**Figure 8**). Similarly, in *M. lignano*, synonymous substitutions toward GC-rich codons are associated with reduced AT codon counts (**Supplementary Figure 6**).

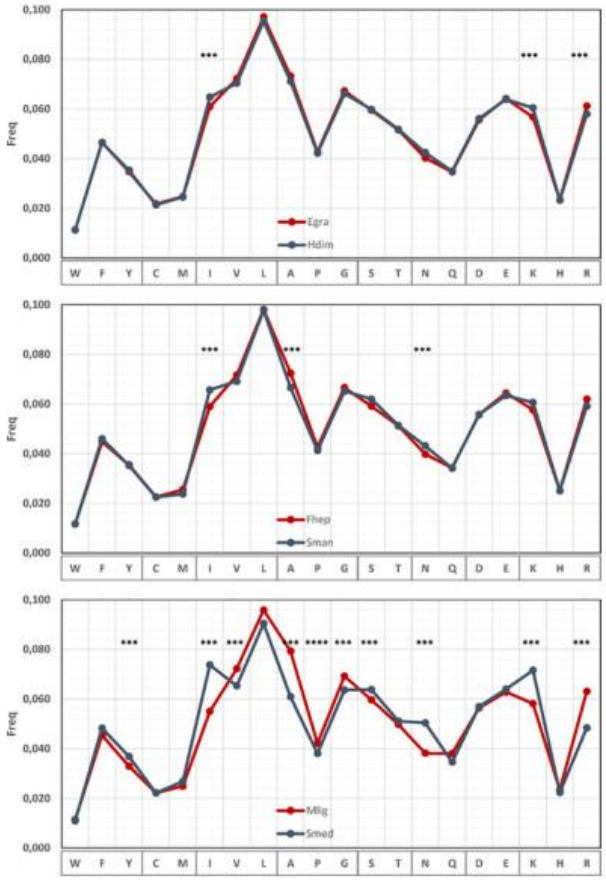


FIGURE 6 | Amino acid usage differences within flatworms. Amino acid usage in a set or 700 orthologous genes compared in pairs of species with different global GC. The amino acids showing significant differences are indicated.

The analysis of the non-synonymous changes at the codon level showed an increased complexity (**Supplementary Table 3**). One striking feature is that amino acid changes involving two substitutions are more common than those explained by simple substitutions. A detailed example is presented in **Figure 9**.

Ala is a relatively GC-rich codon (GCN) that is frequently substituted by the more GC neutral Ser (TCN + AGY) and vice versa (**Figures 7** and **9**). When this substitution takes place, it is expected that the GCN codon would change for the corresponding TCN variant, i.e., that GCA would turn into TTA and GCG into TCG. In 49 positions in the alignment, a GCG coding Ala is present in *F. hepatica*, while Ser codons are present in *S. mansoni* (second row). The simple transversion GCG to TCG is underrepresented with only six occurrences, while the changes toward TCA and TCT are more abundant (23 and 15 occurrences, respectively). Similarly, the GCC to TCC transversion (third row) represents only a 15% of the Ala (GCC) changes to Ser, while the more AT-rich variants (TCA and TCT) represent more than 56% of the substitutions. Notably, when the AT-ending Ala codons (GCA and GCT, 1st and 4th rows) are substituted, the more common codon is the one expected by a single 3rd position change.

The reciprocal Ser (in *F. hepatica*) to Ala (in *S. mansoni*) changes are less common (169 times vs 263 Ala to Ser), but again, in 77% of the cases, they are enforced to 3rd position T or A irrespective of the original Ser codon. Similar effects can be seen when analyzing other amino acid changes (**Supplementary Table 3**) and particularly in those regarded as significant (from **Figure 7**) as Ile to Val or Lys to Arg (**Supplementary Figure 7**).

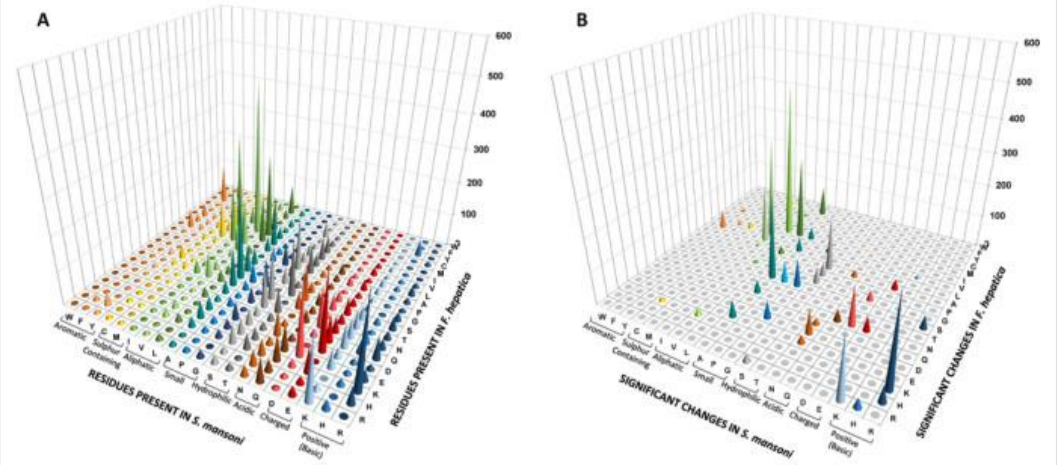


FIGURE 7 | Non-synonymous amino acid changes between *F. hepatica* and *S. mansoni*. **(A)** Graphical representation of a substitution matrix of amino acids present at orthologous positions in alignments. The residues are indicated. Amino acids present on *S. mansoni* are ordered and color coded by property. **(B)** Non-synonymous reciprocal changes that show significant differences.

Taken together, these results are strongly suggestive of a combined effect of mutation and selection in order to maintain both the compositional GC skew of the species and the property of the coded amino acid. In other words, whenever an amino acid change occurs through a simple substitution, this is then rapidly switched to those that follow the GC of the species.

DISCUSSION

Platyhelminthes classes show a wide range of GC composition, even within groups. Our results show that GC3 content explains most of the observed variability in the codon usage as reflected by the variation in the RSCU values. Based on GC3 variability, we found different clusters within the free-living species, trematodes, and cestodes. This can be clearly seen when the species tree is compared with the tree representing codon usage similarity. Indeed, Platyhelminthes show great differences between both trees, while this phenomenon is not seen in other models as different as bacteria and hexapoda (Behura and Severson, 2012; Dilucca et al., 2018). A similar study in nematodes show a comparably wide distribution of GC values, although the variations are more consistent with the phylogeny (Cutter et al., 2006). These results suggest more recent and strong compositional shifts for these groups of organisms. Further work is needed to explain this particular phenomenon in flatworms.

Codon usage bias is a generalized feature of the genomes of many organisms that is deeply influenced by evolutionary phenomena and results basically from the balance between mutational bias and natural selection (see Plotkin and Kudla, 2011, for a review). To assert the relative influence of these two factors, a plot of GC1–2 vs GC3 for all the species taken together was generated. This "pseudo" neutrality plot shows a slight slope, indicating that GC3 shows a different behavior when compared with GC1–2, a result that is generally taken as evidence that selection is acting to shape codon usage (Sueoka, 1988). Furthermore, this plot shows clearly distinctive clusters within trematodes, cestodes, and free-living species based on GC content. Notably, these differences seem to blurrily reflect the diversity of lifestyles and niches of the diverse flatworms. The observed differences question the use of single species as a model for each class; a clear demonstration of this is the differences between the model *S. mansoni* and other trematodes observed in the boxplot of **Figure 4**. Codon usage bias in flatworm mitochondrial genomes has also been reported (Le et al., 2004; Mazumder et al., 2018). Even though considering genic and genomic large differences, they may follow evolutionary pressures independent from the nuclear genomes.

Interestingly, a similar study across nematodes found robust evidence for selection on codon usage bias in free-living species, a feature found marginally in parasitic ones, and particularly in the most compositionally biased (Cutter et al., 2006). The association of selective bias in free-living or parasitic species is not clear-cut in the case of flatworms, which might be reflecting diverse evolutionary strategies.

In agreement with the hypothesis of translational selection driving synonymous codon usage bias, we observe a clear association of gene expression levels with codon usage where highly expressed genes are rich in GC-rich codons, while the opposite is observed for low-expression genes. Similar results have been previously reported for cestodes, among others (Chen et al., 2013; Yang et al., 2014; Yang et al., 2015; Huang et al., 2017; Maldonado et al., 2018). Even in highly AT-biased genomes—as observed for the schistosomes—the GC content

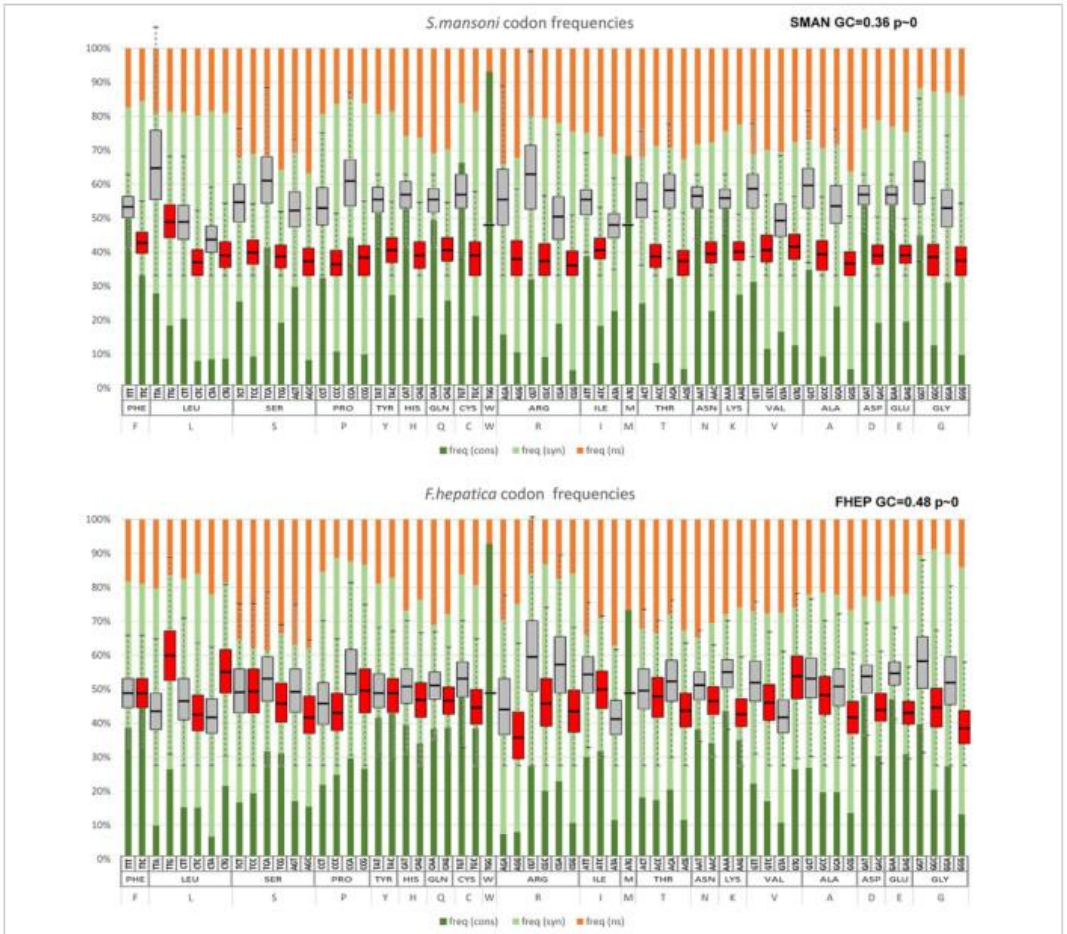


FIGURE 8 | Codon changes between *S. mansoni* and *F. hepatica*. The substitution state for each codon in pairwise alignments is represented in bars, indicating conserved codons (dark green), synonymous substitutions (pale green), and non-synonymous changes (orange). Note the enrichment of synonymous changes in GC3-poor codons in *S. mansoni* (top) in comparison with *F. hepatica* (bottom). For comparison, the figure is overlaid with the codon frequencies of **Figure 3**.

of highly expressed genes is relatively high when compared with that of the general trend. It is worth to mention that the bias in repair mechanisms of actively transcribed DNA has also being proposed to explain this observation.

The observed differences in CG and codon usage among these organisms are also reflected in the amino acid composition. Recently, Li et al. (2015) show the strong relationship of synonymous codon usage and differential amino acid usage, using a strategy based on classifying amino acids in three groups (high, medium, and low GC content) according to the GC composition of their corresponding codons. Our results on amino acid frequencies in the different flatworm species are consistent with these observations.

Furthermore, when orthologues positions are considered, we mainly observed amino acid substitutions conservative of the physicochemical properties as would be expected. However, these changes frequently involve codons of completely different GC content that follow the differences observed in general GC content of the genomes, i.e., Ile (AUH) vs Leu (CCN, UCR). In this way, AT-rich genomes accumulate changes to amino acids in the low GC group, while the opposite is observed in GC-rich genomes.

Remarkably, when the frequency of a certain amino acid substitution is not reciprocal between two given organisms, the amino acids involved belong to the different groups defined by Li et al. (2015). An interesting case is observed for the Lys to

Fhep \ Sman		A				S					
		GCA	GCG	GCC	GCT	TCA	TCG	TCC	TCT	AGC	AGT
A	GCA	204	45	96	261	21	9	11	18	1	9
	GCG	124	32	58	167	23	6	3	15	1	1
	GCC	205	31	86	318	25	10	11	17	3	8
	GCT	269	62	101	396	21	8	8	25	3	6
S	TCA	9	3	2	10	140	31	41	98	4	9
	TCG	12	2	4	17	91	34	24	85	3	18
	TCC	14	4	5	19	132	33	50	134	7	7
	TCT	15	4	6	17	117	33	38	122	3	9
	AGC	0	3	1	7	19	1	2	10	81	155
	AGT	4	2	2	7	13	3	4	18	74	215

FIGURE 9 | Non-synonymous changes between *S. mansoni* and *F. hepatica*. Detail of the substitution matrix of **Supplementary Table 3** of the changes involving Ala and Ser (*S. mansoni* codons in columns, *F. hepatica* in rows). Note the lower-than-expected counts of changes toward GC3-rich codons and the enrichment in synonymous substitutions that imply two substitutions.

Arg substitution. Even though these amino acids have similar physicochemical properties, they belong to opposite groups, being Lys coded by the most AT-rich group of codons, while Arg is on the highest GC content side.

A detailed analysis of the non-synonymous changes showed a higher-than-expected frequency of codon changes involving two nucleotides. This is paralleled by a marked reduction in the counts of the expected codon substitutions involving a single change. A plausible explanation for this phenomenon is offered by a two-hit mechanism, providing a clear example of the combined effect of mutation and selection. The two-hit hypothesis proposed implies that when a mutation changes the coded amino acid, this non-synonymous substitution is rapidly adapted to the general GC content of the genome by a second synonymous change.

CONCLUSIONS

GC bias has a great influence on synonymous codon and amino acid usage across Platyhelminthes, a feature not shared by all metazoans. Both free-living and parasitic species show the phenomena, and no clear correlation with lifestyles or evolutionary closeness is evident so far. The changes introduced by GC bias impact not only in synonymous codon usage but also in amino acid frequencies. The evidence so far suggests that both mutation and selection are acting to shape the coding strategies of the diverse flatworms.

DATA AVAILABILITY

All datasets generated for this study are included in the manuscript/supplementary files.

AUTHOR CONTRIBUTIONS

GL and SF performed the bioinformatics analysis and contributed in writing the manuscript. GR performed bioinformatics analysis. PS and JT participated in the design of the study and the interpretation of data, drafting the manuscript, and critical revision of its content. All authors read and approved the final manuscript.

FUNDING

Comisión Sectorial de Investigación Científica-Universidad de la República (CSIC-UdelaR), Uruguay. Award number: I+D16-516. Specific budget for publication was included in the project.

SF, GL, PS, and JT are researchers of Programa de Desarrollo de las Ciencias Básicas (PEDECIBA) and members of the SNI program of the Agencia Nacional de Investigación e Innovación (ANII) program. GR received a postgraduate scholarship from ANII.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00771/full#supplementary-material.

REFERENCES

Alvarez, F., Garat, B., Musto, H., Picon, M., and Ehrlich, R. (1993). Tendencies in echinococcus sp. codon usage. *Mem. Inst. Oswaldo Cruz* 88, 345–346. doi: 10.1590/S0074-02761993000200029

Behura, S. K., and Severson, D. W. (2012). Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS One* 7 (8), e43111.

Behura, S. K., and Severson, D. W. (2013). Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biol. Rev.* 88 (1), 49–61. doi: 10.1111/j.1469-185X.2012.00242.x

Bernardi, G., and Bernardi, G. (1985). Codon usage and genome composition. *J. Mol. Evol.* 22, 363–365. doi: 10.1007/BF02115693

Caira, J. N., and Littlewood, D. T. J. (2013). "Worms, platyhelminthes," in *Encyclopedia of Biodiversity: Second Edition* (USA, MA: Academic Press) 437–469. doi: 10.1016/B978-0-12-384719-5.00166-0

Charif, D., and Lobry, J. R. (2007). "SeqinR 1.0-2: a contributed package to the r project for statistical computing devoted to biological sequences retrieval and analysis," in *Structural approaches to sequence evolution: Molecules, networks, populations*. Eds. U. Bastolla, M. Porto, H. E. Roman, and M. Vendruscolo, Biological and Medical Physics, Biomedical Engineering (New York: Springer Verlag), 207–232. doi: 10.1007/978-3-540-35306-5_10

Chen, L., Liu, T., Yang, D., Nong, X., Xie, Y., Fu, Y., et al. (2013). Analysis of codon usage patterns in Taenia pisiformis through annotated transcriptome data. *Biochem. Biophys. Res. Commun.* 430, 1344–1348. doi: 10.1016/j.bbrc.2012.12.078

Coghlan, A., Tyagi, R., Cotton, J. A., Holroyd, N., Rosa, B. A., Tsai, I. J., et al. (2019). Comparative genomics of the major parasitic worms. *Nat. Genet.* 51, 163–174. doi: 10.1038/s41588-018-0262-1

Collins, J. J., 3rd (2017). Platyhelminthes. *Curr. Biol.* 27 (7), R252–R256. doi: 10.1016/j.cub.2017.02.016

Cutter, A. D., Wasmuth, J. D., and Blaxter, M. L. (2006). The evolution of biased codon and amino acid usage in nematode genomes. *Mol. Biol. Evol.* 23, 2303–2315. doi: 10.1093/molbev/msl097

Dilucca, M., Cimini, G., and Giansanti, A. (2018). Essentiality, conservation, evolutionary pressure and codon bias in bacterial genomes. *Gene.* 663, 178–188. doi: 10.1016/j.gene.2018.04.017

Egger, B., Lapraz, F., Tomiczek, B., Müller, S., Dessimoz, C., Girstmair, J., et al. (2015). A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. *Curr. Biol.* 25, 1347–1353. doi: 10.1016/j.cub.2015.03.034

Ellis, J., Morrison, D. A., and Kalinna, B. (1995). Comparison of the patterns of codon usage and bias between Brugia, Echinococcus, Onchocerca and Schistosoma species. *Parasitol. Res.* 81, 388–393. doi: 10.1007/BF00931499

Ellis, J. T., and Morrison, D. A. (1995). Schistosoma mansoni: patterns of codon usage and bias. *Parasitology* 110, 53–60. doi: 10.1017/S003118200008104X

Fernandez, V., Zavala, A., and Musto, H. (2001). Evidence for translational selection in codon usage in Echinococcus spp. *Parasitology* 123, 203–209. doi: 10.1017/S0031182001008150

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P., and Berriman, M. (2017). WormBase ParaSite – a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* 215, 2–10. doi: 10.1016/j.molbiopara.2016.11.005

Huang, X., Xu, J., Chen, L., Wang, Y., Gu, X., Peng, X., et al. (2017). Analysis of transcriptome data reveals multifactor constraint on codon usage in Taenia multiceps. *BMC Genomics* 18, 1–13. doi: 10.1186/s12864-017-3704-8

Kalinna, B. H., and McManus, D. P. (1994). Codon usage in echinococcus. *Exp. Parasitol.* 79, 72–76. doi: 10.1006/expr.1994.1063

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Lamolle, G., Protasio, A. V., Iriarte, A., Jara, E., Simón, D., and Musto, H. (2016). An isochore-like structure in the genome of the flatworm schistosoma mansoni. *Genome Biol. Evol.* 8, 2312–2318. doi: 10.1093/gbe/evw170

Larsson, K., and Jondelius, U. (2008). Phylogeny of Catenulida and support for Platyhelminthes. *Org. Divers. Evol.* 8, 378–387. doi: 10.1016/j.ode.2008.09.002

Laumer, C. E., and Giribet, G. (2014). Inclusive taxon sampling suggests a single, stepwise origin of ectolecithality in Platyhelminthes. *Biol. J. Linn. Soc.* 111, 570–588. doi: 10.1111/bij.12236

Laumer, C. E., Hejnol, A., and Giribet, G. (2015). Nuclear genomic signals of the "microturbellarian" roots of platyhelminth evolutionary innovation. *Elife* 4, 1–31. doi: 10.7554/eLife.05503

Le, T. H., McManus, D. P., and Blair, D. (2004). Codon usage and bias in mitochondrial genomes of parasitic platyhelminthes. *Korean J. Parasitol.* 42 (4), 159–167. doi: 10.3347/kjp.2004.42.4.159

Li, J., Zhou, J., Wu, Y., Yang, S., and Tian, D. (2015). GC-content of synonymous codons profoundly influences amino acid usage. *G3 Genes Genomes Genet.* 5, 2027–2036. doi: 10.1534/g3.115.019877

Maldonado, L. L., Stegmayer, G., Milone, D. H., Oliveira, G., Rosenzvit, M., and Kamenetzky, L. (2018). Whole genome analysis of codon usage in Echinococcus. *Mol. Biochem. Parasitol.* 225:54-66. doi: 10.1016/j.molbiopara.2018.08.001

Mazumder, G. A., Uddin, A., and Chakraborty, S. (2017a). Comparative analysis of codon usage pattern and its influencing factors in Schistosoma japonicum and Ascaris suum. *Acta Parasitol.* 62, 748–761. doi: 10.1515/ap-2017-0090

Mazumder, G. A., Uddin, A., and Chakraborty, S. (2017b). Expression levels and codon usage patterns in nuclear genes of the filarial nematode Wucheraria bancrofti and the blood fluke Schistosoma haematobium. *J. Helminthol.* 91, 72–79. doi: 10.1017/S0022149X16000092

Mazumder, G. A., Uddin, A., and Chakraborty, S. (2018). Codon usage pattern of complex III gene of respiratory chain among platyhelminths. *Infect. Genet. Evol.* 57, 128–137. doi: 10.1016/j.meegid.2017.10.018

Meadows, H. M., and Simpson, A. J. G. (1989). Codon usage in Schistosoma. *Mol. Biochem. Parasitol.* 36, 291–293. doi: 10.1016/0166-6851(89)90178-3

Milho, J. L., and Tracy, J. W. (1995). Updated codon usage in schistosoma. *Exp. Parasitol.* 80, 353–356. doi: 10.1006/expr.1995.1046

Mitreva, M., Wendl, M. C., Martin, J., Wylie, T., Yin, Y., Larson, A., et al. (2006).Codon usage patterns in Nematoda: analysis based on over 25 million codonsin thirty-two species. *Genome Biol.* 7 (8), R75. doi: 10.1186/gb-2006-7-8-r75

Musto, H., Romero, H., and Rodríguez-Maseda, H. (1998). Heterogeneity in codon usage in the flatworm Schistosoma mansoni. *J. Mol. Evol.* 46, 159–167. doi: 10.1007/PL00006291

Palidwor, G. A., Perkins, T. J., and Xia, X. (2010). A general model of codon bias due to GC mutational bias. *PLoS One* 5 (10), e13431. doi: 10.1371/journal.pone.0013431

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

Plotkin, J. B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* 12, 32–42. doi: 10.1038/nrg2899

Protasio, A. V., Tsai, I. J., Babbage, A., Nichol, S., Hunt, M., Aslett, M. A., et al. (2012). A systematically improved high quality genome and transcriptome of the human blood fluke schistosoma mansoni. *PLoS Negl. Trop. Dis.* 6 (1), e1455. doi: 10.1371/journal.pntd.0001455

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Sharp, P. M., Emery, L. R., and Zeng, K. (2010). Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 365, 1203–1212. doi: 10.1098/rstb.2009.0305

Sueoka, N. (1988). Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci.* 85, 2653–2657. doi: 10.1073/pnas.85.8.2653

Wei, T., and Simko, V. (2017). R package "corrplot": visualization of a correlation matrix (version 0.84). Available at: https://github.com/taiyun/corrplot.

Wright, F. (1990). The effective number of codons used in a gene. *Gene.* 87, 23–29. doi: 10.1016/0378-1119(90)90491-9

Yang, X., Luo, X., and Cai, X. (2014). Analysis of codon usage pattern in Taenia saginata based on a transcriptome dataset. *Parasit. Vectors* 7, 1–11. doi: 10.1186/s13071-014-0527-1

Yang, X., Ma, X., Luo, X., Ling, H., Zhang, X., and Cai, X. (2015). Codon usage bias and determining forces in Taenia solium genome. *Korean J. Parasitol.* 53, 689–697. doi: 10.3347/kjp.2015.53.6.689

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ANEXO III

Softwares, paquetes y parámetros utilizados durante el Objetivo I

```
# Adapter trimming and quality control
trim_galore -q 20 --length 18 --fastqc infile.fq > outfile.fq

#Collapse reads

collapse_reads.pl reads.fa > collapsed_reads.fa

# Mapping

bowtie -q -v 2 -S -p 10 index infile.fq outfile.sam --al aligned_reads.fa

# Identification and quantification of miRNAs

bwa_sam_converter.pl -i mapped_reads.sam -a mapped_reads.arf

miRDeep2.pl collapsed_reads.fa genome.fa mapped_reads.arf species_mature.fa
other_mature.fa species_precursor.fa -s species_star.fa> report.log

## miRNA target predictions

#TargetScan

perl targetscan_70.pl miR_seeds.txt 3UTR_sequences.txt outfile

#PITA

perl pita_prediction.pl -utr 3UTR_sequences.fa -mir miR_mature.fa -prefix string

#Miranda

miranda miR_mature.fa 3UTR_sequences.fa -out outfile

#Script for comparison of target predictions

#!/usr/bin/python

fm = open("miranda_targets_list", 'r')
```

```
fp = open("pita_targets_list", 'r')
fg = open("tscan_targets_list", 'r')

fc = open("output", 'w')

a = list()
b = list()
d = list()
e = list()

for line in fm:
    line=line.strip("\n")
    a.append(line)

for line in fp:
    line=line.strip("\n")
    b.append(line)

for line in fg:
    line=line.strip("\n")
    d.append(line)

def compare_intersect(x, y):
    return set(x).intersection(y)

c_set=compare_intersect(a,b)
c=list(c_set)

e_set=compare_intersect(c,d)
e=list(e_set)

print(e, end="", file=fc)

fm.close()
fp.close()
fg.close()
fc.close()

#GO enrichment was performed in R

library(topGO)

# gene IDs and background annotation
```

```
file_background <- "GO_background.txt"
Nodes <- 20 # number of processes to show
Ontology <- "GOs"

# Create temp file
data <- read.csv(file_background, sep = "\t", header = TRUE, row.names =
NULL)[,(c('query', Ontology))]
file_temp <- paste0(file_background,"2")
write.table(data, file = file_temp, sep = "\t", quote = FALSE, col.names = FALSE,
row.names = FALSE)

# Get background annotation
GOesByID <- readMappings(file = file_temp)
bg_genes <- names(GOesByID)

# Get gene IDs for the enrichment
file <- "targets.txt" #ID of targeted genes
tg <- read.csv(file, header=F)$V1

# Compare genes vs bg_genes
compared_genes <- factor(as.integer(bg_genes %in% tg))
names(compared_genes) <- bg_genes

# Create topGO object
GO_BP_data <- new("topGOdata", ontology = "BP", allGenes = compared_genes,
                  annot = annFUN.gene2GO, gene2GO = GOesByID)
GO_MF_data <- new("topGOdata", ontology = "MF", allGenes = compared_genes,
                  annot = annFUN.gene2GO, gene2GO = GOesByID)

GO_CC_data <- new("topGOdata", ontology = "CC", allGenes = compared_genes,
                  annot = annFUN.gene2GO, gene2GO = GOesByID)

#### Run Fisher test ####
result_BP_Fisher <- runTest(GO_BP_data, algorithm = "classic", statistic = "fisher")

result_MF_Fisher <- runTest(GO_MF_data, algorithm = "classic", statistic = "fisher")

result_CC_Fisher <- runTest(GO_CC_data, algorithm = "classic", statistic = "fisher")
# Create and print table with enrichment result

all_BP_Res <- GenTable(GO_BP_data, classicFisher = result_BP_Fisher, topNodes = Nodes)
#optionally numChar=1000 prevents trimming of GO descriptions
```

```
all_BP_Res$classicFisher <- as.numeric(all_BP_Res$classicFisher)
all_BP_Res <- all_BP_Res[all_BP_Res$classicFisher<0.01,]
all_BP_Res <- all_BP_Res[,c("GO.ID","Term","classicFisher")]
all_BP_Res
```

Softwares, paquetes y parámetros utilizados durante el Objetivo II

```
# Evaluación de la calidad de los ensamblajes

busco -c 10 -i input.fa -o busco_output -l metazoa_odb10 -m tran

# Predicción de ORFs

TransDecoder.LongOrfs -t input

# Búsqueda por BLAST

blastall -p blastp -b 1 -e 0.00001 -i input -d database -m 8 > output

# Predicción de dominios funcionales

 hmmscan --domE 0.001 --tblout tblout_seq --domtblout tblout_dom --pfamtblout pfamformat_out --cpu 20 Pfam-A.hmm input

# Agrupamiento de secuencias

cd-hit -c 0.9 -i input -o output

# Alineamiento de secuencias

mafft --thread 4 --threadtb 5 --threadit 0 --reorder --dash --originalseqonly --maxiterate 2 --retree 1 --localpair input > output

# Conversión de formato fasta a phylip

t_coffee -convert -infile=input.fas -outfile=output.phy -output=phylip

# Construcción de árboles filogenéticos

phyml -i input.phy -d aa -b -4 -m model --no_memory_check –quiet

# Construcción de matrices con conteo de residuos en cada posición del alineamiento
```

```
library("seqinr")

input=read.alignment(file = "input.fas", "fasta")

matriz <-as.matrix(input) #Convierte los alineamientos a una matriz de secuencias/posicion

perfil<-consensus(matriz,method = c("profile")) #Genera una tabla posicion/frecuencia de aa

perfil<-perfil[-c(1),] #saco la primera linea que tiene la frencuencia de los "-" del
alineamiento

write.table(perfil, file="output.csv",sep=";",dec =",") #Exporta una salida tabular
```